Ab initio studies of dopant-defect complexes in $KNbO_3$

by

E. G. Suter

(Under the Direction of Steven P. Lewis and William M. Dennis)

Abstract

In this work, we explore the effect of various dopant candidates in multiple defect centers in the nonlinear optical material potassium niobate ($KNbO_3$) by means of *ab initio* calculations. We employ Density Functional Theory (DFT) to model different transition metal defect complexes within $KNbO_3$ and explore trends in the band structures and projected densities of states to make qualitative predictions about the optical and infrared response of these systems. We also perform a detailed study of Fe as a substitutional impurity on both K and Nb sites in different local environments and ionization states and make detailed assignments of defect center levels reported in the experimental literature. Using DFT, we can explore a rather large number of dopant candidates and defect complexes identifying promising systems for deeper study. We have begun exploring the application of the GW approximation to some of these systems, using results for the pure phases of $KNbO_3$ as a baseline. While limitations in computational time and resources prevent the extensive use of GW on doped systems, we discuss these limitations and present viable alternative routes of study for extending this methodology.

Index words: Density Functional Theory, GW Approximation, Nonlinear Optics, Defects, Photorefractive Effect

Ab initio studies of dopant-defect complexes in $KNbO_3$

by

E. G. Suter

B.S., University of Georgia, 2014

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2022

Ab initio studies of dopant-defect complexes in $KNbO_3$

by

E. G. Suter

Major Professors: Steven P. Lewis and William M. Dennis

Committee: Dean R. Evans
Heinz-Bernd Schüttler

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2022

Acknowledgments

Foremost, I would like to extend my sincerest gratitude to my advising professors, Dr. Steven Lewis and Dr. William Dennis. Your guidance, knowledge, and advice have been there when I most needed them. I will never forget the myriad acts of kindness, patience, and consideration that you have both shown. Nor will I forget the deluge of puns and witticisms we've shared. I am proud of the researcher I've become under your mentorship and could not have asked for a better advising duo.

I would also like to thank the members of my committee, Dr. Heinz-Bernd Schüttler and Dr. Dean Evans. Dr. Schüttler, you have given me many useful lessons in my undergraduate career and your curiosity and keen questions made my oral exam extremely fruitful. Dean, I am incredibly grateful for the amazing opportunity of attending the COLLA and for your insights as an experimentalist which have helped me view where my work might fit in to the body of knowledge and helped shape the researcher I want to become.

I would also like to extend my thanks to the University of Georgia Department of Physics and Astronomy. The professors here have given me a breadth of knowledge I will take with me wherever I go. I also wish to profusely thank the staff of the Georgia Advanced Computing Resource Center, and in particular to Shan-Ho Tsai. Doubtless the research within this dissertation was only possible by her and the staff's herculean efforts in keeping the server racks running.

And finally, I would like to thank my family and friends. Your love and encouragement throughout this process are responsible for why I have been able to complete this task at

all. You have given me strength when I felt fearful, support when I felt uncertain, and your unerring confidence in me has pushed me to be better than I thought myself capable. It fills me with pride, joy, and wonder to have been fortunate enough to have you in my life.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Ferroelectrics represent an important class of materials whose technological use covers a broad range of applications including as capacitors, sensors, actuators, transducers, and energy harvesters. Recent discoveries of useful ferroelectric materials for possible device fabrication has reinvigorated the field [1]. Ferroelectric materials often also offer a variety of nonlinear optical applications such as harmonic generation and electrooptic modulation. These materials are characterized by the presence of a spontaneous macroscopic electric dipole moment present in the crystal. Given that this dipole moment is temperature dependent, all ferroelectric materials thus exhibit pyroelectricity as well. Similarly, the nature of the spontaneous polarization comes from the relative displacement of ionic centers within the crystal, rendering ferroelectric materials piezoelectric as well. The technological possibilities these materials open is only truly realized when considering the capacity to adjust or engineer the properties of interest. In materials science, this is largely done either through the discovery of new materials with the desired properties or, more pragmatically, by tuning the properties of known materials. It is this latter procedure that our work focuses on.

Potassium niobate (formula $KNbO_3$) is a member of the diverse family of perovskite materials, all of which have the ABX_3 stoichiometry. Perovskite materials have demonstrated numerous industrial applications including catalysis [3], superconductivity [4], colossal mag-

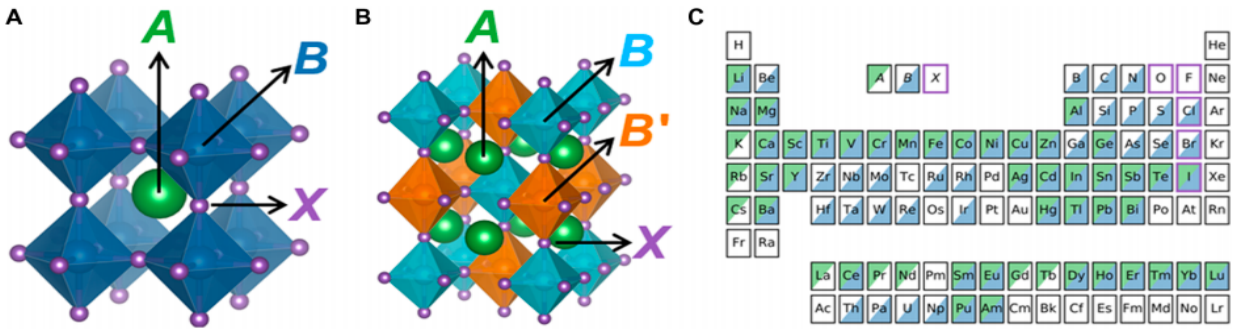

Figure 1.1: Picture demonstrating (a) The arrangement of atoms in the ideal ABX_3 cubic perovskite structure, (b) The atomic arrangement in the more complicated $A_2BB'X_6$ rock salt double perovskite structure, and (c) the many elements which can be utilized to form perovskite structures. Figure reproduced with permission from Bartel *et al.* [2]

netoresistance [5], and linear and nonlinear optical switching [6]. As demonstrated in Fig. (1.1), the incredible breadth of properties available to the perovskite family comes, in part, from their extensive compositional variety. Study of the formability of possible perovskites has centered largely around analysis of the Goldschmidt tolerance factor [7]:

$$t = \frac{1}{\sqrt{2}} \frac{(r_A + r_X)}{(r_B + r_X)} \tag{1.1}$$

Here, r_A, r_B, r_X are the ionic radii of the constituent atoms. In an ideal cubic perovskite structure, twice the B–X bond length is equal to the unit cell edge and twice the A–X bond length is equal to the face diagonal of the cubic cell. Thus, assuming a bond distance to be the sum of the two ionic radii, the Goldschmidt factor would be unity for an ideal cubic perovskite. Experimental data indicates that most ideal cubic perovskite materials occur with a t value in the range of 0.8 – 0.9 with distorted perovskites occurring in a wider range of values. Thus, the Goldschmidt factor does not represent a uniform stability criterion and other models predicting formability of perovskite compounds have been proposed [8, 9].

The ideal perovskite structure belongs to the cubic space group $Pm\bar{3}m$. This cubic structure has the A cations located at the vertices of the cubic cell (and thus 8-fold coordinated

with the X anions), the B cations located at the body center (and thus 6-fold coordinated with the X anions), and the anions (O in the case of oxide perovskites) placed at the midpoint of each of the faces of the cell. This forms an anionic octahedral complex about the B-site cation with A-site cations lying in the interstices of the octahedra. The tilting, stretching, or rotation of these BX_6 complexes, depicted in Fig. 1.1 (a) and (b), are a major structural factor in the manifestation of the possible ferroelectric and nonlinear optical properties of the crystals.

At temperatures below the Curie temperature, which varies for each perovskite species, some perovskite crystals experience structural phase transitions. These transitions can be accompanied by either asymmetrical distortions of the octahedral units or relative displacements between the anionic cages and the cations. In these cases, a spontaneous dipole moment may develop in the unit cell, rendering the crystal ferroelectric. The exact nature of these transitions remains something of an open question. A model espoused by Cochran [10] holds that certain unstable vibrational modes, or soft modes, of the crystal structure are stabilized by a transition to a structure of lower symmetry with a finite dipole moment. Other theories propose that some ions or groups of ions, disordered in the high-temperature phase, become ordered in a low-temperature ferroelectric phase [11].

$KNbO_3$ experiences a set of displacive, ferroelectric phase transitions as the crystal is cooled. These include a tetragonal phase from 500 – 710 K, an orthorhombic phase from 220 – 500 K, and a rhombohedral phase below 220 K. The structures of these phases are shown in Fig. 1.2. In going from the cubic to the tetragonal phase, a small displacement occurs, squeezing the octahedral unit along one of the cubic axes, with a relative shift in the opposite direction of the enclosed Nb ion, leading to a ferroelectric dipole along a [001] cubic axis. Upon transition to the orthorhombic state, the Nb ion displaces along the [011] direction relative to the cubic axes. This displacement leads to a doubling of the unit cell relative to the cubic structure. The transition to the rhombohedral state involves the displacement of the Nb ion along the [111] axis of the original cubic cell [12, 13]. Given that the orthorhombic

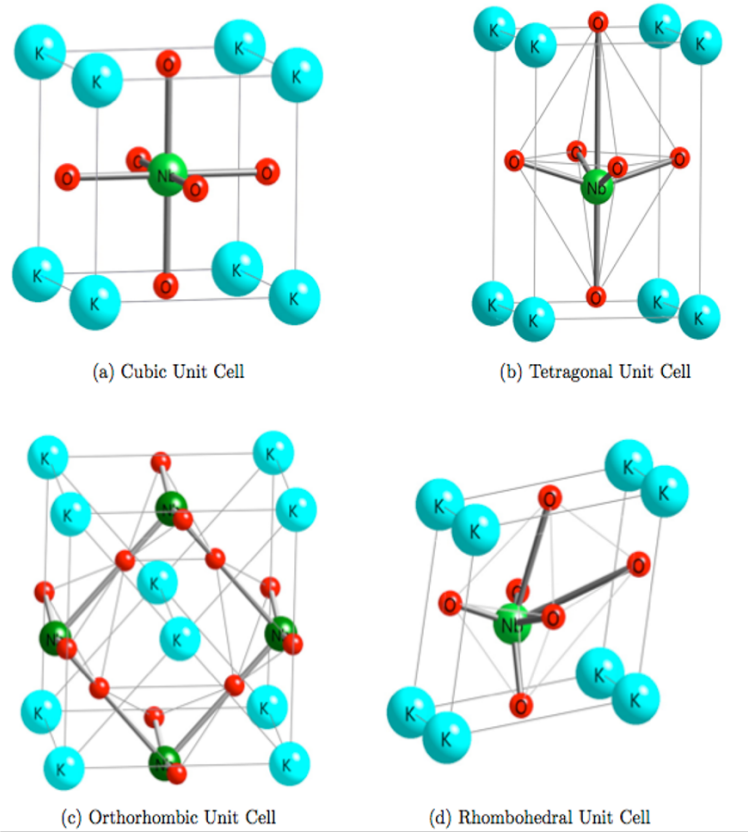

Figure 1.2: Ferroelectric phase transitions of $KNbO_3$. Note that the displacements from one phase to the next are exaggerated here for clarity.

structure is the stable phase at room temperature, it is the phase of largest practical interest and will be the focus of this dissertation.

1.1 Material parameters

We discuss and specify here some of the research concerning the material parameters of $KNbO_3$. We begin with examining how well the Generalized Gradient Approximation (GGA) under standard Kohn-Sham Density Functional Theory (DFT) approximates the structural parameters (lattice parameters and atomic displacements from equilibrium) of the phases of $KNbO_3$. In Table 1.1, structural parameters for each of the phases of $KNbO_3$ are reported.

Phase	Structural Parameters (Theoretical)	Structural Parameters (Experimental)	Percent Error
Cubic	$a = 4.026$ Å	$a = 4.000$ Å	0.0065%
Tetragonal	$a = 3.993$ Å	$a = 3.9937$ Å	0.018% in a
	$c/a = 1.055$	$c/a = 1.0165$	3.79% in c/a
Orthorhombic	$a = 3.983$ Å	$a = 3.971$ Å	0.3% in a
	$b/a = 1.454$	$b/a = 1.4347$	1.35% in b/a
	$c/a = 1.463$	$c/a = 1.4407$	1.55% in c/a
Rhombohedral	$a = b = c = 4.024$ Å	$a = b = c = 4.016$ Å	0.2% in a
	$\alpha = \beta = \gamma = 89.913$	$\alpha = \beta = \gamma = 89.818$	0.1% in α

Table 1.1: Table of lattice parameters for the pure phases of $KNbO_3$

In our calculations of these parameters, the Nb atoms are frozen at their ideal locations, while the rest of the crystal is allowed to relax both the ionic coordinates and the lattice parameters themselves.

As can be seen in the table, compared to experimental results, most parameters tend to fall within the 1 – 2% error as expected of the GGA. We find a slightly higher percent error for the c/a parameter of the tetragonal structure, which is of some interest. This may be a manifestation of the so-called supertetragonality reported by Bilc *et al.*, [14] ascribed to the exchange energy contribution of the GGA. However, all lattice parameters of practical importance (*i.e.* those of the orthorhombic structure) fall within the expected error range of Kohn-Sham DFT.

The strong dielectric and optical responses of $KNbO_3$ have been studied for quite some time. As early as 1973, Wiesendanger [15] performed a thorough investigation of the dielectric, piezoelectric, and elastic constants of single-domain $KNbO_3$ crystals and used these measurements to estimate its electrooptic coefficients. Later, Fukuda *et al.* [16] also explored these properties and Uematsu [17] in the same year measured dispersion curves and temperature dependencies of the refractive indices of orthorhombic $KNbO_3$. Later, Zgonik *et al.*

[18] used these results along with new measurements to determine the relevant parameters for electro- and acousto-optical applications. Zysset *et al.* [19] and Biaggio *et al.* [20] conducted deeper investigations of the refractive indices for the purposes of nonlinear optical applications. This body of evidence all demonstrates that $KNbO_3$ exhibits large electro-optic coefficients, photoconductivity, and high nonlinear optical susceptibilities, making it promising for many optical applications.

1.2 Photorefractive Effect

One of the more intriguing optical properties of $KNbO_3$ is known as the photorefractive effect. It is an interesting nonlinear optical phenomenon which relies on the photoconductivity and electro-optic nature of a material. We present here an abbreviated track of developments in the field of photorefractive studies. A more in-depth discussion is undertaken in the book by Solymar *et al.* [21].

The field of photorefractives traces its origin to the 1966 study by Ashkin *et al.* [22]. In that study the authors noted the optically-induced modulation of the index of refraction in samples of lithium niobate ($LiNbO_3$) and lithium tantalate ($LiTaO_3$). Refractive index modulations had also been observed that same year in potassium tantalate-niobate($KTa_xNb_{1-x}O_3$), albeit that paper focused on the different possible applications of $KTa_xNb_{1-x}O_3$. One year later, Chen [23] published a paper specifically concerning these modulations, which in the case of potassium tantalate-niobate occur only in the presence of a laser and a DC electric field. That paper noted that the photorefractive response, which was then still unlabeled, distorts the shape and polarization of the laser beam and can prohibit second-harmonic generation. Chen termed the effect "optical damage", given its deleterious effects on the other nonlinear optical applications of the material. It was also in this paper that Chen noted that those observations could be explained by the generation of an electric field produced by

charge accumulation around the impinging beam and proposed what would turn out to be a qualitative description of the mechanism of the photorefractive effect.

Numerous applications for the "damage" were conceived of rapidly [24, 25], and new photorefractive materials discovered very quickly thereafter: $Bi_4Ti_3O_3$ [26] in 1968, $Sr_{0.75}Ba_{0.25}Nb_2O_3$ [27] in 1969, and $BaTiO_3$ [28] in 1970. Huignard and Micheron [29] discovered the photorefractive effect in the materials $Bi_{12}SiO_{20}$ and $Bi_{12}GeO_{20}$ in 1976, and in the same year, Günter *et al.* [30] introduced $KNbO_3$. Much later Glass *et al.* [31] and Klein [32] discovered the effect in GaAs and InP.

The early and middle 1970s also saw growth of the understanding of the mechanisms of the effect, with Amodei [33] determining that diffusion can be responsible for the effects in $LiNbO_3$, albeit the model employed was not able to describe the maximum achievable space charge field. A year later Amodei and Staebler [34] investigated the significance of the diffusion and applied fields and their respective phases. A further development on the theoretical front was made by Glass *et al.* [35] who explained the asymmetric photocurrents in lithium niobate as the result of local asymmetries of Nb-Fe^{2+} distances. Many more developments were made throughout the middle and late 1970s, but the equations of the formalism were firmly established in a pair of papers that only gradually made their way out of Kyiv. The two papers by Kukhtarev *et al.* [36, 37] are, by and large, considered the best theoretical foundation for the field. Discussion of the effect is left to the next section where we examine in broad strokes the actual dynamics of the photorefractive effect.

1.2.1 Dynamics

The photorefractive effect relies on a combination of a material's photoconductive nature along with exhibition of the nonlinear Pockels effect. The underpinning process involved is the continual photo-excitation, migration, and eventual trapping of charge carriers. When two beams of appropriate wavelengths are made to interfere within a photorefractive crystal, photo-excitation will occur preferentially at the bright fringes of the interference pattern.

Depending on the occupation of the states within a given crystal, electrons or electron holes may be excited, which we then use to describe those states as either donor or acceptor states. For our purposes in this section we shall assume only electrons as carriers, with the knowledge that there is some symmetry between electron and hole behavior, though the details are necessarily more complicated than just a change of sign.

As the sinusoidal interference pattern is established, electrons excited from donor states in the material create a spatial distribution of excited carriers which matches the interference pattern. This will generate a carrier concentration gradient along which excited carriers will preferentially diffuse from the bright fringes to the dark fringes of the interference pattern, leaving behind ionized donor centers.

These electrons will travel some mean distance according to their lifetime within the material before recombining with the valence band or acceptor states from which they can be re-excited and travel once more. At steady state, the distribution of photo-excited electrons (being trapped at chemical centers in the dark fringes) and the ionized donors, will generate a spatially modulated distribution of positive and negative charge within the bulk of the crystal, aptly termed a space-charge field. This space-charge field will be $\frac{\pi}{2}$ radians out of phase with the interference pattern. The electro-optic nature of the material will locally alter its refractive index in the presence of a local electric field. The space-charge field will thus induce a spatially modulated refractive index hologram, which will in turn also be out of phase with the initial interference pattern. The information contained in the interference of the two beams is thus frozen into the crystal, and can be read out with a separate beam. This process is depicted graphically in Fig. 1.3.

The ability to write and read volume holograms into a crystal has obvious applications as non-volatile memory storage. Depending on the dark conductivity of the crystal, holograms can last from milliseconds to months without significant decay. These holograms can also easily be erased by uniform light exposure. However, many applications exist beyond hologram storage [38]. As the grating builds up, the interfering lasers can themselves diffract off

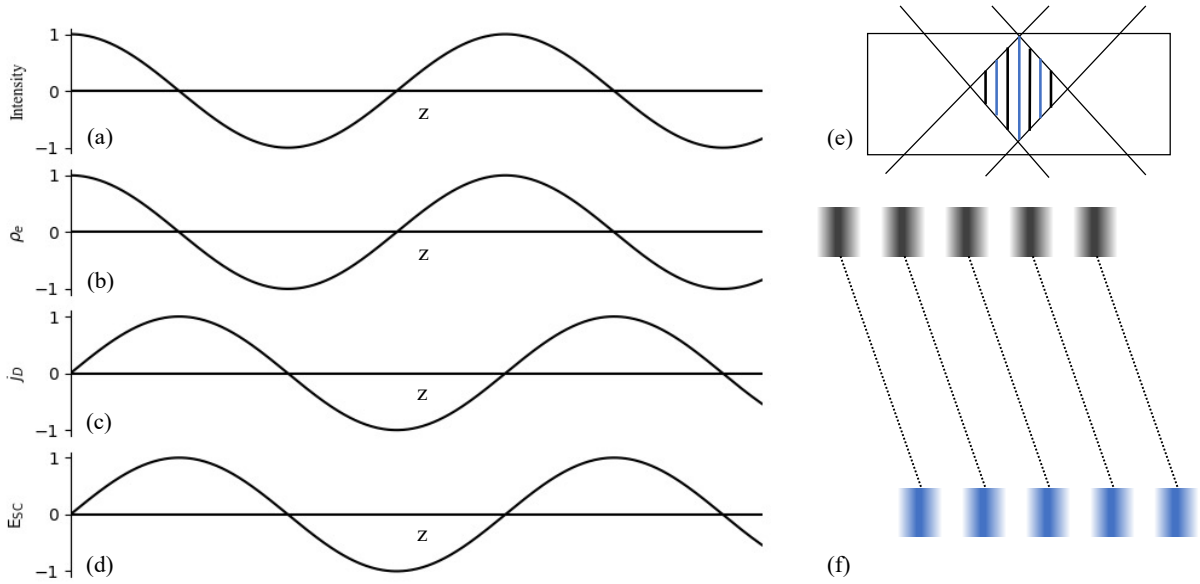

Figure 1.3: Demonstration of the photorefractive process. Interference of two beams within the bulk of a crystal (schematically shown in (e)) generates a sinusoidal intensity pattern along a dimension of the crystal, arbitrarily chosen as the z axis here (shown in (a)). Electrons in the material are preferentially excited at the constructive interference fringes, leaving behind positively charged donor ions, and creating a sinusoidal electron density pattern (b). The photo-excited electrons, now mobile in the crystal, migrate to areas of lower potential. This diffusion current is related to the gradient of the electron density, and thus the current follows a cosine pattern (c). Upon migration, the electrons, now trapped on acceptor sites or chemical features, produce a space-charge electric field which is phased-shifted with respect to the initial interference pattern (d). This space-charge field then induces a change in the material's refractive index via the Pockels effect (f). This refractive index modulation, also phase-shifted from the interference pattern, is frozen into the crystal as long as the electrons remain trapped.

of the hologram which can result in the transfer of power from one of the beams to the other, depending on the sign of the charge carrier. This is a process known as two-beam coupling. Photorefractive materials can be used to make phase-conjugating mirrors [39] which can correct for aberrations along a beam path, given that the distortion is relatively constant. The unique nature of the grating buildup can also be exploited to pick out only features with some time dependent change associated to them, generating so-called novelty filters which

track only changes to an image rather than an image itself. The various applications of this process have distinct demands, however. For example, some applications may require short build-up times for photorefractive gratings while others may require long grating lifetimes. Thus, levers of control for the various photorefractive figures of merit are desirable when considering practical device manufacture.

1.2.2 Important parameters

A key factor in the evolution of the photorefractive effect is the charge generation rate, given by:

$$g \;=\; (\beta_t + sI)(N_D - N_{D^+}) \;=\; \frac{\eta_q \alpha I}{\hbar\omega} \tag{1.2}$$

In this expression, β_t is the rate of thermal generation of charge carriers, s is the excitation cross-section, η_q is the quantum efficiency, α is the absorption constant, I is the intensity of the optical field, and N_D/N_{D^+} are the number densities of the unionized and ionized donors, respectively. Extending from this, the key equation governs the time evolution of the ionized donor concentration, where for simplicity, we will ignore the thermal generation rate:

$$\frac{\partial N_{D^+}}{\partial t} \;=\; sI(N_D - N_{D^+}) - \gamma_R n N_{D^+}, \tag{1.3}$$

where γ_R is the rate of recombination of the charge carrier and n is the charge carrier concentration. The solutions to Eq. (1.3) are presented in terms of combinations of the primary electric fields governing the behavior of the photorefractive effect which are given by:

$$E_D = \frac{k_B T K}{e},\; E_M = \frac{\gamma_R N_{A^-}}{\mu K} = \frac{\Lambda}{2\pi\mu\tau_e},\; E_q = \left(\frac{e N_{A^-}}{\epsilon_s K}\right)\left(1 - \frac{N_{A^-}}{N_D}\right) \tag{1.4}$$

where, $K = 2\pi/\Lambda$ is the wave vector of the grating, e is the electronic charge, k_B is the Boltzmann constant, T the temperature, γ_R is the electron recombination rate, μ is the

electron mobility, τ_e is the electron lifetime, ϵ_s is the static dielectric constant of the material, and N_{A^-} and N_D are the concentrations of ionized acceptor states and unionized donor states, respectively.

The field E_D is referred to as the diffusion field; E_M is known as the drift field, the field which pushes a charge carrier through a grating wavelength in the charge carrier's lifetime; and E_q is known as the saturation field, describing the largest space-charge field which can be achieved. Depending on the properties of the crystal, a photovoltaic field may also be present, with a contributing effect to the migration of the charge carriers.

Eq. 1.3 can be linearized and solved in various limits, but more in-depth discussion goes beyond our scope here. The purpose of this discussion is to introduce the many parameters which can be altered or tuned to affect the various photorefractive figures of merit. In particular, it can be readily seen that the trap density plays a vital role in determining the saturation field. The trap density also affects not only the likelihood of excitation, but also the likelihood of recombination. What is less apparent, but nevertheless true, is that understanding the nature of the trap states introduced by any particular defect can be difficult to determine experimentally.

1.2.3 Importance of Doping

It should be emphasized that the presence of mid-gap states, introduced by either defects in the material (intrinsic defects) or the presence of dopants (extrinsic defects), is vital to the processes which drive the photorefractive effect. While robust gratings can be generated using purely valence band to conduction band transitions [40], precise tuning of the photorefractive performance of materials in this way is limited by intrinsic material properties and can still be influenced by the presence of intrinsic defects. In particular, the high photorefractive sensitivities [18] observed in nominally pure $KNbO_3$ (band gap value of ≈ 3.3 eV [41]) requires the use of UV light in order to excite charge carriers.

Research on extending the spectral response of $KNbO_3$ into the visible and IR regimes has largely centered around experimental observations that doping of the crystal melt with transition metals has noticeable effects on the absorption spectrum of $KNbO_3$ crystals [42–45]. Many experiments have demonstrated the substantial effects that dopants can have on the ferroelectric properties of other oxide systems [46, 47]. However, the precise chemical nature of these defect complexes is poorly understood, with opinion varying on the charge state and location of the transition metal, as well as whether charge-compensating mechanisms such as vacancies are important or dominant [48]. Additionally, the inclusion of dopants often comes with trade-offs in other photorefractive figures of merit, such as grating rise time and photorefractive gain [49–52].

Chemical intuition may give some insight into which dopants could introduce useful states into the host material, but precise characterization of the nature of the resultant defect complexes may not always be straightforward and may require complex characterization techniques in order to determine key parameters. Given the vast experimental possibility space formed by the consideration of different host materials, dopants to be included, dopant concentrations, growth methods, post-growth treatments, and even co-doping procedures, cost and practicality requires experimentalists to be very discerning in the use of their resources.

In this regard, *ab initio* calculations offer a tantalizing means of exploring different possibilities quite cost-effectively, by comparison. The wide availability of large computing clusters and important algorithmic improvements has driven the monetary cost and wall-clock times of first principles calculations down dramatically. Properly executed Density Functional Theory (DFT) calculations can offer a means of examining band structures of defects with various occupation schemes, modeling different charges states that may be realized experimentally. The use of DFT methods and, to lesser extent, post-DFT methods at higher levels of theory to begin this program of theoretical exploration for enhanced photorefractives is the main focus of this dissertation.

The organization of this dissertation is as follows. Chapter 2 provides an introduction and background on the computational methods used, Density Functional Theory and the GW approximation, Chapter 3 covers a strategy for identifying promising dopant candidates using DFT, Chapter 4 presents a study of defect complexes of Fe-doped $KNbO_3$, Chapter 5 presents preliminary GW studies of pure phases of $KNbO_3$ and presents the challenges in applying this method to larger systems, and Chapter 6 presents the larger conclusions of this dissertation.

Chapter 2

Techniques and Methods

2.1 Density Functional Theory

The development of DFT follows a trajectory which extends from some of the earliest quantum chemistry formalisms which sought to calculate electron wavefunctions from first principles. The following background information on DFT is covered very fully by Jorge Kohanoff [53] and a reader interested in a deeper understanding of the DFT information presented in this chapter is directed to that text.

2.1.1 The Full Many-Body Hamiltonian

For a system of N electrons and P atomic nuclei, the full many-body Hamiltonian takes the form

$$\begin{aligned}\hat{H}_{\mathrm{MB}} = & -\sum_{i=1}^{N}\frac{\hbar^2}{2m_e}\nabla_i^2 \; + \; \frac{e^2}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{|\boldsymbol{r}_i \; - \; \boldsymbol{r}_j|} \\ & -\sum_{I=1}^{P}\frac{\hbar^2}{2M_I}\nabla_I^2 \; + \; \frac{e^2}{2}\sum_{j=1}^{P}\sum_{J\neq I}^{P}\frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} \\ & - e^2\sum_{i=1}^{N}\sum_{I=1}^{P}\frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|} \; + \; V_{\mathrm{external}}.\end{aligned} \tag{2.1}$$

In Eq. (2.1), the lower case indices refer to electrons in the system and the upper case indices refer to the nuclei of the system. The sets of coordinates $\{\boldsymbol{r}_i\}$ and $\{\boldsymbol{R}_i\}$ denote the electronic and nuclear coordinates respectively, where we implicitly include the spin coordinate σ_i or σ_I in the set of positions. Z_I refers to the nuclear charge of atom I.

The first line of Eq. (2.1) is then read as the sum of the kinetic energies of the electrons and the sum of the Coulombic electron-electron interaction. Similarly, the second line of Eq. (2.1) then refers to the kinetic energies of the nuclei and the Coulomb potential of the nucleus-nucleus interactions. The final line then consists of the Coulomb interaction of the electrons with the nuclei as well as any external potentials the system may be subjected to (*e.g.* laser fields, static electric fields, *etc.*). External potentials such as laser fields will obviously play an important role when considering things like excitation within the system, however for the purposes of this discussion the system is assumed to be under no external potential.

Eq. (2.1) demonstrates quite expressly that the study of real materials from a fully quantum mechanical description quickly runs into the problem of intractability. In principle, the Hamiltonian of Eq. (2.1) could be used to solve the time-independent Schrödinger equation

$$\hat{H}_{\mathrm{MB}}\Psi_n(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}) = \xi_n\Psi_n(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}), \tag{2.2}$$

where $\Psi_n(\{\boldsymbol{r}_i\}, \{\boldsymbol{R}_I\})$ is the full many-body wavefunction, containing all information regarding properties of the system, and ξ_n is the energy eigenvalue for the many-body state indexed by n. The full wavefunction is dependent on the set of electronic coordinates and the set of nuclear coordinates (including spin) and must be anti-symmetric under exchange of electronic degrees of freedom and either symmetric or anti-symmetric under exchange of nuclear degrees of freedom depending upon whether the nuclear spin is integer or half-integer. Eq. (2.2) depends on $3(N + P)$ coordinates, which presents an important computational bottleneck for calculating all Ψ_n when N and P are large, given that the calculation of the many-body wavefunction scales exponentially with this factor. Further complicating the calculation of the full many-body problem is the Coulombic interaction of the electrons, which couples their motions and thus the many-body wavefunction cannot be represented as a product of single-electron states (*i.e.* the problem cannot be split into independent equations). The Coulomb interaction is also too strong to be treated perturbatively.

Given this intractability, the introduction of well-controlled approximations that retain the essential physics of the many-body problem is necessary. One such means of simplifying the problem is the Born-Oppenheimer approximation [54]. Proposed by Max Born and J. Robert Oppenheimer in 1927, it proposes that one can decouple the nuclear and electronic motions. This is motivated by the masses of the nuclei being much larger than those of the electrons (a proton is $\sim$2000 times the mass of an electron). Under the assumption that they are experiencing the same forces, the kinetic energies of the nuclei become negligible in comparison to those of the electrons and can be ignored in the calculation of the wavefunctions. The kinetic energy contribution due to the nuclei will be a constant value and can be added back to find the total energy of the system. The electron dynamics occur on a much shorter time scale than the nuclear dynamics. Using the Born-Oppenheimer approximation we may represent the total many-body wavefunction as a product of the nuclear and the electronic wavefunctions

$$\Psi_n(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\},t) = \sum_n \Theta_n(\{\boldsymbol{R}_I\},t)\Phi_n(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}), \tag{2.3}$$

where Θ_n is the wavefunction describing the properties of the nuclear subsystem and Φ_n is the equivalent for the electronic system. Importantly, the dependence on the set of nuclear coordinates enters the electronic wavefunction parametrically while all of the explicit time dependence of the wavefunction resides in the nuclear wavefunction. For any evolution in the set of nuclear coordinates, $\{\boldsymbol{R}_I\}$, the electronic wave function will depend on the instantaneous set of nuclear coordinates. The Born-Oppenheimer approximation thus allows us to completely decouple the nuclear and electronic degrees of freedom.

It is worth discussing cases where the Born-Oppenheimer approximation breaks down. Perhaps the most famous example is the so-called "avoided crossing" of two energy levels. In simplest terms, the "no-level-crossing" theorem states that two energy levels linked by a perturbation do not cross as the strength of the perturbation is varied. For molecular systems, this manifests during chemical reactions where the molecule passes through a transition state, that is, a strong perturbation from one atomic configuration to another. During such avoided crossings, the coupling between the nuclear and electronic motions becomes too large to ignore, and the Born-Oppenheimer approximation is no longer valid.

However, assuming that the approximation remains valid, we can focus on the electronic wavefunction, $\Phi_n(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\})$ which satisfies the time-independent Schrödinger equation of the form:

$$\hat{h}_{\mathrm{e}}\Phi_n(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}) = \epsilon_n(\{\boldsymbol{R}_I\})\Phi_n(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}), \tag{2.4}$$

with $\epsilon_n(\{\boldsymbol{R}_I\})$ defining energies of the electronic states for a given nuclear configuration.

The electronic Hamiltonian in Eq. (2.4) is given by

$$
\begin{aligned}
\hat{h}_{\rm e} &= \hat{T} \ + \ \hat{V}_{\rm ee} \ + \ \hat{V}_{\rm ext} \\
&= \ -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 \ + \ \frac{1}{2}\sum_{i=1}^{N}\sum_{i\neq j}^{N}\frac{1}{|\boldsymbol{r}_i-\boldsymbol{r}_j|} \ - \ \sum_{i=1}^{N}\sum_{I=1}^{P}\frac{Z_I}{|\boldsymbol{r}_i-\boldsymbol{R}_I|} \\
&= \ -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 \ + \ \sum_{i<j}^{N}\frac{1}{|\boldsymbol{r}_i-\boldsymbol{r}_j|} \ + \ \sum_{i=1}^{N} v_{\rm ext}(\boldsymbol{r}_i),
\end{aligned}
\tag{2.5}
$$

where Hartree units have been adopted ($\hbar = m_{\rm e} = e = \frac{4\pi}{\epsilon_0} = 1$). The terms in Eq. (2.5) are the kinetic energy of the electrons, the electron-electron interaction and the electron-nuclear interaction respectively. In going from the second to third lines, we have adjusted the sum and coefficient in the $\hat{V}_{\rm ee}$ term to avoid double counting the Coulombic interactions of the pairs of electrons, and we have described the electron-nuclear interaction as an external potential which the electrons experience. We emphasize that this external potential is for an instantaneous, frozen-in, nuclear configuration. On the time scale of the electronic motion, the nuclear motion is completely ignored. Thus, the electronic wavefunction is assumed to follow the nuclear dynamics and track adiabatically to its ground state governed by this Hamiltonian, whose external potential is determined by the specific nuclear configuration.

In determining the many-electron wave function $\Phi_n(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\})$ we re-emphasize two issues that were briefly mentioned before. As with all fermion wavefunctions, Φ_n must be anti-symmetric with respect to exchange of any two electron coordinates. That is,

$$
\Phi_n(\boldsymbol{r}_1,\boldsymbol{r}_2,...\boldsymbol{r}_i,...\boldsymbol{r}_j,...\boldsymbol{r}_N) = -\Phi_n(\boldsymbol{r}_1,\boldsymbol{r}_2,...\boldsymbol{r}_j,...\boldsymbol{r}_i,...\boldsymbol{r}_N). \tag{2.6}
$$

This exchange effect must be obeyed by the electron wavefunctions and is therefore an essential feature that any formalism must reproduce. A more complicated effect comes from the fact that the Coulomb interaction correlates the motion of all of the electrons. It is for this reason that we cannot disentangle the many-body wavefunction as a product of single-

electron wavefunctions. In order to proceed with calculating these functions, we will require the formalisms developed by Walter Kohn and Lu Jeu Sham [55].

2.1.2 The Hohenberg-Kohn Theorems

The approximation in Eq. (2.3) allows us to focus on the electronic portion of the many-body problem. This is extremely valuable given that knowledge of the ground state electronic wavefunction gives access to numerous important properties of the system: equilibrium lattice parameters and atomic positions, relative stabilities of different structures, elastic properties, dielectric properties, phonon modes, polarizations, *etc.*

As previously mentioned, the many-body electronic wavefunction depends on the $3N$ coordinates of the electrons in the system (neglecting the implicit $3P$ coordinates of the nuclei). One essential idea in the formation of DFT is the idea that we can reframe the problem of calculating this complicated function with its many degrees of freedom as instead focused on solving for the electron *density*. The ground state electron density has instead just 3 degrees of freedom, *i.e.* the three Cartesian directions.

The electronic Hamiltonian in Eq. (2.5), along with the electronic many-body state $|\Phi\rangle$ yields an energy of the state

$$\begin{aligned} E &= \frac{\langle\Phi|\hat{h}_{\mathrm{e}}|\Phi\rangle}{\langle\Phi|\Phi\rangle} \equiv \langle\hat{h}_{\mathrm{e}}\rangle = \langle\Phi|\hat{T} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{ext}}|\Phi\rangle \\ &= \int_1\int_2 ... \int N d^3\boldsymbol{r}_1 d^3\boldsymbol{r}_2 ... d^3\boldsymbol{r}_N \Phi^*[-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i<j}^{N}\frac{1}{|\boldsymbol{r}_i-\boldsymbol{r}_j|} + \sum_{i=1}^{N} v_{\mathrm{ext}}(\boldsymbol{r}_i)]\Phi, \end{aligned} \tag{2.7}$$

recognizing that

$$\langle\Phi|\Phi\rangle = \int d^3\boldsymbol{r}\rho(\boldsymbol{r}), \tag{2.8}$$

we can identify that an electronic wavefunction can be used to define the electron density distribution, $\rho(\boldsymbol{r})$ in equation (2.8). This allows us to re-write the external potential term in (2.7) as

$$\langle\Phi|\hat{V}_{\text{ext}}|\Phi\rangle \;=\; \frac{1}{N}\int d^3r v_{\text{ext}}(\boldsymbol{r})\rho(\boldsymbol{r}), \tag{2.9}$$

where N is the number of electrons in the system.

This allows us to further re-write the total energy of the electronic system as

$$\begin{aligned} E \;&=\; \langle\hat{T}\rangle \;+\; \langle\hat{V}_{\text{ee}}\rangle \;+\; \langle\hat{V}_{\text{ext}}\rangle \\ E \;&=\; \langle\hat{T}\rangle \;+\; \langle\hat{V}_{\text{ee}}\rangle \;+\; \int d^3r v_{\text{ext}}(\boldsymbol{r})\rho(\boldsymbol{r}) \\ E \;&=\; \langle\hat{T}\rangle \;+\; \langle\hat{V}_{\text{ee}}\rangle \;+\; V_{\text{ext}}[\rho], \end{aligned} \tag{2.10}$$

where $V_{\text{ext}}[\rho]$ is a *functional* of the electronic density. If we can re-write the first two terms similarly as functionals of the ground state electronic density, then our mapping of the $3N$ degrees of freedom of the wavefunction to the 3 degrees of freedom of the electron density will be complete. This mapping of the ground-state electronic many-body wavefunction problem onto the problem of determining the ground-state electron density distribution is facilitated by two extremely powerful mathematical theorems discovered by Pierre Hohenberg and Walter Kohn in 1964 [56]. The Hohenberg-Kohn theorems form the bedrock of the formalism of DFT and are discussed here.

Hohenberg-Kohn Theorem I: *The external potential of a system of interacting particles is uniquely determined by the ground-state electron density, apart from an additive constant.*

The implication of the theorem is that the electronic Hamiltonian is completely determined (aside from a constant additive term) given a particular external potential (*i.e.* atomic configuration). Thus, a corollary to the theorem is that the many-body wavefunctions for

all states are completely determined with only the ground-state density. The proof of this theorem is given in Appendix A. Although difficult to prove in the other direction (see the V representability problem [57]), the power of this theorem lies in the fact that given a particular external potential in which the electrons reside (*i.e.* the atomic configuration of the material of interest), there is one and *only* one ground-state electron density. So the only *a priori* knowledge that is necessary to calculate the ground-state electron density, and all subsequently determined properties, is just the knowledge of the atoms involved and their geometry.

Hohenberg-Kohn Theorem II: *For any given external potential, a functional of energy in terms of the particle density can be defined. The functional attains its minimum only at the ground-state density*

The importance of the second Hohenberg-Kohn theorem comes from the corollary that since the energy of the system can *always* be written as a functional of the density, we are then capable of ensuring that we can find the ground-state density by means of a variational search for the global minimum of the functional. The proof of this theorem is also detailed in Appendix A. Its importance can be demonstrated quite straightforwardly. The theorem guarantees that we may always write the energy of a system like so:

$$\begin{aligned} E_{\mathrm{HK}}[\rho] &= \langle\Phi[\rho]|\hat{h}_{\mathrm{e}}|\Phi[\rho]\rangle \\ &= \langle\Phi[\rho]|\hat{T} + \hat{V}_{\mathrm{ee}}|\Phi[\rho]\rangle + \int d^3r v_{\mathrm{ext}}(\boldsymbol{r})\rho(\boldsymbol{r}) \\ &= F_{\mathrm{HK}}[\rho] + \int d^3r v_{\mathrm{ext}}(\boldsymbol{r}) \\ &\geq E[\rho_0] = E_0 = \langle\Phi[\rho_0]|\hat{h}_{\mathrm{e}}|\Phi[\rho_0]\rangle. \end{aligned} \tag{2.11}$$

The last line in Eq. (2.11) emphasizes that the ground-state wavefunction (and the density it subsequently determines) is *the* state which minimizes the energy functional. Perhaps the most important facet of this expression is the term $F_{\mathrm{HK}}[\rho]$. This term is a universal func-

tional, true for all systems of interacting electrons, independent of the potential in which they sit. While this universal functional exists in principle, the second Hohenberg-Kohn theorem provides no means of constructing it. It is important to note here that the operators $\hat{T}$ and $\hat{V}_{\mathrm{ee}}$ are the true, fully interacting, kinetic energy and electron-electron interaction operators, whose forms are not perfectly known. To take full advantage of these powerful theorems, Kohn and Sham provided a powerful ansatz that results in the modern incarnation of DFT as we know it today.

2.1.3 Kohn-Sham Ansatz

The central idea behind the Kohn-Sham ansatz is that the complicated many-body interactions are what prevent straightforward descriptions of the operators in the universal functional. If a system of non-interacting electrons could be generated with the same ground-state density as the real system, their description would be much easier. A key realization of Kohn and Sham in their seminal work of 1965 [55] was that the wavefunction of a system of non-interacting electrons can be easily expressed as a Slater determinant of one-electron orbitals. This non-interacting system of electrons would obey a Hamiltonian of the form

$$\hat{h}_{\mathrm{KS}} \;=\; \sum_{i=1}^{N} \left[-\frac{\hbar^2}{2m}\nabla_i^2 \;+\; v_{\mathrm{R}}(\boldsymbol{r}_i) \right], \tag{2.12}$$

with N, the total number of electrons and v_{R} being a potential designed to account for the various interactions among the electrons, as well as the interactions of the electrons with the nuclei and any other external fields. The kinetic energy here contains no effects of correlation. That is to say, it is not the true kinetic energy of the real system. We recognize that the electrons of this fictitious system interact only with this reference potential.

This Hamiltonian yields an energy functional of the form

$$E_{\mathrm{KS}}[\rho(\boldsymbol{r})] \;=\; T_{\mathrm{R}}[\rho(\boldsymbol{r})] \;+\; \int d^3r \rho(\boldsymbol{r}) v_{\mathrm{R}}(\boldsymbol{r}), \tag{2.13}$$

with a density constructed similarly to Eq. (2.9), but now with the single-electron orbitals $\phi(\boldsymbol{r})$ of the non-interacting system

$$\rho(\boldsymbol{r}) \;=\; \sum_{i=1}^{N} |\phi_i(\boldsymbol{r})|^2, \tag{2.14}$$

and the kinetic energy of equation (2.13) begin expressed as

$$T_{\mathrm{R}} \;=\; -\frac{1}{2}\sum_{i=1}^{N} \langle \phi_i | \nabla^2 | \phi_i \rangle. \tag{2.15}$$

In order for this non-interacting system to be of use, it must be ensured to have the same ground-state density as the real system. This is guaranteed by the second Hohenberg-Kohn theorem if $E_{\mathrm{KS}}[\rho] \geq E_{\mathrm{KS}}[\rho_0] = E_0$, the equality being valid only if the density is the true ground-state density, i.e. $\rho = \rho_0$. Therefore the functional derivative of $E_{\mathrm{KS}}[\rho]$ must vanish at this density. Taking the functional derivative of $E_{\mathrm{KS}}[\rho]$ under the constraint that the density gives the correct number of electrons we obtain

$$\begin{aligned} &\frac{\delta}{\delta\rho(\boldsymbol{r})}\left[E_{\mathrm{KS}}[\tilde{\rho}] \;-\; \mu_{\mathrm{R}} \int d\boldsymbol{r}\rho(\boldsymbol{r})\right]_{\rho=\rho_0} \;=\; 0, \\ &\text{resulting in the expression,} \\ &\frac{\delta T_{\mathrm{R}}}{\delta\rho(\boldsymbol{r})} \;+\; v_{\mathrm{R}} \;=\; \mu_{\mathrm{R}}, \end{aligned} \tag{2.16}$$

where the Lagrange multiplier of the constraint, μ_{R}, is interpreted as the chemical potential of the electrons in the reference system.

We now need to apply this same procedure to the energy functional of Eq. (2.11). The previously discussed, non-interacting kinetic energy T_{R} presents a useful idea: the assumption that the non-interacting part of this operator (which can be easily written down under largely sensible approximations) will capture the essential physics of the system while the part due to the complicated many-body interactions may be combined into another term

which can then be separately approximated. We can similarly approach the electron-electron interaction. The central approximation for the non-interacting part is known as the Hartree approximation. It is the electrostatic energy of a system of electrons that all interact with an average potential produced by the entire collection of them. This assumes that each electron interacts with the field without perturbing it and it takes the mathematical form

$$E_{\text{Hartree}}[\rho] \;=\; \frac{1}{2}\int\int d^3\boldsymbol{r}d^3\boldsymbol{r'}\frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|}. \tag{2.17}$$

This approximation does not account for effects arising from electron-electron repulsion, but a deeper error is that any electron interacts with the average potential that it also plays a role in producing. Any individual electron thereby unphysically interacts with itself.

By separating out the non-interacting contributions from the operators which make up the universal functional, all uncertainty about the form of the many-body interactions can be incorporated into a single term, dubbed the exchange-correlation term (which will be discussed in more detail later). The functional of Eq. (2.11) can now be written under the Kohn-Sham ansatz as

$$E[\rho] \;=\; T_{\text{R}} \;+\; E_{\text{Hartree}}[\rho] \;+\; \int d^3\boldsymbol{r}v_{\text{ext}}(\boldsymbol{r})\rho(\boldsymbol{r}) \;+\; E_{\text{XC}}[\rho]. \tag{2.18}$$

We now minimize the functional by setting its functional derivative (with respect to the electron density) to zero to obtain

$$\frac{\delta}{\delta\rho(\boldsymbol{r})}\left[E[\rho] \;-\; \mu_{\text{R}}\int d\boldsymbol{r}\rho(\boldsymbol{r})\right]_{\rho=\rho_0} \;=\; 0, \tag{2.19}$$

resulting in the expression,

$$\frac{\delta T_R}{\delta\rho(\boldsymbol{r})} \;+\; v_{\text{ext}}(\boldsymbol{r}) \;+\; \int d\boldsymbol{r'}\frac{\rho(\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|} \;+\; v_{\text{XC}} \;=\; \mu, \tag{2.20}$$

where again, μ here represents the chemical potential of the electrons, albeit now in the *real* system. We have also introduced the term $v_{\mathrm{XC}} = \frac{\delta E_{\mathrm{XC}}}{\delta \rho(\boldsymbol{r})}$ which is the potential term that captures the effects of exchange and correlation.

Asserting that no charge flow occurs between the real and auxiliary systems allows us to equate the chemical potentials in expressions Eq. (2.20) and Eq. (2.16). After very simple manipulation we get:

$$v_{\mathrm{R}} = v_{ext}(\boldsymbol{r}) + \int d\boldsymbol{r}' \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} + v_{\mathrm{XC}}, \tag{2.21}$$

which establishes the form of the reference potential for the auxiliary Kohn-Sham system. We can thus write down the total energy of a system of non-interacting electrons as

$$E[\rho] = -\frac{\hbar^2}{2m} \sum_i \langle \phi_i | \nabla_i^2 | \phi_i \rangle + \int d\boldsymbol{r} \rho(\boldsymbol{r}) v_{ext}(\boldsymbol{r}) + \frac{e^2}{2} \int \int d\boldsymbol{r} d\boldsymbol{r}' \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} + E_{\mathrm{XC}}[\rho]. \tag{2.22}$$

Importantly, to recover the true total energy of the system, the Coulombic interaction between the nuclei of the specific atomic configuration must be accounted for. However, this represents a constant, as the atomic configuration is frozen in. We are now equipped to proceed with the minimization of this energy functional to find the ground-state electron density, or rather the minimization with respect to the single-electron Kohn-Sham orbitals $\{\phi_i\}$. We re-frame this minimization problem via the Rayleigh-Ritz theorem [58], which allows us to equivalently treat it as the solution to an eigenvalue problem of the following form:

$$\left[-\frac{1}{2}\nabla^2 + v_{\mathrm{ext}}(\boldsymbol{r}) + \int d\boldsymbol{r}' \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} + v_{\mathrm{ext}}(\boldsymbol{r}) \right] \phi_i(\boldsymbol{r}) = \epsilon_i^{\mathrm{KS}} \phi_i(\boldsymbol{r}). \tag{2.23}$$

Equations (2.23) form a set of coupled, nonlinear, Schrödinger-like equations called the Kohn-Sham equations. The eigenvalues ϵ_i^{KS} are known as the Kohn-Sham eigenvalues and correspond to the Lagrange multipliers of the corresponding minimization problem. It is to

be noted that these eigenvalues should not be confused with the excitation spectrum of the real system. These are the energies of the non-interacting reference system and thus, while the ground-state electron density of the real system is accurately represented, excitation is not adequately described by a system whose particles cannot, by design, interact.

An even more important point to recognize is that the single-electron wavefunctions $\{\phi_i\}$ define the density by $\rho(\boldsymbol{r}) = \sum_i |\phi_i|^2$, and thus, referring to Eq. (2.23) we see that the equation both depends upon the density, and aims to solve for the wavefunctions which define the density. Thus the Kohn-Sham equations must be solved self-consistently. This aspect will be discussed in more detail later.

The Kohn-Sham equations represent the final incarnation of the Kohn-Sham ansatz. To summarize, the Kohn-Sham ansatz makes use of the two Hohenberg-Kohn theorems to approach the intractability of the many-body problem. One extremely simplifying insight of the Kohn-Sham ansatz is that the universal functional predicted in the second Hohenberg-Kohn theorem can be broken into non-interacting and many-body interacting contributions. Then, with suitable approximations to the exchange-correlation functional (to be discussed in the next section), a non-interacting system of electrons could be constructed which interact via an appropriate reference potential. Then by the first Hohenberg-Kohn theorem, this system and the true system will experience the same external potential and will subsequently have the same ground-state electron density. By transforming the minimization problem into a set of Schrödinger-like equations, the Kohn-Sham equations can be then be solved self-consistently to find the ground-state density of the system.

2.1.4 Exchange-Correlation Functional

If the true form of the exchange-correlation functional in Eq. (2.22) were known, Kohn-Sham theory would be an exact theory. In practice, however, the form of the exchange-correlation functional comprises the largest approximation of the Kohn-Sham process. As

demonstrated in its name, the exchange-correlation energy accounts for the two most complicated and fundamental effects of exchange and correlation of the electronic motion.

Demonstrated earlier in Eq. (2.6), the exchange phenomenon concerns the Fermionic nature of the electrons. According to the Pauli Exclusion Principle, the wavefunction for a Fermionic system must be anti-symmetric under the exchange of any pair of particle labels. This condition means that an allowed many-body wavefunction must vanish if any two Fermions have the same set of spin *and* spatial coordinates (particles can share the same spatial coordinates if they are of opposite spin states). This also restricts the space of possible wavefunctions that the Kohn-Sham formalism should be able to find. The Pauli Exclusion Principle also results in an effective repulsive force between electrons. This is because electrons of like spin are forbidden from obtaining the same spatial coordinates, meaning they spend more time, on average, separated from one another. Given that the Coulombic interaction between electrons is repulsive, this effective increase in average distance means that the population of like-spin states is favored given that they will incur less Coulombic repulsion and will subsequently be lower in energy. The energetic difference between the like-spin and opposite-spin states is commonly referred to as the "exchange energy" even though it is ultimately Coulombic in origin. The exchange energy has an exact form given orthonormal wavefunctions:

$$E_{\text{ex}} = -\frac{e^2}{2}\sum_{\sigma}\int\int d\boldsymbol{r}d\boldsymbol{r}'\frac{\delta_{\sigma\sigma'}|\sum_i \phi_i^{\sigma *}(\boldsymbol{r})\phi_i^{\sigma}(\boldsymbol{r}')|^2}{|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{2.24}$$

Calculations of a single electron system reveal that the E_{XC} is equal in magnitude to the Hartree energy, producing the extremely fortunate side effect of canceling the spurious self-interaction introduced in the Hartree approximation. However, the approximations of the Local Density Approximation (LDA), to be discussed, in general do not perfectly cancel the self-interaction.

The phenomenon of correlation is much less intuitively described. Rather than a quantum mechanical effect, correlation is essentially a classical effect. The Coulombic repulsion of

electrons with one another produces an obvious tendency of electrons to stay away from one another. In effect, each electron can be thought of as inducing a dearth of electrons in its immediate vicinity as it moves around. While an in-depth explanation of correlation is beyond the scope of this dissertation, correlation may also be thought of as a deeply complicated coupling of every electron's motion with any others with which it locally interacts.

An alternative interpretation of this term is that the presence of an electron with a particular spin-state at some point in space will cause all other like-spin electrons to interact with a so-called "exchange hole" at that point in space. We can express this as a Hartree-like term which accounts for the average interaction of all other like-spin electrons with their respective exchange-holes:

$$E_{\mathrm{ex-hole}} \;=\; \frac{e^2}{2}\int d\boldsymbol{r}\rho(\boldsymbol{r})\int d\boldsymbol{r}'\frac{\rho_{\mathrm{ex-hole}}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{2.25}$$

The effects of exchange and correlation are combined to produce a form for the E_{XC} term

$$E_{\mathrm{XC}}[\rho] \;=\; e^2\int d\boldsymbol{r}\rho(\boldsymbol{r})\epsilon_{\mathrm{XC}}([\rho],\boldsymbol{r}) \;=\; e^2\int d\boldsymbol{r}\rho(\boldsymbol{r})\int d\boldsymbol{r}'\frac{\tilde{\rho}_{\mathrm{XC}}(\boldsymbol{r},\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}, \tag{2.26}$$

with the exchange-correlation hole density, $\tilde{\rho}_{\mathrm{XC}}$, being similar in nature to the density introduced in the previous equation.

The last points to be addressed in the formalism are the actual approximations used to model the E_{XC} term. Eq. (2.26) is useful from a conceptual point of view, but provides no practical means of calculating the energy. The most straightforward energy functional used in modern solid-state calculations is the LDA. In the LDA, small volumes of the material are assumed to have approximately uniform electron density. The exchange-correlation energy of this uniform volume of electron density is then assumed to be equivalent to that of a homogeneous electron gas with the same density. That is,

$$E_{\mathrm{XC}}[\rho] \;=\; e^2\int d\boldsymbol{r}\rho(\boldsymbol{r})\epsilon_{\mathrm{XC}}^{\mathrm{HEG}}[\rho(\boldsymbol{r})]. \tag{2.27}$$

The $\epsilon_{\mathrm{XC}}^{\mathrm{HEG}}$ term has been calculated to great accuracy via quantum Monte-Carlo methods. At first glance, this approximation is quite crude. The LDA inherently models exchange and correlation as local effects, even though the true effects are non-local. However, the LDA has been shown to work quite well when compared to experiment. LDA reproduces lattice constants within about 1–2% of experimental results, while bulk moduli are too large by about 30%. These errors are found to be systematic. The approximation of the exchange-correlation energy causes the resulting charge densities to be somewhat more homogenized than might otherwise be expected, hence bond lengths are typically shorter, resulting in the discrepancy with experiment. In the limit of strongly-correlated materials it fairs less well, but for a wide variety of systems, the LDA produces remarkably useful results.

A first attempt to rectify some of the error introduced by the simplifications of the LDA is provided by the Generalized Gradient Approximation (GGA). Here, the exchange-correlation energy is calculated incorporating terms which depend on the local inhomogeneity of the density, that is, powers of the gradient of the charge density are used to parametrize the energy

$$E_{\mathrm{XC}}[\rho] \;=\; e^2 \int d\boldsymbol{r}\rho(\boldsymbol{r})\epsilon_{\mathrm{XC}}[\rho(\boldsymbol{r}), \nabla\rho(\boldsymbol{r})]. \tag{2.28}$$

While the GGA takes a logical first step towards addressing some of the short falls of the LDA, it does not provide systematically better results. The charge densities tend to be over-localized which results in bond lengths being systematically *overestimated*. However, both functionals have been shown to provide good results over a wide breadth of important systems. A more in-depth discussion of these functionals, their parameterizations, and other useful functionals is found in Appendix B.

2.1.5 Practical Implementation Details

With the formalism of DFT now firmly established, we concern ourselves in this section with the tools required to practically implement the theory in order to calculate properties of real systems.

Basis Sets: The representation of any wavefunction will require a suitable basis set onto which the components of the wavefunction will be projected. For crystal materials, the periodic nature of the potential $v(\boldsymbol{r}) = v(\boldsymbol{r}+\boldsymbol{R})$, with $\boldsymbol{R}$ being a real-space lattice vector in the Bravais lattice, implies that a natural choice of basis sets would be plane waves.

In periodic potentials, the wavefunctions of electrons take the form of Bloch states, that is, plane waves which can be broken down as the product of a lattice-periodic function and a phase factor whose value depends on the dot product of the crystal momentum of the electron and its real-space position vector.

$$\phi_{n,\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n,\boldsymbol{k}}(\boldsymbol{r}) \tag{2.29}$$

We will require the Kohn-Sham functions to take the form of Bloch states, which are characterized now by two indices, n, being the state label or band index of the electron and, $\boldsymbol{k}$ the crystal momentum vector of the electron in reciprocal space which will be a wave vector in the first Brillouin zone. Since $u_{n,\boldsymbol{k}}(\boldsymbol{r})$ has the same periodicity of the lattice, it lends itself to a natural expansion as a Fourier series. We can make a further useful observation that the Bloch functions in different cells of the crystal will be related by a phase factor,

$$\phi_{n,\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R}) = e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\phi_{n,\boldsymbol{k}}(\boldsymbol{r}), \tag{2.30}$$

ergo, the probability densities of states in different unit cells will be the same, meaning we can calculate the properties of the entire crystal simply with knowledge of a single unit cell.

The expansion of the periodic parts of the Bloch wavefunctions into Fourier series takes the form

$$u_{n,\boldsymbol{k}}(\boldsymbol{r}) \;=\; \sum_{\boldsymbol{G}} C_{n,\boldsymbol{k}}(\boldsymbol{G}) e^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{2.31}$$

where $\mathbf{G}$ represents a reciprocal-space lattice vector, defined with respect to the real-space lattice by $\boldsymbol{G}\cdot\boldsymbol{R} \;=\; 2\pi j$, with j an integer. We may use this expression to expand the full Bloch wavefunctions in this same basis like so

$$\phi_{n,\boldsymbol{k}}(\boldsymbol{r}) \;=\; \sum_{\boldsymbol{G}} C_{n,\boldsymbol{k}+\boldsymbol{G}}(\boldsymbol{G}) e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}}. \tag{2.32}$$

In principle, an infinite number of these Fourier terms is required to describe the electronic orbitals perfectly. However, the contributions from plane waves of high kinetic energies decrease exponentially. So the series expansion may be truncated by choosing a sufficiently high energetic cutoff, and only including plane waves with kinetic energy less than this chosen cutoff:

$$\left|\frac{\hbar^2(\boldsymbol{k} \;+\; \boldsymbol{G}_{\max})}{2m}\right| \;\leq\; E_{\text{cut}}. \tag{2.33}$$

Eq. (2.33) can also be thought of as describing a manifold of $\boldsymbol{G}$-vectors whose extent about a given k-point determines the number of planes waves used in the representation of the wavefunction for that k-point. Thus, different points in the Brillouin zone will have different numbers of plane waves representing them.

The cutoff allows us to represent the electron wavefunctions as precisely as necessary, with a means of systematically improving the representation by simply increasing the cutoff energy. The validity of a particular choice of E_{cut} can be established through total energy calculations of a system of interest. The cutoff energy can be increased until the total energy results are within acceptable precision. The intrinsic tradeoff between computational cost and simulational accuracy is thus tailored to the need of any particular study.

While other basis sets may be used, each comes with their own features and drawbacks. The plane wave basis is desirable for its (in principle) completeness and for the fact that it

makes the calculation of energy and its derivatives much less cumbersome than with some other basis sets.

***k*-space sampling**: Bloch's theorem, which helped us find a useful expansion for the electron wavefunctions, introduces an index which labels the crystal momenta, wave vectors (k) in the first Brillouin zone. There are, however, infinite k-points in reciprocal space. However, we need not sample every k-point in the Brillouin zone, since the wavefunctions of k-points sufficiently close to one another are very similar. Thus, we can achieve results over a representative sample of k-points and that will be capable of representing the wavefunctions across most of the Brillouin zone. One of the most famous means of generating a k-point grid which optimally samples the Brillouin zone was devised by Monkhorst and Pack [59]. In this scheme, a generic function, f, to be integrated over the Brillouin zone is expressed as

$$f(\boldsymbol{r}) \;=\; \frac{\Omega}{(2\pi)^3}\int_{\mathrm{BZ}} d\boldsymbol{k}\tilde{f}(\boldsymbol{k}) \;=\; \sum_j w_j \tilde{f}(\boldsymbol{k}_j)e^{i(\boldsymbol{k}\cdot\boldsymbol{r})}, \tag{2.34}$$

with $\tilde{f}(\boldsymbol{k})$ being the Fourier transform of f, and the BZ limit in the integral refers to some mesh of points in the Brillouin zone. The factor Ω is the volume of the real-space unit cell and the factors w_j are factors whose weight sums up to unity. However, the number of relevant k-points may be further reduced by exploiting the point-group symmetry of the lattice. With this consideration

$$f(\boldsymbol{r}) \;=\; \sum_{j=1}^{P(n_j)} w'_j \tilde{f}(\boldsymbol{k}_j), \tag{2.35}$$

where $P(n_j)$ is a number of points within the irreducible wedge of the Brillouin zone that depends on the point symmetry of the crystal. Here the modified weights now reflect the ratio of the group of the individual k-point to that of the full group. The density of the resulting k-point mesh plays a role in determining the accuracy of the reference potential in the Kohn-Sham scheme. The chosen k-mesh may also be validated by examining the convergence of the total energy of the system with increasing density.

Pseudopotentials: While modeling every electron in a system is possible, it is neither computationally efficient nor rigidly necessary. Many of the electrons in the system are core electrons, which reside very close to their parent nucleus. These electrons are largely chemically inert, playing no major direct role in chemical bonding, but they are also extremely computationally inefficient to represent in a plane-wave basis. The short wavelengths of the core electrons' wavefunctions require many Fourier terms in order to be adequately represented and doing so for every core electron in a system would be wasting resources on a rather static part of the problem.

To continue using the plane wave basis, an alternate way to represent these electrons is desired. Given that our interest lies almost exclusively in the valence electrons, we may think to combine the core electrons and the parent nucleus into an ionic core. This pseudo-nucleus must then produce a potential which interacts with the valence electrons to produce the correct all-electron behavior of the wavefunction beyond a certain cutoff radius from the atomic center. The pseudo-wavefunction then represents an eigenfunction of a pseudo-Hamiltonian which reproduces exactly the eigenvalues and behavior of the true wavefunction (beyond a certain radius) but which is nodeless and smooth within that radius. The chosen cutoff radius needs to be varied based on the angular momentum channel of a particular valence electron. This pseudo-wavefunction is then much more easily represented by fewer Fourier terms in our plane wave expansion and produces the correct behavior in the region of interest, alongside the huge computational savings of no longer needing to simulate the individual core electrons. Fig. (2.1) gives a good representation of the mechanics of pseudopotentials.

Two main methods of pseudopotential generation vary the conditions of the pseudo-wavefunction within the cutoff radius. The somewhat more standard group of norm-conserving pseudopotentials require that norm of the true and pseudo-wavefunctions be identical within the cutoff radius, leading to the boundary condition:

$$\frac{1}{R^l_{\mathrm{AE}}(\epsilon, r_{\mathrm{c}})}\left[\frac{dR^l_{\mathrm{AE}}(\epsilon, r_{\mathrm{c}})}{dr}\right]_{r_{\mathrm{c}}} = \frac{1}{R^l_{\mathrm{PS}}(\epsilon, r_{\mathrm{c}})}\left[\frac{dR^l_{\mathrm{PS}}(\epsilon, r_{\mathrm{c}})}{dr}\right]_{r_{\mathrm{c}}}, \tag{2.36}$$

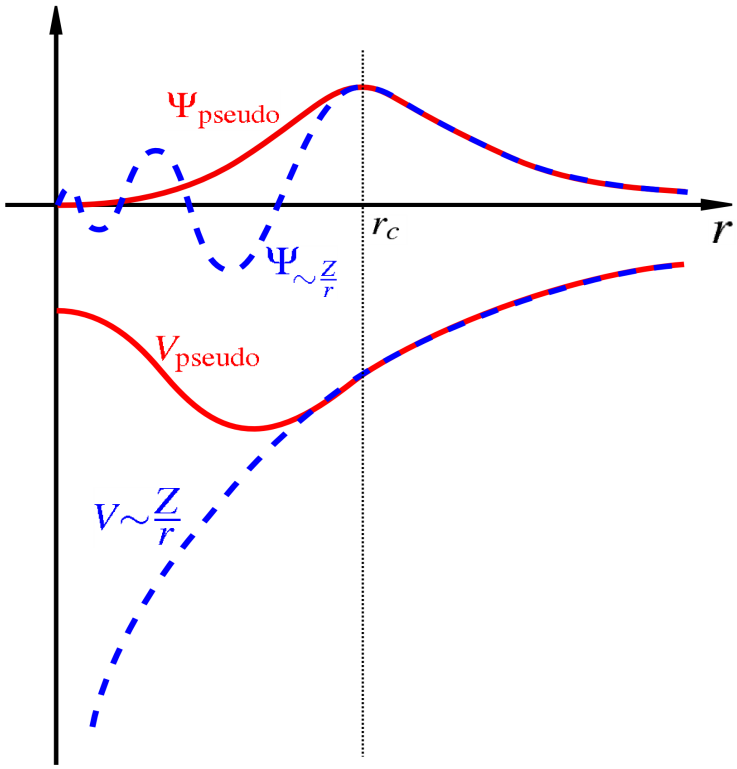

Figure 2.1: Example showing how the use of an ionic core potential may smooth out the wavefunction of a valence electron within the cutoff radius. Note the matching of the all-electron (blue dashed line) wavefunction and pseudo-wavefunction (red solid line) beyond the cutoff radius.

with $R^l_{\text{AE,PS}}$ being the all-electron or pseudopotential values of the radial parts of the wavefunctions for angular momentum channel l. These pseudopotentials deliver high accuracy and demonstrate a remarkable amount of transferability, that is to say the pseudopotentials are generated to match all-electrons calculations for a reference system and are then expected to hold for other slightly different chemical environs. As a tradeoff, these pseudopotentials require quite high cutoff energies to produce such accurate results. Ultrasoft pseudopotentials, on the other hand make use of the fact that there is technically no need for the norm-conserving constraint and one can generate a pseudo-wavefunction with good accuracy and transferability requiring much lower cutoff energies. However, their creation

involves the use of a so-called augmentation charge which makes their use in some types of calculations of material properties slightly more cumbersome.

The pseudopotentials used throughout this dissertation are of the norm-conserving variety and are generated by the Vanderbilt method. Moving forward, we will continue to reference what are actually pseudo-wavefunctions as simply wavefunctions unless the distinction between the two is necessary.

Self-consistency: As mentioned in our discussion of the Kohn-Sham ansatz, the reference potential which is used to calculate the electronic wavefunctions, which in turn define the electronic density, depends itself on the set of electronic wavefunctions, being a functional of the electronic density. This problem is resolved by using a self-consistent scheme, *i.e.* one whose input is an educated guess and whose output is either consistent with the input, or refined until consistent with the input.

One means of approach requires the direct minimization of Eq. (2.22) with respect to the set of electronic wavefunctions using a technique such as the conjugate gradient method. This involves choosing a trial single-particle wavefunction and minimizing the contribution of this state to the total energy while maintaining its orthogonality with all other states. A more widely used technique is instead to opt to solve the system of eigenvalue equations presented by Eq. (2.23). After expanding the wavefunctions in the plane wave basis, these equations take the form:

$$\begin{aligned} &\sum_{\boldsymbol{G}'} \left[\frac{\hbar^2}{2m} |\boldsymbol{k}+\boldsymbol{G}| \delta_{\boldsymbol{G}\boldsymbol{G}'} + \tilde{V}_{\text{ext}}(\boldsymbol{G}-\boldsymbol{G}') + \tilde{V}_{\text{Hartree}}(\boldsymbol{G}-\boldsymbol{G}') + \tilde{V}_{\text{XC}}(\boldsymbol{G}-\boldsymbol{G}') \right] C_{n,\boldsymbol{k}+\boldsymbol{G}'} \\ &= \epsilon_n(\boldsymbol{k}) C_{n,\boldsymbol{k}+\boldsymbol{G}}, \end{aligned} \tag{2.37}$$

where tildes signify the Fourier transform of the corresponding term.

With all the mechanisms hitherto discussed, the self-consistent electron density takes the form:

$$\rho(\boldsymbol{r}) \;=\; \sum_{\boldsymbol{k}\in \mathrm{BZ}} w_{\boldsymbol{k}} \sum_{n=1}^{N_{\boldsymbol{k}}} f_n^{(\boldsymbol{k})} |\phi_{n(\boldsymbol{k})}(\boldsymbol{r})|^2, \tag{2.38}$$

with $N_{\boldsymbol{k}}$ the number of electronic states occupied at that k-point, $f_n^{(\boldsymbol{k})}$ is the occupation number of the particular state indexed by n and k, all of which are weighted by the factors $w_{\boldsymbol{k}}$ introduced in Eq. (2.35). Eq. (2.37) can then be rendered as a large matrix whose number of elements depends on the chosen cutoff energy, k-point mesh, and number of states to be calculated, n. By diagonalizing this matrix, eigenfunctions and eigenvalues can be computed, the former being able to be combined in order to yield the calculated density. The general outline of the procedure goes as follows: we begin with a trial electron density. This is usually approximated as a simple superposition of normal atomic orbitals. The reference potential of the Kohn-Sham scheme is then calculated for this density. This establishes the matrix to be diagonalized (for the given choice of calculation parameters) which is then used to generate eigenvalues and eigenfunctions of the Kohn-Sham system. The electronic density is built from these eigenfunctions and this density is checked against the input density. If they agree within an established tolerance, then that density is the self-consistently found ground-state density of the Kohn-Sham system, which can then be identified with the real system and from which material parameters can be extracted. If the input and output densities disagree, then a new trial density is constructed based on a mixing of the prior iteration's output and the current iteration's output. The new density is then used as input, and the cycle is repeated until self-consistency is reached. A diagram of this process is included in Fig. (2.2).

In this work, the DFT code of choice was Quantum Espresso [60]. It is a well-tested, open-source code suite that is interoperable with other useful solid-state physics code-bases. The details of its computational schema and workflows are documented in Appendix C.

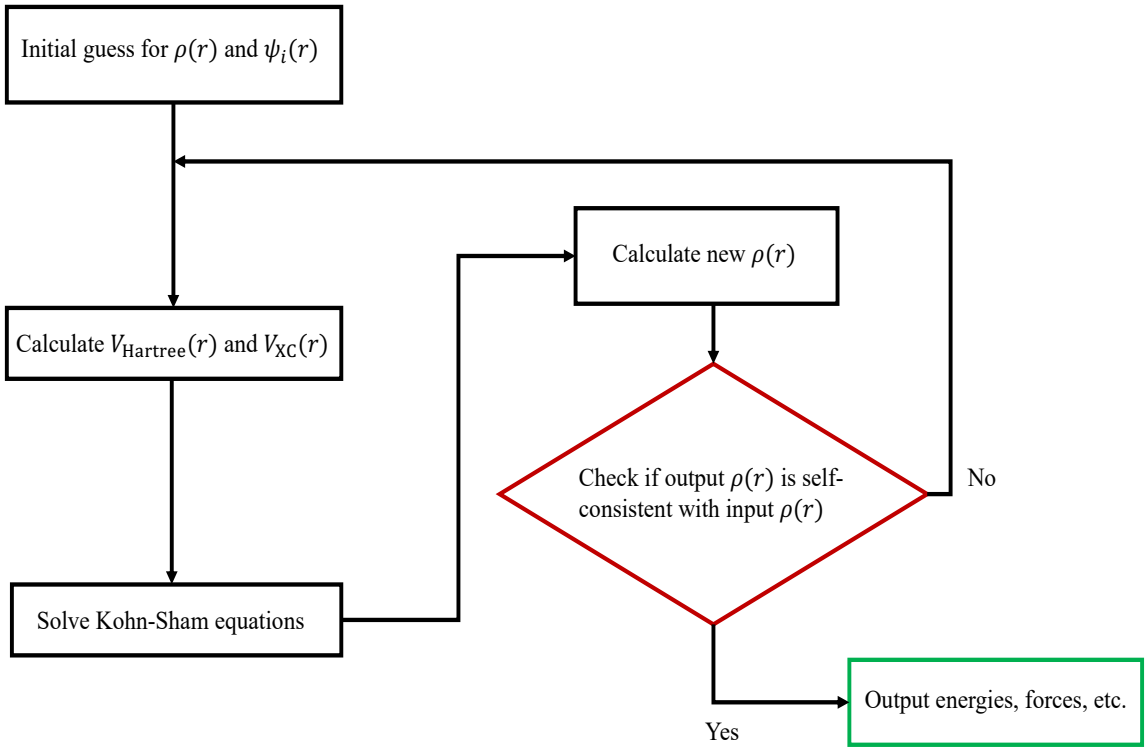

Figure 2.2: A flowchart demonstrating the logic of attaining a self-consistent electron density in DFT.

2.1.6 Limitations

DFT has proven itself to be a remarkable and robust tool for materials science exploration. With only knowledge of the chemistry and geometry of a crystal as input parameters, DFT offers a significantly less expensive way to model or explore new materials than other quantum chemistry methods. Inexpensive core hours and large computational facilities have made *ab initio* calculations possible for an ever-wider range of systems. While Kohn-Sham DFT has more than earned its reputation, the formalism is not all-encompassing. The variety of exchange-correlation functionals, while well-motivated and sensibly executed, still do not provide a completely systematic way to improve on the form of the universal functional. This can hardly be considered a limitation of DFT itself, but the lack of systematically improving the largest approximation within its framework leaves researchers with a variety of different

functionals to test on systems of interest. Some fare generally better than others for certain classes of materials, but this still presents a certain element of empirical study. Functionals could, at some level, be designed and tuned to fare better for one class of materials or another, but this dilutes the true *ab initio*, predictive nature of the method.

More glaringly at the heart of DFT is that it is a ground-state theory. The Kohn-Sham formalism gives the eigenvalues and eigenfunctions of the fictitious reference system, which through the graces of the two Hohenberg-Kohn theorems, are designed to produce a density that coincides with the ground-state density of the real system. No such guarantee is promised for excited states of the system. This precludes a large class of interesting phenomena, such as optical activity and charge-carrier dynamics, from being fully studied at this level. It is well-known that standard DFT underestimates the bandgap of semi-conductor systems by as much as 50%. While qualitative trends about *e.g.* mid-gap states and such can be examined, quantitative predictions about excitation spectra and the like are beyond the scope of standard DFT. Some systems also present pathologically for DFT. NiO is a material that was notoriously difficult given that it was known empirically to be a semi-conductor, but due to some aspects of the formalism, DFT calculations routinely predicted it as metallic. While workarounds such as the Hubbard U correction (and other advances along that vein) managed to alleviate some of these pathologies, they did so at the cost of introducing some amount of empirical judgment to the process.

Additionally, pure systems are often not of much experimental interest, as defects and impurities are almost unavoidable. Intentionally doped systems are often much more fertile grounds of exploration for experimentalists seeking to manipulate materials for application. As such, the periodic boundary conditions, which allow us to make powerful use of the Bloch theorem in DFT, work against us somewhat when it comes to modeling these types of systems. In order to accurately model an isolated defect in a system, a large supercell of the material must be used such that the defect is screened from interacting too strongly with its periodic replicas. While increases in computational resources can help abate this

problem to some extent, the calculations still scale like N^3, where N is representative of the number of electrons in the study, before considerations for basis set size and such are taken into account. This makes study of very large systems still computationally difficult.

All of this is not to say that DFT cannot deliver extremely useful results, from both theoretical and experimental perspectives. In addition to the wealth of knowledge that even standard DFT can offer about a material, post-DFT methodologies exist which use the Kohn-Sham results as a stepping stone to apply higher levels of theory. One such powerful method is known as the GW approximation, which is discussed in the next section.

2.2 The GW Approximation

While necessary to make progress on the essential aspects of the many-body problem, the assumptions of the Kohn-Sham ansatz involve us solving a problem of many non-interacting electrons. To move to higher levels of theory, we need to re-include the interactions in a formally sound way. Moving past these assumptions of standard DFT is the primary means of calculating the dynamics of excited states in materials. In this regard, the GW approximation is a well-founded post-DFT method which may give us access to important experimental quantities [61].

2.2.1 Connection to Spectroscopy

Spectroscopy, in its various forms, is one of the most fundamental means of studying physical phenomena. We introduce some common applications of spectroscopy used in solid-state physics. This background will be used to establish the nuances of many-body states and excitations whose description necessitate this higher level of theory.

- Photoemission Spectroscopy: In photoemission spectroscopy (PES), laser light is used to excite an electron from a bound state within the material into an unbound state where it is ejected from the material. With knowledge of the work function, here

labeled Φ, control over the energy of the input light ($h\nu$, with ν being the frequency of the light), and a measurement of the resulting kinetic energy of the electron E_{k}, the energy of the initial bound state of the emitted electron can be reconstructed using:

$$\mathrm{IP}_s \;=\; -\epsilon_s \;=\; h\nu - E_{\mathrm{k}} - \Phi \tag{2.39}$$

Here IP_s refers to the ionization potential, defined as the energy required to remove the electron from the state s of the neutral material. ϵ_s is the corresponding energy value of that bound state and since PES probes the occupied state, this energy must be *less* than the Fermi level, which can be set to zero (as done here) for convenience. It is worth noting that the IP is defined to be a positive quantity always.

- Inverse Photoemission Spectroscopy: The inverse process (IPES), involves the same quantities, however now an electron with controlled kinetic energy is used to probe the unoccupied states of the material. The electron is injected into the sample and at some time later undergoes a radiative transition into a lower energy state, emitting a photon of energy $h\nu$. Knowing the work function of the material, the initial kinetic energy of the electron, and now measuring the energy of the emitted photon, the energy of the state that the electron ultimately falls into can be reconstructed in a similar manner as to Eq. (2.39)

$$-\,\mathrm{EA}_s \;=\; \epsilon_s \;=\; E_{\mathrm{k}} - h\nu + \Phi \tag{2.40}$$

EA_s is a quantity known as the electron affinity and is defined as the energy needed to detach an electron from a negatively charged species; hence, its mathematical form as the negative of the energy eigenvalue of the state ϵ_s. Given that IPES is probing the unoccupied states of the material, the value of ϵ_s will be *at least* that of the Fermi level. In contrast to the IP_s, the EA_s can be positive or negative depending on whether

the state the electron ends up in is bound or unbound. EA_s is negative for an unbound final state.

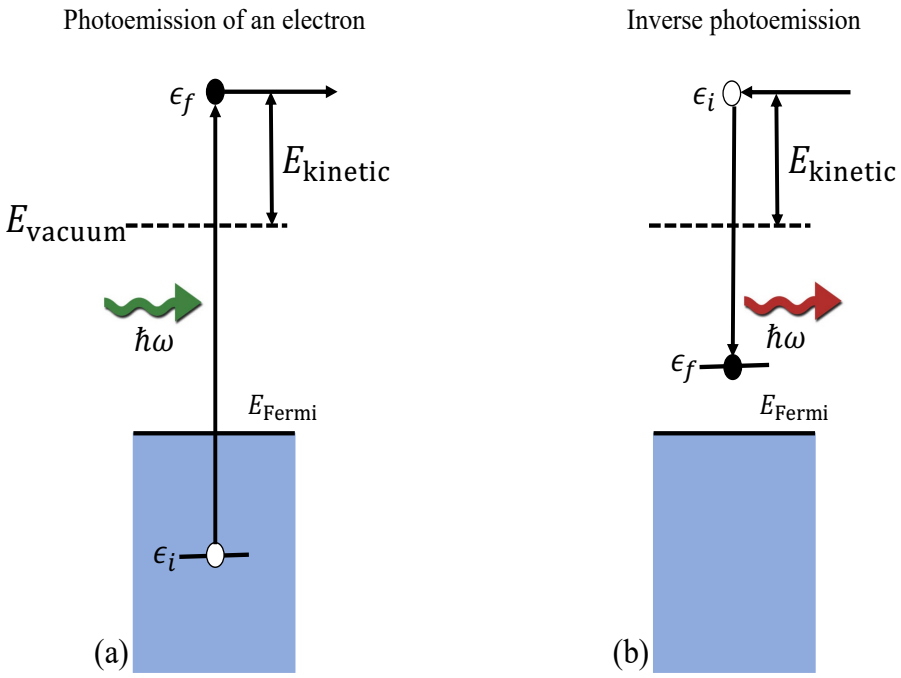

Figure 2.3: A representation of how the PES and IPES processes probe occupied and unoccupied states of the material. The exact energies of these types of excitations are due to complicated many-body effects.

Fig. 2.3 gives a good schematic of these processes. It emphasizes that these types of excitations are fundamentally different from single-particle excitations. PES and IPES involve probing the response of the entire system of electrons to either the excitation of one electron out of the material, or the presence of an extra electron in the material. We immediately see the important role of many-body effects. It is not simply that we can study the dynamics of how one electron behaves; we are, in fact, concerned with how the entire system of electrons responds to a change in the electron density.

In PES experiments, the experimentally observed quantity is the probability that an electron is emitted with a kinetic energy, E_{k}, in a given time interval, which manifests as the measured photocurrent in the material upon exposure to light. This photocurrent is related

in a complicated manner to the intrinsic spectral function of the system which is given by the imaginary part of the single-particle Green's function:

$$A(\boldsymbol{r}, \boldsymbol{r}', \omega) \;=\; \frac{1}{\pi} \operatorname{Im} G(\boldsymbol{r}, \boldsymbol{r}', \omega) \operatorname{sgn}(E_{\mathrm{F}} - \omega). \tag{2.41}$$

We do not concern ourselves here with the details of the mathematical relationship between the photocurrent and the spectral function, as it is yet not rigidly formulated.

2.2.2 The Green's Function

The single particle Green's function, G in Eq. (2.41), represents the probability that the creation of a particle at a point $\boldsymbol{r}$ at some point in time, is correlated with the annihilation of that particle at another point $\boldsymbol{r}'$ at some later point in time. This process can be thought of as the propagation of the particle from its point of creation to its point of annihilation. Hence, the alternative name for the Green's function as the propagator. Understanding of the propagator is vital to our discussion here and will be covered in more detail later. A useful resource which provides significant background on the topic of GW calculations can be found in the work by Golze *et al.* [62].

The creation and annihilation processes alluded to in the concept of the propagator correspond to the excitations of different many-body states. The transition energies of equations (2.39) and (2.40) are the removal and addition energies, respectively, of an electron from one of these states

$$\begin{aligned} \epsilon_s \;&=\; E(N) \;-\; E(N-1, s), \\ \Psi_s(\boldsymbol{r}) \;&=\; \langle N-1, s | \hat{\Psi}(\boldsymbol{r}) | N \rangle, \end{aligned} \tag{2.42}$$

for $\epsilon_s < E_{\mathrm{F}}$. Here $\hat{\Psi}(\boldsymbol{r})$ is an operator which annihilates a particle at point $\boldsymbol{r}$ with $\Psi(\boldsymbol{r})$ being the associated transition probability between the N and $N-1$ many-body states. These

many-body states, $|N, s\rangle$, are the many-body eigenstates of the N-electron Hamiltonian such that $\hat{H}_{\mathrm{MB}}|N, s\rangle = E(N, s)|N, s\rangle$, $E(N, s)$ being the corresponding energy eigenvalue of the state. In the context of a general PES experiment, Eq. (2.42) shows the understanding of these states quite simply. The state $|N\rangle$ describes the ground-state of the N-electron system. In the presence of photoexciting light, an electron is removed from the system, transitioning to an $N-1$ electron state by removing an electron from some particular valence band, s. We can build up an analogous expression for the IPES situation via:

$$\begin{aligned} \epsilon_s &= E(N+1, s) - E(N), \\ \Psi_s(\boldsymbol{r}) &= \langle N|\hat{\Psi}(\boldsymbol{r})|N+1, s\rangle, \end{aligned} \tag{2.43}$$

for $\epsilon_s \geq E_{\mathrm{F}}$. Now, for an IPES experiment, we conceptualize the neutral N-body system transitioning to the $N+1$ state with the addition of the electron, now in some other, unoccupied band s. In both cases, the interest is in calculating the *response* of the system of electrons to either the absence of an electron from the group, or to the presence of another electron.

Having already introduced the concept of the Green's function, we can solidify the intuition built by equations (2.42) and (2.43) with the precise definition of the Green's function:

$$G(\boldsymbol{r}, \sigma, t, \boldsymbol{r}', \sigma', t') = -i \left\langle N \left| \hat{T}[\hat{\Psi}(\boldsymbol{r}, \sigma, t)\hat{\Psi}^{\dagger}(\boldsymbol{r}', \sigma', t')] \right| N \right\rangle, \tag{2.44}$$

with σ representing the spin-state of the electron, and the Hermitian adjoint of the annihilation operator $\hat{\Psi}^{\dagger}$ being instead a creation operator. $\hat{T}$ is the time-ordering operator which arranges the operators such that the operator acting at an earlier time stands to the right and acts on the ground-state ket $|N\rangle$. However, the definition of G allows for both time orderings, $(t > t')$ or $(t < t')$. With the creation operator standing to the right, G describes, as discussed previously, the propagation of an electron from one point in space and time to

another point in space at some *later* time. If instead, the other time ordering is assumed, then the process now describes the creation of a *hole* at one space-time point which propagates *backward* in time to the other point in space. An important point to be made is that taking Eq. (2.44) with argument $(\boldsymbol{r})$ for both operators, at infinitesimally different times, yields a density of the many-body system akin to Eq. (2.9).

The Fourier transform of the Green's function in equation (2.44) turns the time argument into an energy difference and reveals more about the importance of the Green's function:

$$G(\boldsymbol{r},\boldsymbol{r}',\omega) \;=\; \lim_{\eta\to 0^+}\sum_s \Psi_s(\boldsymbol{r})\Psi_s^*(\boldsymbol{r}')\left[\frac{\Theta(\epsilon_s - E_{\rm F})}{\omega-(\epsilon_s - i\eta)} \;+\; \frac{\Theta(E_{\rm F}-\epsilon_s)}{\omega-(\epsilon_s+i\eta)}\right], \tag{2.45}$$

which is referred to as the Lehman representation of the Green's function. Eq. (2.45) assumes that we have summed over the spin degrees of freedom in the definition of G in Eq. (2.44). η here is an infinitesimal factor which allows for the poles of the Green's function to be calculated via contour integration. Θ is the Heaviside function, which is unity for positive arguments and zero for negative arguments, and enforces the time ordering of any creation/annihilation process. The Lehman representation of the Green's function allows us to identify the many-body excitations in equations (2.42) and (2.43) with the poles of the single-particle Green's function. Additionally, the Lehman representation allows us to examine the spectral function in Eq. (2.41). In its diagonal form it becomes:

$$\begin{aligned} A(\boldsymbol{r},\boldsymbol{r},\omega) \;&=\; \frac{1}{\pi}\,\mathrm{Im}\,G(\boldsymbol{r},\boldsymbol{r},\omega)\,\mathrm{sgn}(E_F-\omega) \\ &=\; \sum_s \Psi(\boldsymbol{r})\Psi^*(\boldsymbol{r})\delta(\omega-\epsilon_s). \end{aligned} \tag{2.46}$$

2.2.3 The Quasi-particle Picture

Eq. (2.46) has the form of a many-body density of states. Yet the excitations of the real system do not present as precise delta function peaks. A generic excitation pattern is displayed in Fig. 2.4. In this fiqure, the distinct peak is attributed to a particle-like state, but with finite width. This broadening comes about as the sum of many delta functions, close in energy, merging to form this so-called quasi-particle peak. The smaller peak away from the main peak is referred to as a satellite whose presence in the quasi-particle spectrum is an essential feature of the process.

We can motivate the thinking about quasi-particle excitations by considering a system of particles whose two-body interaction is turned on adiabatically. The spectral function of the non-interacting system will consist purely of delta function peaks and the many-body states will be characterized by the action of a single creation or annihilation operator. The wavefunctions and associated energies of particles in this system will be eigenfunctions and eigenvalues of a single-particle Hamiltonian.

With the interaction turned on, these states are no longer single-particle eigenstates. The matrix elements of the spectral function contributing to a feature will involve transition amplitudes from states which no longer vanish. These contributions can combine into a definable peak centered around one of these single-particle eigenvalues and when this occurs it appears as a broadening of the original, delta-like peak and is referred to as a quasi-particle excitation. This propagation of the extra particle can also excite other collective excitations (plasmons, phonons, magnons, *etc.*) in the medium which are responsible for the satellite peak in the characteristic spectrum in Fig. (2.4).

The width of the quasi-particle peak is associated with the lifetime of the excitation τ. The broadening of the peak is associated with a renormalization factor, symbolized by Z, of the weight relative to the non-interacting peak. A quasi-particle peak then assumes an approximate shape:

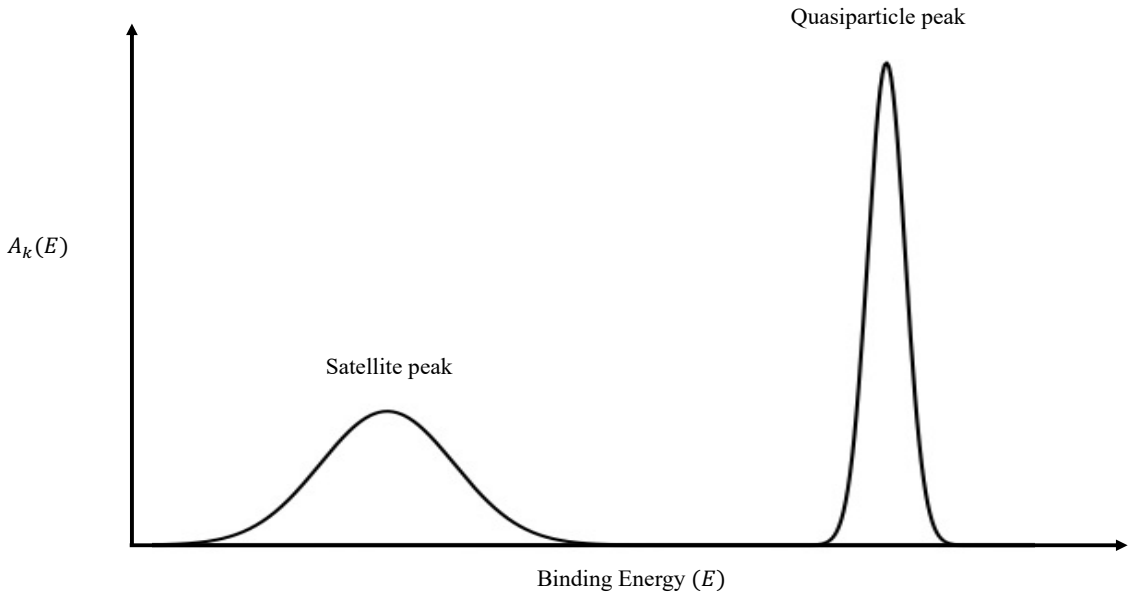

Figure 2.4: A generic scheme of a many-body excitation. The main peak is referred to as a quasi-particle excitation. The smaller peak is referred to as a satellite peak and arises due to other types of collective excitations which occur in response to the main excitation.

$$A_{ss}(\omega) \approx \frac{1}{\pi} \left| \frac{Z_s}{\omega - (\epsilon_s + i\Gamma)} \right|. \tag{2.47}$$

Quasi-particle energies are not eigenvalues of the full many-body Hamiltonian. The poles of the Green's function in equation (2.45) lie exclusively along the real axis whereas quasi-particle excitations carry a complex component to them as well. The real part of the quasi-particle energy is associated with the actual energy of the excitation, while the imaginary component, $\Gamma = \frac{2}{\tau}$ is inversely related to its finite lifetime.

Examining the real-space perspective will give us important insight which will play into the actual implementation of the GW approximation. With a system of N interacting electrons, the repulsive nature of the two-body Coulombic interaction keeps the electrons relatively separated. The addition (or removal) of another electron into this system now

sees the N electrons moving to accommodate this extra electron (or hole). The additional particle carries with it a cloud of locally opposite charge because the N electrons will always move to screen the charge, either getting closer to the extra charge in the case of a hole or moving away from the extra charge in the case of an extra electron. In either case, the extra particle and its polarization cloud are deemed the quasi-particle. The feedback between the extra particle and the perturbation it causes in its surroundings is termed the self-energy and will be a central focus of the GW approximation.

2.2.4 Hedin's Equations

The Green's function formalism is a powerful tool for calculating various response quantities of a system. We have introduced several of the important pieces of this formalism already, and relegate further details of the general formalism to Appendix E.

In 1965, Lars Hedin expanded the Green's function and the self-energy term in powers of the screened Coulomb interaction [63]. This leads to a set of coupled, closed-form integro-differential equations which now bear his name. We will arrive at these equations by a perturbative approach. The intuition for this approach is quite straightforward.

The starting block to build upon would be the non-interacting propagator, usually symbolized by G_0. This is easy to envision, as it involves the particle created in the system propagating unperturbed to the point where it is eventually annihilated. Harkening to the Feynman path integral picture, we must consider all possible paths the particle can take to reach its final point. Each of these paths consists of various interactions which can disrupt the particle. In the case of an electron propagating through a solid, one huge family of these perturbations will be the two-body Coulombic interactions it experiences with other electrons in the medium. However, these perturbations can actually get arbitrarily complex, and all types must be included in the sum to get the true, fully interacting propagator.

A full understanding of this concept requires a more in-depth understanding of Feynman diagrams, however, a full review goes far beyond the purview of this chapter and we relegate

a more complete exposition to Appendix E. We instead focus here on the essential points necessary to understand Hedin's equations and subsequently the GW approximation to them.

The self-energy, mentioned briefly at the end of the previous section, is the actual quantity of interest in the GW approximation. It contains all of the exchange and correlation effects which will generate the true, interacting propagator. We encapsulate this effect in the definition of the fully interacting propagator:

$$G(1,2) \;=\; G_0(1,2) \;+\; \int d(3,4) G_0(1,3) \Sigma(3,4) G(4,2). \tag{2.48}$$

We have introduced a more streamlined notation in Eq. (2.48). The numerals now stand for an entire set of coordinates such that $1 = \{\boldsymbol{r}_1, t_1, \sigma_1\}$. Eq. (2.48) is the Dyson equation and we see that G appears recursviely within its own definition. We could re-insert the definition of G *ad infinitum* to get ever-higher orders of the series. In principle, this process needs to be iterated to infinite order to be perfectly accurate. Obviously this is impossible, and so the series must be truncated at some point. Hedin's approach to doing this can be conceptualized as taking all Feynman diagrams of a characteristic shape, referred to as a diagram's "topology", and renormalizing them into a single approximate object.

In doing so, we "dress" the interaction with the approximate total effect of a family of processes and replace an infinite number of Feynman diagram objects with a single object. There is no information lost in this procedure, but the details of the approximated objects become encoded in the energy dependence of the renormalized object.

A more thorough examination of Hedin's derivation is undertaken in Appendix E, but here we state the end result of Hedin's formalism:

$$
\begin{aligned}
P(1,2) &= -i \int d(3,4) G(4,2) G(2,3) \Gamma(3,4,1) \\
W(1,2) &= v(1,2) + \int d(3,4) v(1,2) P(3,4) W(4,2) \\
\Sigma(1,2) &= i \int d(3,4) G(1,4) W(1^{+},3) \Gamma(4,2,3) \\
\Gamma(1,2,3) &= \delta(1,2)\delta(1,3) + \int d(4,5,6,7) \frac{\delta \Sigma(1,2)}{\delta G(4,5)} G(4,6) G(7,5) \Gamma(6,7,3).
\end{aligned}
\tag{2.49}
$$

The equations in (2.49) are Hedin's equations. The function P is the reducible polarizability, W is the screened Coulombic interaction, Σ is the self-energy, and Γ is known as the vertex function.

The polarizability is a familiar quantity, describing the development of dipoles in the material in response to an electric field. As such, this is an important quantity in describing the dynamics of the polarization cloud of the additional charge. The chief insight of Hedin was to use the term W, the screened Coulomb interaction, instead of the bare interaction in the description of the many-body problem. Screening effects, usually very strong in polarizable materials, reduce the magnitude of the Coulomb interaction and thus facilitates the perturbative expansion we've discussed. Σ, as mentioned earlier, becomes the quantity of focus in Hedin's formulation. It captures all the effects of exchange and correlation and in principle must be non-local and energy dependent, though these points are discussed more deeply later. Γ is perhaps the most difficult to conceptualize. It is a structure which, in the Feynman diagram picture, connects two points of propagation at internal times with an interaction. This can be done with or without internal structure to the vertex itself, but again, these details are relegated to Appendix E. The vertex function depends on three coordinate sets, and thus is also the most difficult to calculate.

Taken together with Eq. (2.48), Hedin's equations become closed. We combine the Hedin equations with several other auxiliary equations to form a full picture of how these equations can be linked together. We present them here:

$$\begin{aligned}\chi(1,2) &= \chi_0(1,2) + \int d(3,4)\chi_0(1,3)v(3,4)\chi(4,2) \\ &= \left\langle N \left| \hat{T}[\delta\hat{n}(1)\delta\hat{n}(2)] \right| N \right\rangle,\end{aligned} \tag{2.50}$$

$$\epsilon^{-1}(1,2) = \delta(1,2) + \int d(3)v(1,3)\chi(3,2),$$

where ϵ^{-1} is the inverse of the dielectric function, which quantifies the electrical response of the material. χ is, in constrast to P, the *irreducible* polarizability, so called because it cannot be broken down into smaller, composite Feynman diagrams of the same topology. Here $\delta\hat{n}(1) = \hat{n}(1) - n(1)$ represents the fluctuation of the density about its average value, with $\hat{n}$ being a number operator. We can use these two expressions to re-write the screened Coulomb interaction as follows

$$W(1,2) = \int d(3)\epsilon^{-1}(1,3)v(3,2). \tag{2.51}$$

The frequency dependence of the screened Coulomb interaction is quite important. The process of screening involves the rearrangement of charges in order to minimize their interaction. This time difference dependence, manifesting equally as a frequency dependence, is what allows the system to respond to the extra charge and generate the quasi-particle. Handling this frequency dependence can be done fully, but approximations for it are discussed in Appendix D.

An interpretation of the effect of Hedin's equations is useful here. The polarizability equation describes the creation of electron-hole pairs. Their interaction is captured by the vertex function, which depends upon the change in potential induced by the creation of the

particle pair. The polarizability determines the dielectric function, which in turn determines the screened Coulomb interaction. The self-energy then describes the energy of the created particle pair due to its interaction with the environment.

Despite the steps taken to reduce the number of Feynman diagrams calculated, Hedin's equations are still formidable to solve. Recognizing this, Hedin suggested the GW approximation. By taking the approximation $\Gamma(1,2,3) = \delta(1,2)\delta(1,3)$, the equation for the self-energy then becomes $\Sigma = iGW$ where the approximation scheme derives its name. This leaves us with the set of equations which form the heart of the GW approximation.

$$\begin{aligned}
G(1,2) &= G_0(1,2) + \int d(3,4) G_0(1,3)\Sigma(3,4)G(4,2) \\
\Gamma(1,2,3) &= \delta(1,2)\delta(1,3) \\
\chi_0(1,2) &= -iG(1,2)G(2,1) \\
W(1,2) &= v(1,2) + \int d(3,4) v(1,3)\chi_0(3,4)W(4,2) \\
\Sigma(1,2) &= iG(1,2)W(1^+,2).
\end{aligned} \tag{2.52}$$

In equation (2.52), a coordinate with a + superscript represents the same space and spin coordinates, but infinitesimally advanced in time, that is $1 = \{\boldsymbol{r}_1, \sigma_1, t_1\}$ and $1^+ = \{\boldsymbol{r}_1, \sigma_1, t_1 + \eta\}$ with η being infinitesimal.

With the simplification to the vertex function, the theory describes a system of non-interacting *quasi-particles*. We can still use these quasi-particle results to build even higher-level interactions, such as describing excitonic behavior. This involves solutions of what is

known as the Bethe-Salpeter equation whose discussion is limited to Appendix D where it is presented as the means by which GW methods can calculate absorption spectra.

2.3 Implementation

The GW equations introduced in the previous section seem to lend themselves to a self-consistent scheme similar to what occurs in DFT. However, this presents both computational and theoretical difficulties. A straightforward expansion of the self-energy in terms of the screened interaction is not guaranteed to converge. Additionally, there is no prescription for how and at what stage in the self-consistent cycle the vertex should be updated.

In practice, a fully self-consistent GW scheme does not seem to generally improve the experimental agreement [64–66] over non-self-consistent applications. Fully self-consistent schemes are also significantly more expensive to execute. In this dissertation, we use exclusively what are known as single-shot GW calculations.

2.3.1 Single-shot GW

Single-shot GW, also known as the G_0W_0 method, is perhaps the most widely used GW scheme. It is also the simplest. The general schema of the G_0W_0 is shown in Fig. (2.5)

We begin with an effective single-particle eigenvalue equation, known as the quasi-particle equation:

$$\hat{h}^{\mathrm{MF}}(\boldsymbol{r})\psi_{s\sigma}(\boldsymbol{r}) \; - \; \int d\boldsymbol{r}' v_{\sigma}^{\mathrm{MF}}(\boldsymbol{r},\boldsymbol{r}')\psi_{s\sigma}(\boldsymbol{r}') \; + \; \int d\boldsymbol{r}' \Sigma_{\sigma}(\boldsymbol{r},\boldsymbol{r}',\sigma_{s\sigma})\psi_{s\sigma}(\boldsymbol{r}') \; = \; \epsilon_{s\sigma}\psi(\boldsymbol{r}), \quad (2.53)$$

with

$$\hat{h}_{\sigma}^{\mathrm{MF}} \; = \; -\frac{1}{2}\nabla^2 \; + \; v_{\mathrm{ext}} \; + \; v_{\mathrm{H}} \; + \; v_{\sigma}^{\mathrm{MF}}. \quad (2.54)$$

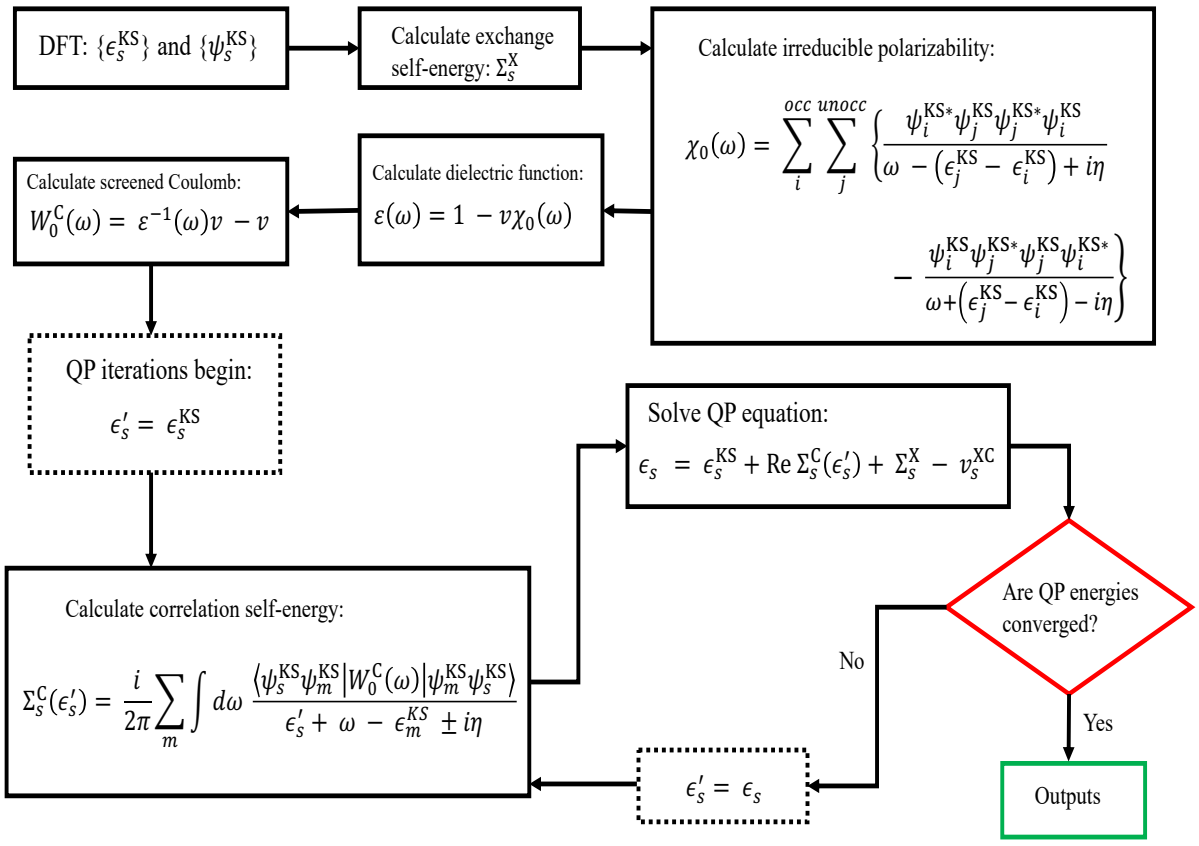

Figure 2.5: A general diagram of how the G_0W_0 scheme works. Using the DFT eigenvalues and wavefunctions, the exchange portion of the self-energy (Σ^{X}) can be approximated. The wavefunctions and eigenvalues are also used to construct the irreducible polarizability, χ_0. The dielectric function can then be generated and subsequently the screened Coulomb interaction, W is approximated. Then the Kohn-Sham eigenvalues are used as an initial value for the quasiparticle energies to generate the correlation part of the self-energy Σ^{C}. The quasiparticle eigenvalue equation can then be solved and iterated until self-consistency is reached in the quasiparticle energies.

The superscript MF refers to the mean-field, in our case DFT, quantities. Here v_σ^{MF} is the DFT level exchange-correlation potential, with the spin channel given in the index σ, v_{H} is the Hartree potential, and v_{ext} is the external potential all as discussed in the section on DFT. The solutions to equation (2.53) yield the quasi-particle wavefunctions $\{\psi_{s\sigma}\}$ and the quasi-particle energies $\{\epsilon_{s\sigma}\}$. These solutions can only be obtained after making our

approximations to solve for the self-energy. We examine the GW self-energy in the frequency domain:

$$\Sigma_\sigma(\boldsymbol{r},\boldsymbol{r}',\omega) \;=\; \frac{i}{2\pi}\int d\omega' e^{i\omega'\eta} G_0^\sigma(\boldsymbol{r},\boldsymbol{r}',\omega+\omega')W_0(\boldsymbol{r},\boldsymbol{r}',\omega'). \tag{2.55}$$

We now need to generate the non-interacting propagator and the screened Coulomb interaction. The initial, non-interacting propagator is created using the set of wavefunctions, $\{\phi_{s\sigma}^0\}$, generated by Kohn-Sham DFT:

$$G_0^\sigma(\boldsymbol{r},\boldsymbol{r}',\omega) \;=\; \sum_m \frac{\phi_{m\sigma}^0(\boldsymbol{r})\phi_{m\sigma}^{0*}(\boldsymbol{r}')}{\omega-\epsilon_{m\sigma}^0 - i\eta\,\mathrm{sgn}(E_F-\epsilon_{m\sigma}^0)}. \tag{2.56}$$

The screened Coulomb interaction is slightly more involved. Its form under the random phase approximation (RPA), as indicated previously in Eq. (2.51) relies on the inverse of the dielectric function and the Coulomb potential. The dielectric function is found by:

$$\epsilon(\boldsymbol{r},\boldsymbol{r}',\omega) \;=\; \delta(\boldsymbol{r},\boldsymbol{r}') \;-\; \int d\boldsymbol{r}'' v(\boldsymbol{r},\boldsymbol{r}'')\chi_0(\boldsymbol{r}'',\boldsymbol{r},\omega), \tag{2.57}$$

and the irreducible polarizability given by:

$$\chi_0(\boldsymbol{r},\boldsymbol{r}',\omega) \;=\; -\frac{i}{2\pi}\sum_\sigma \int d\omega' G_0^\sigma(\boldsymbol{r},\boldsymbol{r}',\omega+\omega')G_0^\sigma(\boldsymbol{r},\boldsymbol{r},\omega), \tag{2.58}$$

which simplifies to [67, 68]:

$$\chi_0(\boldsymbol{r},\boldsymbol{r}',\omega) = \sum_\sigma \sum_i^{\mathrm{occ}} \sum_a^{\mathrm{cond}} \left\{ \frac{\phi_{i\sigma}^{0*}(\boldsymbol{r})\phi_{a\sigma}^0(\boldsymbol{r})\phi_{a\sigma}^{0*}(\boldsymbol{r}')\phi_{i\sigma}^0(\boldsymbol{r}')}{\omega-(\epsilon_{a\sigma^0}-\epsilon_{i\sigma}^0)+i\eta} - \frac{\phi_{i\sigma}^0(\boldsymbol{r})\phi_{a\sigma}^{0*}(\boldsymbol{r})\phi_{a\sigma}^0(\boldsymbol{r}')\phi_{i\sigma}^{0*}(\boldsymbol{r}')}{\omega+(\epsilon_{a\sigma^0}-\epsilon_{i\sigma}^0)-i\eta} \right\}. \tag{2.59}$$

Eq. (2.59) is known as the Adler-Wiser expression. In it, the sum over i runs over occupied states and a includes some number of unoccupied states. In principle, the sum over a should be infinite, but we will return to this point later.

The effects of exchange and correlation are usually separated as two contributions to the self-energy. The correlation part is represented as:

$$\begin{aligned}
\Sigma^{\mathrm{c}}_{\sigma}(\boldsymbol{r},\boldsymbol{r}',\omega) &= \frac{i}{2\pi}\int d\omega' G^{\sigma}_{0}(\boldsymbol{r},\boldsymbol{r}',\omega+\omega')W^{\mathrm{c}}_{0}(\boldsymbol{r},\boldsymbol{r}',\omega'), \\
&\text{where } W^{\mathrm{c}}_{0} \text{ is defined by:} \\
W^{\mathrm{c}}_{0}(\boldsymbol{r},\boldsymbol{r}',\omega) &= W_{0}(\boldsymbol{r},\boldsymbol{r}',\omega) - v(\boldsymbol{r},\boldsymbol{r}'),
\end{aligned} \tag{2.60}$$

and the exchange part can be represented as [69]:

$$\begin{aligned}
\Sigma^{\mathrm{x}}_{\sigma}(\boldsymbol{r},\boldsymbol{r}') &= \frac{i}{2\pi}\int d\omega' e^{i\omega'\eta} G^{\sigma}_{0}(\boldsymbol{r},\boldsymbol{r}',\omega+\omega')v(\boldsymbol{r},\boldsymbol{r}') \\
&= -\sum_{i}^{occ}\phi^{0}_{i\sigma}(\boldsymbol{r})\phi^{0*}_{i\sigma}(\boldsymbol{r}')v(\boldsymbol{r},\boldsymbol{r}').
\end{aligned} \tag{2.61}$$

While an exponential factor is technically necessary to close the contour integration of the Eq. (2.60), the screened Coulomb interaction dies off quite quickly and the limit $\eta \to 0^{+}$ can be taken before the integration.

It should be noted that while we have referred to G_0 as the non-interacting Green's function, this is technically only true if calculated using a non-interacting Hamiltonian. In practice, our G_0 is constructed considering at least the Hartree interaction. The mean-field Kohn-Sham single-particle wavefunctions turn out to be a good starting point to a true, interacting propagator and thus, when referring to G_0 in this work, we will consider it to be this starting point.

There are many computational details that need to be discussed for a full picture of the complexities that are involved in a GW calculation. However, they are also tied to the code suite used. In all studies here, we've used the BerekeleyGW code package [70]. These details are relegated to Appendix D, where we discuss details of BerkeleyGW.

Chapter 3

Effects of Transition Metal Dopants on the Electronic Structure of $KNbO_3$

As discussed in Chapter 1, the breadth of properties of $KNbO_3$ is the result of an interplay of structures in the cationic and anionic sublattices of its various phases. However, given the risk of the sample cracking upon cooling, the series of structural transitions not only makes growing large boules of the crystal difficult, but also complicate defect characterization techniques such as electron paramagnetic resonance (EPR), which work best at low temperatures.

Considering the expense and effort involved in growing crystals of $KNbO_3$, a deep and fundamental understanding of the specific defect complexes and their chemical nature can provide helpful insight to experimentalists in the physical, chemical, and materials sciences, allowing them to focus on the most promising systems. To this end, first-principles calculations present a cost-effective way to narrow down the field of candidate dopant materials and predict the likelihood of different defect complexes.

To the extent that DFT calculations can be used to predict how a given defect complex model affects the band structure of a material, this study is a "first pass" at examining which dopants (under the specific models considered here) might be useful for the purposes

of extending the spectral response range of $KNbO_3$. More stringent quantitatively accurate calculations of these systems under higher levels of theory can then be undertaken to more accurately predict quantitative properties of the top candidates.

In this chapter, we calculate band structures within DFT for $KNbO_3$:X – V_O (where X = Ti, V, Cr, Mn, Fe, Co, Ni and V_O represents an oxygen vacancy) to examine defect states introduced into the band gap by a given transition metal under two types of defect models. The rest of the chapter is organized as such: Section II introduces our methods and the defect models in detail, Section III presents our results and analysis of the band structures and the preliminary viability of a given dopant for visible and IR response applications, and Section IV summarizes our conclusions and compares our predictions to experimental results with outlook toward the next chapters.

3.1 Methodology

For this study, we employ DFT under the generalized gradient approximation (GGA), as implemented in the Quantum Espresso distribution (version 6.2) [60]. All pseudopotentials used are of the optimized norm-conserving Vanderbilt (ONCV) variety [71, 72].

We investigate the orthorhombic structure. In order to perform calculations of the doped systems, we created a $2 \times 2 \times 2$ supercell containing 80 atoms. The pure supercell was then relaxed to establish lattice parameters (reported in Chapter 1) and equilibrium positions for all atoms. For this supercell, relatively tight convergence of the total energy of the supercell (up to $\sim$10 mRy) was achieved using a kinetic energy cutoff of 150 Ry and a charge density cutoff of 600 Ry. Additionally it was determined that a $4 \times 4 \times 4$ Monkhorst-Pack [59] k-point grid was enough to obtain converged total energies of the pure supercell to within a $\sim$1 mRy level. We find that our lattice parameters differ from experimental results by less than 1.5%, [12, 13] which is within the expected accuracy of Kohn-Sham DFT. Calculations on the doped systems included a small smearing of the occupation levels (convolution with

a Gaussian of width 0.02 Ry) in order to account for the small, but non-negligible metallic characteristic imparted by the dopants. All doped systems were treated in the calculation as spin polarized.

To ascertain possible magnetic ground states of these systems, QE allows one to bias the magnetization of a given atomic species at the start of the calculation. This allows the calculation to explore the magnetic subspaces of the possible electron densities. The self-consistent DFT cycle proceeds as normal, but now the code also seeks to minimize the magnetization of the cell in addition to minimizing the energy of the electron density. To ensure that we are finding the true ground states of the systems presented, we ran self-consistent calculations with a fully polarized dopant atom, that is, setting the starting magnetization to its maximum value, as well as a calculation with this value set to zero. That is, we explore the minimization process from both the most polarized case and the least polarized case. In all cases, we find that either both starting points find electron densities with identical magnetizations and total energies, or that the biased starting magnetization results produce a density with a lower total energy than the unbiased starting point. We present results derived from the fully polarized starting point.

3.1.1 Defect Models

While the exact origin of the IR sensitivity introduced by $3d$ transition metals remains an open question [48], one prevailing idea is that the transition metals preferentially occupy Nb sites within the crystal [73, 74]. It is generally thought that at this substitution site the dopants assume a 3+ charge state partly on the basis that their atomic radii in this charge state are closer to that of Nb^{+5}, which should minimize stress on the cell and lead to a more favorable formation energy. Given this assumption, the removal of an oxygen atom in the cell would free two electrons which would presumably bind on the dopant to attain the 3+ charge state [75]. An O vacancy is thus invoked as one possible mechanism of charge compensation. Though other centers are certainly thought to be at play in many samples

[76, 77]. The band structures presented later support the assertion that Cr, Mn, Fe, Co, and Ni all exist in the predicted 3+ charge state.

In the models tested here, we examine a substitution of Nb with Ti, V, Cr, Mn, Fe, Co, and Ni (*i.e.,* the $3d$ transition metals). We then examine the band structures and projected densities of states when we induce the charge-compensating O vacancy in either the first coordination shell of O atoms around the dopant (depicted in Fig. 3.1 (a)) or in a shell that is as far away from the dopant as possible in the supercell. The former model we refer to as the coordinated vacancy model whereas the latter we call the uncoordinated vacancy model. We note that in this latter model, the distance between the dopant and the vacancy is $\sim$7.5 Å which can hardly be considered totally isolated, but is the best that can be done in a $2 \times 2 \times 2$ supercell. Fortunately, $KNbO_3$ has a high static dielectric constant [78] which significantly reduces the electrostatic interaction between periodic replicas. In both models, after the defect is introduced, the atoms are again allowed to relax in response to the presence of the defect.

Our models aim to represent the effects of the dopants when the charge-compensating vacancy either does or does not disturb the crystal field locally surrounding the dopant. The expected effects of this disturbance will be examined in the next subsection.

3.1.2 Expected Trends

As shown in Fig. 3.1(b) and (c), an important factor in determining the population energetics of the dopant $3d$ states involves considerations of the local crystal field experienced by the dopant ion. In particular, the geometry of the surrounding chemical environment introduces an electrostatic perturbation which modifies the position of the dopant states. A dopant ion surrounded by an ideal oxygen octahedron will experience a crystal-field splitting of the d orbitals and the five degenerate $3d$ states will split into a degenerate pair of e_g states and a degenerate triplet of t_{2g} states. In the doped $KNbO_3$ orthorhombic structure, the oxygen octahedron is distorted relative to that of the ideal cubic phase due to the dopant

at the Nb site undergoing a displacement along a dihedral of the octahedron. This implies that the above model of the crystal-field splitting is not strictly correct here, as the sets of symmetry-split defect states will not be exactly degenerate. A second, and as will be shown below, larger perturbation is the exchange splitting, whereby the Coulomb cost of pairing two electrons will make the filling of states with a single spin more energetically favorable, as this configuration will tend to keep the electrons more spatially separated.

The variation of the crystal field and exchange splittings across the $3d$ row is shown in Fig. 3.1 (d). In this figure the crystal-field splitting, Δ_{cf}, was determined by calculating the barycenters of the e_g and t_{2g} manifolds and subtracting, while the exchange splitting, Δ_{ex}, was determined by calculating the barycenters of the spin-up and spin-down 3-d manifolds and subtracting. As can be seen in Fig. 3(c), the crystal-field splitting only varies slightly across the 3-d row with a minimum at Fe^{3+} due to the close matching of the ionic radius of Nb^{5+} with that of Fe^{3+} and its adjacent neighbors in the row [79]. In contrast, the exchange interaction varies more strongly across the 3-d row, with a maximum at Fe^{3+} as expected because of its high spin configuration [80, 81]. As can be seen in Fig 3(c), the detailed interplay of the crystal-field and exchange splitting results in exchange being the dominant interaction with the notable exception of Cr^{3+}. Given the $[\mathrm{Ar}]4s^0 3d^3$ (less than half-full) electron configuration of Cr^{3+}, this effect does not impact the population energetics of the $KNbO_3$:Cr states.

3.1.3 Approximation to Actual Transition Energies

The single-particle electron states determined in standard Kohn-Sham DFT are the eigenstates of the fictitious non-interacting Kohn-Sham system, not the real physical system. As such, the excited states should not be expected to properly represent the excitation dynamics of the real systems. A fairly crude approximation to the actual excitation energies is performed here by assuming that a scaling relationship exists between the calculated and experimental band gap of $KNbO_3$ and that this scaling holds for the calculated and experi-

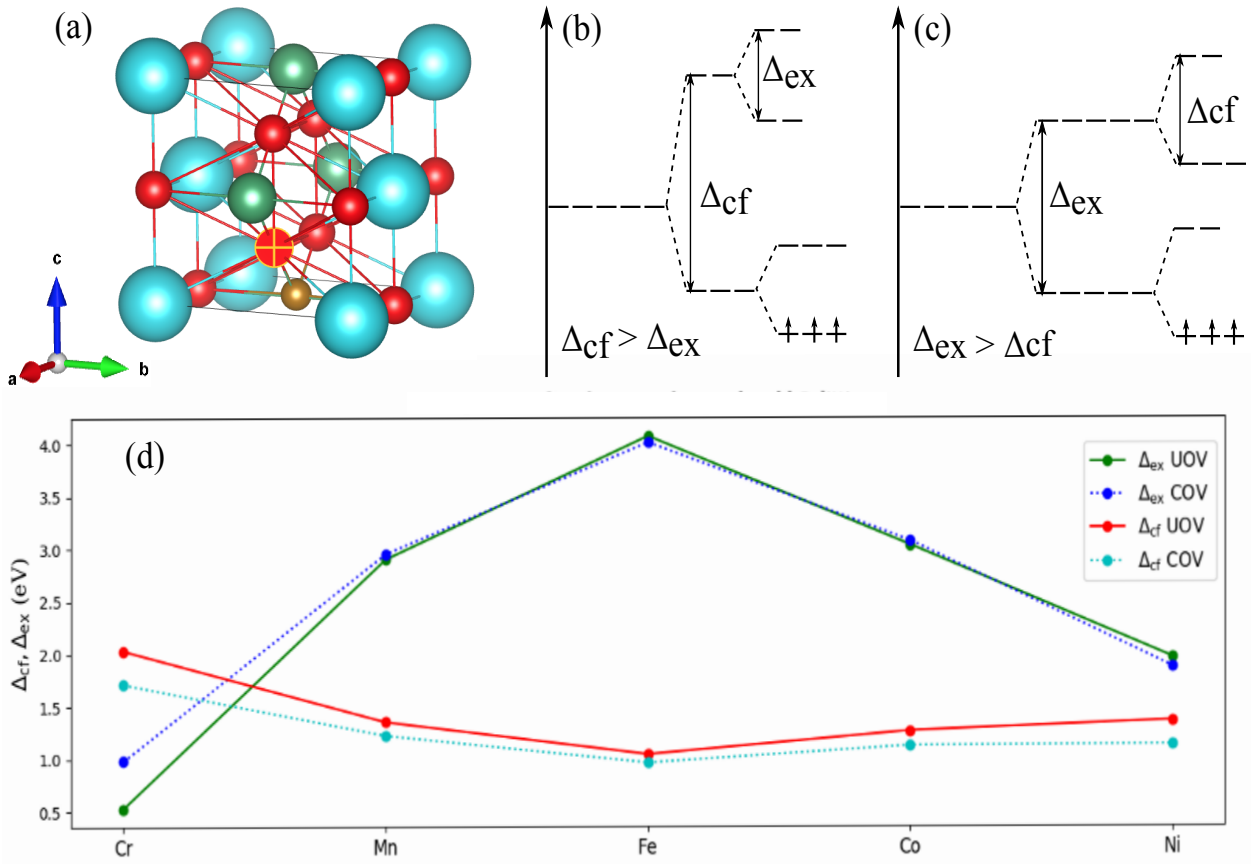

Figure 3.1: (a) Defect structure rendered in a single orthorhombic cell of $KNbO_3$. The cyan spheres are K, the green spheres Nb, the red spheres O, and the single orange sphere is the dopant species. The O atom with the cross-hair and yellow outline is the site of the vacancy. The arrows show the a, b, and c crystallographic axes of the cell. (b) Diagram of energy level splitting in the case where the crystal field splitting Δ_{cf} is the dominant perturbation. The occupation scheme of Cr^{+3} is demonstrated. (c) Diagram of energy level splitting in the case where the exchange splitting Δ_{ex} is dominant instead. The occupation scheme of Fe^{+3} is demonstrated (d) exchange and crystal-field splitting energies as determined by barycenters of peaks in pDOS for the uncoordinated oxygen vacancy (UOV) and coordinated oxygen vacancy (COV) models

mental impurity states as well. This approximation is somewhat reminiscent of the familiar "scissor operator" sometimes used in the literature. However, the scissor operator opens the band gap via a rigid shift in the conduction band relative to the valence band. The choice of where in the band gap to act the scissor operator is thus somewhat arbitrary and applying it to these systems would subsequently shift the impurity bands around arbitrarily. Our

current procedure does not generate a relative shift between the valence, conduction, or any impurity bands but merely scales them all according to the empirical band gap.

Using the value of $E_{\mathrm{g,exp}} = 3.3$ eV for the experimental band gap and $E_{\mathrm{g,calc}} = E_{\mathrm{LU-E_{HO}}} = 2.3$ eV for the calculated band gap (determined as the difference in energy between the lowest unoccupied state and the highest occupied state of the pure system), a ratio $R_{\mathrm{g}} = E_{\mathrm{g,exp}}/E_{\mathrm{g,calc}} = 1.4$ is obtained.

Given the high dopant concentration realized in these calculations, impurity bands, which would be expected to be extremely flat in experimental conditions, exhibit significant dispersion in these band structures due to interaction with their periodic replicas. The high dielectric constant of $KNbO_3$ is expected to significantly dampen long-range electrostatic interactions. This suggests that the unit cell size chosen for reasons of computational expense is sufficient. In analyzing these features, we find that the dispersion does not appear to affect the occupation scheme of any of the defect complexes. Hence, we feel it is sufficient for present purposes simply to calculate the barycenters of these states (or clusters of nearly degenerate states) using the pDOS information to approximate the position of the expected perfectly flat bands in an infinite supercell. These barycenters represent the average energy of the band as determined by integrating the state energies within an energy range capturing the feature of interest, weighted by the pDOS, and then normalized by the total number of states within the energy range. Upon applying the scaling factor R_{g} to these barycenters, the result was subtracted from the relevant band edge (conduction band edge for donor states and the valence band edge for acceptor states) to yield the approximated excitation energies reported later in Table II.

3.2 Results

The band structure and projected density of states (pDOS) of pure orthorhombic $KNbO_3$ are shown in Fig. 3.2. They reveal that the valence band of $KNbO_3$ largely arises from O-

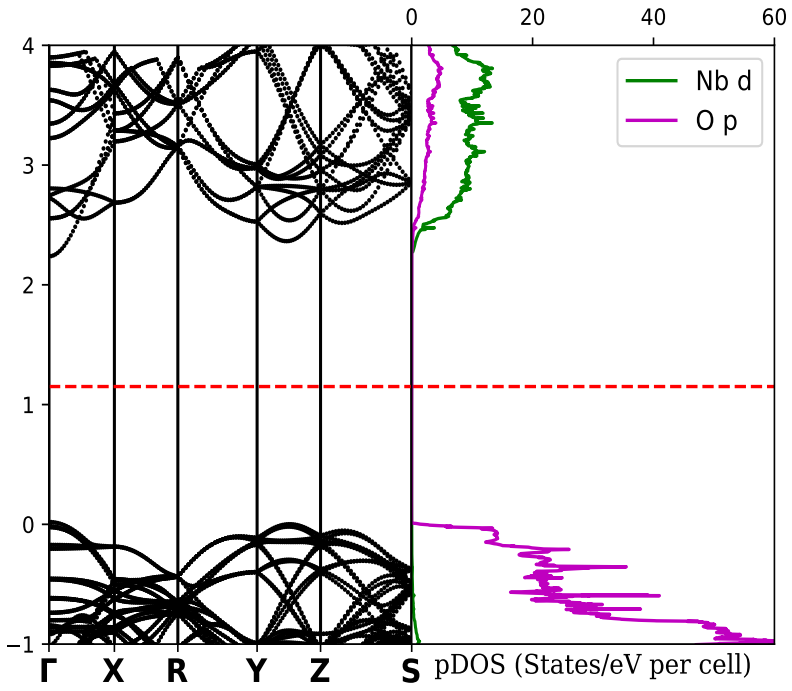

Figure 3.2: Non spin-polarized band structure and projected density of states for pure orthorhombic $KNbO_3$. The red line indicates the Fermi level, conventionally placed in the center of the band gap for semi-conducting systems.

$2p$ states while the conduction band arises from the Nb-$4d$ states, with some evidence of hybridization between the two sets of orbitals. This agrees with most explanations in the literature that the Nb-O bonds in $KNbO_3$ exhibit some covalence [82].

3.2.1 Ti and V

Presented here together are the results for $KNbO_3$:Ti and $KNbO_3$:V (see Fig. 3.3) both for the coordinated vacancy model. Given that these two dopants present the least useful changes to the band structure of the pure material we devote only a small amount of time to their discussion.

For the case of $KNbO_3$:Ti, the 3-d impurity states introduced are well above the band gap and thus present no viable means of extending the spectral range of $KNbO_3$. The presence of Ti induces partial occupation of the conduction band of $KNbO_3$ as indicated by the Fermi level, rendering the material metallic. This is perhaps unexpected considering conventional wisdom which would see a Ti substitution on Nb as donating a hole to the valence band. It is worth noting that the significant self-interaction error in DFT-GGA due to the localized d states of Ti may result in erroneous prediction of this metallic state, but the action of Ti as a hole dopant relies on Ti in a formal 4+ charge state. Thus, our coordination of this dopant with an O vacancy may be responsible for the divergence from this expectation. Given this behavior, even higher levels of theory are not expected to change these states enough to consider Ti^{3+} a viable dopant for photorefractive applications in the visible or IR regimes. The $KNbO_3$:Ti supercell yields an absolute magnetization of 0.0 Bohr magneton (μ_B) for the coordinated vacancy and 0.25 μ_B for the uncoordinated vacancy case.

In the case of $KNbO_3$:V, we see two degenerate states lying just below the conduction band with a less distinct peak of V-d character appearing in the conduction band. We identify these as the set of t_{2g} states while recognizing the presence of some V-d character diffusely present in the conduction band. In similar fashion to Ti, the presence of V shifts the Fermi level to the edge of the conduction band, rendering the material metallic within DFT-GGA. However, the Fermi level here appears to be pinned to these impurity bands. V should be capable of substituting isovalently on a Nb site. While not entirely apparent from the figure, the peaks in the V-d pDOS also show significant presence of O-2p character indicating some degree of hybridization between the impurity orbitals and the oxygen atoms nearby. It is possible that higher levels of theory applied to these results would actually result in shifts of the impurity states somewhere into the band gap. However, given that they occur so close to the conduction band edge already, those shifts are not expected to bring these states into the visible or IR regimes. The $KNbO_3$:V supercell yields an absolute magnetization of 2.37 μ_B for both the uncoordinated vacancy and coordinated vacancy cases.

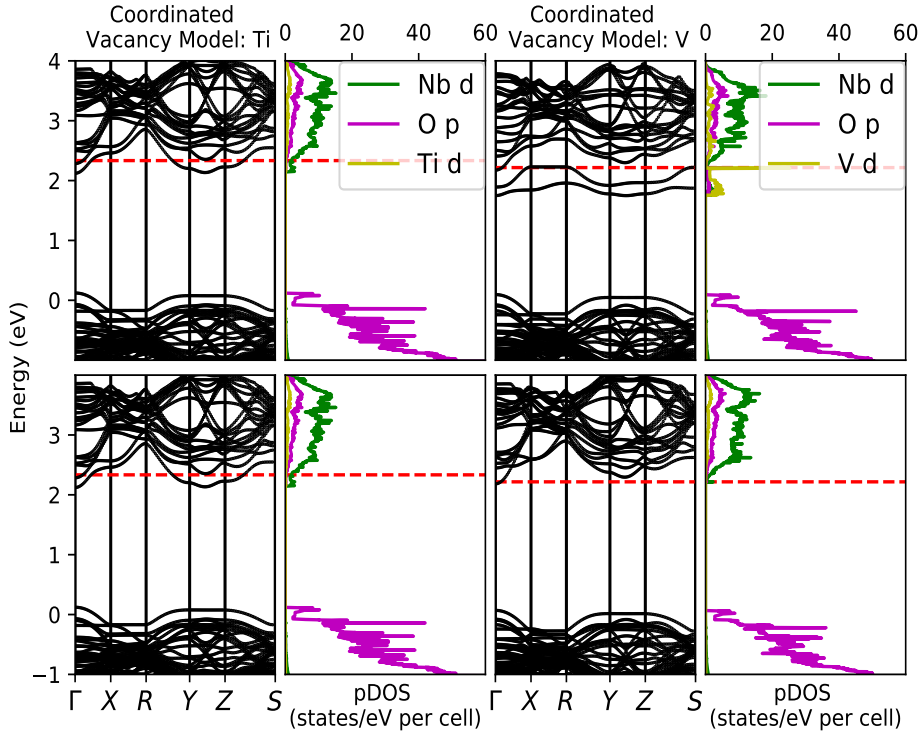

Figure 3.3: Spin-up (top) and spin-down (bottom) band structures and pDOS for $KNbO_3$:Ti (left) and $KNbO_3$:V for the coordinated vacancy model. The red line indicates the Fermi level as determined by density of states integration

3.2.2 Cr

Cr doping has been demonstrated to significantly affect the optical properties of other oxide systems [83] and is expected to represent a strong candidate for $KNbO_3$. The main impurity states of interest in $KNbO_3$:Cr (see Fig. 3.4) are a cluster of three nearly degenerate t_{2g} states located about mid-gap that occur only in the spin-up band structures. Comparing these features to those in the band structures of $KNbO_3$:Ti and $KNbO_3$:V, it is evident that increasing nuclear charge seems to shift the cluster of t_{2g} states to lower energy, as expected. Because these states are below the Fermi level, they may be useful as electron donors. Comparison of the uncoordinated and coordinated vacancy models reveals that the coordination of the O vacancy with the Cr impurity shifts an e_g state just barely down in energy out of the

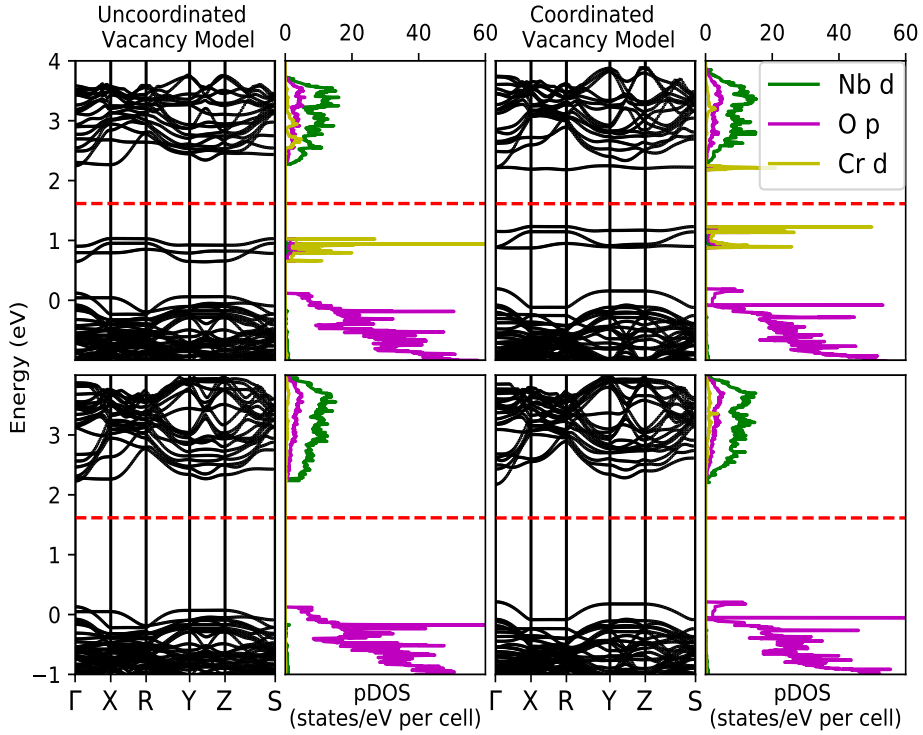

Figure 3.4: Spin-up (top) and spin-down (bottom) band structures and pDOS for $KNbO_3$:Cr for the uncoordinated vacancy (left) and coordinated vacancy (right) models. The red line indicates the Fermi level as determined by density of states integration

conduction band. One noticeable effect of the local vacancy is that it results in a re-ordering of the t_{2g} states. Comparing to the uncoordinated vacancy model, it is apparent that one of the least dispersive t_{2g} states becomes more energetically favorable relative to the most dispersive t_{2g} state. Given that the position of the cluster is practically the same across these two models in $KNbO_3$:Cr, this re-ordering likely arises from the perturbation of the crystal field since the coordinated vacancy further breaks the symmetry of the octahedron. The $KNbO_3$:Cr supercell yields an absolute magnetization of 3.34 μ_B for the coordinated vacancy case and 3.38 μ_B for the uncoordinated vacancy case.

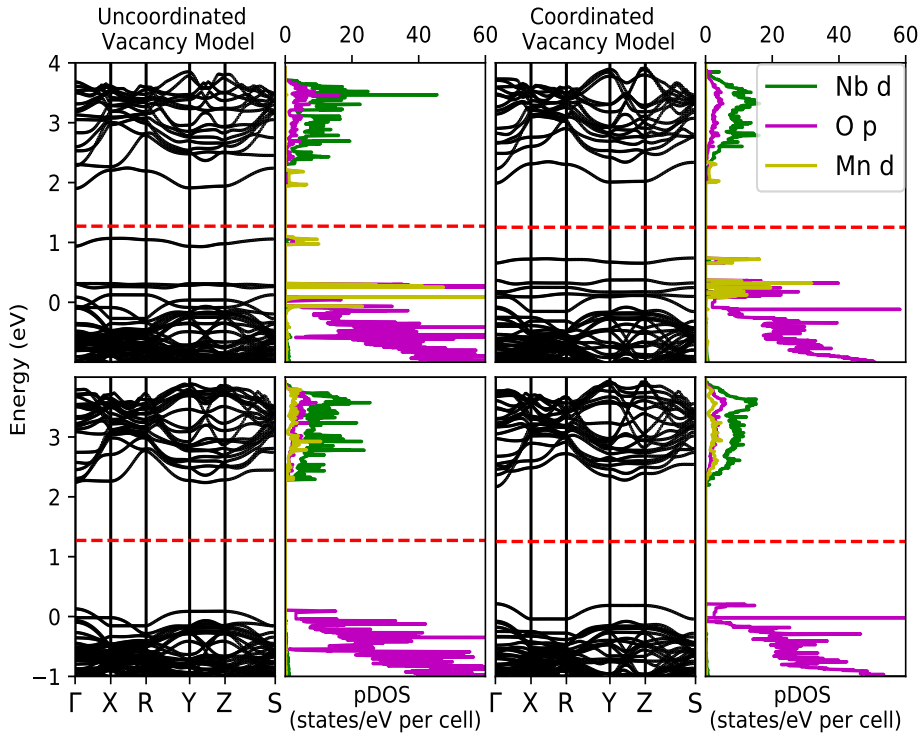

Figure 3.5: Spin-up (top) and spin-down (bottom) band structures and pDOS for $KNbO_3$:Mn for the uncoordinated vacancy (left) and coordinated vacancy (right) models. The red line indicates the Fermi level as determined by density of states integration

3.2.3 Mn

The band structure of $KNbO_3$:Mn (see Fig. 3.5) becomes quite complex. While the useful impurity features from an experimental point of view appear in the spin-up band structures, we observe the beginnings of more noticeable Mn 3-d features in the conduction bands of the spin-down band structures. The $3d$ character of these spin-down features is reminiscent of the diffuse character present in the dopants examined above; however, particularly in the uncoordinated vacancy model, peaks in the pDOS indicate that more localized bands are beginning to present. Examining the spin-up features we notice that the set of occupied t_{2g} states present in the previous systems decreases in energy such that it occurs just barely above the O-$2p$ states making up the valence band. These states, while not useful as midgap

states, may still be useful as trap states. Lying higher in energy, the e_g states straddle the Fermi level in each model. In the coordinated vacancy model, the separation of the two e_g states is increased, shifting the lower lying state closer to the valence band edge. The band re-ordering discussed in the results of $KNbO_3$:Cr is more difficult to examine here. The cluster of t_{2g} states exhibits significant overlap with states of O-2p character and it is difficult to identify the most and least dispersive bands from the figure. However, by examining the norm of the wave functions of these bands, it is confirmed that re-ordering of the t_{2g} states due to the coordinated oxygen vacancy does occur in $KNbO_3$:Mn as well, though it is more complicated than a simple switching of two states. The $KNbO_3$:Mn supercell yields an absolute magnetization of 4.22 μ_B for the coordinated vacancy case and 4.24 μ_B for the uncoordinated vacancy case.

If this occupation scheme would persist at higher levels of theory, it would indicate dual acceptor/donor states in the system. The presence of both types of defect states would make possible both electron and hole excitations. In the context of photorefractive applications, the simultaneous presence of electrons and holes as charge carriers can generate competing gratings in the material thereby weakening the photorefractive response. Thus Mn may be useful in more specialized applications where both electron and hole gratings are desired.

3.2.4 Fe

$KNbO_3$:Fe (see Fig. 3.6) demonstrates a significant change in the band structures observed so far. It is the first system where impurity gap states occur in the spin-down band structures of our models. In particular, the impurity states in the spin-up band structures are just above or within the valence band, rendering them less suitable for the purposes of visible/IR photorefractivity. Examining the spin-down band structures, we note that the separation between the e_g and t_{2g} states seems to decrease relative to previous dopants. While the impurity states in the conduction band are difficult to distinguish, peaks in the pDOS plots seem to show a similar trend across the uncoordinated and coordinated vacancy models

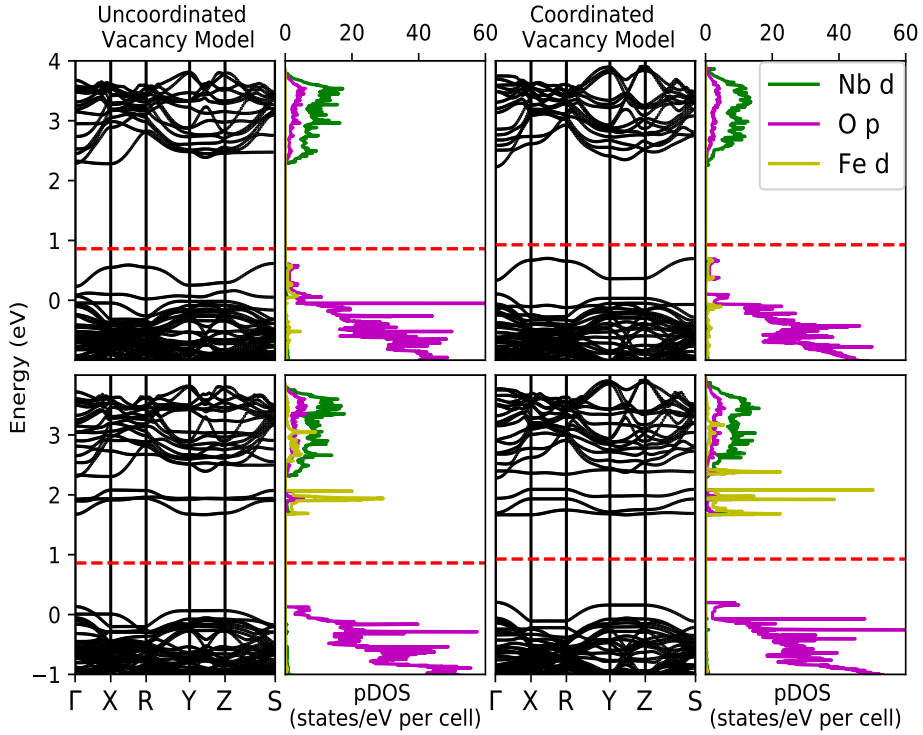

Figure 3.6: Spin-up (top) and spin-down (bottom) band structures and pDOS for $KNbO_3$:Fe for the uncoordinated vacancy (left) and coordinated vacancy (right) models. The red line indicates the Fermi level as determined by density of states integration

as before, one e_g state being more strongly affected by the presence of the vacancy than the other. We similarly recognize that continuation of the trend of the most and least dispersive t_{2g} bands re-order under the presence of a nearest-neighbor oxygen vacancy. The cluster of t_{2g} states in the spin-down midgap are unoccupied, indicating that Fe may introduce acceptor states viable for extending the spectral response of $KNbO_3$ via hole conduction. The $KNbO_3$:Fe supercell yields an absolute magnetization of 5.06 μ_B both vacancy cases.

3.2.5 Co

$KNbO_3$:Co (see Fig. 3.7) shows another noticeable difference from previous systems. Of immediate note is that there is a significant difference in the Fermi level between the two

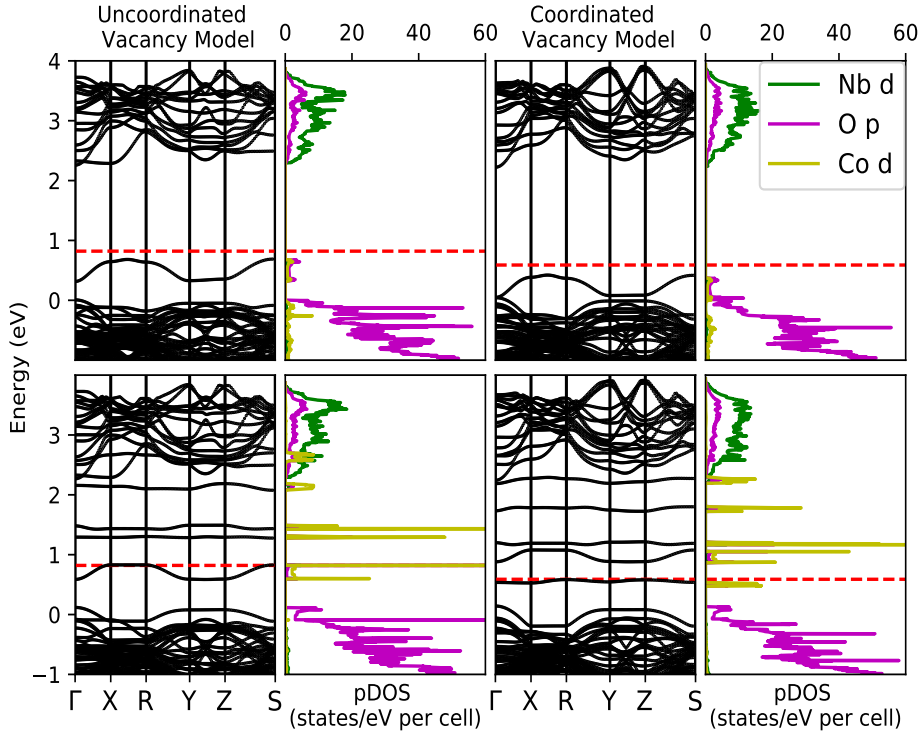

Figure 3.7: Spin-up (top) and spin-down (bottom) band structures and pDOS for $KNbO_3$:Co for the uncoordinated vacancy (left) and coordinated vacancy (right) models. The red line indicates the Fermi level as determined by density of states integration

models. This is a consequence of the band re-ordering pattern that has been established thus far. In $KNbO_3$:Co however, the near-degeneracy of the t_{2g} states is significantly disrupted, even in the uncoordinated vacancy model. The system supports the occupation of the lowest-lying t_{2g} state, but given that the lowest lying state switches from the most dispersive state (in the uncoordinated vacancy model) to the least dispersive state (in the coordinated vacancy model), the Fermi level noticeably shifts as well. However, this shift of approximately 0.23 eV is very similar to the energetic dispersion of the band, which is approximately 0.3 eV. Additionally, in the coordinated vacancy model the presence of the vacancy seems to shift all of the relevant impurity states down in energy relative to the uncoordinated vacancy model. Given the spread of impurity states across the band gap, it is likely that the introduction of

Co as a dopant will introduce viable acceptor/donor states. The precise shift of these levels under higher levels of theory is difficult to predict, however. The $KNbO_3$:Co supercell yields an absolute magnetization of 4.11 μ_B for the coordinated vacancy case and 4.13 μ_B for the uncoordinated vacancy case.

3.2.6 Ni

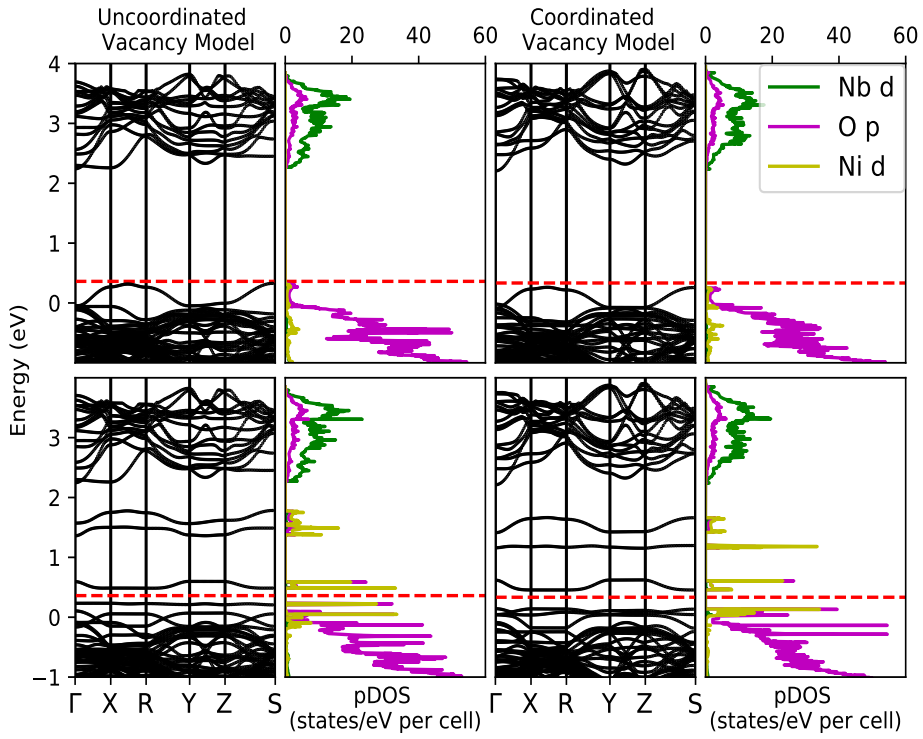

Figure 3.8: Spin-up (top) and spin-down (bottom) band structures and pDOS for $KNbO_3$:Ni for the FN (left) and coordinated vacancy (right) models. The red line indicates the Fermi level as determined by density of states integration

The general Coulombic trend of impurity states lowering in energy with increasing dopant atomic number continues in $KNbO_3$:Ni (see Fig. 3.8). The useful impurity features are only exhibited in the spin-down band structures. Interestingly the two lowest t_{2g} states are occupied and fall well into the valence band. As with $KNbO_3$:Mn, the reordering of the t_{2g} states across the two models is difficult to examine given that the two states in question are within the valence band. Examination of the norms of the wave functions for these bands also

confirm that this reordering does occur as well. The unoccupied e_g states in $KNbO_3$:Ni fall relatively deep within the band gap. These states along with the highest t_{2g} state, which is also unoccupied, may represent a useful set of acceptor states for establishing photorefractive gratings via hole conduction. The $KNbO_3$:Ni supercell yields an absolute magnetization of 3.13 μ_B for the coordinated vacancy case and 3.14 μ_B for the uncoordinated vacancy case.

3.3 Discussion

Shown in Table 3.1 are the predicted transition energies of the identified impurity $3d$ states calculated using the scaling procedure discussed above. They are further identified with their nature as an acceptor or donor state as determined by the occupations indicated by the bands structures presented in the previous section. For the purposes of extending the spectral response of $KNbO_3$ into the visible and IR regimes, transition energies from 2.8 eV to approximately 1.7 eV could be considered viable. Given this range, most of the dopants studied here present at least one state that could extend the spectral response of $KNbO_3$. Based on the pure number of viable states available it would appear Co and Ni present as perhaps the best candidates at first glance.

3.3.1 Limitations

Several limitations ought to be emphasized lest readers over interpret the quantitative predictions. As stated earlier, Kohn-Sham DFT lacks the theoretical basis to accurately calculate excited states. The same shortcoming which prevents the accurate prediction of semiconductor band gaps also affects the energy (and possibly the occupations) of impurity states. Secondly, localized d orbitals can be subject to self-interaction and localization error [84], which can affect their energies and occupations. Corrections to these localization errors, such as the introduction of a Hubbard U parameter can likely help abate some of these issues [85]. The Hubbard U correction, in principle, introduces an extra term in

Impurity	Scaled Transition Energy (eV)			
	Uncoordinated Vacancy		Coordinated Vacancy	
	Spin up	Spin down	Spin up	Spin down
Cr	1.1 (Donor)	–	0.97 (Donor) 3.1 (Acceptor)	–
Mn	2.9 (Donor) 1.9 (Donor) 2.8 (Donor)	–	3.1 (Donor) 2.3 (Donor) 2.9 (Acceptor)	–
Fe	3.2 (Donor) 2.8 (Donor)	2.5 (Acceptor)	2.6 (Donor)	2.3 (Acceptor) 2.7 (Acceptor)
Co	1.7 (Donor)	1.2 (Acceptor) 1.3 (Acceptor) 2.0 (Acceptor)	2.2 (Donor)	1.9 (Donor) 0.86 (Acceptor) 1.1 (Acceptor) 1.7 (Acceptor) 2.1 (Acceptor)
Ni	3.2 (Donor)	3.1 (Donor) 0.59 (Acceptor) 1.9 (Acceptor) 2.2 (Acceptor)	3.2 (Donor)	3.3 (Donor) 0.59 (Acceptor) 1.5 (Acceptor) 2.02 (Acceptor)

Table 3.1: Transition energies of impurity features predicted by scaling method. Behavior as either an acceptor or donor state is determined by occupations (position below or above the Fermi level). Barycenters of the pDOS peaks were calculated for distinguishable features in order to compensate for the artificial dispersion introduced by interactions of the dopant with its periodic images.

the Kohn-Sham Hamiltonian to account for the on-site Coulomb interaction responsible for the correlation-induced gap in strongly correlated materials. In effect, this term penalizes inaccurate occupation of the states in question. This parameter is, however, often introduced only in a semi-empirical sense in that it is adjusted to open the band gap to the experimental value or to achieve a desired occupation of problematic highly-localized orbitals. While there exist methods of determining the appropriate value of the U parameter self-consistently [86, 87], they were not implemented here given that the systems under study did not exhibit

partial occupation of the d orbitals. Examination of the band structures produced in this study show occupations of the relevant dopant d orbitals consistent with the intuition of the model, thus, it was deemed that the extra computational complexity and expense involved in the introduction of a Hubbard U correction would not provide sufficient improvement to these results to warrant the additional time and resources.

Hybrid functionals are another means to open the band gap from its DFT value [88]. These functionals, which include some empirically chosen percentage of exactly calculated exchange energy, have demonstrated very good results in a wide variety of materials. This added precision comes with a drastic increase in computational cost however, and while its effect on the band gap is well-demonstrated, there is no a priori reason that it should be expected to affect the occupation of these d orbitals.

Given the aim of our methodology in this study, which is to examine a suite of dopant defect complexes with as little computational expense as possible to get qualitative trends, we find that well-converged GGA results are an adequate starting point to examine these trends for the defect complexes studied here. We emphasize that this may not be a "one size fits all" approach and that methods such as those mentioned above may be useful to take into consideration for other systems.

It should also be addressed that strongly correlated materials are characterized by a correlation-induced opening of an energy gap at the Fermi level and the presence of satellite structures within the electronic spectrum. As with most exchange-correlation functionals in Kohn-Sham DFT, the GGA is not capable of describing these phenomena. Further calculations capable of accounting for correlation-related phenomena, under the GW approximation for example, can be undertaken for promising systems. These more sophisticated and computationally expensive methods build perturbatively on Kohn-Sham DFT results such as those collected here.

Additionally, photorefractive models involve complicated rate equations that rely on various complicated parameters. Cross-sections of the various excitations, the density of trap

states, recombination rates and hole and electron mobilities all play into the solutions of these equations and our calculations do not address these issues. Here, we have considered only one charge compensating mechanism.

The size of our supercell was also chosen in an attempt to balance the computation cost of these calculations with some degree of accuracy. Larger supercells could be employed to further reduce the interactions of periodic replicas of the defect complexes. However, the use of larger supercells can be difficult to justify when GW calculations are the ultimate end, given that the scaling of those calculations can be prohibitive for large system sizes. Finally, our scaling method implemented here provides a fairly simplistic mechanism to deal with the DFT errors in excited states. As such, we caution against over interpreting the quantitative results. The goal of this study was to examine trends and identify candidate systems for deeper investigation.

3.4 Conclusions

Our results indicate that for these two relatively simple models of charge-compensated defect complexes in doped $KNbO_3$, trends can already be observed in the occupation, splitting, and energetics of impurity states as the atomic number of the dopant is altered across the $3d$ row of transition metals. We find that the distribution of these impurity states changes quite systematically across the $3d$ row. As the atomic number of the dopant is increased, the impurity states are gradually pulled down in energy. The relative effects of exchange splitting and crystal field splitting are seen to affect the ordering and spin state of impurity gap states. We also observe that a coordinating oxygen vacancy can influence the local lattice geometry significantly, which can affect the local crystal field, evidenced by the reordering of some of the t_{2g} states. These results also qualitatively exclude Ti and V as useful dopants for the purposes of IR response. While quantitatively accurate statements about optical response (and furthermore photorefractive response) cannot be made at this level of theory,

our results do indicate that Cr, Fe, Co, and Ni all stand out as distinctly interesting candidate dopants, albeit in different spectral regions and with different donor/acceptor state behaviors, and therefore photorefractive applications.

3.5 Copyright Information

Chapter 4

Investigation of Fe Defect Centers in $KNbO_3$

The nature and behavior of defect complexes in $KNbO_3$ is the subject of continuing debate [73, 74, 76, 89]. In particular, the behavior of defects in different charge states can be vital to experimental applications of this material. Of particular relevance to the photorefractive effect is whether these different charges states exhibit donor or acceptor behavior as this will subsequently relate to the effective trap density and dark conductivity in a material. Both of these factors can greatly impact the photorefractive performance as a high effective trap density can result in larger photorefractive gain and a high dark conductivity will contribute to a shorter grating lifetime. Thus, obtaining a fundamental and detailed understanding of adjacent valence states of dopant-defect complexes is another opportunity to leverage the power of *ab initio* methods to help inform experimental designs. In this chapter we explore such a methodology that extends the capabilities offered by surveys such as the one described in the previous chapter to investigate a range of experimentally relevant, complex defect centers in multiple physically occurring ionization states. In particular, it has been shown that doping $KNbO_3$ with Fe can dramatically increase the photorefractive speed and sensitivity increasing its efficacy for two-beam coupling applications [48]. This improvement

in photorefractive performance can be effected by using a reduction treatment that converts a fraction of the Fe^{3+} ions to Fe^{2+} state. In this chapter, we study Fe as a substitutional impurity on both K and Nb sites in different local environments (defect complexes) and ionization states.

4.1 Defect Complexes

The 2016 results by Basun *et al.* [48] indicated three possible defect complexes in $KNbO_3$, a lone Fe substituting on a Nb site (Fe[Nb]), an Fe substitution on a Nb site with a coordinating O vacancy (Fe[Nb] – V_O), and an Fe substitution on a K site with a nearest-neighbor K vacancy (Fe[K] – V_K). One result of that study is a schematic energy level diagram they propose to explain the effects of reduction treatments on the optical properties of the crystal, reproduced here as Fig. 4.1.. They remark that while both substitution-vacancy complexes are possible, their work supports the Fe[K] – V_K complex as the dominant photorefractive center.

In this study, we employ the DFT-based methodology below to assign each of the defect states in Fig. 4.1 in order to more fully flesh out this picture. We consider all defect complexes with Fe in a 2+/3+ charge state transition only, as the literature indicates this is the predicted most relevant photoactive center for photorefractive purposes.

4.2 Methodology

Much of the methodology used to investigate the Fe doped systems of interest to this chapter has been described in Chapters 2 and 3. Below we describe briefly two additional techniques that have been used to gain additional insight.

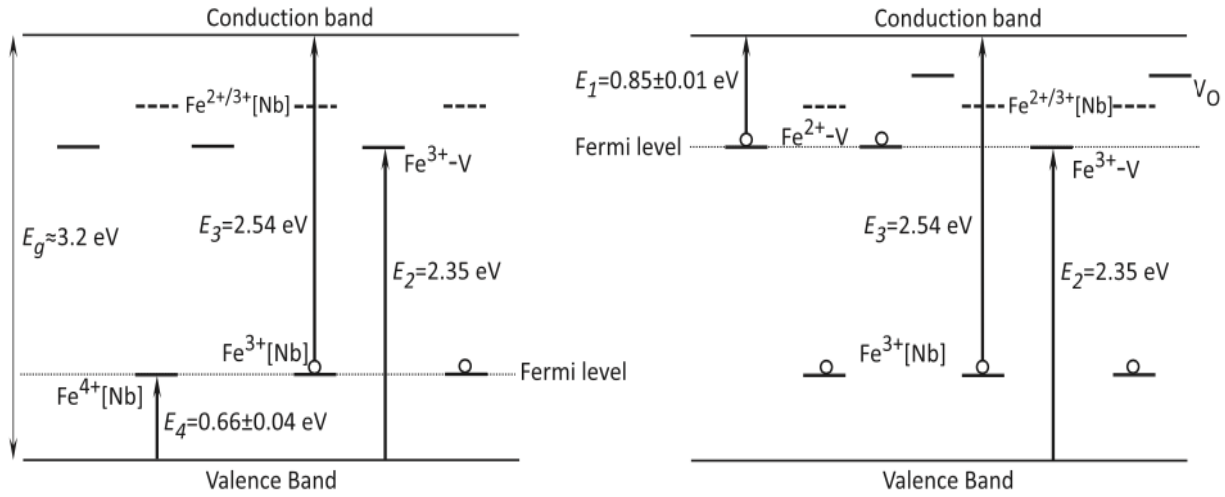

Figure 4.1: Reproduction of a diagram from Basun *et al.* [48] showing proposed energy level diagrams inferred from their results, in the spirit of work done by Zgonik *et al.* [90]. The left part of the diagram shows proposed energy levels and occupations in crystal which were grown and poled, but not reduced. The right part shows the proposed changes to occupations induced by reducing the crystal after growth and poling. The Fe^{X+} – V levels shown here refer to the Fe atom in an X+ charge state in a complex with a vacancy. The dashed lines represent $Fe^{2+/3+}$[Nb] levels proposed by Zgonik *et al.*. The V_O level is taken from the work of Biaggio *et al.* [91]. Reprinted figure with permission from [Basun *et al.*, Physical Review B, 93, 094102 2016.] Copyright (2016) by the American Physical Society.

4.2.1 Addition of a Jellium Background Charge

Upon identifying candidate systems of interest, investigating the behavior of valence states adjacent to the initially enforced state can be undertaken by doping the unit cell with additional charges and determining the occupation of impurity states by inspection of the band structures, as performed in the previous chapter. The introduction of extra charge into a cell with periodic boundary conditions causes a divergence in the electrostatic energy. This can be mitigated by introducing a compensating jellium background equal and opposite of the included charge. In this way, we can dope extra electrons or holes into the unit cell and examine on which atomic site they prefer to localize by examining how the occupations of impurity states in the band structure change with respect to the Fermi level.

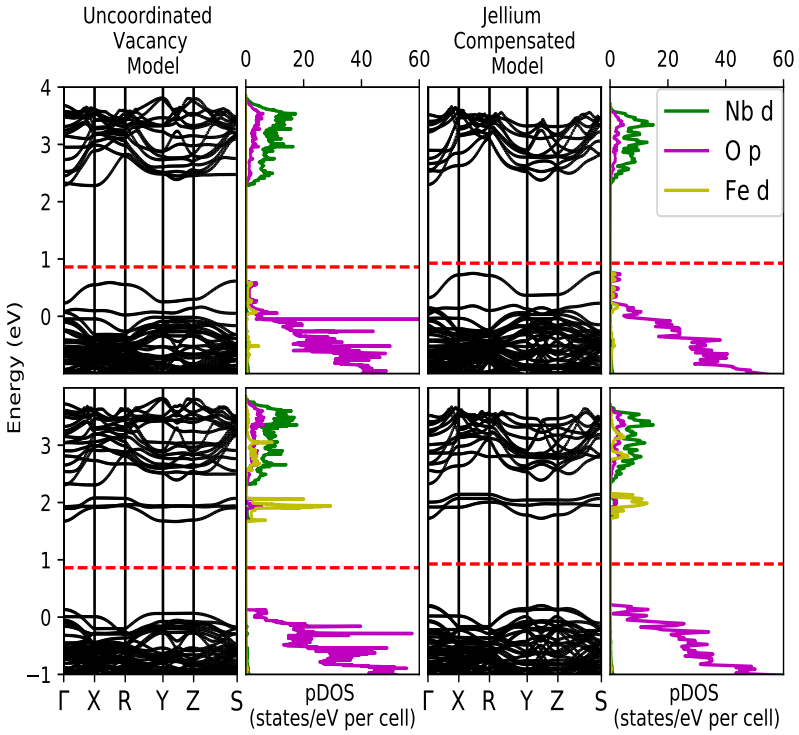

Figure 4.2: A demonstration of the effect of jellium compensation on a band structure. Both systems are Fe^{3+} substituting on a Nb site. In the left panel we show the band structure of the system when charge compensated by an O vacancy as far away in the supercell as possible. In the right panel we show the band structure when the system is externally charged and compensated by a jellium back ground. Though qualitative differences do exist between the two, we see that the use of a jellium background produces the same charge state of the dopant and behavior of the impurity bands.

In Fig. 4.2 we compare band structures for an Fe^{3+} on the Nb site with a second-nearest neighbor charge compensating O vacancy with a Fe^{3+} on a Nb site with a compensating jellium background charge. As can be seen in Fig. 4.2, the two band structures are nearly identical showing that use of a compensating jellium background charge provides an excellent proxy for non-local charge compensation models.

Again, we use GGA exchange correlation functionals with optimized norm-conserving Vanderbilt pseudopotentials. Quantum Espresso is again our tool of choice, albeit we now use version 6.6, which was released after the previous work was completed. We also employ

supercells constructed similarly to those of the previous chapter, that is 2×2×2 supercells of the pure orthorhombic structure relaxed to established equilibrium lattice parameters. That structure is then defected and subsequently the ions are allowed to relax to respond to the defect complex and any additional charge introduced, establishing equilibrium atomic positions. All other convergence parameters (kinetic energy cutoff, charge density cutoff, k-point grid, and Gaussian convolution width) are the same as in the previous study.

4.2.2 Configuration Coordinate Diagrams

A system of atoms with a dopant in a given valence state will, by necessity have a different set of equilibrium atomic positions than an otherwise identical system with the dopant in a different valence state. The electron density will respond to the presence of "extra" holes or electrons, relative to the system with the other valence state, and the atomic configuration will subsequently respond to the altered electronic density, as implied by the first Hohenberg-Kohn Theorem. The two systems will also display different electronic excitation levels. These levels will also be linked in some sense, given that excitation to or from an impurity state in one charge state will result in a change of the charge state of the dopant.

One useful construction to visualize the interplay between these sets of electron energy levels is through configuration coordinate diagrams. In this formalism, the electronic levels are coupled to the states of a harmonic oscillator and each electronic energy level depicted as a harmonic oscillator potential with a ladder of quantum harmonic oscillator states parameterized by a generalized coordinate, Q. For the case of oxide perovskites, the configuration coordinate is usually taken to be the breathing mode of the oxygen octahedron. In this chapter, we will consider equilibrium structures for defect complexes in two different charge states. As can be seen in Fig. 4.3, for the Fe[Nb] and Fe[Nb] – V_O, this will correspond quite well to a mode related to the motion of the oxygen octahedron, though it will not be a perfect mode coordinate. This will also correspond to a type of mode coordinate in the Fe[K] – V_K complex but with a different geometry of nearest neighbors about this site. We

note that while the A site cation in a cubic oxide perovskite would be cubically coordinated with 8 O atoms, our relaxations indicate that Fe displaces relative to this ideal position to push it into a more octahedral coordination. This significant change in geometry results in the Fe ion having its 5 nearest neighbors as O atoms which form an octahedral complex around it, albeit with an apical oxygen "missing".

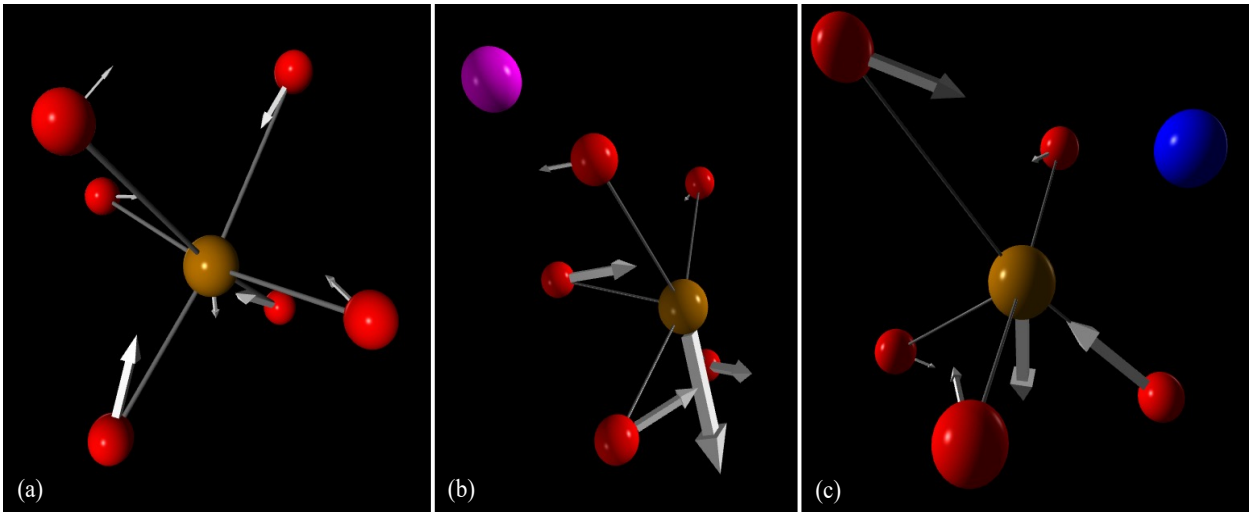

Figure 4.3: Visualizations of the (a) Fe[Nb], (b) Fe[K] – V_K, and (c) Fe[Nb] – V_O defect complexes studied in this work. Each panel only displays the nearest O neighbors of the dopant ion, with small red spheres representing the O atoms, and the orange sphere representing the Fe atom. In cases where a vacancy is present, either a blue or purple sphere has been added which indicates the approximate location of the O vacancy or K vacancy, respectively. White arrows indicate the direction of displacement of each of these atoms in going from the equilibrium structures with the Fe in the 2+ state to those with Fe in the 3+ state and are scaled for visibility in each case separately, though relative lengths of arrows for any panel do correspond to the amount of displacement of the ion. In panel (b), the apparent octahedral coordination is not accidental. This is explained in the relevant section of the text.

Consider the case of an Fe atom substituting on a Nb site. If for instance, the Fe assumes a 3+ charge state, then the energy of the Fe-3d impurity states introduced into this system can be thought of as a set of parabolae whose minima all occur at the set of atomic coordinates

corresponding to the equilibrium configuration of the system. Perturbing these coordinates means that the ground state electronic wavefunction of the unperturbed system will no longer be the lowest energy state, hence the displacement from the minimum of the parabola. In this case, the set of parabolae corresponding to the electronic levels of the system with the Fe ion in a 2+ charge state will have their minima occurring at the corresponding equilibrium atomic configuration. Thus, plotting the energy against this configuration coordinate gives us a picture of how charge transfer processes from one valence state to another may occur.

Configuration coordinate diagrams are useful tools for describing strongly coupled electron-phonon systems, such as transition metal ions with $3d$ electrons, like our target systems here. Qualitatively, configuration coordinate diagrams provide an explanation for the "Stokes shift" between absorption and emission spectra. Given accurate enough configuration coordinate parameters, vibronic sideband spectra (with Pekarian line shapes) can be calculated [92]. Configurational coordinate diagrams can also be used to determine nonradiative relaxation pathways as shown schematically in Fig. 4.4. In addition, if the configuration coordinate diagrams are constructed for different ions, they can also be used to explore and characterize metal to metal [93] and metal to ligand charge transfer transitions [94].

To construct these sorts of diagrams via *ab initio* methods we use the following method. We first find the equilibrium structures for a defect complex in two adjacent charge states, in this work Fe^{3+} and Fe^{2+}. The configurational coordinates for these equilibrium structures are denoted Q_0 and Q_N, respectively. We then create $N-1$ intermediate structures by linearly interpolating between these two sets of coordinates to generate a sequence $Q_i = Q_0, Q_1, Q_2, \ldots, Q_N$, which becomes our generalized coordinate. In this work we use $N = 6$. For each value of Q_i we calculate both the band structure and pDOS for both the 3+ and 2+ ionization states. This allows us to confirm whether or not the predicted occupation scheme is being achieved and, by extension, the charge state of the dopant. With this data for each structure, we can extract the electronic energy levels at a high-symmetry point in the Brillouin zone, the Γ-point for instance, and plot these energies against our configuration

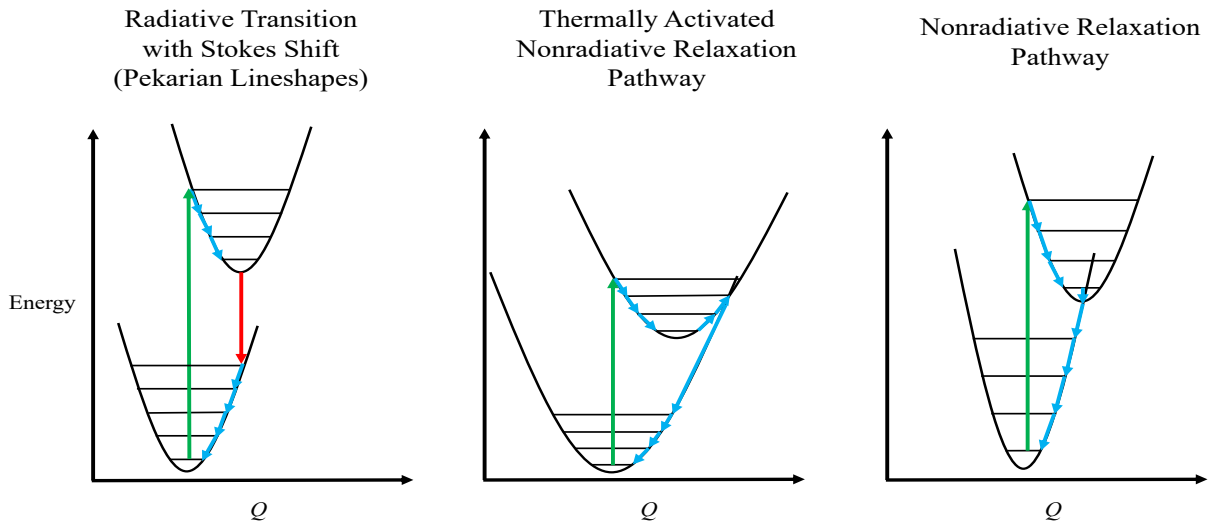

Figure 4.4: A schematic representation of several configuration coordinate diagrams. Panel (a) shows the radiative transition of an electron from the ground state to an excited vibrational state of the system at a different configuration coordinate. The electron then nonradiatively de-excites to the new ground state of the perturbed system, and then radiatively recombines to an excited vibrational state of the unperturbed structure, though now emitting a lower frequency photon than it absorbed, demonstrating what is known as a Stokes shift in its spectrum. Panel (b) shows a system where the electron transitions to a vibrational state of the perturbed system, nonradiatively relaxes to the new ground state and is then thermally excited to another vibrational state of the new perturbed system. Due to over lap in the wavefunctions of this thermally excited state and a vibrational excited state of the unperturbed structure, this electron may transition to the excited state of the unperturbed system and relax to the previous ground state, demonstrating absorption, but now with thermally activated nonradiative relaxation. Panel (c) shows another nonradiative relaxation pathway. In this case, the vibrational ground state wavefunction of the perturbed system overlaps with a vibrational excited state of the unperturbed system, and the electron may absorb a photon, relax to the perturbed ground state and then transition to the vibrational excited state of the unperturbed system without need of an intermediate thermal excitation.

coordinate (now labeled by Q_i). We then choose to linearly interpolate between these points as a guide to the eye.

4.3 Results

4.3.1 Fe[Nb]

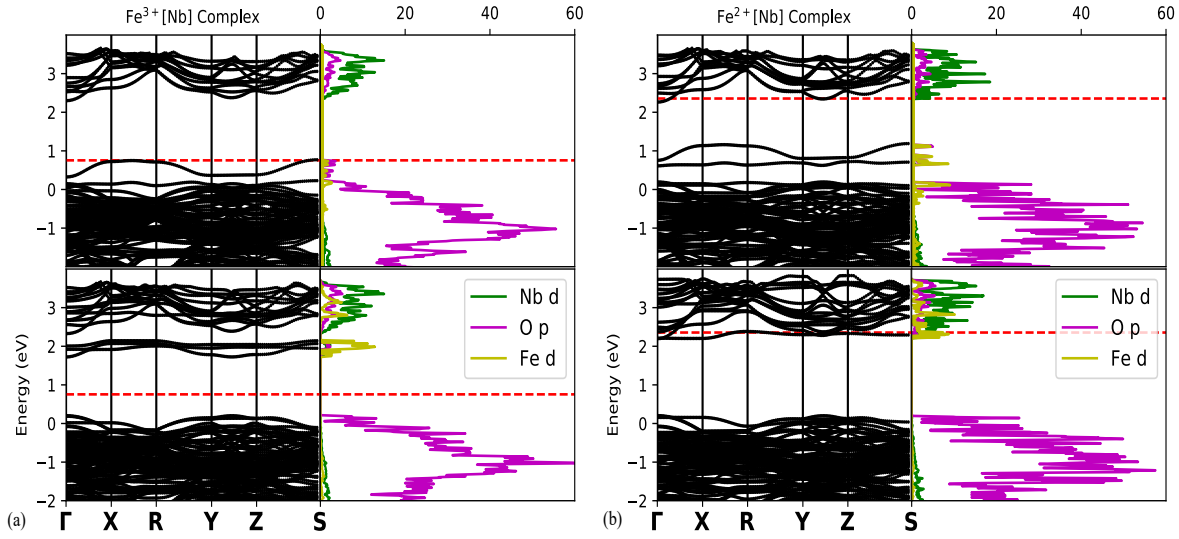

Figure 4.5: Band structure and pDOS of the Fe[Nb] complex with the Fe ion (left) in a 3+ charge state and (right) in a 2+ charge state. For each charge state, the atomic positions are those of the respective equilibrium structure. The top and bottom panels show the spin-up and spin-down quantities, respectively. The dashed red line indicates the position of the Fermi level as determined by integration of the density of states.

Fig. 4.5(a) shows the band structure and pDOS of the Fe[Nb] system with the Fe in a 3+ charge state, with the atomic positions at their corresponding equilibrium positions. Given the position of the Fe-3d states with respect to the Fermi level, we can confirm that Fe is, in fact, in a 3+ charge state. In the spin-up band structure we notice an occupied, slightly dispersive e_g state positioned just above the valence band. The crystal-field split partner to this e_g state lies within the valence band almost exactly at the band edge. The pDOS plots indicate that these e_g states may be hybridized to some extent with the O-2p states which

make up the majority of the valence band. Examining the spin-down band structure, we see a cluster of three unoccupied, nearly-degenerate t_{2g} states within the band gap, though positioned closer to the conduction band edge. We see also from the projected densities of states that the spin-down e_g states are situated well within the conduction band, though the sharpness of the peaks indicates they are quite localized.

In examining panel (b) of Fig. 4.5 we notice that the inclusion of an extra electron, along with the relaxation to the equilibrium atomic positions, raises the energies of all Fe $3d$ states. In the spin-up distribution, we notice that the e_g states increase in energy. The more dispersive e_g state in the 3+ case remains occupied, but now its partner state emerges completely from the valence band. Evidenced by the spin-up pDOS, we now find a t_{2g} state at the valence band edge. A somewhat more drastic change is shown in the spin-down band structure. All but the lowest-lying spin-down t_{2g} states are pushed into the conduction band, with this lowest state now falling just below the Fermi level, indicating its occupation, and hence the Fe 2+ charge state. The Fermi level itself is also pushed up to just barely cut through the bottom of the conduction band.

4.3.2 Fe[K] – V_K

Fig. 4.6 shows the band structure of the Fe[K] – V_K complex with the Fe ion in a 3+ charge state at its equilibrium coordinates. Contrary to the previous complex, we see no midgap impurity states in the spin-up band structure. The presence of an e_g band in the valence band can be inferred from the pDOS and by the pattern established with the previously investigated systems.

The spin down structure reveals two interesting features. The first is a triplet set of unoccupied t_{2g} states. As discussed above, despite K sites being cubically coordinated with 8 O atoms in the ideal perovskite structure, the relaxation of the system tends to prefer placing the Fe dopant into a system bearing some resemblance to the octahedral coordination with 5 of the nearest neighbor O atoms. The near degeneracy of the t_{2g} states indicates that in this

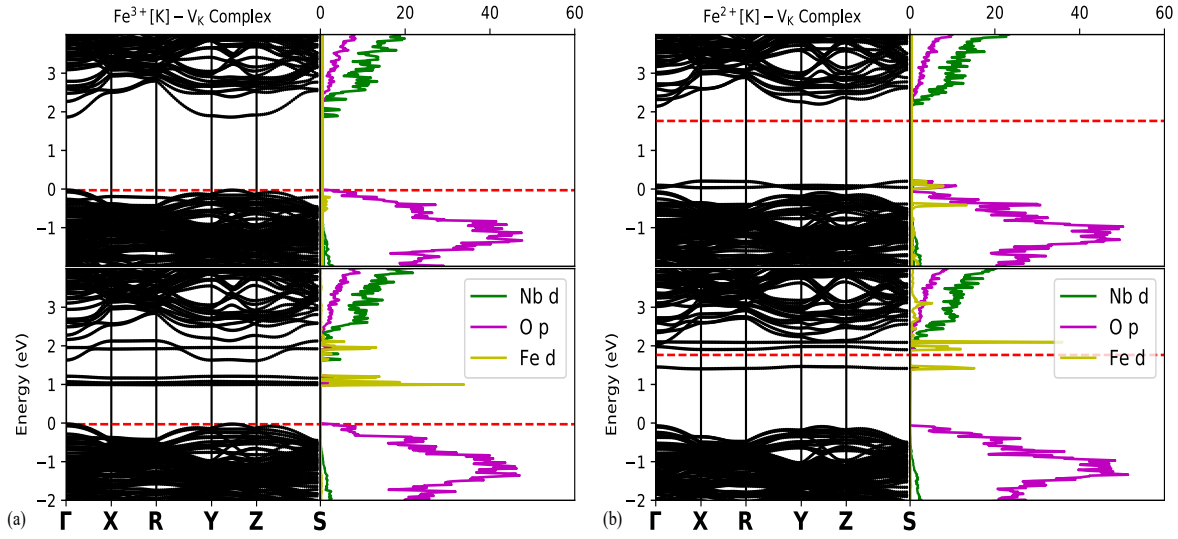

Figure 4.6: Band structure and pDOS of the Fe[K] – V_K complex with the Fe ion (left) in a 3+ charge state and (right) in a 2+ charge state. For each charge state, the atomic positions are those of the respective equilibrium structure. The top and bottom panels show the spin-up and spin-down quantities respectively. The dashed red line indicates the position of the Fermi level as determined by integration of the density of states.

system, the exchange splitting is a relatively weak perturbation compared to the crystal field splitting. Above the t_{2g} states in energy we observe a pair of e_g states close to the conduction band edge. One of these e_g states exhibits significantly more dispersion than the other, a pattern which appears in our previous study as well. It can be seen that the barycenter of this line falls quite close to its partner e_g state, however, which further supports that the exchange splitting effect is significantly weaker than the crystal field effect in this system.

The band structures of the complex with the Fe in the 2+ charge state (shown in panel (b) of Fig. 4.6) exhibit similar behaviors to the previous complex. In the spin-up band structures we see the two nearly degenerate e_g states now placed just above the valence band edge. The spin-down band structures again show the effect of the presence of the extra electron (and relaxation) as forcing the e_g and t_{2g} states up in energy, but now the t_{2g} states are not moved completely into the conduction band. The lowest lying spin-down t_{2g} state is

situated well below the conduction band with the Fermi level lying between this state and the next higher t_{2g} state.

4.3.3 Fe[Nb] – V_O

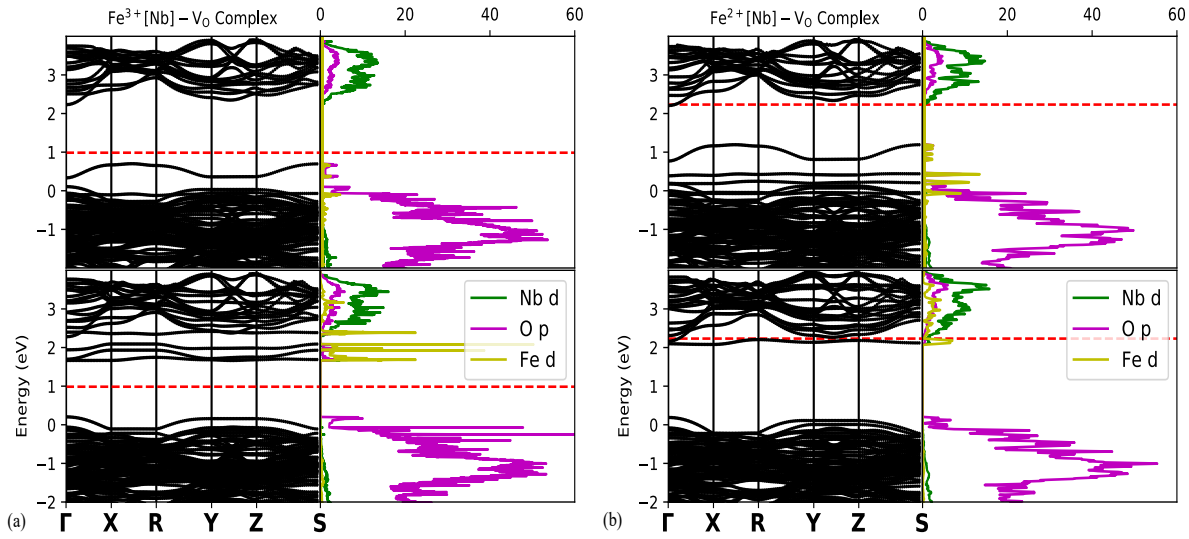

Figure 4.7: Band structure and pDOS of the Fe[Nb] – V_O complex with the Fe ion (left) in a 3+ charge state and (right) in a 2+ charge state. For each charge state, the atomic positions are those of the respective equilibrium structure. The top and bottom panels show the spin-up and spin-down quantities respectively. The dashed red line indicates the position of the Fermi level as determined by integration of the density of states.

Fig. 4.7(a) shows the band structure and pDOS of the Fe[Nb] – V_O in a 3+ charge state, again at the equilibrium structure for this charge state. As with the isolated substitution case, we see in the spin-up band structure an occupied, dispersive e_g state lying just above the valence band, with another e_g state within the valence band but very close to the edge. The spin-down band structure becomes slightly more interesting. We see another triplet of unoccupied t_{2g} states close to the conduction band edge. The splitting between these states is markedly more significant than in the case with the isolated substitution. Discussed in the previous chapter, this arises due to the additional perturbation of the octahedral crystal field induced by the nearest neighbor oxygen vacancy. The presence of this oxygen vacancy

also brings a higher lying e_g state just barely to the edge of the conduction band. This band is present in the Fe[Nb] complex by inspection of the pDOS, but is buried within the conduction band in that case. Given that the largest difference between the two calculations is the coordination of the O vacancy with the Fe atom, we can safely infer that this e_g state is strongly affected by the crystal field splitting.

In going from the 3+ charge state to the 2+ charge state of the Fe[Nb] – V_O complex (band structures in panel (b) of Fig. 4.7), we see similar trends as what occur through this transition in the isolated substitution case. The spin-up e_g states remain occupied and increase in energy. The lower-lying e_g is again lifted out of the valence band, with a spin-up t_{2g} again present at the edge of the valence band edge. The presence of the extra electron again forces most of the spin-down states Fe-3d into the conduction band with the Fermi level positioned just within the conduction band. Again, the lowest lying spin-down t_{2g} is occupied, but now appears to barely stay below the conduction band edge.

4.4 Discussion

4.4.1 Assignment of Impurity Levels

Examination of the band structures presented in the previous sections can be used to assign possible transition processes that may occur. The occupied e_g state in the spin-up band structure of the Fe[Nb] complex can thus be interpreted as a possible deep donor state, given its occupation. An electron in this state may be excited to the conduction band by a photon of sufficient energy, ionizing the dopant to a 4+ charge state, which aligns quite well with the intuition for this level presented in Fig. 4.1. Additionally, a lower-energy excitation of say, an electron in the valence band, to the unoccupied spin-down t_{2g} states for this system would see the system reduced from the 3+ ionization state to the 2+ ionization state. This largely matches the predictions of the literature [48, 95] which indicate that Fe3+/2+ transitions are largely responsible for increases in the photorefractive gain, as they

provide more viable charge donor and acceptor states, respectively. The possibility of intra-ion exchange is low given that the only possible process would involve a spin-flip of the charge carrier, forbidden for direct photon-mediated processes.

Assignment of impurity levels are performed with the aid of the Fig. 4.8. This figure shows a redrafted version of the figure by Basun *et al.* in panel (a) compared with a similar figure depicting the level assignments from the current work. Since the level positions were calculated using DFT, the exact energies of the states are inaccurate, however the relative positions of the states are qualitatively reliable.

The state highlighted in blue in panel (a) of Fig. 4.8, denoted $Fe^{3+/4+}$[Nb] was assigned by Basun *et al.* as an Fe ion on a Nb site, which can be in either a 3+ or 4+ ionization state. This level is directly comparable to the state highlighted in blue in panel (b), which is caused by an Fe ion substituted on a Nb site with nonlocal charge compensation denoted Fe^{3+}[Nb]. It is an occupied, spin up, e_g state denoted by $|e_g, \uparrow\rangle_O$.

The state highlighted in pink in panel (a) denoted $Fe^{2+/3+}$[Nb] was assigned by Zgonik *et al.* [90] as an Fe ion on a Nb site, which can be in either 2+ or 3+ ionization state. This level is directly comparable to the state highlighted in pink in panel (b), which is induced by an Fe ion substituted on a Nb site with nonlocal charge compensation as above. It is an unoccupied, spin down, t_{2g} state denoted by $|t_{2g}, \downarrow\rangle_U$.

The state highlighted in red in panel (a) denoted $Fe^{2+/3+} - V$ was assigned by Basun *et al.* as an Fe ion on either a K or Nb site (though their results support the K site) near either a K or O vacancy, which can be in either a 2+ or 3+ ionization state. This level is directly comparable to the state highlighted in red in panel (b), this is one of two states due to an Fe ion substituted on a K site next to a K vacancy with additional nonlocal charge compensation denoted $Fe^{3+}[K] - V_K$. It is an unoccupied, spin down, t_{2g} state denoted by $|t_{2g}, \downarrow\rangle_U$. Note that for the $Fe^{3+}[K] - V_K$ defect complex there is no $|e_g, \uparrow\rangle_O$ within the band gap as this state has been depressed to a position below the valence band edge.

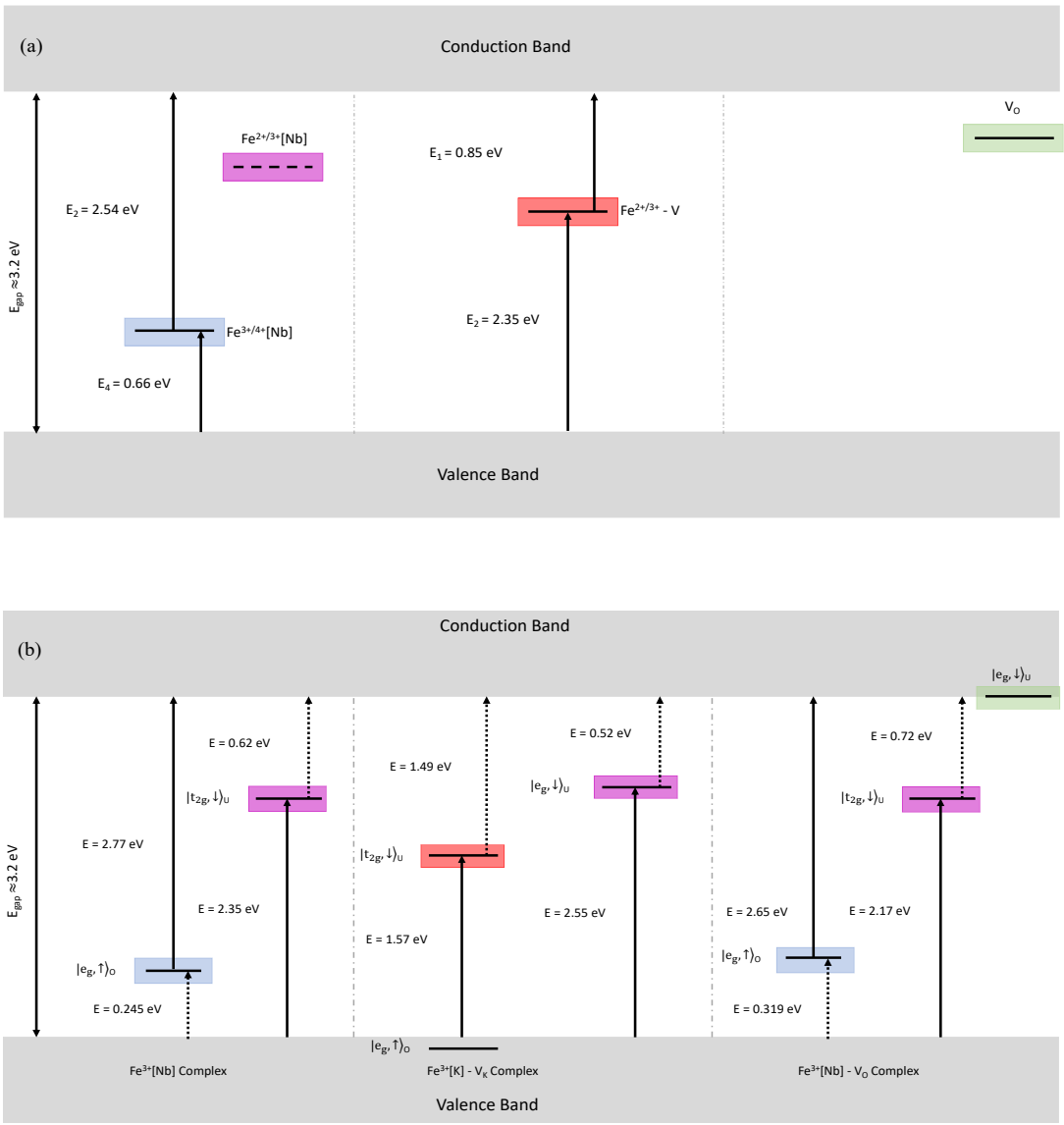

Figure 4.8: (a) A redrafted version of the energy level diagram reproduced from the work of Basun *et al.* [48]. In this version, the assignments made by Basun *et al.* are placed in a single pane and have been ordered by defect complex. Error bars which are present in the original figure are excluded here for clarity. (b) A diagram summarizing the excitations that we expect is based upon the band structures with Fe^3 in different defect complexes. Energies presented here have been scaled by the factor deduced in the previous study (1.4) which brings the DFT bandgap in line with the experimental values. In that chapter, the factor was deduced by comparison of the bandgap of the pure system. The doped supercells exhibit slight closing of the band gap compared to the pure system, hence why the sums of transition energies are slightly less than the experimental band gap. We color-code the assignments we make based on this pane with the assignments shown in (a). For the Fe[Nb] – V_O complex, because the unoccupied, spin-down e_g state is not strictly speaking below the conduction band, we place it at the conduction band edge.

The state highlighted in green in panel (a) denoted V_O was assigned by Biaggio *et al.* [91] as a shallow trap state arising due to an O vacancy. This feature is only observable in the reduced samples studied by Basun *et al.* We believe that we can assign this level to the state highlighted in green in panel (b), induced by an Fe ion substituted on a Nb site with a nearest neighbor O vacancy denoted $Fe^{3+}[Nb] - V_O$. It is an unoccupied, spin down, e_g state denoted by $|e_g, \downarrow\rangle_U$. This state lies just above the lowest point of the conduction band, but as can be seen from the band structure and pDOS in Fig. 4.7, this state is (i) nondispersive, (ii) highly localized, and (iii) would possibly appear as a sharp absorption feature below the main conduction band edge. It is also possible the position of this state may relax to a lower energy when occupied. An investigation of this effect provides partial motivation for the topic of the following section.

The $Fe^{3+}[Nb] - V_O$ has $|e_g, \uparrow\rangle_O$ and $|t_{2g}, \downarrow\rangle_U$ states at positions very similar to the $Fe^{3+}[Nb]$. Given the strong electron-phonon coupling associated with Fe^{3+} ions and the resulting broad absorption and emission features, it thus may be difficult to distinguish these levels using optical spectroscopy.

4.4.2 Configurational Coordinate Diagrams

In this section, we construct and examine DFT-generated configuration coordinate diagrams of each dopant complex with Fe in the 3+ charge state.

In Fig. 4.9, we show our configurational coordinate diagrams (CCD) for the defect complexes. In each of the panels of that figure, the color coordination of the levels has been organized to match as precisely as possible the assignments made in Fig. 4.8. Grey curves represent energy eigenvalues either coming from the valence bands, largely consisting of O-2p bands, or eigenvalues coming from conduction band states, primarily Nb 4-d based.

While normal CCDs model the states as harmonic, of important note here is that in some cases a parabola is an imperfect fit to the configurational coordinate data. This is because (i) the presence of anharmonicity in the atomic bonds of the material will cause the harmonic

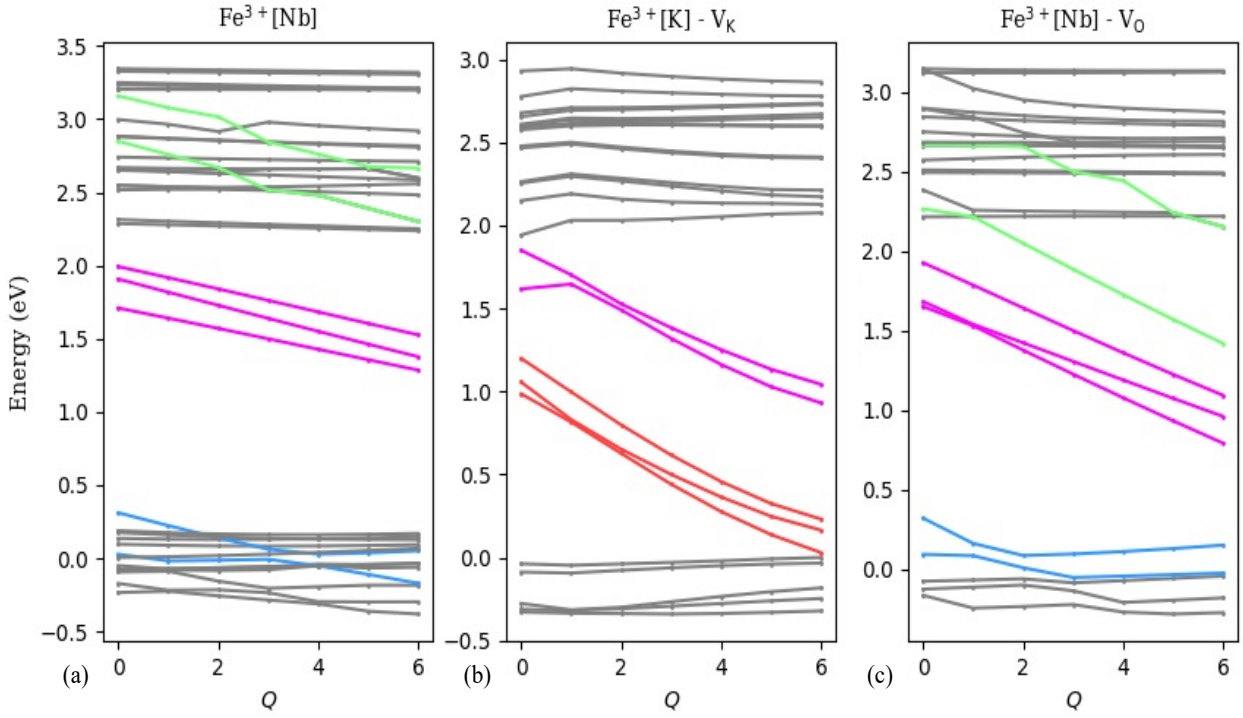

Figure 4.9: Configurational coordinate diagrams for each of the defect complexes with Fe in the 3+ charge state. In each panel, Q represents our configurational coordinate, with $Q = 0$ corresponding to the respective defect complex in the equilibrium configuration stabilizing Fe in the 3+ state. $Q = 6$ then corresponds to the equilibrium configuration which stabilizes Fe in the 2+ state, however, we charge the supercell such that Fe should assume a 3+ ionization state. The energy scales have been shifted with respect to the valence band maximum of the pure system.

oscillator approximation to break down (ii) our choice of configurational coordinate is not a perfect mode coordinate and (iii) the change in the band structures with respect to the configurational coordinate can cause significant changes to the ordering of the band energies which makes the unambiguous assignment of an individual state in a manifold difficult, and (iv) the possibility of impurity levels experiencing an avoided crossing. Having examined the parabolic fits to these levels, we opted instead for a linear interpolation to guide the eye rather than invite over-interpretation of the data here.

Panel (a) of Fig. 4.9 shows the CCD of the Fe^{3+}[Nb] complex. The blue curves in this panel are representative of the change of the occupied, spin-up e_g state, while the pink states are the unoccupied, spin-down t_{2g} states. The green states are indicated here given that their designation according to our level assignments is the same as the correspondingly colored state assigned to the V_O level in the case of Fe^{3+}[Nb] – V_O complex. We do not show them in Fig. 4.8 given their position within the conduction band.

Of immediate note is that the minimum of the occupied spin-up e_g states occurs at $Q \approx 2, 3$ for the upper and lower spin-up e_g states, respectively. This is indicative that the calculated configurational coordinate is not a perfect mode coordinate as discussed above. In order to double check the consistency of the spin-up e_g states, we use a Quantum Espresso post-processing utility that enables us to plot the contribution of selected atomic wavefunctions to the pseudo-charge density. That is, we can map out the norm of the wavefunctions contributing to a specific band. Using this information, we confirm that the nature of each of these bands is consistent across each structural step. That is to say, their bonding structure with respect to the local O atoms stays the same.

The unoccupied t_{2g} states exhibit a much simpler Q dependence, In particular, these states decrease in energy as Q increases. This relaxation is expected given that the lowest t_{2g} state becomes occupied when another electron is introduced into the system.

Panel (b) of Fig. 4.9 shows the CCD of the Fe^{3+}[K] – V_K complex. For this defect center the occupied, spin-up e_g states are situated within the valence band and are therefore not highlighted in this diagram. The red-colored curves represent the triplet of unoccupied t_{2g} Fe-3d states. The higher doublet of e_g states is color-coded in pink given in accordance with Fig. 4.8. Similar to the Fe^{3+}[Nb] complex, the unoccupied states decrease in energy as Q increases, again this relaxation is expected given that the lowest t_{2g} state becomes occupied when another electron is introduced into the system.

Panel (c) of Fig. 4.9 shows the CCD of the Fe^{3+}[Nb] – V_O complex. Similar to the case of the Fe^{3+}[Nb] complex, the blue curves in this panel are representative of the change of the

occupied, spin-up e_g states, while the pink states are the unoccupied, spin-down t_{2g} states. The unoccupied spin-down e_g states are depicted in green. Also similar to the case of the Fe^{3+}[Nb] complex, the minimum of the occupied spin-up e_g states occurs at $Q \approx 3$.

Of particular interest is the behavior of the unoccupied spin-down e_g states. The lower of these states lies at the edge of the conduction band at $Q = 0$, while the upper of these states lies within the conduction band. As Q increases from zero, both of these levels decrease in energy, so that both lie within the band gap at $Q = 6$. The unoccupied spin-down e_g states can therefore act as deep/shallow acceptors. Furthermore, since these acceptor states are associated with the Fe^{3+}[Nb] – V_O complex, the number of these states will increase as the number of O vacancies increase since the probability of finding an O vacancy adjacent to a Fe^{3+}[Nb] increases accordingly. We therefore assign the V_O acceptor level reported by Biaggio *et al.* to a $|e_g, \downarrow\rangle_U$ state, though its observation may depend significantly on the local geometry of the complex.

4.4.3 Limitations

While our method shows promising first results, we note some key limitations and outline some possibilities for refinement below.

A first and foremost limitation is that the localization of extra charge onto the dopant site is by no means guaranteed in any system. We expect tight convergence parameters may help this, but if charge fails to localize appropriately, extensions such as Hubbard U methods may be necessary to enforce the desired occupation of the dopant atom. Also of import is that reasonably-sized supercell calculations are likely to incur some amount of dispersion in these impurity bands as interactions with periodic replicas can only be eliminated in a truly infinite supercell. As mentioned in the previous chapter, we are fortunate in this regard considering that $KNbO_3$ has a relatively high dielectric constant which helps screen the dopants from their images. If the material of interest incurs strong enough dispersion in

the impurity bands, the energy levels of a specific high-symmetry point in the Brillouin zone may give different configuration coordinate diagrams.

Related to this issue is that it can be important to assign some index to each of these bands and keep this index consistent across different configurations. As discovered in the last chapter, significant enough distortion of the local symmetry around a dopant site can cause certain bands to reorder. Given that advancing our configuration coordinate will certainly involve some distortion of the local symmetry in the lattice, attention must be paid to this issue if the diagrams are to be accurate to states of a specific character. This can be done by mapping out the orbital structures of the bands of interest and verifying that their shape stays consistent across the configuration coordinate, or, when it changes, that the corresponding energy is re-indexed accordingly. This process is, however, painstaking and, at the current state, requires human inspection.

As mentioned above, the lack of a true mode coordinate may also negatively impact the results of the method. It may be possible to overcome this, however, if one can be identified. While not pursued in this work, the calculation of phonon properties of these systems would yield a set of stable vibrational modes. Taken together with the knowledge of how the system coordinates change in going from one equilibrium configuration to another, it may be possible to project out which vibrational modes contribute most strongly to the change in configuration, and from this deduce a more rigidly defined configurational coordinate. High quality phonon calculations within DFT are well-established in the literature [96]. In Fig. 4.10 we show a phonon dispersion curve for cubic $KNbO_3$. The calculations to extract this dispersion curve also produce the necessary data to visualize the phonon modes at various points within the Brillouin zone. Using the displacement patterns akin to those gather in Fig. 4.3 along with the breakdown of different vibration modes, it should be possible to associate a given displacement pattern with some linear combination of vibrational modes, which could lead to a more well-defined configurational coordinate.

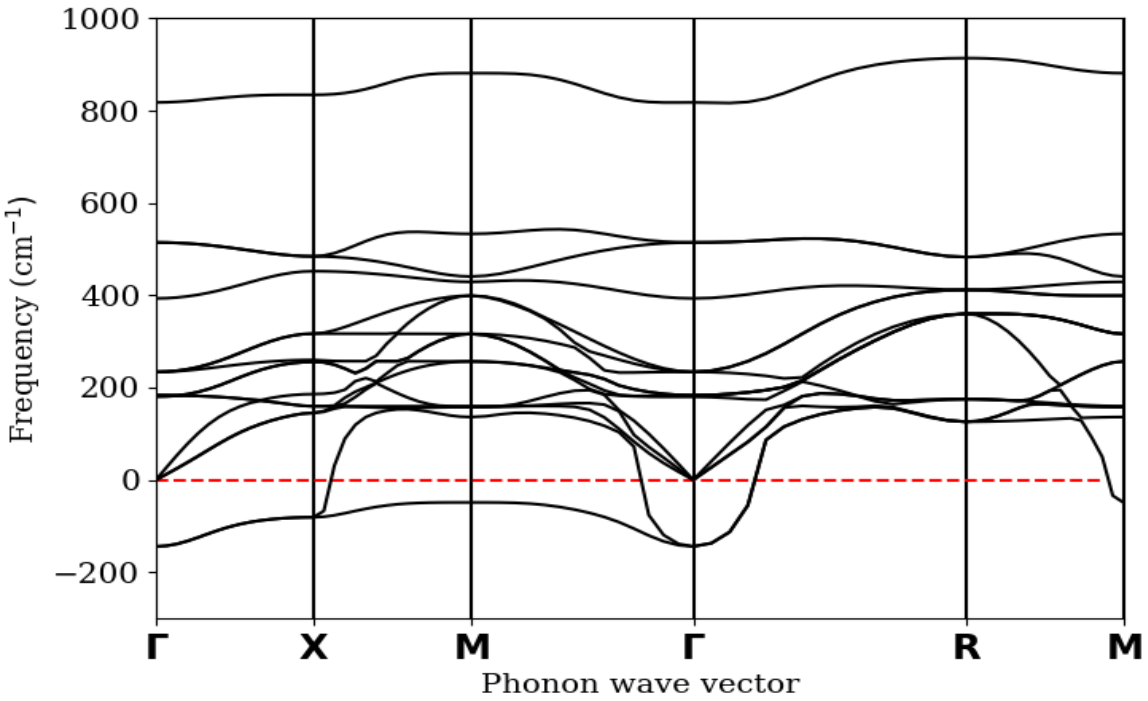

Figure 4.10: A phonon dispersion curve for cubic $KNbO_3$. The existence of imaginary phonon frequency eigenvalues indicates that the structure is unstable against vibrations of this mode, which may be stabilized by a transition to the ferroelectric tetragonal phase.

Finally, the approximation of a given energy level as being harmonic with respect to the configuration coordinate may not strictly be obeyed. Anharmonicity in the response of a particular bond may cause a parabolic fit to be inadequate to describe the energy level. Going beyond this and including some types of anharmonic corrections is in principle possible, but the additional complications of identifying which bands undergo such anharmonicity and fitting such anharmonic factors systematically may be beyond the scope of such a method.

4.5 Conclusions

It has been demonstrated in the literature that $KNbO_3$ doped with Fe in different substitutional sites and different charge states can dramatically enhance that host material's

photorefractive properties by modifying the effective trap density and dark conductivity. In this chapter, we have performed a detailed study of Fe as a substitutional impurity on both K and Nb sites in different local environments and ionization states. In this chapter, we describe a methodology which may prove useful in the examination of defect complexes in adjacent charge states. The structural origin of the differences between electronic energy levels of ions in different ionization states can play an important role in determining their nature as acceptor or donor states. Of equal interest is how the acceptor or donor nature of these states may evolve under the relaxation to or from the equilibrium structure of an adjacent charge state. Our proposed methodology allows for the exploration of these differences at an *ab initio* level, which can be vital for guiding explorations of novel materials.

A useful technique within DFT to vary the ionization state of an ion is to introduce an extra charge into a cell which is compensated by an equal and opposite jellium background charge. We have demonstrated that band structures for an Fe^{3+} on the Nb site with a second-nearest neighbor charge compensating O vacancy and an Fe^{3+} on a Nb site with a compensating jellium background charge exhibit nearly identical band structures showing that use of a compensating jellium background charge provides an excellent proxy for non-local charge compensation models.

We calculate band structures and pDOS for Fe^{3+} on a Nb site with non-local compensation, Fe on a Nb site with a nearest-neighbor O vacancy as charge compensation and Fe on a K site, with a nearest-neighbor K vacancy plus additional non-local charge compensation. In addition, the band structures and pDOS for the corresponding Fe^{2+} are calculated. This data can be used to assign the defect states reported by Basun *et al.*, Biaggio *et al.* and Zgonik *et al.*

This work has also demonstrated the ability of DFT calculations to add to the interpretation of experimental data. While it does not produce reliable energies for these levels, the qualitative behaviors of impurity states can be discerned. This allows us to potentially aid

experimentalists in the assignment of defect levels, giving clues about aspects such as the symmetry of the impurity levels.

In order to investigate how these defect energy levels change their position under electronic excitation, we construct configurational coordinate diagrams for each of the states of interest. The specific configurational coordinate, Q, is defined by a linearly interpolated path between a given ion's 3+ equilibrium position and its 2+ equilibrium position. The limitations of this approach are discussed.

The method in its current form is seen to produce physically reasonable configurational coordinate curves for many of the states studied. However, in the case of $Fe^{3+}[Nb]$ it predicts that the $|e_g, \uparrow\rangle_O$ states are stabilized at some intermediate values of Q. This could either be indicative of a poor choice of configurational coordinate as is discussed above, or it may be an artifact of an avoided crossing of the two levels.

Of particular note is the $Fe^{3+}[Nb] - V_O$ defect complex which exhibits unoccupied, spin-down e_g states whose position relative to the conduction band edge may identify it as the V_O level in Fig. 4.1. Examining the band structure of the Fe^{3+} ion forced into the Fe^{2+} equilibrium coordinates sees these levels relax away from the conduction band edge, placing them more appropriately in line with the corresponding level in Fig. 4.1.

Clearly some limitations of our results are also inherent to the fact that we are limited in our supercell size. However, we believe that we have demonstrated the potential of the method to aid in the interpretation of experimental data. While more case studies are necessary to properly generalize the method and refinements are certainly necessary to develop a more systematic way of identifying the appropriate configurational coordinates for candidate systems, the results presented here speak to a proof of concept which is worth testing on other systems of experimental interest.

Chapter 5

GW Calculations

Going beyond standard Kohn-Sham DFT requires the consideration of many-body effects. This chapter focuses on the use of the BerkeleyGW (BGW) code suite to analyze the effects of many-body interactions in $KNbO_3$ systems. We begin with a preliminary study on the cubic, tetragonal, and orthorhombic phases of $KNbO_3$.

5.1 Preliminary Results

The pure phases of $KNbO_3$ are useful testing grounds as a preface to applying the GW approximation to doped systems of $KNbO_3$. The unit cells for the pure phases are quite small and can be used to establish convergence parameters, the effects of the many-body corrections on the host band structures, as well as providing a baseline of the computational expense necessary to apply these corrections to supercell systems.

5.1.1 Convergence Procedure

As mentioned in Appendix D, the convergence procedure for GW calculations is significantly more complicated than for standard DFT calculations. The two main components of these types of calculations consist of calculating the dielectric function and then using this to

calculate the energetic corrections to individual states. Separately, these are accomplished by the **epsilon.x** and **sigma.x** executables of the BGW suite. These two steps need to be individually converged.

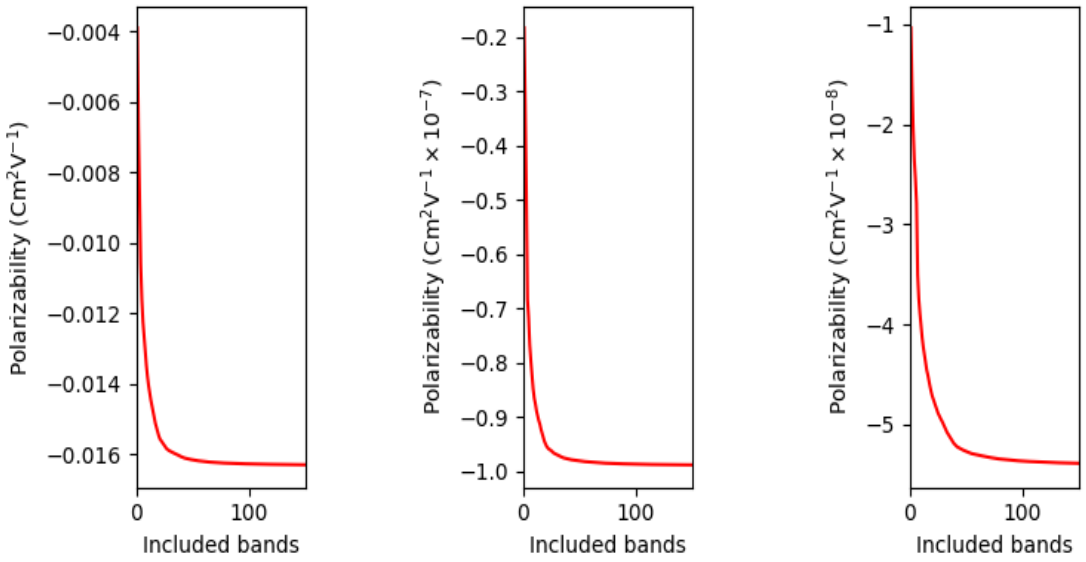

Figure 5.1: A graph demonstrating convergence of the polarizability for $\boldsymbol{G}, \boldsymbol{G'} = 0$ at a single point in the Brillouin zone for the pure phases. Panel (a) shows this curve for the cubic system, panel (b) the curve for the tetragonal system, and panel (c) the curve for the orthorhombic system. Panels (b) and (c) have a quite different scales, indicated in their units.

Appendix D explains that the inclusion of a number of unoccupied bands (or equivalently in the epsilon calculation, an energetic cutoff for the screened Coulomb interaction) acts as the primary convergence parameter for the dielectric function which is calculated over a specified grid of points in the Brillouin zone. For each Brillouin zone point, **epsilon.x** calculates the irreducible polarizability over a set of partial sums with respect to the number unoccupied states included. This provides one metric of convergence. Displayed in Fig. (5.1) is a graph showing this convergence behavior for a Brillouin zone point in cubic $KNbO_3$. By examination of all such graphs for all Brillouin zone points, we find that the dielectric function converges very rapidly in each of the pure systems with the inclusion of approximately 100

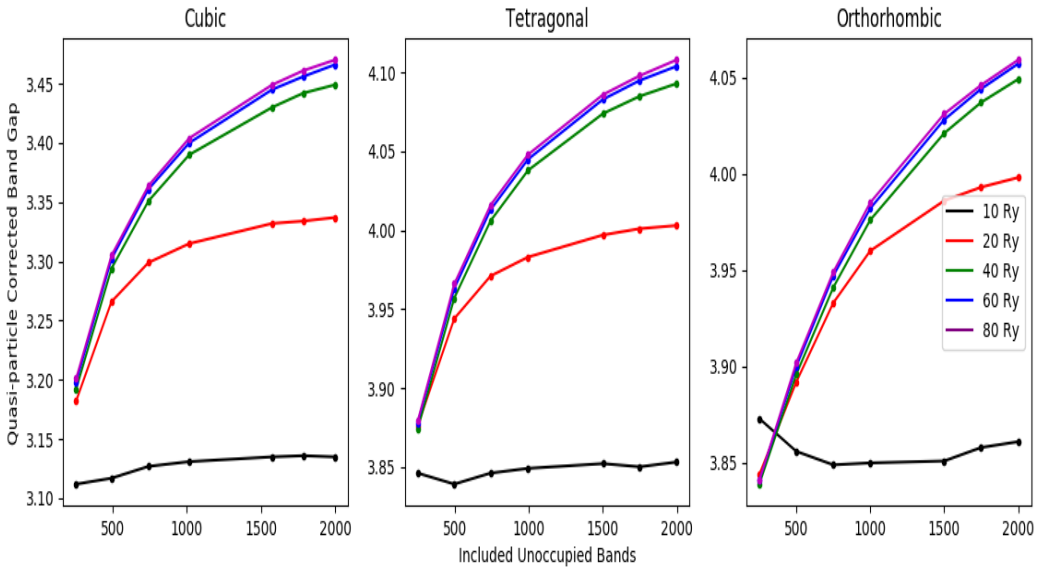

Figure 5.2: Convergence behavior of the calculated Γ point band gap for the pure phases of $KNbO_3$. Panel (a) shows the convergence for the cubic system, panel (b) shows it for the tetragonal system, and panel (c) shows it for the orthorhombic system.

unoccupied states. This information can prove useful in determining appropriate convergence for the supercell, should it follow similar trends.

The convergence of the **sigma.x** calculation relies not only on the appropriate convergence of the previous steps, but also on the number of unoccupied states included in its own summation as well as the energetic cutoff for the screened Coulomb interaction. Unlike in the **epsilon.x** step, these two parameters are independent. In practice, one must set the cutoff energy, vary the number of included virtual bands and observe the convergence in some physical quantity. Then the cutoff energy must be altered, the number of included virtual states varied over the same range, and the convergence of the same physical quantity observed once more. In our case, the band gap is used as the physical quantity to be converged.

Fig. (5.2) demonstrates the necessity of this convergence scheme. The figure shows the behavior of the band gap in cubic, tetragonal, and orthorhombic $KNbO_3$. Each curve

exhibits a monotonic increase with respect to the number of included virtual states, but each curve converges to different values. The amount of error is then considered based on the tightest parameters used, that is, the highest cutoff with the largest number of virtual states. Using a cutoff energy that is too low can yield false convergence, and thus this behavior is important to investigate for each system of interest. This process is, however, by nature of needing to do these calculations multiple times quite expensive in terms of computational resources and wall-clock time. Thus, having some idea of reasonable convergence parameters may conserve significant computational resources.

Fig. 5.2 (a) also demonstrates the diminishing returns of increasing these convergence parameters. Calculating the quasi-particle corrections for cubic $KNbO_3$ with a screened Coulombic cutoff of 80 Ry and including approximately 2000 virtual states requires approximately 50% more total RAM and increases the necessary wall-clock time by a factor of approximately 4 as compared to running the calculation with a screened Coulombic cutoff of 40 Ry with the same number of virtual states. However, the difference in the calculated band gaps of these two schemes is only 0.021 eV, or an error of 0.6% with respect to the tighter convergence scheme. If we consider a 1% error as tolerable, we can reduce this further to a cutoff of 40.0 Ry including only about 1500 bands. Following this same procedure, we determine that the tetragonal and orthorhombic systems also reach 1% error with these values.

A key capacity offered by GW calculations is the application of quasi-particle corrections to DFT level band structures, producing quasi-particle band structures. While these band structures still do not account for electron-hole interactions, they can be expected to correct impurity levels to energies affording closer agreement with experiment. These can provide us with more accurate insights into the validity of promising defect models, with only minimal cost beyond application of the standard single-shot GW approximation. Quasi-particle corrected band structures can be considered as the next step in an in-depth examination of

any given defect complex. These types of band structures are demonstrated for the pure $KNbO_3$ phases in the following section.

5.1.2 Initial Quasiparticle Results

The calculation of the self-energy operator over a sparse grid of point in the Brillouin zone can be used to interpolate the quasi-particle correction to a finer grid of k-points. In particular, this interpolation can be done over the high-symmetry path of the band structure calculation to generate quasi-particle corrected band structures.

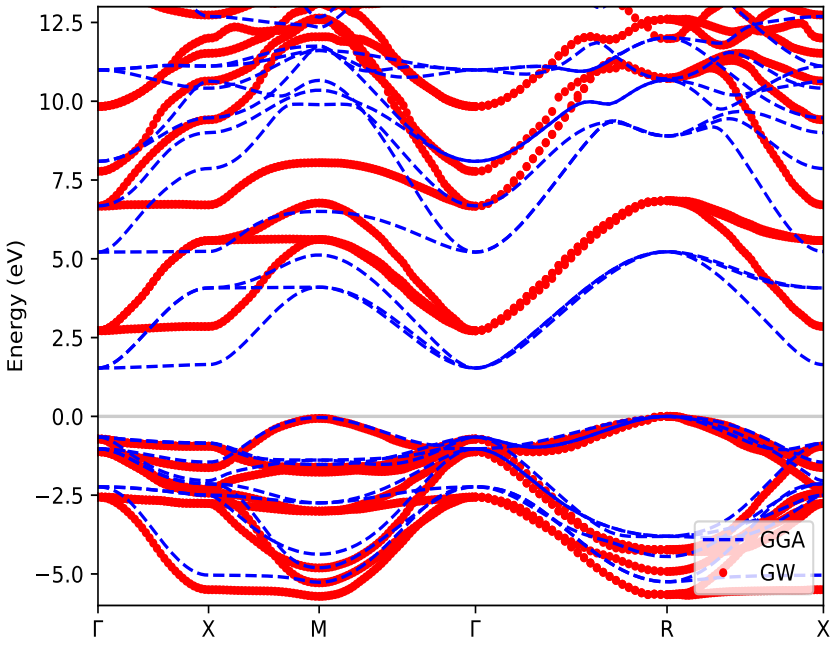

Figure 5.3: The quasi-particle corrected band structure of cubic $KNbO_3$. The DFT bands collected with the GGA are displayed as dashed blue lines. The GW corrected values are shown as red dotted lines

Fig. 5.3 shows the effects of the quasi-particle corrections on the band structure of cubic $KNbO_3$. This figure focuses on the energy window containing the band gap. We notice in this regime that the valence band edge is largely unaffected by the corrections, though with some small shift downward in energy for some bands. The conduction bands, however, shift

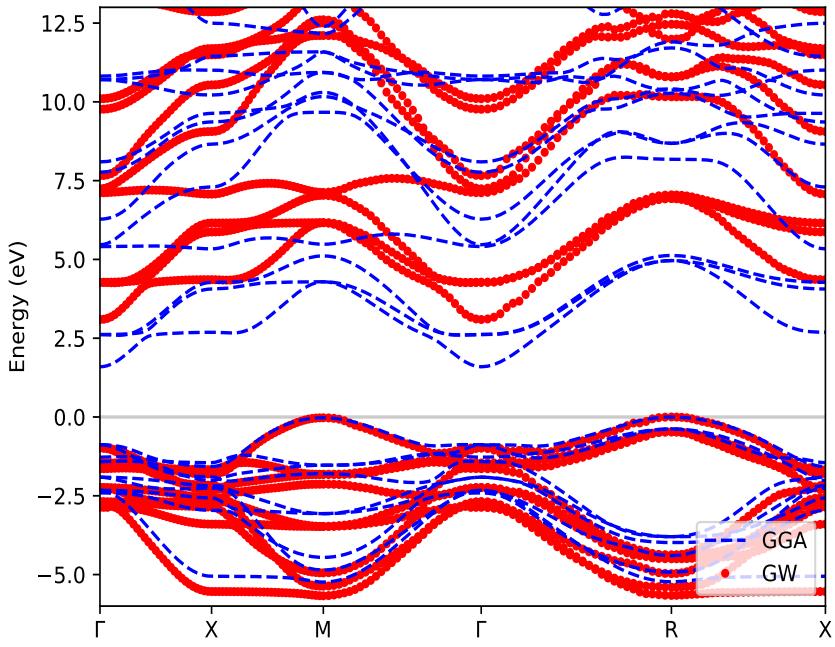

Figure 5.4: The quasi-particle corrected band structure of tetragonal $KNbO_3$. The DFT bands collected with the GGA are displayed as dashed blue lines. The GW corrected values are shown as red dotted lines

quite rigidly upward. We see that the band gap at the Γ point opens from a DFT value of 1.97 eV to a value of approximately 3.4 eV, in much better agreement with values found in experiment, which range from 3.16 – 3.24 eV [97–99]. Examination of bands outside of the chosen window shows that deeper occupied states tend to be corrected downward in energy, while nearly all unoccupied states are shifted upward.

Fig. 5.4 shows the quasi-particle band structure of the tetragonal $KNbO_3$ system. We notice very similar trends as occurs in the cubic system. The valence band edge remains minimally corrected with deeper occupied states correcting downward in energy and the unoccupied states correcting upward. We find that the Γ point band gap opens up from a DFT derived value of 2.47 eV to a value of 4.07 eV, also much closer to experimental values,

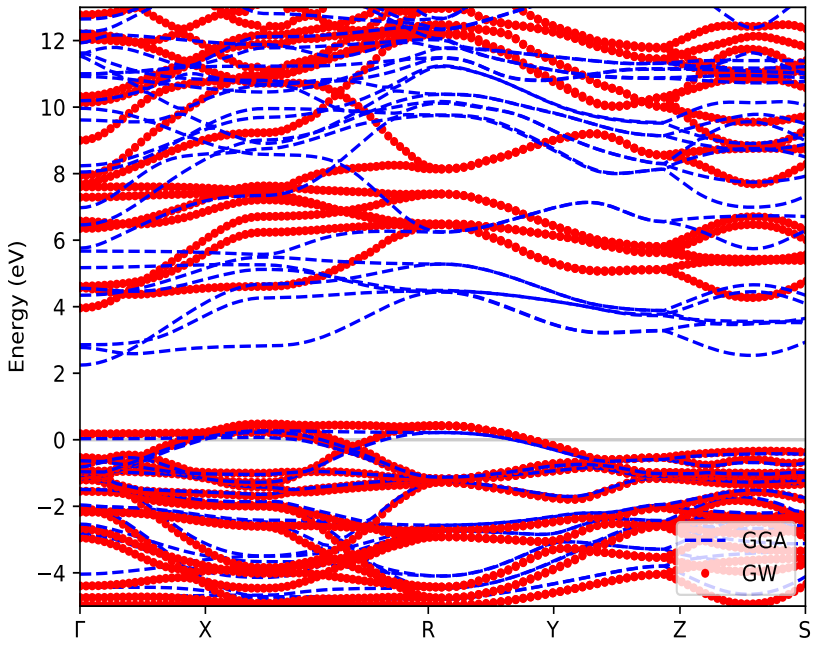

Figure 5.5: The quasi-particle corrected band structure of orthorhombic $KNbO_3$. The DFT bands collected with the GGA are displayed as dashed blue lines. The GW corrected values are shown as red dotted lines

ranging from 3.08 – 3.30 eV [99, 100]. It is interesting to note that the GW approximation appears to significantly overcorrect the band gap.

In Fig. 5.5 we see the quasi-particles band structure of the orthorhombic structure. The trends observed in the cubic and tetragonal systems appear to hold well for the orthorhombic system as well. The valence band edge remains minimally corrected with the Γ point gap opening up largely due to the upward shift of the conductions bands. The DFT band gap of 2.5 eV opens to a value of 4.02 eV. This also compares favorably with experimental values ranging from 3.15 – 3.3 eV [99, 101, 102]. We notice that once again the GW overestimates the band gap.

The value of quasi-particle band structures at this level is largely limited to establishing intuition for these calculations and observing the trends of GW effects. However, the results of our self-energy corrections agree quite well with those of Schmidt *et al.* [103] who applied GW corrections under the Vienna Ab Initio Simulation Package (VASP) to the pure phases of $KNbO_3$. Of note is that we find slightly lower values of the GW correction to the cubic and orthorhombic direct band gaps compared to their results. Schmidt *et al.* conclude that a scissor operator is adequate to describe the effect of the GW corrections on the pure systems of $KNbO_3$ however, there is no guarantee that this assumption will remain true for doped system.

5.2 Future Directions

5.2.1 Necessity of larger computational resources

These initial results provide significant insight into the computational expense of using the GW approximation for the supercells necessary to model defected systems. In going from an orthorhombic unit cell to a defected supercell, we add approximately 70 atoms (the exact number depending on the particular vacancies introduced for charge compensation) but must also consider that our entire computational load is further increased considering that we must use spin-polarized calculations to account for the metallic characteristic of any considered dopants.

Importantly, scaling up GW calculations to apply to supercell calculations is unfeasible with our current resources. The calculation of the dielectric function for the orthorhombic system with a screened Coulomb cutoff of 90.0 Ry and including 2500 virtual states required approximately 1300 GB of total execution RAM running for approximately 57 wall clock hours. While this execution time is quite variable depending on the processors employed, the memory required is fixed. By comparison, the **epsilon.x** calculation of an Fe-doped supercell system with a screened Coulomb cutoff of 80.0 Ry and including 2000 virtual states requires

on the order of 120,000 GB of RAM, an approximately 90-fold increase in computational cost. This step is largely memory-limited. These calculations can be parallelized over the individual k-points and calculated as resources become available.

The calculation of the self-energy operator, on the other hand, is largely time limited. For the orthorhombic structure, with a screened Coulomb cutoff of 80.0 Ry and including 2000 virtual bands, the **sigma.x** executable requires approximately 126 GB of RAM. Comparison with the supercell calculation is unavailable, given the expense of the preceding calculation. However, assuming that a similar scaling exists between the **epsilon.x** and **sigma.x** steps of the orthorhombic structure and the supercell, we can infer a cost on the order of 12,000 GB of total RAM necessary for the self-energy calculations of the supercell system. These calculations can similarly be split per k-point in the specified grid, making this step much more approachable from a memory-usage standpoint. This step is instead largely time-limited. Again, timing for this step is somewhat unreliable as the quality of processor involved will vary the wall-clock time significantly.

Compared to either of the preceding steps, the interpolation to a quasi-particle corrected band structure is largely negligible. In moving forward bringing these calculations to scale, we need to consider means of making the first two steps computationally tenable. The first step to doing this for the **epsilon.x** calculation may be loosening the convergence criteria slightly. We can ostensibly build timing curves from a calculation using a screened Coulomb cutoff of 20.0 Ry and including up to 1500 virtual states which decreases the computational cost by a factor of approximately 12, requiring approximately 9,800 GB of RAM. Starter grants for the Expanse system (of the San Diego Supercomputer Computer) under the Extreme Science and Discovery Environment (XSEDE) program have a maximum limit of 50,000 hours. The Expanse cluster contains 728 nodes using AMD EPYC 7742 processors. Each of these nodes contains 128 cores each with a RAM limit of 256 GB. This type of calculation would require the dedication of 39 full nodes of the Expanse system, using the starting grant hours to run the calculation for 10 hours. This would be an important step in establishing the time to

completion that a more fully converged calculation would require, that is, establishing the amount of resources a future grant would need to request.

5.2.2 Solving BSE for doped systems

The quasi-particle eigenenergies and wavefunctions calculated via **epsilon.x** and **sigma.x** are the basis for the even more complicated task of diagonalizing the Bethe-Salpeter equation (BSE) to find excitonic eigenenergies and wavefunctions. The computational resources required for this step increase significantly, hence its reservation as future step of the investigation. However, these types of calculations would provide spectral data this is the most directly relevant to experimentalists.

As discussed in Appendix D, the calculation of absorption spectra in the BGW code is done in two parts. The first part consists in the use of the **kernel.x** executable, which calculates the interaction kernel of the BSE over a coarse grid of k-points. This step of the absorption calculation requires the dielectric function calculated by **epsilon.x**, as well as the wavefunction information. The second step is performed by the **absorption.x** executable which interpolates the results from the coarse k-point grid, to a finer grid and subsequently diagonalizes the BSE. This step relies on the output of **kernel.x** as well as the wavefunction information for the fine and coarse grids, and additionally the use of the quasi-particle corrected eigenenergies output by a **sigma.x** calculation. In addition to the convergence criteria of each of the previous steps, **absorption.x** also depends on the densities of the fine and coarse grids, and the number of bands included for each of those grids.

The cost of an absorption calculation for the orthorhombic structure totals 14,400 GB of RAM, putting the cost of a supercell calculation well beyond the realm of computational tractability without even greater supercomputer resources. But these may require significant "proof of concept" examples in order to justify these sorts of expenses.

5.2.3 Alternative *ab initio* approaches

Given the expense of GW calculations demonstrated here, alternative investigation methods may be useful for future studies. While the specific applicability of each of these methods to $KNbO_3$ has not been examined in-depth, we present brief examinations of their benefits and difficulties here.

5.2.3.1 Hubbard U

Highly localized orbitals, such as d and f orbitals, experience a higher self-interaction and localization errors as discussed in Chapter 3. Often this causes the LDA and GGA to delocalize these types of orbitals. In Mott-Hubbard materials, this results in DFT predicting the material to be conductive instead of insulating or semi-conducting. Corrections to this behavior came in the form of the Hubbard U correction [85, 104]. This approach involves the introduction of an occupation-dependent energy term into the Kohn-Sham Hamiltonian. This potential energetically penalizes the system if the specified orbitals are found to be unoccupied. This U parameter is frequently tuned in order to open up a band gap in the wrongly-predicted metallic materials. This semi-empirical approach somewhat undercuts the point of *ab initio* approaches, however recent advances have produced self-consistent methodologies for determining the value of the U parameter [86, 105]. Fundamentally however, these calculations are still at a DFT level, meaning that results obtained will still be limited to the non-interacting electron picture, it may however prove necessary for Mott-Hubbard systems or systems whose defects may involve strongly localizing some states.

5.2.3.2 Hybrid Functionals

If agreement with experiment is a high priority, the LDA and GGA functionals can be replaced by hybrid functionals. The representation of the many-electron wavefunction of a system using Slater determinants comes with the unique benefit of exactly accounting for the exchange phenomenon inherent to Fermions. The mathematics of calculating a Slater

determinant makes it computationally intractable for systems larger than a few atoms. The Hartree and Hartree-Fock quantum chemistry methods which are built upon this approach do not account well for the correlation effects in many-electrons systems. This is appropriate for small molecular systems, where correlation effects are largely negligible.

Obviously applying Slater determinant-based methods to crystal systems becomes prohibitively expensive and not ideal given the high degree of correlation in most solid-state systems. However, hybrid energy functionals seek to include a certain amount of the exactly calculated exchange energy of these calculations within the framework of DFT [88]. This involves the user specifying the fraction of the exactly-calculated exchange to be included in the calculation. Some consensus in the literature indicates an inclusion factor of $\alpha = 0.25$ is optimal for a broad range of systems, but this should still be considered as another convergence parameter to be tested. This exact exchange almost universally brings band gaps calculated under Kohn-Sham DFT to fairly good agreement with experiment. This comes at the cost of a dramatically increased computational cost and time. However, for larger systems, these calculations may be significantly more approachable than GW calculations. New functionalities within Quantum Espresso version 7.0 now makes it possible to perform band structure calculations with hybrid functionals. While this approach still does not rise to the level of theory of GW, hybrid functionals do include some degree of screening in the Coulomb potential, meaning that these types of band structures may give useful guidance on the impurity bands which a defect complex may introduce into a system.

5.2.3.3 TDDFT

As mentioned previously, the prediction of absorption spectra is one of the most powerful tools which can used to guide crystal growth projects. The developments of DFT and its successes in treating static systems eventually gave rise to a similarly structured formalism for dealing with dynamical processes such as absorption and scattering processes in the form of time-dependent DFT (TDDFT) [106].

Electron energy loss and inelastic X-ray scattering are both techniques that can be used to explore collective and single-particle excitations in solid-state systems. In TDDFT, these types of spectral information can be obtained from the poles of the susceptibility calculated by generating an independent-electron susceptibility and then solving a Dyson-like screening equation to obtain a full response function. This generally involves the calculation of numerous unoccupied states to build the susceptibility. This is then followed by the computationally expensive step of solving the Dyson-like equation for a given frequency, iterating these steps for each frequency desired. Significant developments to this approach involve a technique which allows one to calculate these quantities without using any virtual states, allowing for the calculation of a spectrum of a solid over a broad range of frequencies with a computational effort only a few times that necessary for a single ground-state DFT calculation. This algorithm has been implemented in Quantum Espresso and as such, this represents a possibly fruitful means of approach. Although these calculations are still at the level of independent electrons under the formalism of DFT, the inclusion of screening has the possibility to be a useful step toward making *ab initio* predictions of systems of experimental interest.

5.3 Conclusions

With this work we have ascertained the effect of quasi-particle corrections to pure $KNbO_3$ systems. We have also begun to form reasonable assumptions about the computational expense for applying GW calculations to doped $KNbO_3$ systems by simple scaling arguments. We have determined useful ranges of convergence parameters for these systems, which may be expected to hold to some extent for the $KNbO_3$ supercell systems as well. While supercomputing power in the modern era has skyrocketed, it is extremely vital to have approximate bounds of the resources necessary to complete calculations of this type. Thus, determining the resources necessary for full-scale GW applications is a vital step. By applying the GW

approximation to simpler systems we may begin to identify convergence parameters which may yield trustable results for supercell systems. If consistent timing information is available for those small-scale calculations, approximate timing curves for full-scale simulations may be able to be generated as well. Additionally, several levels of theory between Kohn-Sham DFT and full GW calculations have been identified which may be useful in guiding experiments. Band structures with hybrid functionals, TDDFT and the combination of TDDFT with hybrid functionals [107] represent only a few of these useful intermediate levels of theory which can offer predictive capabilities useful to experimentalists.

The application of GW calculations to the study of defected supercell systems proves prohibitively expensive for these systems and remains untenable on anything short of large scale supercomputing arrays. While lower levels of theory may still serve as useful guideposts to experimental designs, GW calculations still represent the gold standard of *ab initio* predictive power. As such, they do still stand as a sort of "last resort" if highly accurate *ab initio* predictions are vital. There is also some precedent for the possibility of these calculations on supercell systems. Rinke *et al.* [108] performed GW calculations of a 64-atom supercell of Si to calculate the defect formation energies of the Si interstitial. Lany and Zunger [109] used a 64-atom supercell of ZnO similarly to calculate the defect formation energy of O vacancies in that system, however for computational tractability they employed relatively soft pseudopotentials for O and determined the response functions only at the Γ point. Investigating the formation energy of O vacancies in hafnia (HfO_2) systems, Jain *et al.* [110] applied GW corrections to systems of up to 96 atoms. Their calculations were performed on the powerful TeraGrid supercomputer system. Additionally, the cutoff energy used for the calculation of the dielectric matrix was restricted to 10 Ry, presumably to strike a balance between computational tractability and accuracy with respect to experiment. Malashevich *et al.* [111] studied O vacancies in rutile TiO_2 by employing GW calculations on defected supercell systems consisting of 71 atoms. These calculations were performed at the National Energy Research Scientific Computing (NERSC) Center, another powerful supercomputer

facility, however, the dielectric function was calculated using a cutoff of 40 Ry and with the inclusion of 4000 conduction bands. Thus, while they require some of the most advanced supercomputing capabilities available, the feasibility of applying GW corrections to doped systems has been demonstrated in the literature.

Additionally, application of GW calculations to large systems has been tackled on several algorithmic fronts as well. Furthmüller *et al.* [112] sought to simplify the treatment of the dynamical and local-field effects by employing a model dielectric function. Bruneval and Gonze [113] developed a methodology by which the slow convergence of GW with respect to the number of included virtual states is mitigated by replacing the energetic contribution of all virtual states beyond an explicit threshold by a common energy determined through an adjustable parameter. Approaches by Umari *et al.* [114] and Giustino *et al.* [115] focused on removing the need for empty orbitals altogether by calculating the dielectric response and self-energy with occupied orbitals within a linear-response Sternheimer equation. Most recently, Gao *et al.* [116] have proposed an approach which seeks to approximate the costly summation over unoccupied bands with a more efficient numerical integration over a sparse energy grid. While these approaches vary in conceptual complexity, ease of optimization, and implementation it is worth noting that methods of applying GW applications to large systems are being pursued.

Chapter 6

Summary, Conclusions, and Future Directions

The photorefractive effect is a phenomenon with a diversity of technological applications. The maturation of photorefractive materials for practical device fabrication, however, is yet to be fully realized. To achieve this, promising photorefractive systems must be identified, and methods of tuning their photorefractive properties must be explored. Proper exploitation of the photorefractive effect will vary based on the desired application. If, for instance, long storage times are required, then methods of engineering deep trap centers may be necessary to ensure low dark conductivity and long grating lifetimes. However, dynamic holography applications may require fast reading and writing times in which case short grating rise times are necessary. Judiciously doping photorefractive materials is an established means of achieving these various properties. It has been shown that dopants can be used to extend the spectral response range of materials, increase photorefractive gain, and significantly alter grating rise and decay times.

Growing these materials is often challenging, expensive, and time consuming. Additionally, avoiding unintentional dopant inclusion is often a challenge in and of itself. In such cases, understanding the precise nature and effect of those dopant may further complicate

experiments. Additionally, the parameter space for experimental exploration is vast. The capacity to vary crystal system, dopant species, concentration percentages, growth methods, and post-growth treatments represent only a few degrees of freedom that experimentalists have. Methodologies which can side-step the direct expense of crystal growth are thus desirable as a means of helping to cost-effectively explore this parameter space. *Ab initio* studies represent one possible means of aiding experimental design.

To this end, we have performed simulations of doped $KNbO_3$ systems. We examined the effect of transition metal substitutions on Nb sites with coordinated and second-nearest neighbor O vacancies. By generating DFT band structures and projected densities of states, we explored the usefulness of almost the entire row of $3d$ transition metals, examining most in 3+ charge state. We invoked simple scaling arguments examined the trends within the band structures of the systems to make qualitative predictions about their response. We concluded that a Ti substitution does not prefer to maintain a 3+ charge state, and instead likely attempts to donate an electron back to the conduction band. A V substitution, while exhibiting midgap impurity states, introduced them at a high position relative to the valence band edge and for this reason we assumed that V in a 3+ charge state was also not a viable dopant candidate.

Our scaling arguments in that study showed that, Cr, Fe, Co, Ni may each present midgap states which could be of some experimental use, though with varying acceptor and donor natures and in different spectral regimes. Mn also demonstrated midgap states of mixed donor and acceptor nature, which we took to indicate a possibility of it exhibiting hole-electron grating competition. Our survey mainly served to demonstrate that DFT can be used to efficiently examine qualitative trends which might help as a first screening of interesting dopant candidates for tuning the spectral response of $KNbO_3$.

Defects that exist in multiple charge states are similarly important to the field of photorefractive enhancement. Questions of which states may serve as charge carrier traps and which donate charge carriers under excitation are vital to experimental design. An excellent

experimental study on the states introduced by Fe doping of $KNbO_3$ was conducted by Basun *et al.* [48] and so, in order to utilize DFT to the fullest extent that we can, we developed models to explore the Fe dopant in different defect centers. We studied the Fe[Nb], Fe[K] – V_K, and Fe[Nb] – V_O defect complexes, investigating Fe in the 2+ and 3+ charge states. In order to do so, we used charge-compensating jellium backgrounds to introduced external charge into the system and manipulate the charge state of the Fe dopant. We found that for $KNbO_3$ charge-compensating jellium backgrounds were a useful proxy for non-local mechanisms of charge compensation. We explored a procedure wherein the equilibrium coordinates of the systems with Fe in the 3+ and with Fe in the 2+ state were found, and then a number of intermediate structures were generated by linearly interpolating between the equilibrium positions of the two ionization states. Collecting band structures and projected densities of states for each of these intermediate structures allowed us to examine how these states changed under the perturbation of forcing a 3+ ion into the equilibrium geometry of the 2+ state.

By concerning ourselves only with the relative positions of energy levels, we found that the level assignments made by Basun *et al.* could be matched quite well to various states seen in our simulations. Assignments of the Fe[K] – V_K system agreed quite well with their observations. In particular, they observed states that they ascribed to the Fe in a 2+ and 3+ charge state, and did not observe the presence of any lines which indicated Fe in a 4+ charge state (at variance with the Fe[Nb] case). Our simulations revealed that such a state is buried in the valence band, and thus would be difficult to observe spectroscopically.

In that study, we also found that for the Fe[Nb] complex, Fe in a 3+ charge state is stabilized at some intermediate values of Q. This is unexpected behavior and may indicate a failure of the model, or may be the artifact of those levels experiencing an avoided crossing. However, whether this is interesting physics or a limitation to our method, its discovery could still prove useful. The exact nature of the center reported by Basun *et al.* has not been well understood to date, but our study provides insight into the Fe[Nb] – V_O center. It behaves

similarly to Fe[Nb], with some occupied impurity states being stabilized at intermediate values of Q. It is also observed that an unoccupied, spin down e_g state which was just barely at the conduction band edge was shown to relax away from the conduction band, coming into a relative position that we felt aligned with a proposed V_O level in the work by Biaggio *et al.* [91]

A larger result of these studies is the development of a methodology by which DFT may be employed at several levels to thoroughly explore a candidate dopant-defect system prior to needing to expend time and resources growing such a crystal. A first pass exploring many interesting dopants in one or more defect complexes can be used to get a sort of "lay of the land" concerning what types of impurity states may be introduced and furthermore, how these features change as the dopant is varied. If a promising dopant already exists, the space of possible defect complexes can similarly be explored and defect levels predicted and assigned at a qualitative level. If a system is particularly costly to grow, and enough computational resources are available, these early results may serve as a foundation for the employment of the GW approximation to bring these qualitative trends into the realm of quantitative predictions. It is to be emphasized that this methodology relies on a feedback between experiment and simulation. The models developed must be guided by chemical intuition and constrained by experimental capacity, while in turn, the simulations may be used to guide experimental design and to aid in the interpretation of physical results.

While our methodology produced reasonable alignment with the experimental result of Basun *et al.*, the current holy grail of *ab initio* predictions would be quantitatively accurate predictions of spectra found under the GW approximation. We studied the effects of the GW approximation on the pure phases of $KNbO_3$ as a testbed. This is well warranted, as the convergence procedures for GW calculations become significantly more complex and expensive, and having ideas of the convergence behavior of related systems may give some hint as to starting points for converging supercell systems. We observed the effects of the GW approximation on the band structures of the pure systems. We also used these results

to place estimates on the cost of applying the GW approximation to doped supercells of $KNbO_3$.

We concluded that such calculations were beyond the capacity of the computational resources currently available to us, however many levels of theory exist between standard Kohn-Sham DFT and GW which ought to be examined. Furthermore, we see a number of ways that our methodology may be refined and extended to offer as much predictive capacity as possible to aid in the development and tuning of materials.

Each step of the work done here has a number of refinements which should be considered. These steps should also be taken largely in tandem with experimental design. There is an overabundance of computational effort which could be expended in many avenues of exploration. The questions of how, where, and at what level this effort should be expended are ones which should largely be determined by experimental need.

Chapter 5 described a number of levels of theory in between standard Kohn-Sham DFT and GW approximation methods which could be undertaken to look at interesting systems in more depth. The DFT calculations performed in this work already serve as a viable testing ground for these levels of theory. The application of hybrid functionals requires some additional convergence testing, however the GGA and hybrid functionals tend to produce similar equilibrium lattice parameters [117]. The GGA-derived structures in this work would likely already suffice for use in calculations with hybrid functionals. Research concerning the usefulness of hybrid functionals for calculating defect properties shows promising degrees of accuracy [118, 119]

Additionally, the supercells created in this work could be adapted for use in TDDFT studies. One particularly interesting avenue of study would be the use of hybrid functional band structures combined with TDDFT studies to see if level assignments made in Chapter 4 hold at higher levels of theory and if so, whether these assignments can be made directly to absorption features calculated in TDDFT. Confirming such behavior would be a large step forward for our methodology, as hybrid functionals and TDDFT represent a much more

affordable alternative to GW calculations, and experimentally relevant systems could be studied at this level with more modest computational resources, even as some level of GW application is pursued simultaneously.

For studies requiring the study of a large number of dopants, we have already shown that DFT calculations can be a useful screening procedure. Procedures at the basic DFT level lend themselves quite well to high-throughput computation schemes. Architectures for these types of workflows have been solidly established in the literature [120–122]. A large body of work has already demonstrated the usefulness of high-throughput DFT for research purposes and materials discovery [123–126].

Workflows applied at the scale of many hundred of different systems do require significant optimizations. Performing convergence testing, examining pseudopotential behavior, and even storing and examining data for each individual system would involve a time expenditure which would render the benefits of *ab initio* calculations moot. Fortunately much work has also been done in this vein. A database of pseudopotentials which are designed to optimize softness and transferability was designed by Garrity *et al.* [127]. This move away from "case-by-case" construction is vital to ameliorating the need for pseudopotential testing which could hinder large-scale studies.

We believe that our methodology presented here could well be adapted to such high-throughput schemes, though we note that refinements to our procedures should still be considered at smaller scale.

Finally, the recent rise in machine learning techniques and applications has made analyzing large amounts of complex data feasible. If high-throughput calculation schemes are pursued, they can produce abundant data which can be utilized by machine learning algorithms to aid in predictions of material properties. There is a vast and growing body of research which has been combining these two approaches [128], and it stands to reason that such methods may be useful in extending the methodology we explore in this work.

Appendix A

Proof of the Hohenberg-Kohn Theorems

Here we re-state and present the mathematical proofs of the first and second Hohenberg-Kohn theorems. The proofs presented here follow the notation of Kohanoff [53] and can be found there.

Hohenberg-Kohn I: *The external potential of a system of interacting particles is uniquely determined by the ground-state electron density, apart from an additive constant.*

We proceed by contradiction, assuming that we can have two unique external potentials produce the same ground-state *density*.

Assume two Hamiltonians, H and H' that differ only in the *external* potential they experience, V and V'. Assume also, a set of eigenfunctions for each of these Hamiltonians, $\{|\psi_i\rangle\}$ and $\{|\psi_i'\rangle\}$ with particular ground-states, $|\psi_0\rangle$ and $|\psi_0'\rangle$.

These ground-state wavefunctions have corresponding energies defined through:

$$E_0 \;=\; \langle\psi_0|\hat{H}|\psi_0\rangle \text{ and } \;\; E_0' \;=\; \langle\psi_0'|\hat{H}'|\psi_0'\rangle. \tag{6.1}$$

By definition of the ground state, $|\psi_0'\rangle$ is not the ground-state eigenfunction of $\hat{H}$ and we thus make the following statements:

$$E_0 \;<\; \langle\psi_0'|\hat{H}|\psi_0'\rangle \;+\; \langle\psi_0'|\hat{H}-\hat{H}'|\psi_0'\rangle \;=\; E_0' \;+\; \int d\boldsymbol{r}\rho(\boldsymbol{r})[V(\boldsymbol{r})-V'(\boldsymbol{r})]$$

$$E_0 \;<\; E_0' \;+\; \int d\boldsymbol{r}\rho(\boldsymbol{r})[V(\boldsymbol{r})-V'(\boldsymbol{r})], \tag{6.2}$$

where the last term in the last statement uses Eq. 2.8. We now repeat this process with the primed and unprimed quantities interchanged:

$$\begin{aligned} E_0' \;&<\; \langle\psi_0|\hat{H}'|\psi_0\rangle \;+\; \langle\psi_0|\hat{H}'-\hat{H}|\psi_0\rangle \\ &=\; E_0 \;+\; \int d\boldsymbol{r}\rho(\boldsymbol{r})[V'(\boldsymbol{r})-V(\boldsymbol{r})] \\ &=\; E_0 \;-\; \int d\boldsymbol{r}\rho(\boldsymbol{r})[V(\boldsymbol{r})-V'(\boldsymbol{r})] \end{aligned} \tag{6.3}$$

$$E_0' \;<\; E_0 \;-\; \int d\boldsymbol{r}\rho(\boldsymbol{r})[V(\boldsymbol{r})-V'(\boldsymbol{r})],$$

where our initial assumption allows us to say that the wavefunctions $|\psi_0'\rangle$ and $|\psi_0\rangle$ produce the same density ρ. Now adding the final expressions from both of these procedures, we arrive at the conclusion:

$$E_0 \;+\; E_0' \;<\; E_0' \;+\; E_0 \tag{6.4}$$

This is clearly a contradiction as a quantity cannot be less than itself. Our assumption that the density term in the last lines of equations (6.2) and (6.3) was equivalent has thus been proven to be false. We conclude that the ground-state density produced by the external potentials V and V' must be unique to each of them.

Hohenberg-Kohn Theorem II: *For any given external potential, a functional of energy in terms of the particle density can be defined. The functional attains its minimum*

only at the ground-state density

Let $\rho'(\boldsymbol{r})$ be a non-negative density, normalized to the number of electrons and define the variational energy as a functional of the electron density as:

$$\begin{aligned} E[\rho'] \;&=\; F[\rho'] \;+\; \int d\boldsymbol{r}\rho'\boldsymbol{r}v_{\text{ext}}(\boldsymbol{r}), \\ \text{where } F[\rho'] \;&=\; \langle\Psi[\rho']|\hat{T}+\hat{U}_{\text{ee}}|\Psi[\rho']\rangle, \end{aligned} \tag{6.5}$$

with $|\Psi[\rho']\rangle$ the ground-state wavefunction in a potential which has ρ' as its ground-state density. So $E_0 = E_v[\rho]$ ensures $E_0 < E_v[\rho']$ for any $\rho' \neq \rho$, thus:

$$\langle\Psi[\rho']|\hat{H}|\Psi[\rho]\rangle \;=\; E_v[\rho'] \;\geq\; E_v[\rho] \;=\; E_0 \;=\; \langle\Psi[\rho]|\hat{H}|\Psi[\rho]\rangle. \tag{6.6}$$

Appendix B

Exchange-Correlation Functionals

As mentioned in Chapter 2, the choice of the exchange-correlation functional in DFT comprises the largest approximation with the Kohn-Sham framework. We focus here on the details of the implementation and performance of the two most widely-used exchange-correlation functionals, the Local Density Approximation (LDA) and the Generalized Gradient Approximation (GGA). These details are explored much more fully in [53]

B1 Local Density Approximation

The LDA was developed as the first means of implementing the Kohn-Sham formalism. It is based around the exact calculations for the exchange energy which was calculated in Dirac's formulation of the homogeneous electron gas (HEG). This was an exact result and thus was a helpful starting point to calculate exchange energies in solids.

The central idea of the LDA is to assume that within a small integration volume, the electron density of the material is approximately even. In this case, locally, approximating the electrons as behaving like an HEG can hold quite well. In the limit of uncorrelated materials, the LDA becomes exact. A helpful approach is to split and separately calculate the exchange (X) and correlation (C) energies.

$$\epsilon_{\mathrm{XC}}^{\mathrm{LDA}}[\rho] \;=\; \epsilon_X[\rho(\boldsymbol{r})] \;+\; \epsilon_C[\rho(\boldsymbol{r})]$$

$$\epsilon_{\mathrm{X}}[\rho] \;=\; -\frac{3}{4}\left(\frac{3\rho(\boldsymbol{r})}{\pi}\right)^{\frac{1}{3}}$$

$$\epsilon_{\mathrm{C}}^{\mathrm{PZ}}[\rho] \;=\; \begin{cases} A\ln(r_s) + B + Cr_s\ln(r_s) + Dr_s & r_s \leq 1 \\ \gamma/(1+\beta_1\sqrt{r_s} + \beta_{2r_s}) & r_s > 1 \end{cases},$$

where $r_s = (3/4\pi\rho)^{(\frac{1}{3})}$ is the mean interelectornic distance expressed in atomic units. The correlation energy is based on quantum Monte Carlo results by Ceperley and Alder [129]. The PZ superscript in the correlation energy refers to the parametrization of those results by Perdew and Zunger [130]. The parameters $A, B, C,$ and, D all have characteristic values and the parameters $\gamma, \beta_1,$ and, β_2 are constants.

Despite the initial impression of crudeness, the LDA has demonstrated remarkable success over a wide range of systems. We list here some trends and limitations of the approximation scheme:

- The LDA favors electronic densities which are more homogeneous compared to exact results; as a consequence of this binding energies of molecules and cohesive energies of solids tend to be overestimated.

- Bond lengths tend to be underestimated due to the over-binding problem, however these lengths, bond angles and vibrational frequencies tend to be reproduced to within about 1% of experiment. Elastic constants and phonon frequencies tend to be slightly underestimated, but also well-reproduced. Dielectric properties are overestimated by about 10%.

- For weakly bound system involving hydrogen bonds or van der Waals interactions, bond lengths are far too short and dispersion interactions are poorly reproduced.
- Chemical trends such as ionization potentials and transfer energies between electronic states of different symmetries are usually correct.
- Electronic densities where the electrons are very localized are quite poorly calculated. The LDA fails to perfectly cancel the self-interaction, which becomes very important for strongly localized states.
- Band gaps of strongly correlated materials (e.g. NiO, V_2O_3 etc.) are opened by the correlations within the materials, but the LDA does not always reproduce this behavior.

B2 Generalized Gradient Approximation

A natural extension to the LDA would be to incorporate some amount of information about the variations of the electron density about a local point of integration. To this end, the GGA was developed as an attempted first-order improvement over the LDA.

Under the GGA the exchange-correlation energy takes the form:

$$E_{\mathrm{XC}}^{\mathrm{GGA}}[\rho] \;=\; \int d\boldsymbol{r}\rho(\boldsymbol{r})\epsilon_{\mathrm{XC}}F_{\mathrm{XC}}[\rho,\nabla\rho,\nabla^2\rho,...], \tag{6.7}$$

where, F_{XC} is a factor which modifies the exchange-correlation energy based on derivatives of the electron density. The effect of this factor on the exchange and correlation energies are separately accounted for. The exchange part is modified by:

$$F_{\mathrm{X}}(s) \;=\; 1 \;+\; \kappa\left(1 \;-\; \frac{1}{1+\mu s^2/\kappa}\right), \tag{6.8}$$

where $s = \frac{|\nabla\rho|}{2k_{\mathrm{F}}\rho}$, with k_{F} being the magnitude of the Fermi wave-vector. $\mu = \beta(\frac{\pi^2}{3})$ where β is a parameter with a characteristic value. κ is a parameter that is designed to control the

behavior of F_X in the limit that the density is vanishingly low, that is to say $s \to \inf$ (known as the Lieb-Oxford bound).

The modification to the correlation part of the energy is quite complicated, but comes in the form of an additional contribution to the LDA correlation energy:

$$E_{\mathrm{C}}^{\mathrm{GGA}} \;=\; \int d\boldsymbol{r}\rho(\boldsymbol{r}) \left[\epsilon_{\mathrm{C}}^{\mathrm{LDA}}[\rho,\zeta] \;+\; H[\rho,\zeta,t]\right], \tag{6.9}$$

with

$$H[\rho,\zeta,t] \;=\; \left(\frac{e^2}{a_0}\right)\gamma\phi^3 \ln\left(1 \;+\; \frac{\beta t^2}{\gamma}\left[\frac{1+At^2}{1+At^2+A^2t^4}\right]\right), \tag{6.10}$$

where

$$\begin{aligned} t \;&=\; |\nabla\rho|/(2\phi_{\mathrm{TF}}\rho) \\ \phi \;&=\; [(1\;+\;\zeta)^{2/3} \;+\; (1\;-\;\zeta)^{2/3}]/2 \\ A \;&=\; \frac{\beta}{\gamma}[e^{-\epsilon_{\mathrm{C}}^{\mathrm{LDA}}[\rho]/(\gamma\phi^3 e^2/a_0)} \;-\; 1]. \end{aligned} \tag{6.11}$$

Despite its increased sophistication, the GGA is not without its limitaitons. While the GGA ameliorates some shortcomings of the LDA, it is not a systematic improvement.

- The GGA improves the binding energies and atomic energies over the LDA.
- Bond lengths and angles are improved.
- For $4d$ and $5d$ transition metals, there is not a clear improvement over the LDA.
- There is some improvement in the gap energy, and subsequently the dielectric constant, but this is not substantial.

An extensive benchmarking of a number of different functionals can be found in work done by Goerigk *et al.* [131].

B3 Beyond LDA and GGA

Functionals moving beyond the LDA and GGA have been attempted. The family of meta-GGA functionals seek to consider the fourth-order gradient expansion of the exchange-correlation energy. The central philosophy behind them is that the incorporation of additional semi-local information may improve some properties of the GGA.

Other approaches have focused on requiring the functional to obey as many physical constraints as possible. These strongly constrained, appropriately normalized (SCAN) functionals also fall into the category of meta functionals. A review and comparison of these types of functionals can be found in the work by Zhang *et al.* [132]

These functionals provide some improvements in various properties depending on the characteristics of the system under investigation. We will not expound on their details any more here, but direct an interested reader to studies done by Sun *et al.* [133, 134]

Appendix C

Quantum Espresso Details

Quantum Espresso (QE), or the op**E**n **S**ource **P**ackage for **R**esearch in **E**lectronic **S**tructure, **S**imulation, and **O**ptimization, is an integrated suite of codes for electronic-structure calculations and is our tool of choice for DFT calculations. It consists of a number of core packages with interoperable, third-party add ons. We describe here some of the core executables and the general workflow to produce basic DFT bandstructures.

C1 Executables

- **pw.x**: The heart of the QE package, pw.x is the executable responsible for solving the Kohn-Sham equations. It has several calculation modes which can be run.

  Most important is the self-consistent field, 'scf', option which takes the user-specified input data, generates the Kohn-Sham Hamiltonian and solves the Kohn-Sham equations to self-consistency. Self-consistency is determined when the estimated energy error of the electronic system falls below a certain, user-defined threshold. The output of this calculation is the set of self-consistent single-particle wavefunctions over the specified set of k-points and the self-consistent potential.

  It is also possible to run the program non-self consistently under the 'nscf' calculation mode, which runs the Kohn-Sham mechanism for one iteration. This will not produce

a guaranteed set of solutions to the Kohn-Sham equations, but rather is run using results from an scf calculation. It is useful for interpolating to a finer grid of k-points, yielding approximate wavefunctions over this tighter grid.

The **pw.x** executable also has a number of modes used for structural optimization. Mode 'relax' performs an alternating set of self-consistent iterations to get the Kohn-Sham wavefunctions for a specific ionic configuration and ionic relaxations where the Hellman-Feynman forces on the ions are calculated. The ions are moved according to a quasi-Newtonian algorithm and the self-consistent iterations are performed again for the new configuration. This is repeated until the forces are below a specified threshold and then the equilibrium ion positions are reported. The calculation mode 'vc-relax' performs a similar procedure but it also calculates the stress on the cell and will relax the lattice parameters themselves to find equilibrium lattice parameters. Both of these procedures can also be performed using *ab initio* molecular dynamics [135].

The last calculation mode of importance to us is the 'bands' mode. In this mode, we can non-self consistently solve for Kohn-Sham states over a specified path of k-points in the Brillouin zone. This must obviously be done after obtaining scf results.

- **dos.x**: The **dos.x** executable is a post-processing utility which calculates the total density of states (DOS). A specified smearing method of the occupation function convolves the states with a smearing function (*e.g.* a gaussian) of a specified width in order to produce reasonably smooth DOS curves. For spin-polarized calculations, it calculates the DOs of the spin-up and spin-down electron densities separately.

- **projwfc.x**: The **projwfc.x** is another post-processing utility which projects the electronic states onto the set of atomic orbitals, producing the so-called projected density of states (pDOS). This helps to identify which atoms contribute most strongly to each states. These results must be taken with some amount of caution as the set of atomic orbitals is not complete, and thus the projection is not perfect. This incomplete projec-

tion is captured in the reported Lowdin spilling parameter which reflects the percentage of the states which cannot be represented by this projection.

- **bands.x**: The **bands.x** works on the results of a 'bands' nscf calculation. It takes the eigenvalues along the specified k-point path and orders the results to form the bands. The output files from this utility are easily plotted band structures.

- **pp.x** The **pp.x** program is a multi-faceted post-processing utility which encompasses a number of different functions. It can be used to extract the electron density, the total or ionic potentials, the spin-polarization, the norms of wavefunctions contributing to specified bands, among many other functionalities.

C2 Workflow

We discuss here details concerning the quality of our calculations. In particular, the first issue we focus on is convergence of the results. As mentioned in Chapter 2, this focuses largely on the energetic cutoff for the wavefunctions and the density of the k-point mesh used to sample the Brillouin zone. The increase in either of these parameters will increase the computational cost of the simulation. The challenge then, is to determine parameters which are "good enough" to yield reliable results.

We do this for the wavefunction cutoff (ecutwfc parameter in **pw.x** input files) by simply running scf calculations over a range of cutoff energies with a fixed k-point mesh and comparing the resulting total energies of the system. In this work, we generally choose an energetic cutoff which converges the total energy of the system to within a level of approximately 10 mRy. Fortunately, the convergence of the total energy with respect to the cutoff energy is largely independent of the convergence with respect to the density of k-points. We can repeat this process for an increasingly dense k-point mesh. This is quite simple as QE needs only a simple set of directives to automatically generate a Monkhorst-Pack grid of a specified dimension.

The cutoffs chosen are discussed in the relevant chapters, but it should be noted that they depend on the type and quality of pseudopotentials used. In particular, elements with high angular momentum channels will generally require higher cutoff energies.

Once parameters for good convergence are established, our workflow for producing band structures is as follows: a self-consistent calculation is run with the chosen convergence parameters. Then a non-self-consistent calculation at a denser k-point mesh is run. From the results of the nscf calculation, we can obtain the charge density map, DOS, and pDOS. We then run a bands calculation over a specific path of high-symmetry points in the Brillouin zone and the **bands.x** post-processing using those results.

Appendix D

BerkeleyGW Details

D1 Introduction

The BerkeleyGW code (BGW) is a suite of programs for GW calculations. It is interoperable with a number of DFT code bases, including Quantum Espresso. A flow chart of how the various executables generate results is presented in Fig. D1. We explore the details of each step of the process here, but a full explanation is found in the paper by Deslippe *et al.* [70]

D2 Executables

As explored in Chapter 2, the single-particle wavefunctions produced by DFT calculations are a useful starting point for GW calculations. As such, the first step in the workflow of BGW is to produce well-converged DFT results. From here, the executables of the BGW suite are used to perturbatively arrive at GW corrected results.

- **epsilon.x**: The first step of the BGW process uses the plane-wave wavefunctions of the DFT results to build and invert either the static or dynamic Random Phase Approximation (RPA) dielectric function. For the static dielectric function, the RPA polarizability is first calculated by:

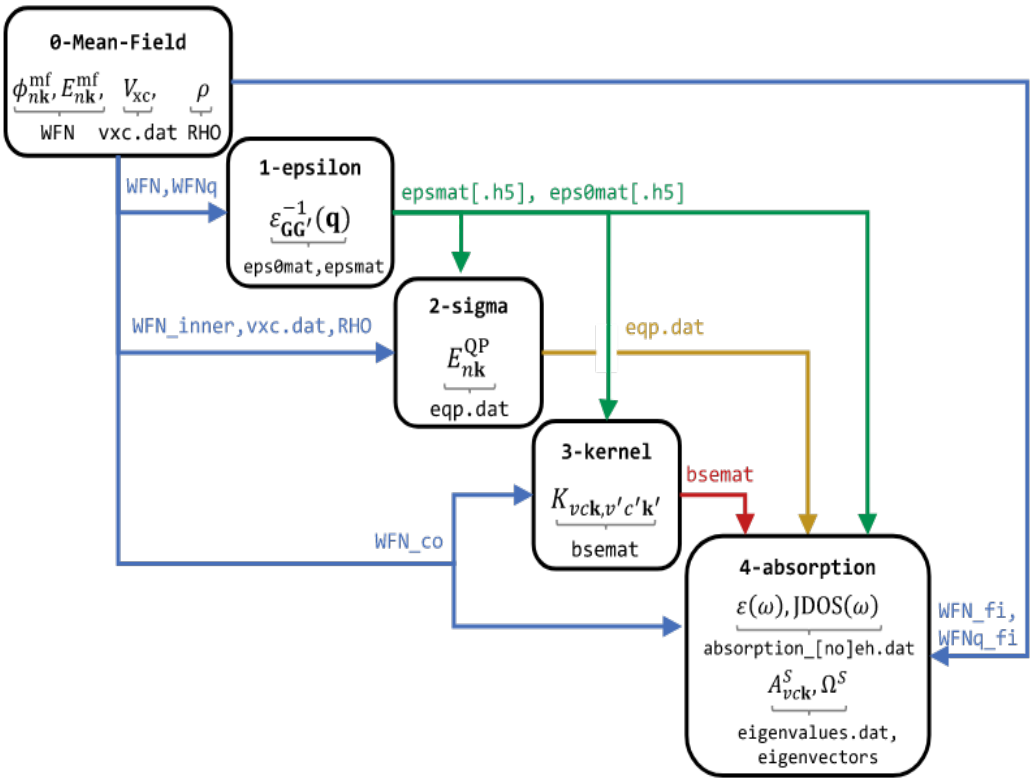

Figure D1: A chart demonstrating the workflow for GW calculations within the BerkeleyGW code suite. Necessary inputs and created outputs are shown for each of the 4 main executables.

$$\chi_{\boldsymbol{GG'}}(\boldsymbol{q};0) \;=\; \sum_{n}^{\text{occ}}\sum_{n'}^{\text{cond}}\sum_{\boldsymbol{k}} M_{nn'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G})(M^{*}_{nn'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G'}))\frac{1}{E_{n\boldsymbol{k}+\boldsymbol{q}}-E_{n'\boldsymbol{k}}}, \tag{6.12}$$

with

$$M_{nn'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G'}) \;=\; \langle n\boldsymbol{k}+\boldsymbol{q}|e^{i(\boldsymbol{q}+\boldsymbol{G}\cdot\boldsymbol{r})}|n'\boldsymbol{k}\rangle. \tag{6.13}$$

The $M_{nn'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G'})$ factors are plane-wave transition elements, $\boldsymbol{q}$ is a vector in the first Brillouin zone, $\boldsymbol{G}$ is a reciprocal lattice vector, and the $E_{n\boldsymbol{k}}$ and the kets are the eigenvalues and eigenstates respectively of the DFT calculation. The number of $\boldsymbol{G}$-vectors for which Eq. (6.13) is evaluated is determined by a user-defined energy cutoff, called now the dielectric cutoff in BGW.

The RPA polarizability allows for the construction of the the dielectric function through:

$$\epsilon_{\boldsymbol{GG'}}(\boldsymbol{q};0) \;=\; \delta_{\boldsymbol{GG'}} \;-\; v(\boldsymbol{q}+\boldsymbol{G})\chi_{\boldsymbol{GG'}}(\boldsymbol{q};0), \tag{6.14}$$

where $v(\boldsymbol{q}+\boldsymbol{G}) = \frac{4\pi}{|\boldsymbol{q}+\boldsymbol{G}|^2}$ is the Fourier transform of the bare Coulomb interaction.

We note that the extension for calculating the frequency dependence of the dielectric function involves the calculation the retarded and advanced dielectric functions:

$$\begin{aligned}\epsilon^{\text{r/a}}_{\boldsymbol{GG'}}(\boldsymbol{q};\omega) \;&=\; \delta_{\boldsymbol{GG'}} - v(\boldsymbol{q}+\boldsymbol{G})\sum_{n}^{\text{occ}}\sum_{n'}^{\text{cond}}\sum_{\boldsymbol{k}} M_{nn'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G})M^{*}_{nn'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G'}) \\ &\times\frac{1}{2}\left[\frac{1}{E_{n\boldsymbol{k}+\boldsymbol{q}}-E_{n'\boldsymbol{k}}-\omega\mp i\delta}+\frac{1}{E_{n\boldsymbol{k}+\boldsymbol{q}}-E_{n'\boldsymbol{k}}+\omega\pm i\delta}\right].\end{aligned} \tag{6.15}$$

The expressions in Eq. (6.15) are evaluated over a grid of frequencies, with δ being a broadening parameter which must in be tuned, ideally minimizing it. The upper signs are taken for the retarded function and the lower signs for the advanced function. In the code, a lower and upper frequency cutoff are specified as well as a frequency spacing. One must refine this grid in order to truly converge the calculations if the dynamical dielectric function is required.

The process involves three computationally intensive steps: (i) the calculation of the matrix elements in Eq. (6.13); (ii) the summations therein; and (iii) the inversion of the dielectric function. The first step of this process may be made to scale as $\sim N^3 \log(N)$, with N the number of atoms in the system. With the matrix elements calculated, the next step scales formally as N^4, however in practice, this step typically takes up a small percentage of the calculation. The inversion of the dielectric function, necessary for the calculation of the of the screened Coulomb interaction, scales like N^3, but the process can be parallelized well on up to hundreds of processors and also represents only a small percent of the calculation time.

Importantly, the calculation of the dielectric function at $\boldsymbol{q} = 0$ for the $\boldsymbol{G} = \boldsymbol{G}' = 0$, that is the head of the dielectric function, clearly diverges. The BGW code handles this by requiring a DFT calculation over a k-grid that is shifted by a $\boldsymbol{q}$-vector with a very small magnitude. The limit $\boldsymbol{q} \to 0$ is then taken numerically. Metals and semi-conductors diverge differently in this limit, and this behavior is handled by a user-specified tag in the input file.

The **epsilon.x** requires two different DFT calculations. The first of these calculations is done with the desired number of conduction bands and uses a k-point grid which is generated by an included BGW utility, **kgrid.x**. This produces the main information that BGW will use to calculate the dielectric function, collected by another utility program, **pw2bgw.x** which converts the information from the Quantum Espresso data files into a form that can be used by BGW. The second calculation is performed using

only the occupied bands of the system but with a k-grid shifted by the aforementioned small $\boldsymbol{q}$-vector. The input file for **epsilon.x** then requires the user to specify the number of conduction bands to be included in the summation in Eq. (6.13), the grid of points over which the unshifted wavefunction information is collected, and a tag determining whether the system is semi-conducting or metallic in order to handle the $\boldsymbol{q} \to 0$ limit.

Convergence of these epsilon calculations is quite complicated. There is, in fact, only one convergence parameter. One fixes either the dielectric cutoff energy or the number of bands and adjusts the other parameter to match. The highest band included should match the dielectric cutoff. Fortunately, BGW provides an easily plotted data file which shows the value of the polarizability using partial sums up to the number of conduction bands specified in the input. This file contains the convergence information for each $\boldsymbol{q}$-point specified.

- **sigma.x**: With the dielectric function, or rather, its inverse, now in hand, the next step of the process is to construct the self-energy operator. It computes a representation of the Dyson's equation in the basis of the DFT wavefunctions:

$$\langle \psi_{n\boldsymbol{k}} | H^{\mathrm{QP}}(\omega) | \psi_{m\boldsymbol{k}} \rangle \;=\; E^{\mathrm{MF}}_{n\boldsymbol{k}} \delta_{n,m} \;+\; \langle \psi_{n\boldsymbol{k}} | \Sigma(\omega) - \Sigma^{\mathrm{MF}}(\omega) | \psi_{m\boldsymbol{k}} \rangle, \tag{6.16}$$

with ω an energy parameter that should be set self-consistently to the quasi-particle eigenvalues, however obtaining self-consistency in the GW approximation requires some effort (see Appendix E). $E^{\mathrm{MF}}_{n\boldsymbol{k}}$ are the DFT eigenvalues, $\psi_{n\boldsymbol{k}}$ are the DFT eigenfunctions, and Σ^{MF} is an approximation to the self-energy operator, such as the exchange-correlation potential from the DFT step. Though the full equation is presented in Eq. (6.16), often the DFT eigenvalues and eigenfunctions are a good enough approximation to the quasi-particle solutions that only the diagonal elements need be considered, and in practice the off-diagonal elements by and large do not significantly affect the results.

The **sigma.x** program requires DFT results from a grid which is identical to that used for **epsilon.x**, but with some differences in the Quantum Espresso to BGW conversion utility input file.

The matrix elements in Eq. (6.16) can be evaluated by **sigma.x** within various approximations (*e.g.* the Hartree-Fock scheme, the static Coulomb-hole screened-exchange method, GW under a generalized plasmon pole model, or full-frequency GW calculations). The static Coulomb-hole screened-exchange (COHSEX) method breaks the many-body excitation into two components, $\Sigma = \Sigma_{\mathrm{CH}} + \Sigma_{\mathrm{XC}}$, one which captures the interaction of the excited electron and the hole its excitation generates and the other which captures the effects of the screening of the hole by the medium and the exchange energy. This method does not dynamically screen the hole, hence the static prefix. In the GW and static-COHSEX, the Σ_{SX} and Σ_{CH} parts are defined as:

$$\begin{aligned}\langle n\boldsymbol{k}|\Sigma_{\mathrm{SX}}(\omega)|n'\boldsymbol{k}\rangle &= -\sum_{n''}^{\mathrm{occ}}\sum_{\boldsymbol{qGG'}} M^*_{n''n}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G})M_{n''n'}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G'}) \\ &\times [\epsilon^r_{\boldsymbol{GG'}}]^{-1}(\boldsymbol{q};\omega - E_{n''\boldsymbol{k}-\boldsymbol{q}})v(\boldsymbol{q}+\boldsymbol{G'})\end{aligned} \tag{6.17}$$

and

$$\begin{aligned}\langle n\boldsymbol{k}|\Sigma_{\mathrm{CH}}(\omega)|n'\boldsymbol{k}\rangle &= \frac{i}{2\pi}\sum_{n''}^{\mathrm{occ}}\sum_{\boldsymbol{qGG'}} M^*_{n''n}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G})M_{n''n'}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G'}) \\ &\times \int_0^\infty d\omega' \frac{[\epsilon^r_{\boldsymbol{GG'}}]^{-1}(\boldsymbol{q};\omega) - [\epsilon^a_{\boldsymbol{GG'}}]^{-1}(\boldsymbol{q};\omega')}{\omega - E_{n''\boldsymbol{k}-\boldsymbol{q}} - \omega' + i\delta} v(\boldsymbol{q}+\boldsymbol{G'}).\end{aligned} \tag{6.18}$$

In practice, the code calculates the matrix elements of the bare exchange Σ_X, related to Eq. (6.18) by replacing the retarded dielectric function with a delta function, $\delta_{\boldsymbol{GG'}}$,

and also calculates $\Sigma_{\rm SX} - \Sigma_{\rm X}$. The frequency integral is done numerically over the same frequency grid that was specified in the frequency-dependent **epsilon.x** calculation.

One may also account for the frequency dependence in an approximate way which saves the need to converge over a grid of frequencies. The main method of doing this is known as the generalized plasmon pole (GPP) model. This model attempts to account for the energy dependence of the polarizability in terms of a discrete set of energy bands [136]. In the **sigma.x** code of BGW, the GPP equations corresponding to Eqs. (6.17 and 6.18) are:

$$\begin{aligned}\langle n\boldsymbol{k}|\Sigma_{\rm SX}(\omega)|n'\boldsymbol{k}\rangle &= \frac{i}{2\pi}\sum_{n''}^{\rm occ}\sum_{\boldsymbol{q}\boldsymbol{G}\boldsymbol{G}'} M^*_{n''n}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G})M_{n''n'}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G}') \\ &\times\left[\delta_{\boldsymbol{G}\boldsymbol{G}'} + \frac{\Omega^2_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})(1-i\tan\phi_{\boldsymbol{G}\boldsymbol{G}'(\boldsymbol{q})})}{(\omega-E_{n''\boldsymbol{k}-\boldsymbol{q}})^2-\tilde{\omega}^2_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})}\right]v(\boldsymbol{q}+\boldsymbol{G}')\end{aligned} \tag{6.19}$$

and

$$\begin{aligned}\langle n\boldsymbol{k}|\Sigma_{\rm CH}(\omega)|n'\boldsymbol{k}\rangle &= \frac{i}{2\pi}\sum_{n''}^{\rm occ}\sum_{\boldsymbol{q}\boldsymbol{G}\boldsymbol{G}'} M^*_{n''n}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G})M_{n''n'}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G}') \\ &\times\frac{\Omega^2_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})(1-i\tan\phi_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q}))}{\tilde{\omega}_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})(\omega-E_{n''\boldsymbol{k}-\boldsymbol{q}}-\tilde{\omega}_{\boldsymbol{G}\boldsymbol{G}'})},\end{aligned} \tag{6.20}$$

where $\Omega_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})$ is the effective bare plasma frequency, $\tilde{\omega}_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})$ is the GPP mode frequency, and $\phi_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})$, is the phase of the renormalized $\Omega^2_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})$, respectively. They have the forms:

$$\begin{aligned}\Omega^2_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q}) &= \omega_{\rm p}^2\frac{(\boldsymbol{q}+\boldsymbol{G})(\boldsymbol{q}+\boldsymbol{G}')}{|\boldsymbol{q}+\boldsymbol{G}|^2}\frac{\rho(\boldsymbol{G}-\boldsymbol{G}')}{\rho(\boldsymbol{0})}, \\ \tilde{\omega}^2_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q}) &= \frac{|\lambda_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})|}{\cos\phi_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})}, \\ |\lambda_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})|e^{i\phi_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})} &= \frac{\Omega^2_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q})}{\delta_{\boldsymbol{G}\boldsymbol{G}'}-\epsilon^{-1}_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q};0)},\end{aligned} \tag{6.21}$$

with $\lambda_{\boldsymbol{GG'}}$ as the amplitude of the renormalized $\Omega^2_{\boldsymbol{GG'}}(\boldsymbol{q})$ and ρ is the electron charge density in reciprocal space and $\omega_{\mathrm{p}}^2 = 4\pi\rho(\boldsymbol{0})e^2/m$ is the classical plasma frequency. For the GPP model, the frequency integral of Eq. (6.18) is reduced to a single term using an analytical approximation to the frequency dependence of the dielectric matrix which requires only the static dielectric matrix $\epsilon_{\boldsymbol{GG'}^{-1}}(\boldsymbol{q};0)$. This approximation is done using the f-sum rule [137] which relates the imaginary part of the inverse dielectric function to the first term in Eq. (6.21). It should be noted that for systems without inversion symmetry ρ and V_{XC} are complex functions in reciprocal space.

For static COHSEX calculations, the self-energy operator parts can be obtained by by setting $\omega - E_{n'\boldsymbol{k}-\boldsymbol{q}}$ to zero in Eqs. (6.20). They then take the form:

$$\begin{aligned}\langle n\boldsymbol{k}|\Sigma_{\mathrm{SX}}(0)|n'\boldsymbol{k}\rangle &= -\sum_{n''}^{\mathrm{occ}}\sum_{\boldsymbol{qGG'}} M^*_{n''n}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G})M_{n''n'}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G'}) \\ &\times [\epsilon^r_{\boldsymbol{GG'}}]^{-1}(\boldsymbol{q};0)v(\boldsymbol{q}+\boldsymbol{G'})\end{aligned} \tag{6.22}$$

and

$$\begin{aligned}\langle n\boldsymbol{k}|\Sigma_{\mathrm{CH}}(0)|n'\boldsymbol{k}\rangle &= \frac{1}{2}\sum_{n''}^{\mathrm{occ}}\sum_{\boldsymbol{qGG'}} M^*_{n''n}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G})M_{n''n'}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G'}) \\ &\times \left[\epsilon^{-1}_{\boldsymbol{GG'}}(\boldsymbol{q};0)-\delta_{\boldsymbol{GG'}}\right]v(\boldsymbol{q}+\boldsymbol{G'}) \\ &= \frac{1}{2}\sum_{\boldsymbol{qGG'}} M_{nn'}(\boldsymbol{k},\boldsymbol{q}=0,\boldsymbol{G'}-\boldsymbol{G})\left[\epsilon^{-1}_{\boldsymbol{GG'}}(\boldsymbol{q};0-\delta_{\boldsymbol{GG'}})v(\boldsymbol{q}+\boldsymbol{G'})\right],\end{aligned} \tag{6.23}$$

where in the expression for $\Sigma_{\mathrm{CH}}(0)$, the completeness relation for the sum over empty states is used to obtain the last equation which is free any dependence on the empty orbitals.

For Hartree-Fock calculations, only the bare exchange is computed via:

$$\begin{aligned}
&\langle n\boldsymbol{k}|\Sigma_{\mathrm{X}}|n'\boldsymbol{k}\rangle \\
&= -\sum_{n''}^{\mathrm{occ}} \sum_{\boldsymbol{q}\boldsymbol{G}\boldsymbol{G}'} M^*_{n''n}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G}) M_{n''n'}(\boldsymbol{k},-\boldsymbol{q},-\boldsymbol{G}') \delta_{\boldsymbol{G}\boldsymbol{G}'} v(\boldsymbol{q}+\boldsymbol{G}').
\end{aligned} \tag{6.24}$$

Convergence of the summations in the above expressions with respect to unoccupied states is quite generally very slow. A complicating factor is that the convergence must also be checked with respect to the dielectric cutoff used as well. These issues are approached in Chapter 4 where preliminary GW calculations are explored.

- **kernel.x** and **absorption.x**: As mentioned in Chapter 2, the GW approximation effectively solves for the properties of the system characterized by non-interacting quasi-particles. Going beyond this involves using the quasi-particle information to solve for the electron-hole excited states using a Bethe-Salpeter equation (BSE):

$$(E^{\mathrm{QP}}_{c\boldsymbol{k}} - E^{\mathrm{QP}}_{v\boldsymbol{k}}) A^s_{vc\boldsymbol{k}} \; + \; \sum_{v'c'\boldsymbol{k}'} \langle vc\boldsymbol{k}|K^{\mathrm{eh}}|v'c'\boldsymbol{k}'\rangle \; = \; \Omega^s A^s_{vc\boldsymbol{k}}, \tag{6.25}$$

where s is an excitonic state label, $A^s_{vc\boldsymbol{k}}$ is an excitonic wavefunction, Ω^s is the excitation energy, and K^{eh} is the electron-hole interaction kernel. In BGW, the Tamm-Dancoff approximation is taken [138], that is, only valence to conduction band transitions are included. The exciton wavefunction is then expressed in real-space by:

$$\Psi(\boldsymbol{r}_{\mathrm{e}},\boldsymbol{r}_{\mathrm{h}}) \; = \; \sum_{v,c,\boldsymbol{k}} A^s_{vc\boldsymbol{k}} \psi^*_{v,\boldsymbol{k}}(\boldsymbol{r}_{\mathrm{h}}) \psi_{c,\boldsymbol{k}}(\boldsymbol{r}_{\mathrm{e}}), \tag{6.26}$$

with the imaginary part of the dielectric function being expressed as:

$$\begin{aligned}\epsilon_2(\omega) &= \frac{16\pi^2 e^2}{\omega^2}\sum_s |\boldsymbol{e}\cdot\langle 0|\boldsymbol{v}|s\rangle|^2\,\delta(\omega-\Omega^s)\\ \langle 0|\boldsymbol{v}|s\rangle &= \sum_{vc\boldsymbol{k}} A^s_{vc\boldsymbol{k}}\langle v\boldsymbol{k}|\mathbf{v}|c\boldsymbol{k}\rangle,\end{aligned} \tag{6.27}$$

where $\mathbf{v}$ is the velocity operator along the direction fo the polarization of light, $\boldsymbol{e}$.

The **kernel.x** executable constructs the kernel of Eq. (6.25) over a coarse k-point grid. This kernel is limited to the static approximation and is broken in to two terms, a screened direct interaction and a bare exchange interaction:

$$\langle vc\boldsymbol{k}|K^{\mathrm{d}}|v'c'\boldsymbol{k}'\rangle = -\int d\boldsymbol{r}d\boldsymbol{r}'\psi^*_c(\boldsymbol{r})\psi_{c'}(\boldsymbol{r})W(\boldsymbol{r},\boldsymbol{r}')\psi^*_{v'}(\boldsymbol{r}')\psi_{c'}(\boldsymbol{r}') \tag{6.28}$$

and

$$\langle vc\boldsymbol{k}|K^{\mathrm{x}}|v'c'\boldsymbol{k}'\rangle = -\int d\boldsymbol{r}d\boldsymbol{r}'\psi^*_c(\boldsymbol{r})\psi_{v}(\boldsymbol{r})v(\boldsymbol{r},\boldsymbol{r}')\psi^*_{v'}(\boldsymbol{r}')\psi_{c'}(\boldsymbol{r}') \tag{6.29}$$

The k-grid over which the kernel is calculated is often the same one used in the **sigma.x** executable. The matrix elements are calculated in reciprocal space using:

$$\langle vc\boldsymbol{k}|K^{\mathrm{d}}|v'c'\boldsymbol{k}'\rangle = \sum_{\boldsymbol{G}\boldsymbol{G}'} M_{cc'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G})W_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{q};0)M^*_{vv'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G}') \tag{6.30}$$

and

$$\langle vc\boldsymbol{k}|K^{\mathrm{x}}|v'c'\boldsymbol{k}'\rangle = \sum_{\boldsymbol{G}\boldsymbol{G}'} M_{cv}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G})v(\boldsymbol{q}+\boldsymbol{G})\delta_{\boldsymbol{G}\boldsymbol{G}'}M^*_{c'v'}(\boldsymbol{k},\boldsymbol{q},\boldsymbol{G}'). \tag{6.31}$$

Thus for each $\boldsymbol{k},\boldsymbol{k}'$ point, the matrix elements $M_{vv'}, M_{cc'}, M_{vc}$ must be calculated. In practice, the summations involved in Eq. (6.31) require less time than the matrix elements to calculate and the number of bands needed to converge these calculations

is often far less than those used in the preceding steps. An energy window of ∼10 eV is often enough to converge a spectrum beyond the visible region.

Excitonic binding energies and absorption spectra require sampling over a fine k-point grid in order to be properly represented. The steps involved in the calculation of the kernel of the BSE would be prohibitively expensive. This is remedied in BGW by interpolating the calculated kernel to a finer grid before solving the BSE.

The **absorption.x** executable expands the wavefunctions collected in a calculation over a fine grid in terms of the wavefunctions of the nearest k-point in the coarse grid through:

$$u_{n\boldsymbol{k}_{\rm fi}} = \sum_{n'} C^{\boldsymbol{k}_{\rm co}}_{n,n'} u_{n'\boldsymbol{k}_{\rm co}} \tag{6.32}$$

with the coefficients defined by

$$C^{\boldsymbol{k}_{\rm co}}_{n,n'} = \int d\boldsymbol{r} u_{n\boldsymbol{k}_{\rm fi}} u^*_{n'\boldsymbol{k}_{\rm co}}(\boldsymbol{r}), \tag{6.33}$$

where the coefficients of the expansion normalized, such that $\sum_{n'}|C^{\boldsymbol{k}_{\rm co}}_{n,n'}|^2 = 1$. One can improve the interpolation systematically by including more valence and conduction bands. Importantly, n and n' are restricted to both be either valence or conduction bands. Using the coefficients in Eqs. (6.33) the interpolation is completed via:

$$\langle vc\boldsymbol{k}_{\rm fi}|K|v'c'\boldsymbol{k}'_{fi}\rangle = \sum_{n1,n2,n3,n4} C^{\boldsymbol{k}_{\rm co}}_{c,n1} C^{*\boldsymbol{k}_{\rm co}}_{v,n2} C^{*\boldsymbol{k}'_{\rm co}}_{c',n3} C^{\boldsymbol{k}'_{\rm co}}_{v',n4} \langle n_2 n_1 \boldsymbol{k}_{\rm co}|K|n_4 n_3 \boldsymbol{k}'_{\rm co}\rangle. \tag{6.34}$$

With the kernel constructed on the fine grid, the last part to be tackled is the diagonalization of the BSE. This results in the excitonic eigenvalues and eigenfunctions which

can then be used to create the absorption spectrum. The BGW suite contains some post-processing utilities which allow for the plotting of these functions.

- **inteqp.x**: As long as the quasi-particle information is collected, that is, having run the **sigma.x** executable, the quasi-particle corrections may be used to obtain the corrections to band structure along an arbitrary path. This is done with the **inteqp.x** executable. The quasi-particle corrections are much smoother functions of k and ω than the quasi-particle eigenvalues themselves.

Appendix E

On the GW Approximation

Here we focus on some details of the Green's function formalism, Hedin's equations, and Feynman diagrams to give the reader more insight into the theoretical background. We follow the derivation presented in [62] to which we direct an interested reader for more information.

E1 Equation of Motion for G

We derive the expressions for the propagator, beginning from the Heisenberg equation of motion for the field operator:

$$i\frac{\partial}{\partial t}\hat{\psi}(\boldsymbol{x},t) \;=\; \left[\hat{\psi}(\boldsymbol{x},t),\hat{H}_{\text{elec}}\right], \tag{6.35}$$

where $\boldsymbol{x}$ contains the space and spin coordinates and $\hat{H}_{\text{elec}}$ is the electronic part of the many-body Hamiltonian, which can be broken down itself as: $\hat{H}_{\text{elec}} = \hat{h}_0 + v(\boldsymbol{r},\boldsymbol{r}')$. The potential in $\hat{H}_{\text{elec}}$ is the two-body interaction. Noting that the field operators $\hat{\psi},\hat{\psi}^\dagger$ are anti-commuting operators, we can evaluate the commutator in the Heisenberg equation to get:

$$\begin{aligned}\left[i\frac{\partial}{\partial t} \;-\; \hat{h}_0(\boldsymbol{r})\right] G(\boldsymbol{x}t,\boldsymbol{x'}\boldsymbol{t'}) \;&=\; \delta(t-t')\delta(\boldsymbol{x}-\boldsymbol{x}') \\ -\; i\int d\boldsymbol{x}'' v(\boldsymbol{x},\boldsymbol{x}'') \left\langle N\left|\hat{T}\left[\hat{\psi}^\dagger(\boldsymbol{x}'',t)\hat{\psi}(\boldsymbol{x}'',t)\hat{\psi}(\boldsymbol{x},t)\hat{\psi}^\dagger(\boldsymbol{x}',t')\right]\right|N\right\rangle,\end{aligned} \tag{6.36}$$

the integrated expectation value is the two-particle Green's function $G_2(\boldsymbol{x}t, \boldsymbol{x}''t, \boldsymbol{x}''t^+, \boldsymbol{x}'t')$. The equation of motion for G_2 would then introduce the three-particle Green's function. Thus, iterating the procedure would generate higher and higher orders of Green's functions to describe all possible many-body interactions. We encapsulate the sum of these interactions in the full, non-local self-energy using:

$$\begin{aligned} -i \int d\boldsymbol{x}'' v(\boldsymbol{x}, \boldsymbol{x}'') G_2(\boldsymbol{x}t, \boldsymbol{x}''t, \boldsymbol{x}''t^+, \boldsymbol{x}'t') \;\equiv \\ \int dt'' \int d\boldsymbol{x}'' \bar{\Sigma}(\boldsymbol{x}t, \boldsymbol{x}''t'') G(\boldsymbol{x}''t'', \boldsymbol{x}'t'), \end{aligned} \tag{6.37}$$

where $\bar{\Sigma}$ refers to the sum of the self-energy and the Hartree potential given by $\Sigma = \bar{\Sigma} - v_H$. Using the self-energy in equation (6.37) we arrive at the equation of motion for the Green's function:

$$\left[i\frac{\partial}{\partial t} \; - \; \hat{h}_0(\boldsymbol{r}) \; - \; v_H(\boldsymbol{r}) \right] G(\boldsymbol{x}t, \boldsymbol{x}'t') \; = \; \delta(t-t')\delta(\boldsymbol{r}-\boldsymbol{r}') + \int dt'' \int \boldsymbol{x}'' \Sigma(\boldsymbol{x}t, \boldsymbol{x}''t'') G(\boldsymbol{x}''t'', \boldsymbol{x}'t'). \tag{6.38}$$

If we take G_0 as the solution to part of the electron Hamiltonian, $\hat{h} = \hat{h}_0 + v_H$, then this equation simplifies further to:

$$G(1,2) \; = \; G_0(1,2) \; + \; \int d(3,4) G_0(1,3) \Sigma(3,4) G(4,2). \tag{6.39}$$

Which is the expression of equation (2.48) presented in Chapter 2, now with the streamlined expression for the variable dependences described in that chapter.

E2 Derivations of Hedin's Equations

In this section we focus on the actual derivation of Hedin's equations. We follow here a somewhat simplified derivation as presented in [62], but the interested reader can follow Hedin's actual expansion in [63]

We introduce a small perturbing field ϕ which will later be set to zero. We can use the identity put forward by Schwinger [139] to represent the two-particle Green's function as:

$$G_2(1,3,2,3^+) \;=\; G(1,2)G(3,3^+) \;-\; \frac{\delta G(1,2)}{\delta\phi(3)}. \tag{6.40}$$

We can substitute this expression into equation (6.37) and multiply both sides of the resulting equation by G^{-1} to get:

$$\begin{aligned}\bar{\Sigma}(1,2) \;&=\; \delta(1,2)\int d(3)v(1,3)G(3,3^+) \;-\; i\int d(3,4)v(1,3)G(1,4)\frac{\delta G^{-1}(4,2)}{\delta\phi(3)} \\ &=\; \delta(1,2)v_H(1) \;-\; i\int d(3,4)v(1,3)G(1,4)\frac{\delta G^{-1}(4,2)}{\delta\phi(3)}.\end{aligned} \tag{6.41}$$

We introduce now a set of quantities and identities which we combine to get Hedin's equations:

$$\begin{aligned}&\text{Total Potential: } V(1) \;=\; \phi(1) \;+\; v_H(1), \\ &\text{Inverse Dielectric Function: } \epsilon^{-1}(1,2) \;=\; \frac{\delta V(1)}{\delta\phi(2)}, \\ &\text{Screened Coulomb Interaction: } W(1,2) \;=\; \int d(3)\epsilon^{-1}(1,3)v(3,2), \\ &\text{3-point vertex: } \Gamma(1,2,3) \;=\; -\frac{\delta G^{-1}(1,2)}{\delta V(3)}, \\ &\text{Irreducible Polarizability: } P(1,2) \;=\; -i\frac{\delta G(1,1^+)}{\delta V(2)} \;=\; \frac{\delta n(1)}{\delta V(2)}, \\ &\text{Reducible Polarizability: } \chi(1,2) \;=\; -i\frac{\delta G(1,1^+)}{\delta\phi(2)} \;=\; \frac{\delta n(1)}{\delta\phi(2)}.\end{aligned} \tag{6.42}$$

Using these expressions along with the equation for G, we get Hedin's equations:

$$\begin{aligned}
P(1,2) &= -i\int d(3,4)G(4,2)G(2,3)\Gamma(3,4,1),\\
W(1,2) &= v(1,2) + \int d(3,4)v(1,2)P(3,4)W(4,2),\\
\Sigma(1,2) &= i\int d(3,4)G(1,4)W(1^{+},3)\Gamma(4,2,3),\\
\Gamma(1,2,3) &= \delta(1,2)\delta(1,3) + \int d(4,5,6,7)\frac{\delta\Sigma(1,2)}{\delta G(4,5)}G(4,6)G(7,5)\Gamma(6,7,3).
\end{aligned} \tag{6.43}$$

E3 Self-consistency in the GW approximation

Going beyond the "single-shot" GW method (discussed in Chapter 2) requires some form of self-consistency to be applied to the GW equations. It is to be noted that a straightforward expansion of the Green's function in powers of the screened Coulomb potential is not guaranteed to converge [61]. Additionally, the storage of all the quantities involved in the calculation can be very taxing depending on the system of study.

Despite these difficulties, different self-consistent schemes within the GW approximation have been devised. Conceptually, the purest of these schemes is fully self-consistent GW (scGW). Under this scheme, all four quantities in the GW approximation are iterated until self-consistency is achieved.

E4 Feynman Diagrams

An understanding of Feynman diagrams helps elucidate some details of Hedin's expansion. We cover some fundamentals here, but direct readers interested in a deeper understanding toward the book by Mattuck [140].

A Feynman diagram is a way to graphically represent the expansion of the Green's function into an infinite series. Each piece of a diagram corresponds to a piece of the mathematics that go into the expansion. The fully interacting Green's function is usually

represented as a double line with an arrow indicating the propagation in time. Time is generally understood to increase in the direction of the propagator. We remind the reader that the propagator technically satisfies both time orderings, although in one we consider the anti-particle propagating backward in time. In our diagrams here we consider time increasing vertically.

The starting and ending points of the propagator represent the starting and ending space-time points of the propagation. We note that the particle leaving the system need not be the same particle that entered. Along a path of propagation, the particle will interact with other particles in the medium. These interactions are the potentials (sometimes referred to as "self-energy" parts) in the definition of the Green's function (equation 6.39). An interaction is then said to take place at some "internal" time, which is any point in time between the starting and ending time. Arbitrarily many interactions may be taken to occur and the number of interactions is taken as the order of the diagram.

When equation (6.39) was introduced, we discussed that itself to the notion of self-consistency. We can iteratively substitute the Green's function back into itself to generate higher and higher orders of the expansion. We approach the expansion through Feynman diagrams in the same way, considering increasingly complicated interactions. The first step in this process is to build up an intuition about what sorts of interactions the particle may undergo.

As a simplified case, we consider an electron moving ballistically in a lattice of ions. The electron could be inserted into the system at time and place $\{t_1, \boldsymbol{r}_1\}$, propagate freely to its destination without scattering at all, and then leave the system at time and place $\{t_2, \boldsymbol{r}_2\}$. This would be the non-interacting propagator term G_0, displayed by a single, arrowed line. The electron may instead be inserted into the system at the same time and place $\{t_1, \boldsymbol{r}_1\}$, propagate freely for some time, then, at some *internal* time t' (with $t_1 < t' < t_2$), scatter off one of the ions in the lattice and then continue freely propagating until it leaves the system at time and place $\{t_2, \boldsymbol{r}_2\}$. This is represented by a propagator ending at time t',

some representation of the interaction, and then another propagator beginning at t' and continuing to t_2. In our simplified case here, the particle merely interacts with the scattering center, and no further structure need be introduced into the interaction. But we must build the diagrammatic expansion of the Green's function by considering *all* possible pathways and interactions. The electron could scatter off the first ion, into the second, and continue on. It could scatter off the first ion twice and continue on. It could scatter off the first ion twice, the second ion once, the third ion twice and so on, ostensibly including every combination of interactions possible. This would establish the full diagrammatic expansion of the propagator. Each diagram represents a mathematical term in the expansion as integrals of products of Green's functions and potentials, and the sum of all diagrams is then the fully-interacting propagator.

The above scenario makes it clear that the expansion is difficult to achieve even for relatively simple examples. We make this problem more tractable by making assumptions about which diagrams contribute the most to the sum. More complicated sequences of interactions are expected to have lower probabilities of occurring and thus contribute less to the propagator. We can, by organizing different families of diagrams, partially sum over these diagrams, assuming they form a sequence which has a finite sum.

One more important note about Feynman diagrams is that the processes represented in any given diagram do not necessarily carry the physical interpretation our introduction has ascribed to them. Processes in these diagrams may violate laws such as the Pauli exclusion principle or introduce interactions which occur in zero time, and thus present unphysical *virtual* processes. However, the sum of all of these processes is *required* to produce physical results. While this is provable, it is well beyond the scope of this appendix. We now introduce several diagrams of serious importance to the many-body electronic problem. We will describe the processes in these diagrams using physical intuition, but it should be understood that they are really only quasi-physical in nature.

E5 Bubble Diagrams

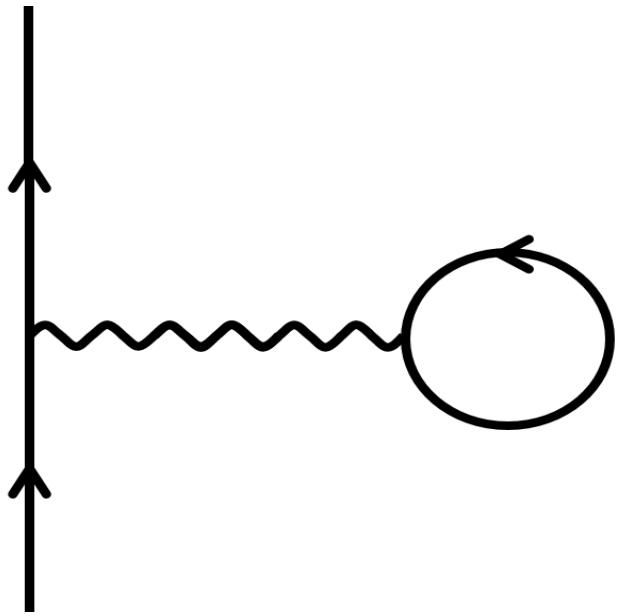

Figure E1: A Feynman diagram known as a bubble diagram.

The diagram depicted in Fig. (E1) is what is known as a bubble diagram. It represents a process wherein the propagating particle excites a particle initially in an occupied state, forming a particle-hole pair for an instant, but the propagating particle then interacts with the excited particle scattering it back into the hole. This process is referred to as "forward scattering", given that the initially propagating particle is scattered back into its initial state.

This diagram is an example of an unphysical process. The interaction takes place in zero time, that is, the particle which is excited and its hole are created and annihilated in the same instant. Assuming that only bubble diagrams contribute to the propagator yields a geometric series whose infinite sum can be achieved and is equivalent to taking the well-known Hartree approximation. This approximation sees the particle as interacting with an average potential produced by the particles in all occupied states.

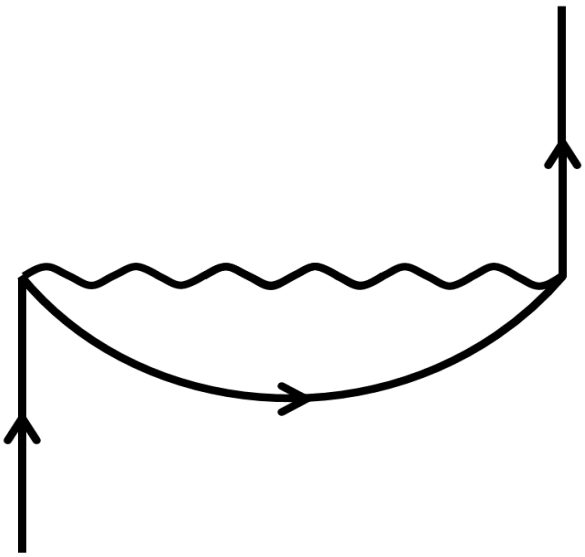

Figure E2: A Feynman diagram known as an open oyster diagram.

E6 Open Oyster Diagrams

The open oyster diagram, depicted in Fig. (E2), is a similar process to the bubble diagram. The propagating particle excites a particle in a state below the Fermi surface, creating a particle-hole pair. However, different from the bubble diagram, the propagating particle scatters into the state of the excited particle with the excited particle scattering into the state of the propagating particle. The original particle then recombines with the hole. This diagram is referred to as a first-order exchange diagram, the process sometimes referred to as "exchange scattering".

When forward scattering and exchange scattering processes are considered as the only contributions, the diagrams can be summed again, this time yielding the next higher rung on the ladder of approximations the Hartree-Fock approximation. The Hartree-Fock approximation involves the particle interacting with the same averaged external potential as in the standard Hartree approximation, however now the phenomenon of exchange is accounted for as well.

E7 Ring Diagrams

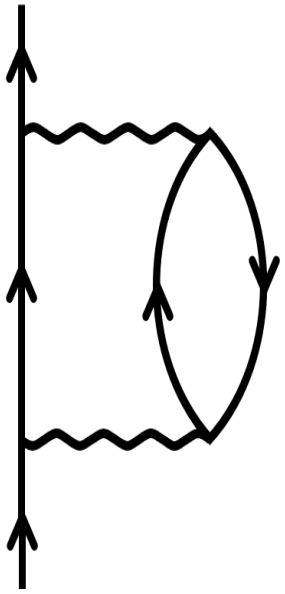

Figure E3: A Feynman diagram known as a ring diagram.

The last diagram we will discuss is shown in Fig. (E3), called a ring diagram. The ring diagram depicts a process wherein the propagating particle kicks a particle out of an occupied state, generating yet again a particle-hole pair. However, in a ring diagram, the particle-hole pair propagates as well until at some point the pair annihilates due to interaction with the initial particle. Higher orders of this type of diagram describe the particle-hole pair exciting other particle-hole pairs. This family of processes encapsulates the idea of the propagating particle exciting a cloud of particles as it moves through the medium. As explored in Chapter 2, this idea is central to the quasi-particle picture. The dynamically-adjusting polarization cloud that an extra particle generates is precisely of this nature.

Going one step further on the ladder of approximations for the many-electron system, we consider the the open-oyster and ring diagrams as the diagrams contributing most to the propagator. Summing these two families of diagrams yields the well-known Random Phase Approximation. This also gives us a very convenient means of thinking about the correlation in the system. We can consider the 'bare' interaction (or the non-interacting propagator)

as now being 'dressed' by the cloud of virtual interactions with the many particles of the medium. This feedback between the medium and the extra particle is exactly what is involved in the self-energy which is key to the GW approximation.

Bibliography

[1] T. Mikolajick, S. Slesazeck, H. Mulaosmanovic, M. H. Park, S. Fichtner, P. D. Lomenzo, M. Hoffmann, and U. Schroeder, "Next generation ferroelectric materials for semiconductor process integration and their applications", Journal of Applied Physics **129**, 100901 (2021).

[2] C. J. Bartel, C. Sutton, B. R. Goldsmith, R. H. Ouyang, C. B. Musgrave, L. M. Ghiringhelli, and M. Scheffler, "New tolerance factor to predict the stability of perovskite oxides and halides", Science Advances **5**, eaav0693 (2019).

[3] P. A. De and A. A. Barresi, "Deep Oxidation Kinetics of Trieline over LaFeO3 Perovskite Catalyst", Ind. Eng. Chem. Res. **40**, 1460–1464 (2001).

[4] J. Stahn, J. Chakhalian, C. Niedermayer, J. Hoppler, T. Gutberlet, J. Voigt, F. Treubel, H. U. Habermeier, G. Cristiani, B. Keimer, and C. Bernhard, "Magnetic proximity effect in perovskite superconductor/ferromagnet multilayers", Physical Review B **71**, 140509 (2005).

[5] M. V. Lobanov, A. M. Balagurov, V. J. Pomjakushin, P. Fischer, M. Gutmann, A. M. Abakumov, O. G. D'Yachenko, E. V. Antipov, O. I. Lebedev, and G. Van Tendeloo, "Structural and magnetic properties of the colossal magnetoresistance perovskite $La_{0.85}Ca_{0.15}MnO_3$", Physical Review B **61**, 8941–8949 (2000).

[6] D. F. Xue, S. X. Wu, Y. C. Zhu, K. Terabe, K. Kitamura, and J. Y. Wang, "Nanoscale domain switching at crystal surfaces of lithium niobate", Chemical Physics Letters **377**, 475–480 (2003).

[7] V. M. Goldschmidt, "Die gesetze der krystallochemie", Naturwissenschaften **14**, 477–485 (1926).

[8] H. Zhang, N. Li, K. Li, and D. F. Xue, "Structural stability and formability of ABO_3-type perovskite compounds", Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials **63**, 812–818 (2007).

[9] C. H. Li, X. G. Lu, W. Z. Ding, L. M. Feng, Y. H. Gao, and Z. G. Guo, "Formability of ABX_3 (X = F, Cl, Br, I) halide perovskites", Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials **64**, 702–707 (2008).

[10] W. Cochran, "Crystal stability and the theory of ferroelectricity", Advances in Physics **9**, 387–423 (1960).

[11] R. Comes, M. Lambert, and A. Guinier, "Chain structure of $BaTiO_3$ and $KNbO_3$", Solid State Communications **6**, 715–719 (1968).

[12] A. W. Hewat, "Cubic-tetragonal-orthorhombic-rhombohedral ferroelectric transitions in perovskite potassium niobate - neutron powder profile refinement of structures", Journal of Physics C-Solid State Physics **6**, 2559–2572 (1973).

[13] A. W. Hewat, "Soft modes and structure, spontaneous polarization and curie constants of perovskite ferroelectrics: tetragonal potassium niobate", Journal of Physics C-Solid State Physics **6**, 1074–1084 (1973).

[14] D. I. Bilc, R. Orlando, R. Shaltaf, G. M. Rignanese, J. Iniguez, and P. Ghosez, "Hybrid exchange-correlation functional for accurate prediction of the electronic and structural properties of ferroelectric oxides", Physical Review B **77**, 165107 (2008).

[15] E. Wiesendanger, "Domain-structures in orthorhombic $KNbO_3$ and characterization of single domain crystals", Czechoslovak Journal of Physics **B 23**, 91–99 (1973).

[16] T. Fukuda, H. Hirano, Y. Uematsu, and T. Ito, "Dielectric-constant of orthorhombic $KNbO_3$ single domain crystal", Japanese Journal of Applied Physics **13**, 1021–1022 (1974).

[17] Y. Uematsu, "Nonlinear optical-properties of $KNbO_3$ single-crystal in orthorhombic-phase", Japanese Journal of Applied Physics **13**, 1362–1368 (1974).

[18] M. Zgonik, R. Schlesser, I. Biaggio, E. Voit, J. Tscherry, and P. Gunter, "Materials constants of $KNbO_3$ relevant for electrooptics and acoustooptics", Journal of Applied Physics **74**, 1287–1297 (1993).

[19] B. Zysset, I. Biaggio, and P. Gunter, "Refractive-indexes of orthorhombic $KNbO_3$. I. dispersion and temperature-dependence", Journal of the Optical Society of America B-Optical Physics **9**, 380–386 (1992).

[20] I. Biaggio, P. Kerkoc, L. S. Wu, P. Gunter, and B. Zysset, "Refractive-indexes of orthorhombic $KNbO_3$. II phase-matching configurations for nonlinear-optical interactions", Journal of the Optical Society of America B-Optical Physics **9**, 507–517 (1992).

[21] L. Solymar, D. J. Webb, and A. Grunnet-Jepsen, *"The Physics and Applications of Photorefractive Materials"* (Oxford University Press, 1996).

[22] A. Ashkin, G. D. Boyd, J. M. Dziedzic, R. G. Smith, A. A. Ballman, J. J. Levinstein, and K. Nassau, "Optically-induced refractive index inhomogeneities in $LiNbO_3$ and $LiTaO_3$", Applied Physics Letters **9**, 72–74 (1966).

[23] F. S. Chen, "A laser-induced inhomogeneity of refractive indices in KTN", Journal of Applied Physics **38**, 3418– (1967).

[24] F. S. Chen, R. T. Denton, K. Nassau, and A. A. Ballman, "Optical memory planes using $LiNbO_3$ and $LiTaO_3$", Proceedings of the Institute of Electrical and Electronics Engineers **56**, 782–783 (1968).

[25] F. S. Chen, J. T. Lamacchia, and D. B. Fraser, "Holographic storage in lithium niobate", Applied Physics Letters **13**, 223–225 (1968).

[26] L. H. Lin, "Holographic measurements of optically induced refractive index inhomogeneities in bismuth titanate", Proceedings of the IEEE **57**, 252–253 (1969).

[27] J. B. Thaxter, "Electrical control of holographic storage in strontium-barium niobate", Applied Physics Letters **15**, 210–212 (1969).

[28] R. L. Townsend and J. Lamacchi, "Optically induced refractive index changes in $BaTiO_3$", Journal of Applied Physics **41**, 5188–5192 (1970).

[29] J. P. Huignard and F. Micheron, "High-sensitivity read-write volume holographic storage in $Bi_{12}SiO_{20}$ and $Bi_{12}GeO_{20}$ crystals", Applied Physics Letters **29**, 591–593 (1976).

[30] P. Gunter, U. Fluckiger, J. P. Huignard, and F. Micheron, "Optically induced refractive-index changes in $KNbO_3$:Fe", Ferroelectrics **13**, 297–299 (1976).

[31] A. M. Glass, A. M. Johnson, D. H. Olson, W. Simpson, and A. A. Ballman, "4-wave mixing in semi-insulating InP and GaAs using the photorefractive effect", Applied Physics Letters **44**, 948–950 (1984).

[32] M. B. Klein, "Beam coupling in undoped GaAs at 1.06-μm using the photorefractive effect", Optics Letters **9**, 350–352 (1984).

[33] J. J. Amodei, "Electron diffusion effects during hologram recording in crystals", Applied Physics Letters **18**, 22–24 (1971).

[34] J. J. Amodei, D. L. Staebler, and W. Phillips, "Improved electrooptic materials and fixing techniques for holographic recording", Applied Optics **11**, 390–396 (1972).

[35] A. M. Glass, D. V. D Linde, and T. J. Negran, "High-voltage bulk photovoltaic effect and the photorefractive process in $LiNbO_3$", Applied Physics Letters **25**, 233–235 (1974).

[36] N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electrooptic crystals. I. steady-state", Ferroelectrics **22**, 949–960 (1979).

[37] N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electrooptic crystals. II. Beam Coupling - Light Amplification", Ferroelectrics **22**, 961–964 (1979).

[38] M. Zgonik, C. Medrano, M. Ewart, H. Wuest, and P. Gunter, "$KNbO_3$ crystals for photorefractive applications", Optical Engineering **34**, 1930–1935 (1995).

[39] P. P. Banerjee, H. L. Yu, D. A. Gregory, and N. Kukhtarev, "Phase conjugation, edge detection and image broadcasting using two-beam coupling configurations in photorefractive $KNbO_3$", Optics and Laser Technology **28**, 89–92 (1996).

[40] G. Montemezzani, P. Rogin, M. Zgonik, and P. Gunter, "Interband photorefractive effects - theory and experiments in $KNbO_3$", Physical Review B **49**, 2484–2502 (1994).

[41] E. Wiesendanger and G. Guntherodt, "Optical anisotropy of $LiNbO_3$ and $KNbO_3$ in interband transition region", Solid State Communications **14**, 303–306 (1974).

[42] C. Medrano, E. Voit, P. Amrhein, and P. Gunter, "Optimization of the photorefractive properties of $KNbO_3$ crystals", Journal of Applied Physics **64**, 4668–4673 (1988).

[43] C. Medrano, M. Zgonik, I. Liakatas, and P. Gunter, "Infrared photorefractive effect in doped $KNbO_3$ crystals", Journal of the Optical Society of America B-Optical Physics **13**, 2657–2661 (1996).

[44] D. R. Evans, G. Cook, J. L. Carns, M. A. Saleh, S. A. Basun, J. M. Seim, and G. J. Mizell, "Major improvements of the photorefractive and photovoltaic properties in potassium niobate", Optics Letters **31**, 89–91 (2006).

[45] A. E. Clement and G. C. Gilbreath, "Photorefractive effects in tantalum-doped potassium niobate", Applied Optics **30**, 2458–2464 (1991).

[46] X. S. Gao, J. M. Xue, and J. Wang, "Ferroelectric behaviors and charge carriers in nd-doped $Bi_4Ti_3O_{12}$ thin films", Journal of Applied Physics **97**, 034101 (2005).

[47] M. M. Kumar and Z. G. Ye, "Dielectric and electric properties of donor- and acceptor-doped ferroelectric $SrBi_2Ta_2O_9$", Journal of Applied Physics **90**, 934–941 (2001).

[48] S. A. Basun and D. R. Evans, "Identification of the specific Fe centers and associated defect structure responsible for enhanced dynamic holography in photorefractive $KNbO_3$:Fe", Physical Review B **93** (2016).

[49] E. Voit, M. Z. Zha, P. Amrhein, and P. Gunter, "Reduced $KNbO_3$ crystals for fast photorefractive nonlinear optics", Applied Physics Letters **51**, 2079–2081 (1987).

[50] E. Erunal, R. A. Eichel, S. Korbel, C. Elsasser, J. Acker, H. Kungl, and M. J. Hoffmann, "Defect structure of copper doped potassium niobate ceramics", Functional Materials Letters **3**, 19–24 (2010).

[51] M. Ewart, R. Ryf, C. Medrano, H. Wuest, M. Zgonik, and P. Gunter, "High photorefractive sensitivity at 860 nm in reduced rhodium-doped $KNbO_3$", Optics Letters **22**, 781–783 (1997).

[52] C. Medrano, M. Zgonik, N. Sonderer, C. Beyeler, S. Krucker, J. Seglins, H. Wuest, and P. Gunter, "Photorefractive effect in Cu-doped and Ni-doped $KNbO_3$ in the visible-infrared and near-infrared", Journal of Applied Physics **76**, 5640–5645 (1994).

[53] J. Kohanoff, *"Electronic Structure Calculations for Solids and Molecules"* (Cambridge University Press, 2006).

[54] J. R. Max Born Oppenheimer, "Zur Quantentheorie der Molekeln [On the Quantum Theory of Molecules]", Annalen der Physik **389**, 457–484 (1927).

[55] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects", Physical Review **140**, A1133–A1138 (1965).

[56] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", Physical Review B **136**, B864–B871 (1964).

[57] W. Kohn, "Upsilon-representability and density functional theory", Physical Review Letters **51**, 1596–1598 (1983).

[58] W. Ritz, "Über eine neue methode zur lösung gewisser variationsprobleme der mathematischen physik", Journal für die Reine und Angewandte Mathematik **135**, 1–61 (1909).

[59] H. J. Monkhorst and J. D. Pack, "Special points for brillouin-zone integrations", Physical Review B **13**, 5188–5192 (1976).

[60] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "Quantum espresso: a modular and open-source software project for quantum simulations of materials", Journal of Physics-Condensed Matter **21**, 395502 (2009).

[61] F. Aryasetiawan and O. Gunnarsson, "The GW method", Reports on Progress in Physics **61**, 237–312 (1998).

[62] D. Golze, M. Dvorak, and P. Rinke, "The GW Compendium: A Practical Guide to Theoretical Photoemission Spectroscopy", Frontiers in Chemistry **7**, 377 (2019).

[63] L. Hedin, "New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem", Physical Review **139**, A796–A823 (1965).

[64] B. Holm and U. von Barth, "Fully self-consistent GW self-energy of the electron gas", Physical Review B **57**, 2108–2117 (1998).

[65] B. Holm, "Total energies from GW calculations", Physical Review Letters **83**, 788–791 (1999).

[66] P. Garcia-Gonzalez and R. W. Godby, "Self-consistent calculation of total energies of the electron gas using many-body perturbation theory", Physical Review B **63**, 075112 (2001).

[67] S. L. Adler, "Quantum theory of dielectric constant in real solids", Physical Review **126**, 413–420 (1962).

[68] N. Wiser, "Dielectric constant with local field effects included", Physical Review **129**, 62–69 (1963).

[69] M. J. van Setten, F. Weigend, and F. Evers, "The GW-Method for Quantum Chemistry Applications: Theory and Implementation", Journal of Chemical Theory and Computation **9**, 232–246 (2013).

[70] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, "BerkeleyGW: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures", Computer Physics Communications **183**, 1269–1289 (2012).

[71] D. R. Hamann, M. Schluter, and C. Chiang, "Norm-conserving pseudopotentials", Physical Review Letters **43**, 1494–1497 (1979).

[72] D. R. Hamann, "Optimized norm-conserving vanderbilt pseudopotentials", Physical Review B **88**, 085117 (2013).

[73] O. Beck, D. Kollewe, A. Kling, W. Heiland, and F. Hesse, "Lattice site location of iron in potassium niobate", Nuclear Instruments Methods in Physics Research Section B-Beam Interactions with Materials and Atoms **85**, 474–478 (1994).

[74] F. M. Michelcalendini, M. Peltier, and F. Micheron, "Electron-paramagnetic resonance of Fe^{3+} in orthorhombic $KNbO_3$", Solid State Communications **33**, 145–150 (1980).

[75] S. Brulisauer, D. Fluck, P. Gunter, L. Beckers, and C. Buchal, "Controlled reduction of Fe-doped $KNbO_3$ by proton-irradiation", Optics Communications **153**, 375–386 (1998).

[76] E. Siegel, W. Urban, K. A. Mueller, and E. Wiesendanger, "Electron-paramagnetic resonance of Fe3+ in $KNbO_3$ at 300 K", Physics Letters A **53**, 415–416 (1975).

[77] E. Siegel, "EPR-experiments on Fe-3+ in $KNbO_3$ at room-temperature", Ferroelectrics **13**, 385–387 (1976).

[78] P. D. Durugkar, "Dielectric dispersion in pure, samarium, samarium and tin doped potassium niobate single crystals", Ferroelectrics **540**, 54–64 (2019).

[79] R. D. Shannon, "Revised effective ionic-radii and systematic studies of interatomic distances in halides and chalcogenides", Acta Crystallographica Section A **32**, 751–767 (1976).

[80] C. K. Jørgensen, *"Atomic Spectra and Chemical Bonding in Complexes"* (Pergamon Press, 1962).

[81] J. C. Slater, "Exchange in spin-polarized energy bands", Physical Review **165**, 658–669 (1968).

[82] A. V. Postnikov, A. I. Poteryaev, and G. Borstel, "First-principles calculations for Fe impurities in $KNbO_3$", Ferroelectrics **206**, 69–78 (1998).

[83] J. S. McCullough, A. L. Harmon Bauer, C. A. Hunt, and J. J. Martin, "Photochromic response of bismuth germanium oxide doped with chromium", Journal of Applied Physics **90**, 6017–6021 (2001).

[84] A. Floris, I. Timrov, B. Himmetoglu, N. Marzari, S. de Gironcoli, and M. Cococcioni, "Hubbard-corrected density functional perturbation theory with ultrasoft pseudopotentials", Physical Review B **101**, 16 (2020).

[85] V. I. Anisimov, J. Zaanen, and O. K. Andersen, "Band theory and Mott insulators: Hubbard-U instead of Stoner-I", Physical Review B **44**, 943–954 (1991).

[86] M. Cococcioni and S. de Gironcoli, "Linear response approach to the calculation of the effective interaction parameters in the LDA+U method", Physical Review B **71**, 035105 (2005).

[87] H. J. Kulik, M. Cococcioni, D. A. Scherlis, and N. Marzari, "Density functional theory in transition-metal chemistry: A self-consistent Hubbard U approach", Physical Review Letters **97**, 103001 (2006).

[88] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened coulomb potential", Journal of Chemical Physics **118**, 8207–8215 (2003).

[89] E. Possenriede, O. F. Schirmer, H. J. Donnerberg, and B. Hellermann, "ESR investigation of transition-metal defects in $KNbO_3$", Journal of Physics-Condensed Matter **1**, 7267–7276 (1989).

[90] J.-P. Huignard and P. Günter, *"Photorefractive Materials and Their Applications 2"* (Springer, 2007).

[91] I. Biaggio, M. Zgonik, and P. Gunter, "Buildup and dark decay of transient photorefractive gratings in reduced $KNbO_3$", Optics Communications **77**, 312–317 (1990).

[92] B. Henderson and G. Imbusch, *"Optical Spectroscopy of Inorganic Solids"* (Oxford University Press, 1989).

[93] Z. Barandiaran, A. Meijerink, and L. Seijo, "Configuration coordinate energy level diagrams of intervalence and metal-to-metal charge transfer of dopant pairs in solids", Physical Chemistry Chemical Physics **17**, 19874–19884 (2015).

[94] C. Wegeberg and O. S. Wenger, "Luminescent first-row transition metal complexes", Czechoslovak Journal of Physics **1**, 1860–1876 (2021).

[95] D. Fluck, S. Brulisauer, P. Gunter, C. Buchal, and L. Beckers, "Improvement of the photorefractive response of Fe-doped $KNbO_3$ crystals by MeV proton irradiation", Nuclear Instruments Methods in Physics Research Section B-Beam Interactions with Materials and Atoms **148**, 678–682 (1999).

[96] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, "Phonons and related crystal properties from density-functional perturbation theory", Reviews of Modern Physics **73**, 515–562 (2001).

[97] J. W. Liu, G. Chen, Z. H. Li, and Z. G. Zhang, "Hydrothermal synthesis and photocatalytic properties of $ATaO_3$ and $ANbO_3$ (A = Na and K)", International Journal of Hydrogen Energy **32**, 2269–2272 (2007).

[98] T. T. Su, H. Jiang, H. Gong, and Y. C. Zhai, "An alternative approach of solid-state reaction to prepare nanocrystalline $KNbO_3$", Journal of Materials Science **45**, 3778–3783 (2010).

[99] T. T. Zhang, K. Zhao, J. G. Yu, J. Jin, Y. Qi, H. Q. Li, X. J. Hou, and G. Liu, "Photocatalytic water splitting for hydrogen generation on cubic, orthorhombic, and tetragonal $KNbO_3$ microcubes", Nanoscale **5**, 8375–8383 (2013).

[100] Y. Shiozaki, E. Nakamura, T. Mitsui, and Landolt-Börnstein, *"New Series III/36AI"* (Pergamon Press, 2001).

[101] T. T. Zhang, W. Y. Lei, P. Liu, J. A. Rodriguez, J. G. Yu, Y. Qi, G. Liu, and M. H. Liu, "Insights into the structure-photoreactivity relationships in well-defined perovskite ferroelectric $KNbO_3$ nanowires", Chemical Science **6**, 4118–4123 (2015).

[102] P. G. Kang, B. K. Yun, S. Shin, J. H. Ko, Y. S. Lee, and J. H. Jung, "Possible origin of stabilized monoclinic structure of $KNbO_3$ nanomaterials at room temperature", Materials Science and Engineering B - Advanced Functional Solid-State Materials **210**, 19–23 (2016).

[103] F. Schmidt, A. Riefer, W. G. Schmidt, A. Schindlmayr, M. Imlau, F. Dobener, N. Mengel, S. Chatterjee, and S. Sanna, "Quasiparticle and excitonic effects in the optical response of $KNbO_3$", Physical Review Materials **3**, 054401 (2019).

[104] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, "First-principles calculations of the electronic structure and spectra of strongly correlated systems: The LDA+U method", Journal of Physics-Condensed Matter **9**, 767–808 (1997).

[105] I. Timrov, N. Marzari, and M. Cococcioni, "Self-consistent hubbard parameters from density-functional perturbation theory in the ultrasoft and projector-augmented wave formulations", Physical Review B **103**, 045141 (2021).

[106] E. Runge and E. K. U. Gross, "Density-functional theory for time-dependent systems", Physical Review Letters **52**, 997–1000 (1984).

[107] I. Timrov, N. Vast, R. Gebauer, and S. Baroni, "turboEELS-A code for the simulation of the electron energy loss and inelastic X-ray scattering spectra using the Liouville-Lanczos approach to time-dependent density-functional perturbation theory", Computer Physics Communications **196**, 460–469 (2015).

[108] P. Rinke, A. Janotti, M. Scheffler, and C. G. Van de Walle, "Defect Formation Energies without the Band-Gap Problem: Combining Density-Functional Theory and the GW Approach for the Silicon Self-Interstitial", Physical Review Letters **102**, 026402 (2009).

[109] S. Lany and A. Zunger, "Many-body GW calculation of the oxygen vacancy in ZnO", Physical Review B **81**, 113201 (2010).

[110] M. Jain, J. R. Chelikowsky, and S. G. Louie, "Quasiparticle Excitations and Charge Transition Levels of Oxygen Vacancies in Hafnia", Physical Review Letters **107**, 216803 (2011).

[111] A. Malashevich, M. Jain, and S. G. Louie, "First-principles DFT plus GW study of oxygen vacancies in rutile TiO_2", Physical Review B **89**, 075205 (2014).

[112] J. Furthmuller, G. Cappellini, H. C. Weissker, and F. Bechstedt, "GW self-energy calculations for systems with huge supercells", Physical Review B **66**, 045110 (2002).

[113] F. Bruneval and X. Gonze, "Accurate GW self-energies in a plane-wave basis using only a few empty states: Towards large systems", Physical Review B **78**, 085125 (2008).

[114] P. Umari, G. Stenuit, and S. Baroni, "Optimal representation of the polarization propagator for large-scale GW calculations", Physical Review B **79**, 201104 (2009).

[115] F. Giustino, M. L. Cohen, and S. G. Louie, "GW method with the self-consistent Sternheimer equation", Physical Review B **81**, 115105 (2010).

[116] W. W. Gao, W. Y. Xia, X. Gao, and P. H. Zhang, "Speeding up GW Calculations to Meet the Challenge of Large Scale Quasiparticle Predictions", Scientific Reports **6**, 36849 (2016).

[117] Y. Matsushita, K. Nakamura, and A. Oshiyama, "Comparative study of hybrid functionals applied to structural and electronic properties of semiconductors and insulators", Physical Review B **84**, 075205 (2011).

[118] H. P. Komsa and A. Pasquarello, "Assessing the accuracy of hybrid functionals in the determination of defect levels: Application to the As antisite in GaAs", Physical Review B **84**, 075207 (2011).

[119] P. Deak, M. Lorke, B. Aradi, and T. Frauenheim, "Optimized hybrid functionals for defect calculations in semiconductors", Journal of Applied Physics **126**, 130901 (2019).

[120] W. Setyawan and S. Curtarolo, "High-throughput electronic band structure calculations: Challenges and tools", Computational Materials Science **49**, 299–312 (2010).

[121] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, "A high-throughput infrastructure for density functional theory calculations", Computational Materials Science **50**, 2295–2310 (2011).

[122] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, "AiiDA: automated interactive infrastructure and database for computational science", Computational Materials Science **111**, 218–230 (2016).

[123] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, "Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)", JOM **65**, 1501–1509 (2013).

[124] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Ruhl, and C. Wolverton, "The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies", NPJ Computational Materials **1**, 15010 (2015).

[125] G. Hautier, A. Jain, S. P. Ong, B. Kang, C. Moore, R. Doe, and G. Ceder, "Phosphates as Lithium-Ion Battery Cathodes: An Evaluation Based on High-Throughput ab Initio Calculations", Chemistry of Materials **23**, 3495–3508 (2011).

[126] Z. Z. Li, Q. C. Xu, Q. D. Sun, Z. F. Hou, and W. J. Yin, "Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning", Advanced Functional Materials **29**, 1807280 (2019).

[127] K. F. Garrity, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, "Pseudopotentials for high-throughput DFT calculations", Computational Materials Science **81**, 446–452 (2014).

[128] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, and A. Fazzio, "From DFT to machine learning: recent approaches to materials science: A review", Journal of Physics-Materials **2**, 032001 (2019).

[129] D. Ceperley and B. Alder, "Quantum Monte-Carlo", Science **231**, 555–560 (1986).

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

[131] L. Goerigk and S. Grimme, "A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions", Physical Chemistry Chemical Physics **13**, 6670–6688 (2011).

[132] Y. B. Zhang, J. W. Sun, J. P. Perdew, and X. F. Wu, "Comparative first-principles studies of prototypical ferroelectric materials by LDA, GGA, and SCAN meta-GGA", Physical Review B **96**, 035143 (2017).

[133] J. W. Sun, A. Ruzsinszky, and J. P. Perdew, "Strongly Constrained and Appropriately Normed Semilocal Density Functional", Physical Review Letters **115**, 036402 (2015).

[134] J. W. Sun, R. C. Remsing, Y. B. Zhang, Z. R. Sun, A. Ruzsinszky, H. W. Peng, Z. H. Yang, A. Paul, U. Waghmare, X. F. Wu, M. L. Klein, and J. P. Perdew, "Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional", Nature Chemistry **8**, 831–836 (2016).

[135] R. Car and M. Parrinello, "Unified approach for molecular-dynamics and density-functional theory", Physical Review Letters **55**, 2471–2474 (1985).

[136] G. E. Engel and B. Farid, "Generalized plasmon-pole model and plasmon band structures of crystals", Physical Review B **47**, 15931–15934 (1993).

[137] M. S. Hybertsen and S. G. Louie, "Electron correlation in semiconductors and insulators: Band-gaps and quasi-particle energies", Physical Review B **34**, 5390–5413 (1986).

[138] M. Rohlfing and S. G. Louie, "Electron-hole excitations and optical spectra from first principles", Physical Review B **62**, 4927–4944 (2000).

[139] J. Schwinger, "On the Green's functions of quantized fields. I", Proceedings of the National Academy of Sciences of the United States of America **37**, 452–455 (1951).

[140] R. D. Mattuck, *"A Guide to Feynman Diagrams in the Many-body Problem"* (McGraw-Hill, 1976).