Theoretical and Computational Studies Related to Solar Water Splitting with Semiconductor Alloys

A Dissertation Presented

by

Li Li

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Physics

Stony Brook University

May 2011

Stony Brook University

The Graduate School

Li Li

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Philip B. Allen – Dissertation Advisor
Professor, Department of Physics and Astronomy

Maria V. Fernandez-Serra – Chairperson of Defense
Professor, Department of Physics and Astronomy

Robert E. Shrock
Professor, Department of Physics and Astronomy

Matthew Dawber
Professor, Department of Physics and Astronomy

Mark S. Hybertsen
Group Leader (Theory and Computation)
Center for Functional Nanomaterials, Brookhaven National Laboratory

This dissertation is accepted by the Graduate School.

Lawrence B. Martin
Dean of the Graduate School

Abstract of the Dissertation

Theoretical and Computational Studies Related to Solar Water Splitting with Semiconductor Alloys

by

Li Li

Doctor of Philosophy

in

Physics

Stony Brook University

2011

The $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution (or, alloy) is a visible-light-driven photo-catalyst for water splitting. Its reduced band gap is a main advantage for harvesting solar energy. Because the synthesized samples are in the powder form, the understanding of the bulk structures and the surfaces are hindered. In this thesis, we address both the bulk and the surfaces of this material through simulations based on the density-functional theory (DFT) version of quantum electronic structure theory.

The ordering of the atoms in the alloy is the key information to understand the bulk properties, especially the band gap reduction mechanism. Using the cluster expansion formalism, we construct an accurate model from DFT calculations. The subsequent Monte Carlo simulation reveals a phase diagram which has a wide miscibility gap and an $x=0.5$ ordered compound. The disordered phase displays strong short-range order (SRO) at synthesis temperatures.

To study the influences of SRO on the lattice and electronic properties, we conduct DFT calculations on snapshots from the Monte Carlo simulation. Consistent with previous theoretical and experimental findings, lattice parameters were found to deviate from Vegard's law with small upward bowing. Bond lengths depend strongly on local environment, with a variation much larger than the difference of bond length between ZnO and GaN. The downward band gap bowing deviates from parabolic by having a more rapid onset of bowing at low and high concentrations. An overall bowing parameter of 3.3 eV is predicted from a quadratic fit to the compositional dependence of the calculated band gap. Our results indicate that SRO has significant influence over both structural and electronic properties.

Recent experiments showed that the semi-polar $(10\bar{1}1)/(10\bar{1}\bar{1})$ surfaces dominate the powder samples. To search for stable reconstructions of these two surfaces, we use an evolutionary algorithm to explore the surface structures. To simplify the study, we only consider the pure GaN bulk with various numbers of Ga, N, and O atoms allowed to bond to surfaces. A few stable reconstructions at different Ga, N, and O chemical potentials are found. The consequences for the water splitting catalysis are discussed.

In this thesis, I also include a chapter on electron transfer during a non-adiabatic process. The relevance to the water splitting project is that a photo-excited hole must transfer across the semiconductor/water interface to initiate the oxidation of water. Similarly, a photo-excited electron must transfer to the H^+ in the liquid to cause hydrogen reduction (H_2 formation). The transfer process is ignored in the next of the thesis. Since the Born-Oppenheimer approximation does not apply, it is a challenging problem for numerical simulations. A few approximate methods have been proposed, which greatly reduce the calculation complexity, but still take the non-adiabaticity into account. To test these methods, we study a simple model, in which the nuclei can be treated quantum mechanically. Numerically exact solutions are obtained and compared with these popular approximations. We find that these methods do produce correct trends in general. But caution must be taken since they break down in some scenarios.

Contents

List of Figures

ix

xi

List of Tables

Acknowledgements

Chapter 1

Background

1.1 Photo-catalytic Water Splitting

A way of storing renewable energy sources is to use sun light to split water
into oxygen and hydrogen. In this process, the solar energy is converted into
chemical energy directly, as shown in the following reaction

$$2H_2O \; + \; 4h\nu \; \rightarrow \; 2H_2 \; + \; O_2 \qquad (1.1)$$

In this reaction, the absorbed photons first create electron-hole pairs in the
dye or absorber (in our case, semiconductor),

$$4h\nu \; \rightarrow \; 4h^+ \; + \; 4e^-. \qquad (1.2)$$
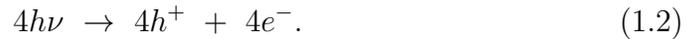
The electrons and holes then drive the two half reactions,

$$4H^+ \; + \; 4e^- \; \rightarrow \; 2H_2, \qquad (1.3)$$
$$2H_2O \; + \; 4h^+ \; \rightarrow \; O_2 \; + \; 4H^+. \qquad (1.4)$$

Each half reaction requires a catalyst and a minimum energy, as shown in
fig. 1.1. To produce hydrogen, the conduction band minimum must lie higher
than the standard reduction potential of the proton reduction reaction. To
produce oxygen, the valence band maximum must lie lower than the standard
reduction potential of the water oxidation reaction. However, not all materials
that satisfy these two prerequisites are good photo-catalysts for water splitting.
Another constraint is the energy of solar photons. At sea level, most of the
solar energy lies in the visible range, from 1.65 eV to 3.1 eV. The desired
material must have a band gap in the visible range to utilize the solar energy
efficiently. There are also other prerequisites for a material to be a good photo-
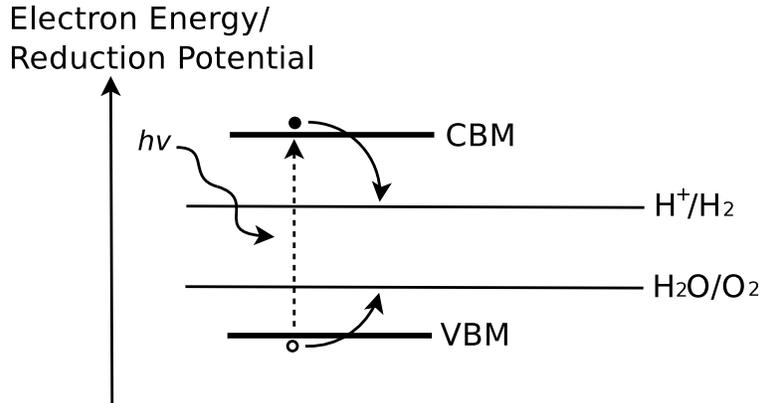
Figure 1.1: The schematic diagram for the photon absorption and the water splitting process. CBM, conduction band minimum; VBM, valence band maximum; H^+/H_2, hydrogen reduction potential; H_2O/O_2, water oxidation potential.

catalyst for water splitting. For example, the material must be stable during the reaction. The electron-hole pair must be able to separate and migrate to the reaction sites before they recombine.

1.2 The $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ Solid Solution

The $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution was synthesized to produce a reduced band gap photo-catalyst [3, 4]. The powder sample by Domen's group, when loaded with rhodium-chromium mixed oxide nano particles as a co-catalyst for hydrogen production, splits water into hydrogen and oxygen under visible light irradiation. The quantum efficiency, the percentage of irradiated photons that participate in the reaction, is 2.5% for wavelength in the range of 420∼440 nm. The reaction is stable. The solid solution functions continuously for 35 hours without degradation during the experiment.

This solid solution is a good candidate because of its reduced band gap [3, 4], which is crucial for improving solar photon absorption efficiency. Diffuse reflectance spectroscopy measurements indicate a band gap bowing parameter of 3∼4 eV [1, 2]. The band gap of the 50% alloy is around 2.4 eV [1], much more efficient for solar applications than either GaN and ZnO, with E_g equal to 3.4 and 3.2 eV, respectively.

There are theoretical discussions on the origin of the band gap reduction [5–9]. It is accepted that the alloy adopts the wurtzite lattice, but the atom site

occupacy is unknown. Lacking this information, previous studies used different approaches including the dilute limits [5, 6], long-periodic superlattices [6], special quasi-random structures [7] and/or structures based on intuition and enumeration [8]. Wang *et al* designed a rather simple Hamiltonian for this system [9]. They used snapshots from the Monte Carlo simulation based on this Hamiltonian to suggest the solid solution has no miscibility gap at 1100K but exhibit a strong short-range ordering (SRO). The SRO causes the highest occupied molecular orbitals (HOMO) and lowest unoccupied molecular orbitals (LUMO) to locate at different regions of the material. The spatial separation of HOMO and LUMO reduces the overlap between their wave functions, causing reduced efficiency for light absorption, and also, perhaps, reducing the undesired radiative recombinations.

Lacking direct measurements on the surface structures, the alloy surfaces have not been extensively studied by theorists. The X-ray photoelectron spectroscopy measurements indicate that the surface region is dominated by Ga, O, and N atoms, with virtually no Zn atoms, very different from the bulk concentrations [2]. Recent unpublished transmission electron microscopy measurements reveal that the dominant surfaces of the powder samples are $(10\bar{1}1)/(10\overline{11})$ surfaces. (Communication from Dr. James Ciston at BNL) Until now, there is no systematic theoretical studies on these two surfaces of the alloy.

1.3 Organization of this Thesis

Given the current studies on this material, a few important questions are left unresolved. This thesis tries to answer the following questions. (1) what is the atom site occupancy in the bulk of this alloy? (2) What is the properties of this alloy and their dependence on the atom site occupancy? (3) Given the recent experiments on the alloy, what might the surface look like?

The thesis is organized as follows. Chapter 2 introduces the density functional theory, and the methods used in this study. Chapter 3 presents our extensive studies on the bulk of this alloy. Chapter 4 presents our preliminary studies on the surfaces. Chapter 5 is a more distantly related project about electron transfers during non-adiabatic processes.

Chapter 2

Methods

2.1 Density Functional Theory

The systems we are interested in contain many nuclei and electrons. Nuclear motions can be treated classically (except for the lightest elements like H). Although the nuclei and the electrons interact with each other, we can *usually* separate their motion, because of the huge differences between their masses. Therefore, the nuclei can be treated as fixed when we are solving the electrons. This is the Born-Oppenheimer approximation [10].

Under the Born-Oppenheimer approximation, the electrons see an "external" potential $V_{\text{ext}}(r)$ from the nuclei. The Hohenberg-Kohn theorems [11] state that the ground state electron density $n(r)$ is uniquely determined by the external potential $V_{\text{ext}}(r)$. And the energy of the system $E[n]$ is a functional of the density $n(r)$. The Hohenberg-Kohn theorems are exact, but we do not know the form of the functional $E[n]$. The Kohn-Sham ansatz [12] converts the many-body system of the interacting electrons, $\Psi(r_1, r_2, ...)$, into a set of non-interacting one-body wave functions, $\{\psi_i(r)\}$. The density of the electrons is now the sum of the fictitious non-interacting electrons,

$$n(r) = \sum f(\epsilon_i - \epsilon_f)|\psi_i(r)|^2. \tag{2.1}$$

The occupancy function $f(\epsilon_i - \epsilon_f)$ can be derived from a certain thermal distribution, where ϵ_f is the fermi energy. And the energy of the system

becomes

$$E_{\text{tot}} = \sum_i f(\epsilon_i - \epsilon_f) \int d^3r \; \psi_i^*(r) \left(T_i + V_{\text{ext}}\right) \psi_i(r)$$
$$+ \frac{1}{2} \int d^3r d^3r' \frac{n(r)n(r')}{|r - r'|} + E_{\text{xc}}[n(r)]$$
$$+ T_{\text{nuclei}} + V_{\text{nuclei-nuclei}}. \tag{2.2}$$

The first term is the kinetic energy of the non-interacting electrons and their potential energy due to the external field. The second term is the Hartree term. The third term is the exchange-correlation energy. We also included the fourth and the fifth term, the kinetic energy and the Coulomb energy of the nuclei. Each wave function $\psi_i(r)$ and eigen energy ϵ_i is the solution to the Schrödinger equation

$$\left(-\frac{\nabla_i^2}{2} + V_{\text{ext}}(r) + \int d^3r' \frac{n(r')}{|r - r'|} + V_{\text{xc}}(r; n)\right) \psi_i(r) = \epsilon_i \psi_i(r). \tag{2.3}$$

Here the exchange-correlation potential is defined as $V_{\text{xc}}(r) = \frac{\delta E_{\text{ex}}[n]}{\delta n(r)}$. However, the exact form of the exchange-correlation functional, $E_{\text{xc}}[n]$, is unknown. Approximate exchange-correlation energy densities have been proposed. Under the Local Density Approximation (LDA), the exchange-correlation energy density only depends on the local electron density. The representative one is the Perdew-Zunger [13] (PZ) form. Later, more sophisticated functionals have been proposed, which also depend on the gradient of the local electron density (the Generalized Gradient Approximations, or GGA). The popular ones include Perdew-Burke-Ernzerhof [14] (PBE) and Becke-Lee-Yang-Parr [15–17] (BLYP).

A very important idea for efficient implementation of DFT is the pseudo-potential. In practice, electrons are usually separated into core and valence electrons. The core electrons do not participate in chemical bonding. There is no need to treat them explicitly. Therefore, the ion, a collection of the nucleus and its core electrons, is treated as a single entity. Furthermore, the ion potential can be simplified into a pseudo-potential according to the norm-conserving condition [18]. The pseudo-potential $V_i(r)$ from ion core i, is usually angular-momentum dependent, written as

$$V_i(r) = \sum_l U_{i,l}(r)\hat{P}_l, \tag{2.4}$$

where $\hat{P}_l$ is the projector operator for angular momentum l. The pseudo-

potential reproduces the true potential outside the core region, but is much smoother inside the core. The oscillations of the electron wave function inside the core region are eliminated by the pseudo-potential. This is a great advantage for numerical calculations. A further approximation, the ultra-soft pseudo-potential [19], uses more than one projector for each momentum, and further smooths the electron wave function. The projector-augmented wave (PAW) used in this study is a more general form of the ultra-soft pseudo-potential [20, 21].

2.2 Methods in this Study

Our first principles calculations use the VASP package [22]. We choose the Perdew-Burke-Ernzerhof [14] (PBE) implementation for the exchange-correlation functional and the projector-augmented wave [20, 21] (PAW) basis set for the expansion of wave functions. The plane wave cutoff is 500 eV. An $8 \times 8 \times 6$ k-point mesh is used for the wurtzite GaN and ZnO primitive cell. For supercells, the k-point meshes are adjusted to have as similar density to the primitive cell k-mesh as possible. All self-consistent calculations are converged to 0.1 meV. For structural relaxation, a conjugate gradient algorithm reduces the force on each atom to less than 0.05 eV/Å. Gallium and zinc $3d$ electrons are treated explicitly as valence electrons.

It is well known that DFT tends to underestimate the band gaps of semiconductors. This band gap problem is due to the incomplete cancellation of the self-interaction energy in the Hartree term and the exchange-correlation term in Eq. 2.2. The incomplete cancellation is more pronounced for systems with very localized orbitals, e.g., gallium and zinc $3d$ orbitals. To partially prevent the alloy from incorrectly becoming metallic in the DFT calculation, we apply an on-site Coulomb interaction [23, 24] U to the $3d$ orbitals of gallium and zinc. The values of U (from Ref. [7]) are 3.9 and 6.0 eV, respectively. These values were shown [7] to be the best to reproduce lattice parameters and band gaps. After the correction, the $3d$ band positions and the band gaps of GaN and ZnO (2.4 and 1.6 eV, respectively) lie closer to the experimental values (3.4 and 3.2 eV, respectively).

Chapter 3

$(\text{Ga}_{1-x}\text{Zn}_x)(\text{N}_{1-x}\text{O}_x)$ Solid Solution: Bulk

In this chapter, we first introduce our model for the bulk alloy. Our model assumes the alloy is on a wurtzite lattice. With no information of atom site occupancies on the lattice, we start our investigation with the enumeration of possible $2 \times 2 \times 1$ supercells and a few long-periodic superlattices. These results show that the short-range order plays a key role in determining the properties of the alloy. Motivated by this finding, we employ the cluster expansion formalism to build an accurate model. We use this model to predict the formation energy for any atom site occupancy. Subsequent Monte Carlo simulations and DFT calculations confirm the finding. Based on the simulation, we predict the quantitative dependence of the alloy properties on synthesis temperature.

3.1   Model

We model the solid solution as a wurtzite lattice with equal composition of Ga and N, and no atom exchange between cation and anion sublattices, similar to the approach adopted in previous work [7]. Thus the formula is $(\text{Ga}_{1-x}\text{Zn}_x)(\text{N}_{1-x}\text{O}_x)$. These assumptions are consistent with experimental results [1, 25, 26]. In first-principles calculations, we assume the atoms reside on this lattice with bond lengths and bond angles allowed to relax. Recent diffraction data for a sample near $x = 1/8$ was best fitted with a split-site anion model indicative of significant deviations from a uniform wurtzite structure [27]. In this study, we restrict the lattice sites for the anions to those from the wurtzite structure, although this assumption may need to be reconsidered when more complete experimental information becomes available. Point defects such as vacancies, interstitials, and cation/anion substitutions

are also beyond the scope of this study. Our goal is to understand the atom site occupancy of the crystalline alloy at thermal equilibrium as a function of temperature, and its influence on the lattice parameters, bonds, and band gaps.

The most important quantities in this study are the formation energy and the band gap bowing. The formation energy E for a specific alloy structure is defined as

$$E = E_{\text{alloy}} - xE_{\text{ZnO}} - (1 - x)E_{\text{GaN}}, \tag{3.1}$$

where x is the ZnO concentration. It shows the energy required to form the alloy from GaN and ZnO. The band gap E_g for a specific alloy structure, usually deviates from the linear interpolation of the two end points (GaN and ZnO). The band gap bowing parameter b describes this deviation in parabolic approximation. It is defined as

$$E_g(\sigma) = (1 - x)E_g(\text{GaN}) + xE_g(\text{ZnO}) - bx(1 - x). \tag{3.2}$$

A positive bowing parameter corresponds to a band gap bowing downward. We will introduce other quantities when they appear.

3.2 Preliminary Studies

3.2.1 $2 \times 2 \times 1$ Supercell

Formation energy, bandgap and volume

A complete enumeration of structures within a 2x2x1 supercell provides valuable estimate of the relationships between alloy properties and the atom site occupancies. There are 376 symmetrically inequivalent structures in a $2 \times 2 \times 1$ supercell, including pure GaN and ZnO. The calculated formation energies are shown in fig. 3.1. Almost all the formation energies are positive, implying that mixing must be driven by entropy at finite temperature. The average energy at each concentration follows a parabolic relationship. A truly random alloy should approximately follow this curve. The formation energy for a completely random alloy with 50% ZnO is around 0.1 eV. Therefore we are expecting the atom site occupancies to significantly deviate from random at synthesis temperatures around 1000K. In fig. 3.1, there is one structure with negative formation energy, suggesting that there might exist a stable compound in the phase diagram. We will visit this later in our Monte Carlo simulations.

Bandgap bowings are shown in fig. 3.2. They are positively correlated with the formation energy. If the formation energy for the 50% alloy is 0.1 eV/atom,
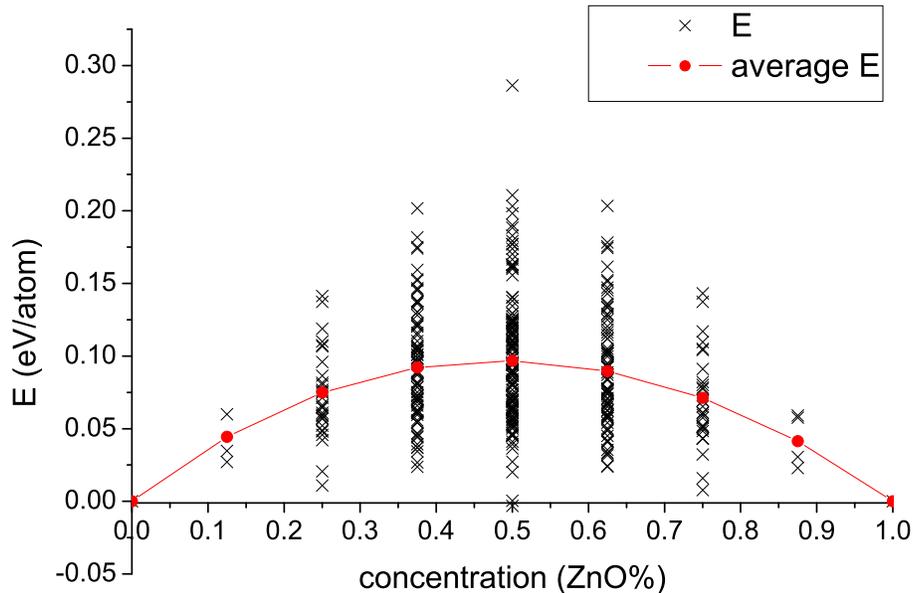
Figure 3.1: Formation energies *versus* concentration for structures within the $2 \times 2 \times 1$ supercell. Also shown is the average formation energy at each concentration.

the bowing parameter is predicted to be 6 eV from the figure, much higher than the experimental values (around 3∼4 eV). Therefore, the synthesized samples are more likely to be low energy structures, different from the random structures.

Volumes of the relaxed structures are shown in fig. 3.3. They depend linearly on ZnO concentration and formation energy. High energy structures tend to have larger volume. Structural relaxation, especially increasing volume, can change the formation energy significantly. We calculated formation energy of constrained structures for comparison. In these constrained structures, all atoms are located at the ideal coordinates of the wurtzite lattice. All the lattice parameters are linear interpolated between GaN and ZnO, according to the ZnO concentration. The relaxation reduces the formation energy by 30% (see fig 3.4). Therefore, in our following studies, we always let the atoms and the lattice parameters to be fully relaxed.

Charge, bond length and electrostatic potential

We also calculate atomic properties within this supercell, including charge, bond length and average electrostatic potential of the ion core. These quanti-

9
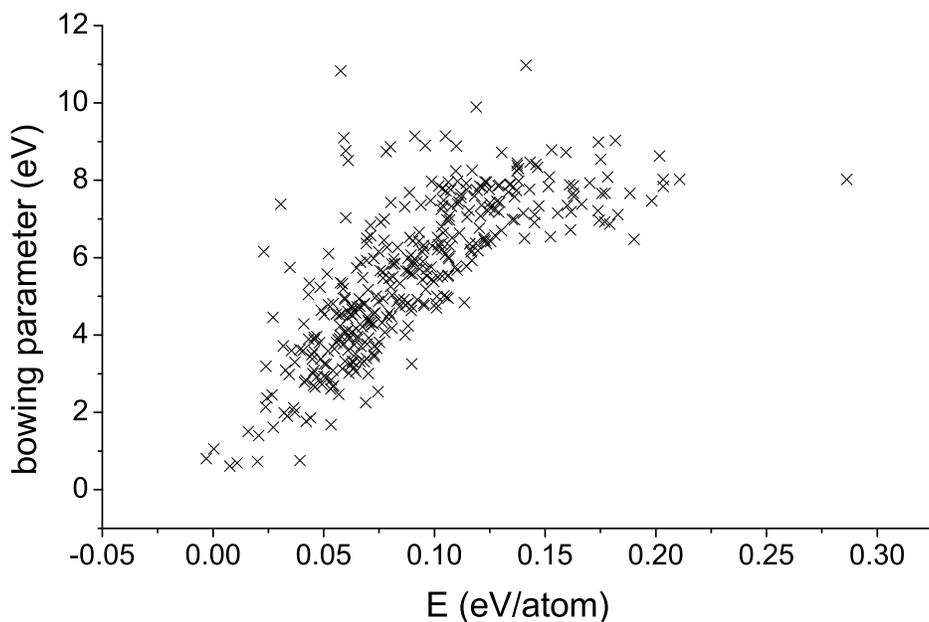
Figure 3.2: Bowing parameters *versus* formation energies within $2 \times 2 \times 1$ supercell
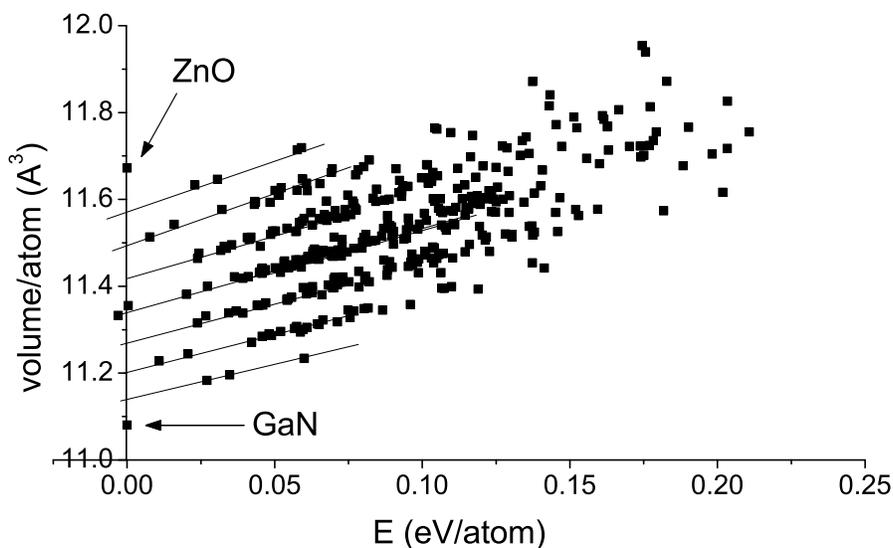
Figure 3.3: Volumes of relaxed structures *versus* formation energy within $2 \times 2 \times 1$ supercell (2 structures which relax into a non-wurtzite lattice are ignored). The data points are clearly grouped by concentrations. Solid lines are drawn for each concentration to guide the eye.
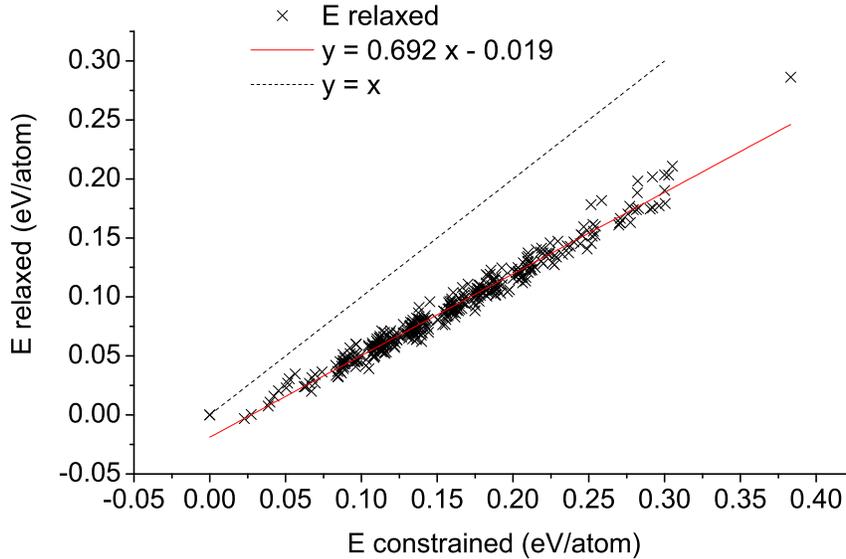
Figure 3.4: Formation energy of relaxed structures *versus* that of constrained structures. The linear relationship is calculated by least square fitting.

ties show very strong dependence on the atoms' first and second neighbors.

The Bader charge is shown in fig. 3.5. Oxygen and nitrogen charges are inert with respect to their environment. Both gallium and zinc loose more electrons when there are more oxygen neighbors. By comparing relaxed structures and unrelaxed structures, we find that the structural relaxation has very little effect on the magnitude of the Bader charge on each ion. Therefore the charge accumulation on each ion is decided primarily by species of, not distance from, the neighboring atoms. Since anion charge tends to be constant, the change in cation charge must be accomplished by transferring electrons within the cation sublattice. Compared to the cations in the GaN and ZnO crystals, the Zn atom in the alloy has more electrons. Ga atom has less electrons. Therefore, the Zn atom effectively drains electrons from nearby cations, while the Ga atom donates electrons.

The bond length changes according to the charges of the constituent ions. Atoms with more electrons tend to occupy more space, leading to longer bond length. Since the charges on both constituent ions depend on their own neighbors, the bond length therefore also depend on the neighbors. For example, for the Zn-O bond, the Zn atom has 3 extra anion neighbors besides the O atom. The O atom has 3 extra cation neighbors besides the Zn atom. All types of bonds tend to decrease when the bond has more O or Zn neighbors

11
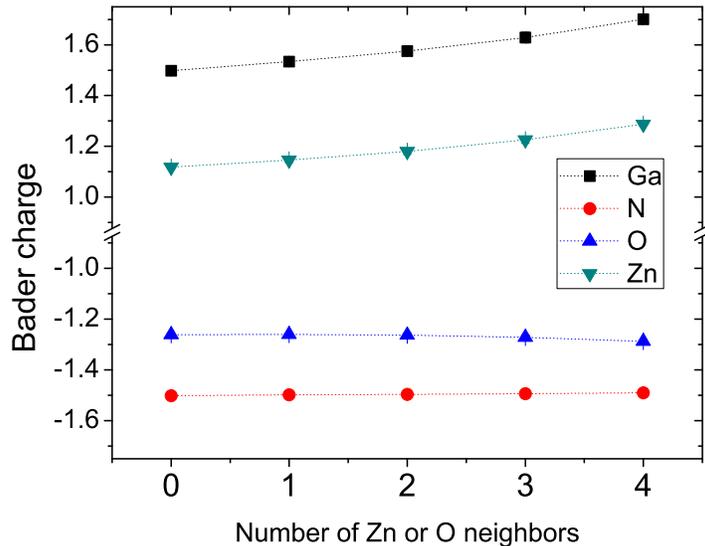
Figure 3.5: Bader charges calculated from DFT, averaged over the same species and the number of Zn (for N and O) or O (for Ga and Zn) nearest neighbors. The standard deviation is small. The error bar is less than the size of the symbol.

(fig. 3.6). The O neighbors reduce the charge on the cation and shrink the bond length. A possible explanation for the effect of the Zn neighbors is as following. The Zn atom in the alloy retains more charge than in ZnO crystal, leading to greater volume, effectively pushing the nearby bonds shorter.

The average electrostatic potential of each ion core also shows a strong dependence on its environment. For O and N atoms, the potential mainly depends on the species of first and second neighbors (fig. 3.7 and fig. 3.8). More Zn first neighbors increases the core potential. The second neighbors influence the core potential by altering the charges on the first neighbors. For Ga and Zn atoms, the first neighbors (anions) have the dominant effect (fig. 3.9 and fig. 3.10), because the charges on the first neighbors are relatively constant.

3.2.2 Long Periodic Superlattices

A computationally convenient way to study the long-range interactions, strain and electrostatic energies, is to calculate long-periodic superlattices (LPSL). These structures have alternating thick layers of GaN and thick layers of ZnO at different orientations. They are easy to construct and relatively cheap to
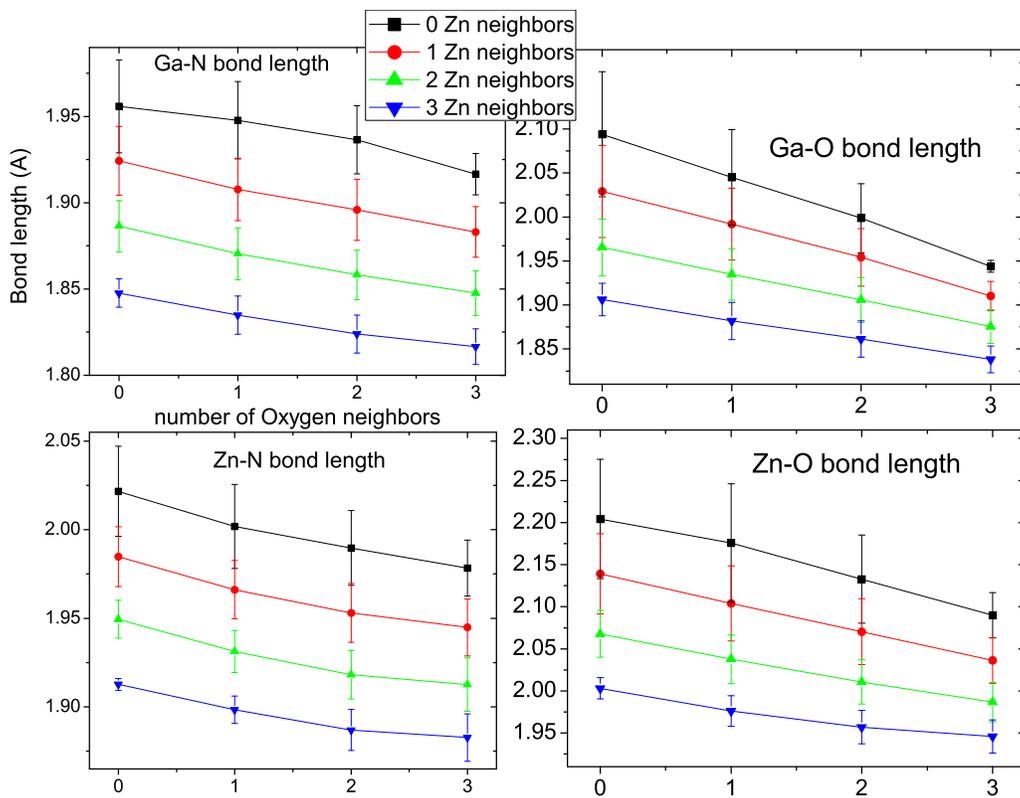
Figure 3.6: Average bond length from relaxed structures. The width of the error bars corresponds to two times the standard deviation.
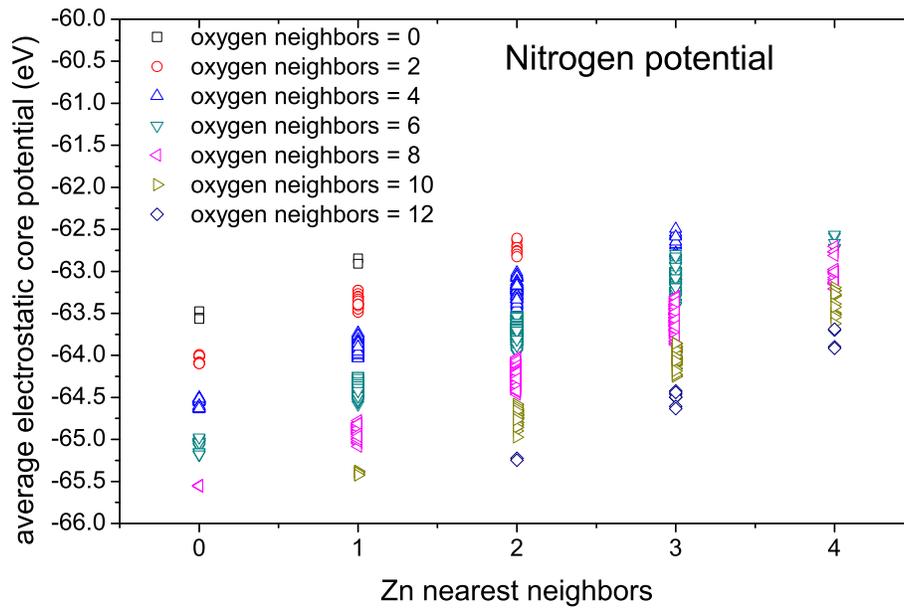
Figure 3.7: Average electrostatic potentials of N atoms, within a $2 \times 2 \times 1$ supercell. Data points are categorized by the number of first neighbor zinc atoms and second neighbor oxygen atoms.
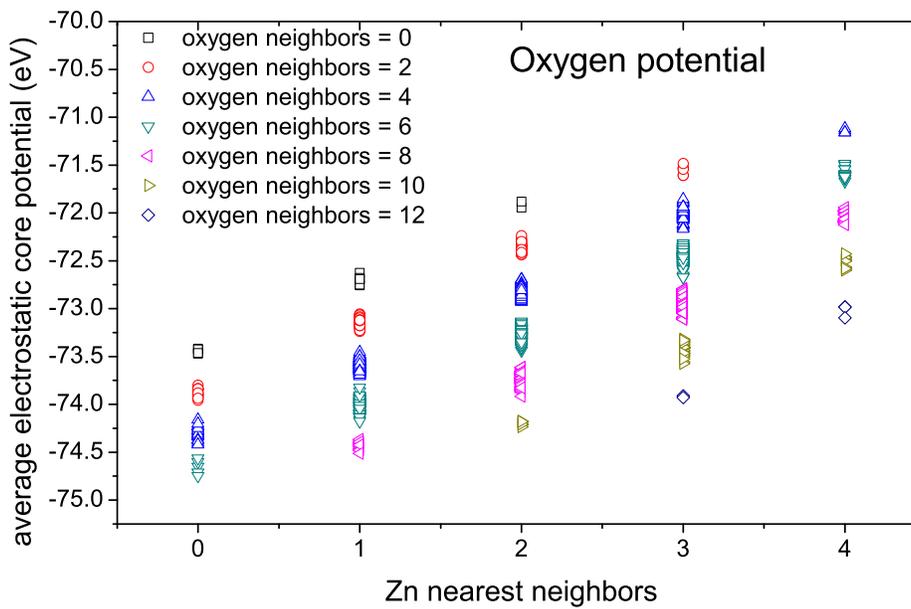
Figure 3.8: Average electrostatic potentials of O atoms.
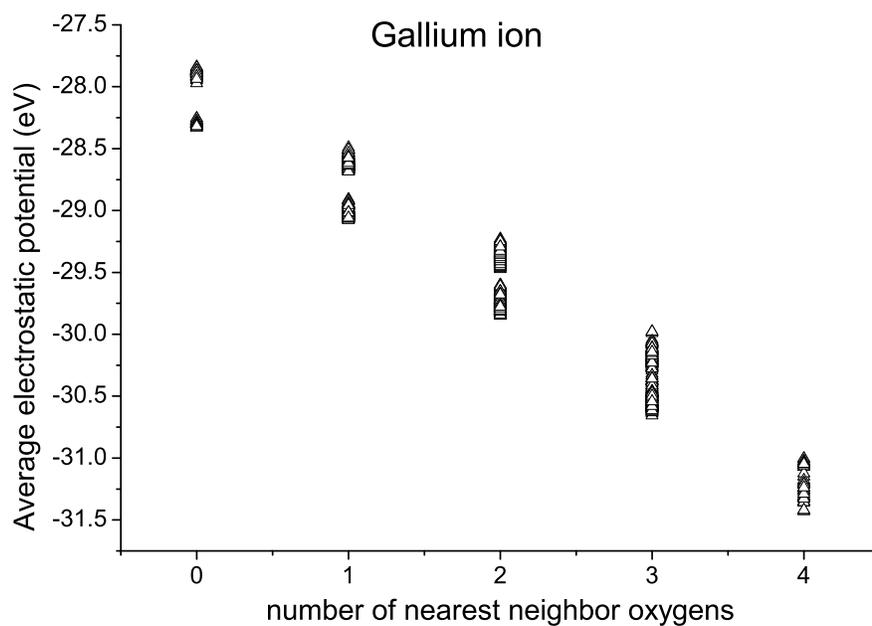
Figure 3.9: Average electrostatic potentials of Ga atoms.
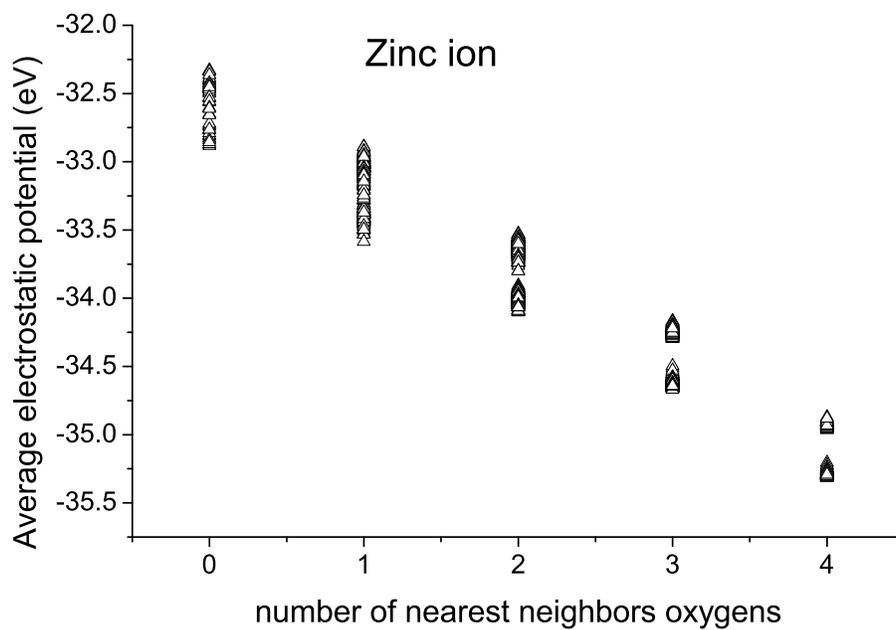
Figure 3.10: Average electrostatic potentials of Zn atoms.

15
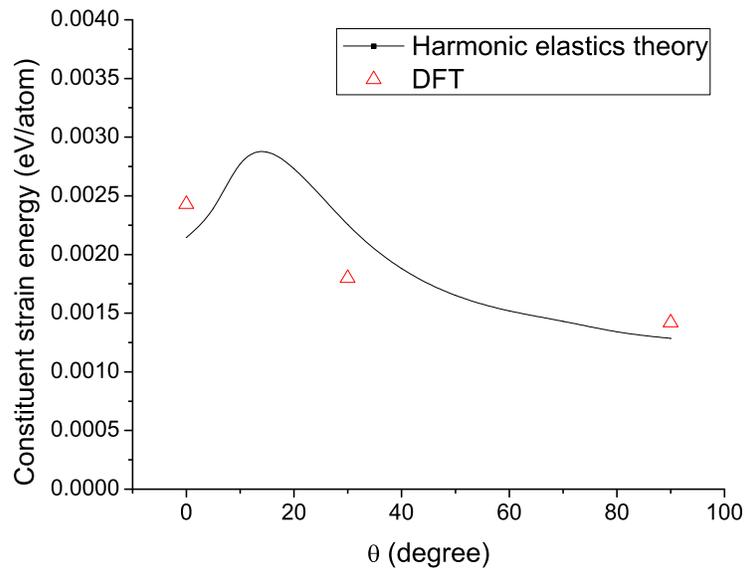
Figure 3.11: Predicted and calculated constituent strain energies for 50% ZnO LPSL at various inclination angles with respect to the c-axis. The predicted curve is obtained from harmonic elastic theory. The calculated values are from DFT.
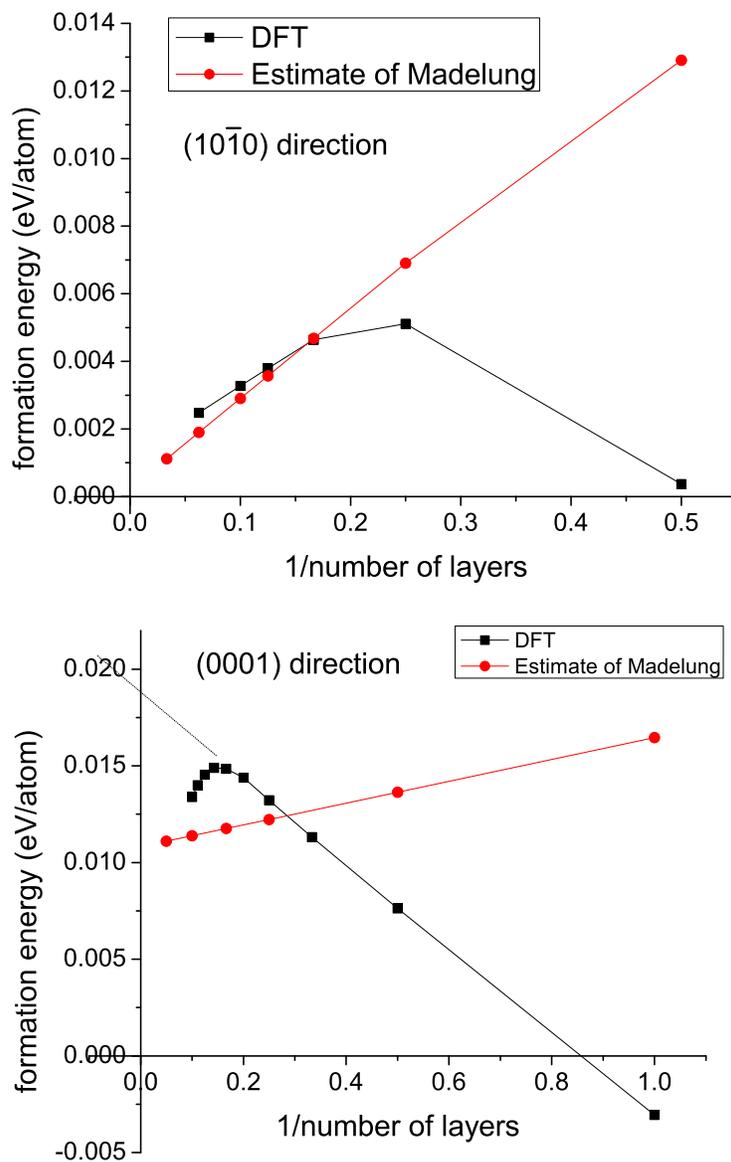
Figure 3.12: Formation energy of LPSL in (10$\bar{1}$0) and (0001) direction, calculated by DFT. An estimate of Madelung contribution is also shown in the graph. Please be aware that the Madelung contribution to the formation energy is very sensitive to the values of the ion charges and coordinates

17

calculate using DFT. We can get the asymptotic formation energy by extrapolating finite size LPSL to infinity.

The strain energy reaches a maximum if GaN and ZnO form layered structures. The constituent strain energy is an estimated strain energy for these types of structures. It is the minimum energy required to force GaN and ZnO to have the same perpendicular lattice size at a specific orientation. Therefore, it is angular dependent. We first estimate the strain energy of 50% ZnO LPSL by harmonic elastic theory, using published elastic constants for GaN and ZnO. Because of the small lattice mismatch between GaN and ZnO, this energy is only 1∼3 meV/atom, as shown in fig. 3.11. This result is confirmed by DFT calculations using deformations predicted from the harmonic elastic theory. For more random structures, or other ZnO concentrations, the strain energy is much smaller. Therefore, ignoring the strain energy will not cause errors in the formation energy beyond 3 meV/atom.

To estimate the electrostatic contribution to the formation energy, we calculate the Madelung energy using charges from the Bader analysis, on an ideal wurtzite lattice. The absolute magnitude of the Madelung energy is on the order of 10 eV/atom. However, by subtracting GaN and ZnO Madelung energy from the alloy, the contribution to the formation energy is only on the order of 10 meV/atom. Therefore, it is very sensitive to the volume and charges used. But it turns out to be of the same magnitude as the total DFT formation energy (see fig. 3.12). In the non-polar direction ($10\bar{1}0$), the Madelung energy approaches zero for a very thick LPSL. DFT will converge to the constituent strain energy of this direction (∼2 meV/atom). However, for the polar (0001) direction, the Madelung energy converges to 10∼20 meV/atom. It does not approaches zero for infinitely thick slabs. This is due to the polar nature of the wurtzite lattice in this orientation.

In conclusion, the dominant long-range interaction in this alloy is the Coulomb interaction. The resulting Madelung energy can be as high as 10∼20 meV/atom in extreme cases. For alloys with no long-range ordering, it is safe to ignore the long-range interactions.

3.3   Cluster Expansion (CE) Construction

Previous results suggest the importance of the atomic short-range order. To get a quantitative understanding, we use the cluster expansion formalism to build an accurate model to predict the formation energy for any atom site occupancy.

3.3.1 Method

The CE [28–32] is a standard tool in thermodynamic studies of alloys. Once constructed, it only requires the site occupancy as input to predict the formation energy of a specific configuration. Positional relaxation is implicitly included in our CE parameters but does not appear explicitly. This method uses an Ising-like model with spins σ_i on site i to represent occupation. If site i is a cation site, then $\sigma = 1$ denotes Zn and $\sigma = -1$ denotes Ga. Similarly, if site i is an anion site, then $\sigma = 1$ denotes O and $\sigma = -1$ denotes N. The total energy per four-atom primitive cell is the sum of the relevant one-, two-, and many-body interactions:

$$E = \sum_{\alpha} m_{\alpha} J_{\alpha} \langle \prod_{i \in \alpha'} \sigma_i \rangle \tag{3.3}$$

The index α is used to enumerate symmetry-inequivalent clusters, with multiplicity m_{α} per primitive unit cell. The angular bracket gives the average spin product for all clusters which are symmetrically equivalent to each other. The effective cluster interactions (ECI) J_{α} are obtained by fitting to a database of DFT energies of fully relaxed structures. The initial database contains randomly generated structures. It gives an initial CE model, which is then used to generate new trial structures in the low- and medium-energy range, which are then relaxed by DFT and added to the fitting database. This method has been successfully applied to a wide range of systems including metals, semiconductors, oxides, etc. It has also been generalized to treat multisublattice systems [33]. The $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution is a two-sublattice example, which contains not only clusters belonging to a single (cation or anion) sublattice, but also clusters containing both (see Fig. 3.13). The two-sublattice cluster expansion, if all clusters are taken into account, gives a complete basis set for the site occupancy space. The error of the cluster expansion construction is measured by the "leave-many-out" cross validation score (LMO-CV) [34–38]. Following the procedure described in Ref. [38], we split the database into construction data sets and validation data sets. The validation data set contains 30% of the entire database. For a specific selection of clusters, we fit the CE parameters using the construction data set, then calculate the mean squared error of prediction (MSEP) for the validation data set. The final LMO-CV is estimated by averaging the MSEP over $2N$ random splits of the input database of size N. To select appropriate clusters, a range of basis set cutoffs is examined to minimize the prediction error [39]. Routines in the ATAT package [40–45] are used to do the cluster expansion construction and the subsequent Monte Carlo simulation.

Table 3.1: Values of ECI in meV. The indexing of the two-body clusters is shown in Fig. 3.13. The zero-body term is normalized to one primitive cell.

	One-body		Two-body						
Zero-body	Cation	Anion	1	2	3	4	5		
495.69	−2.20	−2.20	−134.19	−112.95	29.29	29.82	31.24		
			Two-body						
6	7	8	9	10	11	12	13	14	
39.25	−6.89	−4.96	−4.19	−0.88	−1.55	−3.59	−5.24	−4.88	

3.3.2 Constructed Model

Figure 3.13 and Table 3.1 show the selected clusters and calculated effective cluster interactions for all the relevant clusters in the present paper. We construct the cluster expansion using a database of 120 structures calculated by DFT (up to a $4 \times 4 \times 3$ supercell). The CE contains 1 zero-body cluster, 2 one-body clusters (cation site and anion site), and 14 pair clusters. The ECIs for the 2 one-body clusters are degenerate due to the constraint of equal number of Ga and N atoms. This selection of clusters gives the minimum leave-many-out cross validation score of 3 meV/atom. Our tests show that including longer-range pair clusters or many-body clusters does not further reduce the LMO-CV. Like a well-behaved CE construction, the magnitude of the effective cluster interactions J_α decreases as the separation between the constituent atoms increases. Nearest-neighbor interactions (clusters 1,2 in Fig. 3.13) give the dominant contributions to the formation energy. The negative sign indicates a strong *clustering* tendency, e.g., Ga prefers N neighbors rather than O neighbors. This is due to the matching valence charge in Ga-N and Zn-O bonds rather than Ga-O and Zn-N bonds in a tetrahedrally coordinated environment. The difference between the ECIs of pair 1 and pair 2 shows that the *clustering* tendency in the *ab* plane is stronger than along the *c* axis. All of the second-neighbor interactions are positive, indicating an *ordering* tendency, which represents a repulsion between the same species, e.g., Ga prefers Zn as a second neighbor rather than Ga. These two competing tendencies determine the short-range order we will discuss later.
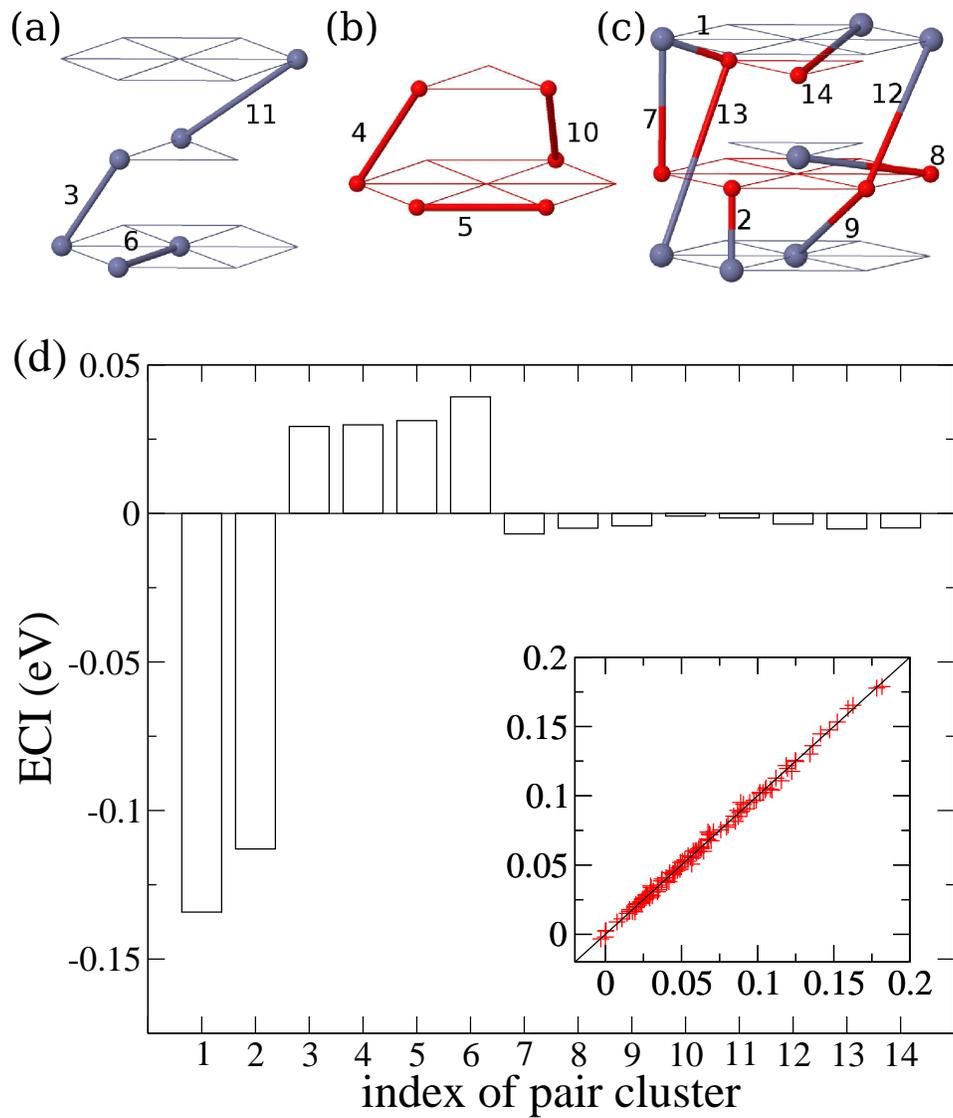
Figure 3.13: Numbering of clusters and calculated effective cluster interactions (ECIs). Zero- and one-body clusters are not shown in the figure. The ECIs are indexed by the separation of their constituent atoms. The distance of pair 14 is 5 Å. (a) Cation-cation clusters. (b) Anion-anion clusters. (c) Cation-anion clusters. (d) Effective cluster interactions. Inset: comparison of formation energy between CE prediction (y) and DFT calculation (x), in units of eV/atom.

3.4 Monte Carlo Simulation

3.4.1 Method

Monte Carlo simulation is used to investigate the thermodynamic properties and phase diagram. The simulation uses a $14 \times 14 \times 8$ supercell with periodic boundary conditions. We only allow MC moves that change the number of Ga and N atoms by the same amount, so that the stoichiometric constraint is satisfied. The equilibration of the structure and averaging of thermodynamic quantities takes at least 50,000 steps/atom. Convergence tests suggest that the accuracy of the energy averaging is better than 0.2 meV/atom.

3.4.2 Thermodynamics, Phase Diagram, and Short-Range Order

Monte Carlo simulations are performed to investigate the equilibrium thermodynamic properties. Figure 3.14(a) shows the formation energy averaged over thermal ensembles of configurations as a function of temperature. At $x = 0.5$, the alloy is predicted to undergo a first-order phase transition from an ordered compound to the disordered phase as T increases above 870 K. At $x = 0.25$, the disordered phase is predicted to exist above 760 K, and to become phase separated at lower temperature. Actual samples have not been found with these long-range orders, presumably because 870 K is too low for equilibration to occur.

Based on the MC simulation, we propose a theoretical phase diagram (Fig. 3.15) for the $(\mathrm{Ga}_{1-x}\mathrm{Zn}_x)(\mathrm{N}_{1-x}\mathrm{O}_x)$ solid solution. It has a wide miscibility gap and an $x = 0.5$ stable compound. The stable compound has layered ordering in the (0001) direction as shown in Fig. 3.15(b), with the same periodicity as the wurtzite structure. The atoms are arranged so that, among the four first neighbors of Ga, there are three N atoms and one O; among the twelve second neighbors of Ga, there are six Zn and six Ga atoms. Zn, N, and O atoms experience a similar environment. This structure is a delicate compromise between the *clustering* tendency for first neighbor and the *ordering* tendency for second neighbor. Its formation energy is about -3 meV/atom, barely stabilized against phase separation into GaN and ZnO.

In our simulations, the disordered phase displays strong short range *clustering* [Fig. 3.14(c) and 3.14(e)]. This effect can be quantified by the Warren-Cowley short range order (SRO) parameter α_{lmn}, defined as

$$\alpha_{lmn}(x, T) = 1 - \frac{P_{lmn}^{A(B)}(x, T)}{x}, \tag{3.4}$$

(a)

(b) x=0.25,T=700K   (c) x=0.25,T=900K

(d) x=0.5,T=700K   (e) x=0.5,T=900K
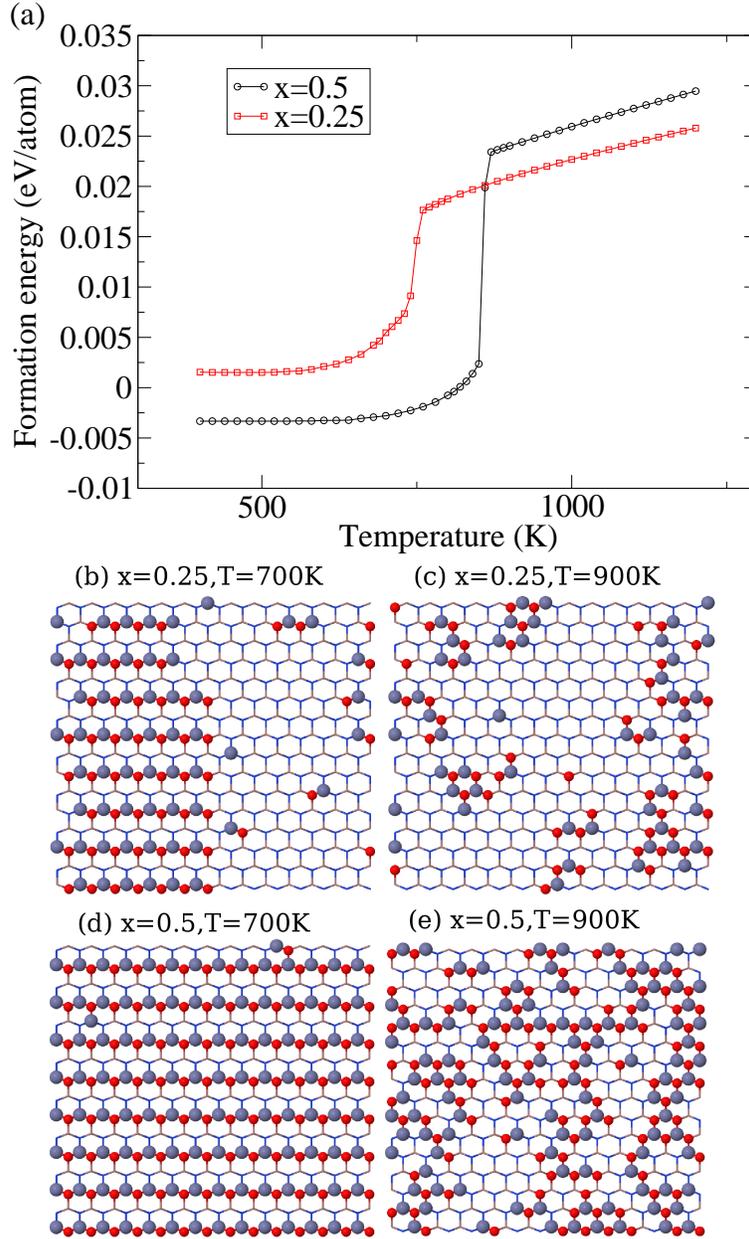
Figure 3.14: (a) Formation energies of the solid solution $(\mathrm{Ga}_{1-x}\mathrm{Zn}_x)(\mathrm{N}_{1-x}\mathrm{O}_x)$ calculated from Monte Carlo simulation at concentrations $x = 0.5$ and $x = 0.25$. (b)(c)(d)(e) Snapshots from the Monte Carlo simulation. Only a $14{\times}1{\times}8$ slice of the $14 \times 14 \times 8$ simulation cell is shown. In the graph, the horizontal direction is the wurtzite a lattice vector. The vertical direction is the c vector. Small (red) balls, oxygen; large (blue) balls, zinc; gallium and nitrogen atoms are hidden.

23

(a)

(b)
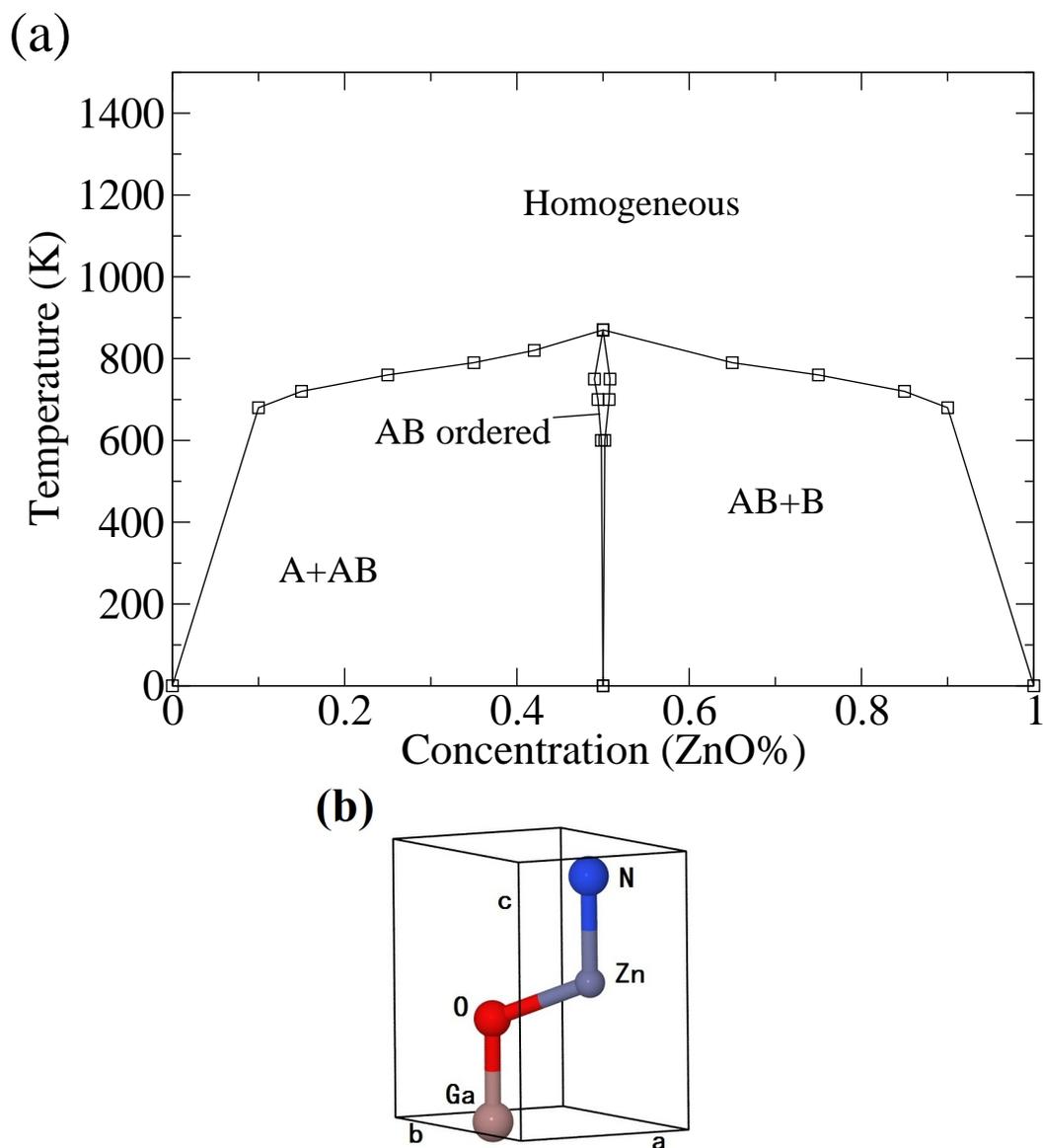
Figure 3.15: (a) Computed phase diagram of $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution. Straight lines are guides to the eye. Phase A is mostly GaN. Phase B is mostly ZnO. Phase AB is the ordered superlattice structure. (b) Ball-stick model of the ordered AB compound.
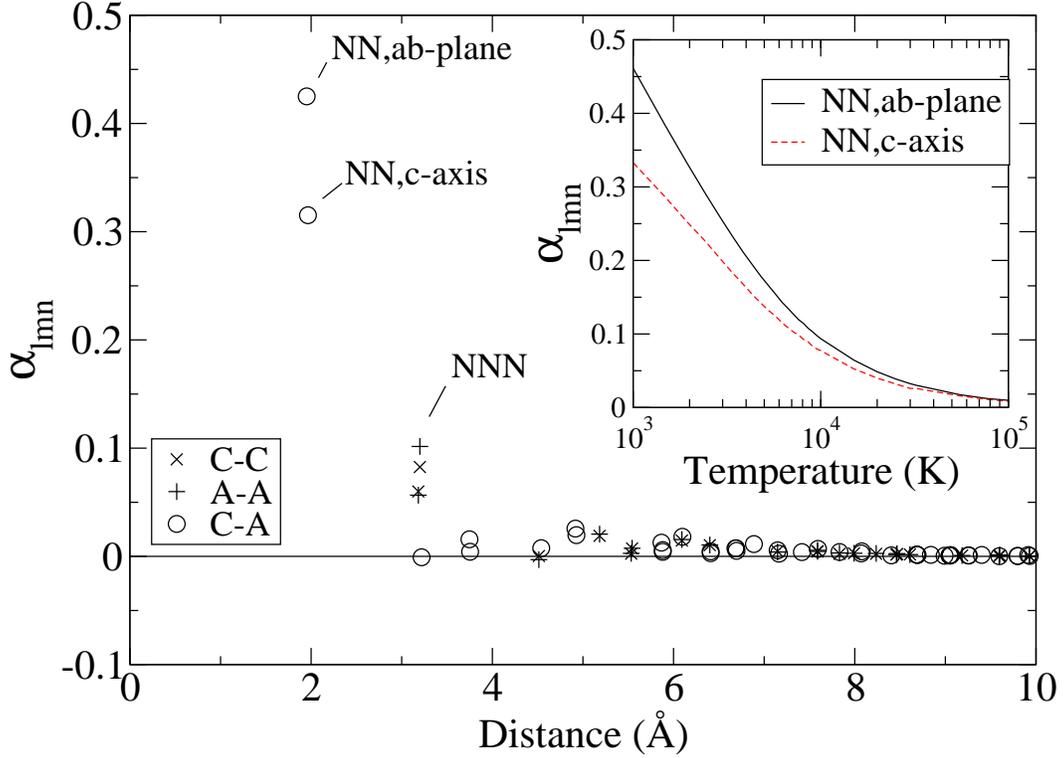
Figure 3.16: Calculated SRO for x=0.2 and T=1200 K. Each point represents a different type of pair (C-C, cation-cation pair; A-A, anion-anion pair; C-A, cation-anion pair; NN, nearest neighbor; NNN, next-nearest neighbor). Inset: temperature dependence of SRO for the first neighbors in ab plane and c axis.

where x is the concentration of ZnO, T is the equilibration temperature, and $P_{lmn}^{A(B)}$ is the conditional probability of finding a B atom in the lmn shell, given that the center atom is A. There are three types of SRO in the $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution. For a pair of cation sites, A is Ga and B is Zn. For a pair of anion sites, A is N and B is O. For a pair of cation and anion sites, A is Ga and B is O. Positive SRO indicates *clustering* and negative indicates *ordering*. Figure 3.16 shows the calculated SRO at $x = 0.2$ and $T = 1200$ K. The SRO is positive for first- and second-neighbor shells; it quickly decays to zero at and beyond the third neighbor. This *clustering* tendency persists to very high temperatures (see inset in Fig. 3.16). Therefore, the SRO is an inherent characteristic of the $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution. It remains relatively constant within the range of synthesis temperatures and can not be removed.

3.4.3 Lattice Parameters, Bond Lengths, and Band Gaps

The Monte Carlo simulation based on the cluster expansion can only predict site occupancies. It can not provide direct information about coordinate relaxation or electronic structure. However, we can obtain this information from DFT. The investigation of lattice parameters, bond lengths and band gaps contains two steps. First, we conduct Monte Carlo simulation and equilibrate the structure at a specific temperature and concentration. Then, we randomly draw snapshots from the simulation and use them to do DFT calculations. Due to DFT's limited capability of handling large structures, we restrict the supercell to be $4 \times 4 \times 3$, with 192 atoms. To average over the fluctuations due to the finite size of the simulation cell, four snapshots are taken at each temperature and concentration. We estimate the quantities of interest, e.g., the band gap, from DFT calculations of these snapshots.

Actual $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ samples at room temperature do not show the ordered binary or phase-separated structures, because low atom mobility below 900 K inhibits equilibration. Since the temperature dependence of SRO is relatively weak (see Fig. 3.16), we adopt 1200 K as a reasonable effective equilibration temperature characterizing actual samples at lower temperature. Although the measurements of band gaps, etc., are conducted at room temperature, it is appropriate to compare with theory at the higher effective equilibration temperature.

Figure 3.17 shows the lattice parameters extracted from DFT calculations of these snapshots. As comparisons, we also considered snapshots from a MC temperature of 5000 K, which exhibits half as much SRO (see inset of Fig.3.16). In reality, the sample would decompose at such a high temperature; we use it here simply to study the influence of ordering. We find the upward bowing predicted from snapshots at 5000 K to be approximately twice that found at 1200 K. Greater disorder causes the lattice parameters to increase. Experimentally, c bows more than a [1], whereas DFT gives equal bowing. The overall agreement on the magnitude of upward bowing suggests that SRO exists in the samples reported by Chen *et al* [1].

Figure 3.18 shows the analysis of cation-anion (nearest neighbor) bond lengths. Although the bond in the ZnO crystal is longer than that in GaN, the difference becomes even larger in the alloy. The Ga-N bond shrinks further and the Zn-O bond expands upon mixing. This unusual bond relaxation is a consequence of the non-isovalent nature of the alloy. The ZnSe-GaAs system shows similar behavior [46], in which the Zn-Se bond expands and the Ga-As bond shrinks. However, the average bond length for all cation-anion bonds follows approximately a linear relationship. This is due to the change in the proportion of different types of bond, i.e., there are more Zn-O bonds in a
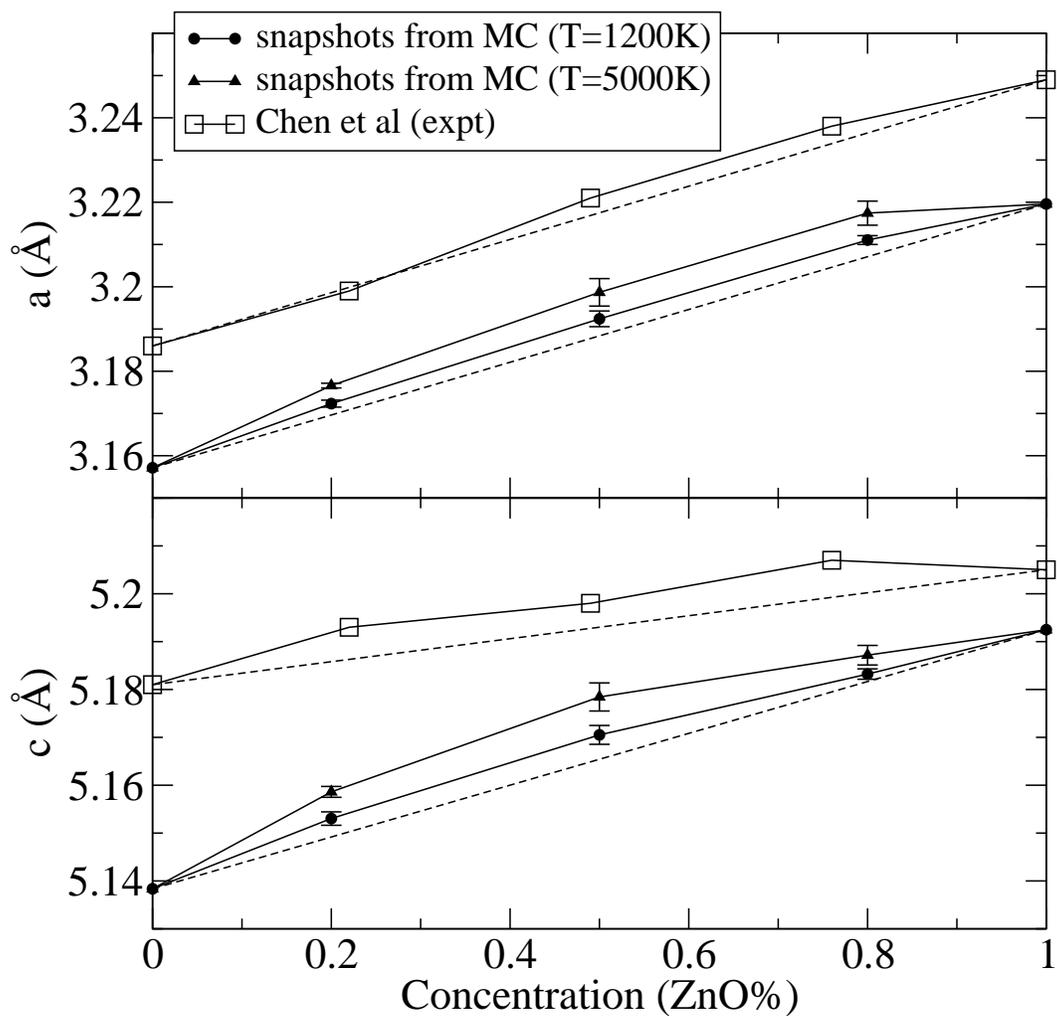
Figure 3.17: Filled symbols show the lattice parameters from DFT based on MC snapshots. The error bars cover twice the standard deviation of underlying snapshots (four per point). Open symbols show the experimental results (Ref. [1]). Dashed lines show the predictions from Vegard's law.
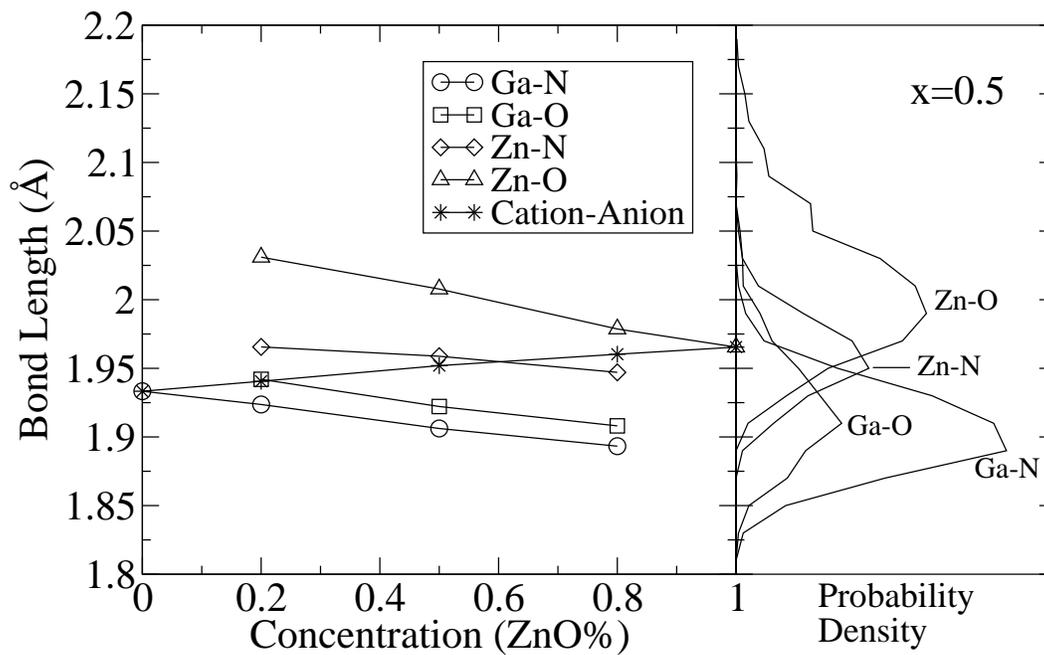
Figure 3.18: (Left) Average bond lengths of snapshots from T=1200 K MC simulation. (Right) Probability distribution of bond types and bond lengths at x=0.5

ZnO-rich condition.

Figure 3.19 shows the comparison between calculated and measured band gaps. To correct for the well-known errors in the band gap as calculated with DFT, a composition-dependent adjustment is included. For any structure σ with composition $x(\sigma)$, the adjusted band gap is

$$E_{g,\text{adjusted}}(\sigma) = E_{g,\text{DFT}}(\sigma) + \Delta(x), \tag{3.5}$$

where

$$\begin{aligned} \Delta(x) = \;\; &(1-x)(E_{g,\text{expt}}(\text{GaN}) - E_{g,\text{DFT}}(\text{GaN})) \\ &+ x(E_{g,\text{expt}}(\text{ZnO}) - E_{g,\text{DFT}}(\text{ZnO})). \end{aligned} \tag{3.6}$$

A useful quantity in the analysis of alloy band gaps is the bowing parameter b, defined as

$$E_g(\sigma) = (1-x)E_g(\text{GaN}) + xE_g(\text{ZnO}) - bx(1-x), \tag{3.7}$$

which, for any configuration σ, describes its deviation (in parabolic approximation) from linear interpolation between the two end points. The band gap from snapshots of 1200 K MC simulation is symmetric but not perfectly parabolic. The bowing is slightly greater at low and high ZnO concentrations. Compared to 1200 K, the snapshots from 5000 K MC simulation have much larger bowing parameters, indicating a red shift of the band gap, induced by disorder, consistent with results of Wang *et al.* [9]. The asymmetric behavior is due to the different band-gap-reducing mechanism at the dilute limit [6]. Using Eq. (3.2), the fitted bowing parameter at 1200 K is 3.34 eV. Previous work, which did not take the strong short-range order into account, predicted the bowing parameter to be 4.05 eV [7]. Compared to experiments, the band gap of 1200 K MC snapshots closely follows the value from the high-temperature and high-pressure synthesized samples (Chen *et al.* [1]). It also agrees well with the 22% ZnO sample from Maeda *et al.* [2]. In the regime of lower ZnO concentration, the trend of our calculated data requires a bowing parameter greater than 3.34 eV. Indeed, the theoretical investigation by Wang *et al.* [9]. predicted the bowing parameter for the 12.5% alloy to be 4.8 eV (at 1100 K). The experimental results from Maeda *et al.* [2]. indicate that the bowing parameter increases with decreasing ZnO concentration, from ∼4 eV at 22%, to ∼12 eV at 5%. In summary, we find that both concentration and disorder cause band gap bowing to deviate from a simple T-independent parabola.
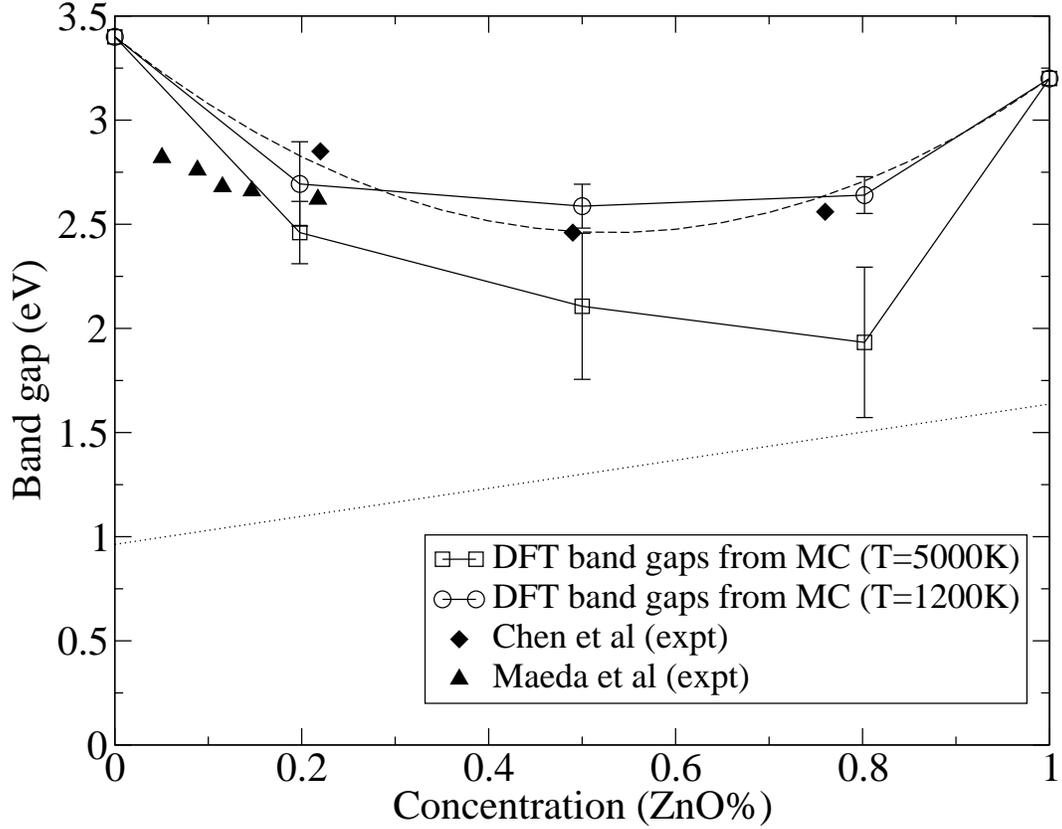
Figure 3.19: Comparison of band gaps between DFT and experiment. Open symbols represent DFT results on snapshots described in previous figures. DFT results have been shifted up by the dotted line (Eq. (3.6)) to enable comparison with experimental data. The dashed line is the parabolic fitting of the DFT results, with a bowing parameter of 3.34 eV. Diamonds are the experimental values from Chen *et al.* (Ref. [1]). Triangles are the experimental results from Maeda *et al.* (Ref. [2]).

3.5 Conclusion

Our investigation of the $2 \times 2 \times 1$ supercells show the cation charge strongly dependent on its environment. The anion charge remain constant. The bond length and the ion core potential change as the ion charge changes. To gain quantitative understanding of the alloy site occupancies, we construct a cluster expansion model for $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solutions which accurately extrapolates DFT energies. Monte Carlo simulation reveals a phase diagram with a wide miscibility gap and an $x = 0.5$ stable compound below 870 K. At synthesis temperatures, the solid solution is in the disordered phase. Strong short-range order is an inherent property and remains relatively constant within the likely range of equilibration temperatures. Based on snapshots from MC simulation, we investigate the structure and electronic properties by DFT. The lattice parameters are found to deviate from Vegard's law. The upward bowing is increased by randomness. The relaxation of bond lengths is unusual and can be attributed to the different valences of GaN and ZnO. Short-range order also induces a blue shift in the band gap.

Chapter 4

$(\mathbf{Ga}_{1-x}\mathbf{Zn}_x)(\mathbf{N}_{1-x}\mathbf{O}_x)$ Solid Solution: Surfaces

Surfaces play a key role in the photo-catalytic water-spitting process. In this chapter, we first introduce the fundamental theories and methods. We simulate the surface with the slab model. Various aspects of the simulation, including surface passivation, spontaneous polarization and chemical potentials are carefully discussed. Then, we study the reconstructions of the alloy surface. Only the GaN bulk with Ga, N and O ad-atoms are considered. An evolutionary algorithm is used to explore the surface structures. A few stable structures are found at different chemical potentials.

4.1 Theories and Methods

In this section, we introduce the key theories and methods used in the simulation of surfaces.

4.1.1 Slab Model

One way to simulate the surface within the 3D periodic boundary condition is using the slab model, as shown in fig. 4.1 (a GaN $(10\bar{1}1)/(10\overline{11})$ slab). Vacuum is inserted to separate the slab from its images. The two exposed surfaces are separated by a few layers of the bulk material. In practice, only atoms near the surfaces are relaxed during the calculation. Atoms in the middle of the slab are fixed to the 3D crystal coordinates. The slab must be thick enough to prevent two surfaces from interacting with each other. For the non-polar directions, e.g., a GaN $(10\bar{1}0)$ slab, the exposed surfaces can often be chosen to be equivalent. Otherwise, the exposed surfaces are different, e.g.,
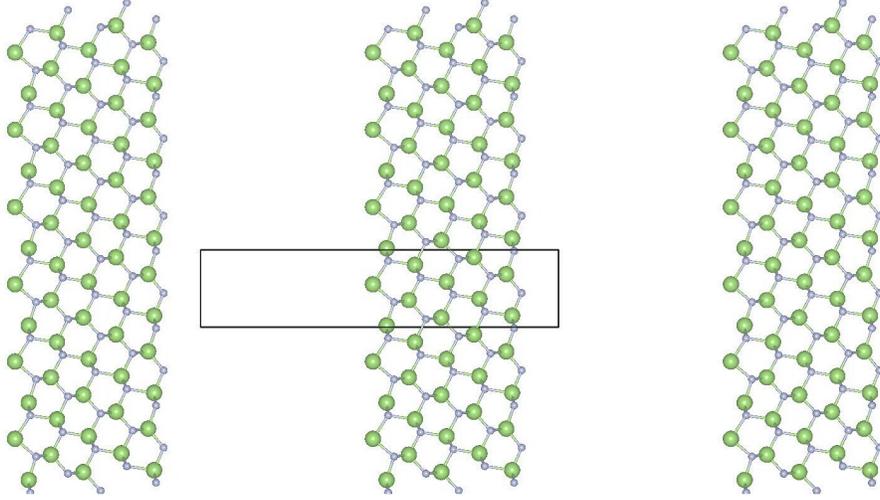
Figure 4.1: The slab model used in the first-principles simulations. The GaN $(10\bar{1}1)$ and $(10\overline{11})$ are exposed. The black rectangle indicates the unit cell. it is periodic in 3D. The vacuum is inserted to separate the slab (middle) with its images (e.g., slabs on the left and right). Large balls, Ga atoms; Small balls, N atoms.

a GaN $(10\bar{1}1)/(10\overline{11})$ slab. If we are interested in studying a polar surface, an inequivalent (and sometimes complementary) surface always exists at the back of the slab. Care must be taken to make sure that the back surface does not interfere with our study.

4.1.2 Macroscopic Averaging

We are usually interested in the behavior of a quantity along the direction perpendicular to the surface (the z direction, from now on). For example, we can average the electrostatic potential $V(x, y, z)$ over the xy directions,

$$\overline{V}(z) = \frac{1}{S} \int_S V(x, y, z) \ dx \ dy, \tag{4.1}$$

where S is the surface area of the unit cell. The planar averaged quantity, $\overline{V}(z)$, gives the trend of the potential along the z direction. However, it contains oscillations due to the microscopic structure of the slab (fig. 4.2). To further eliminate these oscillations, we can perform the macroscopic averaging, or
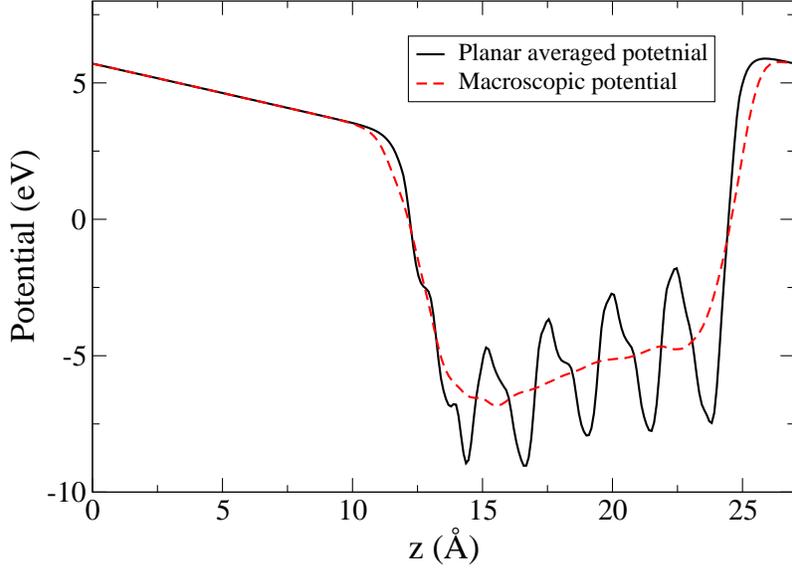
Figure 4.2: The planar averaged potential $\overline{V}(z)$ and the macroscopic potential $\overline{\overline{V}}(z)$ for the slab model shown in fig. 4.1.

window-averaging [47],

$$\overline{\overline{V}}(z) = \frac{1}{d}\int_{z-d/2}^{z+d/2}\overline{V}(z')dz', \tag{4.2}$$

where d is the periodicity of the 3d crystal along the z direction. The macroscopic electric field inside the slab is straight forward to extract from $\overline{\overline{V}}(z)$. Similarly, we can also calculate the macroscopic electron density $\overline{\overline{\rho_e}}(z)$ from $\rho_e(z)$.

4.1.3 Dipole Correction

Even with the inserted vacuum layer, the artificial periodic images can still interact with the slab through Coulomb interaction. Increasing the thickness of the vacuum can reduce such interaction, at the price of more computational costs. The dominant interaction between the slab and its images is the dipole-dipole one. By inserting a fictitious dipole layer inside the vacuum, the dipole interaction can be cancelled [48, 49]. The inserted dipole layer introduces a jump in the electrostatic potential in the vacuum, as shown in fig. 4.3. The amount of the inserted dipole is determined self-consistently to ensure zero E
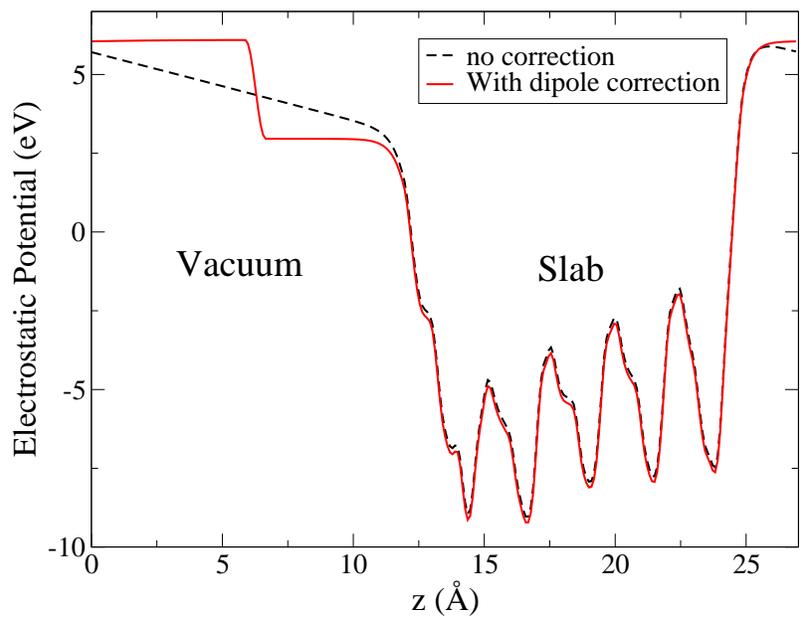
Figure 4.3: Comparison of the electrostatic potential, with and without the dipole correction, for the slab model shown in fig. 4.1. During this calculation, all atoms are kept fixed. The electrostatic potential is averaged over the xy directions (Eq. 4.1).

field in the vacuum region. With the correction, both the total energy and the forces on the surface atoms converge more rapidly with respect to the vacuum thickness (fig. 4.4). Therefore, in our surface calculations, we always use the dipole correction. The discontinuity changes the electric field inside the vacuum. It can also be used to generate an external E field applied on the slab.

4.1.4   Electron Counting Rule and Passivation

The electron counting rule emerges from the study of semiconductor surfaces [50]. This rule, although not exact, is very useful in comparing surface energies, estimating surface charges and guessing surface reconstructions. It states that the surface structures with filled dangling bonds on the electronegative elements and empty dangling bonds on the electropositive elements are the lowest in energy. The surface then is likely to be semiconducting. Surfaces not satisfying this rule will be metallic and higher in energy. The rule assumes that the dangling bonds from the cations (the electropositive elements) are higher in energy than those from the anions (the electronegative elements). Redistributing electrons between these two can lower the surface energy. The energy reduction is maximized if there is no net charge left.

The surface on the back of the slab can have partially filled cation dangling bonds or partially empty anion dangling bonds. It can interact with the top surface by transferring electrons. To prevent that, fractional hydrogen can be used to passivate these dangling bonds, making sure that the electron counting rule is satisfied [51]. For example, the GaN $(000\bar{1})$ surface has one N dangling bond for each surface unit cell. Since the N contributes $5/4$ electrons to each bond, this dangling bond needs $3/4$ electron to become filled. A ficticious hydrogen of $Z = 3/4$ with $3/4$ electron can bond to this N, making the surface semiconducting. In practice, the coordinates for the fractional hydrogen are chosen so that the N$-$H surface state is filled and moved into the bulk valence band, leaving no states in the fundamental band gap of the GaN crystal.

4.1.5   Spontaneous Polarization

For polar materials like GaN, the spontaneous polarization plays an important role in the slab calculations. It alters the electrostatic potential inside the slab along the polar direction. The spontaneous polarization is a bulk property. Here I briefly introduce the theories.

The absolute polarization of a 3D infinite crystal is not easy to define. The
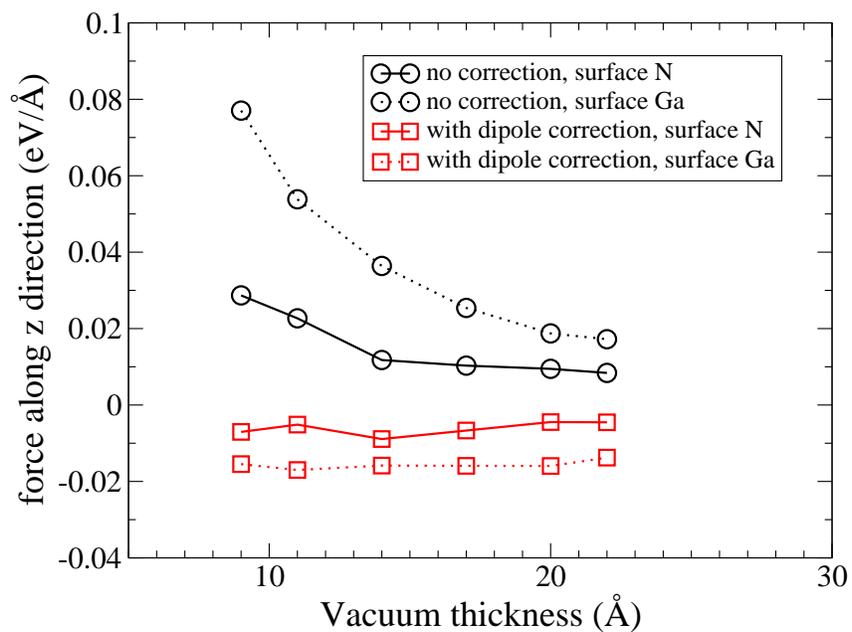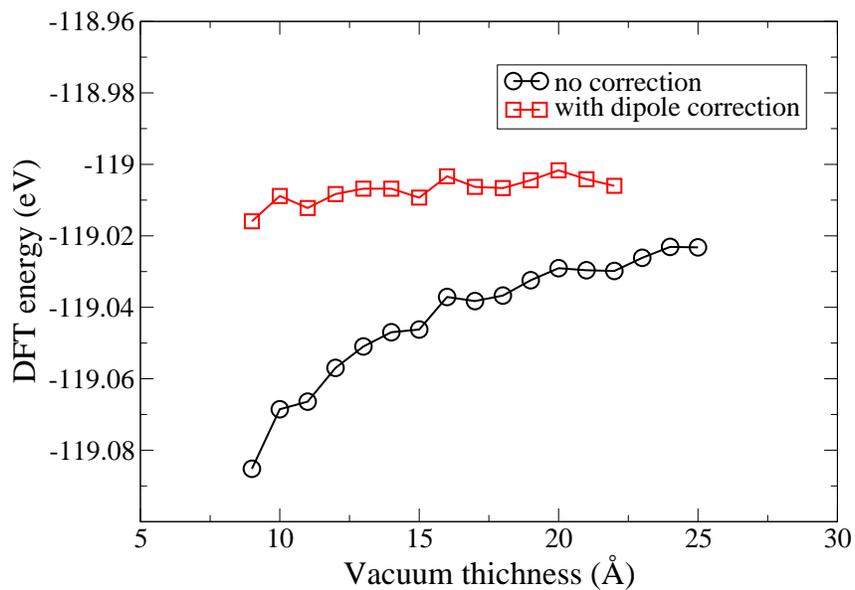
Figure 4.4: Convergence of the energy (upper) and the forces on the surface atoms (lower) with respect to the vacuum thickness. Surface N refers to the rightmost atom of the slab in fig. 4.1. Surface Ga refers to the leftmost atom.

intuitive approach is to sum the contribution from electrons and nuclei, as

$$\mathbf{P} = \frac{1}{\Omega} \sum_i Z_i \mathbf{R}_i + \int_\Omega \mathbf{r} \rho_e(\mathbf{r}) d\mathbf{r}, \tag{4.3}$$

where Ω is the unit cell volume, Z_i and $\mathbf{R}_i$ are the charges and coordinates of the ions, $\rho_e(r)$ is the electron charge density. This approach does not work, since it depends on the choice of the unit cell boundary. The alternative approach is to focus on the change of the polarization [52],

$$\Delta \mathbf{P} = \int_{\lambda_1}^{\lambda_2} \frac{\partial \mathbf{P}}{\partial \lambda} d\lambda = \mathbf{P}(\lambda_2) - \mathbf{P}(\lambda_1), \tag{4.4}$$

where λ is the parameter on a path to transform structure 1 into structure 2. The change of the total polarization is the sum of changes of the ionic and the electronic parts,

$$\Delta \mathbf{P} = \Delta \mathbf{P}_{\mathrm{nuc}} + \Delta \mathbf{P}_e. \tag{4.5}$$

The ionic part is trivial to calculate. The electronic part is well defined if the transformation is adiabatic and the system remains insulating along the path [52]. In fact, it has been shown that $\mathbf{P}_e$ follows

$$\mathbf{P}_e(\lambda) = -\frac{ie}{(2\pi)^3} \sum_{n=1}^M \int_{\mathrm{BZ}} d\mathbf{k} \langle u_{n\mathbf{k}}^{(\lambda)} | \nabla_\mathbf{k} | u_{n\mathbf{k}}^{(\lambda)} \rangle, \tag{4.6}$$

where M is the number of occupied bands, BZ stands for the Brillouin zone, $u_{n\mathbf{k}}^{(\lambda)}$ is the cell periodic function defined as $u_{n\mathbf{k}}^{(\lambda)} = e^{-i\mathbf{kr}} \psi_{n\mathbf{k}}$ [53, 54]. The $\mathbf{P}_e(\lambda)$ is only well-defined *modulo* a quantum $fe\mathbf{R}/\Omega$, where $f = 2$ for the spin-degenerated case. Since the result (Eq. 4.6) for the electronic polarization does not depend on the path, only the initial and final states are needed in the calculation. However, if the change of the polarization is greater than the quantum, more structures are needed along a path to obtain the exact result.

Both the wurtzite GaN and ZnO crystals have spontaneous polarization. To calculate the absolute magnitude, their metastable zinc blende structures, which have no polarization as required by the symmetry, can be used as the reference structures. Our calculated results for GaN are shown in Table 4.1 along with available experimental data and previous theories. The ion-clamped dielectric constants ϵ_∞ are calculated through finite difference using small electric fields (by VASP) [60, 61]. It follows

$$\mathbf{P} = \mathbf{P}_0 + \chi \epsilon_0 E, \tag{4.7}$$

Table 4.1: The GaN spontaneous polarization $\mathbf{P}_0$, static (ϵ_{static}) and ion-clamped (ϵ_∞) dielectric constants.

	$\mathbf{P}_0(C/m^2)$	$\epsilon_{\text{static}} \parallel c$	$\epsilon_{\text{static}} \perp c$	$\epsilon_\infty \parallel c$	$\epsilon_\infty \perp c$
Expt (Ref. [55])		10.04	9.5	5.35	
Expt (Ref. [56])				5.70	
Theory (Ref. [57])	-0.029				
Theory (Ref. [58])		10.28		5.69	
Theory (Ref. [59])	-0.032 (LDA) -0.034 (GGA)				
This study	-0.038			5.45	5.28

where E is the macroscopic electric field, χ is the susceptibility with $\chi = \epsilon - 1$. This equation ignores higher order terms in the E field. The spontaneous polarization is defined under the zero field condition ($E = 0$). Our calculated spontaneous polarization is greater than that in Ref. [57, 59]. The discrepancy may caused by the different exchange-correlation functional, our on-site Coulomb interactions, or different lattice parameters and atom coordinates. The calculated ion-clamped dielectric constants reproduce fairly well previous experiments and theories.

4.1.6 Slabs with Spontaneous Polarization

Slabs with spontaneous polarization are slightly more complicated in the simulation than those without. Since the zinc blende GaN has no spontaneous polarization, we use it as a comparison to study the wurtzite GaN slab. The zinc blende $(111)/(\overline{111})$ slab is very similar to the wurtzite $(0001)/(000\overline{1})$ slab. Their surface structures are the same, only the stacking of atoms in the bulk is different (see fig. 4.5). To study the effects of the surfaces and external E fields, four slabs (wurtzite versus zinc blende, semiconducting versus metallic surfaces) are calculated under three external E fields.

Figure 4.5 shows the slab model used in the DFT calculation. We construct the slab by cleaving the bulk crystal. Both the wurtzite $(0001)/(000\overline{1})$ slab and the zinc blende $(111)/(\overline{111})$ slab contain 5 layers of Ga and N. These two have similar surface structures. Both top surfaces are terminated by Ga atoms, while both bottom surfaces are terminated by N atoms.

The cleaved surfaces are metallic due to the partially filled dangling bonds. To create semiconducting surfaces, we use fractional hydrogen to passivate the dangling bonds as describe before. A $Z = 5/4$ hydrogen atom is placed 1.55 Å on top of the Ga to compensate its one dangling bond, and a $Z = 3/4$
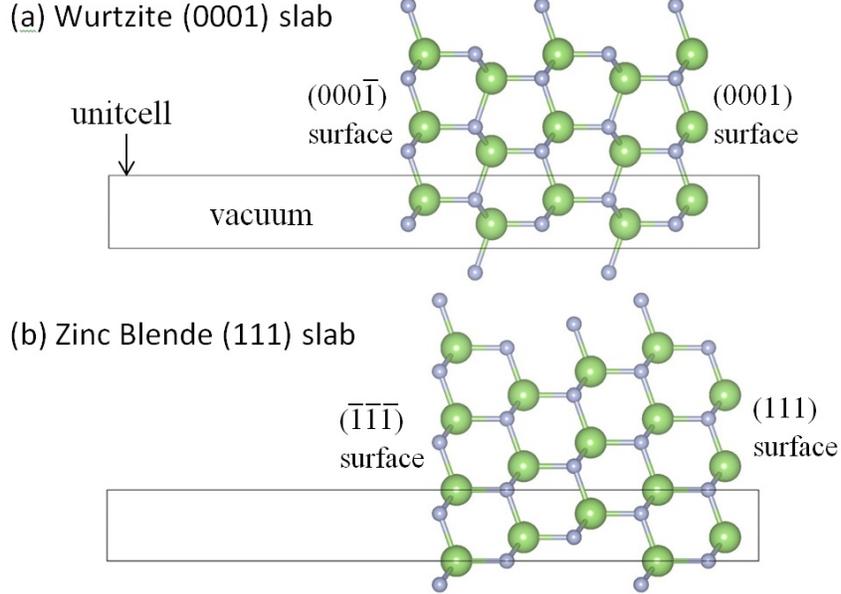
Figure 4.5: The wurtzite $(0001)/(000\bar{1})$ and the zinc blende $(111)/(\bar{1}\bar{1}\bar{1})$ slabs constructed by cleaving their bulk crystal.

hydrogen atom is placed 1.05 Å on top of each N to compensate its dangling bond. All surface bands are either filled or empty, creating a band gap (around 2 eV) close to the DFT+U value of the fundamental band gap of the GaN crystal. Therefore, without free electrons or holes, the passivated surfaces are semiconducting.

In our DFT calculation, the slab and its images are separated by at least 10 Å vacuum. To simplify the calculation and its interpretation, all ions are fixed at their ideal coordinates. To realize the predefined external E field, a dipole layer is introduced in the vacuum region self-consistently, to achieve the desired external E field. The DFT calculation uses the VASP package, with the PBE functional and the PAW basis set. The plane wave cutoff is 400 eV. An on-site Coulomb interaction $U = 3.9$ eV is applied to the Ga $3d$ orbital.

The internal E fields are extracted from the macroscopic electrostatic potential. The surface charge density is obtained by integrating the macroscopic charge density of both electrons and nuclei. The slabs with semiconducting or metallic surfaces display very different behaviors. Here I will study them separately.
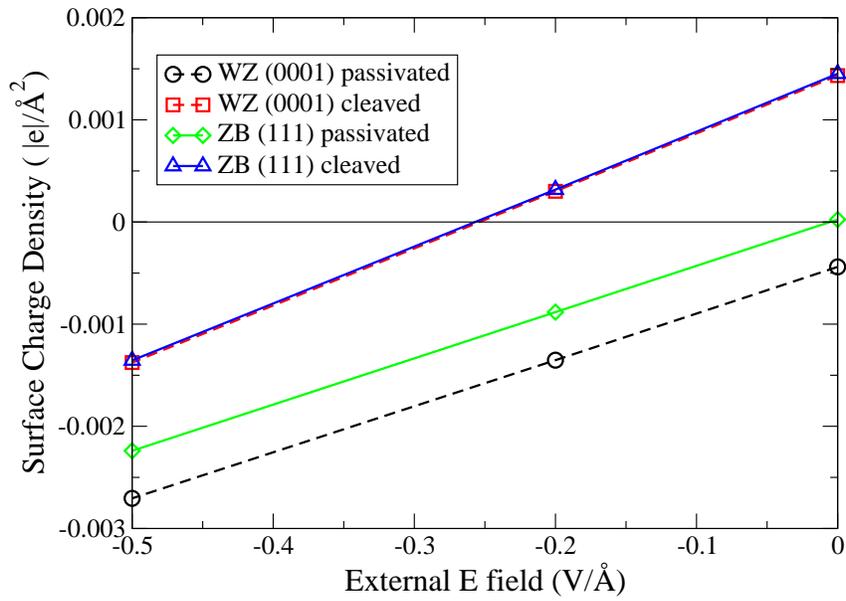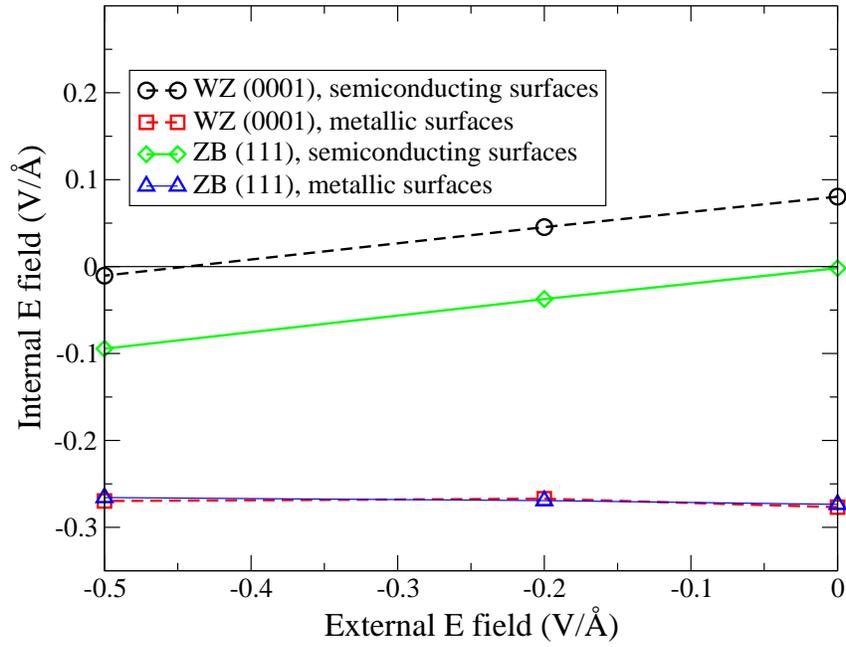
Figure 4.6: Comparison of the internal E field (upper) and surface charge density (lower) between the two slabs shown in fig. 4.5.

Slabs with two metallic surfaces

For the GaN slab with metallic surfaces, the internal E field does not change significantly with the external E field. This internal E field is related to the transferring of the free charge between the two metallic surfaces. Further tests show that it mainly depends on the thickness of the slab and the properties of the specific surface structures. The following equation describes the behavior,

$$E_{\text{internal}} = \frac{\Delta V}{L}, \tag{4.8}$$

where L is the thickness of the slab. The potential difference ΔV is the property of the two surfaces. Because both surfaces are metallic, if they have different abilities to attract electrons, electrons will transfer between the two, generating a potential difference ΔV across the slab. The transfer will stop when the energy cost to overcome ΔV offsets the gain due to the transfer. The value of E_{internal} for both wurtzite and zinc blende slab is -0.27 V/Å, despite of their different bulk structures. This shows that ΔV is mostly determined by the surfaces being exposed. Our further tests show that the surface energies and structures converge very fast with thickness. Because in thicker slabs less charge transfer is needed to generate the same ΔV, hence there is less disruption to the electronic structures of the surface.

The surface charge is the sum of free charge and bound charge,

$$\sigma = \sigma_0 + \sigma'. \tag{4.9}$$

The free charge is the divergence of the electric displacement,

$$\sigma_0 = \nabla \cdot \mathbf{D}. \tag{4.10}$$

The bound charge is the negative of the divergence of the polarization,

$$\sigma' = -\nabla \cdot \mathbf{P}. \tag{4.11}$$

For slabs with metallic surfaces, the free charge responds to the external E field, keeping ΔV constant.

Slabs with two semiconducting surfaces

For the GaN slabs with semiconducting surfaces (i.e. passivated), if there is no free charge in the system, the bound charge equals the total charge. Therefore, we have

$$\sigma' = \sigma = \epsilon_0(E_{\text{external}} - E_{\text{internal}}), \tag{4.12}$$

And
$$\sigma' = \mathbf{P} \cdot \hat{n} = \mathbf{P}_0 \cdot \hat{n} + (\epsilon - 1)\epsilon_0 E_{\text{internal}}. \tag{4.13}$$

Therefore, the internal E field approximately follows the equation

$$E_{\text{internal}} = \frac{E_{\text{external}}}{\epsilon} - \frac{\mathbf{P}_0 \cdot \hat{n}}{\epsilon\epsilon_0}, \tag{4.14}$$

In these calculations, the dielectric constant ϵ is the high frequency one (ϵ_∞) since all of the ions are fixed, where $\hat{n}$ is the normal direction of the surface. The quantity $\mathbf{P}_0$ is the spontaneous polarization under the $E = 0$ condition.

Under zero external E field, the Zinc Blende (111) passivated slab has zero internal E field, because it has zero spontaneous polarization. However, wurtzite GaN has finite polarization along the (0001) direction. Our bulk calculation yields $\mathbf{P}_0 = -0.038 C/m^2$ and $\epsilon_\infty = 5.45$ (Table 4.1). According to Eq. 4.14, these gives $E_{\text{internal}} = 0.08$ V/Å, exactly the same as we observed in fig. 4.6.

For the semiconducting slab, the potential drop across the slab is caused by the polarization and the external E field. When the drop is greater than the band gap, breakdown happens. Electrons transfer from the VBM of one surface to the CBM of another, lowering the system's total energy. In this case, the two surfaces become metallic. The accumulated surface charge now contains the free charge and the bound charge. The potential drop across the slab is pinned to the band gap. The semiconducting slab then behaves like a metallic one after breakdown.

Discussion

In the DFT studies of a specific surface, the slab usually has a passivated (semiconducting) bottom surface, and a metallic or semiconducting top surface. If both surfaces are semiconducting, the above mentioned discussions apply. If one surface is semiconducting and another one is metallic, further calculations indicate that its behavior is closer to the semiconducting slab scenario, if the breakdown does not happen. This is the most favored situation, because transfer of free charge between the two surfaces causes slow convergence in DFT calculations. It also causes the total energy to converge much slower with respect to the thickness of the slab.

4.1.7   Surface Energy and Chemical Potential

It is easy to compute surface energy only in the case where the slab is terminated by two identical surfaces. This is natural only in special cases such

as wurtzite ($10\bar{1}0$) and zinc blende (110). For the simple polar surfaces, i.e., wurtzite (0001)/($000\bar{1}$), zinc blende (111)/($\bar{1}\bar{1}\bar{1}$), the natural cleavage yields two "conjugate" surfaces, one cation-terminated and the other anion-terminated. The surface energy for each one of the two conjugate surfaces can not be calculated through the slab model. Only the cleavage energy, which is the total energy of the two conjugate surfaces, can be calculated this way. The energy for zinc blende polar surfaces are uniquely defined because of its symmetry [62]. In fact, by constructing wedge-shaped geometries, the energy of the zinc blende polar surfaces, i.e., (111) and ($\bar{1}\bar{1}\bar{1}$), can be calculated [63]. The wurtzite polar surfaces, i.e., (0001) and ($000\bar{1}$), do not have absolute energies. Adding a function,

$$C_0 \cos(\theta), \tag{4.15}$$

to the energy of all surfaces at the same time, does not affect physically observable properties [62]. Here θ is the angle between the surface and the c direction. C_0 is an arbitrary constant.

For the slab model, the sum of the energies for the two exposed surfaces is well defined,

$$E_{\text{surf1}} + E_{\text{surf2}} = E_{\text{total}} - \sum_i N_i \mu_i, \tag{4.16}$$

where E_{surf1} and E_{surf2} are surface energies, E_{total} is the total energy of the slab, N_i and μ_i are the number and chemical potential for each species, respectively. Therefore, the absolute energy differences can be calculated for different reconstructions on one surface, as long as another surface is kept unchanged during the comparison.

The chemical potential is the energy to add or decrease one atom from the system, assuming there is a reservoir for each species to equilibrate with. If two surface reconstructions on the same surface have different number of atoms, their relative energy difference depends on the chemical potential of those species. The chemical potentials must follow some constraints. For example, the chemical potential of Ga is bounded above by the chemical potential of Ga metal,

$$\mu(\text{Ga}) \leq E(\text{Ga metal}). \tag{4.17}$$

The chemical potential of N is bounded above by the chemical potential of the N_2 molecule,

$$\mu(\text{N}) \leq \frac{1}{2}E(\text{N}_2). \tag{4.18}$$

Also, if the bulk of the slab is GaN, the Ga and N atoms must be in equilibrium with it,

$$\mu(\text{Ga}) + \mu(\text{N}) = E(\text{GaN}). \tag{4.19}$$

44

These conditions lead to the range of the Ga chemical potential to be

$$E(\text{GaN}) - \frac{1}{2}E(\text{N}_2) \leq \mu(\text{Ga}) \leq E(\text{Ga metal}). \tag{4.20}$$

The range for the N chemical potential is

$$E(\text{GaN}) - E(\text{Ga metal}) \leq \mu(\text{N}) \leq \frac{1}{2}E(\text{N}_2). \tag{4.21}$$

In our following studies, we also introduce O atoms into the system. Besides the constraint on oxygen chemical potential from the O_2 molecule,

$$\mu(\text{O}) \leq \frac{1}{2}E(\text{O}_2), \tag{4.22}$$

we must also considered possible compounds like Ga_2O_3 as the boundary,

$$2\mu(\text{Ga}) + 3\mu(\text{O}) \leq E(\text{Ga}_2\text{O}_3). \tag{4.23}$$

These two conditions combined, lead to

$$\mu(\text{O}) \leq \frac{1}{2}E(\text{Ga}_2\text{O}_3) - \frac{3}{2}\mu(\text{Ga}). \tag{4.24}$$

There is no lower bound for the O chemical potential; $\mu(\text{O}) = -\infty$ means no O presence.

Eq. 4.19, Eq. 4.20, Eq. 4.21 and Eq. 4.24 are the conditions we considered in our following studies of the GaN surface reconstructions with Ga, N and O ad-atoms.

4.2   Reconstructions

4.2.1   Background

According to the X-ray photoelectron spectroscopy measurement, the surface composition of the alloy is very different from the bulk [2]. The Zn content is reduced while O content enhanced (see Table. 4.2). The Zn atom is volatile under the nitridation condition. The presence of O atoms on GaN surfaces has been studied for the $(10\bar{1}0)$ surface. The high binding energy is caused by the strong Ga-O bond and the electron counting rule [64].

From the transmission electron microscopy measurements, the $(10\bar{1}1)$ and $(10\bar{1}\bar{1})$ surfaces dominate the sample. (Communication from Dr. James Ciston at BNL) These semi-polar surfaces have many equivalents such as $(1\bar{1}01)$,

Table 4.2: The measurements of bulk and surface composition from Ref. [2]. In the experiment, the nitridation process is to improve the crystal quality. This process also decreases the ZnO composition in the bulk.

nitridation	bulk Zn/Ga	surface atomic ratio		
time (hours)	(O/N) ratio	Zn/Ga	O/Ga	N/Ga
5	0.28	0.19	0.43	0.54
10	0.17	0.18	0.38	0.44
15	0.13	0.11	0.41	0.59
20	0.10	0.05	0.51	0.53
30	0.05	0.02	0.56	0.48

($\bar{1}101$), etc. The alloy particle tends to form a polyhedron with three to six ($10\bar{1}1$)-equivalent facets covering half of the particle and three to six ($10\bar{1}\bar{1}$)-equivalent facets covering another half.

Based on this information, we focus on the reconstructions of these two surfaces with only Ga, N and O ad-atoms, using wurtzite GaN as the bulk instead of the alloy. As an ongoing investigation, I only present our results on the ($10\bar{1}1$) surface. These results are based on the calculations from our collaborators Prof. Artem R. Oganov and Qiang Zhu.

4.2.2 Method

The slab model for the GaN ($10\bar{1}1$) surface simulation is shown in fig. 4.7. Our DFT calculation uses the same parameters as our bulk alloy calculations in the previous chapter, except that the plane wave cutoff is 400 eV and the k point mesh is $8 \times 4 \times 1$ for the surface unit cell. We use the dipole correction on the electrostatic potentials to partly cancel the interactions between the slab and its images. The bottom surface is passivated by fractional hydrogens. The passivated surface is semiconducting, with no surface states in the fundamental band gap of GaN. For trial calculations, the slab contains 5 layers of GaN, with the top 2 layers being relaxed. To obtain accurate surface energies, we use a slab containing 10 layers with top 5 layers being relaxed. Our testing calculations suggest that the surface energy is converged to better than 0.05 eV per surface cell. Our surface calculations are restricted to within the 2×2 surface cell.

We employ an evolutionary algorithm to explore the surface structures, with the collaboration of Prof. Artem R. Oganov and his student Qiang Zhu. The USPEX package developed by our collaborators has been successfully applied to various bulk materials [65, 66]. It can find low energy structures

Vacuum

(10$\bar{1}$1) surface

N

Ga

H, Z=3/4

Vacuum   H, Z=5/4
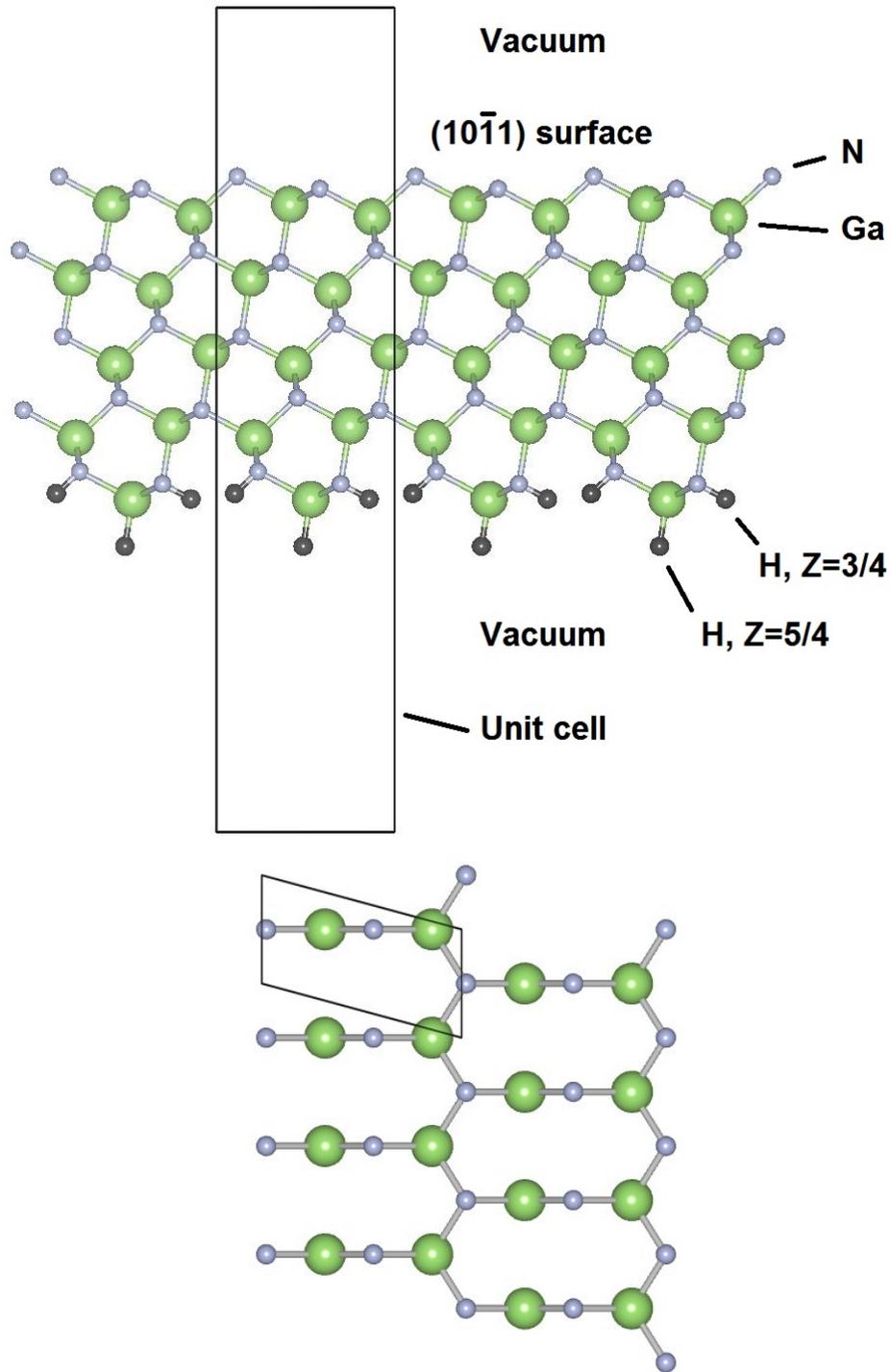
Unit cell

Figure 4.7: (Upper) The slab model used for the GaN (10$\bar{1}$1) surface. The bottom (10$\overline{11}$) surface is passivated by fractional hydrogens. (Lower) Top view of the simple cleaved surface, showing the primitive 1×1 unit cell.

efficiently. The evolutionary algorithm initially generates random structures as the first generations (e.g. 20 structures). These initial structures are relaxed by DFT. They are then ranked by a fitness function, based on the energy. Structures with better fitness are more likely to be selected to generate new structures (off-springs). For our surface searches, we considered two ways to produce off-springs. (1) Heredity. Two structures are chosen from the previous generation. They are randomly sliced at the same place. Then pieces from each of the two structures are combined to generate the off-spring. (2) Mutation. One structure is chosen. The atom coordinates are randomly moved by finite amounts, or one atom is randomly added, or one atom is deleted. These off-springs, combined with a few best structures from the previous generation, produces the next generation. This process is repeated until no new low energy structures can be found.

The calculations for searching stable reconstructions are conducted by our collaborators Prof. Artem R. Oganov and Qiang Zhu. For our surface searches, the fitness function is defined as the energy of the structure relative to the convex hull (see fig. 4.8). From the convex hull, we can build the phase diagram.

4.2.3   GaN ($10\bar{1}1$)

For the GaN ($10\bar{1}1$) surface, the evolutionary algorithm explores approximately 500 structures (see fig. 4.8). The stable structures found during our search are shown in fig. 4.9. Compared to the cleaved surface (fig. 4.7), structure (a) has two Ga adlayers. Structure (b) has one Ga adlayer. Structure (c) has the top N and half of the second N removed. Structure (d) has only the top N removed. Structure (e) has an additional N at the bridging position of the two top N atoms. The first 4 structures have been found by Akiyama *et al* [67]. Structure (e), however, is non-intuitive. This demonstrates the power of the automated searching by the evolutionary algorithm. All of the stable structures are simple to construct. Compared to the cleaved surface, the stable structures can be constructed through either vacancies or ad-atoms. No complicated reconstructions have been found. For the Ga rich condition, there are many intermediate structures between structure (a) and (b), lying very close to the convex hull (see fig. 4.8). These structures can appear under finite temperatures. For the N rich condition, however, most structures are much higher in energy than the convex hull. None of the stable reconstructions satisfies the electron counting rule (EC). Large reconstructions may satisfy the EC rule. But the charge separation at long distance causes a penalty through the Coulomb interaction, making large reconstructions less stable compared to small reconstructions. From the convex hull, we construct the phase diagram (fig. 4.10).
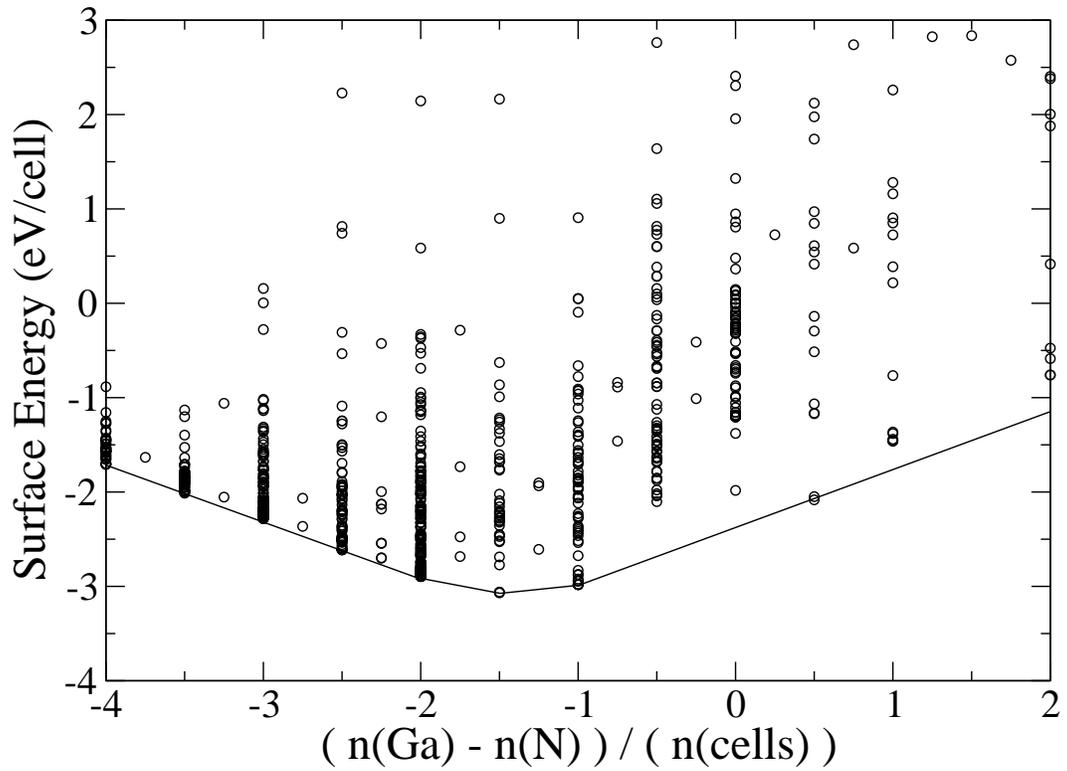
Figure 4.8: Energies of structures explored by the evolutionary algorithm. The vertices of the convex hull are the stable structures appearing in the phase diagram. The slope of each section of the convex hull is where stable structures can coexist. The zeros of the horizontal and vertical axes correspond to the simple cleaved surface with relaxation. The chemical potentials are arbitrarily chosen so the graph appears symmetrical.

(a)

(b)

(c)

(d)

(e)
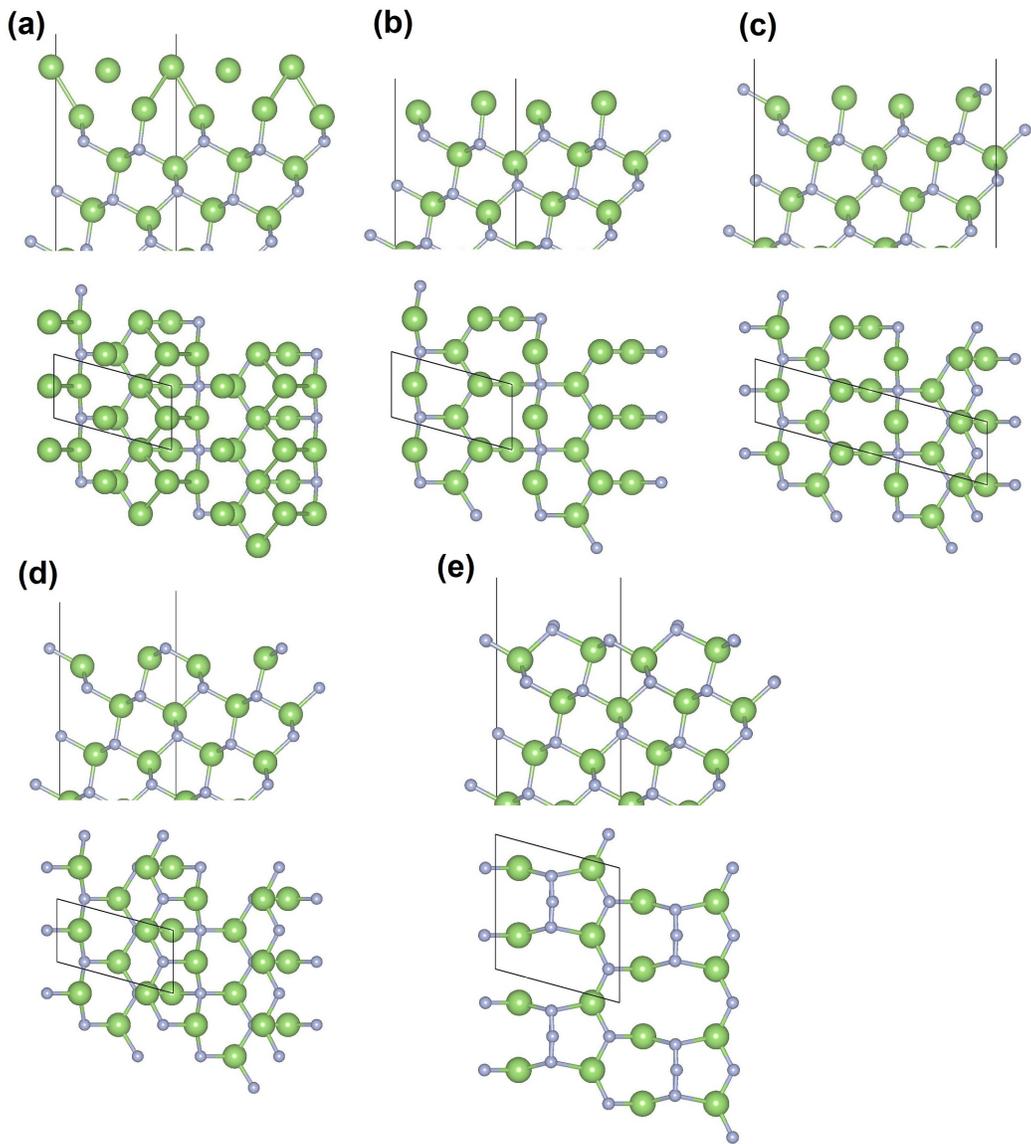
Figure 4.9: Side view and top view of the stable reconstructions. Large balls, gallium; small balls, nitrogen.
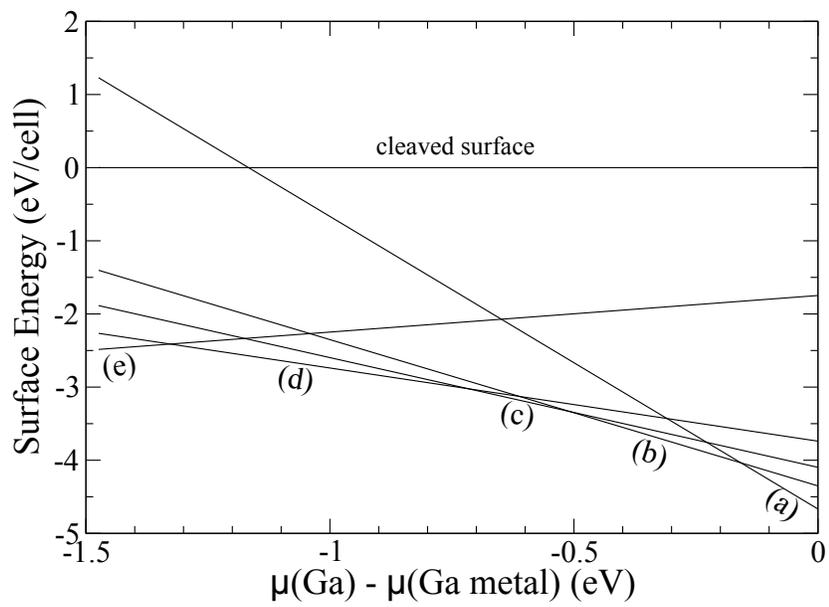
Figure 4.10: Zero temperature phase stability diagram of the GaN $(10\bar{1}1)$ surface. Each stable (minimum energy) section (i.e. (a), (b), ...) corresponds to a stable structure shown in fig. 4.9.

4.2.4 GaN ($10\bar{1}1$) with Oxygen

We employ the same method to search for reconstructions with O atoms. The oxygen chemical potential is a new degree of freedom. Therefore, the phase diagram is two-dimensional (fig. 4.11). The two major reconstructions are structure (f) and (g). Compared to the cleaved surface, structure (f) has half of the top N removed, half of the top N and all of the second N replaced by O. Structure (g) has the top two N replaced by O. Similar reconstructions for the ($10\bar{1}0$) surface have been reported [64]. These reconstructions are favored because of the strong Ga-O bond and the electron counting rule. For example, the simple cleaved surface has 3 nitrogen dangling bonds per surface cell. Because each N dangling bond needs 3/4 electron to become filled, the surface needs 9/4 electrons to satisfy the electron counting rule. For structure (g), substituting 2 N with 2 O introduces 2 extra electrons. Therefore, it only needs 1/4 electron to satisfy the electron counting rule.

More properties of this surface are still under investigation. Previous study proposed a possible reaction mechanism for the water oxidation process on the GaN ($10\bar{1}0$) surface [68]. Given these reconstructions, we are interested to test if this mechanism works for the ($10\bar{1}1$) surface.
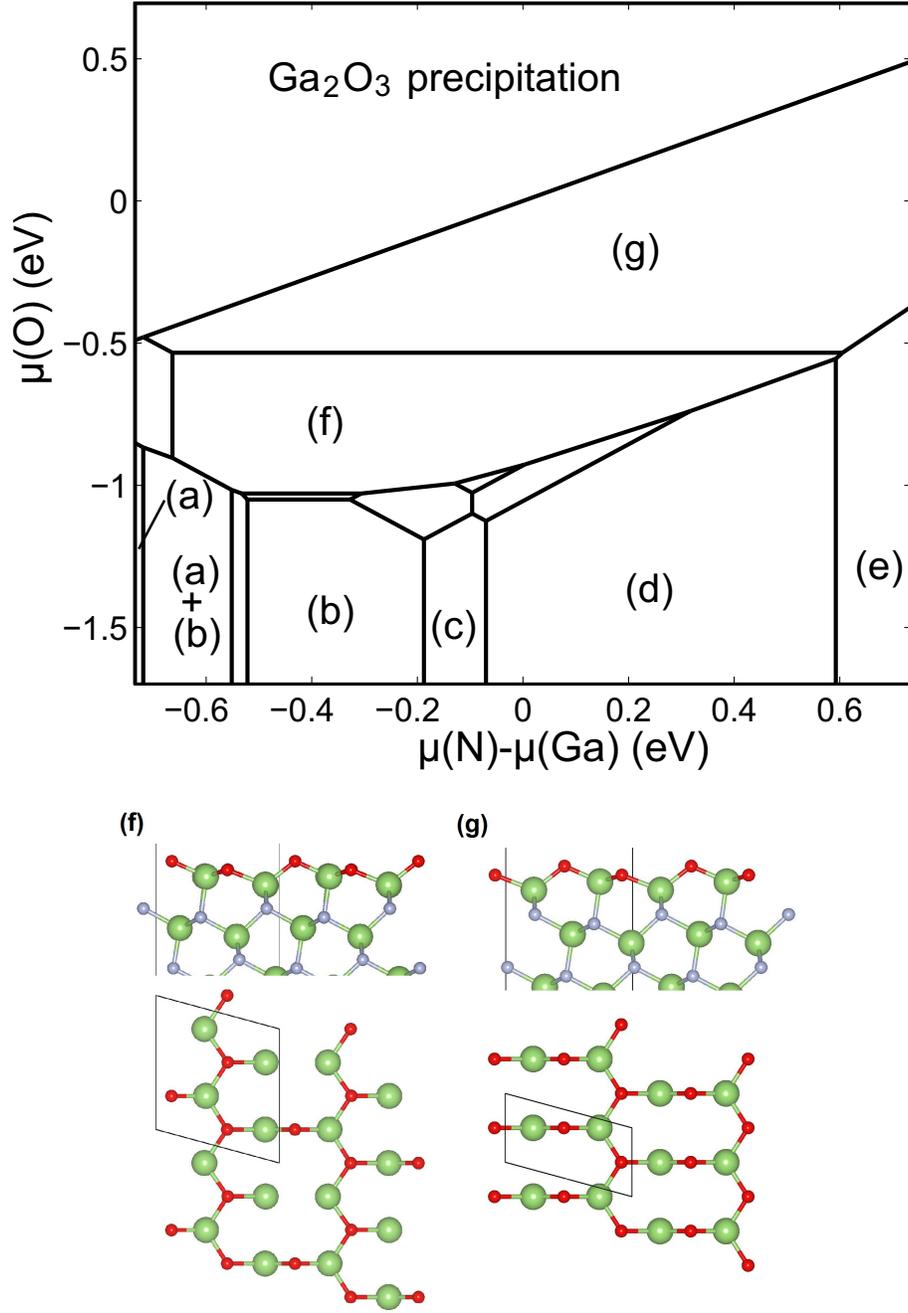
Figure 4.11: (Upper) Phase diagram of the GaN ($10\bar{1}1$) surface with oxygen present. Important reconstructions are labelled from (a) to (g). Structures (a) to (e) are the reconstructions shown in fig. 4.9. (Lower) Side view and top view of structure (f) and (g). Large balls, gallium; small gray balls, nitrogen; small red balls, oxygen.

Chapter 5

A Toy Model for Charge Transfer

5.1 Introduction

When two atoms or molecules collide with each other, electrons might be transfered from one to another. Sometimes the electron transfer means the electronic state changes from one potential energy surface (PES) to another, as shown in fig.5.1. If initially the electron comes from the nucleus with higher binding energy, after the collision, there should be some probability for the electron to be transfered to another nucleus, as long as the kinetic energy permits. In this situation, non-adiabatic effects play an important role. Within the adiabatic approximation, the system will always stay on one PES, no electron transfer will happen.

The whole problem is too complicated to solve exactly. There are quite a few approximation methods proposed, which are computationally affordable. The key points are how to separate the system into a classical part and a quantum part, and how these two subsystems interact with each other. Usually, the actions of the classical nuclei on the electrons are simple, while how the electrons act on the nuclei are not clear. In this project, two methods are investigated. In the *Ehrenfest* method, the nuclei evolve under an single effective potential. In the *Surface Hopping* method, a stochastic method is used to describe the nuclear motion.

To compare these approximate methods, we proposed a very simple model, and treat the nuclei by exact quantum mechanics. Then the problem becomes solving a time-independent Schrodinger equation. The scattering state we get will give the electron transfer probability. We can compare the results with various approximate methods.
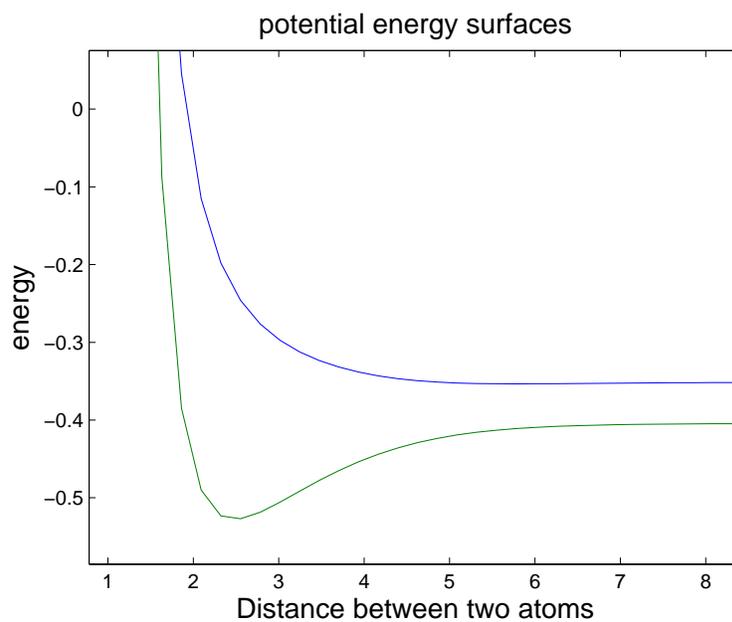
Figure 5.1: The schematic potential energy surfaces of the collision process. The lower curve is the bound state. The upper curve is the antibonding state. The two nuclei have slightly different binding energies. At large separation, the adiabatic states are localized atomic orbitals. Thus, electron transfer means change of the potential energy surface.
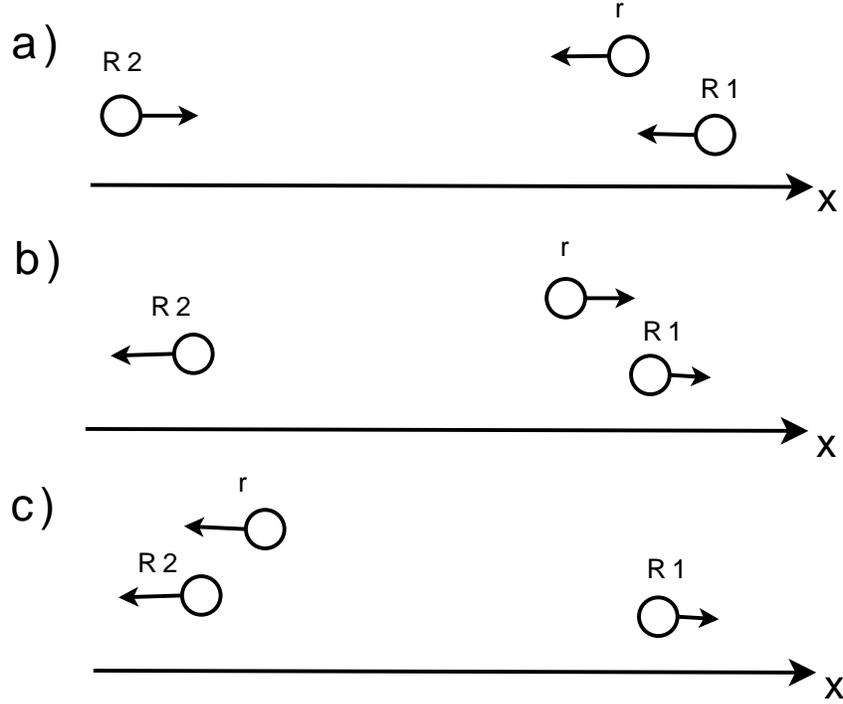
Figure 5.2: Schematic illustration of the collision process. (a) For the initial state, an electron comes with the nucleus from the right. (b) For one component of the final state, the electron is reflected. (c) For the other component of the final state, the electron is transferred to the other nucleus.

5.2   Description of the Model

This one-dimensional toy model has an electron transfer probability during a collision, as shown in the fig. 5.2. In this model, there are two inequivalent nuclei. The spin is completely ignored. Initially, the electron is coming with one nucleus; after the collision, there is a possibility for the electron to be transfered to the other nucleus. Assuming the potentials between the electron and nuclei are V_1 and V_2 , and the potential between the two nuclei is U , the Hamiltonian can be written as :

$$H = \frac{p^2}{2m} + \frac{P_1^2}{2M_1} + \frac{P_2^2}{2M_2} + V_1(r - R_1) + V_2(r - R_2) + U(R_1 - R_2). \quad (5.1)$$

The nuclear repulsion U is chosen large enough at short range that both nuclei reverse their velocities after colliding. Their probability to tunnel through each other is very small. The potentials V_1 and V_2 are chosen so that only one bound

56

state is relevant. Their bound states can be written as ϕ_1 and ϕ_2:

$$\left(\frac{p^2}{2m} + V_1(r - R_1)\right)\phi_1(r - R_1) = \epsilon_1\phi(r - R_1), \qquad (5.2)$$

$$\left(\frac{p^2}{2m} + V_2(r - R_2)\right)\phi_2(r - R_2) = \epsilon_2\phi(r - R_2). \qquad (5.3)$$

To simplify the problem, the initial kinetic energy of the system is small enough, so that electron must bind to one or the other nucleus. That is, there is not enough kinetic energy to ionize the system. Assuming that the electron comes with the left nucleus, after the collision, there are two possibilities: the electron can either go to the right, or go to the left, if the kinetic energy permits. There are a few parameters we can change: the initial kinetic energy, the repulsion distance between the two nuclei, and the binding potential of the electron.

5.3 Computational Methods

5.3.1 Full Quantum Mechanical Method

We only want to calculate the scattering eigenstate for this Hamiltonian. The scattering state should have only one incoming component and two outgoing components. The asymptotic form for the scattering solution of the Hamiltonian(Eq. 5.1) at $R_1 - R_2 \to \infty$ is :

$$\psi = \psi_1 + r(E) \cdot \psi_2 + t(E) \cdot \psi_3. \qquad (5.4)$$

The asymptotic form $(R_1 - R_2 \to \infty)$ for the incoming and outgoing components are

$$\psi_1 = \phi_1(r - R_1)\ \exp\left[-ik\frac{mr + M_1 R_1}{m + M_1}\right]\exp\left[ikR_2\right], \qquad (5.5)$$

$$\psi_2 = \phi_1(r - R_1)\ \exp\left[ik\frac{mr + M_1 R_1}{m + M_1}\right]\exp\left[-ikR_2\right], \qquad (5.6)$$

$$\psi_3 = \phi_2(r - R_2)\ \exp\left[ik'R_1\right]\exp\left[-ik'\frac{mr + M_2 R_2}{m + M_2}\right]. \qquad (5.7)$$

Here

$$\hbar k = \sqrt{(E - \epsilon_1) \Big/ \left(\frac{1}{2(m + M_1)} + \frac{1}{2M_2} \right)}, \qquad (5.8)$$

$$\hbar k' = \sqrt{(E - \epsilon_2) \Big/ \left(\frac{1}{2(m + M_2)} + \frac{1}{2M_1} \right)}. \qquad (5.9)$$

The coefficients $r(E)$ and $t(E)$ are for reflection and transmission. According to the definition of the probability flux (note: this is the definition of single particle probability flux, in our case, there are three particles)

$$j = \frac{\hbar}{2mi} \left(\psi^* \Delta \psi - \psi \Delta \psi^* \right). \qquad (5.10)$$

The reflection and transmisson probabilities $R(E)$ and $T(E)$, and conservation of probability flux relation, are

$$R(E) = |r(E)|^2, \qquad (5.11)$$

$$T(E) = |t(E)|^2 \cdot \frac{k'}{k}, \qquad (5.12)$$

$$R(E) + T(E) = 1, \qquad (5.13)$$

respectively.

5.3.2   Ehrenfest Method

We prefer to call this method "the average classical trajectory method." We assume that a single Newtonian trajectory $R(t)$ describes the nuclei. The classical nuclei move on a single effective potential. The effective potential is a self-consistently weighted average of different adiabatic potential energy surfaces. This approach has limitations, especially when the adiabatic potential surfaces are very different. Then the nuclear motion can never be described by a single trajectory. But still it might provide some useful results.

Here is a simplified derivation. The classical nuclear coordinate is R. The total Hamiltonian is

$$H = T_N + H_0(r, R). \qquad (5.14)$$

T_N is the kinetic energy of nuclei. H_0 contains all the electronic part H_{el} and the nuclear potential energy $U_N(R)$. The nuclei evolve as a single trajectory $R(t)$ on an effective potential H_0.

First we construct the electronic wave function. Suppose we have an adiabatic basis $\{\psi_j(r; R)\}$, and H_0 is diagonal on this basis. We assume the existence of an electronic wavefunction $\psi(r; R)$. In full generality, it can be expanded as

$$\psi(r, R(t), t) = \sum c_i(t) \cdot \psi_j(r; R(t)). \tag{5.15}$$

Put it into $i\hbar \frac{\partial}{\partial t}\psi = H_0\psi$, and project onto ψ_k. We get

$$i\hbar \, \dot{c}_k + i\hbar \sum_j c_j \left\langle \psi_k \left| \frac{\partial}{\partial t} \right| \psi_j \right\rangle = \sum_j c_j \left\langle \psi_k \left| H_0 \right| \psi_j \right\rangle. \tag{5.16}$$

After making the substitutions:

$$\left\langle \psi_k \left| \frac{\partial}{\partial t} \right| \psi_j \right\rangle = \left\langle \psi_k \left| \frac{\partial}{\partial R} \right| \psi_j \right\rangle \cdot \dot{R} = d_{kj} \cdot \dot{R} \tag{5.17}$$

$$\left\langle \psi_k \left| H_0 \right| \psi_j \right\rangle = V_{kj}, \tag{5.18}$$

we get

$$i\hbar \, \dot{c}_k = \sum_j \left(V_{kj} - i\hbar\dot{R} \cdot d_{kj} \right) c_j. \tag{5.19}$$

Note that in the adiabatic representation $\{\psi_j(r; R)\}$, the matrix element $V_{kj}(R)$ is diagonal,

$$V_{kj} = \delta_{kj}E_j, \tag{5.20}$$

where E_j is the PES for the ground state ($j = 1$) and the excited states ($j > 1$). The non-adiabatic coupling $d_{kj}(R)$ is anti-symmetric,

$$d_{kj}(R) = -d_{jk}(R). \tag{5.21}$$

The "Coupling" matrix d_{kj} plays the role of the leading term in the non-adiabatic coupling. The second order term, which involves the second derivative of R, does not appear here. Usually this is enough.

The trajectory $R(t)$ is still unknown. To calculate the nuclear trajectory, one way is to use the weighted "effective" potential:

$$\overline{V}(R) = \left\langle \psi_{el} \left| H_0 \right| \psi_{el} \right\rangle. \tag{5.22}$$

In the adiabatic representation,

$$\overline{V}(R) = \sum_j |c_j|^2 \cdot E_j(R). \tag{5.23}$$

Then, Newtonian evolution is used,

$$\ddot{R} = \frac{1}{M_{\text{eff}}} \left(-\frac{\partial \overline{V}}{\partial R} \right). \tag{5.24}$$

Equation 5.19 and 5.24 can be integrated together, giving the time dependence of the probabilities $|c_j(t)|^2$ on different surfaces.

5.3.3 Surface Hopping Method

The surface hopping method has a few variants, here we only considered the *Fewest Switches* Algorithm proposed by Tully *et al.* [69] The argument is that, it is too expensive to treat the nuclei quantum mechanically, but it is over simplified to just considered a single effective potential. Since the dynamics of the nuclei on different potential energy surfaces may be very different, they need to be treated accordingly. The energy and momentum for each branch must be conserved separately, not just the total energy and momentum.

The surface hopping method achieves this requirement by considering a swarm of trajectories instead of only one trajectory. In this method, the trajectory evolves on only one PES at any particular time, interrupted by random switches to another PES. It is essentially a stochastic method. The nuclear wave function is replaced by a distribution of trajectories; the phase of the nuclear wave function is completely ignored. This might cause some problem when the coherence of nuclei is important. The surface hopping method should provide better results than the Ehrenfest method. But, it is very expensive, since a large number of trajectories is calculated instead of only one; and turns out not to be always better.

For each trajectory, the electronic wave functions are treated the same way as in the Ehrenfest method. But the nuclear motions are governed by only one PES at a each time. For each time step Δt, the trajectory has a certain probability g_{kj} to hop from its current PES k to another PES j. The g_{kj} must ensure that, for a large number of trajectories, the state population of nuclei

approximates the correct one. For the fewest switches algorithm,

$$g_{kj} = \frac{\Delta t b_{jk}}{a_{kk}}, \tag{5.25}$$

$$\text{where } b_{kj} = \frac{2}{\hbar} \text{Im} \left(a_{kj}^* \cdot V_{kj} \right) - 2\text{Re} \left(a_{kj}^* \dot{R} \cdot d_{kj} \right), \tag{5.26}$$

$$\text{and } a_{kj} = c_k \cdot c_j^*. \tag{5.27}$$

In the numerical calculation, at each time step, a random number generator is used to decide if a surface hopping will happen. And if the hopping happened, the kinetic energy of the nuclei is adjusted to conserve the energy and momentum. Thus, in this algorithm, for each branch, the energy and momentum are guaranteed to be conserved.

After running a lot of trajectories, the transfer probability is given by counting how many trajectories are on the excited potential energy surface.

5.4 Numerical Results and Discussion

5.4.1 the Approximate Adiabatic Representation

To proceed, we must first calculate the PES and relevant non-adiabatic coupling term. The ideal way is to solve the 1d Schrodinger equation exactly. However, for large nuclear separation, it is hard to solve exactly. Also, it is even harder to calculate the non-adiabatic coupling term, because it involves the derivative of the nuclear coordinates.

Thus, we contruct the approximate adiabatic eigenstates by the linear combination of the atomic eigenstates. In this project, the mass of the nuclei is chosen to be 30 electron masses. The binding potentials for the nuclei are $V_1 = -0.9e^{-|x|}$ and $V_2 = -1.0e^{-|x|}$. The binding energies are -0.3513, -0.4043a.u., respectively. The repulsion between the two nuclei is $\frac{1}{(R/R_0)^6}$. The calculated PES (E_j, Eq. 5.20) and non-adiabaic coupling term (d_{kj}, Eq. 5.17) are shown in fig. 5.3.

To evaluate the accuracy of this approximation, the exact eigenstates are calculated at $R = 2.0$a.u. for $H_{el} = T_{el} + V_1(r - R/2) + V_2(r + R/2)$ (the nuclear repulsion potential $U_N(R)$ is excluded). The exact eigenenergy for ground and excited states are -0.653 and -0.255, the approximate energies are -0.639 and -0.243, the error is about 2% and 5%, respectively. The projection $\langle \psi_{\text{exact}} | \psi_{\text{approxi}} \rangle = 0.99$. For most calculations in this project, the area of interest is $R > 2.0$a.u., so it is an acceptable approximation. The only problem is that, I don't know the accuracy of the non-adiabatic coupling term.
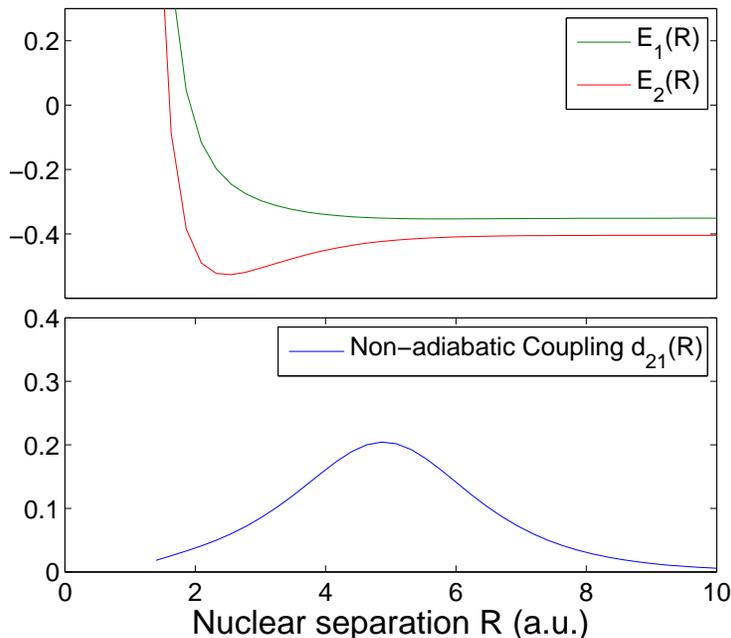
Figure 5.3: The calculated PES (E_j, Eq. 5.20) and non-adiabaic coupling term (d_{kj}, Eq. 5.17). The binding potential for the nuclei are $V_1 = -0.9e^{-|x|}$ and $V_2 = -1.0e^{-|x|}$. The binding energies are -0.3513, -0.4043a.u., respectively. The repulsion between these two nuclei is $1/(R/R_0)^6$, with the repulsion parameter $R_0 = 1.5$a.u.. the mass of the nuclei is chosen to be 30 electron masses. The coupling is the strongest around $R = 5a.u.$.

In fig. 5.3, the coupling term is small for very small nuclear separation, I don't know if this is due to the approximation or if it is the correct behavior.

For the Ehrenfest and surface hopping method, only the matrix elements and coupling terms are needed. Details of the electronic wave function are unimportant.

5.4.2   Full Quantum Mechanical Method

If two nuclei have different binding potentials, as long as the kinetic energy permits, the electron transfer is permitted, even if the final state has higher energy than the initial state. In the following, we study the case where the electron starts from the lower PES.

A typical exact scattering solution is shown in fig. 5.4. The exact solution correctly reproduced the plane-wave behavior of each component: electron
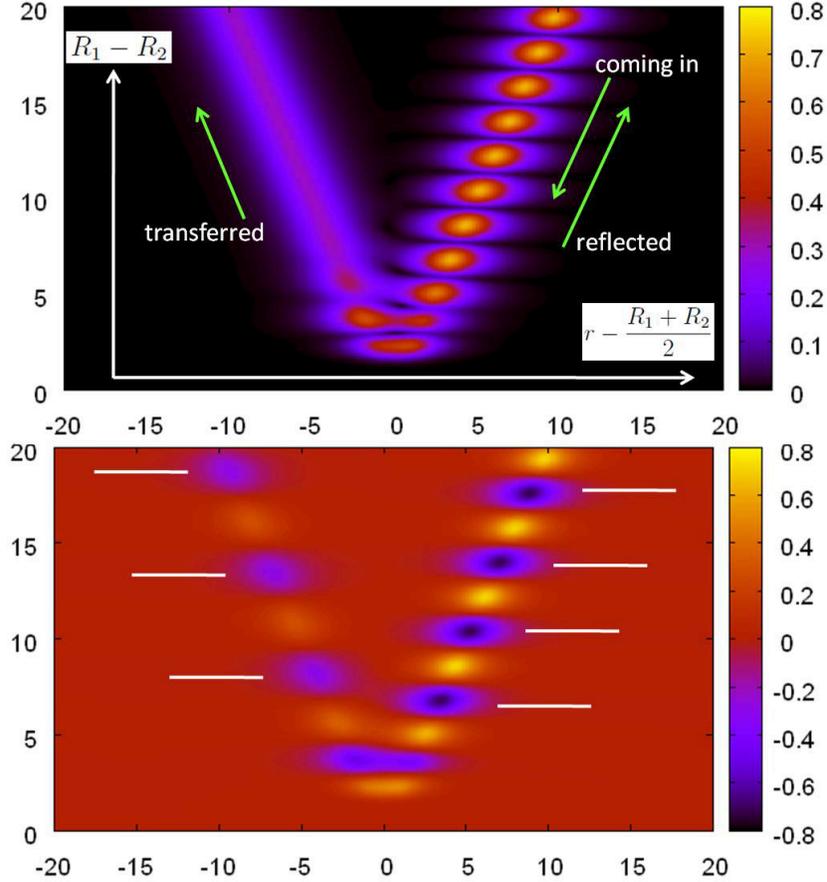
Figure 5.4: The absolute value and the real part of an exact solution. The y-axis is the distance between these two nuclei. The x-axis is the distance from the electron to the center of these two nuclei. On the right side, the incoming and out going components interfere with each other, causing the oscillation of the of absolute value. Because these components have different kinetic energies, the periods of the left and right sides are different (illustrated as white lines in the second graph). The parameters here are the same as fig. 5.3, except that the repulsion parameter $R_0 = 1.6a.u.$. The total initial kinetic energy T_0 is $0.10a.u.$, while the binding energies to these two nuclei are $-0.3513, -0.4043$a.u., respectively.
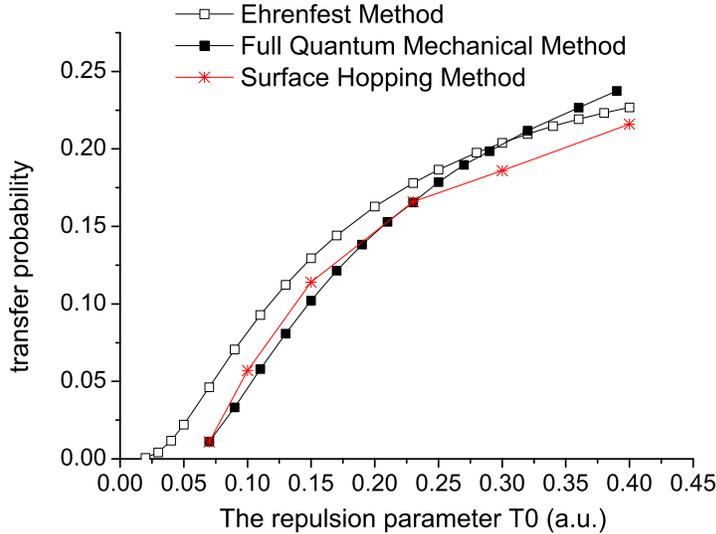
Figure 5.5: The electron transfer probability as a function of initial kinetic energy T_0(a.u.). The repulsion parameter for this graph is $R_0 = 4.0(a.u.)$. Other parameters are the same as fig. 5.3. Generally, both Ehrenfest and surface hopping methods agree with the exact solution. Note the cutoff around $T_0 = 0.5(a.u.)$, which is the gap between the binding energies of these two nuclei. The Ehrenfest method gives non-zero transfer probability below this energy, where the transfer is forbidden by energy conservation.

coming in, reflected, and transferred, as shown in Eq. 5.5 to Eq. 5.7. The electron transfer probabilities are calculated for different repulsion parameters(fig. 5.6), and different initial kinetic energies(fig. 5.5). These results are used to compare the various approximate method.

5.4.3 Ehrenfest Method

In the Ehrenfest Method, when integrating the equation, we can get the time dependence of the population distribution and the effective potential(fig. 5.7). Nuclei evolve on this single effective potential. The two physically distinct branches are replaced by a single trajectory with a single value of nuclear kinetic energy which is an average of the actual kinetic energies of the two physical branches.

The comparison with the full quantum mechanical solution is shown in fig. 5.5 and fig. 5.6. Generally, it agrees well. But for small kinetic energy, it gives non-zero transfer probability, even when the transfer is forbidden by
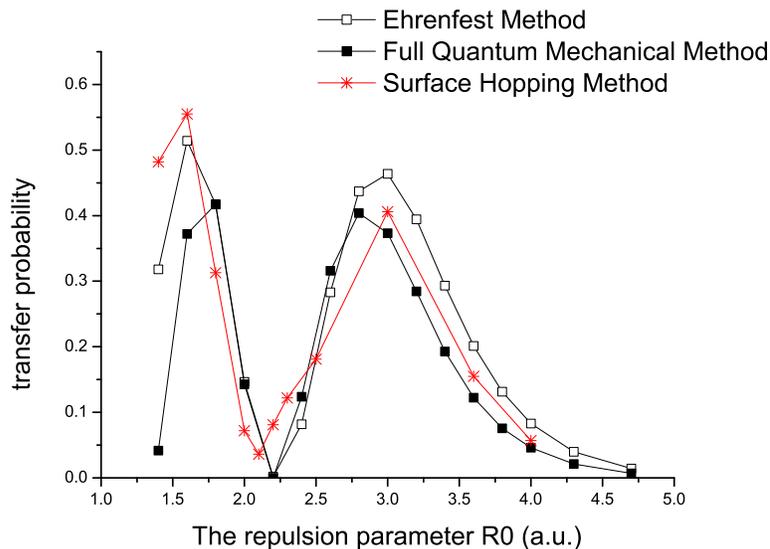
Figure 5.6: The electron transfer probability as a function of repulsion parameter R_0. The total initial kinetic energy for this graph is $T_0 = 0.10a.u.$. Other parameters are the same as fig. 5.3. For Ehrenfest and surface hopping method, it is hard to tell which one is better.

energy conservation. This is because it only conserves average energy and momentum.

The oscillation in fig. 5.6 is shown in detail in fig. 5.8. For smaller R_0, the system will spend more time in the strong coupling region, the population will oscillate between these two states in time.

5.4.4 Surface Hopping Method

The surface hopping method is very expensive. A large number of trajectories is needed to get a statistically significant results. The convergence is very slow, as shown in fig. 5.9. In this calculation, 10000 trajectories are calculated, giving the transfer probability to be 0.057 ± 0.003. Thus, if the number of trajectories is not large enough, the results are really not that accurate.

Consistency is also a problem for the surface hopping method. We use a swarm of trajectories to describe the nuclei, and a wave function to describe the electron. There are two ways to get the population distribution. One way is to calculate the ratio of trajectories that are on the upper PES. The other way is to average the probabilities on the upper PES from the electronic wave function for all trajectories. Consistency requires that the population

65

The effective potential
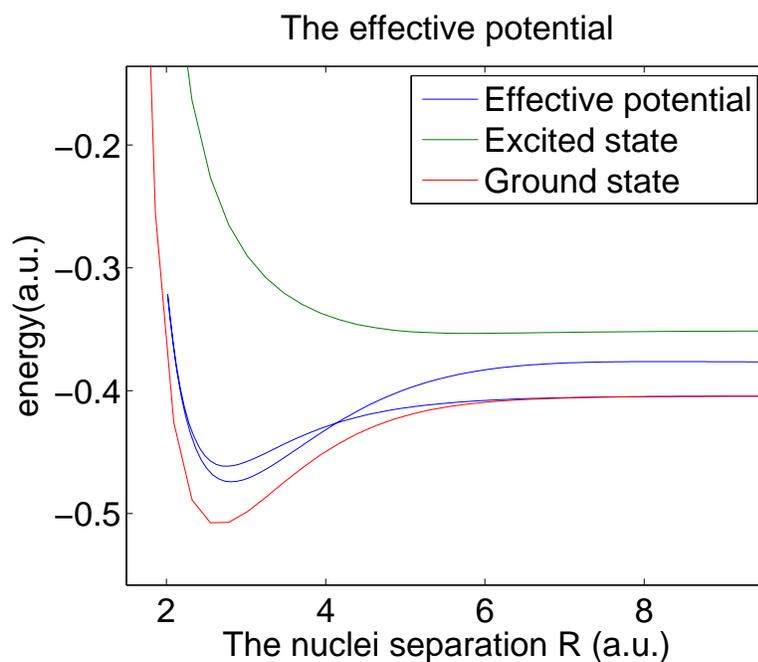
Figure 5.7: The effective potential relative to the ground and the excited PES. It is calculated self-consistently. The two physically distinct branches (the ground and the excited PES) are replaced by a single trajectory (the effective potential). The effective potential depends on the evolution of the electronic wave function. Therefore the inbound and the outbound parts do not overlap.
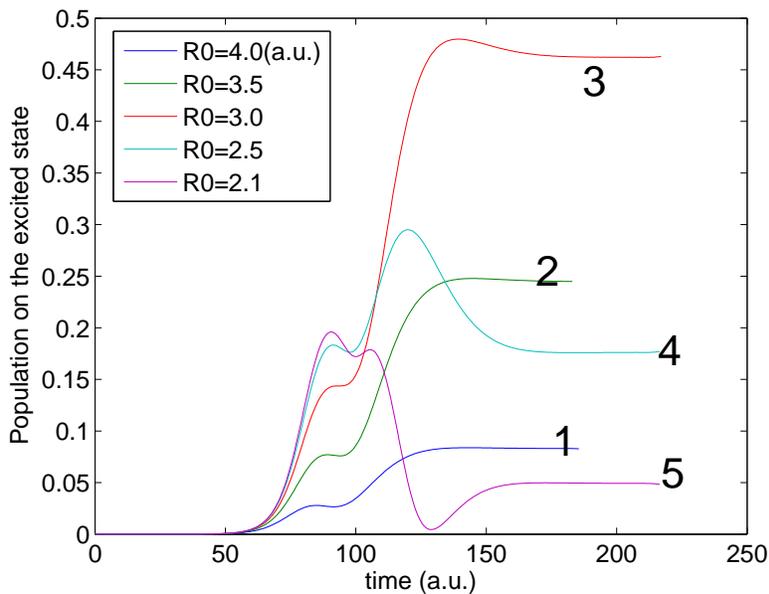
Figure 5.8: The time dependence of the population of the excited state. Numbers 1 to 5 correspond to decreasing R_0. The population oscillates between the ground and excited states.
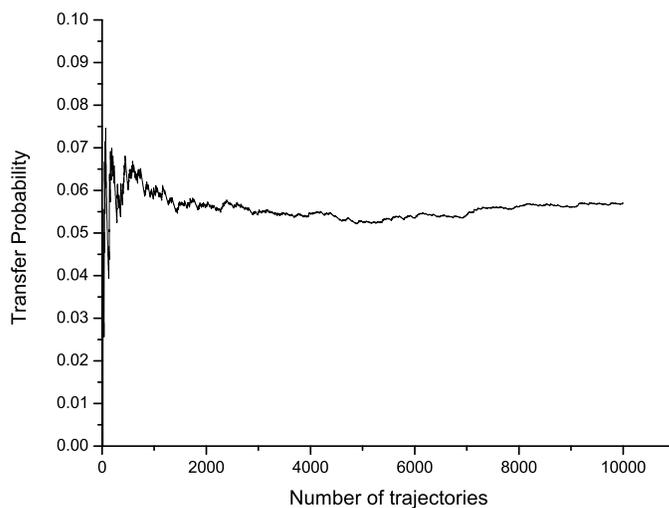
Figure 5.9: The transfer probability calculated from the collection of trajectories. KE= $0.10(a.u.)$, $R_0 = 4.0(a.u.)$. It converges slowly.
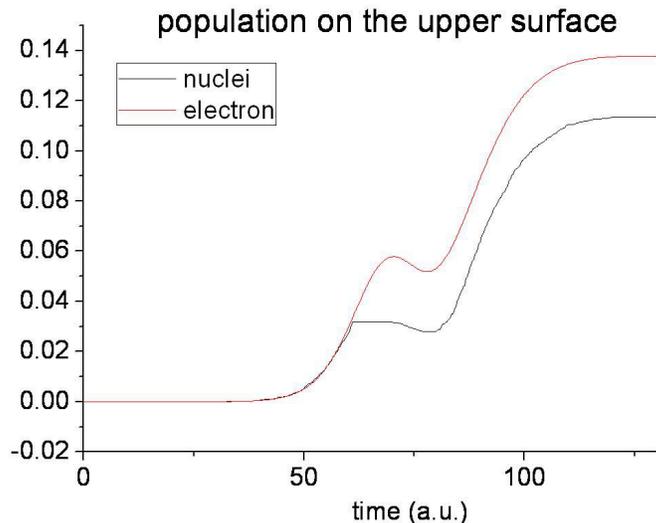
Figure 5.10: The state population of the trajectories and the electrons on the excited PES. ($KE = 0.10, R_0 = 1.5(a.u.)$). They are different, which causes the consistency problem.

distribution of the trajectory must equal the average state population of the electron. This is not accurately obeyed, as shown in fig. 5.10. Surprisingly, we found that the consistancy is not related to the accuracy of the results.

Actually, because of the "frustrated" hops (hopping forbidden by the energy conservation), the consistency is not guaranteed. Also, there are issues about the decoherence effect. Thus, it is really hard to tell when surface hopping method is better than the Ehrenfest method(see fig. 5.5 and fig. 5.6).

5.5 Conclusion

In this project, a simple toy model is investigated. The results from various approximate methods are compared with the exact solution. Generally, the surface hopping method guarantees the energy and momentum conservation. The Ehrenfest method is easy to calculate. In our model, for small kinetic energy and weak coupling, the surface hopping method gives much better results. But for general cases, we don't know which one is better.

5.6 Appendix: Full Quantum Mechanical Method

The general process is as follows: first, by setting up a coordinate transformation, this problem can be reduced to a 2D partial differential equation. With the known asymptotic form of the solution, proper boundary conditions can be constructed. By using a standard finite element method, I can calculate the solution of this equation, then extract the coefficients $r(E)$ and $t(E)$.

5.6.1 Coordinate Transformation

A coordinate transformation will separate the center of mass coordinate of the system, reducing the problem into a 2D equation.

$$x = \sqrt{\frac{2m(M_1 + M_2)}{m + M_1 + M_2}} \left(r - \frac{M_1 R_1 + M_2 R_2}{M_1 + M_2} \right), \tag{5.28}$$

$$y = \sqrt{\frac{2M_1 M_2}{M_1 + M_2}} \left(R_1 - R_2 \right), \tag{5.29}$$

$$z = \sqrt{2(m + M_1 + M_2)} \left(\frac{mr + M_1 R_1 + M_2 R_2}{m + M_1 + M_2} \right). \tag{5.30}$$

Note that the x coordinate is proportional to the distance between the electron and center of the two nuclei, the y coordinate is proportional to the distance between the two nuclei, the z coordinate is proportional to the center of mass coordinate of the whole system. The scaling factors in x, y, z make the kinetic energy operator a Laplacian operator. The unit for x, y, z is $\sqrt{mass} \cdot distance$. Under this transformation, the Hamiltonian becomes

$$
\begin{aligned}
H = &- \hbar^2 \left(\frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y} + \frac{\partial^2}{\partial^2 z} \right) \\
&+ V(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x - \sqrt{\frac{M_2}{2M_1(M_1 + M_2)}} \, y) \\
&+ V(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x + \sqrt{\frac{M_1}{2M_2(M_1 + M_2)}} \, y) \\
&+ U(\sqrt{\frac{M_1 + M_2}{2M_2 M_2}} \, y).
\end{aligned}
\tag{5.31}
$$

The asymptotic form of the solution becomes

$$\psi = \psi_1 + r(E) \cdot \psi_2 + t(E) \cdot \psi_3. \tag{5.32}$$

Where

$$\psi_1 = \phi_1 \left(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x - \sqrt{\frac{M_2}{2M_1(M_1 + M_2)}} \, y \right)$$

$$\times \exp \left[-ik \times \sqrt{\frac{(m + M_1 + M_2)m}{2(M_1 + M_2)}} \frac{1}{m + M_1} \times \left(x + \sqrt{\frac{(m + M_1 + M_2)M_1}{mM_2}} \, y \right) \right]$$

$$= \phi_1 \left(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x - \sqrt{\frac{M_2}{2M_1(M_1 + M_2)}} \, y \right) \times \exp \left[-i \vec{q} \cdot \vec{r} \right],$$

$$(5.33)$$

$$\psi_2 = \phi_1 \left(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x - \sqrt{\frac{M_2}{2M_1(M_1 + M_2)}} \, y \right)$$

$$\times \exp \left[ik \times \sqrt{\frac{(m + M_1 + M_2)m}{2(M_1 + M_2)}} \frac{1}{m + M_1} \times \left(x + \sqrt{\frac{(m + M_1 + M_2)M_1}{mM_2}} \, y \right) \right]$$

$$= \phi_1 \left(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x - \sqrt{\frac{M_2}{2M_1(M_1 + M_2)}} \, y \right) \times \exp \left[i \vec{q} \cdot \vec{r} \right],$$

$$(5.34)$$

$$\psi_3 = \phi_2 \left(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x + \sqrt{\frac{M_1}{2M_2(M_1 + M_2)}} \, y \right)$$

$$\times \exp \left[-ik' \times \sqrt{\frac{(m + M_1 + M_2)m}{2(M_1 + M_2)}} \frac{1}{m + M_2} \times \left(x - \sqrt{\frac{(m + M_1 + M_2)M_2}{mM_1}} \, y \right) \right]$$

$$= \phi_2 \left(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x + \sqrt{\frac{M_1}{2M_2(M_1 + M_2)}} \, y \right) \times \exp \left[-i \vec{q'} \cdot \vec{r} \right].$$

$$(5.35)$$

The $\vec{q}$ and $\vec{q'}$ are just wave vectors in the new coordinates.
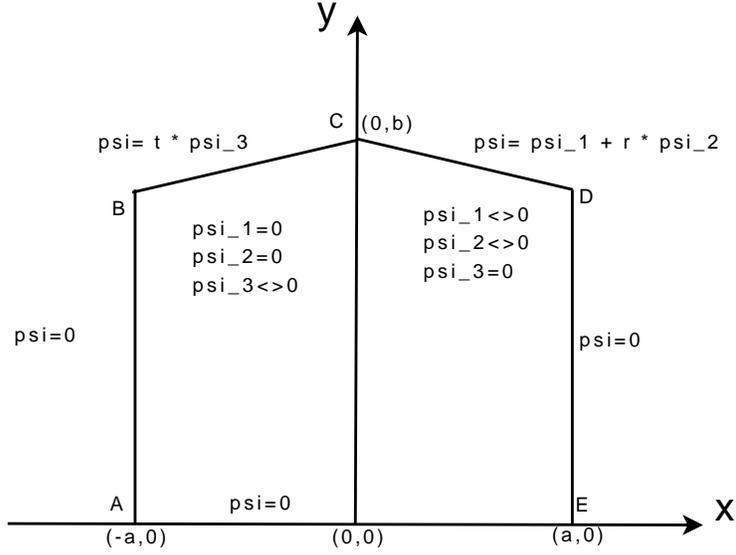
Figure 5.11: The boundary is marked as ABCDE. The parameters a and b should be large enough that the asymptotic form is accurately obtained

5.6.2 Construction of the Boundary Conditions

In the asymptotic form $\psi = \psi_1 + r(E) \cdot \psi_2 + t(E) \cdot \psi_3$, there are two unknowns. However, the boundary conditions can be constructed in a way that these two unknowns do not appear.

Considered the region shown in figure. 5.11. On boundary AE, $y \to 0$, the two nuclei overlap with each other. If the repulsion between them is huge, then $\psi \to 0$. On boundaries AB and DE, if parameter a is large enough, which means the electron is far away from the nuclei, we also get $\psi \to 0$.

On boundary CD, as shown on fig. 5.2(a) and fig. 5.2(b), the electron is far away from the left nucleus, $r - R_2$ is large, so $\phi_3(r - R_2) \to 0$, which means $\psi_3 \to 0$. If we choose CD as $y = -\sqrt{\frac{mM_2}{(m+M_1+M_2)M_1}} \, x + b$, then, on CD, ψ becomes

$$\psi|_{CD} \to \psi_1 + r(E)\psi_2 \tag{5.36}$$

$$= \phi_1\left(\sqrt{\frac{m + M_1 + M_2}{2m(M_1 + M_2)}} \, x - \sqrt{\frac{M_2}{2M_1(M_1 + M_2)}} \, y\right) \times \left(e^{-iq_y b} + r(E)e^{iq_y b}\right) \tag{5.37}$$

$$= \phi_1(...x - ...y) \times C(b, E). \tag{5.38}$$

71

The coefficient $C(b, E)$ is a constant depending on b and E, it can be assigned as 1, because the wave function can be scaled by an arbitrary constant. Thus, on CD, the boundary values are known.

On boundary BC, similar to CD, the electron is far away from the right nucleus, $\phi_1(r - R_1) \to 0$. This leads to $\psi_1 \to 0, \psi_2 \to 0$, and ψ becomes

$$\psi|_{BC} \to t(E) \cdot \psi_3. \tag{5.39}$$

Choose BC as $y = \sqrt{\frac{mM_1}{(m+M_1+M_2)M_2}} \, x + b$, then, on BC

$$\frac{\partial \psi}{\partial \overrightarrow{n}} \to t(E) \frac{\partial \psi_3}{\partial \overrightarrow{n}} \tag{5.40}$$

$$= t(E)\psi_3 \times (ik') \left(\frac{m + M_1 + M_2}{m + M_2} \sqrt{\frac{M_2}{2(M_1 + M_2)M_1}} \right) \tag{5.41}$$

$$= \psi \times (ik') \left(\frac{m + M_1 + M_2}{m + M_2} \sqrt{\frac{M_2}{2(M_1 + M_2)M_1}} \right) \tag{5.42}$$

$$= D(E) \times \psi. \tag{5.43}$$

So, on BC, the value of the normal logarithmic derivitive $(\partial \psi / \partial \overrightarrow{n}) / \psi$ is known.

In conclusion, on $AB, DE, AE, \psi \to 0$, on BC, $\frac{\partial \psi}{\partial \overrightarrow{n}} \to D(E) \times \psi$, on CD, $\psi = \phi_1(...x - ...y) \times C(b, E)$. For a definite energy E, $D(E)$ is fixed, only $C(b, E)$ is unknown, but it can be set to any constant like 1.

5.6.3   Discretization of the Differential Equation

The partial differential equation is discretized by a standard finite element method. Here are some details in the discretization process. First, the domain is decomposed into triangles shown in fig. 5.12. Each triangle has six nodes. The nodes are shared by the adjacent triangles. (Normally, 3-nodes triangle is used. But in this calculation, 6-nodes triangle is much better.) The values of $\psi(r)$ on each node form an unknown vector ψ_i, where i is used to label each node on the entire domain. By choosing a proper set of basis function $\{v_i(r)\}$, $\psi(r)$ can be written as

$$\psi(r) \approx \sum_i \psi_i \cdot v_i(r). \tag{5.44}$$
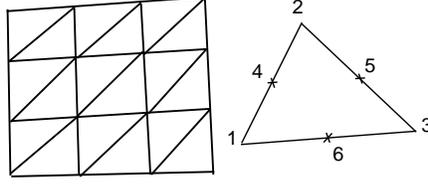
Figure 5.12: An example of decomposition of the domain. In our case the domain shown in fig. 5.11 is a little more irregular, but it can be decomposed into triangles in a similar way. Each triangle is a 6-nodes quadratic element, with half of its nodes located at the center of the edges.

Apparently, $v_i(r)$ satisfy

$$v_i(r_j) = \delta_{ij}. \tag{5.45}$$

In our calculation, the basis function $v_i(r)$ is quadratic on each triangles. By Eq. 5.45, they can be uniquely decided. Thus, $\sum_i \psi_i \cdot v_i(r)$ is unique, continuous and piecewise-quadratic throughout the domain.

The differential equation is discretized by projecting on to each basis function. Starting from

$$\nabla^2 \ \psi(r) + u(r) \ \psi(r) = 0, \tag{5.46}$$

project on to $v_i(r)$,

$$\iint v_i \ \nabla^2 \ \psi + \iint v_i \ u \ \psi = 0, \tag{5.47}$$

apply the Gauss divergence theorem,

$$\oint v_i \ \frac{\partial \psi}{\partial \overrightarrow{n}} - \iint \nabla v_i \cdot \nabla \psi + \iint v_i \ u \ \psi = 0, \tag{5.48}$$

and put in the discretized form of $\psi(r) = \sum_j \psi_j \cdot v_j(r)$,

$$\oint v_i \ \frac{\partial \psi}{\partial \overrightarrow{n}} - \left(\iint \nabla v_i \cdot \nabla v_j \right) \cdot \psi_j + \left(\iint v_i \ u \ v_j \right) \cdot \psi_j = 0 \ \ \text{(sum over j)}. \tag{5.49}$$

The exact form of all $v_i(r)$ are known. By putting in the boundary conditions,

73

finally we get

$$\sum_j A_{ij} \cdot \psi_j = b_i. \tag{5.50}$$

This is the discretization process.

5.6.4 Numerical Solution and Calculation of $R(E)$ and $T(E)$

In our calculation, typically, the linear equation (Eq. 5.50) contains a half million unknowns, the sparse matrix A_{ij} contains ten million non-zeros. This equation is solved by direct LU decomposition method using PETSC and SuperLU on Seawulf Cluster.

Figure 5.13 is a typical solution of this equation, with the potential between the nucleus and the electron as $-e^{-|x|}$. and the potential between the nuclei as $\frac{1}{(r/R_0)^6}$. Both nuclei have the same mass, 30 atomic unit. The total kinetic energy is 0.1 Hartree, which is much lower than the gap between the ground state and excited state.

When the solution is ready, coefficients $r(E)$ and $t(E)$ can be calculated by fitting the solution to the asymtotic form as shown in fig.

$$\psi = \frac{\psi_1 + r(E) \cdot \psi_2 + t(E) \cdot \psi_3}{e^{-iq_y b} + r(E)e^{iq_y b}}. \tag{5.51}$$

In the top right area in fig. 5.11, the solution is

$$\begin{aligned}
\psi_{\text{asymptotic}} &= \frac{\psi_1 + r(E) \cdot \psi_2}{e^{-iq_y b} + r(E)e^{iq_y b}} \\
&= \phi_1(...x - ...y) \times \frac{\exp\left[-i\overrightarrow{q} \cdot \overrightarrow{r}\right] + r(E)\exp\left[i\overrightarrow{q} \cdot \overrightarrow{r}\right]}{\exp\left[-iq_y b\right] + r(E)\exp\left[iq_y b\right]}.
\end{aligned} \tag{5.52}$$

The only parameter that needs to be fitted is r(E). (I also fit/relax the $\overrightarrow{q}$. It is proportional to the momentum, because $\overrightarrow{q}$ is related to the phase in Eq. 5.52. A very small difference will result in huge error. It turns out that if I relax $\overrightarrow{q}$ a little, usually less than 1% , the results can be improved a lot). The fitting is done by minimizing the sum $\sum_i |\psi_{\text{asymptotic}} - \psi_{\text{calculated}}|^2$ in the top right area in fig. 5.11. The minimization is done by using the multidimensional minimization function in the GSL library. The transfer amplitude $t(E)$ can be calculated in a similar way.

Typically, the differences between the calculated solution and the asymp-

t(E) is calculated by
fitting this area

r(E) is calculated by
fitting this area
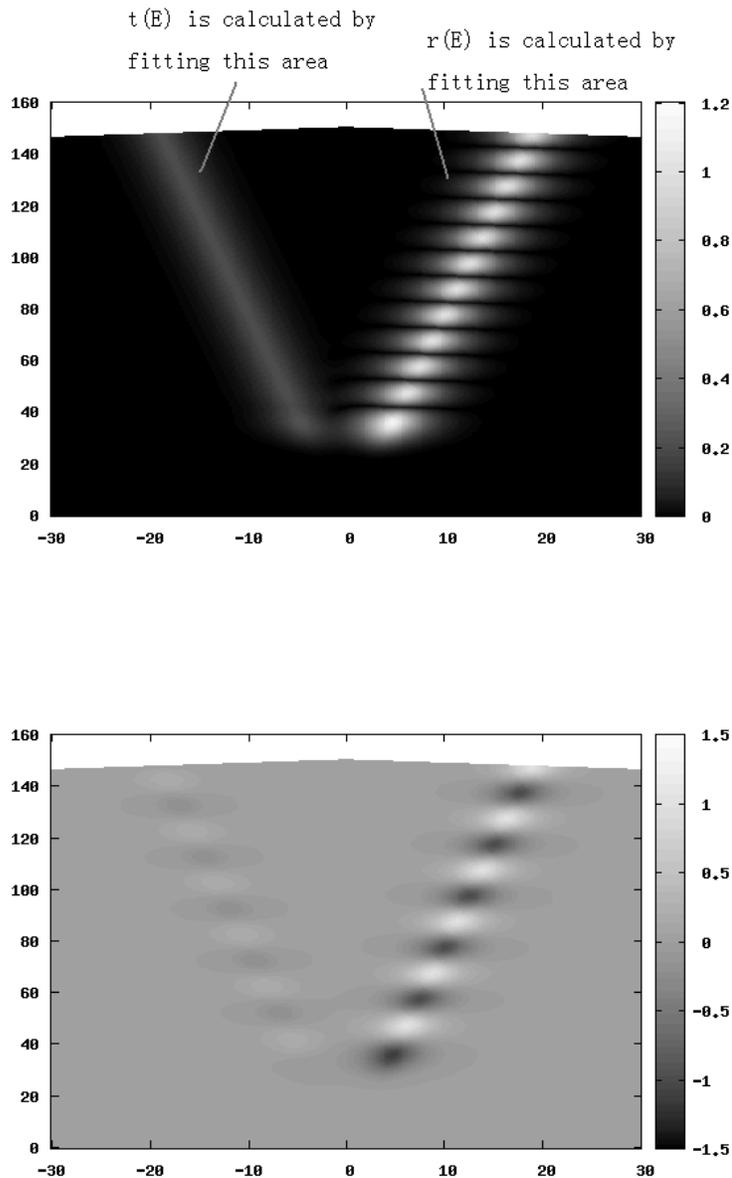
Figure 5.13: A typical solution, shown as abs(ψ) and real(ψ), respectively. Note that the coordinates are the transformed ones (Eq. 5.30). The potential between the nucleus and the electron is $-e^{-|x|}$. The potential between the nuclei is $\frac{1}{(r/R_0)^6}$. Both nuclei have the same mass, 30 electron masses. The total kinetic energy is 0.1 Hartree, which is a little higher than the gap between the ground state and excited state(0.05 Hartree).

totic form can be kept under 1%, namely

$$\max(|\psi_{\text{calculated}}(r) - \psi_{\text{asymptotic}}(r)|) \, / \, \max(|\psi_{\text{calculated}}(r)|) < 1\%. \qquad (5.53)$$

Both the magnitude and phase of $r(E)$ and $t(E)$ can be calculated. However, the magnitude is more stable, while the phase is less stable.

When $r(E)$ and $t(E)$ are known, the calculation of reflection and transmission probability $R(E)$ and $T(E)$ is straightforwad using Eq. 5.11 and Eq. 5.12. Then the conservation law $R + T = 1$ can be tested.

Bibliography

[1] H. Chen, L. Wang, J. Bai, J. C. Hanson, J. B. Warren, J. T. Muckerman, E. Fujita, and J. A. Rodriguez. *J. Phys. Chem. C*, 114(4):1809, 2010.

[2] K. Maeda, K. Teramura, T. Takata, M. Hara, N. Saito, K. Toda, Y. Inoue, H. Kobayashi, and K. Domen. *J. Phys. Chem. B*, 109(43):20504, 2005.

[3] K. Maeda, T. Takata, M. Hara, N. Saito, Y. Inoue, H. Kobayashi, and K. Domen. *J. Am. Chem. Soc.*, 127(23):8286, 2005.

[4] K. Maeda, K. Teramura, D. Lu, T. Takata, N. Saito, Y. Inoue, and K. Domen. *Nature*, 440(7082):295, 2006.

[5] W. Wei, Y. Dai, K. Yang, M. Guo, and B. Huang. *J. Phys. Chem. C*, 112 (40):15915, 2008.

[6] M. N. Huda, Y. Yan, S. Wei, and M. M. Al-Jassim. *Phys. Rev. B*, 78(19): 195204, 2008.

[7] L. L. Jensen, J. T. Muckerman, and M. D. Newton. *J. Phys. Chem. C*, 112(9):3439, 2008.

[8] C. Di Valentin. *J. Phys. Chem. C*, 114(15):7054, 2010.

[9] S. Wang and L. Wang. *Phys. Rev. Lett.*, 104(6):065501, 2010.

[10] M. Born and R. Oppenheimer. *Ann. Phys. (Leipzig)*, 84(20):457, 1927.

[11] P. Hohenberg and W. Kohn. *Phys. Rev.*, 136(3B):B864, 1964.

[12] W. Kohn and L. J. Sham. *Phys. Rev.*, 140(4A):A1133, 1965.

[13] J. P. Perdew and A. Zunger. *Phys. Rev. B*, 23(10):5048, 1981.

[14] J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77(18):3865, 1996.

[15] A. D. Becke. *Phys. Rev. A*, 38(6):3098, 1988.

[16] C. Lee, W. Yang, and R. G. Parr. *Phys. Rev. B*, 37(2):785, 1988.

[17] B. Miehlich, A. Savin, H. Stoll, and H. Preuss. *Chem. Phys. Lett.*, 157 (3):200, 1989.

[18] D. R. Hamann, M. Schlüter, and C. Chiang. *Phys. Rev. Lett.*, 43(20): 1494, 1979.

[19] D. Vanderbilt. *Phys. Rev. B*, 32(12):8412, 1985.

[20] P. E. Blöchl. *Phys. Rev. B*, 50(24):17953, 1994.

[21] G. Kresse and D. Joubert. *Phys. Rev. B*, 59(3):1758, 1999.

[22] G. Kresse and J. Furthmüller. *Phys. Rev. B*, 54(16):11169, 1996.

[23] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen. *Phys. Rev. B*, 52(8): R5467, 1995.

[24] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton. *Phys. Rev. B*, 57(3):1505, 1998.

[25] M. Yashima, K. Maeda, K. Teramura, T. Takata, and K. Domen. *Chem. Phys. Lett.*, 416(4-6):225, 2005.

[26] H. Chen, W. Wen, Q. Wang, J. C. Hanson, J. T. Muckerman, E. Fujita, A. I. Frenkel, and J. A. Rodriguez. *J. Phys. Chem. C*, 113:3650, 2009.

[27] M. Yashima, H. Yamada, K. Maeda, and K. Domen. *Chem. Commun.*, 46:2379, 2010.

[28] D. De Fontaine. *Solid State Phys.*, 47:33, 1994.

[29] J. W. D. Connolly and A. R. Williams. *Phys. Rev. B*, 27(8):5169, 1983.

[30] J. M. Sanchez, F. Ducastelle, and D. Gratias. *Physica A*, 128(1-2):334, 1984.

[31] F. Ducastelle. *Order and Phase Stability in Alloys*. Elsevier Science, New York, 1991.

[32] A. Zunger. In *First principles statistical mechanics of semiconductor alloys and intermetallic compounds*, NATO Advanced Studies Institute, Series B: Physics. Plenum, New York, 1994.

[33] P. D. Tepesch, G. D. Garbulsky, and G. Ceder. *Phys. Rev. Lett.*, 74(12): 2272, 1995.

[34] S Geisser. *J. Amer. Statistical Assoc.*, 70(350):320, 1975.

[35] M. Stone. *J. Roy. Statist. Soc. Ser. B*, 39:44, 1977.

[36] J. Shao. *J. Am. Stat. Assoc.*, 88(422):486, 1993.

[37] P. Zhang. *Ann. Statist.*, 21:299, 1993.

[38] K. Baumann. *TrAC, Trends Anal. Chem.*, 22(6):395, 2003.

[39] N. A. Zarkevich and D. D. Johnson. *Phys. Rev. Lett.*, 92(25):255702, 2004.

[40] A. van de Walle, M. Asta, and G. Ceder. *Calphad*, 26(4):539, 2002.

[41] A van de Walle and M Asta. *Model. Simul. Mater. Sc.*, 10(5):521, 2002.

[42] A. van de Walle and G. Ceder. *J. Phase Equilib.*, 23:348, 2002.

[43] A. van de Walle. *Calphad*, 33(2):266, 2009.

[44] G. L. W. Hart and R. W. Forcade. *Phys. Rev. B*, 77(22):224115, 2008.

[45] D. B. Laks, L. G. Ferreira, S. Froyen, and A. Zunger. *Phys. Rev. B*, 46 (19):12587, 1992.

[46] L. G. Wang and A. Zunger. *Phys. Rev. B*, 68(12):125211, 2003.

[47] A. Baldereschi, S. Baroni, and R. Resta. *Phys. Rev. Lett.*, 61(6):734, 1988.

[48] G. Makov and M. C. Payne. *Phys. Rev. B*, 51(7):4014, 1995.

[49] J. Neugebauer and M. Scheffler. *Phys. Rev. B*, 46(24):16067, 1992.

[50] M. D. Pashley. *Phys. Rev. B*, 40(15):10481, 1989.

[51] K Shiraishi. *J. Phys. Soc. Jpn.*, 59(10):3455, 1990.

[52] R Resta. *Ferroelectrics*, 136(1-4):51, 1992.

[53] R. D. King-Smith and D. Vanderbilt. *Phys. Rev. B*, 47(3):1651, 1993.

[54] D. Vanderbilt and R. D. King-Smith. *Phys. Rev. B*, 48(7):4442, 1993.

[55] A. S. Barker and M. Ilegems. *Phys. Rev. B*, 7(2):743, 1973.

[56] P. Perlin, I. Gorczyca, N. E. Christensen, I. Grzegory, H. Teisseyre, and T. Suski. *Phys. Rev. B*, 45(23):13307, 1992.

[57] F. Bernardini, V. Fiorentini, and D. Vanderbilt. *Phys. Rev. B*, 56(16): R10024, 1997.

[58] F. Bernardini, V. Fiorentini, and D. Vanderbilt. *Phys. Rev. Lett.*, 79(20): 3958, 1997.

[59] F. Bernardini, V. Fiorentini, and D. Vanderbilt. *Phys. Rev. B*, 63(19): 193201, 2001.

[60] R. W. Nunes and X. Gonze. *Phys. Rev. B*, 63(15):155107, 2001.

[61] I. Souza, J. Íñiguez, and D. Vanderbilt. *Phys. Rev. Lett.*, 89(11):117602, 2002.

[62] E. Arbel and J. W. Cahn. *Surf. Sci.*, 51(1):305, 1975.

[63] S. B. Zhang and S. H. Wei. *Phys. Rev. Lett.*, 92(8):086102, 2004.

[64] J. E. Northrup. *Phys. Rev. B*, 73(11):115304, 2006.

[65] A. R. Oganov and C. W. Glass. *J. Chem. Phys.*, 124:244704, 2006.

[66] C. W. Glass, A. R. Oganov, and N. Hansen. *Comp. Phys. Comm.*, 175: 713, 2006.

[67] T. Akiyama, D. Ammi, K. Nakamura, and T. Ito. *Phys. Rev. B*, 81(24): 245317, 2010.

[68] X. Shen, Y. A. Small, J. Wang, P. B. Allen, M. V. Fernandez-Serra, M. S. Hybertsen, and J. T. Muckerman. *J. Phys. Chem. C*, 114(32):13695, 2010.

[69] J. C. Tully. *J. Chem. Phys.*, 93(2):1061, 1990.