First Principles Investigation of Entropy and Equilibrium Atomic Structure at the Surface of III–V Semiconductors

by

John C. Thomas

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Materials Science and Engineering)
in The University of Michigan
2012

Doctoral Committee:

Associate Professor Anton Van der Ven, Co-chair
Professor Joanna Mirecki Millunchick, Co-chair
Assistant Professor Emmanouil Kioupakis
Assistant Professor Donald Jason Siegel

ACKNOWLEDGEMENTS

There are a number of people without whom my graduate studies would have been less successful; there are others without whom it would have been less pleasant. It has been my great fortune that my thesis advisors, Professor Joanna Millunchick and Professor Anton Van der Ven, have been both of these. Their mentorship has been inspiring and enriching, and they have provided me freedom and support to tread the paths I find most rewarding, which truly is a luxury.

Dr. Normand Modine has also been a guiding presence (from afar) in the course of my thesis research. His infectious curiosity for surface science and solid state physics has broadened my understanding and played an instrumental role in inspiring me to develop the methods described in Chapter III.

My committee members, Professor Emmannouil Kioupakis and Professor Donald Siegel, though their influence on my research has begun only recently, have refreshed my perspective on this work through their comments, questions, and suggestions. I thank them for their time and the enthusiasm with which they have approached this commitment.

I would like to acknowledge the Rackham Pre-doctoral Fellowship Program for financial support in the final year of my studies and the Center for Integrated Nanotechnologies at Sandia National Laboratories for usage of computing resources without which this work would not have been possible.

A number of my graduate student colleagues were additionally valuable mentors early in my studies, including Jishnu Bhattacharya and Jessica Bickel. All of my workmates have contributed to the relaxed yet scholarly environment that I have enjoyed, though I owe large parts of my sanity to Kevin Grossklaus and Denis Nothern for their useful conversations and occasional commiseration.

I am very grateful to my parents, who have always given me their support, encouragement, and steadfast reassurance, often in greater measure than I request—but always in the amount that I find is needed.

Most of all, I must thank Elizabeth Canavan-Palermo, my companion in all things. Despite distance, fatigue, frustration, and the burdens of our busy lives, she is the salve that always soothes me and the smile that I cannot hide.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

ABSTRACT

First Principles Investigation of Entropy and Equilibrium Atomic Structure at the Surface of III–V Semiconductors

by

John C. Thomas

Co-Chairs: Anton Van der Ven and Joanna Mirecki Millunchick

The cleaved surfaces of III–V semiconducting compounds can assume a wide variety of surface structures over a range composition. Because III–V alloys, heterostructures, and devices are synthesized epitaxially, under carefully controlled conditions, the ability to reproducibly control surface structure of these materials could enable powerful optimization pathways for III–V applications. Unfortunately, significant challenges prevent accurate prediction of thermodynamic stability and equilibrium properties of multicomponent crystalline surfaces. This dissertation describes rigorous methods that address these challenges and the application of these methods to the surfaces of III–V materials generally, and to the (001) surface of GaAs and its dilute alloys in particular. The methods presented build upon one another, following a progression that begins with systematic enumeration of physically plausible surface structures, proceeds with the use of rigorous energy models and simulation techniques to account for different forms of disorder, and culminates in applications of these techniques to explore alloy ordering and structural disorder of more complex systems.

First, we present a new algorithm to enumerate all possible surface structures that obey basic observed structural trends, which significantly simplifies and systematizes the determination of equilibrium surface structure. The algorithm enables the generation of a large database of potential structures that can be explored using phenomenological or first principles methods to determine the equilibrium behavior of a particular system. We use this method to study the (001)-oriented GaAs surface and identify several new low-energy surface structures that may explain observations on GaAs and its alloys.

Starting from the low-energy structure identified by enumeration, we use first principles energy calculations to parameterize an effective Hamiltonian that can be used in Monte

Carlo simulation to account for thermal vibrations and configurational disorder. Monte Carlo free energies are used to predict the As-rich surface reconstructions of GaAs(001) that are stable at finite temperature. Experimental results are used to calibrate the error in calculated surface free energies arising from DFT approximations. With error bounds taken into account, the calculated phase diagram indicates existence of a stable (4×3) surface reconstruction on GaAs (001), in agreement with experiment.

In order to explore the role of configurational order on surface structure stability, we use the cluster expansion to study surface alloying of In and Ga at the GaAs (001) surface. The ordering behavior of the alloy species in these semiconductor surfaces is important due to its prominent effect on the nucleation and layer-by-layer growth of nanostructured materials for novel devices. We use the cluster expansion formalism to describe In/Ga substitution and As_2 dimer adsorption in the top two surface layers of the experimentally-observed GaAs (2×4) surface structure. Monte Carlo simulations of the alloyed surface rigorously predict finite temperature ordering phenomena. In order to place these results in context, we use the cluster expansion formalism to perform a comprehensive search for ground-state reconstructions of the alloyed InAs/GaAs(001) surface and construct a zero-temperature phase diagram describing the stability of (2×4), (4×3), c(4×4) and (4×2) phases of the alloyed surface.

CHAPTER I

Introduction

Material surfaces, by their very nature, mediate interactions of a material with the outside environment. As such, surfaces rouse essential considerations in the course of engineering and optimizing material behavior, especially for micro- or nano-scale applications, where much of the bulk material is near an interface. Optimization goals for surfaces may include enhancement or suppression of chemical reactivity or of thermal, electronic, or mass exchange with the environment. Because the surface is also the point at which accretion or erosion occur during material synthesis and processing, the instantaneous state and behavior of the surface during these procedures can play a significant role on the properties of the resulting bulk material or device.

At the surface of crystalline materials in particular, even the atomic-scale details of surface structure play important roles in materials synthesis and application. Atomic-scale surface structure has been shown to have significant effects on catalysis[3], oxide formation[4], and magnetic domain ordering[5]. Crystalline oxides and semiconductors for advanced electronic applications require precisely controlled defect and impurity densities, requiring that they be synthesized layer-by-layer in near-vacuum using techniques such as molecular beam epitaxy (MBE) or atomic-layer deposition (ALD). Due to the sensitive nature of epitaxial synthesis, the structure of the surface during epitaxy can broadly influence the properties of the resulting material[6], and evidence suggests that the atomic-scale surface structure is thermodynamically and kinetically linked to surface morphology at longer length-scales[7], with significant potential implications for nanostructural self-assembly.

The atomic structure at the surface of semiconductor alloys during their synthesis can induce bulk alloy ordering in the plane of the surface by dominating the order of alloy species at the surface, which subsequently becomes kinetically trapped in the growing crystal[8]. These ordered alloys have an altered band-gap relative to the disordered bulk alloy, as well as anisotropic electronic properties[9]. Atomic surface structure can likely enhance surface segregation in alloys as well, thereby improving interface abruptness at semiconductor

heterojunctions, where one material is grown coherently on top of another[10]. The enhanced abruptness can measurably improve I–V characteristics of the resulting device[11].

The characteristics of a particular crystallographic plane of a material often allow it to accommodate a range of equilibrium surface structures that are qualitatively different from the bulk. Although the availability of many structural degrees of freedom can offer increased control over synthesis and reveal paths towards optimizing surface properties, high dimensionality of the parameter space also presents significant challenges in the quest to resolve atomic-scale surface features experimentally, and characterization is only further complicated for multicomponent and alloyed surfaces, in which two or more species may arrange in different ways on the surface.

At the atomic scale, many ways of arranging the surface may differ only slightly in energy, compared to the thermal energy of the system. Thermal excitations allow the surface to fluctuate among its low-energy arrangements, causing disorder at finite temperature, which is an additional impediment to characterization. If two qualitatively different structures are energetically similar, but one structure has a greater availability of low-energy arrangements, the structure with more arrangements will be statistically preferred at a given temperature. This preference for a disordered surface is due to it having greater entropy and consequently, lower *surface free energy* than an energetically similar, but more ordered, surface. The importance of entropy to the free energy of atomic-scale surface structure complicates attempts to predict the surface that is thermodynamically stable in real experiments. Although widely-employed techniques, such as density functional theory (DFT), exist to calculate energies of structures in absence of thermal energy, calculating the finite-temperature free energy of a structure requires additional considerations and manipulation of zero-temperature energy data.

The complicated nature of crystalline surfaces and the vast parameter space in which their equilibrium phenomena are described make the construction and confirmation of hypotheses for equilibrium surface structure not only a practical endeavor, but also a compelling theoretical puzzle. **The central purpose of this dissertation is to outline and describe rigorous and systematic methods to predict, beginning from first principles, the finite-temperature thermodynamic stability and equilibrium properties of crystalline surfaces.** These methods build upon each other, following a progression that begins with a codified enumeration procedure for physically plausible surface structures, proceeds with the use of rigorous energy models and simulation techniques to account for the different forms of disorder that affect finite-temperature surface structure, and culminates in applications of these techniques to explore alloy ordering and structural disorder of more complex alloy systems. Emphasis in the application of these methods is given to III–V materials

generally, and to the (001) surface of GaAs and its dilute alloys in particular. Because III–V materials have polar, directional bonds, and because III–V alloys, heterostructures, and devices are predominantly synthesized using MBE, the atomic details of their surface structure have broad relevance to applications that utilize III–V materials.

1.1 Differences Between Atomic Structure of Surface and Bulk: Surface Reconstructions

The creation of a surface changes the local environment of atoms in its vicinity by breaking symmetries and altering the effective potentials experienced by near-surface atoms. In effect, surface atoms are subject to chemical, elastic, and electrostatic interactions that do not exist in the bulk material system. These interactions drive local relaxations that result in differences in structure and atomic coordination relative not only to the bulk material but also to the ideal, planar surface of cleavage. The local relaxation and rebonding of surface atoms is called a surface reconstruction. Unfortunately, a central consequence of surface reconstruction is that the resulting surface structures and properties cannot easily be predicted based on bulk material properties. This section describes the driving forces for surface reconstruction and the relationship of surface structure to the bulk crystallography.

1.1.1 Crystallography of Zincblende III–V Materials Relative to the (001) Surface

Most III–V compounds have a zincblende bulk crystal structure that can be described by the conventional unit cell shown in Fig. 1.1. The zincblende structure takes its name from the ZnS prototype crystal and has space group symmetry $F\bar{4}3m$ (space group number 216). The zincblende crystal is comprised of two interpenetrating FCC sublattices; each atom of one sublattice shares four tetrahedrally-oriented bonds with atoms of the other sublattice. In this sense, the zincblende crystal is a derivative structure of the diamond crystal, where all nearest-neighbor pairs are composed of two different species, thereby destroying the inversion symmetry of the diamond crystal that exists between nearest-neighbor pairs. As illustrated in Fig. 1.1(b)-(c), the zincblende crystal exhibits an $ABCD$ stacking of square $2D$ lattices along [001], with A and C layers belonging to the Group V, or anion, sublattice and B and D layers belonging to the Group III, or cation, sublattice. Each layer is shifted by $\sqrt{2}a_B/4$ relative to the previous layer, where a_B is the bulk lattice parameter of the cubic unit cell.

Due to the $ABCD$ stacking of layers along [001], a perfectly cleaved (001) surface has pmm plane-group symmetry, consisting of intersecting mirror planes along [110] and [$1\bar{1}0$] and a two-fold rotation axis perpendicular to the plane. The semi-infinite bulk below the

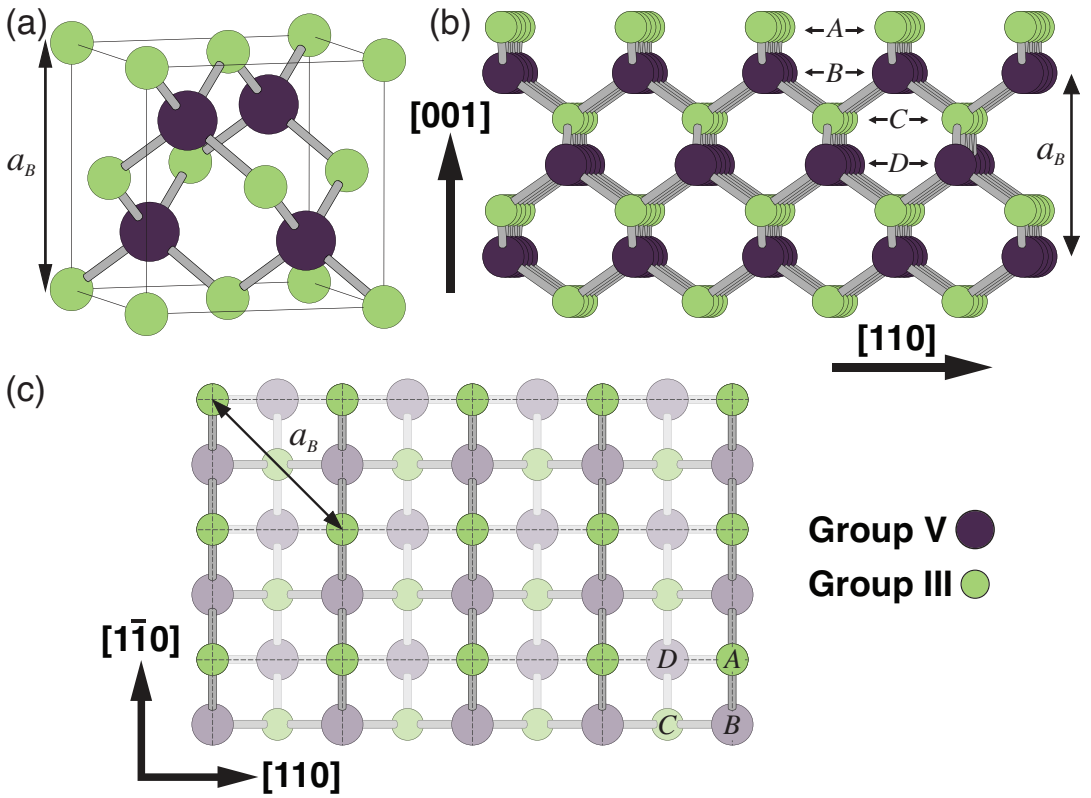


Figure 1.1. Crystallographic overview of zincblende III–V materials. (a) The zincblende conventional unit cell. (b) The zincblende crystal, cleaved at the (001) surface, viewed along [1$\bar{1}$0]. (c) The (001) cleaved surface viewed from above, with dashed lines to indicate the surface primitive cell. In all cases, a_B indicates the bulk lattice parameter of the conventional cell. (b) and (c) label the $ABCD$ layer stacking.

surface plane breaks other symmetries that would otherwise exist for the square lattice of the atomic monolayer. The natural choice of $2D$ primitive lattice vectors for the surface plane are along [110] and [1$\bar{1}$0], both having length $a_s \equiv \sqrt{2}a_B/2$. Each atom at the cleaved surface has two fewer tetrahedral nearest neighbors relative to the bulk atoms but retains the two tetrahedral nearest neighbors in the direction of the bulk. These remaining neighbors lie either in a (110) plane with the surface atom (in the case of A and C layers) or a (1$\bar{1}$0) plane (in the case of B and D layers).

1.1.2 Electronic Structure and Surface Reconstruction in III–V Materials

When the constituent Group III and Group V species have sufficiently low atomic number Z, the III–V compounds can be approximated fairly well with a simple valence-shell model[2]. Owing to their tetrahedral coordination, gross features of the III–V crystal electronic structure can be conceptualized as arising from a linear combination of sp^3-hybridized orbitals of the constituent atoms. III–V compounds are polar, and the sp^3 orbitals of an isolated Group III species are higher in energy than the sp^3 orbitals of a Group V species. Within the III–V compound, the linear combination of orbitals forms valence band states, which are even lower in energy than the Group V sp^3 orbitals, and anti-bonding conduction band states. The relative energies of the various electronic-state configurations are depicted in Fig. 1.2. The bonding states of III–V combinations are significantly lower in energy than either III–III or V–V combinations, resulting in the robust stability of the zincblende structure relative to other crystal structures (the exception being several of the Group III–nitride compounds, which tend to have the crystallographically related wurtzite structure).

At the cleaved (001) surface plane, the broken bonds result in each surface atom having two "dangling bonds" or, more accurately, "dangling hybrids". These orbitals, which cannot participate directly in bonding, assume energies near those of the hybridized orbitals of the isolated atom. Electrons from the bulk and surface atoms will tend to fill all the dangling hybrid states that have energies below the fermi level. If there are either too few or too many surface valence band states, charge buildup will occur at the surface, resulting in large electrostatic energies. The energy of surface electronic states, as well as the electrostatic energy, can be reduced if the surface further reduces its symmetry by rearranging to form new bonds and "reconstruct" the surface, even if the new bonds are III–III or V–V bonds. The reconstructed surface not only has reduced plane-group symmetry but also has reduced translational symmetry and, consequently, a larger unit cell than the unreconstructed primitive surface cell. Figure 1.3 shows how two well-accepted III–V surface reconstructions are related to the unreconstructed (001) surface plane.

The electrostatic and chemical driving forces for surface reconstruction are balanced by

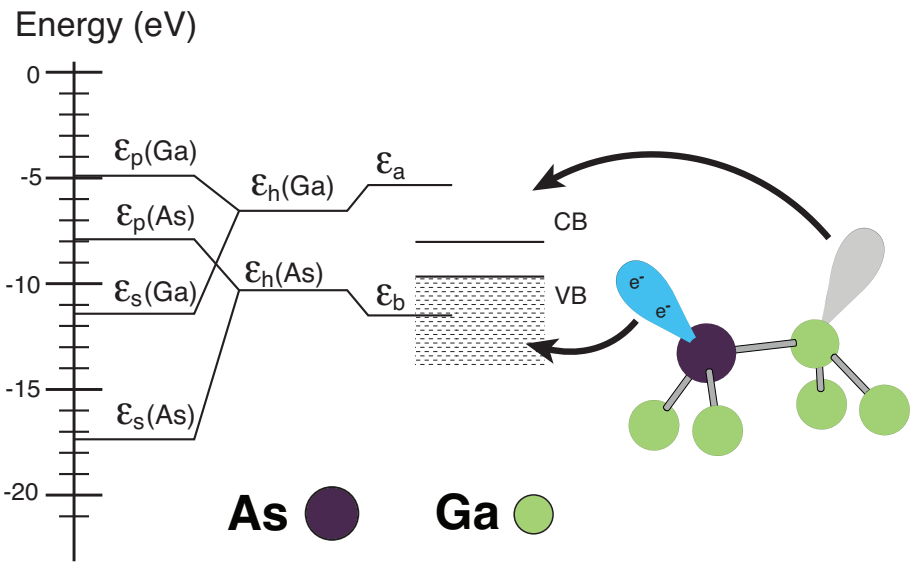


Figure 1.2. Illustration of electronic states relevant to GaAs. ε_p and ε_s are energies of s and p orbitals, depicted for both Ga and As. ε_h is the energy of the sp^3 hybridized orbitals. ε_b and ε_a are energies of the bonding and antibonding, respectively, estimated for the Ga–As covalent bond. The approximate conduction band minimum and valence band maximum of GaAs are shown for reference. The relative energetics of the various electronic structure regimes drive surface reconstruction and the filling of surface states. Figure adapted from [1] and [2].

energy penalties for bond-stretching and bond-bending. Even the minimum-energy reconstruction will still have some atoms that are undercoordinated relative to the tetrahedral bulk atoms, although atoms at the reconstructed (001) surface nearly always has at least three-fold coordination. Electrons fill the remaining dangling hybrids of the undercoordinated atoms according to the position of the dangling surface states relative to the Fermi level. Reconstructions that cannot balance the number of electrons contributed from the bulk with the number of valence-band surface states are said to violate the electron counting rule (ECR)[1], which is a phenomenological rule used to determine whether a surface structure will be semiconducting or, in the event that it has a partially filled valence or conduction band due to unbalanced surface states, that it is conducting. Reconstructed surfaces that violate the ECR tend to be high in energy, due to the resultant charge buildup at the surface. There is some limited experimental evidence of ECR violations at the (001) surface of III–V surfaces. The extremely Sb-rich $c(2\times10)$ reconstruction on GaSb (001)[12] and the Bi-stabilized (2×1) reconstruction observed on the (001) surface of GaAs and InP[13] are notable cases. The ECR is also applicable to a number of other chemistries of important polar and non-polar semiconductors.

Considerations related to the ECR reveal how additional degrees of freedom, that are generally negligible in the bulk, can become important at the surface. If a structure obeys

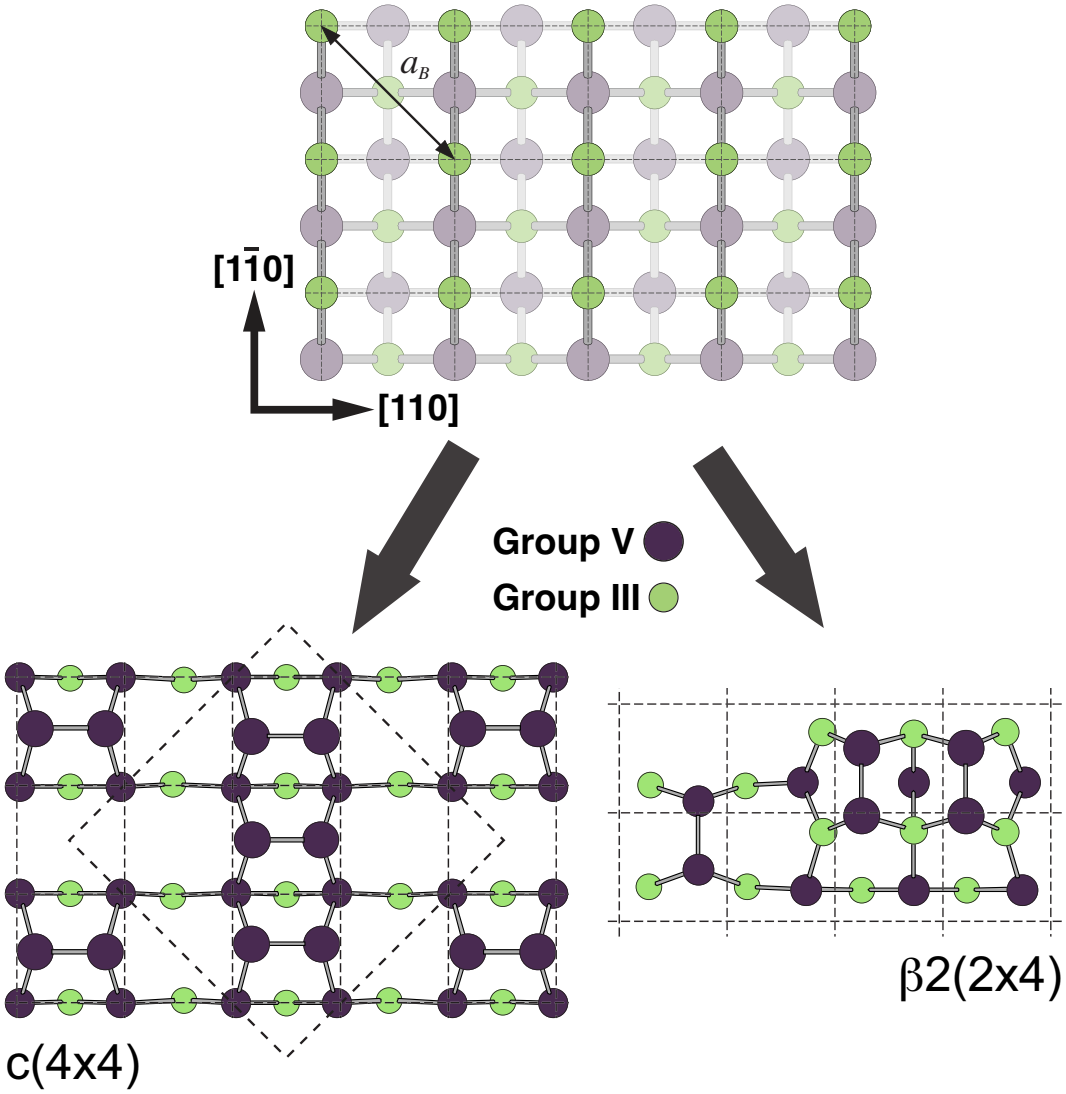


Figure 1.3. The (001) surface of zincblende III–V materials (top) can reconstruct in different ways, do to the directionality of its polar-covalent bonds, to form different types of surface structure. The $c(4\times4)$ and $\beta2(2\times)$ are two well-known examples, but many others have been observed on the various III–V compounds and alloys.

the ECR, charge balance arguments suggest that the energy of substituting Group III for Group V, or visa versa, at one of its tricoordinate (i.e., three-fold coordinated) surface sites should be significantly lower than performing such a substitution in the bulk. When non-isovalent substitution is performed in the bulk, electrons are added or removed without changing the number of valence band states. If the same substitution is performed at a tricoordinate surface site, however, the subsitution will remove electrons while also elevating the dangling-bond state of the site into the conduction band (in the case of III for V), or it will add electrons while also lowering the dangling bond state into the valence band (in the case of V for III). The fact that III/V species substitution at tricoordinate surface sites can occur without altering the local charge neutrality suggests that such substitutions can be sufficiently low in energy to be excited at finite temperature, leading to increased disorder and entropy of the surface.

1.2 Difficulties of Characterizing Surface Structure

Due to the complications posed by thermal disorder as well as the difficulties in general of resolving atomic-scale structural detail, full characterization of the atomic-scale details of surface reconstructions has proven difficult in systems where significant disorder is present or when details of electronic structure cause accepted characterization techniques to perform poorly. Even if the periodicity and symmetries of a surface reconstruction can be determined experimentally, the structure may still evade description by a well-defined atomistic model. Examples include the $(1\times5)/c(2\times10)$ reconstruction on GaSb (001)[14] and a number of perplexing oxygen adsorbate structures on Ag (111)[15].

Experimental techniques used to characterize surface reconstructions can be separated into two broad categories: reciprocal-space diffraction techniques and real-space microscopy techniques. Reciprocal-space methods measure the diffraction of a radiation source as it interacts with the surface. Electrons are typically used as the radiation signal, as in the case of reflection high-energy electron diffraction (RHEED), since their shallow penetration depth make them suitable as a surface-sensitive probe. Reciprocal-space techniques such as RHEED can very easily measure periodicity of the surface resulting from long-range order by effectively performing a spatially-averaged Fourier transform over a region of the surface. Figure 1.4(a)-(b) show the data obtained from a RHEED measurement of a GaAs(001) surface on which a thin layer of InAs has been deposited. Although reciprocal-space measurements provide crucial information about the periodicities and symmetries of surface structure, the spatial averaging that implicitly occurs when a diffraction measurement is made means that diffraction techniques are ill-suited to characterization of short-lengthscale features on a dis-

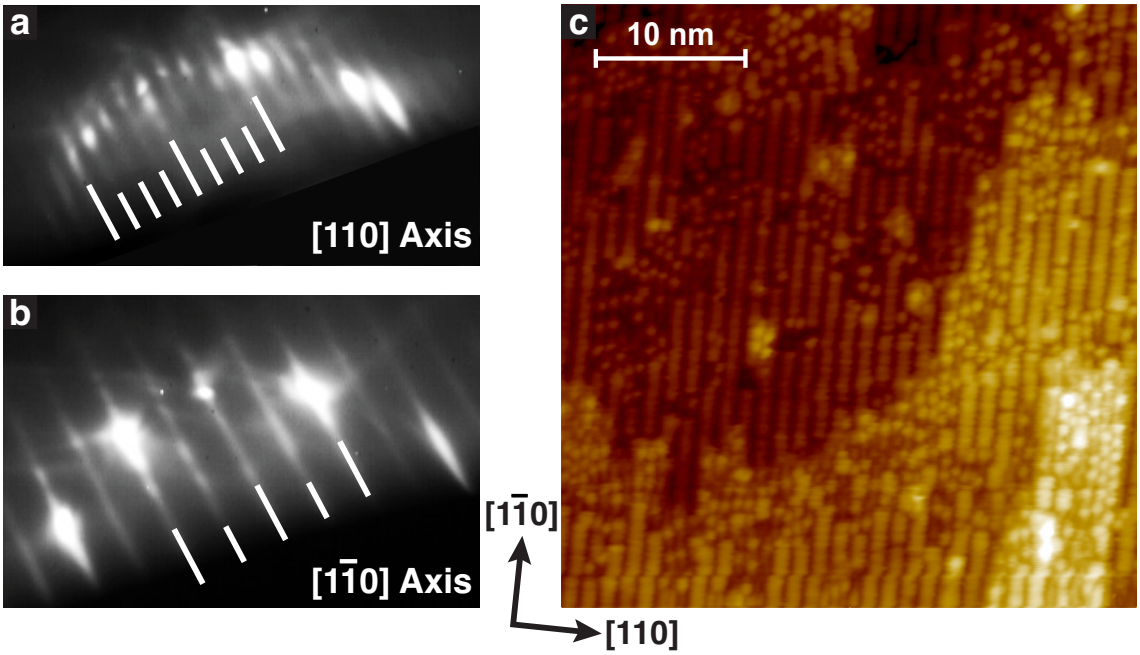


Figure 1.4. Experimental characterization measurements of GaAs (001) after deposition of 1/5 monolayer of InAs at 500°C. (a) and (b) are RHEED diffraction patterns, immediately after deposition, that indicate (2×4) periodicity. (c) is an STM micrograph, taken after cooling, showing that the surface is disordered and exhibits multiple reconstructions.

ordered surface. As an example, if disorder occurs at the length-scale of several surface lattice vectors, as when a reconstruction unit cell does not tile the surface ideally, RHEED measurements will indicate a reduced periodicity that does not describe the shape of the actual unit cell[16]. Moreover, diffractive measurements can be difficult to interpret when multiple types of reconstruction share the surface, causing either the majority reconstruction to dominate the measurement, or the features of both structures to be superposed.

Real-space measurements can be used to fill in details of the surface structure to which reciprocal-space methods are insensitive. If the surface is sufficiently planar, real-space measurements can be performed using scanning probe techniques, such as scanning tunneling microscopy (STM) or atomic force microscopy (AFM). STM measures the occupation and real-space position of surface states by tunneling electrons between the surface and an atomically-sharp metallic tip. By varying the bias voltage between surface and tip, electrons either tunnel out of filled surface states or into empty surface states. Since the ECR specify that cation dangling hybrids and anion dangling hybrids fill differently, STM can reveal some information about the species configuration of the surface. However, on alloyed surfaces, such as $In_xGa_{1-x}As$ (001), where isovalent substitution occurs, differences between distinct cation species are very difficult to distinguish, As are differences between anion species. Additionally, STM is extremely sensitive to noise, tip conditions, and the electronic properties and

dopant profiles in the sample. Although STM can reveal atomic-scale details under idea conditions, its effective resolution is frequently lower in practice. The juxtaposition of RHEED and STM measurements of a single InAs/GaAs (001) surface in Fig. 1.4(a)-(c) demonstrate how both real- and reciprocal-space measurements complement each other. The RHEED measurements clearly reveal long-range periodicity of the surface that is not apparent from STM images; the STM images reveal that there is a minority reconstruction on the surface, of which the RHEED measurements give no indication.

Due to the complex nature of III–V surfaces, there are a number of observed III–V surface reconstructions for which no verifiable structural models exist, despite an abundance of real- and reciprocal-space characterization data. In addition to the $(1\times5)/c(2\times10)$ reconstruction on GaSb (001)[14] is the class of $(n \times 3)$ reconstructions observed over a range of x on $In_xGa_{1-x}As$ (001)[17, 18]. The development of the methods in this dissertation and the application of them to the GaAs (001) and InAs/GaAs(001) surface can be contextualized by the puzzle presented by the $(n \times 3)$ surface. The naming of the $(n \times 3)$ surface arises from its indeterminate periodicity along $[1\bar{1}0]$. RHEED measurements show an apparent periodicity of $n = 1$ or $n = 2$, although STM measurements indicate that the periodicity is actually ill-defined, ranging from $n = 3$ to $n = 6$, but that on average $n = 4$ [18]. STM of the $(n \times 3)$ surface shows that it sometimes shares the surface with nano-scale domains of a well-characterized (2×4) reconstruction, and evidence suggests that this coexistence is strain-stabilized via interaction with the substrate[19]. Consequently, the range of parameters over which the $(n \times 3)$ is thermodynamically stable as a pure phase may be quite small.

1.3 Motivation and Organization of Topics

Where poorly characterized surfaces, such as the $(n\times3)$ exist, these gaps in understanding prevent the formulation of comprehensive thermodynamic and kinetic models of important phenomena (e.g., epitaxial growth, morphological evolution, alloy segregation, or surface diffusion). In addition to experimental characterization techniques, theoretical and computational methods can be used to inform the characterization process either by predicting the thermodynamic stability of hypothesized structural models with respect to well-accepted models, or by predicting the characterization measurements for a hypothesized model and comparing these to experimental results. Chapter II describes the theoretical and computational tools that can be used to compare thermodynamic stability of structural models and predict the finite-temperature properties of a specific structural model.

Regardless of the methods used, the path to identifying a structural hypothesis as the correct model that explains a poorly-characterized surface reconstruction is arduous and

paved by trial and error. Starting from experimental measurements, it is sometimes difficult to determine much more than the lattice vectors of the surface unit cell and general compositional trends. Even guided by the electron counting rule, the number of hypothetical structures that satisfy these minimal constrains can be insurmountable. Predicting thermodynamic stability reliably depends on computationally-intensive DFT calculations; without a systematic and physically-informed method to narrow the search for structural models, theoretical determination of equilibrium surface structure can become intractable. Chapter III describes a new approach that systematizes the enumeration of structural hypotheses and then narrows them down to the most likely candidates. Applying this method to the GaAs (001) surface reveals low-energy candidates for the $(n \times 3)$ surface. Much of the work described in Ch. III has been published previously in the journal *Physical Review B*[20].

Predicting the thermodynamic stability of a surface structure at finite temperature becomes complicated by the presence of thermal disorder, which must be accounted for in a rigorous way. Moreover, if the difference in surface free energies of two hypothesized structures is small, small changes in the surface free energy arising, for example, from calculation error can have a large effect on transition temperatures. Chapter IV describes the use of coarse-grained first principles energy calculations to account for thermal vibrations and configurational disorder and predict the finite temperature phase diagram for As-rich surface reconstructions of GaAs(001). Experimental results, in conjunction with the codified structure enumeration of Ch. III, are used to calibrate the error in calculated surface free energies arising from DFT approximations. The comprehensive analysis of surface reconstruction stability predicts that a (4×3) surface reconstruction is stable on GaAs (001) within the calculation error bounds, in agreement with experimental observation.

Deposition of an alloy species on the surface can induce alloy order and enhance structural order, especially if the alloying species have different equilibrium bond-lengths in the host material. This atomic-size mismatch effect can cause certain types of surface order to persist at temperatures beyond that of the pure material. Chapter V describes predictions of alloy-induced surface order when an InAs wetting layer is deposited on GaAs(001). Although alloying only occurs on the cation sublattice, it interacts strongly with the As sublattice by destablizing As dimer adsorption and enhancing As dimer order in the class of III–V (001) (2×4) reconstructions. Much of the work described in Ch. V has been published previously in the journal *Physical Review B*[21].

Although the analysis in Ch. V yields valuable insights into the role of finite-temperature alloying in an individual structural surface phase, it is important to place these results in context of the other stable reconstructions of the InAs/GaAs(001) wetting layer. Chapter VI describes use of the cluster expansion formalism with first principles methods to perform a

comprehensive search for ground-state reconstructions of the alloyed InAs/GaAs(001) surface and construct a zero-temperature phase diagram to predict the stability of the (2×4), (4×3), c(4×4) and (4×2) phases of the alloyed surface. Application these comprehensive search tool to such a complex system reveals the significant qualitative changes to structure and alloying induced by alloying at the surface.

Chapter VII discusses the relevance of the methods explored in the other chapters to surface systems other than III–V (001) and avenues for future developments.

CHAPTER II

Linking Thermodynamics of Multi-component Surfaces to First Principles Calculations

Many of the major results of surface thermodynamics were derived in the nineteenth century by Josiah W. Gibbs[22]. In fact, the treatment by Gibbs is still cited as the definitive source on the thermodynamics of an isotropic surface. Later contributions have extended the formalism of Gibbs to include considerations for anisotropic surfaces and multiple components. Of these, the description of surface energy advanced by Cahn is of particular importance[23], as it incorporated much of the understanding of surface thermodynamcs up to that point, and specified in clear language how to define properties of the surface in an unambiguous way. We shall follow this outline presented by Cahn in our development of the basic thermodynamic theory for surfaces.

A central concern of this dissertation is to rigorously calculate equilibrium properties of crystalline surfaces from first principles in order to compare the relative stability of similar atomic surface structures. Unfortunately, this application of surface thermodynamics is somewhat outside the scope of authoritative sources in the surface thermodynamics literature. This chapter presents a consistent theory and methodology for investigating the finite-temperature equilibrium behavior of surfaces using first principles methods. First, we derive the basic thermodynamic theory for a multicomponent solid surface. Then we describe the cluster expansion formalism and demonstrate how it can be used to construct an effective Hamiltonian for the calculation of the finite-temperature surface properties. Finally, we describe the first-principles methods used to parameterize effective Hamiltonians for the (001) surface of GaAs and its surface alloys.

2.1 Thermodynamics of a Multicomponent Surface

Although interface problems arise in homogeneous phases, as occurs at the grain boundaries of a polycrystalline material, the presence of a material interface generally suggests the presence of two coexisting phases that are in equilibrium with each other. In the case of immediate interest here the coexisting phases consist of a M–component substitutional crystal and a vapor phase containing the elemental precursors of the crystal. The vapor and the solid each has the $M+2$ degrees of freedom associated with addition/removal of thermal energy, mechanical work via changes in volume, and chemical work associated with adding/removing atoms of the M elemental species.

From these considerations, we can use the first and second laws of thermodynamics to write an equation for the reversible change in total energy of the vapor phase as

$$\mathrm{d}E^{(v)} = T^{(v)}\,\mathrm{d}S^{(v)} + \sum_m \mu_m^{(v)}\,\mathrm{d}N_m^{(v)} - P^{(v)}\,\mathrm{d}V^{(v)}, \tag{2.1}$$

where T is temperature, S is entropy, μ_m and N_m are the chemical potential and number of species m, P is pressure, and V is volume. Superscript v indicates that the variables are defined for the vapor phase. The reversible change in the energy of the solid can be similarly written as

$$\mathrm{d}E^{(s)} = T^{(s)}\,\mathrm{d}S^{(s)} + \sum_m \mu_m^{(s)}\,\mathrm{d}N_m^{(s)} - P^{(s)}\,\mathrm{d}V^{(s)}, \tag{2.2}$$

where superscript s indicates quantities of the solid phase.

For simulating solid materials, and crystalline solids in particular, it is convenient to introduce a change of variables. Defining $N_{(tot)} = \sum_{m=1}^{M} N_m$, we can write $N_M = N_{tot} - \sum_{m=1}^{M-1} N_m$ and $\mathrm{d}N_M = \mathrm{d}N_{tot} - \sum_{m=1}^{M-1} \mathrm{d}N_m$. Thus, Eq. (2.1) becomes

$$\mathrm{d}E^{(v)} = T\,\mathrm{d}S^{(v)} + \mu_M\,\mathrm{d}N_{tot}^{(v)} + \sum_{m=1}^{M-1} \Delta\mu_m\,\mathrm{d}N_m^{(v)} - P\,\mathrm{d}V^{(v)}. \tag{2.3}$$

and Eq. (2.2) becomes

$$\mathrm{d}E^{(s)} = T\,\mathrm{d}S^{(s)} + \mu_M\,\mathrm{d}N_{tot}^{(s)} + \sum_{m=1}^{M-1} \Delta\mu_m\,\mathrm{d}N_m^{(s)} - P^{(s)}\,\mathrm{d}V^{(s)}, \tag{2.4}$$

where $\Delta\mu_m = \mu_m - \mu_M$. This change of variables allows us to control the size of the system and its stoichiometry independently, since $\mu_M = (\partial E/\partial N_{tot})_{N_{m\neq M},S,V}$ while $\Delta\mu_{m'} =$

$(\partial E/\partial N_{m'})_{N_{tot},N_{m\neq m'\neq M},S,V}$. Subscripts of the partial derivatives indicate quantities that are held constant as the independent variable is changed. For simulations of crystalline solids, such a distinction has clear advantages by allowing the definition of thermodynamic boundary conditions that fix N_{tot} (corresponding, for example, to a fixed number of lattice sites) and leave the compositions $c_m = N_m/N_{tot}$ free to vary. Under such conditions, N_{tot} and the various $\Delta\mu_m$ would be the independent variables, while μ_M and c_m would be determined by equilibrium criteria imposed by the boundary conditions.

In order to specify criteria for solid/vapor equilibrium we will make a few simplifying assumptions about the geometry of the coexisting system. First, we assume that the surface is planar. The original description of surface tension for fluid interfaces set forth by Gibbs allowed for curved interfaces, with the result that a curved interface is associated with a discontinuity in pressure across the interface. We shall consider only atomically flat surfaces, for which curvature effects are vanishingly small, allowing us to disregard them. Second, we assume a fixed surface orientation. Although Cahn neglected interface curvature, he included instead surface orientation as an extensive variable. The results derived here will be applied to low-index (001) surface of zincblende III–V materials, which is predicted to be a thermodynamically allowed facet over the range of surface composition[24], so faceting should not occur.

Criteria for two-phase equilibrium within these constraints can be readily obtained for the solid and vapor, with an interface along the z-direction. Criteria can be derived by performing a constrained minimization of the total energy for a thermally, mechanically, and chemically isolated heterogeneous system containing both solid and vapor[25, 26]. The minimization yields criteria for the intensive variables specifying $T = T^{(v)} = T^{(s)}$, $\mu_m = \mu_m^{(v)} = \mu_m^{(s)}$, and $P = P^{(v)} = P^{(s)}$. Together, these criteria specify conditions on all of the degrees of freedom of the two phases.

The equilibrium criteria have physical ramifications on our ability to control the system. We can integrate Eqs. (2.3) and (2.4) by recognizing that the equations are homogeneous equations of degree one with respect to the extensive variables S, N_{tot}, $N_{m\neq M}$, and V. In other words, the energy of each homogeneous phase scales linearly with its quantity. Integration produces the Euler homogeneous forms

$$E^{(v)} = TS^{(v)} + \mu_M N_{tot}^{(v)} + \sum_{m=1}^{M-1} \Delta\mu_m N_m^{(v)} - PV^{(v)} \tag{2.5}$$

and

$$E^{(s)} = TS^{(s)} + \mu_M N_{tot}^{(s)} + \sum_{m=1}^{M-1} \Delta\mu_m N_m^{(s)} - P^{(s)} V^{(s)}. \tag{2.6}$$

We can take the total differentials of Eqs. (2.5) and (2.6), where the total differential of each conjugate variable pair X and Y is $\mathrm{d}(XY) = X\,\mathrm{d}Y + Y\,\mathrm{d}X$. By forcing equality with Eqs. (2.3) and (2.4), we arrive at the Gibbs-Duhem equations

$$0 = S^{(v)}\,\mathrm{d}T + N_{tot}^{(v)}\,\mathrm{d}\mu_M + \sum_{m=1}^{M-1} N_m^{(v)}\,\mathrm{d}\Delta\mu_m - V^{(v)}\,\mathrm{d}P \tag{2.7}$$

and

$$0 = S^{(s)}\,\mathrm{d}T + N_{tot}^{(s)}\,\mathrm{d}\mu_M + \sum_{m=1}^{M-1} N_m^{(s)}\,\mathrm{d}\Delta\mu_m - V^{(s)}\,\mathrm{d}P. \tag{2.8}$$

The two equations in $M+2$ variables can be reduced to a single equation in $M+1$ variables by eliminating a single variable. For example, by eliminating $\mathrm{d}\mu_M$, we obtain

$$0 = \Delta s\,\mathrm{d}T + \sum_{m=1}^{M-1} \Delta c_m\,\mathrm{d}\Delta\mu_m - \Delta\Omega\,\mathrm{d}P. \tag{2.9}$$

where s is the entropy per atom, c_m is the concentration of species m, and Ω is the volume per atom; $\Delta x = x^{(s)} - x^{(v)}$ for each quantity x. Eq. (2.9) is related to the Clausius-Clapeyron relations and is essentially a statement of the Gibbs phase rule. Mathematically, (2.9) specifies an M-dimensional surface in the thermodynamic space of intensive variables on which coexistence of vapor and solid is allowed. In other words, when two phases are in equilibrium, as is *always* the case for the equilibrium interface of a heterogeneous system, two intensive parameters are linearly dependent on the others and therefore cannot be controlled.

The physical system has, in addition to a region of homogeneous solid and one of homogeneous vapor, a region in the vicinity of the interface separating the two regions that behaves differently from either homogeneous region. Although the matter within this surface layer is influenced by the interface, the equilibrium criteria are still well-defined[25, 26]; the temperature, pressure, and chemical potentials have the same values near the interface as they do in the homogeneous systems. We can write the energy differential of the interface

region from the first and second laws of thermodynamics as

$$\mathrm{d}E^{(i)} = T\,\mathrm{d}S^{(i)} + \mu_M\,\mathrm{d}N^{(i)}_{tot} + \sum_{m=1}^{M-1} \Delta\mu_m\,\mathrm{d}N^{(i)}_m - P\,\mathrm{d}V^{(i)} + \gamma a_0 \mathrm{d}N_A, \tag{2.10}$$

where γ is the surface energy per unit area a_0 of the solid-vapor interface. The extensive variable N_A is the amount of interface, in terms of a_0, so that the total area $A = N_A a_0$.

Equation (2.10) is a homogeneous equation of degree one with respect to scaling due to increasing or decreasing the area. However, it is *not* homogeneous with respect to scaling in the direction perpendicular to the interface. Scaling in the perpendicular direction could be achieved by extending the interface farther into vapor phase, farther into the substrate, or some combination of the two. Consequently, the Euler integral form of Eq. (2.10) has the form

$$\tilde{E}^{(i)} = T\tilde{S}^{(i)} + \mu_M \tilde{N}^{(i)}_{tot} + \sum_{m=1}^{M-1} \Delta\mu_m \tilde{N}^{(i)}_m - P\tilde{V}^{(i)} + \gamma, \tag{2.11}$$

where the tilde denotes that the quantities are measured per unit area. The extensive quantities in Eq. (2.11) necessarily depend on the thickness $L^{(i)}$ of the interface region and the definitions of its boundaries. In fact, $\tilde{V}^{(i)}$ is simply $V^{(i)}/A = L^{(i)}$. However, Euler's theorem for homogeneous equations does not reveal any information about this dependence, due to the inhomogeneity of $E^{(i)}$ perpendicular to the surface.

If we choose $L^{(i)}$ sufficiently large, the inhomogeneity of $E^{(i)}$ perpendicular is due entirely to the matter that is directly influenced by the surface. Outside of some thickness $\lambda < L^{(i)}$, the material either behaves homogeneously as the vapor phase or as the solid phase. This concept is illustrated in Fig. 2.1. Extending $L^{(i)}$ farther into the vapor side by the infinitesimal amount $\mathrm{d}\ell$ increases $\tilde{E}^{(i)}$ by $(\mathrm{d}\tilde{E}^{(i)}/\mathrm{d}\ell)\,\mathrm{d}\ell$, or simply $\mathcal{E}^{(v)}\,\mathrm{d}\ell$, where $\mathcal{E}^{(v)}$ is the internal energy per unit volume of the vapor phase. Similarly, extending the boundary of the interface region further into the substrate side increases $\tilde{E}^{(i)}$ by an amount $\mathcal{E}^{(s)}\,\mathrm{d}\ell$. Thus we can separate $\tilde{E}^{(i)}$ as the sum of three integrals along the length coordinate ℓ perpendicular to the surface:

$$\tilde{E}^{(i)} = \int_0^{L^{(s)}} \mathrm{d}\ell\,\mathcal{E}^{(s)} + \int_{L^{(s)}}^{L^{(s)}+\lambda} \mathrm{d}\ell\,\frac{\mathrm{d}\tilde{E}^{(i)}}{\mathrm{d}\ell} + \int_{L^{(s)}+\lambda}^{L^{(i)}} \mathrm{d}\ell\,\mathcal{E}^{(v)}, \tag{2.12}$$

where the second integral is taken over the inhomogeneous layer of thickness λ that encloses the surface.

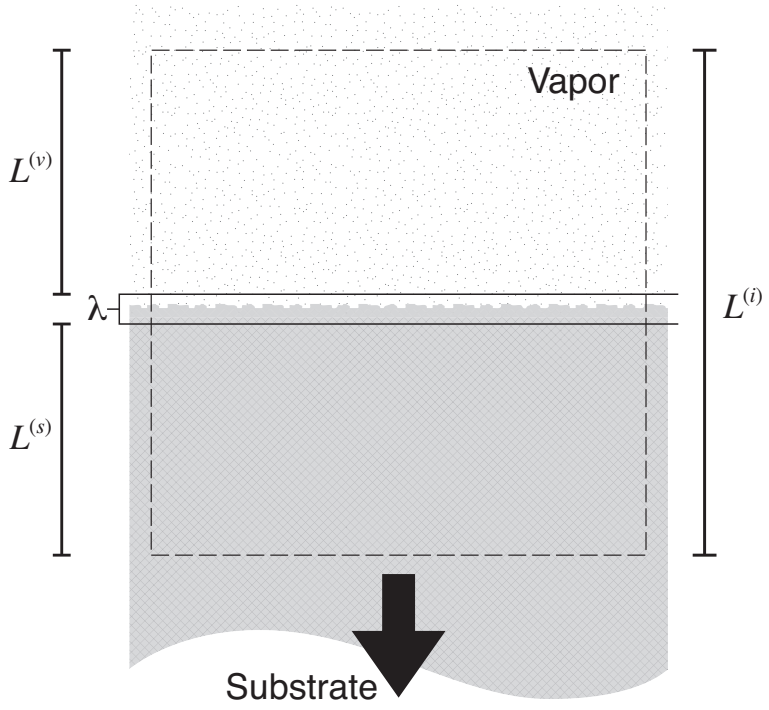


Figure 2.1. Schematic diagram of the idealized interface region, which has a well-defined solid-like region and vapor-like region, in addition to the surface region, which has thickness λ. The value of λ depends on the extensive quantities used to partition the interface region.

The integrals in Eq. (2.12) can be expressed simply as the sum

$$\tilde{E}^{(i)} = L^{(s)}\mathcal{E}^{(s)} + L^{(v)}\mathcal{E}^{(v)} + \tilde{E}^{(\lambda)} \tag{2.13}$$

where $\tilde{E}^{(\lambda)}$ is the energy per area of the surface layer. Rearranging yields

$$\tilde{E}^{(\lambda)} = \tilde{E}^{(i)} - L^{(s)}\mathcal{E}^{(s)} - L^{(v)}\mathcal{E}^{(v)}, \tag{2.14}$$

which has units of energy per unit area. We can determine the integral form of $\tilde{E}^{(\lambda)}$ by inserting Eqs. (2.5), (2.6), and (2.11) into Eq. (2.14) to obtain

$$\tilde{E}^{(\lambda)} = T\tilde{S}^{(\lambda)} + \mu_M \tilde{N}_{tot}^{(\lambda)} + \sum_{m=1}^{M-1} \Delta\mu_m \tilde{N}_m^{(\lambda)} - P\tilde{V}^{(\lambda)} + \gamma. \tag{2.15}$$

The extensive quantities in the surface layer are also measured per unit area and are defined as

$$\tilde{X}^{(\lambda)} = \tilde{X}^{(i)} - L^{(v)}x^{(v)} - L^{(s)}x^{(s)} \tag{2.16}$$

for a quantity X, where $x^{(v)}$ and $x^{(s)}$ are measured per unit volume in the current state of the respective phases.

Because the thickness of the surface layer is ill-defined, Eq. (2.15) is of limited usefulness in its present form, and its relation to the surface energy γ is difficult to discern. We can isolate γ from Eq. (2.15) to obtain

$$\gamma = \tilde{E}^{(\lambda)} - T\tilde{S}^{(\lambda)} - \mu_M \tilde{N}_{tot}^{(\lambda)} - \sum_{m=1}^{M-1} \Delta\mu_m \tilde{N}_m^{(\lambda)} + P\tilde{V}^{(\lambda)} \tag{2.17}$$

Since $\mathrm{d}\tilde{E}^{(i)}$, $\mathrm{d}\mathcal{E}^{(v)}$, and $\mathrm{d}\mathcal{E}^{(s)}$ are known, we can also write its differential,

$$\mathrm{d}\gamma = -\tilde{S}^{(\lambda)}\,\mathrm{d}T - \tilde{N}_{tot}^{(\lambda)}\,\mathrm{d}\mu_M + \sum_{m=1}^{M-1} \tilde{N}_m^{(\lambda)}\,\mathrm{d}(\Delta\mu_m) + \tilde{V}^{(\lambda)}\,\mathrm{d}P, \tag{2.18}$$

which indicates that γ explicitly depends on all the intensive parameters of the system. However, as demonstrated by Eq. (2.9), allowed values of the intensive variables exist implicitly on an M-dimensional surface, such that two intensive variables must be determined by the other M in order to equilibrate the heterogeneous system. Since the equilibrium surface free energy cannot explicitly depend on all $M+2$ intensive parameters, it should be expressed in terms of only M intensive variables, and by requiring this for the expression of γ we can assert the equilibrium criteria for phase coexistence.

Reducing the dependence of Eq. (2.18) can be achieved by requiring that the *extensive* quantities conjugate to two of the intensive parameters go to zero. In other words, we require

$$\tilde{Y}^{(\lambda)} = \tilde{Y}^{(i)} - L^{(v)}y^{(v)} - L^{(s)}y^{(s)} = 0 \tag{2.19}$$

and

$$\tilde{Z}^{(\lambda)} = \tilde{Z}^{(i)} - L^{(v)}z^{(v)} - L^{(s)}z^{(s)} = 0 \tag{2.20}$$

for the two extensive quantities Y and Z. Equations (2.19) and (2.20), along with Eq. (2.16) form a system of three equations in the three unknowns, $L^{(v)}$, $L^{(s)}$, and $\tilde{X}^{(\lambda)}$. We can use

Cramer's rule to solve for $\tilde{X}^{(\lambda)}$ in terms of determinants as

$$[X]_{Y,Z} \equiv \tilde{X}^{(\lambda)} = \frac{1}{A} \frac{\begin{vmatrix} \tilde{X}^{(i)} & x^{(v)} & x^{(s)} \\ \tilde{Y}^{(i)} & y^{(v)} & y^{(s)} \\ \tilde{Z}^{(i)} & z^{(v)} & z^{(s)} \end{vmatrix}}{\begin{vmatrix} y^{(v)} & y^{(s)} \\ z^{(v)} & z^{(s)} \end{vmatrix}}. \tag{2.21}$$

We have borrowed the notation $[X]_{Y,Z}$ from Cahn to denote the extensive quantity X contained in the surface layer when the intensive parameters conjugate to $\tilde{Y}^{(\lambda)}$ and $\tilde{Z}^{(\lambda)}$ are eliminated from Eq. (2.17). The quantity $[X]_{Y,Z}$ is called the *surface excess* of X. It is clear from properties of determinants that both $[Y]_{Y,Z}$ and $[Z]_{Y,Z}$ are zero, and thus their intensive conjugates no longer appear Eq. (2.18). The choice of Y and Z implicitly specifies the thickness and placement of the surface layer within the interface region, since $L^{(v)}$ and $L^{(s)}$ are

$$L^{(v)} = \frac{\begin{vmatrix} \tilde{Y}^{(i)} & y^{(s)} \\ \tilde{Z}^{(i)} & z^{(s)} \end{vmatrix}}{\begin{vmatrix} y^{(s)} & y^{(v)} \\ z^{(s)} & z^{(v)} \end{vmatrix}}, \text{ and } \qquad L^{(s)} = \frac{\begin{vmatrix} \tilde{Y}^{(i)} & y^{(v)} \\ \tilde{Z}^{(i)} & z^{(v)} \end{vmatrix}}{\begin{vmatrix} y^{(v)} & y^{(s)} \\ z^{(v)} & z^{(s)} \end{vmatrix}}. \tag{2.22}$$

However, the geometric interpretation of $L^{(v)}$ and $L^{(s)}$ does not affect the physical meaning of the measurable quantity γ.

With $L^{(v)}$ and $L^{(s)}$ now defined in a way that is consistent with solid-vapor equilibrium, the physical relevance of γ is more readily apparent. Using the notation of Eq. (2.21), $\mathrm{d}\gamma$ is written

$$\mathrm{d}\gamma = -[S]_{Y,Z}\,\mathrm{d}T - [N_{tot}]_{Y,Z}\,\mathrm{d}\mu_N - \sum_{m=1}^{M-1} [N_m]_{Y,Z}\,\mathrm{d}(\Delta\mu_m) + [V]_{Y,Z}\,\mathrm{d}P. \tag{2.23}$$

The quantities $[X]_{Y,Z}$ can be interpreted as the excess amount of X contained in the interface region due to the presence of the surface, when the surface layer is defined such that $\tilde{Y}^{(\lambda)}$ and $\tilde{Z}^{(\lambda)}$ are zero. For example, $[V]_{Y,Z}$ is the surface excess volume. According to this interpretation, it is readily apparent that $\lambda \equiv [V]_{Y,Z}$, and for $Y = V$ or $Z = V$, the surface excess quantities are defined for an idealized two-dimensional surface, with $\lambda = 0$.

Although surface excess quantities are in general dependent on the choice of independent variables, the definition of γ itself is not arbitrary. If we choose for example $Y = N_{(tot)}$ and

$Z = V$, then Eq. (2.17) becomes

$$\gamma = [E]_{N_{tot},V} - T[S]_{N_{tot},V} - \sum_{m=1}^{M-1} \Delta\mu_m [N_m]_{N_{tot},V}. \tag{2.24}$$

This expression is in fact equivalent to the surface excess of the grand potential

$$\gamma = [\Phi]_{N_{tot},V} = \tilde{\Phi}^{(i)} - L^{(v)}\varphi^{(v)} - L^{(s)}\varphi^{(s)}, \tag{2.25}$$

where the grand potential per unit area of the interface region is

$$\tilde{\Phi}^{(i)} = \tilde{E}^{(i)} - T\tilde{S}^{(i)} - \sum_{m=1}^{M-1} \Delta\mu_m \tilde{N}_m^{(i)}. \tag{2.26}$$

The grand potential is a useful definition of thermodynamic free energy in many atomistic simulations of solids. It is the characteristic potential for boundary conditions that fix the total total number of atoms and volume, which can often simplify the implementation of a simulation. In equilibrium at fixed volume and number of atoms, $\Omega^{(v)}\varphi^{(v)} = \Omega^{(s)}\varphi^{(s)}$, where Ω is volume per atom. $\varphi^{(v)}$ is the grand canonical potential of the vapor per unit volume, given by

$$\phi^{(v)} = \varepsilon^{(v)} - Ts^{(v)} - \sum_{m=1}^{M_1} \Delta\mu_m c_m^{(v)}, \tag{2.27}$$

where $\varepsilon^{(v)}$ and $s^{(v)}$ are the energy and entropy of the vapor, per volume; $c_m^{(v)} = N_m^{(v)}/V^{(v)}$ is the concentration of species m in the vapor. $\varepsilon^{(s)}$ for the solid is defined similarly. Equation (2.25) consequently reduces to

$$\gamma = [\Phi]_{N_{tot},V} = \tilde{\Phi}^{(i)} - \frac{\Omega^{(s)}L^{(v)} + \Omega^{(v)}L^{(s)}}{\Omega^{(v)}\Omega^{(s)}}\varphi^{(s)} \tag{2.28}$$

By calculating $L^{(v)}$ and $L^{(s)}$ from Eqs. (2.22), we can eliminate the dependence of γ on $\Delta\mu_M$ and P and obtain the simplified expressions

$$\begin{aligned} \gamma &= \tilde{\Phi}^{(i)} - \tilde{N}^{(i)}\Omega^{(s)}\varphi^{(s)} \\ &= \tilde{\Phi}^{(i)} - \tilde{N}^{(i)}\Omega^{(v)}\varphi^{(v)}. \end{aligned} \tag{2.29}$$

In other words, to determine γ for the surface it is sufficient to calculate the characteristic potential for the interface region, using some simulation method, and then take the difference

with the same characteristic potential calculated for one of the pure phases, at the same set of intensive parameters.

It is generally true that if the characteristic potential depends on either one or two extensive variables only, then we can express γ as the appropriate surface excess of the characteristic potential. For example, in a binary system with independent variables N_1, N_2, T, and P, we find $\gamma = [G]_{N_1,N_2}$, where $[G]_{N_1,N_2}$ is the surface excess Gibbs free energy. For a system with characteristic potential Λ and only one fixed extensive parameter (for instance, $N_{(tot)}$), $\gamma = [\Lambda]_{N_{tot},Z}$, where Z is any other extensive parameter. As indicated by Eq. (2.29), the characteristic potential need only be calculated for one of the pure phases to determine γ in such cases. If surface excess of other extensive quantities are desired, however, they need to be known for both phases.

In many cases, the vapor phase has a negligible density relative to the bulk solid. In such situations, an important limiting case can be obtained for Eq. (2.21). When we eliminate P and μ_M from Eq. (2.17), we can rewrite $[X]_{N_{tot},V}$ as

$$[X]_{N_{tot},V} = \tilde{X}^{(i)} - \tilde{N}^{(i)}\Omega^{(s)}\frac{x^{(s)} - x^{(v)}}{1 - \Omega^{(s)}\rho^{(v)}} - \tilde{V}^{(i)}\frac{x^{(v)} - x^{(s)}\Omega^{(s)}\rho^{(v)}}{1 - \Omega^{(s)}\rho^{(v)}}, \tag{2.30}$$

where $\rho^{(v)}$ is the volume density of atoms in the vapor phase. As the density of the vapor phase approaches zero, $x^{(v)}$ (which scales as $\rho^{(v)}$) and $\Omega^{(s)}\rho^{(v)}$ become vanishingly small, leaving only

$$[X]_{N_{tot},V} = \tilde{X}^{(i)} - \tilde{N}^{(i)}\Omega^{(s)}x^{(s)}, \tag{2.31}$$

which is the approximation of a solid–vacuum interface. Although such an interface cannot exist in equilibrium, it is nearly comparable to a solid surface exposed to a very low-density vapor phase. Such situations arise frequently in epitaxial crystal growth, and as a result, Eq. (2.31) is widely used in the literature to express excess quantities calculated for epitaxial and similar systems, often without calling to attention that it is an approximation. Equation (2.29), however, remains valid despite the approximation demonstrates the advantage of working in terms of the characteristic potential.

In general, γ is not the only thermodynamic potential that can be used to describe the surface region. We may take a Legendre transform of $[E]_{N_{tot},V}$ such as

$$\begin{aligned}\eta = [G]_{N_{tot},V} &= [E]_{N_{tot},V} - T[S]_{N_{tot},V} + P[V]_{N_{tot},V} \\ &= \gamma + \sum_{m=1}^{M-1}\Delta\mu_m[N_m]_{N_{tot},V},\end{aligned} \tag{2.32}$$

in which $P[V]_{N_{tot},V} = 0$, and find

$$\mathrm{d}\eta = -[S]_{N_{tot},V}\,\mathrm{d}T + \sum_{m=1}^{M-1} \Delta\mu_m\,\mathrm{d}[N_m]_{N_{tot},V}, \tag{2.33}$$

allowing us to compare the energies of different surfaces under the somewhat artificial constraint of fixed surface composition. Equilibrium criteria are then satisfied by explicitly requiring equality of $\Delta\mu_m$ everywhere in the interface. Determining equilibrium at fixed surface composition is equivalent to the common-tangent construction used to determine multicomponent phase equilibria. Equation (2.32) can significantly simplify analysis and visualization of surface energy data when many different surface structures are under consideration, as will become clear in Ch. III.

2.2 Relating Surface Thermodynamics to Atomic Surface Structure and Configuration

The thermodynamic results of the previous section are rigorously correct for macroscopic systems measured at equilibrium. In principle, Sec. 2.1 implied that, in the low vapor density limit, we can predict properties of the surface by separately simulating equilibrium behavior of the material in two different geometries. The first is simply the ideal bulk crystal; the second is a region of material that contains the interface but also extends to a depth in the bulk substrate at which the material behaves approximately as the equilibrium crystal. However, Sec. 2.1 made no prescription for how to parameterize the constitutive relations of the bulk solid or interface region. The constitutive relations link the independent thermodynamic parameters of the system, as determined by its boundary conditions, to its measurable properties. As observed in Eq. (2.29), if constitutive relations are known for the entire interface region as well as for the solid bulk, the surface free energy is easily obtained. Functional forms for constitutive relations can be derived for idealized cases to relate, for example, entropy to composition, but this is often not possible even in simple cases for real materials. One strategy to obtain constitutive relations for a real material is to parameterize these idealized models using experimental data. Increasingly, however, calculating constitutive relations outright, from the atomistic scale, is a tractable alternative that can often cover a wider region of thermodynamic phase space while also providing significant insight into the mechanisms from which equilibrium phenomena arise.

Although equilibrium thermodynamics is based on the concept of thermodynamic "state", which is independent of the path taken to reach equilibrium and depends only on independent

macroscopic parameters, an atomistic system also has an instantaneous state that depends on its microscopic degrees of freedom. The instantaneous energy of this microscopic state is an eigenvalue of the Schrödinger equation corresponding to the inseparable many-body eigenfunction that describes the state of all ions and electrons in the material. For most solid-state materials (and especially semiconductors and insulators) we can accept the validity of the adiabatic approximation, which assumes that electronic excitations are sufficiently far away from the ground state that the electronic wavefunction always remains in its lowest eigenstate, which is wholly determined by the ionic positions. Moreover, in a crystalline material, the ions can be thought of as trapped within a sufficiently deep local minimum of the potential energy surface that the coordinates of the minimum energy state map onto well-defined crystal sites that can be labeled. The ionic excitations of a crystal are small perturbations of the ionic positions and momenta about the crystal sites. One of M chemical species resides at each site, and because large excitations of the sites, including exchanges of atoms between sites, are assumed to be infrequent, the instantaneous microstate of the crystal is associated with a well-defined configuration specifying the species at each site. To describe configuration mathematically, we assign each site j an occupation variable, σ_j, that can take one of M discrete values, depending on the species occupying the site. The instantaneous microstate energy of the crystal can therefore be written

$$E_{\vec{\sigma},\vec{n}} = E_0(\vec{\sigma}) + \Delta E(\vec{\sigma},\vec{n}), \tag{2.34}$$

where $\vec{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_j, \ldots, \sigma_{N_{tot}})$ is the vector of occupation variables for the N_{tot} crystal sites, and $\vec{n}$ is the set of quantum numbers that describe the quantized eigenstates of the many-body ionic Hamiltonian for configuration $\vec{\sigma}$. $E_0(\vec{\sigma})$ is the ground-state energy at the local minimum of ionic positions for configuration $\vec{\sigma}$, and $\Delta E(\vec{\sigma},\vec{n})$ is an excited eigenstate of the Hamiltonian in the vicinity of the local minimum.

The actual instantaneous microstate of the crystal at any given time is unknown, but when temperature is fixed we can assign a probability to the crystal being in a microstate with energy E via the canonical distribution

$$P(E) = \frac{g_E e^{-\beta E}}{Z}, \tag{2.35}$$

where $\beta = 1/k_B T$, and k_B is the Boltzmann constant. g_E is the number of microstates having energy E, and Z is the partition function, which, in order to normalize $P(E)$ is given

by

$$Z = \sum_E g_E e^{-\beta E}, \tag{2.36}$$

or equivalently

$$Z = \sum_{\vec{\sigma}} \sum_{\vec{n}} e^{-\beta E_{\vec{\sigma},\vec{n}}}. \tag{2.37}$$

The partition function has the important property that the Helmholtz free energy is given by

$$A = \langle E_{\vec{\sigma},\vec{n}} \rangle - TS = -k_B T \ln Z, \tag{2.38}$$

where angle brackets indicate an average over the canonical distribution.

When a system, due to its boundary conditions, can exchange energy with its environment via chemical or mechanical work the ensemble of microstates is described by a slightly different probability distribution, such that the partition function becomes

$$Z = \sum_{\vec{\sigma}} \sum_{\vec{n}} e^{-\beta H_{\vec{\sigma},\vec{n}}}. \tag{2.39}$$

The partition function of the new ensemble is expressed in terms of the generalized enthalpy, which is the Legendre transform

$$H_{\vec{\sigma},\vec{n}} = E_{\vec{\sigma},\vec{n}} - \sum_\eta \eta X^{(\eta)}_{\vec{\sigma},\vec{n}}, \tag{2.40}$$

where η is a fixed intensive parameter of the system and $X^{(\eta)}_{\vec{\sigma},\vec{n}}$ is the value in the $\{\vec{\sigma},\vec{n}\}$ microstate of its conjugate extensive variable. When the system moves from one microstate to another, the system performs an amount of work on its environment equal to $-\sum_\eta \eta \Delta X^{(\eta)}$. With this definition, the relation in Eq. (2.38) then becomes

$$\Phi = \langle H_{\vec{\sigma},\vec{n}} \rangle - TS = -k_B T \ln Z, \tag{2.41}$$

where Φ is the characteristic thermodynamic potential for the given boundary conditions. If Φ describes a chemically open system that can perform chemical work on the outside

environment, we can recover Eq. (2.26) by using

$$H_{\vec{\sigma},\vec{n}}(\Delta\mu) = E_{\vec{\sigma},\vec{n}} - \sum_{m=1}^{M-1} \Delta\mu_m N_{m,\vec{\sigma}}, \tag{2.42}$$

where the number of atoms of species m in a particular microstate is determined by the configuration.

If the excitations away from the ground-state eigenfunction of the ions in configuration $\vec{\sigma}$ occur at a much shorter timescale than the atom-exchange excitations that alter the configuration, the system will pass through many ionic excitations that can all be described with respect to a configuration $\vec{\sigma}$ before the system passes into a new configuration $\vec{\sigma}'$. This allows us to separate the sum over microstates in Eqn. (2.39) in a useful way[27]. Writing Eq. (2.39) explicitly as

$$Z = \sum_{\vec{\sigma}} \sum_{\vec{n}} e^{-\beta\left(E_0(\vec{\sigma}) + \Delta E(\vec{\sigma},\vec{n}) - \sum \Delta\mu_m N_{m,\vec{\sigma}}\right)}, \tag{2.43}$$

we can bring the terms that do not depend on the excitation eigenstates $\vec{n}$ outside the second summation to obtain

$$Z = \sum_{\vec{\sigma}} e^{-\beta\left(E_0(\vec{\sigma}) - \sum \Delta\mu_m N_{m,\vec{\sigma}}\right)} \sum_{\vec{n}} e^{-\beta\Delta E(\vec{\sigma},\vec{n})}. \tag{2.44}$$

The sum over $\vec{n}$ is essentially the partition function for ionic excitations of configuration $\vec{\sigma}$, which we will call $Z_\Delta(\vec{\sigma})$. If we define

$$F(\vec{\sigma}) = E_0(\vec{\sigma}) - k_B T \ln Z_\Delta(\vec{\sigma}) \tag{2.45}$$

as the free energy of configuration $\vec{\sigma}$ and note that $\exp(-\beta F) = Z_\Delta \exp(-\beta E_0)$, then we obtain

$$Z = \sum_{\vec{\sigma}} e^{-\beta\left(F(\vec{\sigma}) - \sum \Delta\mu_m N_{m,\vec{\sigma}}\right)}. \tag{2.46}$$

Equation (2.46) is the partition function for a chemically open crystalline material whose microstates are specified solely by the configuration of species over its crystal sites. The excitations occurring at time-scales shorter than that of species exchange, which are accounted for in the configuration free energy $F(\vec{\sigma})$, are assumed to arise predominantly from lattice vibrations, although any "fast" degree of freedom that leaves the configuration largely unchanged can be accounted for in $F(\vec{\sigma})$.

In general, the partition function cannot be calculated directly, even if $F(\vec{\sigma})$ is known, since the sum in Eq. (2.46) is taken over all possible configurations of an infinitely large crystal, and the number of possible configurations grows as M^N for M possible species arranged on N distinct sites. Fortunately, derivatives of the characteristic potential can be calculated directly via Metropolis Monte Carlo, which is used to numerically integrate moments of the ensemble probability distribution[28]. Explicit differentiation of Eq. (2.41) yields

$$\begin{aligned} -\left(\frac{\partial \Phi}{\partial \Delta\mu_m}\right)_{T,V,N_{tot},\Delta\mu_{m'\neq m}} &= \langle N_m \rangle \\ &= \frac{1}{Z}\sum_{\vec{\sigma}} N_{m,\vec{\sigma}}\, e^{-\beta\left(F(\vec{\sigma})-\sum \Delta\mu_m N_{m,\vec{\sigma}}\right)}, \end{aligned} \tag{2.47}$$

and

$$\begin{aligned} -\left(\frac{\partial \Phi}{\partial T}\right)_{V,N_{tot},\Delta\mu_m} &= S \\ &= \frac{1}{T}\left\langle F\left(\vec{\sigma}\right) - \sum \Delta\mu_m N_{m,\vec{\sigma}} \right\rangle + k_B \ln Z + \left\langle S_{\vec{\sigma}} \right\rangle, \end{aligned} \tag{2.48}$$

where $S_{\vec{\sigma}} = -(\partial F/\partial T)_{\vec{\sigma},V,N_{tot}}$ is the entropy of configuration $\vec{\sigma}$ due to short-timescale excitations. $\langle N_m \rangle$, $\langle F\left(\vec{\sigma}\right) \rangle$, and $\langle S_{\vec{\sigma}} \rangle$ can be calculated directly from Monte Carlo, since they are first moments of the ensemble probability distribution. However, $(k_B \ln Z)$ cannot be calculated directly. Therefore, in order to calculate Φ from Monte Carlo, we may integrate

$$\mathrm{d}\Phi = -S\,\mathrm{d}T - \sum_{m=1}^{M-1} N_m\,\mathrm{d}(\Delta\mu_m) + \mu_M\,\mathrm{d}N_{tot} - P\,\mathrm{d}V, \tag{2.49}$$

which is most easily accomplished along a path where $\mathrm{d}T = \mathrm{d}V = \mathrm{d}N_{tot} = 0$, in which case

$$\int_{\Phi_0}^{\Phi} \mathrm{d}\Phi = -\sum_{m=1}^{M-1} \int_{t_0}^{t} N_m \frac{\mathrm{d}(\Delta\mu_m)}{\mathrm{d}t'}\,\mathrm{d}t', \tag{2.50}$$

where $\Delta\mu_m(t')$ is the parameterization of $\Delta\mu_m$ along the path taken in thermodynamic space. Φ_0 is the value of the characteristic potential at the initial point on the path and is chosen to be a point where Φ is known—either in the limit $T \to 0$ or at an extreme of composition, where only one configuration is allowed.

2.3 Configurational Dependence of Free Energies Within the Cluster Expansion Formalism

When we described in the previous section how the moments of the ensemble probability distribution can be used to relate thermodynamic constitutive relations for the macroscopic variables of the interface region to its instantaneous states of species configuration, we neglected to consider how the free energy $F(\vec{\sigma}, T)$ of the configuration $\vec{\sigma}$ might be expressed in terms of the single-site occupation variables σ_j. In this section we describe the cluster expansion formalism[29] and demonstrate how it can be used to construct an accurate and compact effective Hamiltonian for describing the dependence of a material property on its configuration. In order for a system to be amenable to the cluster expansion formalism, we require that its crystal sites be well-defined and can be labeled $\{1, 2, \ldots, j, \ldots, N\}$. The occupation of the site j is described by the occupation variable σ_j, which maps the occupation state of the site onto a discrete value. If M different species are allowed to reside at site j, then

$$\sigma_j \in \begin{cases} \{-(M-1)/2, \ldots, 0, \ldots, (M-1)/2\}, & \text{if } M \text{ is odd, and} \\ \{-M/2, \ldots, -1, 1, \ldots, M/2\}, & \text{if } M \text{ is even.} \end{cases}$$

We can express an arbitrary function $f(\sigma_j)$ acting on σ_j as a linear combination of basis functions $\phi_j^{(k)}(\sigma_j)$ such that

$$f(\sigma_j) = \sum_k \alpha_k \phi_j^{(k)}(\sigma_j), \tag{2.51}$$

where α_k are scalar coefficients. Clearly a function of a M-valued variable can itself take at most M values. Thus Eq. (2.51) follows if we require

$$\left\langle \phi_j^{(k)}, \phi_j^{(k')} \right\rangle = \delta_{kk'}, \tag{2.52}$$

where $\left\langle \phi_j^{(k)}, \phi_j^{(k')} \right\rangle$ is an inner product defined on the space spanned by all functions of σ_j. The definition of the inner product is somewhat arbitrary, there has been a historical preference for one in particular[29], which is given by

$$\langle f(\sigma_j), g(\sigma_j) \rangle = \frac{1}{M} \sum_{\sigma_j} f^*(\sigma_j) g(\sigma_j), \tag{2.53}$$

where the sum is over the M allowed values of σ_j. Taken together, Eqs. (2.52) and (2.53)

Table 2.1. Basis function completeness table for orthonormal basis of a ternary site listed in Eqs. (2.54).

	$\phi_j^{(1)}$	$\phi_j^{(2)}$	$\phi_j^{(3)}$
Species 1	1	$-\sqrt{3/2}$	$-1/\sqrt{2}$
Species 2	1	0	$+\sqrt{2}$
Species 3	1	$+\sqrt{3/2}$	$-1/\sqrt{2}$

require that $\alpha_k = \left\langle \phi_j^{(k)}, f \right\rangle$ in Eq. (2.51). These concepts can be illustrated more clearly in Table 2.1, which shows a choice of orthonormal basis for a ternary site (i.e., $M = 3$) corresponding to the basis functions

$$\begin{aligned} \phi_j^{(1)}(\sigma_j) &= 1 \\ \phi_j^{(2)}(\sigma_j) &= \sqrt{\frac{3}{2}}\sigma_j \\ \phi_j^{(3)}(\sigma_j) &= \sqrt{2} - \frac{3}{\sqrt{2}}\sigma_j^2. \end{aligned} \tag{2.54}$$

If we use the entries in Table 2.1 to fill the rows and columns of a matrix A, then the completeness of the basis is guaranteed if A is full rank, and the orthonormality of the basis with respect to the inner product of Eq. (2.53) is guaranteed if $A^{-1} = A^\dagger/M$.

The set of orthonormal basis functions for site j, along with the inner product in Eq. (2.53) define an M-dimensional Hilbert space. The choice of $\phi_j^{(1)} = 1$ simplifies the construction of a basis set for the configuration of the entire N-site crystal, which defines the $N \times M$-dimensional Hilbert space formed by taking the tensor product

$$\left\{\Gamma^{(\kappa)}(\vec{\sigma})\right\} = \begin{Bmatrix} 1 \\ \phi_1^{(2)} \\ \vdots \\ \phi_1^{(M)} \end{Bmatrix} \otimes \begin{Bmatrix} 1 \\ \phi_2^{(2)} \\ \vdots \\ \phi_2^{(M)} \end{Bmatrix} \otimes \cdots \otimes \begin{Bmatrix} 1 \\ \phi_N^{(2)} \\ \vdots \\ \phi_N^{(M)} \end{Bmatrix}, \tag{2.55}$$

where $\Gamma^{(\vec{\kappa})}(\vec{\sigma})$ are basis functions of the entire crystal, in terms of its configuration $\vec{\sigma}$. The crystal basis functions take the form of the product

$$\Gamma^{(\vec{\kappa})}(\vec{\sigma}) = \prod_{j=1}^{N} \phi_j^{(\kappa_j)}(\sigma_j), \tag{2.56}$$

where κ_j specifies which site basis function of site j is included in the product. The crystal basis functions preserve the completeness relations of site basis function via

$$\langle \Gamma^{(\vec{\kappa}')}, \Gamma^{(\vec{\kappa}')} \rangle = \frac{1}{M^N} \sum_{\{\vec{\sigma}\}} \Gamma^{(\vec{\kappa}')*}(\vec{\sigma}) \Gamma^{(\vec{\kappa}')}(\vec{\sigma}) = \delta_{\vec{\kappa}'\vec{\kappa}}, \tag{2.57}$$

where the sum is taken over all M^N possible configurations of the N-site lattice.

For a given $\vec{\kappa}$, many $\phi_j^{(\kappa_j)}$ are equal to one, and do not need to be explicitly accounted for in the product. If we specify by α the set of sites $j \in \alpha$ for which $\kappa_j \neq 1$, we can rewrite the crystal basis functions as

$$\Gamma_\alpha^{(\vec{\kappa})}(\vec{\sigma}) = \prod_{j \in \alpha} \phi_j^{(\kappa_j)}(\sigma_j). \tag{2.58}$$

α may be a single site, a pair, a triplet, etc., or it may be the empty set, $\emptyset$, in which case $\Gamma_\emptyset^{(\vec{\kappa})} = 1$. The sorting of the crystal basis set into clusters is made possible by the initial choice of $\phi_j^{(1)} = 1$. By sorting the crystal basis in this manner, we may choose a restricted set of simple basis functions initially (corresponding to site clusters and geometrically compact pairs, for instance), and iteratively add more complex basis functions as needed. It is due to this property that the mathematical framework presented here is referred to as the cluster expansion formalism.

The completeness of the cluster expansion basis functions make it possible to rigorously expand any property that depends on the species configuration of a crystal as a linear combination of these basis functions.[29, 30, 31] Assuming that the configuration free energy $F(\vec{\sigma})$ can be calculated for an arbitrary configuration of the crystal, we can write it in terms of the $\Gamma_\alpha^{(\vec{\kappa})}$ as

$$F(\vec{\sigma}) = \sum_{\alpha, \vec{\kappa}} V_\alpha^{(\vec{\kappa})} \Gamma_\alpha^{(\vec{\kappa})}(\vec{\sigma}) \tag{2.59}$$

where the scalar coefficient $V_\alpha^{(\vec{\kappa})}$ is the effective cluster interaction (ECI) associated with $\vec{\kappa}$ basis function of cluster α.

Only a small fraction of the ECI are independent, since many clusters are symmetrically equivalent due to the translational periodicity of the crystal, as well as its factor group of symmetry elements. For most physical systems, it is not possible to make any further *a priori* assertions about the ECI. The cluster expansion in Eq. (2.59) is an exact representation of the configurational dependence of $F(\vec{\sigma})$, due to the completeness of the crystal basis. However, all basis functions must be included in the sum for the representation to remain exact.

Experience has borne out, however, that the cluster expansion can be truncated to only include terms associated with clusters containing no more than four or five sites and that are spatially compact relative to some physically relevant length scale of the system.

Although the ECI are formally described as a projection of $F(\vec{\sigma})$ onto the crystal basis functions via the inner product in Eq. (2.57), the inner product cannot be directly calculated for arbitrary functions, since the sum over all configurations of the crystal is infinite. Parameterizing ECI of a restricted basis is instead performed by performing a least-squares fit to a set of values of $F(\vec{\sigma})$ calculated at a finite collection of crystal configurations. A meaningful criterion for the fitness of the restricted basis set can be used to optimize the truncation of the expansion. A frequently used metric for the fitness of restricted basis set is the leave-one-out cross-validation score

$$CV^2 = \frac{1}{P}\sum_{p=1}^{P}\left(F(\vec{\sigma}_p) - \tilde{F}^{(p)}(\vec{\sigma}_p)\right)^2, \tag{2.60}$$

where $F(\vec{\sigma}_p)$ is calculated for the configuration p and $\tilde{F}^{(p)}(\vec{\sigma}_p)$ is the predicted value of $F(\vec{\sigma}_p)$ obtained by performing a least-squares fit to the data from the remaining $P-1$ configurations, not including $\vec{\sigma}_p$, and then evaluating the resulting cluster expansion at $\vec{\sigma}_p$. As the size of the data set grows large, the optimal set of crystal basis functions is that which minimizes the CV score[32].

2.4 Calculating Properties of a Surface Microstate from First Principles

In order to obtain and accurate value for the configuration free energy $F(\vec{\sigma})$ of configuration $\vec{\sigma}$, we must calculate energies for at least the most important short-time-scale excitations of the fixed configuration, given by $E_{\vec{\sigma},\vec{n}}$ in Eq. (2.34). The allowed excitations of the configuration and their energies are determined from the solutions to the many-body Schrödinger equation for the crystal[33],

$$H\psi_{\vec{\sigma},\vec{n}} = E_{\vec{\sigma},\vec{n}}\psi_{\vec{\sigma},\vec{n}}, \tag{2.61}$$

where $H(\mathbf{R_1},\ldots,\mathbf{R_I},\mathbf{r_1},\ldots,\mathbf{r_J})$ is the many-body Hamiltonian that describes interactions among the I ions, whose spacial coordinates are given by $\mathbf{R_i}$ and J electrons, having coordinates $\mathbf{r_j}$. $\psi_{\vec{\sigma},\vec{n}}(\mathbf{R_1},\ldots,\mathbf{R_I},\mathbf{r_1},\ldots,\mathbf{r_J})$ is the many-body quantum wavefunction of the configuration $\vec{\sigma}$ when in eigenstate $\vec{n}$.

Although the allowed eigenstates of the configuration and their energies can be deter-

mined by diagonalization of Eq. (2.61), this is a computationally-demanding procedure, even for small J and I. We can make a few simplifying assumptions for Eq. (2.61) to ease computation. The first is the Born-Oppenheimer approximation, which states that, due to the large difference in mass between ions and electrons, their wavefunctions are independent, and because the electrons, owing to their small mass, react so much more quickly than the ions, the ion positions can be assumed fixed with respect to the electronic wave function. Additionally, we assume that the ions can be treated as classical point particles, since their thermal wavelengths are quite small relative to the nearest-neighbor ionic distance. Given these assumptions, the J-body electronic Hamiltonian becomes

$$H = T + V_{ei} + V_{ee} + V_{ii}, \tag{2.62}$$

where, in atomic units,

$$T = -\frac{1}{2}\sum_{j} \nabla^2_{\mathbf{r_j}} \tag{2.63}$$

is the kinetic energy operator,

$$V_{ei} = -\sum_{i,j} \frac{Z_i}{|\mathbf{r_j} - \mathbf{R_i}|} \tag{2.64}$$

is the electron–ion Coulomb potential,

$$V_{ee} = \frac{1}{2}\sum_{j,j' \neq j} \frac{1}{|\mathbf{r_j} - \mathbf{r_{j'}}|} \tag{2.65}$$

is the electron–electron Coulomb potential, and

$$V_{ii} = \frac{1}{2}\sum_{i,i' \neq i} \frac{Z_i Z_{i'}}{|\mathbf{R_i} - \mathbf{R_{i'}}|} \tag{2.66}$$

is the fully classical ion–ion Coulomb potential, where Z_i is the atomic number of ion i.

Even when the ionic degrees of freedom are excluded from the many-body problem, the many-body electronic wavefunction is insoluble for real materials systems. Mean-field approaches can be used to further simplify the problem, and the Hartree-Fock method is one such approach that has enjoyed some popularity[34]. In the Hartree-Fock method, the many-body electronic wavefunction is constructed as an antisymmetric Slater determinant of single-electron "orbital" wavefunctions. The Hamiltonian that yields the single-electron wavefunctions treats the V_{ee} as an interaction of the electron with the mean-field effective

potential caused by the other electrons in the system. While Hartree-Fock exactly accounts for the Pauli-exclusion exchange energy in the mean-field potential, it does not account for higher-order many-body interactions (so-called *correlation* effects).

2.4.1 Density Functional Theory

In addition to providing an inadequate treatment of correlation effects, the Hartree-Fock method entails solving a non-linear equation in terms of the $3J$ electronic degrees of freedom. The pioneering contributions of Hohenberg, Kohn, and Sham[35, 36] laid the groundwork for rigorously calculating quantum properties of real materials systems by recasting the ground-state solution of the many-body Schrödinger equation in terms of the electron density. This reduces the number of independent degrees of freedom from the $3J$ coordinates of the individual electrons to the three spatial coordinates in terms of which the electron density is described. Hohenberg and Kohn showed[35] that by applying the variational method to the expectation of the many-body Hamiltonian for a trial wavefunction $\psi^{(t)}$, such that

$$E^{(t)} = \frac{\langle \psi^{(t)} | H | \psi^{(t)} \rangle}{\langle \psi^{(t)} | \psi^{(t)} \rangle}, \tag{2.67}$$

there is an equivalent energy functional, $E^{(t)}[\rho]$, of the electron density, and that the electron density that minimizes $E^{(t)}$ uniquely maps onto the ground-state wavefunction of the many-body system. These concepts provide the basis density functional theory (DFT).

Using the results of Hohenberg and Kohn, a form of the density functional can be obtained by positing a separable electronic wave function for a fictitious system of non-interacting electrons and using the variational method to obtain the energy functional[36]

$$E[\rho] = T_s[\rho] + \int \mathrm{d}\mathbf{r}\, v_{ext}(\mathbf{r})\rho(\mathbf{r}) + V_H[\rho] + E_{xc}[\rho], \tag{2.68}$$

where $\rho(\mathbf{r}) = \langle \psi^{(t)} | \sum_j \delta(\mathbf{r} - \mathbf{r_j}) | \psi^{(t)} \rangle$ is the electron density. The only term in Eq. (2.68) that is rigorously known is the integral term, which depends on the external potential v_{ext}, which arises from the electron–ion Coulomb interaction, as well as the presence of any external field. Both $T_s[\rho]$, which is the kinetic energy functional for system of non-interacting electrons, and $V_H[\rho]$, which is the mean-field Hartree energy due to electron–electron interaction (equivalent to the classical Coulomb energy of an electron density ρ) are approximations to the comparable contributions of the ground-state energy. However, Eq. (2.68) remains exact due to the presence of the exchange-correlation functional, $E_{xc}[\rho]$, which accounts for exchange and many-body affects not captured by the other three functionals. Although E_{xc} is not known, various schemes exist for its approximation.

The two predominant approximations used for $E_{xc}[\rho]$ are the local density approximation (LDA) and the generalized gradient approximation (GGA). In LDA, the electron density is assumed to behave locally like a homogeneous electron gas with respect to exchange and correlation[36], the properties of which are well known. In practice, LDA works well but tends to predict over-binding for most systems, with bond lengths that are shorter than experimentally-measured values. GGA treats $E_{xc}[\rho]$ as depending locally on the gradient (and sometimes higher-order spatial derivatives) of the electron density, in addition to its value[37]. Compared to experiment, GGA tends to predict longer bond lengths and lower binding energies.

The utility of the Kohn-Sham approach is that the variational method can be applied to energy functional in Eq. (2.68) to obtain a Schrödinger-like equation for each allowed single-electron state ϕ_j, while the allowed states that are filled with electrons comprise the total wavefunction. The resulting equations are the self-consistent Kohn-Sham equations, which take the form[36]

$$\left[-\nabla^2 + v_{ext}(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r})\right]\phi_j = \epsilon_j \phi_j, \tag{2.69}$$

where the effective potential is a variational derivative of the energy functional in Eq. (2.68). In particular, the effective exchange-correlation potential is given by

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}, \tag{2.70}$$

whose form depends on the precise approximation used for exchange and correlation. Due to the explicit dependence of $v_{xc}(\mathbf{r})$ and $v_H(\mathbf{r})$ on the total electron density, the Kohn-Sham equation for each electron is dependent on all the others, and the set of equations can be solved iteratively until a self-consistent electron density and ground-state energy are obtained.

Self-consistent solution of the Kohn-Sham equations is analogous to solving a system of second-order ordinary differential equations. In the case of crystalline materials, which consist of a periodically-repeating unit cell, the problem becomes particularly well-suited to spectral methods. In fact, Bloch's theorem states that in a periodic potential, the allowed states of a single-electron Schrödinger-type equation, such as Eq. (2.69) take the form[33]

$$\phi_{n,\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}), \tag{2.71}$$

which have the natural quantum numbers n, corresponding to the electron band, and $\mathbf{k}$, corresponding to a vector contained within the Brillouin zone of the crystal lattice. $u_{n,\mathbf{k}}(\mathbf{r})$

is a function having the periodicity of the lattice. Given this constraint, $u_{n,\mathbf{k}}(\mathbf{r})$ can be expanded in terms of a plane-wave basis so that

$$u_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} \alpha_{n,\mathbf{k}}^{(\mathrm{G})} e^{i\mathbf{G}\cdot\mathbf{r}}. \tag{2.72}$$

Because the periodic crystal is assumed infinite, it contains an infinite number of single-electron states, and thus the expansion coefficient $\alpha_{n,\mathbf{k}}^{(\mathbf{G})}$ are a continuous function of $\mathbf{k}$. Moreover, the sum over $\mathbf{G}$ in Eq. (2.72) extends over all vectors from the origin to a lattice point in reciprocal space (in order for the corresponding plane wave to have periodicity of the lattice). In practice, the expansion coefficient is calculated only on a finite grid of k-points, and only extends to values of $\mathbf{G}$ such that $|\mathbf{k}+\mathbf{G}|$ lies within a well-defined radius of the origin measured by the "plane-wave energy cut-off". The plane-wave energy cut-off and the density of the k-point grid must be chosen carefully in any DFT simulation to ensure sufficient accuracy while minimizing computational cost.

2.4.2 Calculation Procedure

This simplest approximation one can make for $F(\vec{\sigma})$ is that it is simply the ground-state energy of the electronic wave-function, which can be calculated directly from DFT. This approximation also implies that the ions are sufficiently massive (or that the bonds are sufficiently stiff) that the zero-point vibrational energy associated with the ionic ground state is negligible. In order to apply the thermodynamic results of Sec. 2.1, we must calculate energies both in the ideal bulk material, and in some calculation geometry that approximates the interface region. All DFT calculations in this dissertation are performed in the self-consistent Kohn-Sham framework as implemented in the Vienna *ab-initio* Simulation Package (VASP)[38], which uses a plane-wave basis set. Except where noted, calculations are performed using the Ceperley-Alder LDA correlation functional[39], as parameterized by Perdew and Zunger[40]. We use ultrasoft pseudopotentials[41] to describe the effective potential of ions and core-shell electrons, allowing energy convergence of GaAs calculations using a plane-wave energy cut-off 203 eV.

Calculating the energy of the interface region, which encloses the reconstructed surface, is somewhat more involved than calculations for the ideal bulk. We use as our interface structure a slab of bulk-like material sufficiently thick so as to approximate the continuation of bulk below the surface. This slab is terminated above the top layer by the reconstructed surface. The bottom atomic bilayer is fixed at the bulk GaAs lattice parameter, as determined by LDA DFT ionic relaxation of the bulk. Atoms in all other layers, and in the reconstructed surface, are allowed to relax to their stable energy minima. The slab is sep-

arated from its periodic image by approximately 12 Å of vacuum, and its bottom surface is passivated by a layer of pseudohydrogen, with nuclear charge Z=0.75, to passivate the dangling bonds of the bottom cleaved surface, in accordance with ECR. As discussed in Sec. 1.1.2, the unpassivated dangling bonds results in charge separation at the interface. Because codes, such as VASP, that express the electron density in terms of a plane-wave basis set must use periodic boundary conditions, charge separation within the slab can cause long-range dipole interactions between the slab and its periodic images. In order to study surfaces that induce charge separation in the slab, it is typically necessary to correct for this nonphysical dipole interaction explicitly[42]. The VASP-implemented dipole correction was tested on a range of low-energy reconstructions on GaAs to determine if dipole interactions across the vacuum layer must be explicitly accounted for, but no meaningful difference in energy was observed for the charge-neutral surfaces being considered here.

DFT calculations for the interface region typically range in size from 90-280 atoms, which is relatively large for a DFT calculation. We can nevertheless obtain good convergence of relative energies, to within a few meV, with respect to the size of the Gamma-centered k-point grid by using a k-point density of at least 144 k-points per reciprocal primitive cell. We only use a single k-point value in the reciprocal direction perpendicular to the slab, as the bands are very flat in this direction due to the vacuum separation (e.g., similar to an electron in a very deep potential well).

CHAPTER III

Enumerating Surface Structure for Systematic Prediction of Reconstruction Stability

3.1 Introduction

While in Ch. II we described how to rigorously calculate the thermodynamic properties of atomic surface structures and compare their relative stabilities, those methods are ultimately parameterized by density functional theory. Although DFT is very accurate for examining relative properties of different structures, it is an insufficient tool to predict structure. Put another way, DFT and the methods presented in Ch. II are appropriate for testing hypotheses of equilibrium structure, but they cannot generate hypotheses. A robust and versatile framework for predicting surface phenomenon from minimal initial information must prescribe methods to both generate and test hypotheses. A number of promising approaches that achieve these goals have been implemented for *bulk* crystal structure prediction. The most prominent among these are genetic programming strategies[43] and data mining of materials databases[44, 45]. Unfortunately, significantly less progress has been made to formalize the prediction of surface structure. Genetic algorithms have been used to determine equilibrium surface structure in single-component systems[46], but since genetic optimization typically requires many hundreds or thousands of energy calculations, it is ill-suited for multicomponent surfaces or when first principles accuracy is desired. Data mining approaches, on the other hand, are difficult to envision for surface systems since significantly less data is available for well-characterized surface structures, especially in comparison to the well-curated databases that exist for bulk crystal structures. Moreover, the range of possible surface structures depends largely on the crystallography of the bulk and how it is related to the surface plane of interest. Consequently, even a single bulk crystal structure may permit many possible surface terminations. Because the range of possible surface structures is so varied, many surface characterization problems likely arise from previously unencountered

structures, thus limiting the usefulness of a data-mining approach.

In some instances the degrees of freedom that dominate surface reconstruction can be mapped onto a 2-dimensional lattice model. The cluster expansion technique can then be employed to efficiently search for thermodynamically stable reconstructions[29, 30]. This approach has been used to study a simple model of As adsorption and surface alloying in InAs/GaAs (001)[21]. Unfortunately, because it is an on-lattice model, the cluster expansion in its typical form is ill-suited to describe the more complicated structural degrees of freedom or symmetry-breaking local relaxations that may occur at the surface of materials with highly directional bonding.

Our particular focus in this dissertation on the (001) surface of zincblende III–V semiconductors provides a prime example of the need for a more systematic approach for predicting equilibrium surface structure. The observed (001) surface structures on III–V materials include a number of poorly understood surface reconstructions, both on the pure compounds and on their alloys. As discussed in Ch. I, even for the GaAs (001) surface, which is relatively well characterized, there are longstanding experimental observations of surface symmetries that have yet to be adequately explained by accurate structural models. In 1989 Däweritz and Hey[47] determined a reconstruction phase diagram from reciprocal space measurements of the GaAs (001) surface that includes (2×3) and (1×3) periodicities at low temperature, and (4×6) and (3×6) periodicities at Ga-rich conditions, none of which can be explained within the current theoretical understanding of the system.

The necessary first step to resolving some of these mysteries presented by experimental surface characterization is to generate plausible hypotheses for ground-state structures having the experimentally determined symmetries. If we can obtain a database of structurally promising candidates, we can subsequently screen them for relative stability using either physically-derived heuristics or the first principles methods described in Ch. II. In this chapter we present a systematic approach for determining equilibrium atomic surface structure in III–V compound semiconductors. On multicomponent surfaces such as GaAs, a fully defined reconstruction consists of a *reconstruction prototype* (i.e., the set of atomic-site coordinates that describe the topology of the surface reconstruction) together with its *species configuration* (i.e., the distinct arrangement of the M atomic species on the sites of the reconstruction prototype). Due to this two-part aspect of multicomponent surface reconstruction categorization, our approach consists of two steps: (i) Explore structural degrees of freedom to generate candidate reconstruction prototypes and use screening methods to identify low-energy prototypes. (ii) Use the cluster expansion formalism[29] to describe the energetics of species configuration for low-energy prototypes in order to identify the overall ground-state reconstructions, each consisting of a reconstruction prototype plus a particular

species configuration of that prototype.

For the first step we introduce an algorithm to generate structurally plausible reconstruction prototypes based on observed or proposed structural motifs. A distinguishing feature of III–V (001) surfaces is the in-plane bonding between neighboring atoms, which gives rise to the "surface dimer" and "backbond" motifs. In compound materials, some sites of a surface reconstruction prototype may be able to host more than one species of atom. In III–V systems this is most readily observed by the formation of *heterodimers*, where a terminating dimer consists of one cation species and one anion species. The possibility of multiple distinct heterodimer arrangements necessitates the second step of our approach in which we use density functional theory energy calculations to determine the energetic role of these configurational degrees of freedom. The cluster expansion formalism from alloy theory[29, 30], described in Ch. II, is then applied to these data to predict energies across the entire configuration space of a single reconstruction prototype, making Monte Carlo methods accessible as a means to conduct comprehensive thermodynamic analysis at finite temperature.

We apply this method to the (001) surface of the zincblende III–V compounds, with the ultimate goal of identifying the energetically competitive reconstructions of this surface that either are stable for a pure binary compound or are likely to become stable upon perturbing the binary system (e.g., by introducing strain or alloying). Several specific surface reconstructions of the III–V binary compounds have been well-characterized, allowing us to develop a set of structural rules that recreate these and similar reconstruction prototypes, while excluding prototypes with unobserved and nonphysical features. An implementation of our enumeration algorithm incorporating these structural rules is used to generate all likely surface reconstructions for user-specified unit cell geometry. We present a statistical characterization of the generated reconstruction prototypes and examine their excess charge characteristics, which are determined by an efficient formulation of the electron counting rule (ECR)[1]. The database of generated reconstructions provides a new and valuable tool for identifying the structure of poorly characterized binary and alloyed III–V semiconductor surfaces.

We examine reconstruction stability by calculating surface free energies from first principles for all charge-neutral reconstruction candidates having experimentally-observed surface lattice vectors on GaAs (001). These results verify the accepted GaAs (001) phase diagram and reveal near-stable reconstructions, including a class of (4×3) surface reconstructions that is nearly stable for GaAs(001). Finite temperature Monte Carlo free energies are compared for this reconstruction prototype and the well-known GaAs $c(4\times4)$ reconstruction in order to determine relative stability in a way that accounts for the entropic contribution of species configuration.

3.2 Methodology

3.2.1 Structural trends of III–V structures

The bulk zincblende crystal of III–V compounds is comprised of two sublattices—the group V, or anion, sublattice and the group III, or cation, sublattice. Each atom of one sublattice shares tetrahedrally-oriented bonds with four atoms of the other sublattice. Along the [001] direction, the crystal exhibits an ABCD stacking of square $2D$ lattices; A and C layers belonging to the anion sublattice and B and D layers belonging to the cation sublattice. The stability of zincblende phases of many III–V compounds is due to the sp^3 orbital hybridization of the constituent atoms, as well as the strong energetic preference for III–V bonds over V–V or III–III bonds.

Although the surface structure is necessarily less well-determined than the bulk, consistent structural traits of III–V (001) surfaces can be discerned from careful consideration of the most likely III–V reconstruction models, as viewed in the context of the experimental and computational literature. We focus here on the class of III–V reconstructions that are Group V-rich. Bonding orbitals at these surfaces are predominantly sp^3 in character, since full filling of $\sigma(sp^3)$ bonding orbitals can be easily realized by V–V and III–V bonds. The resulting bond angles are typically close to tetrahedral, leading to surface structural motifs that can easily maintain coherence with the substrate crystal without requiring bond-lengths significantly different from those in the bulk.

The predominant dimer and back-bond motifs of the most well-substantiated reconstruction models form only along well-specified direction. As illustrated in Fig. 3.1(a), we can define two directions specific to each sublattice—the "dimer" axis and the "non-dimer" axis. An atom in the unreconstructed surface has two substrate-oriented tetrahedral bonds that link it to the bulk. If we imagine these substrate-oriented bonds as the legs of a triangle, it is reasonable to postulate that relaxations of the surface atom perpendicular to the plane of the triangle can be much larger than relaxations within the plane. We therefore define the dimer axis as the direction normal to the plane spanned by the two substrate-oriented tetrahedral bonds, and this is the axis along which in-plane dimers and back-bonds form. The non-dimer axis is perpendicular to the dimer axis in the (001) plane. These axes and the [001] axis are mutually orthogonal, and the dimer and non-dimer axes are defined separately for each sub-lattice. Thus, in terms of the directions defined in Fig. 3.1(a) for GaAs, [110] is the dimer axis of the Ga sublattice and it is the non-dimer axis of the As sublattice; $[1\bar{1}0]$ is the dimer axis of the As sublattice and it is the non-dimer axis of the Ga sublattice.

With these definitions in place, we can enumerate the observed III–V structural traits, which we have organized in a hierarchy, beginning with the least likely to be violated:

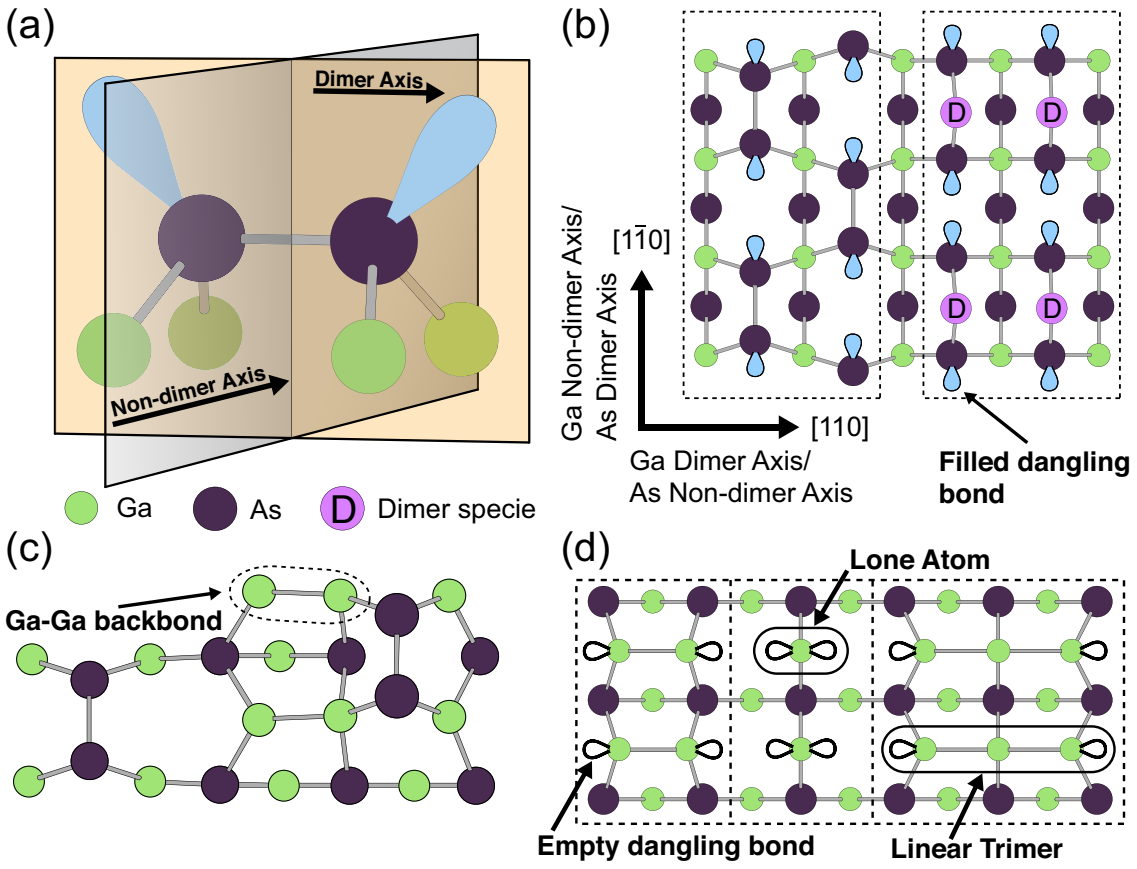


Figure 3.1. III–V (001) surface structure trends shown for GaAs. (a) As dimer, in perspective, with filled dangling bonds. The non-dimer axis is normal to the plane spanned by the dimer and dangling bond. (b) The As-terminated (001) surface, with two possible dimerizations. Dimers on the left are depicted in the usual manner; dimers on the right are depicted using the dimer "specie" abstraction. (c) The $\alpha 2(2\times4)$ reconstruction with Ga-Ga backbond indicated by a dashed oval. (d) The Ga-terminated (001) surface, with three different bonding motifs: surface dimers, with two three-fold coordinated atoms; lone atoms, with two dangling bonds each; and linear trimers, with a central four-fold coordinated atom. (b)-(d) share the same crystal axes.

1. *There are no vacancies below surface layers.* The presence of a vacancy significantly distorts surrounding bonds alters the local charge state, leading to high-energy strain and electrostatic interactions. This observation implies that any atom of the structure must sit above and between two other atoms, along the dimer axis of the two lower atoms.

2. *All atoms have 3-fold or 4-fold coordination.* This requirement is necessary to enable anions to have filled valence shells and cations to have empty valence shells, assuming bonding is polar. Three-fold coordinated anions typically have a filled valence shell orbital which does not participate in bonding, referred to as a "dangling bond".

3. *Anions and cations form dimers along the dimer axis.* If no atom sits in the layer above two atoms that are in-plane neighbors along their dimer axis, these two atoms may bond to form a dimer. Dimers reduce the number of high-energy dangling bonds at the expense of introducing some degree of strain due to the relaxation of the dimerized atoms toward each other. The prototypical III–V anion dimer is illustrated in detail in Fig. 3.1(a). Dimerization is essentially a way to increase the coordination of atoms via local relaxations. This, along with the first two rules, implies that every atom must have at least one neighbor along the dimer axis in the same (001) plane. Figure 3.1(b) shows dimerization in the anion layer, and Fig. 3.1(d) shows dimerization in the cation layer.

4. *Backbonds may form between a lone atom and a neighboring three-fold coordinated atom.* Backbonds are the second type of in-plane bonding. Unlike dimers, a backbond joins a three-fold coordinated atom to a four-fold coordinated atom in the same (001) plane along the plane's dimer axis. Because relaxation of the four-fold coordinated atom is limited, the local relaxation associated with backbonds often results in more highly strained motifs than are caused by the dimer in-plane bond. A III–III backbond is a prominent feature of the $\alpha 2(2\times4)$ reconstruction, which is shown in Fig. 3.1(c) (backbond is circled).

5. *A four-fold coordinated atom may not have an in-plane bond.* Such a bond would essentially amount to a "bridge" bond between larger structural features, or to linear trimers, tetramers, and so on. As noted for dimers or backbonds, neighboring atoms must locally relax in order to obtain a separation near the equilibrium in-plane bond length. Such relaxation is significantly more difficult to achieve if the in-plane bond joins two four-fold coordinated atoms. Consequently, these motifs are largely unphysical from an energetic standpoint. Although a linear tetramer structure has

been proposed for the GaN (001) (4×1) surface[48], later work has cast doubt on this model[49]. The illustration in Fig. 3.1(d) compares motifs with different numbers of in-plane bonds, including a linear trimer motif.

3.2.2 Algorithm to generate surface reconstruction prototypes

The above rules are incorporated in an algorithm to enumerate III–V reconstruction prototypes. The specific form of the algorithm presented here is based on knowledge of the III–V (001) system (e.g., the bulk crystal and the structural rules we have outlined), although it can be easily adapted to study a range of other systems, provided that basic structural information is known. Surface atoms are assumed to reside at well-defined lattice sites, to within small local relaxations. For anion-rich III–V surfaces these sites are assumed to coincide with the bulk zincblende lattice positions, although additional interstitial sites may be necessary to describe other systems. The algorithm takes as input the surface lattice vectors of the desired unit cell.

The surface unit cell has an associated integral unit area, N_A, which is the number of atoms in a complete bulk-like monolayer of the specified unit cell. Because the 1×1 surface lattice vectors lie along the face diagonals of the zincblende cubic unit cell, the surface lattice parameter is given by $a_{surf} \equiv a_{bulk}/\sqrt{(2)}$, so that the unreconstructed surface has one atom per unit area a_{surf}^2. In general, a unit cell with lattice vectors $\mathbf{v} = (v_{[110]}, v_{[1\bar{1}0]})$ and $\mathbf{w} = (w_{[110]}, w_{[1\bar{1}0]})$ in terms of the [110] and [1$\bar{1}$0] unit vectors has an integral unit area

$$N_A \equiv \det \begin{pmatrix} v_{[110]} & v_{[1\bar{1}0]} \\ w_{[110]} & w_{[1\bar{1}0]} \end{pmatrix} / a_{surf}^2. \tag{3.1}$$

Using the structural rules outlined above, it might be possible at this point to perform a brute-force enumeration of reconstruction prototypes; however, many prototypes would be visited that are either symmetrically equivalent or structurally incompliant (i.e., they do not obey our structural rules). The goal of our algorithm is to find the subset of reconstruction prototypes that are both structurally compliant and symmetrically distinct while limiting the total number of structures that must be evaluated. This is done efficiently by constructing the reconstruction prototypes layer-by-layer, from the substrate up, screening at each layer for structural compliance and symmetrical uniqueness. Consider the $n = 0$ atomic layer, which represents the first layer of substrate, sitting at the surface of a semi-infinite bulk crystal. Atop this layer, the $n = 1$ layer is placed, which consists of some arrangement of atoms and vacancies on the N monolayer sites of the unreconstructed unit cell. All 2^N possible layer configurations are considered, keeping only one instance of each symmetrically

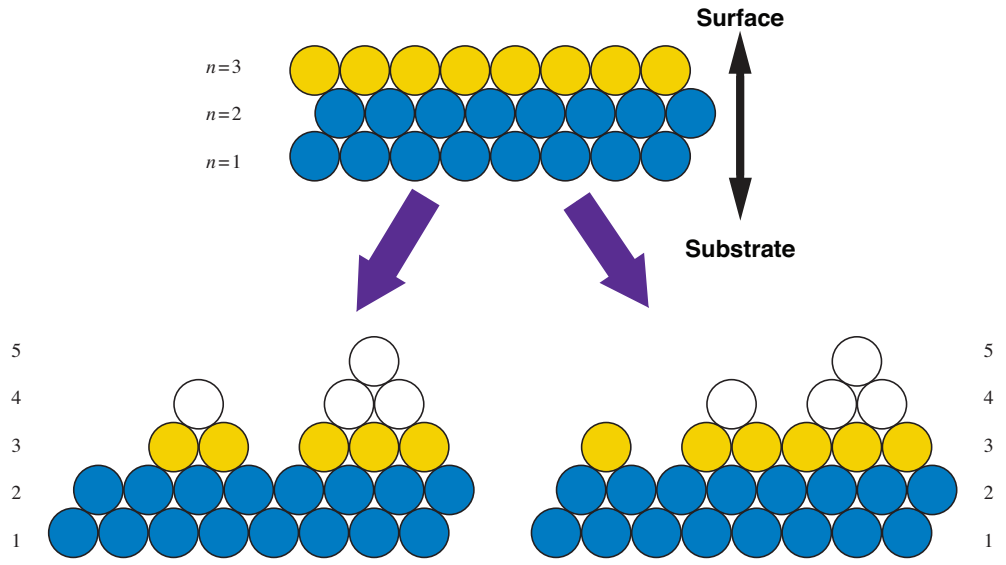


Figure 3.2. Schematic illustration of enumeration procedure for a close-packed cartoon crystal, viewed from the side. Starting from the surface in the top figure, all possible configurations of atoms and vacancies in the current layer $n = 3$ are enumerated. For each $n = 3$ configuration, the procedure is repeated in layer 4. Since sub-surface vacancies are not allowed, the procedure is self-terminating, and the bottom prototypes illustrate possible outcomes. If two arrangements of layer 3 are different, they cannot produce equivalent prototypes in subsequent layers.

distinct monolayer configuration and subsequently excluding all configurations that are not allowed in a structurally compliant prototype.

On top of each compliant $n = 1$ layer configuration we specify a $2D$ crystal of potential sites for the $n = 2$ layer, selected from ideal bulk lattice positions. Each site of this new layer must have two nearest neighbors in the $n = 1$ layer, in accordance with the first structural rule. Structurally compliant configurations of atoms and vacancies in the $n = 2$ layer are enumerated, and each subsequent layer is constructed in this manner. When the n^{th} layer configuration results in all n layers being structurally compliant, this n-layer structure is a valid reconstruction prototype and is added to the database. This procedure is illustrated schematically in Fig. 3.2.

We limit the possibility of constructing fully-filled substrate-like layers in order to prevent the generation of prototypes that are identical upon c-axis translation (normal to the surface). With this criterion in place there are always more atoms in layer $n - 1$ than in layer n, so that the enumeration process will eventually terminate when no new layer can be constructed that obeys the structural rules. For most systems in which there are no bonds normal to the surface plane, which include zincblende (001), structural rule one is sufficient to ensure self-termination. In other systems, such as the simple cubic (001), additional criteria would

have to be specified. The impossibility of c-axis translational symmetry of two generated structures also ensures that if two structures-in-progress are identical up to layer $n-1$ but differ in layer n, all prototypes spawned from the first structure will be distinct from all prototypes spawned from the second structure.

We abstract the in-plane bond as a distinct structural element which we will refer do as the "dimer specie". This structural abstraction is an essential feature of the algorithm that is necessary to distinguish between distinct arrangements of dimers and backbonds on otherwise identical monolayer configurations of atoms and vacancies. In the implementation of the algorithm the dimer specie is treated identically to an atomic species for the purpose of identifying symmetries and enumerating monolayer configurations. A dimer specie can be placed on a regular lattice site in layer n to indicate that its two nearest neighbor atoms in layer $n-1$ are bonded in-plane along the $n-1$ layer dimer-axis. The algorithm can be implemented as described previously, except that layers are enumerated as configurations of atoms, vacancies, and dimer species. Additional restrictions prevent atoms or dimer species in layer n having dimer specie nearest neighbors in layer $n-1$, and dimer species within a layer are not allowed to neighbor along the layer's dimer axis.

Constructing the candidates layer-by-layer from the substrate simplifies the enumeration problem significantly by restricting the number of relevant configurations at each step. With the inclusion of a dimer specie, there are at most 3^{N_A} configurations for the first layer, and significantly fewer for each subsequent layer. The main computational bottleneck is the filtering of equivalent configurations during enumeration, for which we utilized an algorithm similar to that described by Ferreira, Wei and Zunger[50], whereby a set of configurational basis functions are used to construct a mathematical description of configuration that is invariant under space-group operations of the semi-infinite crystal . This method is significantly faster than direct geometric comparison, though faster methods exist[51].

3.3 Electron Counting Rule and its application to analyzing generated prototypes

The formation of dimers and backbonds at the III–V (001) surface inevitably leads to 3-fold coordinated atoms, implying that some sp^3 hybridized orbitals do not participate in bonding. These so-called "dangling bonds" cannot be completely filled with electrons while simultaneously maintaining the local charge balance of the crystal. As a result, the dangling bonds must be either partially or selectively filled. Electron energy states of the dangling hybridized orbitals can be calculated from the energies of the atomic s and p orbitals and compared to the conduction band maximum and valence band minimum energies of the bulk

crystal[2]. For most III–V materials, the electronic states associated with the cation dangling bond are estimated to lie in the conduction band and are therefore empty, while the anion dangling bond states are estimated to lie in the valence band, and thus are filled. Based on these estimates, a heuristic rule was formulated in the form of the electron counting rule (ECR)[1], which demands that the number of electrons required to fill all bonding orbitals and anion dangling bonds is exactly equal to the number of valence-shell electrons donated by the constituent atoms. Experimental and theoretical evidence for the ECR heuristic is prevalent, and we have encountered little compelling theoretical evidence in the literature for violations of ECR in polar III–V materials at equilibrium. The experimental evidence of ECR violation is limited, with notable examples being the extremely Sb-rich $c(2\times10)$ reconstruction on GaSb (001)[12] and the Bi-stabilized (2×1) reconstruction observed on the (001) surface of Bi/GaAs and Bi/InP[13]. The ECR is also applicable to a number of other chemistries important to semiconductors, and it is the heuristic we use as a first-pass screen to determine thermodynamic stability.

If we take the first three structural rules enumerated in Sec. 3.2.1 as given, we can construct an expression for the ECR charge balance that is general for all candidate structures generated by our algorithm (as well as those with "bridged" motifs, which we have excluded). To do this we examine a single atomic bilayer comprised of monolayer n, which is on the anion sublattice, and monolayer $n+1$, which is on the cation sublattice and is nearer to the surface; n increases along the direction pointing away from the solid. In that case the excess number of electrons in the bilayer j containing layers n and $n+1$ is

$$(\Delta q)_j = 2(B_n + B_{n+1}) - 5N_{n+1} - 3N_n, \tag{3.2}$$

where B_n is the number of bulk-oriented bonds, in-plane bonds, and filled dangling bonds in layer n, and N_n is the number of atoms in layer n. The coefficients account for the five valence electrons contributed by an anion, the three valence electrons contributed by a cation, and the two electrons required to form a bond. We assume for now that there are no antisites.

By applying considerations from the first three structural rules, the bonding term in Eq. (3.2) can be expressed as

$$B_n + B_{n+1} = [2(N_n + N_{n+1}) + D_n + D_{n+1}] + [2(N_n - N_{n+1}) - 2D_n]\,, \tag{3.3}$$

where D_n denotes the number of in-plane bonds between atoms in layer n. The left bracketed term accounts for all bulk-oriented bonds (of which each atom has two, according to structural rule 1) and all in-plane bonds formed by atoms in the bilayer. The right bracketed

term accounts for all dangling bonds in the anion layer, which, according to the ECR, must be filled. The expression for the number of dangling bonds is obtained by noting that in the absence of in-plane bonds the number of filled dangling bonds created is twice the difference between the number of anions in layer n and the number cations in layer $n+1$, due to the constraint of structural rule 3. We can subsequently account for the elimination of two filled dangling bonds by each in-plane bond formed in the anion layer. Substituting this expression in Eq. (3.2) we obtain

$$(\Delta q)_j = 3(N_{n+1} - N_n) + 2(D_n - D_{n+1}). \tag{3.4}$$

The expression in Eq. (3.4) can be used to find the total number of excess electrons by summing over all bilayers of the structure. An additional charge of $\frac{3}{2}N_A$ is contributed from the first bulk-like layer of cations below the surface (the distinction of "bulk-like" is arbitrary, as long as all cations in the layer are tetrahedrally coordinated). The integral unit area N_A of the unit cell is equivalent to the number of atoms in a full monolayer. The total excess charge of a particular unit cell is therefore given by the sum

$$\Delta Q = \frac{3}{2}N_A + \sum_{n=1}^{N}(-1)^n(2D_n - 3N_n), \tag{3.5}$$

with sum taken over the N monolayers closest to the surface, starting with an anion layer at $n = 1$. This is a very simple expression for ECR charge balance, and when we abstract the in-plane bond, as we must do for the structure enumeration, it can be used to calculate excess charges for a database of thousands of candidate prototypes in a matter of seconds. First principles calculations bear out the presumed energy penalty imposed by excess surface charge, and thus charge neutrality as predicted by the ECR should be considered as an additional structural criterion. Accordingly, we focus on charge-neutral structures in this work, although the screening is performed as a final step of prototype generation so that completeness is preserved.

3.3.1 Identifying Low Energy Reconstructions and Exploring Configurational Degrees of Freedom

Up to this point, we have discussed the generation of reconstruction prototypes without explicitly specifying the species occupying their lattice sites. In order to ultimately determine the relative stability of the generated structures we must identify the low-energy configurations of atomic species on the prototype lattices. The simplified expression for ECR charge-balance in Eq. (3.4) was derived under the assumption that the surface is composed

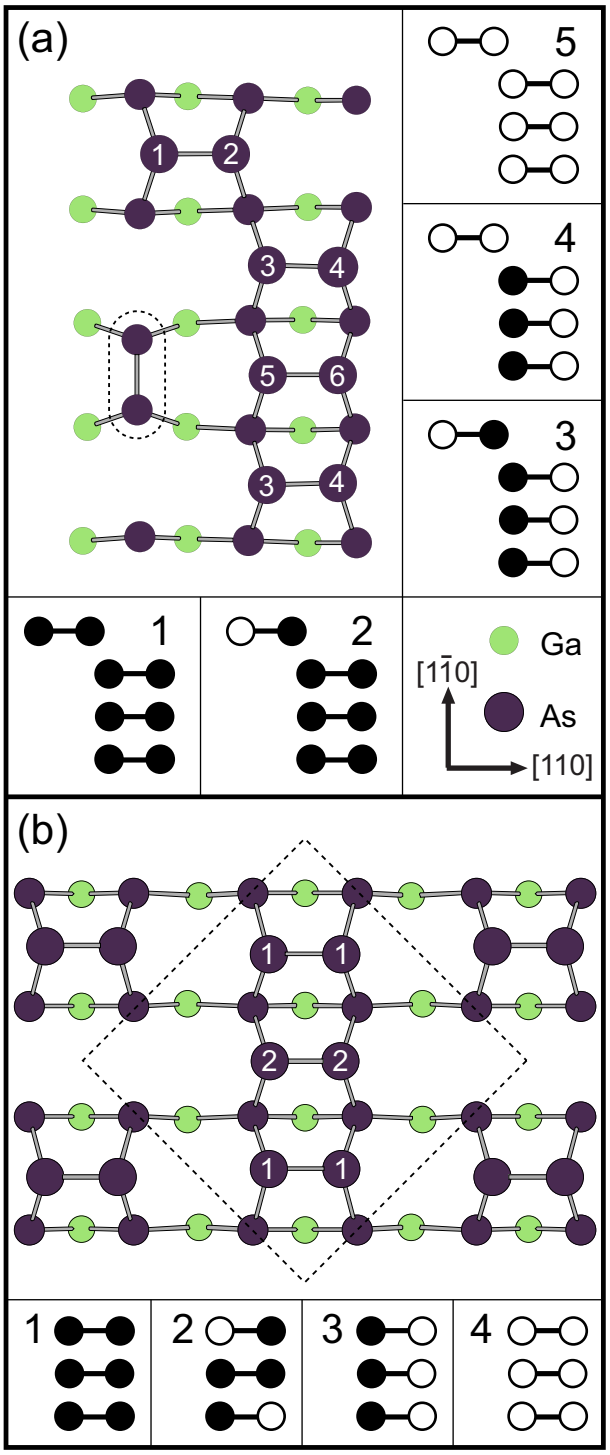


Figure 3.3. Depiction of dimer sites considered for species substitution in the cluster expansions, as well as low-energy configurations of these sites (schematic insets). The sites included in the cluster expansion are indicated by white numbers, where equivalent sites share the same number. Low-energy configurations larger than the unit cell are not depicted. In configuration schematics, black are As, white are Ga. (a) The $h0(4\times3)$ prototype and five low-energy dimer site configurations. (b) The $c(4\times4)$ prototype and four low-energy dimer site configurations.

of single-species monolayers that alternate between Group III and Group V, as in the substrate. Species substitution in the bulk, where atoms are four-fold coordinated, is considered an antisite defect, resulting in a net change of $+2$ or -2 electrons from the substitution of an anion for cation, or visa versa. There is no mechanism to transfer electronic states between the bulk valence and conduction bands to accomodate the change in valence associated with a bulk antisite, and therefore local charge balance must be disrupted. Conversely, if an atom at a three-fold coordinated surface site is substituted, the dangling bond of that site moves from the conduction band to the valence band, or visa versa, accommodating the difference in charge. Consequently, while the formation energy of antisite defects at four-fold coordinated sites, even near the surface, is quite high, the energy of species substitution at *three-fold coordinated sites* is comparatively low. STM studies of III–V surfaces suggest that anion–cation configurational fluctuations of the three-fold sites of surface dimers, resulting in their conversion to heterodimers, do frequently occur[52].

Because of the large difference in substitutional energy between three- and four-fold coordinated sites, we assume that the configuration of four-fold coordinated sites is fixed to the bulk-like configuration. Each three-fold coordinated site, however, introduces an additional configurational degree of freedom associated with its anion–cation occupancy state. The introduction of these degrees of freedom have the consequence of allowing 2^{N_3} possible unit cell configurations of cations and anions on the N_3 three-fold coordinated sites of a given prototype unit cell. Testing all possible configurations of all generated prototypes is a hopelessly intensive and impractical task. Thus, the second step in determining the equilibrium GaAs (001) surface structure focuses on refining the search to a reduced set of reconstruction prototypes and then identifying ground-state configurations of Ga and As on their three-fold coordinated surface sites.

We refine the search for a reconstruction ground state (i.e., a reconstruction prototype plus it lowest energy configuration of species at three-fold coordinated sites) in two steps. In the first step, we calculate an energy for each prototype, assuming an As-rich surface, and identify the lowest energy prototypes (e.g., those within 50 meV of being stable). Once a low-energy reconstruction prototype is identified, we proceed to the second step by collecting the three-fold coordinated sites of the prototype and mapping them onto a two dimensional substitutional lattice model. This lattice contains the sites that have a Ga–As substitutional degree of freedom. Figure 3.3 shows the lattice of selected substitutional sites for two specific reconstruction prototypes (sites where Ga–As substitution is allowed are indicated by numbers).

Once the substitutional surface lattice has been defined we can use the cluster expansion formalism to construct an effective Hamiltonian for the surface in terms of the configurational

degrees of freedom. Instead of the typical site basis function used for a binary, which is simply $\phi_j^{(2)} = \sigma_j$, we choose the site basis function $\phi_j^{(2)} = (1+\sigma_j)/2$. $\phi_j^{(2)} = 1$ indicates Ga occupancy and $\phi_j^{(2)} = 0$ indicates As. Although $\phi^{(1)}$ and $\phi^{(2)}$ are no longer orthonormal with respect to the inner product defined in Sec. 2.3, they are linearly independent. Moreover, the inner product can be redefined such that they are orthonormal[53]. The "correct" inner product for this choice of site basis function corresponds to a measure that gives overwhelming weight to configurations that are As-rich. As a result, this choice of basis set is particularly well-suited to the As-rich regime[54].

The cluster expansion is a valuable tool for identifying configurational ground states, which can be found either by directly enumerating lattice configurations and evaluating their energy via the cluster expansion or by simulated annealing within Monte Carlo simulation. Monte Carlo can also be used to determine the finite-temperature surface free energy of a reconstruction prototype, which is affected by thermal excitations of configuration. By comparing the free energies of various low-energy reconstruction prototypes, a finite temperature phase diagram can be determined.

3.4 Results

3.4.1 Statistical Analysis

Using the algorithm described above along with our structural rules for III–V surfaces, we have enumerated all reconstruction prototypes of the III–V (001) zincblende surface up to $N_A = 12$ (i.e., all supercells containing 12 or fewer (1×1) surface cells) within the Group V-rich regime. Fig. 3.4(a) shows the number of generated structures for each supercell area. The total for each volume is subdivided by supercell shape. Interestingly, the supercell shapes that account for the largest portion of candidate structures (nearly 50%) have aspect ratio much higher than 1 with perpendicular (or nearly perpendicular) lattice vectors. These supercells, which have one lattice vector of length $\sqrt{2}a_{surf}$ that is oriented along either [100] or [010], yield prototypes with prominent (111) facets and may be useful in identifying low-energy reconstructions of the edge along adjoining {111} surfaces.

Although the number of generated reconstruction prototypes increases exponentially with unit cell area, any characterization information about an observed reconstruction, such as the surface unit cell lattice vectors, significantly limits the number of prototypes that must be considered in order to determine its structure. Likewise, if STM data are available, a visual comparison between micrographs and structural models is useful to identify the most plausible candidates before performing energy calculations. Also, it is a convenient feature of the zincblende (001) surface cell that, if a unique unit cell can be formed under a 90° rotation

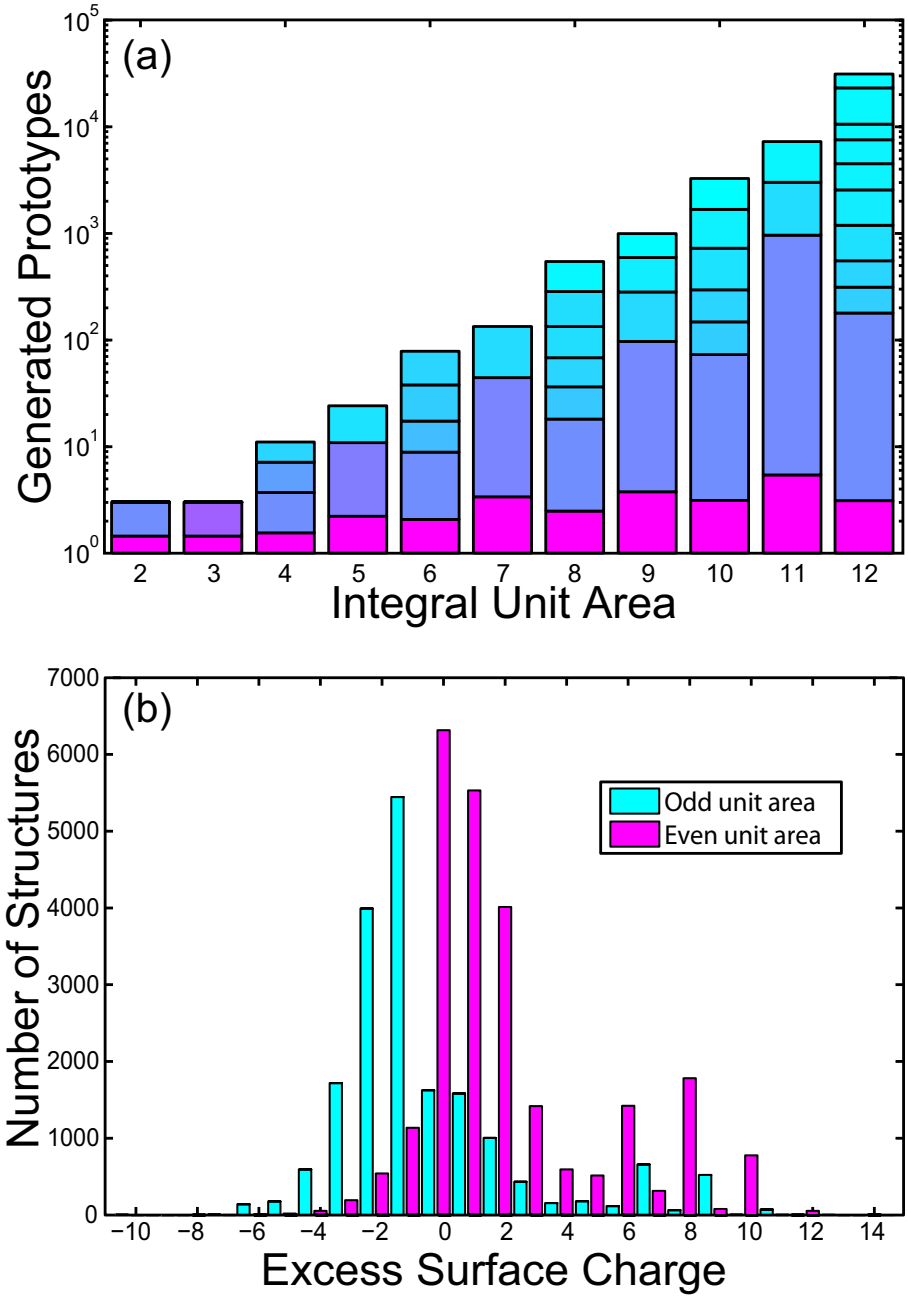


Figure 3.4. Statistical characterization of the III–V surface reconstructions prototype database generated by the algorithm presented in Sec. 3.2.2. (a) Total number of generated structures for each unit cell area, up to $N_A = 12$. Bar subdivisions are proportional to the number of generated prototypes for each supercell shape. Topmost sections correspond to square (or near-square) supercells, and aspect ratio increases down the bar. Color indicates area-normalized aspect ratio for comparison of supercell shapes with different area. (b) Histogram of excess surface charge incidence for the generated prototype database, as determined by the ECR. Distinction between supercells with odd or even integral unit area is indicated.

of lattice vectors, the two unit cells produce the same set of reconstruction prototypes, to within an exchange of species. Thus, nearly half of the possible prototypes can be generated by inspection from the other half.

The excess surface charge was calculated for each generated structure using Eq. (3.4). The excess charge histogram, which shows the distribution of excess charge values for the entire set of generated configurations is shown in Fig. 3.4(b). The distribution is bimodal, with odd-unit-area structures centered at -1.5e and even-unit-area structures centered at charge balance.

3.4.2 Reconstruction Stability of GaAs(001)

We use our database of reconstruction candidates to study GaAs (001) by first identifying the stable and near-stable surface reconstructions in the As-rich regime from first principles. Although GaAs (001) is relatively well studied, until now there has been no comprehensive and systematic method to make theoretical predictions of its phase stability. We apply our approach to the (4×2), (2×4), c(4×4), c(8×2), and (4×3) unit cells. The first four are widely reported in the experimental literature to have stable reconstructions[55, 56, 52, 57]. We consider the (4×3) reconstruction due to the importance of predicting near-stable reconstructions that may may manifest in kinetically limited regimes or upon small changes to the material system. A number of experimental results strongly suggest that a (4×3) reconstruction is near-stable or even stable on GaAs (001), including observation of a (4×3) or $(n \times 3)$ reconstruction on the pure GaAs surface at low temperature[47] and during reconstruction transitions[58]. Additionally, the occurrence of a (4×3) reconstruction on the chemically similar GaSb(001) is well-known[59], and a $(n \times 3)$ reconstruction can be induced on GaAs (001) by submonolayer deposition of InAs[60]. Using our algorithm, we determine there to be 23 (4×2), 23 (2×4), 20 c(4×4), 20 c(8×2), and 124 (4×3) reconstruction prototypes that obey the electron counting rule.

To analyze reconstruction stability we calculated both the surface excess 0-K Gibbs free energy, which is defined as $\eta = [G]_{V,N_{tot}}$, and the surface free energy, specified by $\gamma = [\Phi]_{V,N_{tot}}$. γ is a Legendre transform of η, such that $\gamma = \eta - \Delta\mu_{\mathrm{As}} x_{\mathrm{As}}^{(xs)}$. $x_{\mathrm{As}}^{(xs)}$ is the surface excess As concentration and, in terms of the quantities defined in Sec. 2.1, $x_{\mathrm{As}}^{(xs)} = A_{(1\times1)}[N_{\mathrm{As}}]_{V,N_{tot}}$. Although μ_{As} is generally considered to be the parameter controlled experimentally, first principles energy analyses at zero-K are conducted using static atomic configurations, each having a fixed composition. To visualize these data it is advantageous to use the surface excess Gibbs free energy, which is the natural thermodynamic potential to use when holding $x_{\mathrm{As}}^{(xs)}$ constant.

Figure 3.5 depicts the zero-temperature (i.e., without entropic contribution) surface ex-

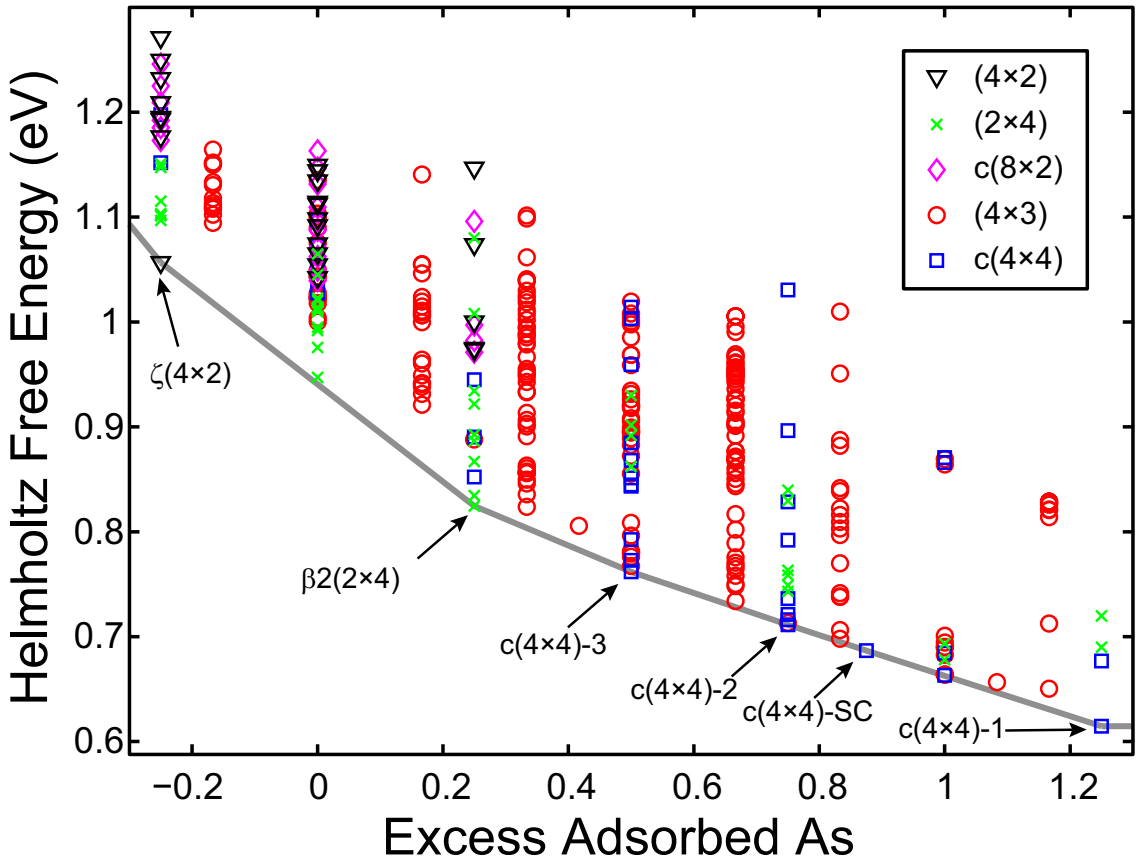


Figure 3.5. Surface excess Gibbs free energies of all 362 calculated reconstructions are shown with respect to excess adsorbed As. 18 configurations lie outside the depicted composition range. Ground states are indicated by arrows and labeled. $c(4\times4)$ variants correspond to configurations depicted in Fig. 3.3, except for $c(4\times4)$-SC which is a supercell configuration larger than the $c(4\times4)$ unit cell. Datapoint markers correspond to unit cell shape, as indicated.

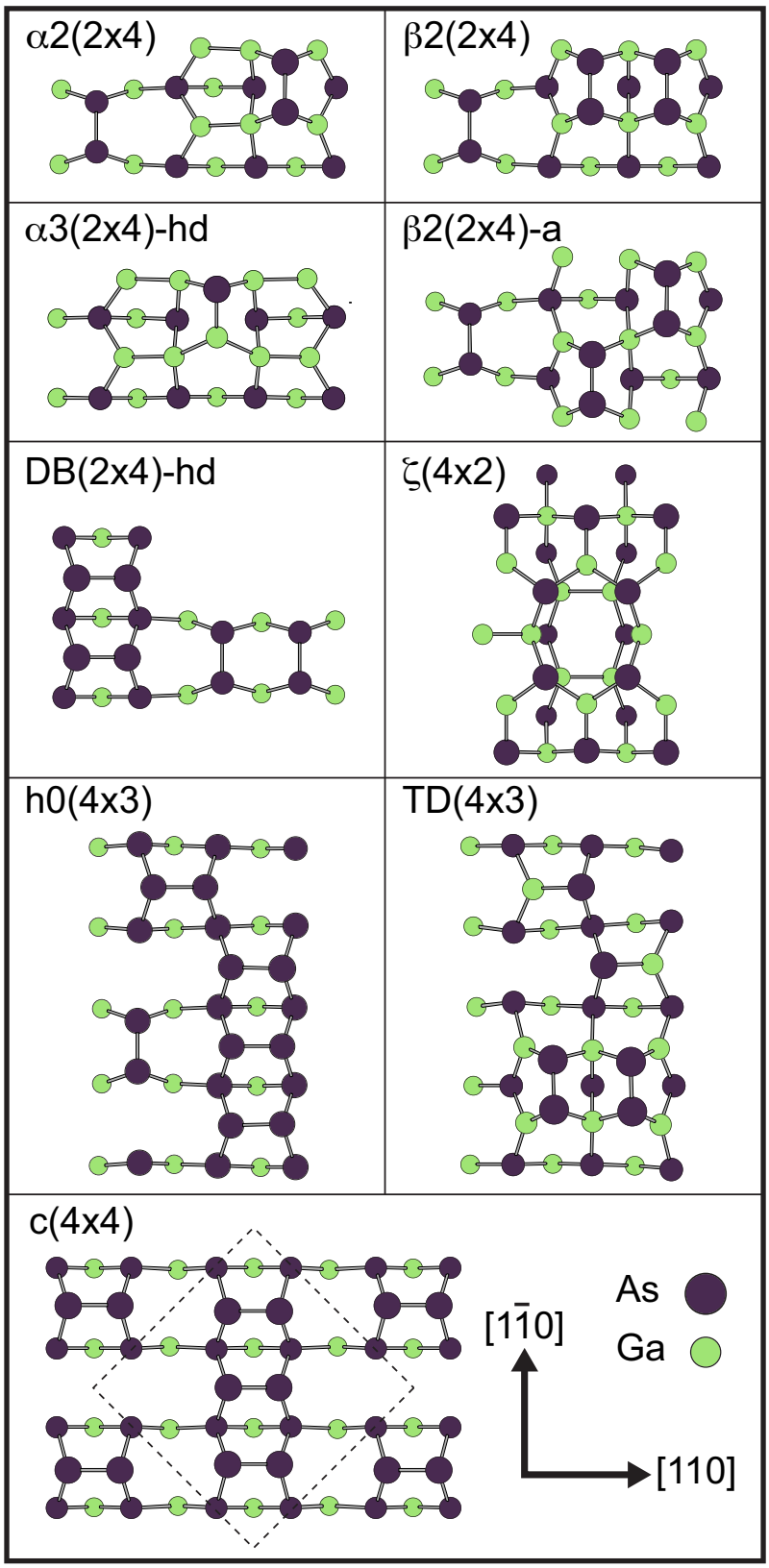


Figure 3.6. Structural models of stable and near-stable reconstructions of GaAs (001), as determined by our calculations.

cess Gibbs free energies plotted versus $x_{\mathrm{As}}^{(xs)}$ for all calculated species configurations of each reconstruction prototype considered. The convex hull of ground-state reconstructions is also shown. In a binary system the convex hull is the curve that describes the lowest possible energy of the system. In Fig. 3.5 it is the lowest-lying convex curve that passes through a subset of datapoints and connects these ground-state reconstructions by straight "common tangent" lines. A ground state is the equilibrium configuration of the system at a given composition; at compositions between two ground states the system exhibits phase separation between the two proximal ground states, with an excess Gibbs free energy on their common tangent line. The convex hull provides an unambiguous reference to determine how close a reconstruction is to stability, which is measured as the difference in η between the reconstruction the convex hull. We focus on those reconstruction prototypes within less than 50 meV of the convex hull, of which there are 9. Structural diagrams of these prototypes are shown in Fig. 3.6.

Our ability to limit the field of 210 reconstruction prototypes to the 9 most likely candidates near the convex hull demonstrates the importance of an objective measure of relative stability that is compatible with the computational process. Having limited our consideration to only 9 reconstruction prototypes it is reasonable to continue with additional first principles analysis to for the energetic effect of species configurations. As database of confiug each prototype by substituting Ga for As (or visa versa when possible) at three-fold coordinated sites. The database of these calculations can be used to construct and refine a cluster expansion for each reconstruction prototype that describes the energetics of Ga-As disorder over the three-fold coordinated sites.

We explored the range of species configuration on the three-fold coordinated sites of the low-energy $h0(4\times3)$ and $c(4\times4)$ prototypes from Fig. 3.6, focusing on the sites indicated in Fig. 3.3. Each site belongs to a surface dimer and, because it has three-fold coordination, can undergo Ga–As substitution. Initial calculations demonstrated that the sites on the $h0(4\times3)$ trench dimer (circled in Fig. 3.3(a)) exhibit a large energy increase upon exchanging As for Ga. Since thermal excitations of occupancy at these sites would be relatively rare events at moderate temperatures in the As-rich regime, they were excluded from further analysis. The remaining three-fold coordinated sites of the $h0(4x3)$ and $c(4\times4)$ prototypes (indicated in Fig. 3.3) were considered in constructing the cluster expansion for each prototype. The cluster expansion for the $h0(4\times3)$ prototype was constructed from a set of 49 configurational energies to parameterize 6 site, 11 pair, and 3 triplet ECI with a cross-validation score of 1.8 meV per considered dimer site of the unit cell. The cluster expansion for the $c(4\times4)$ prototype was constructed from a set of 17 configurational energies to parameterize 2 site, 5 pair, and 2 triplet ECI with a cross-validation score of 0.7 meV per dimer site. Like the

atomic sites of the $h0(4\times3)$ trench dimer, the dimer sites of the $\beta(2\times4)$ have high As-Ga substitution energies. Since the most stable species configuration identified for the $\beta(2\times4)$ prototype is also the most As-rich, we did not construct a $\beta(2\times4)$ cluster expansion.

Using the two cluster expansions to perform a convex hull analysis, we identified 10 low-energy site configurations of the $h0(4\times3)$ prototype and 5 low-energy site configurations of the $c(4\times4)$ prototype. The predicted energies of these site configurations were confirmed by DFT calculations, and no configurations with lower energies were predicted by the cluster expansions. Figure 3.3 shows depictions of the identified low-energy configurations that can be described within the prototype unit cells. Of the low-energy configurations found, 4 $c(4\times4)$ site configurations are reconstruction ground states of the GaAs (001) surface. The lowest-energy configuration of the $h0(4\times3)$ comes within about 2 meV per integral unit are of being stable and six of its ten low-energy species configurations are within 10 meV of the total convex hull. In addition to the four $c(4\times4)$ reconstruction ground states, the total convex hull of GaAs (001) reconstructions includes the $\beta2(2\times4)$ at slightly lower $x_{\mathrm{As}}^{(xs)}$ and the highly Ga-enriched $\zeta(4\times2)$ reconstruction, the latter of which was not generated by our algorithm but was included as a Ga-rich limiting case.

To study the intermediate-temperature thermodynamics of the $h0(4\times3)$ and $c(4\times4)$ prototypes, we applied Monte Carlo simulation techniques to the optimized cluster expansion of each reconstruction within the semi-grand canonical ensemble. The surface excess grand potential can then be obtained via integration. Figure 3.7(b) shows the surface excess grand potential, σ, for the $h0(4\times3)$ and $c(4\times4)$ prototypes at 525° C in the As-rich regime, with 0 K DFT results shown for comparison. The $c(4\times4)$ becomes significantly more stable relative to the $h0(4\times3)$ as temperature is increased. Additionally, the dimer sites of the two prototypes exhibit a solid solution behavior, where there are no ordering phase transitions at physically relevant temperatures. Trends in the relative stability are illustrated more clearly in Fig. 3.8, which shows $\Delta\sigma = \sigma_{h0(4\times3)} - \sigma_{c(4\times4)}$ in the As-rich regime. We compare 0 K results for the $\beta2(2\times4)$ to the finite temperature results to approximate the chemical potential range over which the $c(4\times4)$ prototype is stable at finite temperature. This is not an unreasonable approximation, since the $\beta2(2\times4)$ cannot become more As-rich without forming antisite defects. Thus, at finite temperatures its stability will not encroach significantly upon that of the As-rich $c(4\times4)$.

3.5 Discussion

A database of the likely reconstruction prototypes of the III–V (001) surface systems will significantly expedite the future study of these material systems. Especially as focus of

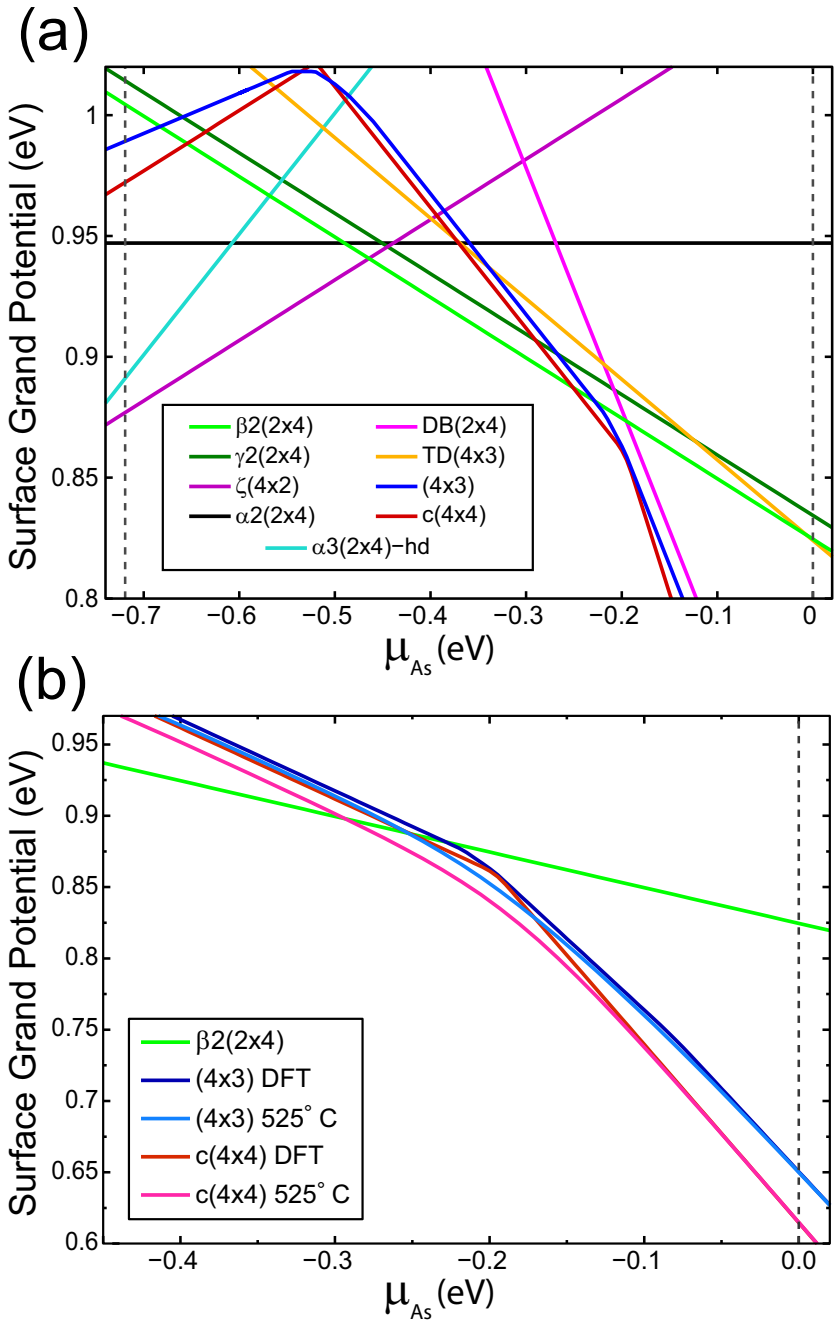


Figure 3.7. (a) Surface excess grand potential energies of GaAs(001) reconstructions within 20 meV of the convex hull, calculated from DFT. For simplicity and consistency, dimer-site configurations of the $h0(4\times3)$ and $c(4\times4)$ prototypes are each shown as a single line. (b) Surface excess grand potential of the $c(4\times4)$ and $h0(4\times3)$ prototypes and the $\beta2(2\times4)$ reconstruction in the As-rich regime. Finite temperature surface grand potentials of $c(4\times4)$ and $h0(4\times3)$ are also shown, calculated with Monte Carlo at 525° C using the optimized cluster expansions. Dashed lines indicate the chemical potential at which the surface becomes unstable relative to bulk As (on the right) and bulk Ga (on the left).

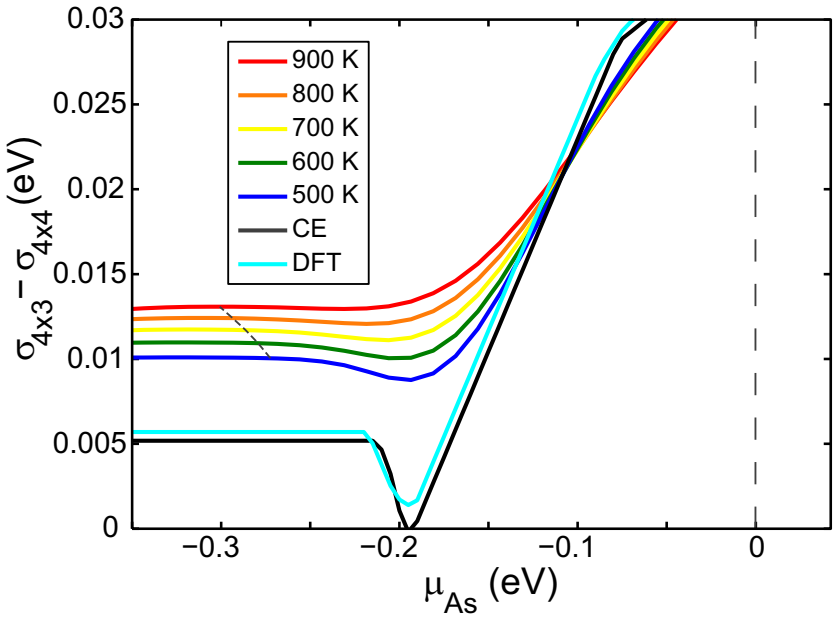


Figure 3.8. The difference in surface grand potential energies of the $h0(4\times3)$ and $c(4\times4)$ prototypes are plotted with respect to As chemical potential. Differences are shown at several temperatures, calculated from Monte Carlo using the optimized cluster expansions. 0 K results are shown comparing cluster expansion predictions to DFT energy calculations.

study within III–V surface science continues to expand into systems with larger numbers of alloying components and wider ranges of alloy compositions, a trial and error approach to constructing reconstruction prototypes is insufficient to characterize these complex surfaces, and the likelihood for redundancy of effort is high. Having a database of prototypes will enable automated characterization using real- and reciprocal-space experimental data along with predictive-adaptive models of electronic structure effects. Additionally, the possibility of exploring structural and configurational degrees of freedom in an efficient and systematic way in order to exhaustively identify likely reconstruction prototypes demonstrates the significant opportunity that exists for constructing similar databases for many other material systems.

For the generated III–V (001) prototypes, a surprising distinction between two basic classes of generated arises from Eq. (3.4) that determines the ability of a reconstruction to satisfy the electron counting rule. It is clear that the excess bilayer charge in Eq. (3.4) must always take an integer value. By contrast, the excess charge donated by the bulk is $\frac{3}{2}N_A$, where N_A is the integral unit area defined Eq. (3.1); this contribution has an integer value for even unit area but has a half-integer value for odd unit area. This implies that a III–V (001) surface reconstruction with odd unit area can never obey the electron counting rule if it follows the basic structural rules we have outlined and has no four-fold coordinated antisites. This unexpected result may account for the lack of any verified (001) III–V surface

reconstructions with odd unit area. Although references to (1×3) and (1×5) reconstructions can be found in the literature, structural models for these reconstructions that are verifiable from first principles are lacking. Experimental characterization techniques that lead to such nomenclature use reciprocal space measurements, which can be significantly affected by thermal excitations of the surface. Thermal disorder along one or both axes can cause phase decoherence at smaller Bragg angles, shortening apparent lattice periodicities by a factor of two or more of the actual periodicity that exists at short length scales, where the ECR may be satisfied.

After considering all (4×3) reconstruction prototypes for that obey the anion-rich structural guidelines, we have found that the $h0(4\times3)$ reconstruction prototype, which was originally proposed by Barvosa-Carter, *et al.*[59] to explain observations on GaSb and AlSb, is very nearly stable on pure GaAs (001). We have determined that the $h0(4\times3)$ structural model is the only candidate that is near the convex hull. As such, it is by far the most likely candidate prototype for the (4×3) reconstruction that is observed in some regimes on $In_xGa_{1-x}As(001)$ and the InAs wetting layer on GaAs(001). It is important to note that the energy differences are within the error range expected from DFT approximations and implementation. As such, the $h0(4\times3)$ prototype is sufficiently close to the convex hull that previously reported experimental observations of a (4×3) or $(n \times 3)$ reconstruction on pure GaAs[47, 58] indicate that dimer-site configurations of the $h0(4\times3)$ prototype may actually be stable at equilibrium over a small region of temperature and chemical potential. Also, differences between prototypes in vibrational free energy, which has not been considered here, is another possible stabilization mechanism. Even if it is not stable, observations of the (4×3) reconstruction on pure GaAs may be due to sample preparation along a thermodynamic path that limits growth kinetics, resulting in a metastable $h0(4\times3)$ surface.

Although the finite temperature behavior of the $c(4\times4)$ has been investigated previously[61], the physical model used was not predictive and did not explore the entire configurational phase space and thus would be unsuitable for comparing the stability of two different reconstructions at finite temperature. Additionally, the very small energy difference between $h0(4\times3)$ and $c(4\times4)$ prototypes was apparently not realized at the time, and no comparison was made between the two prototypes. Our method identifies the same $c(4\times4)$ ground states predicted by the trial-and-error enumerative method employed in that work, as well as a $N_A = 16$ supercell that is a ground-state configuration containing four Ga-As heterodimers.

In addition to the two reconstructions on which we have focused most intently, our methods have yielded a number of other low-energy reconstructions that are less than 50 meV from the convex hull (with surface excess grand potentials plotted in Fig. 3.7). These represent structures which might occur as metastable phases, along surface phase domain

boundaries, or as isolated defects of more stable reconstructions. Alternatively, they might occur as intermediate structures during layer-by-layer growth. Structural illustrations of these reconstructions, as well as the ground states, are shown in Fig. 3.6.

The non-ground-state structures that are shown are the cation-rich $\alpha 3(2\times4)$ with heterodimer termination;[62] the $\beta 2(2\times4)$-a, which is an alternate dimerization of the $\beta 2(2\times4)$ and may occur as a low-energy defect; the $\alpha 2(2\times4)$, which is well-known as a ground-state reconstruction of InAs(001); the $DB(2\times4)$, which has a prominent dimer backbone; and the $TD(4\times3)$, which is similar in structure to the $h0(4\times3)$ but taller, with a prominent double dimer feature. A careful survey of the literature shows that most of these structures have been considered previously, either as dilute defects or as equilibrium reconstructions of other materials systems. However, in the case of the $DB(2\times4)$[59] or the $TD(4\times3)$[17] the structures were proposed only to explain local defects of $(n\times3)$ reconstructions in STM micrographs; no names or first principles energy analyses could be found for these reconstructions.

In the Ga-rich regime of the GaAs (001) surface we confirmed that the $\zeta(4\times2)$ reconstruction is stable. This result is consistent with published results for the Ga-rich GaAs(001) surface[63]. The $\zeta(4\times2)$ is the only well-accepted reconstruction ground state that violates the rules we have enumerated, due to the sp^2 bond character of its Ga-Ga bonds. It should be noted that the $\zeta(4\times2)$ was originally derived from a previously proposed structure, the $\beta(4\times2)$, which does obey our structural rules. The related $\zeta(4\times2)$ structure can be obtained by perturbing a six-atom structural motif of the $\beta(4\times2)$ by $a/4$ along the [110] direction and then using DFT energy minimization to relax the ion positions[63].

At the Group III-rich extreme, the necessary III–III bonds result in under-filling of some $\sigma(sp^3)$ orbitals, leading to a combination of $\sigma(sp^3)$ and $\sigma(sp^2)$ bonding orbitals to be energetically preferred. The resultant sp^2 character of the surface Group III bonds manifests in more trigonal bond angles, giving rise to fundamentally different surface motifs, such as those seen in the $\zeta(4\times2)$. Although our method could be applied to these bonding environments, for now we have chosen to focus on the Group V-rich regime, where more application-driven interest is focused. Considering that the $\zeta(4\times2)$ can be constructed by perturbing and relaxing the atom positions of the $\beta(4\times2)$, we suggest that a possible approach to exploring the Group III-rich regime would be to attempt various translational perturbations of promising structural candidates that obey our proposed structural rules. Specific perturbations should be chosen in an attempt to free the structures from shallow minima of the potential energy surface and promote trigonal bonding of locally Group III-enriched features. Physically motivated guidelines similar to those we have laid out for the Group V-rich regime could be used to identify structures most likely to benefit from perturbation toward trigonal-type bond-

ing and to identify appropriate perturbations based on the symmetry of the reconstruction prototype.

3.6 Summary

We have presented a comprehensive and tractable approach to resolving gaps in understanding of surface reconstruction phase diagrams where an experimentally observed but poorly characterized reconstruction phase exists, or even in lieu of such evidence. This approach was demonstrated by conducting a study of the zincblende III–V (001) surface in the Group V-rich regime, for which we have codified the structural rules obeyed by observed structural motifs. A comprehensive database of likely reconstructions was efficiently constructed using our algorithm, providing a valuable resource for analyzing surface phenomena of these systems and for predicting new equilibrium reconstructions for poorly-understood systems.

Our in-depth study of the GaAs (001) surface using this database of reconstruction prototypes represents the most comprehensive analysis of GaAs (001) reconstruction stability to date, not only by taking into consideration all likely reconstruction prototypes, but also by rigorously including the effects of configurational disorder of homo-/hetero-dimer site occupancy at finite temperature. From this analysis we have demonstrated that the $h0(4\times3)$ reconstruction prototype is much more energetically competitive with the $c(4\times4)$ than previously thought. Moreover, because our analysis included all charge-neutral (4×3) reconstruction prototypes that are likely, given our understanding of the III–V (001) surface, we have demonstrated that the $h0(4\times3)$ reconstruction prototype is the most likely candidate for the (4×3) reconstruction observed on $In_xGa_{1-x}As$ (001) and at low temperature on GaAs (001).

CHAPTER IV

The Combined Roles of Configurational and Vibrational Disorder on Predicted Surface Reconstruction Stability

The broad importance of III–V compound semiconductors to device applications is due to their wide variety of realizable, tunable alloys. However, because useful III–V alloys and heterostructures require precisely controlled layer-by-layer synthesis, their resultant properties and quality are largely limited by our understanding of structure and ordering phenomena at the crystalline growth surface. The structure and composition of the surface play an important role in the injection of point defects and antisites, particularly at low temperatures, as has been indicated by the link between photoluminescence intensity and growth surface[64]. Increasingly, these surface-induced defects are exploited for their beneficial consequences, as in low-temperature grown (LTG) GaAs, which is an important material for THz emitters and detectors[65]. LTG GaAs is typically grown in the [001] orientation at temperatures below 300°C and under As-rich conditions[66, 67]. In this regime excess surface As becomes kinetically trapped in the growing film, incorporating at up to 1 at-% above bulk stoichiometry in the form of both antisite defects and metallic As precipitates[66, 68]. The high charge mobility and very low carrier lifetime in defected LTG GaAs make it particularly well suited for THz-range heterodyne photomixers[69, 70]. This low-temperature regime is also used for growing alloy films in low-solubility systems, such as $Ga_{1-x}Bi_xAs$[71] and the ferroelectric semiconductor $Ga_{1-x}Mn_xAs$[72].

Significant theoretical study of GaAs(001) has previously identified only two stable As-rich surface reconstructions, relative to bulk stoichiometry: the $c(4\times4)$ and $\beta2(2\times4)$, which are illustrated in Figs. 4.1(a) and 4.1(b), respectively[73, 62, 20]. However, theoretical predictions of reconstruction stability contradict experimental observations, which indicate the existence of a stable "×3" surface reconstruction on GaAs(001)[47, 58, 64]. Scanning tunneling microscopy (STM) experiments strongly suggest that the GaAs "×3" surface is actually

comprised of a (4×3) reconstruction[58]. Any complete description of GaAs(001) surface stability must account for this (4×3) reconstruction.

In this chapter we conduct a rigorous and comprehensive theoretical analysis of surface reconstruction stability on GaAs(001). We identify the low-energy As-rich GaAs(001) reconstruction prototypes and calculate their finite-temperature surface free energies from first principles, taking into account the combined effects of configurational disorder and vibrational excitations. By relating the surface free energies to the finite-temperature partial pressures of As_4, and estimating errors arising from approximations of DFT, we construct the GaAs(001) surface phase diagram from first principles, finding good agreement with experiment. By revealing the mechanisms that govern reconstruction stability on GaAs(001), our results provide crucial insight about the role of thermal excitation on the finite temperature surface reconstruction stability.

4.1 Entropy Considerations on the GaAs(001) Surface

Traditionally, surface reconstruction stability has been determined by comparing energies obtained from electronic structure calculations for a collection of candidate reconstructions, all of which are conjectured *a posteriori*, sometimes based on scant evidence. Constructing a candidate structure and verifying its stability is complicated by a number of factors. On a multicomponent surface, a well-specified surface reconstruction is comprised of a reconstruction prototype, which defines the bonding topology, along with a species configuration that decorates it; each reconstruction prototype can be decorated with many different, but nearly degenerate, species configurations. Energy differences between reconstructions are calculated using density functional theory (DFT). Despite reliably predicting many groundstate properties of III–V compounds, conventional DFT provides no direct information about thermally excited behavior. These effects, including lattice vibrations and fluctuations in species configuration, contribute an entropic component to the surface free energy that may alter reconstruction stability. Consequently, zero-K surface enthalpies calculated via DFT are an insufficient predictor of reconstruction stability. At typical synthesis temperatures ($k_BT \sim$ 50–80 meV) a small entropy difference between surface reconstructions can overwhelm the difference in their surface enthalpies, which is sometimes less than 10 meV/$A_{(1\times1)}$, where $A_{(1\times1)}$ is the area of the (1×1) surface cell. Such a situation can result in *entropic* stabilization of one reconstruction relative to another.

Using the enumeration procedure presented in Ch. III, we have succesfully catalogued all plausible III-V surface reconstructions. This presents us with the unprecedented capability to conduct an exhaustive search for a low energy (4×3) reconstruction on GaAs(001). Among

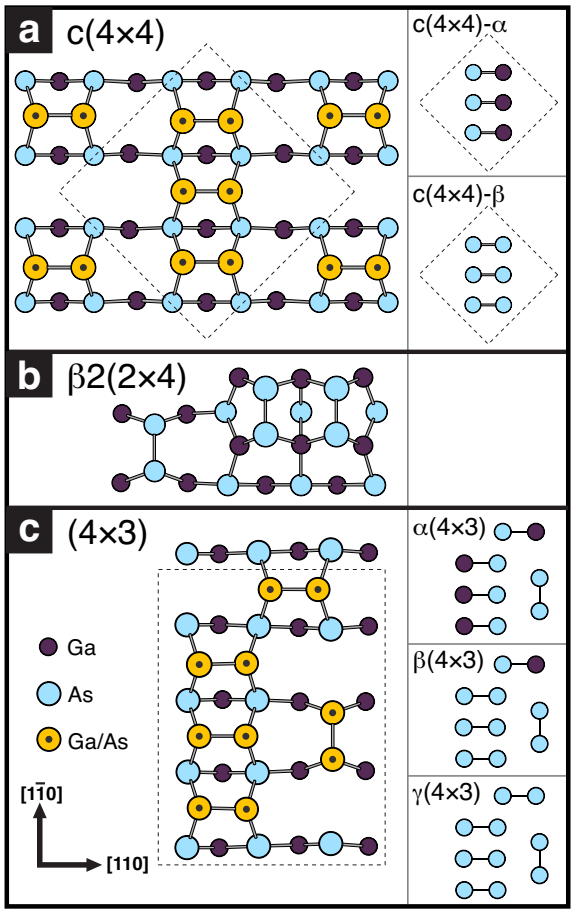


Figure 4.1. The low-energy As-rich reconstruction prototypes of GaAs(001). In (a) and (c), gold circles with black dots indicate variable sites, which can be either Ga or As. Inset images illustrate the most stable Ga/As species configurations of these sites.

the 124 conceivable (4×3) reconstruction prototypes that are charge balanced, DFT calculations indicate that the prototype that is lowest in energy is the one depicted in Fig. 4.1(c). Two species configurations of this prototype, the $\alpha(4\times3)$ and $\beta(4\times3)$, have been predicted to be stable on the GaSb(001) and AlSb(001) surfaces[59]. These are illustrated in Fig. 4.1(c) inset. However, DFT predicts all species configurations of this (4×3) to be metastable on GaAs(001) relative to either the $\beta2(2\times4)$ or a species configurations of the $c(4\times4)$ prototype. The energy difference between the (4×3) and $c(4\times4)$ prototypes is very small, though; they are separated by 6 meV/$A_{(1\times1)}$ or less over a wide range of chemical potential. This energy difference is much smaller than exists between the $c(4\times4)$ and any other charge-balanced (4×3) prototype. Consequently, the only (4×3) prototype that merits consideration is the one depicted in Fig. 4.1(c).

The multicomponent surface is an open system at fixed temperature, such that the surface

free energy $\gamma(T, \mu_{\mathrm{As}})$ is minimized at equilibrium, where T is temperature, and μ_{As} is the As chemical potential. γ is comprised of contributions from electronic structure (i.e., the zero-K surface enthalpy), configurational excitations, and lattice vibrations. Electronic structure calculations were performed using a surface/slab geometry within the DFT local density approximation (LDA).

On GaAs(001), configurational entropy arises from the many possible ways of arranging Ga and As over the undercoordinated surface sites of the (4×3) and c(4×4). The undercoordinated sites have three sp^3-hybridized interatomic bonds; the fourth sp^3 orbital becomes a "dangling bond". As suggested by the electron counting rule heuristic[1], the dangling bond allows Ga and As to substitute at these sites without affecting charge balance. Consequently, thermal excitations can sample many different configurations of Ga and As on the lattice of tricoordinate sites. Figs. 4.1(a) and (c) indicate the sites that can undergo low-energy species substitution in the (4×3) and c(4×4) prototypes. Energies of many Ga/As configurations were calculated and, via the cluster expansion (CE) formalism[29, 30, 31] described in Ch. II, were used to construct an effective Hamiltonian that describes the energy of any Ga/As configuration of the undercoordinated sites. (4×3) and c(4×4) reconstructions were considered separately, resulting in two prototype-specific cluster expansions.

The surface vibrational free energy of each species configuration is evaluated using the Einstein model of treating each ion as an isolated harmonic oscillator centered at its lattice site. This approximation has been used previously to study other covalently-bonded surface systems[74]. Vibrational frequencies were obtained from the diagonalized site Hessian tensors, calculated using the finite-difference implementation in VASP. Calculated vibrational frequencies are nearly independent of surface species configuration, allowing the vibrational free energy of each surface substitution site to depend only on the on-site and nearest-neighbor species occupations. DFT calculations show the vibrational free energy of lower layers is independent of Ga/As surface configuration to within ∼1 meV/$A_{(1\times1)}$ at synthesis temperatures. The CE effective Hamiltonian for the combined configurational energy and vibrational free energy of each reconstruction prototype can be evaluated efficiently, making it well-suited to Metropolis Monte Carlo (MC).

4.2 Predicting GaAs (001) Reconstruction Phase Stability at Finite Temperature

Surface free energies were calculated from equilibrium MC simulations independently for the As-rich (4×3) and c(4×4) reconstruction prototypes by integrating $\langle x_{\mathrm{As}}^{(xs)} \rangle = (\partial\gamma/\partial\mu_{\mathrm{As}})_T$ with respect to μ_{As} at constant T, where $x_{\mathrm{As}}^{(xs)}$ is the surface excess As coverage. We also

considered $\beta2(2\times4)$, depicted in Fig. 4.1(b), in order to bound the range of μ_{As} relevant to As-rich growth. DFT calculations indicate a large energy penalty for Ga/As substitution in the $\beta2(2\times4)$, due to the consequent formation of high-energy Ga–Ga bonds. Configurational contributions to the temperature dependence of $\gamma_{\beta2(2\times4)}(\mu_{\mathrm{As}}, T)$ were thus neglected, leaving only vibrational excitations. The GaAs(001) surface phase diagram was constructed by minimizing the surface free energy over $\gamma_{\beta2(2\times4)}$, $\gamma_{(4\times3)}$, and $\gamma_{c(4\times4)}$, with respect to μ_{As} and T. First-order phase boundaries occur where minimal surface free energies cross. From the equilibrium conditions among bulk metallic As, bulk GaAs, and polyatomic As_m gas, the phase boundaries can be expressed as a function of normalized As_m partial pressure

$$\tilde{p}_{\mathrm{As}_m} = \frac{p^{(\mathrm{As}_m)}}{p_0^{(\mathrm{As}_m)}} = \exp\left[m\frac{\mu_{\mathrm{As}} - g_{\mathrm{As}}^{(bulk)}}{k_B T}\right], \tag{4.1}$$

where $p_0^{(\mathrm{As}_m)}$ is the vapor pressure of As_m over bulk As at temperature T, and $g_{\mathrm{As}}^{(bulk)}$ is the Gibbs free energy of bulk As, per atom. We considered only As_4; use of As_2 would only alter the slopes of phase boundaries on an Arrhenius depiction of the predicted phase diagram. At $\tilde{p}_{\mathrm{As}_m} \geq 1$, GaAs becomes thermodynamically unstable relative to solid As, yielding a natural upper limit for $\tilde{p}_{\mathrm{As}_m}$.

Monte Carlo free energies predict metastability of the (4×3) relative to the $c(4\times4)$ and $\beta2(2\times4)$ over the entire range of $\tilde{p}_{\mathrm{As}_m}$ and T. However, the (4×3) is nearly stable at low temperature, lying only 6 meV/$A_{(1\times1)}$ above the $c(4\times4)$ at 400 K. Despite this, the calculated finite-temperature phase diagram fails to explain established results; not only is the (4×3) observed experimentally in the As-rich regime[47, 58], but also our exhaustive search for a stable (4×3) reconstruction found no other plausible candidates.

4.2.1 Estimating the Effect of DFT Error

In order to explore this contradiction between theory and experiment, we established a lower bound on the error of DFT surface energies by comparing the relative surface energies of 9 low-energy $c(4\times4)$ and (4×3) reconstructions calculated using LDA to the same relative surface energies calculated using the generalized gradient approximation (GGA) for quantum exchange and correlation, instead of LDA. Specificially, we considered the quantity

$$\epsilon_i = \left(\gamma_i^{(LDA)} - \gamma_{ref}^{(LDA)}\right) - \left(\gamma_i^{(GGA)} - \gamma_{ref}^{(GGA)}\right), \tag{4.2}$$

where γ_i and γ_{ref} are the surface free energies calculated or reconstruction i and for the reference reconstruction, respectively. The reference reconstruction was taken to be the re-

construction with the lowest calculated energy, according to LDA. To estimate the error bounds, we calculated the standard deviation of ϵ_i at $\mu_A s = 0$ and $T = 0$, for the reconstructions considered. The standard deviation $\Delta\epsilon$ can be shown mathematically to be independent of both $\mu_A s$ and the choice of reference reconstruction. It essentially describes the scatter in predicted energy between GGA and LDA, irrespective of an average shift in predicted energy. Our calculations yield a value for $\Delta\epsilon$ of 9.1 meV/$A_{(1\times1)}$. Because the variance of a n-trial sample is statistically underestimated by a factor of $(n-1)/n$, we estimate an error of at least $\pm$9.7 meV/$A_{(1\times1)}$ for our calculated surface energies.

4.2.2 The Calculated GaAs(001) Phase Diagram

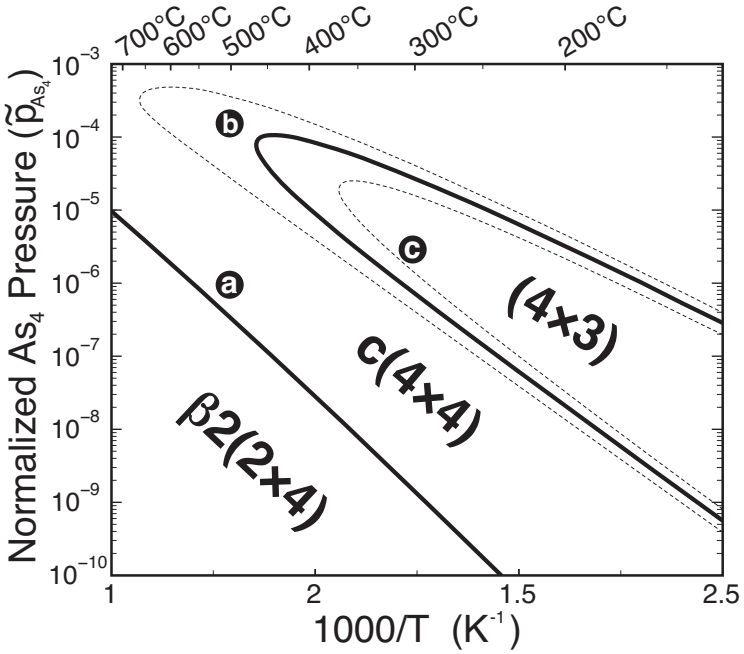


Figure 4.2. Calculated GaAs(001) surface reconstruction phase diagram, as a function of inverse temperature and normalized partial pressure. The (4×3) surface free energy has undergone a negative shift of 8.5 meV/$A_{(1\times1)}$ relative to that of the c(4×4). Alternate phase boundaries are shown for 8-meV/$A_{(1\times1)}$ and 9-meV/$A_{(1\times1)}$ negative shifts. Circles (a), (b), and (c) indicate the thermodynamic parameters corresponding to MC snapshots in Fig. 4.4.

Within the estimated surface free energy error bounds, we analyzed the sensitivity of the computed GaAs(001) phase diagram to a constant shift of the (4×3) surface free energy. Figure 4.2 shows the surface reconstruction phase diagram obtained with a 8.5-meV/$A_{(1\times1)}$ negative shift of $\gamma_{(4\times3)}$, which most closely resembles experimental results. The stability of the (4×3) is very sensitive to small shifts in its surface free energy, as demonstrated by the alternate phase boundaries in Fig. 4.2, shown for negative shifts of 8.0 and 9.0 meV/$A_{(1\times1)}$.

A negative shift larger than 9.5 meV/$A_{(1\times1)}$ qualitatively changes the phase diagram, as the (4×3) overwhelms the region of $c(4\times4)$ stability bordering the $\beta2(2\times4)$.

4.2.3 Relative importance of vibrational and configurational excitations

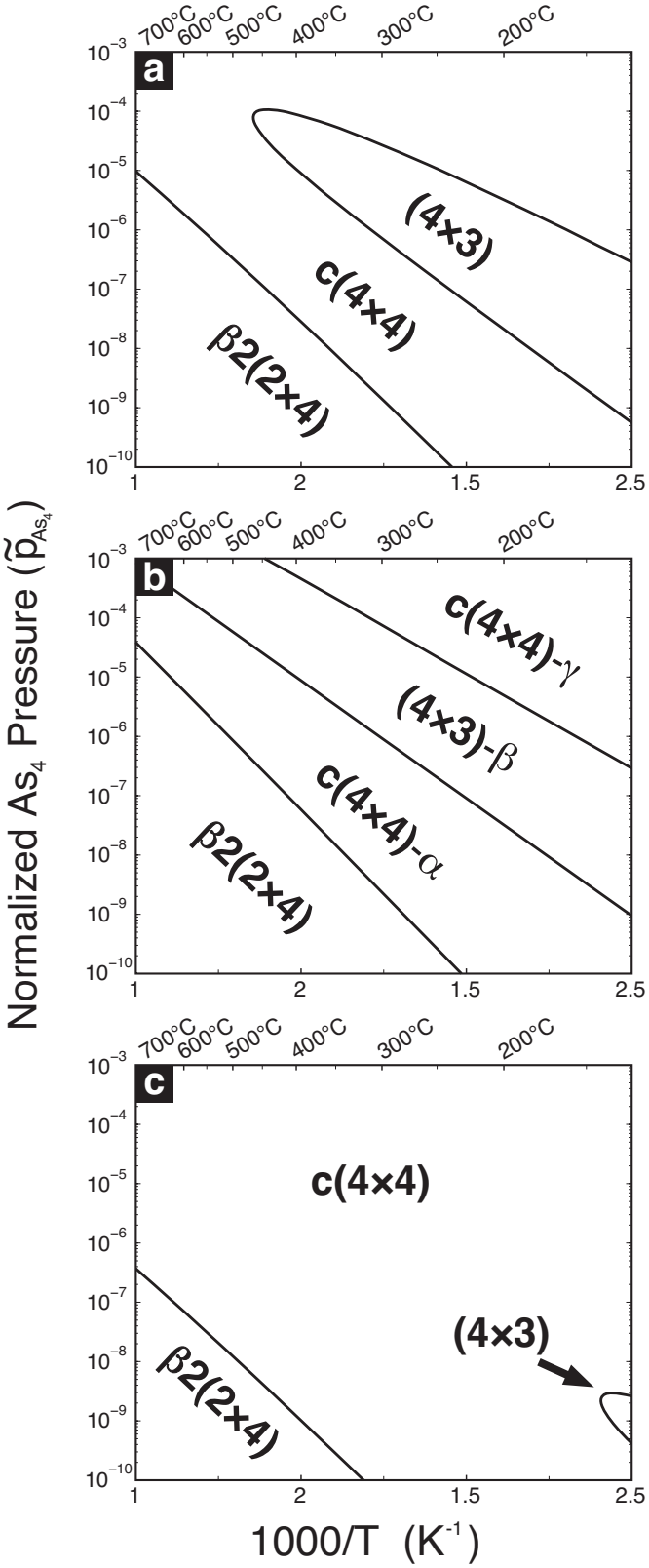
We can contrast the effects of vibrational and configurational excitations on GaAs(001) reconstruction stability to better understand their individual roles in determining the predicted GaAs(001) surface phase diagram, shown in Fig. 4.3(a), by considering each effect independently. To study the affect of vibration alone, we calculated slab vibrational free energies for the configurational groundstates predicted by the cluster expansion Hamiltonians, and then calculated the surface free energy for each, following the formalism described in Ch. II. The surface free energy γ was calculated as the surface excess of characteristic potential Φ. The bulk reference energy $\phi^{(bulk)}_{\mathrm{GaAs}}$ also includes the bulk vibrational free energy, calculated using the Einstein model. The reference state for μ_{As} is set by the per-atom Gibbs free energy of bulk As, which also includes vibrational contributions. The reconstruction phase diagram produced by this vibration-only model is shown in Fig. 4.3(b), where the surface free energy of each (4×3) configuration has a negative shift of 8.5 meV/$A_{(1\times1)}$). The (4×3) is well-stabilized in this model, even at high temperature, where the (4×3) is not experimentally observed.

To study the affect of configuration alone, we calculated slab configurational free energies using Monte Carlo to integrate the partition function without vibrational contributions. This is simply the traditional method for using a cluster expansion Hamiltonian to predict thermodynamic properties. Surface free energies were again calculated as the surface excess characteristic potential, but the bulk GaAs and bulk As reference energies did not include vibrational contributions in this case. The reconstruction phase diagram thus obtained, and with a 8.5 meV/$A_{(1\times1)}$) negative shift to $\gamma_{(4\times3)}$, is shown in Fig. 4.3(c). The (4×3) is stable only at low temperature, as is observed experimentally, but its region of stability is greatly diminished without vibrational effects, such that the (4×3)–$c(4\times4)$ transition occurs at a much lower temperature than has been measured.

4.3 Discussion

The phase diagram in Fig. 4.2 bears a strong resemblance to the comparable region of the experimental phase diagram reported by Daeweritz, *et. al*[47]. In that work, the phase diagram was measured with reflection high-energy electron diffraction (RHEED) during epitaxial growth. In particular, our phase diagram calculated with the 8.5-meV/$A_{(1\times1)}$ negative shift indicates that the (4×3) is stable only at low temperatures, and the (4×3)-

Figure 4.3. GaAs(001) phase diagrams obtained by considering (a) vibrational and configurational excitations simultaneously, (b) vibrational excitations alone, and (c) configurational excitations alone. In all three cases, (4×3) free energies have been negatively shifted by 8.5 meV/$A_{(1\times1)}$.



$c(4\times4)$ transition temperature of 460°C agrees with the maximum temperature at which the (4×3) has been observed experimentally. Also, both the experimental and computed phase diagrams exhibit a range of As_4 isobars that pass, with increasing temperature, from (4×3) to $c(4\times4)$ and then to (2×4). Nevertheless, it is difficult to draw precise comparisons with the experimental phase diagram. Due to the use of diffraction techniques to measure surface periodicities, experimental measurements of various $(n\times3)$ reconstructions, with $n<4$, are likely due to a (4×3) reconstruction interrupted by long-range disorder[16]. Also, the experimental phase diagram likely exhibits metastable reconstructions because it was measured during growth, when the physical surface is away from equilibrium. This is especially likely considering that the small energy difference between $c(4\times4)$ and (4×3) likely provides a relatively weak driving force to nucleate reconstruction domains once the $c(4\times4)$–(4×3) boundary is crossed. This may also explain how the $c(4\times4)$ surface is so easily preserved during experimental sample quenching. Moreover, a combination of kinetic effects and lack of long-range order may account for the observations by Daeweritz, *et. al* of a (2×1) reconstruction along the $c(4\times4)$–$\beta2(2\times4)$ phase boundary, a region where the very small energy differences among all three reconstruction prototypes suggest a high likelihood of disorder and/or hysteretic effects.

As the first study to combine vibrational and configurational excitations to analyze finite-temperature surface reconstruction stability, our results elucidate the importance of thermal disorder in reconstruction stabilization. As is evident from Fig. 4.3, configurational and vibrational excitations have opposing effects. While vibrational excitations improve stability of the (4×3) prototype, particularly at low temperature, configurational excitations favor stability of the $c(4\times4)$ reconstruction. Vibrational effects favor the (4×3) due to its high density of undercoordinated surface sites, which have lower vibration frequencies than subsurface sites. The strong influence of configuration on $c(4\times4)$ stability is partly due to its ordering phenomena at low temperature, where $c(4\times4)$–β short-range order occurs on the As-rich side of Fig. 4.2 and $c(4\times4)$–α short-range order occurs proximal to the $\beta2(2\times4)$ phase boundary, as shown in Fig. 4.4(a). The metastable $c(4\times4)$ would transition between these two regimes in the region where (4×3) is stable. At low temperature the $c(4\times4)$ ordering transition is abrupt (though not sufficiently so as to be first- or second-order), exacting an energetic penalty on the $c(4\times4)$. At higher temperature, the $c(4\times4)$ ordering transition occurs more gradually, over a region of configurational disorder, as demonstrated by Fig. 4.4(b). As Fig. 4.4(c) shows, the (4×3) predominantly exhibits $\beta(4\times3)$ short-range order over this same region. These findings suggest strategies to control surface disorder and thereby influence defect incorporation during growth. For example, a $c(4\times4)$ surface obtained by decreasing As overpressure along an isotherm originating from a (4×3) surface will result in a relatively

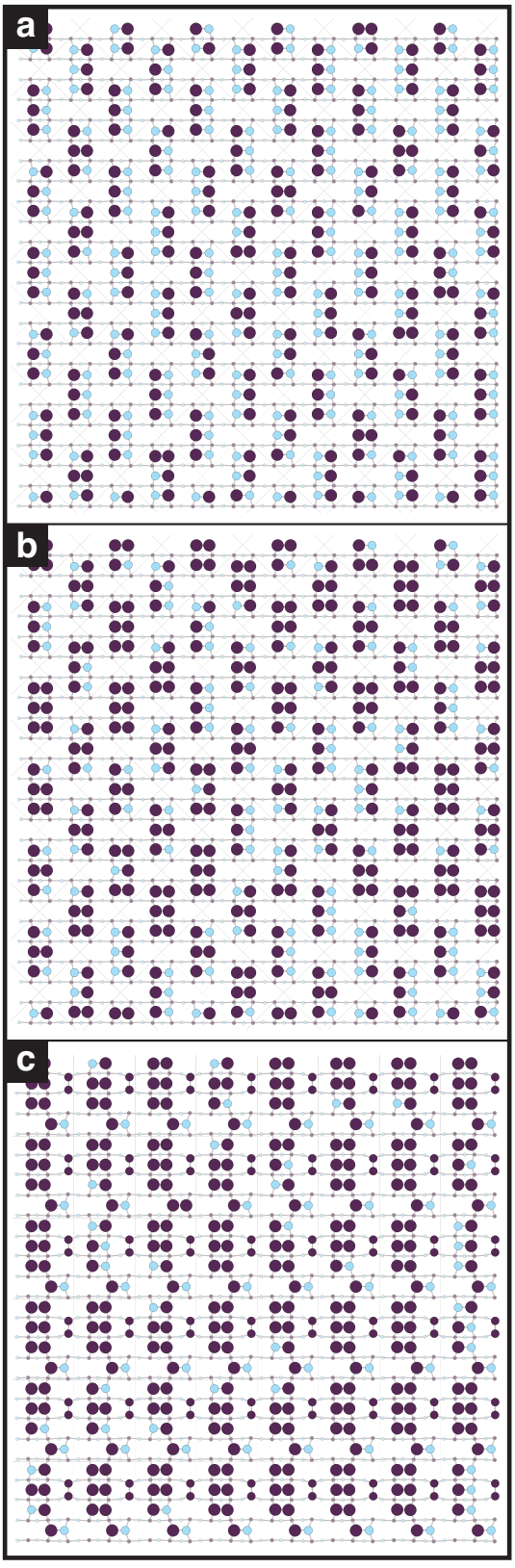


Figure 4.4. Instantaneous snapshots of the MC simulation cell at the three thermodynamic points indicated in Fig. 4.2. The illustrated regimes are (a) $c(4\times4)$–α short-range order, (b) $c(4\times4)$ disorder, and (c) $\beta(4\times3)$ short-range order.

well-ordered $c(4\times4)$-α surface configuration. Conversely, approaching the $c(4\times4)$ by increasing temperature along an isobar originating on a (4×3) surface should produce a disordered $c(4\times4)$ surface.

4.4 Summary

We have constructed a surface phase diagram for As-rich GaAs(001) from first principles, assuming a well-defined lower bound on DFT error. In the course of doing so, we considered all possible semiconducting (4×3) reconstruction prototypes, and found only one prototype that was stable, within the established error bounds. Finite-temperature surface free energies of the stable As-terminated surface phases, calculated by rigorously accounting for vibrational and configurational excitations, demonstrated that our proposed (4×3) prototype can be stabilized only at low temperatures, while the $c(4\times4)$ prototype is entropically stabilized at elevated temperatures, in excellent agreement with experiment. These results finally resolve the characterization problem of the GaAs(001) (4×3) surface phase. Furthermore, they provide a solid thermodynamic foundation from which to understand and manipulate incorporation of point defects and antisites during low temperature growth by controlling degrees of configurational disorder via the predominant surface reconstruction.

CHAPTER V

Effects of Alloying on InAs/GaAs(001) (2×4) Reconstruction

5.1 Introduction

In device design and other applications it is often necessary to alloy two or more compounds to engineer a material within a set of desired parameters. This significantly increases the dimensionality of the associated phase space by creating new configurational degrees of freedom associated with the many possible ways of arranging the alloy constituents over their sublattice sites. In addition, alloying may introduce interactions arising from atomic size mismatch strain, which is present if the alloy constituent species have significantly different bonding radii[75]. The strain fields arising from atomic size mismatch may interact strongly over distances of several bond lengths to induce correlations between the occupancies of non-neighboring sites. Because of these and other effects that necessitate the sampling of many configurations and large length scales, comprehensive stability studies of realistic alloyed covalent surface systems have only recently become possible.

In this chapter, we study disorder in surface reconstructions arising from alloying and finite temperature effects with a first principles cluster expansion and Monte Carlo simulations. We focus on a prototypical surface of a ternary III-V semiconductor alloy in order to understand the role that alloying and atomic size mismatch strain plays in surface reconstruction stability and short-range order at covalent surfaces. We choose to consider a thin layer of InAs alloyed on a zincblende GaAs (001) substrate, since this system is relatively well studied and allows us to examine the role of atomic size mismatch on surface order at finite temperature, due to the large cation size difference between In and Ga.

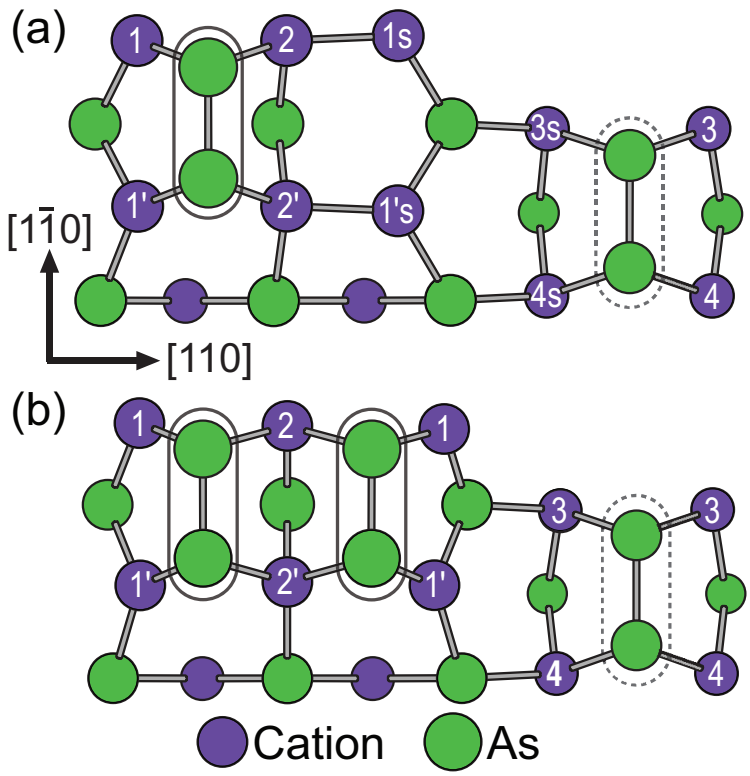


Figure 5.1. (a) The $\alpha 2(2\times4)$ and (b) $\beta 2(2\times4)$ reconstructions. Cation sites under consideration are labeled according to symmetric equivalence, assuming periodicity of the unit cell. Solid ellipses indicate row dimers, and dashed ellipses indicate the trench dimer.

5.1.1 InGaAs (2×4) Reconstruction

To model alloying and local strain effects we focus on a specific class of reconstruction observed in GaAs and InAs, encompassing the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ reconstructions[76, 62] (Fig. 5.1). They are common to many III-V compounds and possess several shared structural features. The most notable of these are the *dimer row* and the *trench*. The dimer row is the highest feature of the unit cell and consists of either one or two As dimers (which we will call *row dimers*), while the trench is one atomic bilayer below the dimer row and contains a single As dimer (the *trench dimer*). The $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ differ from each other only in that the $\alpha 2(2\times4)$ has one row dimer, whereas the $\beta 2(2\times4)$ has two.

The $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ reconstructions have been shown experimentally and via first principles calculations to be stable over a continuous range of arsenic chemical potential on InAs (001)[77] and on GaAs (001)[63], although $\alpha 2(2\times4)$ is predicted to be only marginally stable on the GaAs surface. Alloying GaAs with In, however, appears to stabilize the $\alpha 2(2\times4)$. In work by Krzyzewski *et al.*, which images the InAs wetting layer on GaAs, the $\alpha 2(2\times4)$ appears to occur, along with the $\beta 2(2\times4)$[60]. Also, the $\alpha 2(2\times4)$ is predicted to be stabilized upon surface alloying at the $In_{0.5}Ga_{0.5}As$ lattice parameter[78]. Because of these observations and predictions, we can consider these two reconstructions independently from other stable reconstructions within chemical potential intervals near the $\alpha 2(2\times4)$-$\beta 2(2\times4)$ transition.

To study the transition between $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ at finite temperature it is necessary to first identify important configurational degrees of freedom. We can distinguish between an anion sublattice and a cation sublattice at the surface, as is typically done for the bulk crystal. The anion sublattice, which we shall refer to here as the *dimer sublattice*, consists of the As row dimers. The $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ reconstructions are formed by placing one and two dimers per unit cell, respectively, on the row dimer sites. The $\beta 2(2\times4)$ reconstruction does not exhibit configurational degrees of freedom on the dimer sublattice as all its row dimer sites are occupied. In the $\alpha 2(2\times4)$ reconstruction, however, half the row dimer sites are unoccupied, allowing for a large number of surface dimer arrangements and thereby creating the potential for substantial disorder at finite temperature. A row dimer coverage of less than 50% must have some unit cells without any row dimers, in violation of the electron counting rule[1]. Regions where this occurs are assumed to be unphysical, and are indicated as such in the results that follow. The cation sublattice is comprised of the six surface cation sites (those directly below the dimer row) and four trench cation sites (those directly below the trench dimer), which correspond to the labeled cation sites of Fig. 5.1. These sites take either a Ga or In, providing the potential for surface alloying. We limit consideration to these ten cation sites because the substitutional energy of other subsurface

sites is significantly higher, as determined from density functional theory (DFT) calculations by Cho *et al.*[79] and confirmed by our own calculations. This result reflects experimental observations that show a tendency for In to surface-segregate on GaAs (001)[80].

Together the dimer and cation sublattices form our surface model and can be viewed as a thermodynamic system that equilibrates with a gas phase and a bulk GaAs phase having a dilute concentration of In. The thermodynamic boundary conditions of the surface are therefore constant temperature and chemical potentials.

5.1.2 Atomic Size Mismatch Strain

Atomic size mismatch strain arises when an atomic species with a large bond radius is substituted on the crystal sites of a species with a smaller bond radius, or vice versa. In the case of large mismatch, the substitution typically causes phase separation or ordering of the bulk alloy, except in the dilute limit. However, at two-dimensional interfaces or surfaces, mismatch strain enhances intermixing between species which are otherwise immiscible[81, 82], to the extent of inducing surface order[83]. In the case of InGaAs/GaAs, In is the larger species, with a bond radius 7% larger than that of Ga. Substituting In for Ga on a GaAs surface compresses surrounding bonds, creating a local strain field. The strain fields resulting from alloy substitution interact over distances of several bond-lengths to give rise to short-range site correlations or periodic orderings not present in the unalloyed system.

As an illustration, consider the substitution of an In atom between the two dimers of the $\beta 2(2\times4)$ (Fig. 5.2(b)). Density functional theory (DFT) calculations predict that it distorts the lattice, moving the dimers out and away from each other. Alternatively, In atoms substituted opposite the lone row dimer of the $\alpha 2(2\times4)$ (Fig. 5.2(a)) relieve tensile stress inherent in the Ga-Ga back-bonds, allowing the row dimer and the surface Ga atoms below it to relax toward their corresponding positions in the $\beta 2(2\times4)$. Thus, at dilute concentrations inserting an In atom into the $\alpha 2(2\times4)$ is energetically preferred to inserting an In atom into the $\beta 2(2\times4)$.

Work by Bickel *et al.*[75] examined this mechanism as an explanation for experimentally observed local order of the $\alpha 2(2\times4)$. In that first principles study, they found that the $\alpha 2(2\times4)$ ground states at low and intermediate In concentration have a "zig-zag" arrangement of the dimer row along the $[1\bar{1}0]$ axis. This was attributed to the strong repulsive interaction between the size-mismatched In atoms and the lone $\alpha 2(2\times4)$ row dimer, which causes In pairs to occupy the sites opposite the dimer. Furthermore, the tendency of In atoms to maximize their distance from other In atoms results in groupings of In pairs and row dimers to orient in an alternating pattern along $[1\bar{1}0]$. This chapter builds upon that work by considering the trench cation sites in addition to the surface cation sites and by explor-

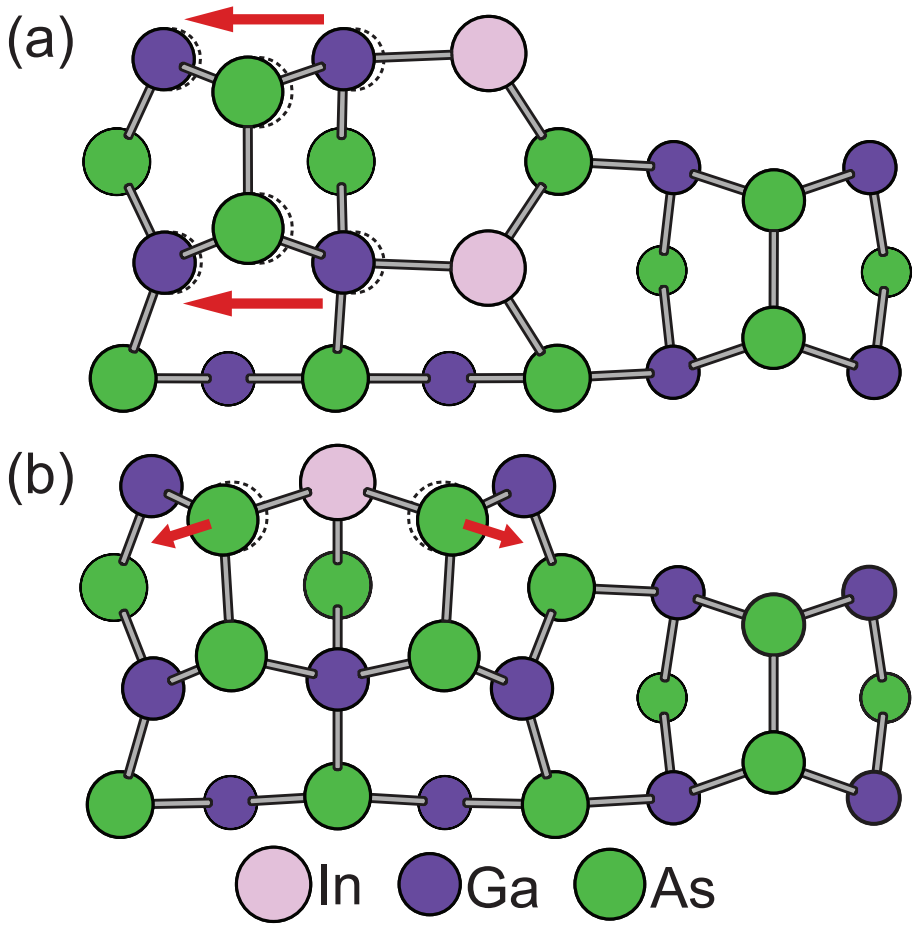


Figure 5.2. An illustration of lattice distortion due to atomic size mismatch upon the substitution of In on the cation sublattice; dashed circles represent atomic positions before substitution. Atom positions were determined via energy minimization from first principles. (a) In the $\alpha 2(2\times4)$ size mismatch can relieve strain inherent in cation-cation backbonds, allowing the surrounding structure to relax to positions approximate to those of the pure GaAs $\beta 2(2\times4)$. (b) In the $\beta 2(2\times4)$ size mismatch leads to significant distortion of the lattice.

ing the transition between the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ reconstructions at finite temperature. We consider interactions among row dimers, In, and Ga, and examine finite temperature effects on surface stability using a cluster expansion parameterized by first principles energy calculations within Monte Carlo simulations.

By using these techniques to explore the configurational phase space of the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ on Inas/GaAs (001), we have identified several basic rules for size mismatch interactions that play a significant role in determining the equilibrium configuration of the system. Understanding these prevailing effects is useful for developing intuition about the system, as one or more of them are evident in all of the results presented in this chapter. They are discussed in greater detail In Sec. 5.3, but are summarized as follows:

It is energetically unfavorable for row dimers to directly bond to In atoms. Because the adsorption of an As dimer constrains the positions of underlying sites, substitution of larger In atoms is energetically unfavorable in surface cation sites bonding to the dimer.

High symmetry cation sites accept In more readily. While Rule I takes precedence at lower As chemical potential and when In is relatively dilute, in the $\beta 2(2\times4)$ reconstruction the sites that sit along the $(1\bar{1}0)$ mirror plane are energetically preferred for In substitution. Because they have higher symmetry than the other cation sites, the cation and dimer sublattices relax more uniformly to accommodate In at these sites while maximizing the distance between In atoms.

Row dimers cause compressive strain at a subset of cation trench sites. Row dimers force the underlying cation sites to relax toward each other, resulting in compressive strain around certain sites in the trench, labeled by a 3 in Fig. 5.1. This significantly increases the substitution energy of these sites relative to similar sites with no nearby row dimer, as is the case near the dimer vacancy of the $\alpha 2(2\times4)$.

Low energy configurations maximize distance between In atoms. Because of their large size relative to Ga, interactions between In atoms are repulsive. For many surface stoichiometries the previous three guidelines usually dominate, but the mutual repulsion of In is important in the dilute In limit or over the range of several unit cells, where dimers and In atoms can arrange themselves such that the first guideline above is satisfied while simultaneously ensuring maximal or near-maximal distance between In atoms.

5.2 Methodology

Of primary interest is the dependence of surface order on alloy stoichiometry and temperature, which motivates the development of an effective Hamiltonian within the cluster

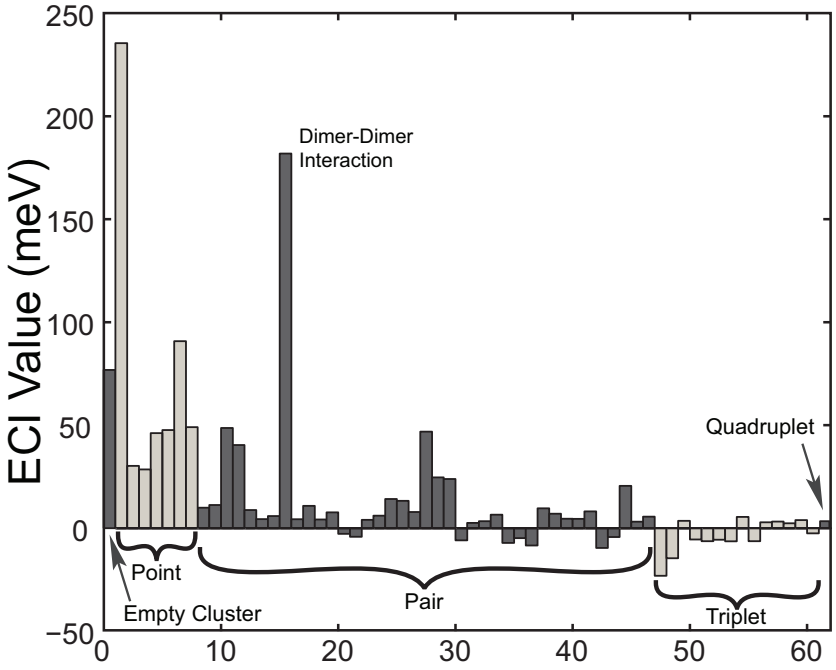


Figure 5.3. The relative magnitudes of all ECI included in the cluster expansion. ECI are grouped by number of sites included; within each group they are ordered by increasing interaction length. Site ECI are ordered with the dimer first, and then 1, 1′, etc., as labeled in Fig. 5.1(a).

expansion formalism to account for the configurational degrees of freedom identified above. The cluster expansion of the configurational energy is chosen as it is a compact and easily evaluated effective Hamiltonian and, with Monte Carlo, allows full consideration of thermal effects due to configurational entropy. These entropic effects, which have hitherto been unaccounted for in most models of this and similar alloyed compound semiconductor systems, has considerable ramifications for phase stability and order in multicomponent systems at typical growth temperatures.

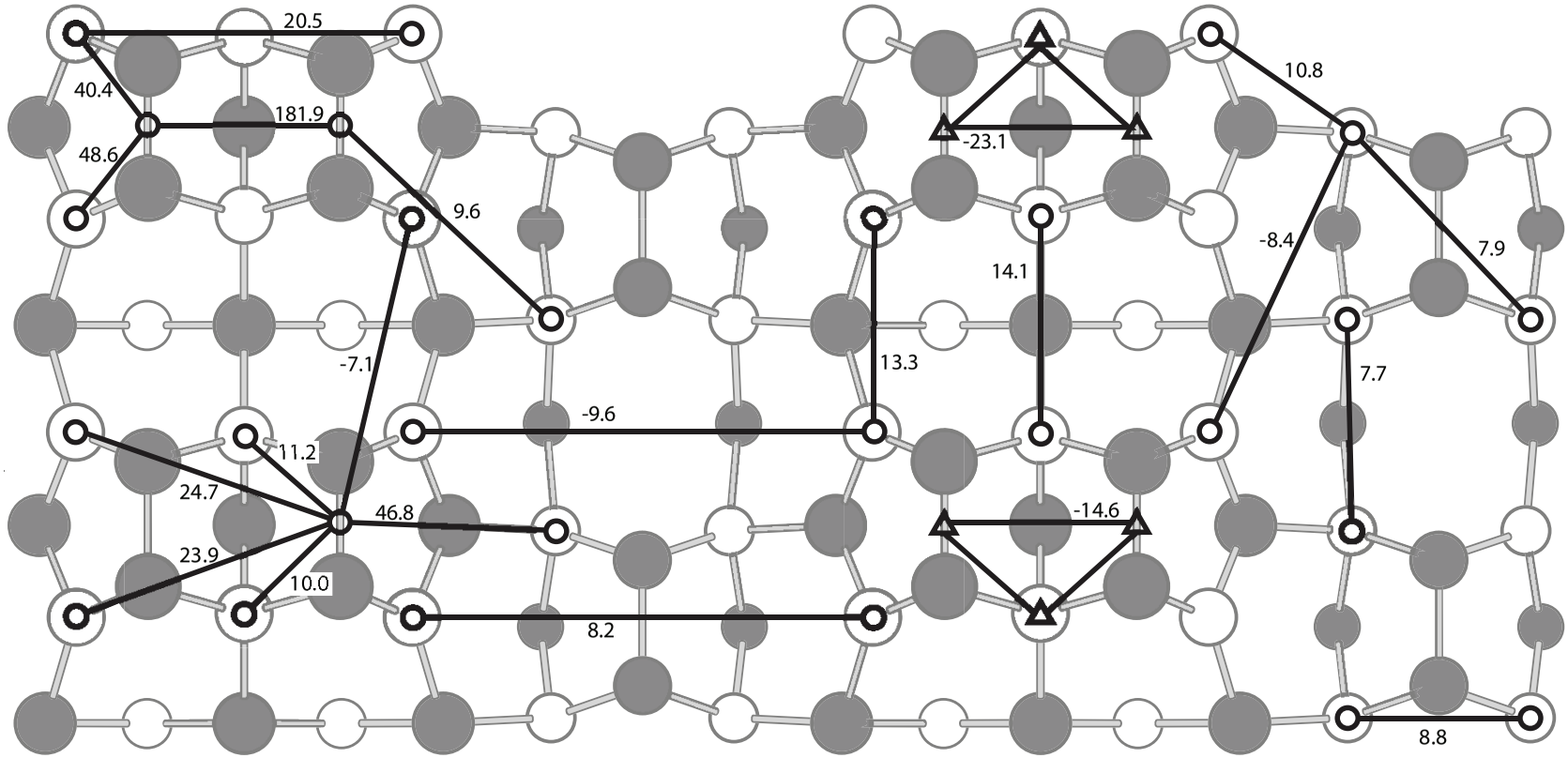


Figure 5.4. Diagram of the 22 strongest cluster interactions, with ECI indicated in meV. Lines joining open circles represent pair interactions; lines joining open triangles represent triplet interactions. Refer to Fig. 5.1(a) for symmetrically equivalent cation sites. All dimer sites are symmetrically equivalent.

We parameterized a cluster expansion of the form given in Eq. (2.59) using a training set of 378 configurational energies calculated using density functional theory (DFT) as implemented in the VASP code[38]. Fig. 5.3 illustrates the values of the resulting ECI for site occupancy, as well as pair, triplet, and quadruplet interactions. The low symmetry of the combined dimer and cation sublattices necessitate inclusion of 62 clusters in the expansion, including the empty cluster, 7 point clusters, 39 pair clusters, 14 triplet clusters, and one quadruplet cluster. The cross-validation score of the basis set is 2.02 meV per surface site and the final least-squares fit has a root-mean-square error of 1.55 meV per surface site. The only inter-dimer interaction retained in the optimized crystal basis is that of the nearest-neighbor dimer pair that exists within the primitive cell.

This pair interaction is the strongest one found for the system, and describes a large repulsive interaction between dimer nearest neighbors (i.e., a larger formation enthalpy for the $\beta 2(2\times4)$ than for the $\alpha 2(2\times4)$). Since the dimer row is the highest feature of the unit cell, row dimers are quite isolated from row dimers of neighboring unit cells, thereby minimizing chemical and elastic interactions between them. Hence longer range inter-dimer interactions are negligible. Of the remaining multi-body interactions, 31 involve only cation sites and 22 involve both cation and dimer sites. Fig. 5.4 shows the 22 strongest interactions in the system, as measured by the magnitude of their ECI. In interpreting values of ECI it is useful to note that a negative point cluster ECI indicates an energetic preference for the positive-"spin" species (i.e., In or dimer), and a negative pair ECI indicates an energetically preferred interaction between like species. The physical significances of ECI of triplet and larger clusters are somewhat harder to interpret, since there are eight possible configurations of a triplet cluster, and 2^n posibilities for n-body clusters. Of particular interest among the interactions shown in Fig. 5.4 are the dimer-dimer interaction, the interactions of the row dimer with the cation sites that bond directly to it, and the interaction of the row dimer with the nearest trench cation site; the effects of these interactions are especially apparent in the Monte Carlo results. Also, contrary to naïve expectations that the presence of the trench might minimize interactions between neighboring unit cells along [110], there are some significant interactions which act directly across the trench.

The training set used to fit the ECI consists of a range of configurations of variously shaped (2×4), (2×8), and (4×4) supercells, selected using a number of metrics intended to ensure that the configurational phase space was evenly sampled and that the ECI of candidate clusters were sufficiently overdetermined. Additionally, configurations that were found to be ground states of preliminary cluster expansions were added to the training set in an iterative process to benchmark and improve the accuracy of the cluster expansion and optimize its ability to predict ground-state configurations. For the (2×4) supercell, brute

force enumeration of configurations was used to identify ground-state structures. However, because the number of possible configurations grows exponentially with supercell size ($\propto 2^{12n}$, where n is the number of (2×4) primitive cells), simulated annealing must be used to identify ground-state configurations of larger supercells. This is accomplished by applying Monte Carlo simulations to a preliminary cluster expansion in an appropriately sized supercell at fixed chemical potentials and starting at a sufficiently high temperature excite fluctuations in site occupancy. The temperature is gradually reduced, until the supercell configuration no longer changes. The configuration with the lowest energy during the simulated annealing run is recorded, and after repeating the process for the relevant range of chemical potentials, a comparison of all low-energy configurations reveals the configurational ground states.

5.2.1 Simulation Details

Although zero-Kelvin phase diagrams can be constructed directly from first principles energies of a library of configurations, or by finding ground states of the cluster expansion, these phase diagrams are based on energies of relatively well-ordered reconstructions separated by abrupt boundaries in chemical potential space. Entropy arising from configurational disorder becomes important at elevated temperatures, causing phase boundaries to shift or disappear altogether relative to zero-Kelvin predictions. In systems with significant thermally-induced disorder, these effects complicate experimental characterization informed by these predictions. In order to incorporate temperature effects and use calculated ensemble averages to make physically useful predictions, we have used the cluster-expanded effective Hamiltonian within the Metropolis Monte Carlo algorithm[84]. Monte Carlo simulations are conducted within the grand canonical ensemble, where the thermodynamic boundary conditions independently fix the temperature and the chemical potentials of the two sublattices. As an adsorptive species, the As chemical potential, μ_{As}, is independently controlled. Conversely, since In and Ga are substitutional species, only the *relative* cation chemical potential, $\Delta\mu_{cat} = \mu_{In} - \mu_{Ga}$, has an effect on the equilibrium state of the system if the number of cation sites remain fixed.

To facilitate comparison with experiment, we use as reference states the rhombohedral $A7$ phase of bulk As, zincblende GaAs, and zincblende InAs. With this choice of reference states, $\mu_{As} > 0$ implies the instability of bulk GaAs relative to bulk As. For the cation sublattice, we can associate the range of $\Delta\mu_{cat}$ to two very different physical regimes. A very low (negative) value of $\Delta\mu_{cat}$ corresponds to a regime where the surface is in equilibrium with bulk GaAs below the In solubility limit. The surface can equilibrate at a significantly higher In concentration than exists in the bulk because at low $\Delta\mu_{cat}$ in an epitaxially grown alloy film, In will tend to segregate to the surface due to atomic size effects, as well as the

lower surface free energy of InAs. As $\Delta\mu_{cat}$ increases and approaches zero the surface must equilibrate with an increasingly non-dilute $In_xGa_{1-x}As$ alloy. However, since the total energy calculations used to parameterize the cluster expansion were performed on GaAs slabs, and In is allowed to occupy only surface cation sites, high values of $\Delta\mu_{cat}$ in our model system correspond more accurately to experimental conditons where a partial monolayer of InAs is directly deposited on GaAs. This is a metastable equilibrium where excess In at the surface is prevented from diffusing into the bulk due to sluggish kinetics.

Grand canonical Monte Carlo simulations were performed at several temperatures and on a dense grid of chemical potential values for μ_{As} and $\Delta\mu_{cat}$ that stabilize the full range of As and In compositions in our surface model. The Monte Carlo simulation cell consisted of 288 primitive cells (a square 24×12 lattice with periodic boundary conditions), for a total of 2880 cation sites and 576 dimer sites. At each point the system was equilibrated over 3000 MC passes (each site was visited 3000 times on average) and ensemble averages were tracked over the following 5000 MC passes. High-precision data were collected along chemical potential trajectories corresponding to pure $\alpha2(2\times4)$ and $\beta2(2\times4)$ surfaces. For the former, the trajectory along which dimer coverage is exactly 50% was interpolated from exploratory MC runs. For the latter, the same contour was used, but shifted to a higher As chemical potential such that only the $\beta2(2\times4)$ was sampled. Data along the trajectories were acquired using a 72×36 lattice over 11000 MC steps, after 5000 MC steps of equilibration.

In Ch. II we showed that various first derivatives of the characteristic potential could be obtained as ensemble averages of conserved quantities taken over the Boltzmann distribution. Various second derivatives of the characteristic potential with respect to its intensive parameters can also be calculated as fluctuations of the conjugate extensive quantities. The relation between fluctuations and second derivatives of the free energy is obtained by explicitly taking second derivatives of the characteristic potential, as expressed in terms of the partition function Z. For the characteristic otential Φ, which is a function with independent variables T, μ_{As}, $\Delta\mu_{cat}$, N_{cat}, and V, we can write

$$\begin{aligned}\Phi &= E - TS - \mu_{As}N_{As} - \Delta\mu_{cat}N_{In} \\ &= -k_BT\ln(Z),\end{aligned} \tag{5.1}$$

where, as before, k_B is the Boltzmann constant. For the InGaAs/GaAs (2×4) system, the partition function, defined by the sum over all configurational microstates $\vec{\sigma}$, takes the form

$$Z = \sum_s \exp\left[-\beta(E(\vec{\sigma}) - \mu_{As}N_{As}(\vec{\sigma}) - \Delta\mu_{cat}N_{In}(\vec{\sigma}))\right], \tag{5.2}$$

where β is $1/(k_B T)$.

Explicit differentiation of the statistical mechanical free energy yields an expression relating its second derivatives to the variances of extensive quantities. As an example, the chemical susceptibility, χ_{ij}, defined as the derivative of N_j with respect to the chemical potential μ_i, where i and j are either As or In, is equal to the second derivative of the free energy Φ and can be expressed in terms of fluctuations in the number of atoms of species i and j according to

$$\chi_{ij} = \frac{\partial^2 \Phi}{\partial \mu_i \partial \mu_j} = \frac{1}{(k_B T)} \left(\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle \right). \tag{5.3}$$

The various second derivatives of the free energy are useful in identifying phase boundaries. A discontinuity of the second derivative indicates a first order transition, while a diverging second derivative indicates a second order transition.

5.3 Results

5.3.1 Surface Stability and Temperature Dependence

The most readily interpreted results of the Monte Carlo simulations is the relative stability of the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ reconstructions. Figure 5.5 shows a surface stability map of the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ reconstructions of InGaAs/GaAs (001) within the physically meaningful chemical potential ranges. The quantity used to determine the boundary in Fig. 5.5 is the As dimer chemical susceptibility, $\chi_{As,As}$, as defined in Eq. (5.3). We determine from the various susceptibilities and ensemble average compositions that there is no formal first- or second-order phase transition between the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$, since since all the ensemble averages are continuous. Instead, there is a smooth transition, indicated by a non-singular maximum in $\chi_{As,As}$, between $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ along lines of constant $\Delta\mu_{cat}$. As expected, the $\beta 2(2\times4)$ is stable at higher μ_{As}, and the $\alpha 2(2\times4)$ is stable at lower μ_{As}, since the $\beta 2(2\times4)$ is more As-rich. The occluded region of Fig. 5.5 corresponds to the chemical potential ranges where the $\alpha 2(2\times4)$ becomes unstable relative to the unreconstructed surface, which is the limiting case of the configurational lattice model. For this unreconstructed surface the electron counting rule is violated, indicating a likely unphysical regime. Other reconstructions and/or bulk phases not explicitly considered in this work will form in this chemical potential range.

An intriguing trend highlighting the sublattice coupling occurs at fixed μ_{As} with cation chemical potential, $\Delta\mu_{cat}$, allowed to vary freely. Raising $\Delta\mu_{cat}$ results in increased In concentration over the surface sites. A high surface In concentration in turn destabilizes the

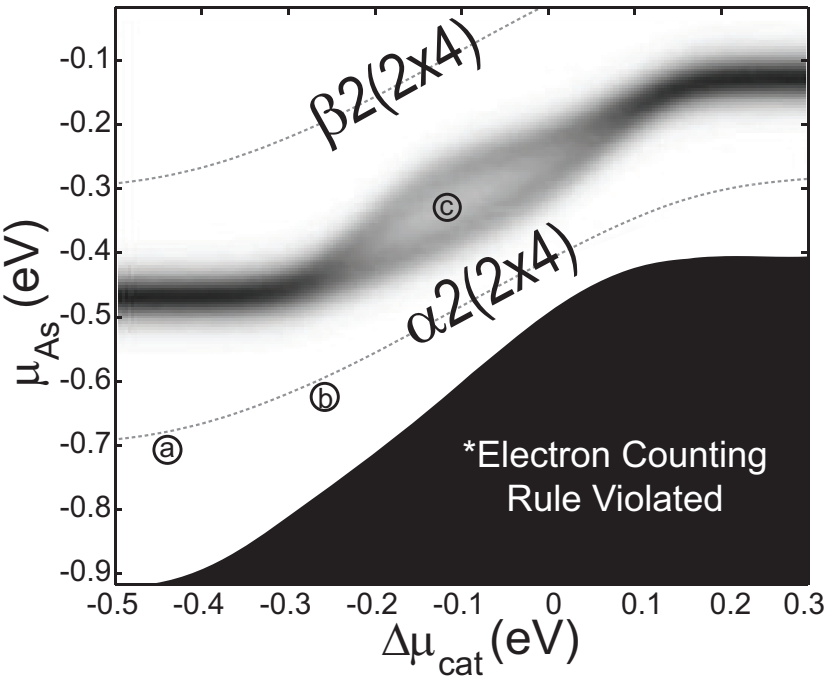


Figure 5.5. Surface stability map of the InGaAs/GaAs (001) surface showing the chemical potential domains of stability for the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ at 100°C. A small region, labeled (c), exists between the two larger ones where the $h_{\alpha\beta}$, a hybrid of the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$, is stable. Boundaries are calculated from the dimer chemical susceptibility, with width corresponding to the chemical potential interval over which the smooth transition between reconstructions occurs. The obscured region is an unphysical limit of the configurational model and corresponds to relative stability of the bulk-terminated surface. Points a, b, and c correspond to microstate snapshots shown in Fig. 5.10.

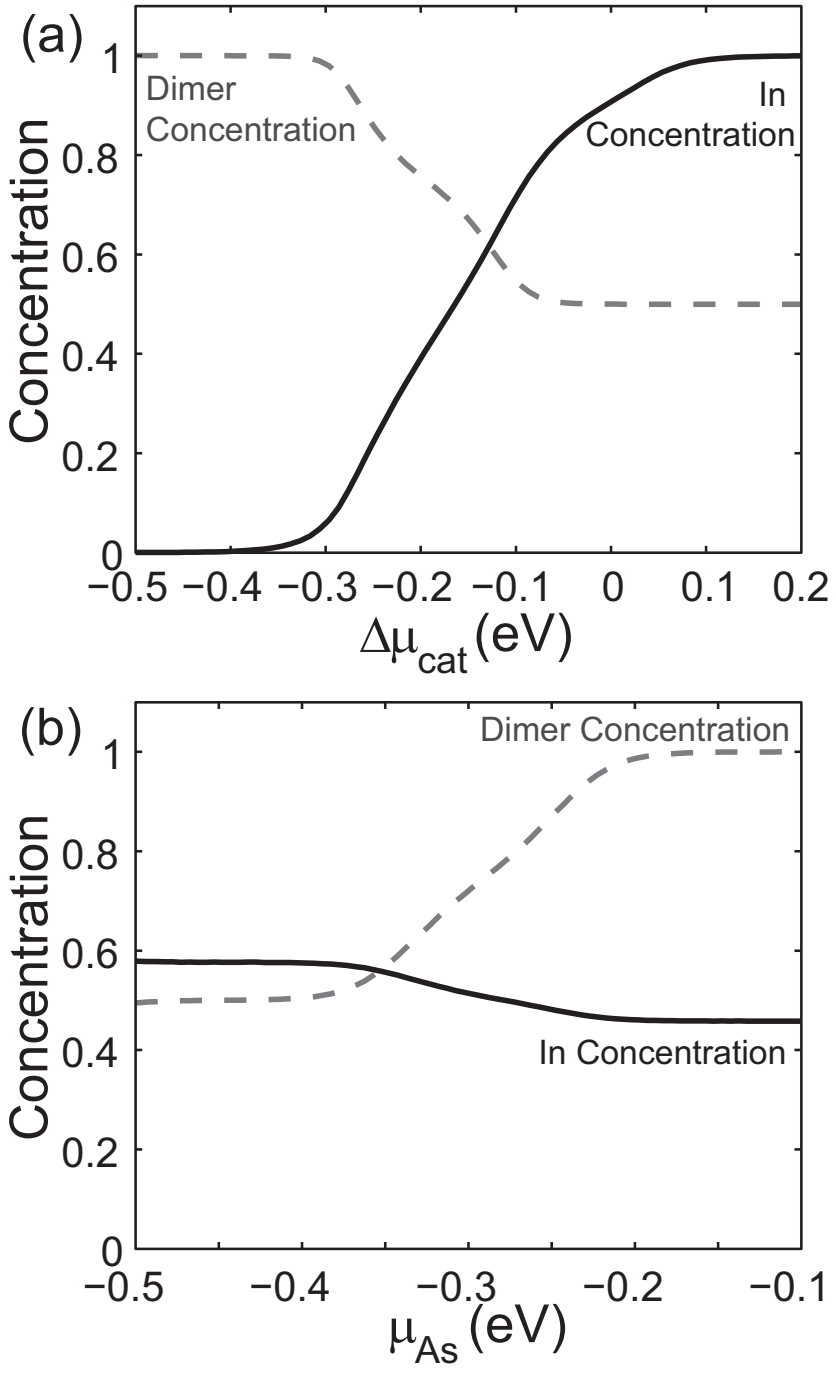


Figure 5.6. The dimer and In concentrations at 100°C (a) along a line of constant cation chemical potential with $\Delta\mu_{In} - \Delta\mu_{In}^{bulk}$ fixed at -6 meV and (b) along a line of constant As chemical potential with $\mu_{As} - \Delta\mu_{As}^{bulk}$ fixed at -313 meV. Chemical potentials are measured on the same axes as Fig. 5.5.

$\beta 2(2\times4)$ via atomic-size mismatch strain interactions in favor of an $\alpha 2(2\times4)$ surface. This phenomenon is seen most clearly in Fig. 5.6(a), which shows the increase in In concentration and resultant decrease in dimer coverage as $\Delta\mu_{cat}$ is increased and μ_{As} is held constant. Fig. 5.6(b) shows a related but weaker effect, where there is a slight decrease in In concentration during the transition from the $\alpha 2(2\times4)$ (50% dimer coverage) to the $\beta 2(2\times4)$ (100% dimer coverage) at fixed $\Delta\mu_{cat}$. The reconstruction destabilization is due to the coupling between the cation and As sublattices as accounted for by the cluster expansion, and is attributed to the atomic size mismatch strain as illustrated in Fig. 5.2. As was described by the first guideline for size mismatch interaction detailed in Sec. 5.1.2, the substitution of In in the $\beta 2(2\times4)$ results in a nontrivial displacement of the row dimers, straining the bonds between row dimers and cations. By contrast, substitution of In in the $\alpha 2(2\times4)$ results in comparatively little distortion of the lattice relative to the pure GaAs $\beta 2(2\times4)$ lattice positions, reducing energetically unfavorable strain fields.

In addition to the regions of $\alpha 2(2\times4)$ and $\beta 2(2\times4)$, Fig. 5.5 also shows a region within the $\alpha 2(2\times4)$-$\beta 2(2\times4)$ transition where $\chi_{As,As}$ is not maximal. This corresponds to a hybrid reconstruction comprised of both the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ unit cells, which we refer to as $h_{\alpha\beta}$. Although the $h_{\alpha\beta}$ bears passing resemblance to a nano-domain phase coexistence, no such coexistence is possible in this system since it lacks first-order phase boundaries. Instead, the $h_{\alpha\beta}$ is a category of actual ordered configurations of the surface that have corresponding zero-temperature configurational ground states. There is in fact a series of zero-temperature hybrid reconstruction ground states near 75% dimer coverage, over a range of In compositions. As an example, the particular hybrid reconstruction ground state having 75% dimer coverage and 37.5% In composition has an enthalpy of formation, as predicted by the cluster expansion, that is 70 meV/unit cell lower than the stoichiometircally comparable two-phase mixture of pure $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ configurations. This comparison neglects the influence of boundary energies, coherency effects, and substrate-mediated elastic interactions, all of which can act to increase the energy of a system in true coexistence. The preference of intermediate-range order over two-phase coexistence in this system implies that the mismatch strains due to In substitution can act to enhance mixing and structure refinement at the surface in regimes where mixing is otherwise suppressed, such as near phase boundaries.

At finite temperature the hybrid surface is stable over a range of In and dimer composition, extending from approximately 30-55% surface In concentration and 70-80% dimer coverage at 100°C; however, it is most stable at 45% surface In and 75% dimer coverage, where there χ_{As_2} has a local minimum. The predicted ground-state configurations of the hybrid reconstruction are formed by alternating $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ unit cells along both [110] and $[1\bar{1}0]$. Two variants are predicted, depending on the relative orientations of the

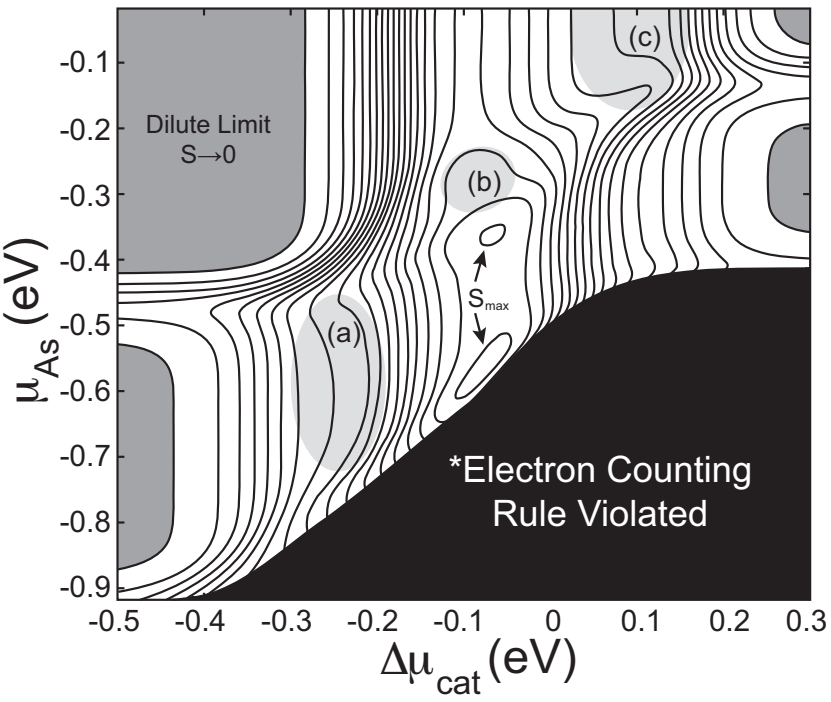


Figure 5.7. Adiabatic contours, obtained from the entropic contribution to free energy. Regions of low contour density indicate either significant order or disorder. Strong ordering occurs as a limiting case at chemical potential extremes (shaded dark grey), but several other nontrivial cases occur (light grey): (a) the "zig-zag" ordering of the $\alpha 2(2\times4)$, (b) the $h_{\alpha\beta}$ hybrid reconstruction, and (c) local ordering of the cation sublattice in the $\beta 2(2\times4)$. As in Fig. 5.5, the black region indicates an unphysical limiting case of the configurational model.

$\alpha 2(2\times4)$ row dimers; one has a (4×8) unit cell and the other has a centered (4×4) primitive cell. However, at typical growth temperatures ($\tilde{5}00$°C), the $h_{\alpha\beta}$ surface shows significant disorder, although the $h_{\alpha\beta}$ unit cell is still relatively well-defined at the length-scale of several (2×4) unit cells at finite temperature, suggesting that the hybrid surface can readily tolerate thermally excited defects to its ground-state configurations.

Using the ensemble averages obtained from grand canonical Monte Carlo simulations, we used Eq. (2.50) from Ch. II to integrate the free energy Φ and determine the entropy S. Because the configurational entropy vanishes at the extremes of composition (e.g., at $x_{As} = x_{In} = 1$), the integration constant Φ_0 is chosen at one of these extremes, since the the grand canonical energy becomes simply $\Phi_0 = E - \mu_{As}\langle N_{As}\rangle - \Delta\mu_{cat}\langle N_{In}\rangle$ at the extremes. The resulting value of TS, is the entropic contribution to the free energy arising from configurational excitations.

Fig. 5.7 illustrates the configurational entropy of the surface as a function of μ_{As} and $\Delta\mu_{cat}$. The contours follow lines of constant entropy and thus form adiabatic chemical potential contours. The entropy is highest near 50% composition on both sublattices, due

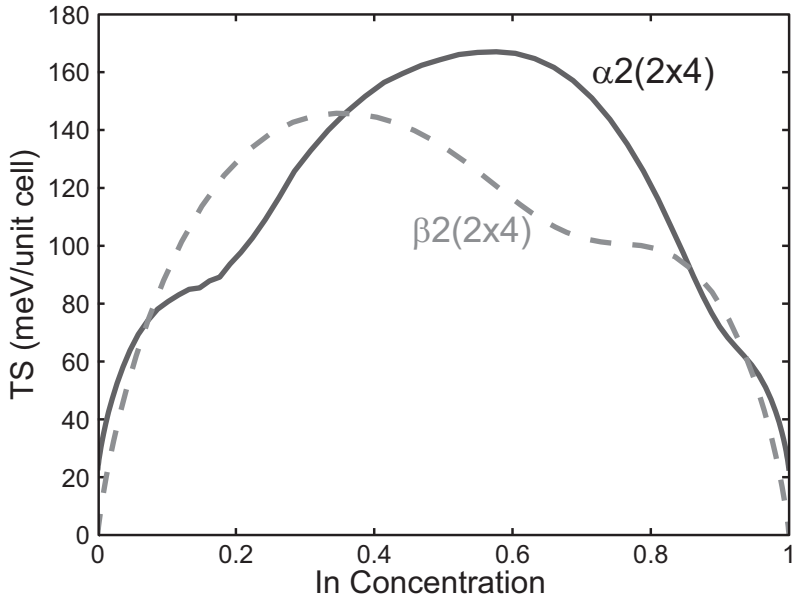


Figure 5.8. Entropic contribution to free energy, as a function of In concentration, for the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ reconstructions at 100°C. The contribution from the $\alpha 2(2\times4)$ is higher than that of the $\beta 2(2\times4)$ over most of the range of concentration. Intervals where the entropic contribution is suppressed correspond to short-range order.

to purely combinatoric reasons. The two entropy maxima shown in Fig. 5.7 are slightly off-stoichiometry since at low temperatures the system is constrained to the $\alpha 2(2\times4)$ at 50% dimer coverage. By going slightly above and below 50% dimer concentration, and thus approaching the transitions away from the $\alpha 2(2\times4)$, fluctuations in dimer concentration increase, reflected by an increase in entropy.

Generally, regions of significant order or disorder are expected where the density of adiabatic contours is low, as the entropy is relatively flat with respect to changes in μ_{As} and $\Delta\mu_{cat}$ in these areas. Fig. 5.8 shows the entropy calculated along the chemical potential trajectories, shown as dotted lines in Fig. 5.5, corresponding to pure $\alpha 2(2\times4)$ and pure $\beta 2(2\times4)$. It is evident from the entropy along these trajectories, where either the $\alpha 2$ or $\beta 2$ structural variant is sampled independently, that certain surface cation compositions correspond to a local change in the convexity of the entropy. The most prominent of these composition intervals correpond chemical potential regions of visibly decreased adiabat density in Fig. 5.7 (shaded light grey), further suggesting the occurrence of increased long- or short-range order. These regions deserve special attention and are discussed in detail in Sec. 5.3.3. It is also important to note that over most of the range of cation composition, the $\alpha 2(2\times4)$ has higher entropy than the $\beta 2(2\times4)$, due to the greater number of possible dimer arrangements on the $\alpha 2(2\times4)$ surface. The directional nature of the $\alpha 2(2\times4)$, caused by its dimer/dimer-vacancy pair, allows many distinct arrangements of the dimer sublattice in the $\alpha 2(2\times4)$, whereas

the $\beta2(2\times4)$ allows only a single dimer configuration. This difference in entropy enhances the stability of the $\alpha2(2\times4)$ near the smooth transition between the two reconstructions as temperature increases. Because there is no way to measure an exact boundary of stability between the two reconstructions, it is difficult to measure this trend quantitatively.

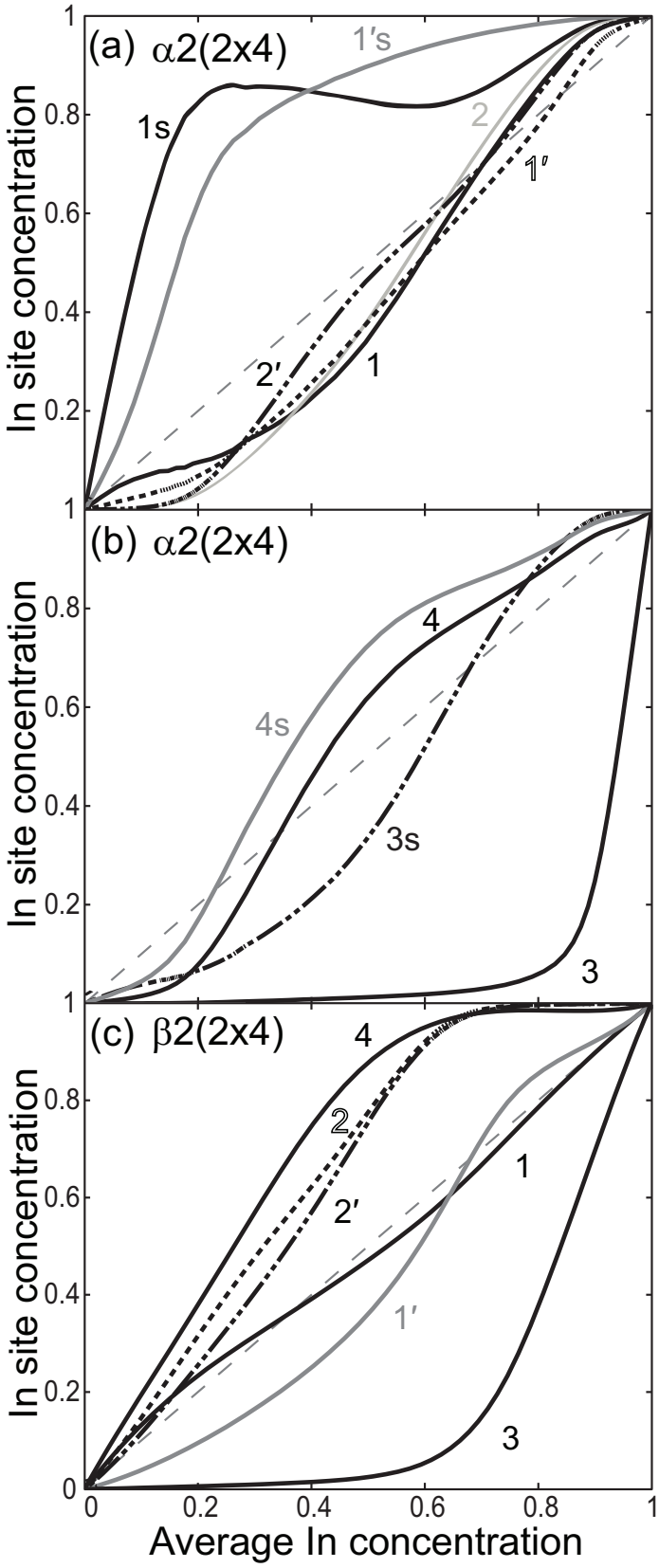
5.3.2 Preferential Cation Site Filling

To better understand the connection between composition and short-range order, we can use calculate ensemble averages of the individual site occupancies on the cation sublattice during Monte Carlo simulations. The $\beta2(2\times4)$ possesses mirror plane that bisects the unit cell across the [110] direction, allowing us to identify six symmetrically distinct classes of cation substitution sites, labeled in Fig. 5.1(a). The position of the trench dimer, which is shifted by one quarter of a unit cell relative to the row dimers, breaks the $[1\bar{1}0]$ mirror symmetry that would otherwise exist. Sites whose degeneracy is broken only by the trench dimer tend to be energetically similar despite being symmetrically distinct and are distinguished by a prime symbol, according to their position relative to the trench dimer. The removal of a dimer from the $\beta2(2\times4)$ to form the $\alpha2(2\times4)$ breaks the [110] mirror symmetry, leading to ten fully distinct cation sites. In the $\alpha2(2\times4)$ we therefore distinguish between those sites that are adjacent to a row dimer and those that are opposite a row dimer, the latter of which we denote by a 's', as shown in Fig. 5.1(b).

Figure 5.9 shows a comparison of the In concentration of the various symmetrically distinct cation sites for both the $\alpha2(2\times4)$ (Figs. 5.9(a) and 5.9(b)) and $\beta2(2\times4)$ (Fig. 5.9(c)) along the dashed lines in Fig. 5.5, corresponding to chemical potential trajectories single reconstruction variant coverage. The sites are labeled as shown in Fig. 5.1. The trends observed for the various sites enable significant intuition about the role of atomic size mismatch and sublattice coupling. We can categorize the symmetrically distinct cation substitution sites into four types, each of which exhibits qualitatively distinct occupancy trends.

Type 1 Sites. The four surface cation sites that sit on the outside of the dimer row comprise the type 1 sites, as indicated in Fig. 5.1. Due to the mismatch strain effects illustrated in Fig. 5.2, and detailed in the first guideline for strain interaction in Sec. 5.1.2, sites $1s$ and $1's$ accommodate In more readily than other sites in the $\alpha2(2\times4)$, relieving the strain inherent in the cation-cation backbonds of the reconstruction and allowing the single row dimer and surrounding sites to sit closer to the positions occupied in the $\beta2(2\times4)$. This is reflected in the average site occupancy for the $\alpha2(2\times4)$, shown in Fig. 5.9(a). Here, the $1s$ and $1's$ sites (which are the 1 and $1'$ sites sitting opposite the lone row dimer) are nearly fully occupied by In at low average In composition. At low average composition the 1 and $1'$ site occupancy is slightly below average, due to the very large contribution to the average

Figure 5.9. In occupancy of various cation sites at 100°C. Site labels correspond to those in Fig. 5.1. (a) Surface sites and (b) trench sites of the $\alpha 2(2\times4)$, and (c) all sites of the $\beta 2(2\times4)$. In each figure, the dashed line along the 45° is the average In concentration.



from the $1s$ and $1's$ sites. At intermediate and high average composition the 1 and $1'$ site occupancy closely follows the average.

In the $\beta 2(2\times4)$ (Fig. 5.9(c)), there is less preference for type 1 sites to accept In than in the $\alpha 2(2\times4)$ (Fig. 5.9(a)). Instead, the site filling remains close to or below the average composition over a wide range of In composition. For average In compositions below approximately 65%, 1 sites are slightly preferred over $1'$ sites. This is likely due to the relative proximity of the $1'$ to the trench dimer. The increase in $1'$ site filling is concurrent with increased filling of site 3, and the strong influence of type 3 sites on the local relaxation of the trench dimer implies a potential causal relation between increased site 3 filling and the change in In filling preference from the type 1 site to the type $1'$ site.

Type 2 Sites. The type 2 sites are comprised of the two cation sites that sit between the row dimer sites. Because the type 2 sites are bisected by a mirror plane, the many short-range pair interactions affecting them have mirror symmetry. This leads to a strong coupling with neighboring cation sites, as well as with the row dimers.

In the $\alpha 2(2\times4)$ (Fig. 5.9(a)), In-filling of type 2 sites is suppressed at low In concentration as the highly preferred 1 and $1'$ sites fill. At intermediate and high In concentration, the filling trends of 2 and $2'$ sites closely follow the average In concentration. In the $\beta 2(2\times4)$ (Fig. 5.9(c)), the type 2 sites exhibit preferential In filling over the entire range of In concentration. This is due mainly to the relatively large separation between neighboring type 2 sites, as well as the large separation between type 2 sites and the trench dimer. This association linking mismatch species localization at high-symmetry, spatially dispersed sites is summarized in the second and fourth strain interaction guideline of Sec. 5.1.2. There is minimal anisotropy between the 2 and $2'$ sites, and any that does exist is likely due to interaction with the trench dimer mediated by the nonuniform occupancy of intervening sites.

Type 3 Sites. The type 3 sites lie in the trench, and are directly in-line with and adjacent to the row dimers along $[1\bar{1}0]$. Although type 3 sites are three atomic monolayers below the row dimer sites, the position of the type 3 sites relative to the row dimer sites results in a strong coupling between row dimer occupation and type 3 substitution energy. The row dimer compressively strains the bonds above and surrounding the adjacent type 3 site, significantly increasing the energy of substituting a large In atom at the site. This is most clearly demonstrated for the $\alpha 2(2\times4)$, where Fig. 5.9(b) shows a strong anisotropy between the $3s$ and 3 sites. The 3 site exhibits almost no filling until it is forced to at high $\Delta\mu_{cat}$ due to the lack of other unsaturated substitution sites. By comparison, the $3s$ site is quite low in energy, as evidenced by a site occupation approaching and exceeding the average at moderate to high In composition. In the $\beta 2(2\times4)$, where both sites adjoin a row dimer,

the type 3 sites exhibit almost no substitution until all other sites are more than half-filled with In.

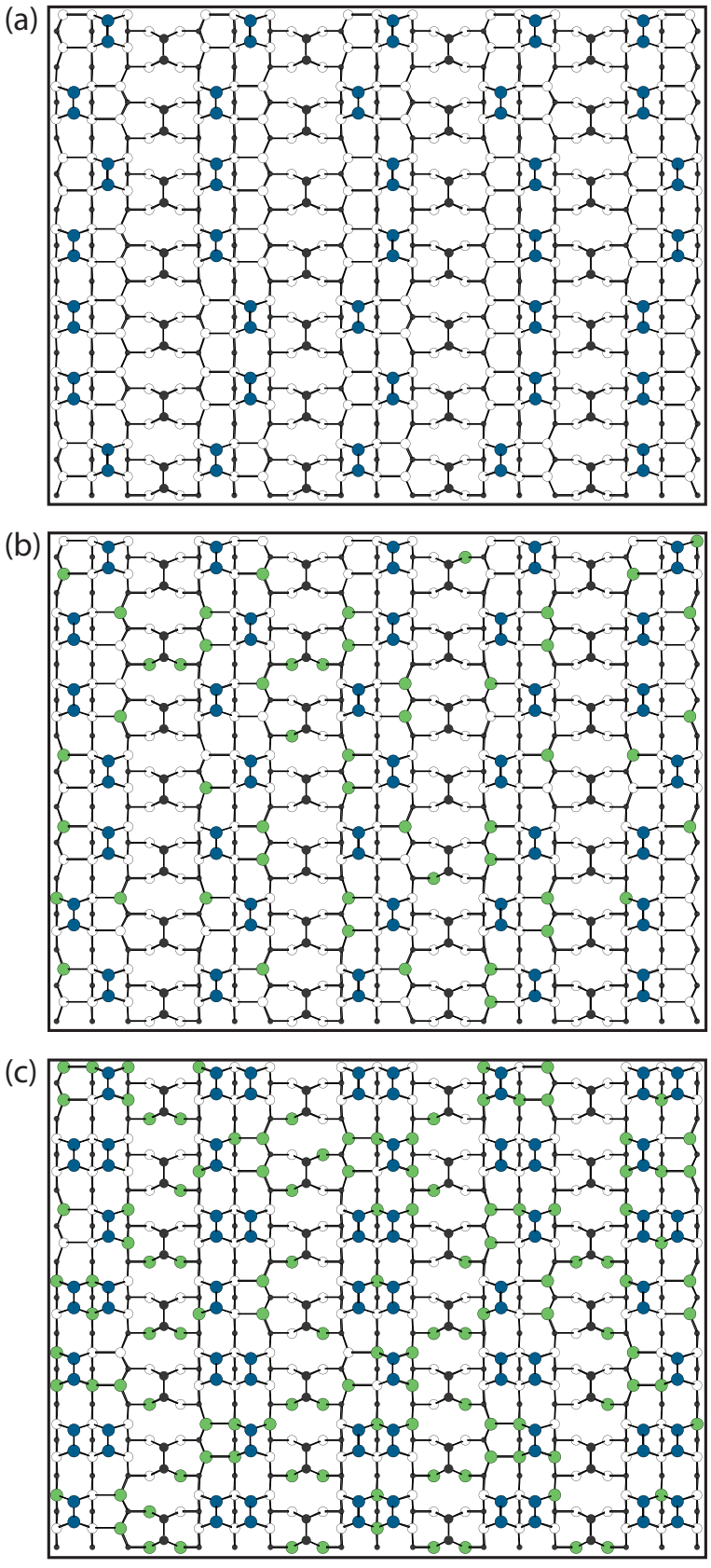
Type 4 Sites. Although type 4 sites also lie within the trench, they are relatively low in energy and exhibit higher than average In occupation at most In concentrations in both the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$. In the $\alpha 2(2\times4)$, the type 4 sites readily accept In, but filling is delayed at lower concentrations, as filling of the low-energy $1s$ and $1's$ sites takes precedence. There is only minimal interaction with the row dimer, as seen by the very similar filling profiles for $4s$ and 4 sites. The relatively weak coupling of type 4 sites to the row dimer sublattice is also very noticeable in the $\beta 2(2\times4)$, where the type 4 sites fill with even higher precedence, reaching full occupancy at 60% average In concentration. We can attribute this large difference to the much higher substitutional energy of site types 1–3 in the $\beta 2(2\times4)$ in general, compared to the $\alpha 2(2\times4)$, while the substitution energies of the type 4 sites remain largely unchanged.

5.3.3 Alloy-Driven Ordering

Having demonstrated the strong differences in average cation site occupancy in Sec. 5.3.2, we surmise that in regions of chemical-potential space where these differences are strongest we are likely to observe either long- or short-range ordering on at least the cation sublattice, and likely the row dimer sublattice as well. These regions also coincide with changes in the convexity of the entropy illustrated in Figs. 5.7 and 5.8. For the $\alpha 2(2\times4)$ there are two regions where increased short-range order is expected, as determined by these two indicators. One is near 20% In concentration and the other is near 90%; both are easily visible in Fig. 5.8. An even stronger suppression of the entropy is observed in the $\beta 2(2\times4)$ between 60% and 65% In concentration. In addition to these ordered regions in the pure $\alpha 2(2\times4)$ or $\beta 2(2\times4)$, there is the thermodynamically stable $h_{\alpha\beta}$ hybrid reconstruction. As stated previously, the $h_{\alpha\beta}$ occurs over a range of In and dimer concentrations and, like the ordered regions of the pure reconstructions, is associated with a change in the convexity of the entropy. By examining short-range order parameters and instantaneous microstate snapshots of the MC simulation cell, we are able to better understand the significance of these regions.

We obtain a quantitative measure of the short-range order of the dimer sublattice by constructing an order parameter that quantifies the tendency of row dimers in the $\alpha 2(2\times4)$ to arrange themselves in straight lines along the $[1\bar{1}0]$ axis, or to form a "zig-zag" arrangement along $[1\bar{1}0]$. Zig-zag ordering has been studied previously in this system using DFT energy calculations[75] and, in the absence of defects, is described by the (4×4) unit cell shown in Fig. 5.11 (inset). The dimer sites of the Fig. 5.11 inset are numbered from 1 to 4, and the occupation variable of a dimer site, as described generally in Sec. 2.3, is $\sigma_i = +1$ for a dimer

Figure 5.10. Microstate snapshots for the various points of chemical potential space specified in Fig. 5.5. To emphasize order, In are green, Ga are white, and row dimers are blue; all other As are black. (a) The $\alpha 2(2\times4)$ with no In, (b) the zig-zag ordering of $\alpha 2(2\times4)$ with 15% In, and (c) the $h_{\alpha\beta}$ hybrid reconstruction at 75% dimer coverage and 37% In.



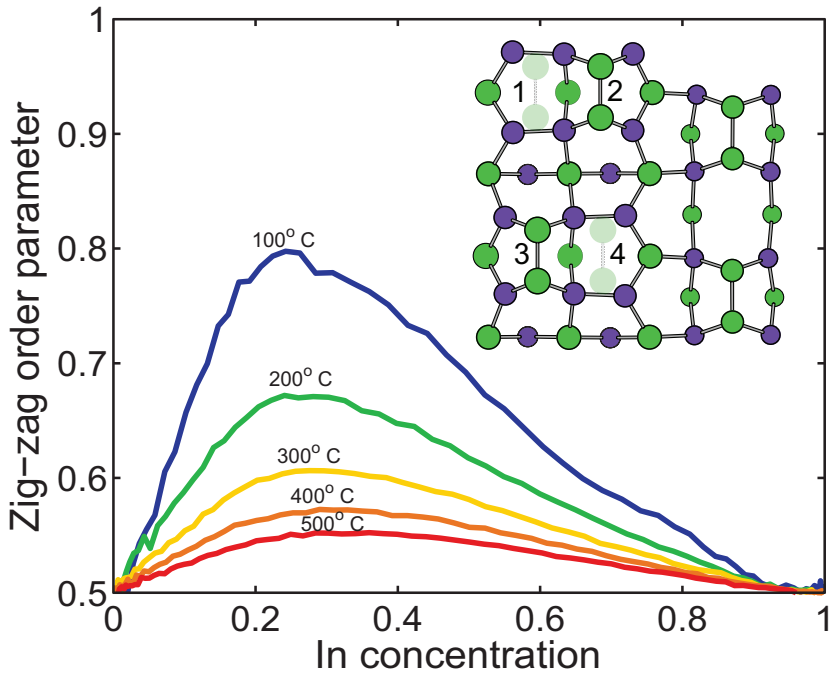


Figure 5.11. The short-range dimer order parameter for zig-zag ordering of the $\alpha 2(2\times4)$ along [110]. It is shown with respect to surface In concentration at temperatures ranging from 100 − 500°C along the chemical potential trajectory corresponding to pure $\alpha 2(2\times4)$. The fully ordered configuration is shown as an inset, with row dimer sites numbered and vacant sites indicated by grey dimers.

occupying site i and $\sigma_i = -1$ for a vacancy. We write an expression

$$p_z = (1+\sigma_1)(1-\sigma_2)(1-\sigma_3)(1+\sigma_4)/16, \tag{5.4}$$

such that perfect zig-zag ordering (5.11 inset) yields a value of $p_z = 1$, while $p_z = 0$ otherwise. The ensemble average of p_z then gives the probability that a (4×4) cell has the dimer configuration shown in Fig. 5.11 for a given set of thermodynamic parameters. Expanding Eq. (5.4) and taking the ensemble average, we find

$$\langle p_z \rangle = \left(1 - 2\langle\sigma_1\sigma_2\rangle - 2\langle\sigma_1\sigma_3\rangle + 2\langle\sigma_1\sigma_4\rangle + \langle\sigma_1\sigma_2\sigma_3\sigma_4\rangle\right)/16, \tag{5.5}$$

where terms corresponding to symmetrically equivalent clusters have been combined or cancel each other. Noting that the configuration of Fig. 5.11 (inset) is symmetrically equivalent to its mirror across (110) (i.e., the pattern of black dimers versus the pattern of greyed dimers), we multiply Eq. (5.5) by a factor of two to obtain the average probability that a (4×4) unit cell has a zig-zag configuration.

An expression similar to Eq. (5.5) can also be obtained for the (4×4) in-line dimer configuration (i.e., dimer sites 1 and 3 are occupied, with 2 and 4 vacant, or visa versa). Denoting this quantity $\langle p_r \rangle$ we renormalize p_z via

$$\widetilde{p}_z = \frac{\langle p_z \rangle}{\langle p_z \rangle + \langle p_r \rangle}, \tag{5.6}$$

giving the order parameter of interest, $\widetilde{p}_z$. By renormalizing with respect to only the dimer configurations consistent with a 50% dimer concentration in each (2×4) unit cell, we prevent any contribution to the order parameter arising from local fluctuations in dimer concentration, which occasionally create instances of the $\beta 2(2\times4)$, especially at high temperature. The quantity $\widetilde{p}_z$ then takes a value of one for perfect zig-zag order, zero for perfect in-line order, and 0.5 for a fully randomized configuration.

The quantity $\widetilde{p}_z$ is shown in Fig. 5.11 for a number of temperatures, measured along the chemical potential trajectory shown in Fig. 5.5 that passes only through the $\alpha 2(2\times4)$ phase field. We observe a strong preference for $[1\bar{1}0]$ zig-zag order, peaking near 25% In composition, with increasing disorder at higher In concentrations. The peak in the order parameter decreases in magnitude with increasing temperature due to thermally-induced disorder, and the order parameter is almost flat at 500°C, which is near typical growth temperatures for this system. We can therefore use the order parameter to predict the degree of order that may be expected experimentally during layer-by-layer growth, after careful annealing, or subsequent to rapid quenching. Increased temperature also shifts the

order parameter peak to slightly higher In concentration (approaching 30% at 500°C). The maximum is induced primarily by the preferential filling of $1s$ and $1's$ sites with In at low average In concentration. On average, a single type 4 site per (4×4) cell also fills with In so that the low-temperature maximum occurs at 25% In concentration.

In conjunction with the propensity of In to occupy sites opposite the $\alpha2(2\times4)$ row dimer, atomic size mismatch strain is further reduced by maximizing the distance between In atoms. In the [110] direction, this has little effect due to the separation of neighboring unit cells by the trench, but along $[1\bar{1}0]$ In atoms alternate their relative positions within the unit cells, forming a zig-zag pattern. Concurrently, the single $\alpha2(2\times4)$ row dimer tends to maximize its distance from In-occupied sites. The net effect is two opposing "zig-zags" along $[1\bar{1}0]$, one consisting of In pairs and the other consisting of row dimers. Fig. 5.10(b) shows a microstate snapshot from a MC simulation illustrating zig-zag ordering of $\alpha2(2\times4)$ at 100°C and 15% In composition. For comparison, Fig. 5.10(a) shows the pure GaAs surface, which assumes a random dimer arrangement even at low temperature.

The other region of greatest interest corresponds to the $h_{\alpha\beta}$ hybrid reconstruction, which occurs in the small region of low dimer susceptibility, $\chi_{As,As}$, along the $\alpha2(2\times4)$-$\beta2(2\times4)$ transition in Fig. 5.5. It is also highlighted as a prominent region of low adiabat density (labeled in Fig. 5.7). A microstate snapshot at a point within this region is shown in Fig. 5.10(c). Predicted and calculated ground-state configurations in this region have a strong tendency to alternate $\alpha2(2\times4)$ and $\beta2(2\times4)$ along both [110] and $[1\bar{1}0]$, and this unit cell ordering is visible in the microstate snapshot. Due to the much higher In substitution energy in cation sites of $\beta2(2\times4)$ relative to those of $\alpha2(2\times4)$, a *lateral modulation* of surface In composition occurs between the In-rich $\alpha2(2\times4)$ and In-poor $\beta2(2\times4)$ when $h_{\alpha\beta}$ is stabilized. At lower In chemical potentials within the $h_{\alpha\beta}$ chemical potential region, a large majority of In atoms migrate to unit cells of $\alpha2(2\times4)$, with 56% of $\alpha2(2\times4)$ cation sites filled with In in contrast to 22% of $\beta2(2\times4)$ cation sites at the chemical potential coordinate corresponding to Fig. 5.10(c). The 3, 4, and $4s$ sites do not play a significant role in this process due to the very low In occupancy of 3 sites and very high In occupancy of 4 and $4s$ sites, regardless of cell type. This composition modulation can be understood by considering the effects of alloying on reconstruction stability demonstrated in Sec. 5.3.1. As shown there, the $\alpha2(2\times4)$ is much more stable under the addition of In than the $\beta2(2\times4)$. This, along with the tendency of In atoms to mutually repel across unit cell boundaries, helps to explain the $h_{\alpha\beta}$ as it appears in Fig. 5.10(c).

In addition to the regions of zig-zag ordering and $h_{\alpha\beta}$, there are two additional features arising from short-range ordering in Fig. 5.8 that merit discussion. One occurs at 90% In in the $\alpha2(2\times4)$, and the other occurs at 60-65% in the $\beta2(2\times4)$. We note though that

these features are not necessarily of particular interest physically as they occur for chemical potentials where other reconstructions not considered here are likely to be stable relative to the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$. In each instance, In fills the available sites of the cation sublattice with the exception of those characterized by a high substitution energy. This results in many fewer arrangements of the cation sublattice than the combinatoric ideal, and thus a change in convexity of the entropy. In the $\alpha 2(2\times4)$ this occurs when all cation sites of the unit cell are filled except the very high-energy 3 sites. In the $\beta 2(2\times4)$ it occurs when all but the 3 sites and one each of the 1 and $1'$ sites are unfilled in the unit cell. These plateaus in filling are seen to some degree in Figs. 5.9(b) and 5.9(c), respectively. There is some ordering of the 1 and $1'$ sites in the $\beta 2(2\times4)$, which show a weak correlation along [210], but in large part, both orderings are caused by differences in site substitution energies rather than interaction between cation sites.

5.4 Summary

We have performed a first-principles study of surface reconstruction stability and order on the covalently bonded InGaAs/GaAs (001) surface at finite temperature, using a cluster expansion of the configurational energy together with Monte Carlo simulations. We have restricted our study to an examination of temperature and alloying on the relative stability between $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ as well as the atomic size mismatch strains arising from In substitution at GaAs surfaces and how these strains effect As dimer ordering. Because of the similarities observed among many III-V compounds, and the tendency in general of directionally-bonded surfaces to develop reconstructions, we can propose several phenomena that may deserve consideration in the engineering of nanostructured surfaces in these varied systems.

Size mismatch-driven surface destabilization. As seen in Fig. 5.5, substituting In atoms for the smaller Ga atoms destabilizes $\beta 2(2\times4)$, resulting in the stability of $\alpha 2(2\times4)$. It is expected that reconstruction stability exhibits a compositional dependence, but in the case of size-mismatched alloys this effect is significantly more pronounced. This has been clearly observed experimentally as x is increased in $In_xGa_{1-x}As$ (001) alloys, causing a destabilization of both the $\beta 2(2\times4)$ and the $c(4\times4)$[85], another III-V surface reconstruction with prominent anion dimers. The resulting surface reconstruction, which is not observed on either pure InAs or GaAs, has a $(n\times3)$ surface repeat unit and has been observed to be robust over a wide range of growth conditions[18, 86]. The precise atomic structure of this reconstruction, however, remains unclear[86]. The results presented here suggest that likely models for the reconstruction incorporate structural elements which would readily accept In,

such as the cation back-bonds found in the $\alpha 2(2\times4)$.

Strain-induced order. We have shown the strong tendency to order that arises from alloying and size mismatch, as evidenced by preferential cation site filling, changes in convexity of the configurational entropy, ensemble averaged site correlations, and Monte Carlo microstate snapshots. Surface order has been suggested previously as a mechanism for long-range bulk order in epitaxially grown ternary III-V films[10]. Although unstrained InAs and GaAs are completely immiscible over most temperatures of interest, and only chalcopyrite-type ordering is predicted to be stable for epitaxially strained ternary III-V alloys, both CuPt- and CuAu-type alloy ordering have been observed on the FCC bulk cation sublattice[87, 88]. The proposed mechanism suggests that equilibrium or near-equilibrium surface order becomes kinetically trapped during layer-by-layer growth, yielding a reconstruction-dependent bulk order. Previous studies have focused on simple dimerized surfaces to examine the effect of dimers on surface order[8] or have studied pair correlations between substitutional sites in fully reconstructed surfaces[79], but computational limitations have largely prevented the full consideration of both temperature effects and multi-site interactions for realistic surfaces. By considering the role of temperature, we have determined the temperature range over which surface order is expected to be most pronounced, as reflected in the temperature dependence of the zig-zag order parameter (Fig. 5.11). Additionally, the consideration of multi-site interactions has enabled the identification of a number of stable In-Ga orderings specific to the (2×4) reconstructions considered here, which may have relevance to bulk alloy order. While short-range order is quite strong in the alloyed surface studied here, long-range order on both the cation and dimer sublattices is weak at even moderate temperatures (200-500°C), as determined from the lack of any strong minima in the configurational entropy and from analysis of microstate snapshots of the various surface orderings. If order on the bulk cation sublattice is indeed due to the kinetic trapping of surface cation order during growth[8], we would expect the observed effect to be relatively short-range, capturing only those correlations which are too strong to be overcome by thermal disorder at synthesis temperatures (e.g., the strong difference in occupancy of the 4 sites and 3 sites in the $\beta 2(2\times4)$). However, since we have only considered a thin InGaAs layer at the GaAs lattice parameter, it is unclear how the order would propagate during film growth, or how surface order would manifest in lattice matched alloys.

Lateral composition modulation. We have also demonstrated the occurrence of a lateral composition modulation on the cation sublattice, driven by mutual ordering of the cation and dimer sublattices. This is most evident in the vicinity of 75% dimer coverage, where the $h_{\alpha\beta}$ hybrid reconstruction occurs, characterized by locally increased In concentrations at $\alpha 2(2\times4)$ cells, where substitution is energetically preferred. Yet again a consequence

of the strong inter-sublattice interaction, this ordering results in relatively strong cation concentration fluctuations over larger length scales than any of the other predicted orderings. Conceivably, similar regimes may exist for other reconstructions, but composition modulation may manifest differently in other systems, depending on symmetry and repeat units. Hence other systems may exhibit hybrid reconstructions with different geometries and modulation length scales. Although atomic-scale cation concentration is difficult to measure in practice, TEM studies of InGaP films, for example, have shown what are thought to be regions of high and low In concentration resulting from variations in atomic surface structure[89]. These composition fluctuations, termed branch defects, can have a measurable effect on device performance by pinning dislocations, which act as non-radiative recombination sites. Additionally, reconstructions have been directly implicated in lateral composition modulation in work by Pearson *et al.*, which demonstrated a correlation between lateral composition modulation using TEM with periodic variations in surface reconstruction observed using STM in the InGaAs system[90].

CHAPTER VI

The Comprehensive InAs/GaAs(001) Surface Phase Diagram

6.1 Introduction

The results of the previous chapter elucidated, using the (2×4) as an example case, the strong influence that alloying can have on the order and phase stability of the surface reconstruction. Because the model used in Ch. V coupled alloy degrees of freedom to structural degrees of freedom, we were able to suggest basic guidelines to anticipate the ways in which alloying the surface with a size-mismatched species may affect its atomic structure. However, because the model used to obtain these results only considered the GaAs (2×4) reconstruction, it cannot be used to make detailed predictions outside the range of chemical potential over which the (2×4) is stable, and that chemical potential range itself cannot be determined without also comparing the (2×4) reconstruction phase to other potentially stable reconstructions.

Due to the low surface energy of InAs and large In–As bond length, relative to GaAs, InAs exhibits surfactant behavior at the GaAs(001) surface. Despite the immiscibility of bulk InAs and GaAs, InAs deposited on GaAs wets the surface to form a two-dimensional alloy[91] that tends to segregate as a floating layer if additional GaAs is subsequently deposited[80]. At a critical wetting layer thickness in the range of 1.2–1.5 ML, depending on temperature, the surface morphology undergoes a transition from two-dimensional to three-dimensional islands[92]. When properly encapsulated within a thin-film heterostructure, these self-assembled islands behave as quantum dots[93], whose tunable electronic properties can be harnessed for cutting-edge device applications. Experiments indicate that the the surface undergoes a characteristic sequence of reconstruction transitions[94, 95]. Starting from the $c(4\times4)$ surface of GaAs(001), deposition of InAs induces a transition first to a $(n\times3)/(4\times3)$ and then to a (2×4) phase before three-dimensional islands begin to form in the vicinity

of 1.4 ML InAs. Quantum dots can also form via an As-induced conversion reaction of In nanocrystals to InAs, in which case the reconstruction present prior to In deposition seems to significantly impact the resulting film morphology and dot uniformity[96].

In this chapter, we use the first principles methods described in previous chapters to better understand the progression of surface reconstructions that is observed on the InAs/GaAs(001) as In is deposited. To this end, we consider the effects of a atomic-size mismatch strain on the structural and configurational phase stability over the entire range of anion-rich surface reconstructions of GaAs(001). We do not consider finite temperature effects but instead use first principles surface energies at 0K, in conjunction with the cluster expansion formalism, to perform a comprehensive search for the ground-state reconstructions of the surface and reveal the alloying response of the system. We use these results to construct a first principles 0-K surface phase diagram of the InAs/GaAs(001) wetting layer.

6.1.1 Reconstructions of the InAs/GaAs(001) Wetting Layer

Experimental evidence indicates that the alloying degrees of freedom created by the introduction of In to the GaAs(001) surface extend the range of thermodynamically accessible surface reconstructions. On pure GaAs(001), as described in detail in Chs. III and IV, the surface can generally be described as passing through a number of distinct surface phases—from $\zeta(4\times2)$ to $\beta2(2\times4)$ to $\beta(4\times3)$ to $c(4\times4)$–α and finally to $c(4\times4)$–β—as the surface is taken from Ga-rich to As-rich conditions. Observations of the InAs wetting layer, however, indicate that it departs from this progression in a number of ways. Starting from the GaAs-(2×4) surface under relatively low As overpressure, alloying with In tends to disorder surface while maintaining its (2×4) periodicity[94, 95], and both $\beta2(2\times4)$ and $\alpha2(2\times4)$ can be observed with STM on the resulting surface[97]. Under As-rich conditions, where the starting surface is GaAs-$c(4\times4)$, In deposition induces a transition to a poorly-ordered $(n\times3)$ surface, according to RHEED measurements[94]. High-resolution STM of this alloyed $(n\times3)$ phase on the InAs/GaAs wetting layer indicate that, at least under some conditions, it is comprised of (4×3) unit cells that are misaligned along the $[1\bar{1}0]$ direction[95, 97].

Although few theoretical findings exist for the structure of the 2D InAs/GaAs wetting layer, previous theoretical investigation of alloying on the InAs/GaAs surface at the $In_{0.5}Ga_{0.5}As$ lattice parameter did predict alloy-driven stabilization of the $\alpha2(2\times4)$ reconstruction at 0K[78]. Calculations also show that, relative to $\beta2(2\times4)$ and $\zeta(4\times2)$, the $\alpha2(2\times4)$ can be stabilized by isotropic compressive strain on GaAs(001)[98]. If the surface is sufficiently enriched in In relative to the bulk, as is certainly true for the InAs/GaAs wetting layer, the surface is effectively placed under compressive strain, suggesting that the observation of $\alpha2(2\times4)$ on the alloyed surface may be due to an increase in the equilibrium

lattice parameter of the surface reconstruction as it becomes enriched in In. Few theoretical studies of (n×3) reconstructions have been performed for GaAs-based surface systems. However, the stability of (n×3) has been tested on $In_{0.5}Ga_{0.5}As$ (001) using density functional theory, and the only (n×3) reconstruction predicted to be stable is restricted to a small region of chemical potential at extremely As-rich conditions[78]. It is described by a (2×3) unit cell that is not charge-balanced, according to the electron counting rule.

6.2 Methodology

The preferred strategy to predict the reconstruction phase stability of the InAs/GaAs(001) wetting layer is based on the same approach taken in previous chapters: we first select the set of most likely surface reconstruction prototypes for the system and then for each prototype explore the dependence of its surface energy on configuration. In Ch. III, we explored in detail the prototypes most likely to occur on the pure GaAs(001) surface and confirmed the well-accepted models for the GaAs-(2×4) [dubbed $\beta2$(2×4)] and GaAs-c(4×4). We additionally proposed a model for the GaAs-(4×3), which is observed at low temperature. These prototypes are all shown in Fig 6.1, in addition to the $\alpha2$(2×4), which is a near-stable variant of the $\beta2$(2×4). Figure 6.1 also include the ζ(2×4), which exhibits planar sp^2-type bonding not considered in Ch. III. It is observed on GaAs(001) under very Ga-rich conditions[63].

For each of the prototypes illustrated in Fig. 6.1, we specify the accessible alloying degrees of freedom. In other words, we select sites of the prototype that can undergo In–Ga species substitution or, in some cases, even As–cation substitution. Encompassing both As–In and As–Ga substitution, As–cation substitution is most likely to occur at sites with three-fold coordination, which are under-coordinated with respect to bulk atoms. As discussed in detail in Ch. I, these sites can undergo non-isovalent substitution while maintaining perfect filling of the valence-band surface electronic states. Of the tricoordinate sites available, the ones most amenable to non-isovalent substitution are the dimer sites, mainly due to their enhanced ability to accommodate the large relaxations caused by non-isovalent substitution. In this study we have limited As–cation substitution to the dimer sites of the (4×3) and c(4×4). Initial energy calculations of As–cation substitution on the GaAs-(2×4) prototypes showed a very large substitution energy, likely due to the necessity of forming high-energy cation–cation bonds. As such, As–cation substitution on the (2×4) was not considered. The sites that can undergo non-isovalent substitution are denoted in Fig. 6.1.

Due to the large bonding radius of In relative to Ga and the strong surfactant effect of InAs on GaAs(001), there is a strong thermodynamic driving for for In to segregate to the surface. Although In could isovalently substitute for Ga at any bulk cation site,

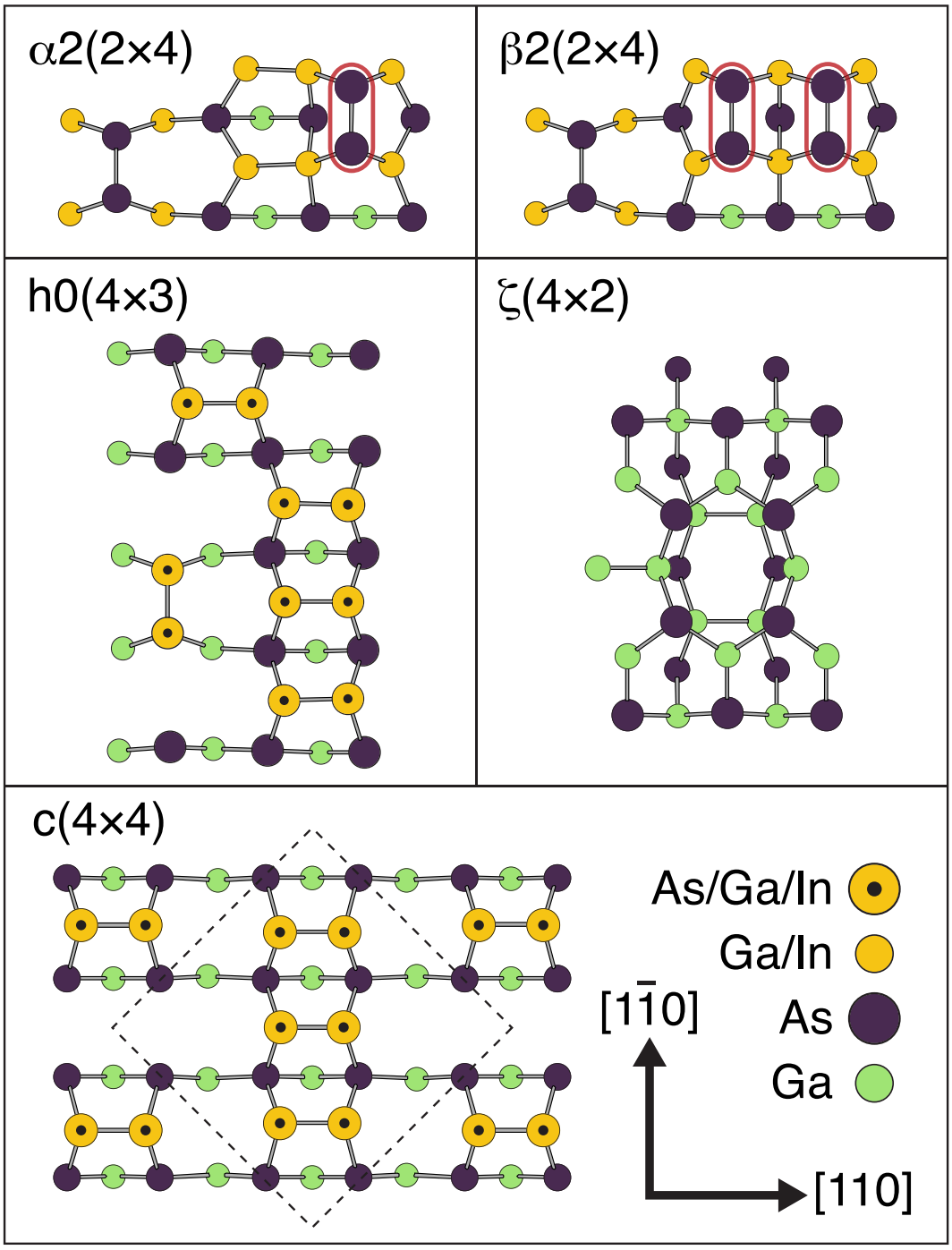


Figure 6.1. Illustration of the surface reconstruction prototypes considered when constructing the InAs/GaAs(001) surface phase diagram. Green (Ga) and purple (As) circles indicate sites whose species is fixed in all configurations. Gold circles indicate sites that can undergo Ga/In substitution, and gold circles with black centers indicate sites that can go As/Ga/In substitution. Circled dimers in the (2×4) prototypes indicate that these As dimers can undergo chemisorption/desorption.

the substitution energy increases significantly for sites below the surface[79]. As such, we only consider In–Ga substitution in the first subsurface layer. Since only the (2×4) prototypes possess cation sites in their first subsurface layer, they are the only ones for which isovalent substitution is considered. For the (2×4) prototypes, we also consider As-dimer chemisorption/desorption (i.e., varying configuration of As-dimer/dimer-vacancy disorder), thus allowing both the $\alpha 2(2\times4)$ and $\beta 2(2\times4)$ variants to be described on the same lattice model. This combined substitution/adsorption model is identical to the one used to study (2×4) alloying in detail in Ch. V.

We do not consider alloying in the $\zeta(4\times)$, interpreting it instead as a Ga-rich reference state. Although $\zeta(4\times2)$ may be able to accommodate some degree of In substitution, evidence suggests that alloying should not enhance its stability. The $\zeta(4\times2)$ is predicted to become less stable, relative to the $\alpha 2(2\times4)$ under compressive strain on GaAs(001)[98], and experimental characterization the (4×2) surface of InAs(001) suggests that when the surface becomes In-rich it is likely described by an altogether different reconstruction prototype[99].

The alloy degrees of freedom were explored by calculating DFT surface energies for a number of In/As/Ga configurations of each reconstruction prototype using the calculation scheme described in Ch. II. We performed a gound-state reconstruction search by fitting a cluster expansion Hamiltonian for each reconstruction prototype, resulting in separate Hamiltonians for the (2×4), (4×3). and $c(4\times4)$ unit cells, each of which was used to screen for ground-state configurations of its respective unit cell. The calculation–screening procedure was continued until self-consistency, meaning that all known ground-state configurations of the final cluster expansions are contained in the database of calculated configurations. Surface energies of the ground-state configurations were used to find the minimal 0K surface free energy over the range of As chemical potential μ_{As} and In chemical potential μ_{In}.

The collection of chemical potential pairs $(\mu_{\mathrm{As}}, \mu_{\mathrm{In}})$ at which surface energies of two minimal-energy ground states cross specify a reconstruction phase boundary. The chemical potentials are defined such that a phase boundary perpendicular to the μ_{In} axis indicates a boundary with fixed excess As composition, $x_{\mathrm{As}}^{(xs)}$; only the surface In composition $x_{\mathrm{In}}^{(xs)}$ changes as the boundary is crossed at 0K. Boundaries perpendicular to the μ_{As} axis indicate transitions at fixed $x_{\mathrm{In}}^{(xs)}$. $x_{\mathrm{In}}^{(xs)}$ is equivalent to the surface In coverage $N_{\mathrm{In}}/A_{(1\times1)}$, since In is completely insoluble in bulk GaAs at 0K. μ_{As} is defined such that the GaAs bulk is unstable relative to bulk As for all $\mu_{\mathrm{As}} > 0$. Similarly, $\mu_{\mathrm{In}} > 0$ corresponds to instability of GaAs relative to bulk InAs. Due to the surfactant nature of InAs on GaAs, bulk GaAs is stable relative to both bulk InAs and bulk In over the entire range of wetting layer compositions considered, according to DFT.

6.3 Results and Discussion

To construct the phase diagram, we performed a total of 840 first principles surface energy calculations corresponding 378 configurations of the (2×4), 296 configurations of the (4×3), and 166 configurations of the c(4×4). However, using the cluster expansion as a tool for performing the ground-state search allows us to screen a much larger set of configurations for new groundstates. In fact, with the cluster expansion to guide our search, we are limited mainly by our ability to enumerate candidate configurations and to manage the very large data sets that are generated. Using the cluster expansion for each prototype we screened over 30,000 (4×3) configurations and 200,000 c(4×4) configurations. Simulated annealing, which also screens many thousands of configurations using the cluster expansion, was used to sample (4×4) and (2×8) configurations of the (2×4) prototypes. Given the number of configurations that we have sampled, we can be relatively certain to have identified the most important configurational ground states of the prototypes under consideration.

The predicted surface phase diagram of the InAs/GaAs(001) wetting layer was obtained by identifying the configuration with the lowest surface free energy, $\gamma_i(T = 0, \mu_{\text{As}}, \mu_{\text{In}})$, at each chemical-potential pair for which bulk GaAs is thermodynamically stable with respect to the bulk forms of In, Ga, As, and InAs. The resulting phase diagram is shown in Fig. 6.2 (a). It is evident from the phase diagram that DFT predicts every considered prototype to have stable configurations within at least a small region of chemical potential. Notably, both the (4×3) and α2(2×4) are predicted to be alloy-stabilized, in that they are not predicted stable at the In-poor extreme of μ_{In} but are stabilized as μ_{In} is increased. Interestingly, the (4×3) is predicted stable only within a small "island" of chemical potential within the larger region of c(4×4) stability. This behavior suggests that, similarly to the unalloyed case, the (4×3) and c(4×4) are very close in energy, and their stoichiometries have very similar dependencies on chemical potential.

The predicted (2×4) reconstruction ground states are unchanged from the results of Ch. V. However, unlike in Ch. V the phase diagram in Fig. 6.2 reveals the relation of the (2×4) to its surrounding phases. In particular, the comprehensive wetting layer phase diagram shows that a large range of In-containing β2(2×4) configurations are inaccessible, or at least metastable with respect to the c(4×4). By comparison, the hybrid (2×4) configurations, which consist of tiled unit cells of both α2(2×4) and β2(2×4), are stable over a relatively large region. The proximity of configuration phase boundaries in the region where the mixed (2×4) is stable implies that it will disorder easily at finite temperature, which was demonstrated by the analysis performed in Ch. V. The α2(2×4) is stable over a large section of the As-poor, In-rich region of chemical potentials. This result contextualizes many of the

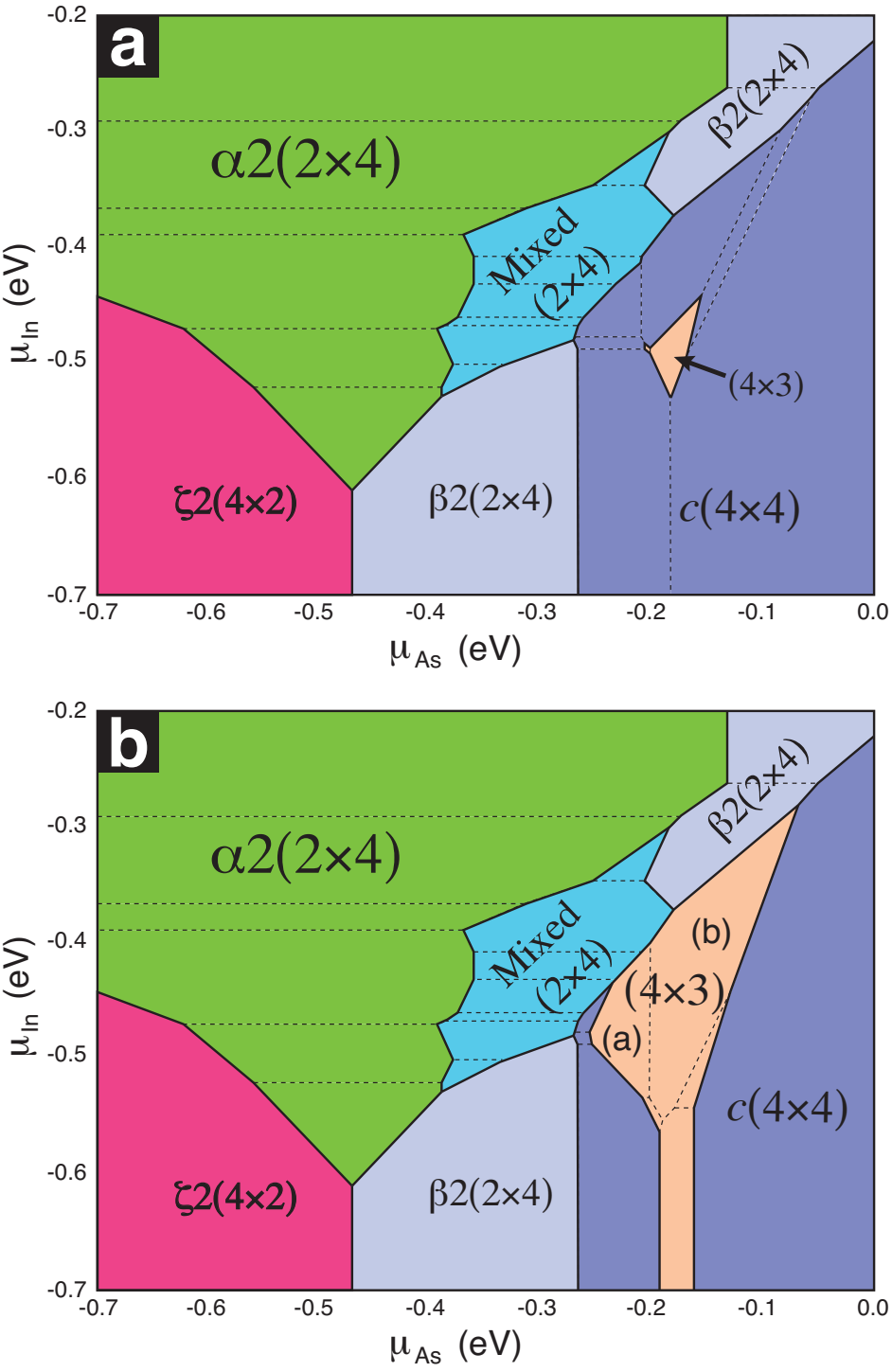


Figure 6.2. Surface phase diagrams of the InAs/GaAs(001) wetting layer, constructed by minimizing the 0K DFT surface energy. Two cases are shown: (a) the phase diagram obtained directly from DFT calculations and (b) the phase diagram obtained by considering an 8-meV negative shift to the (4×3) surface energies. In both cases, solid lines indicate transitions between distinct structural phases, while dashed lines indicate transitions between different configurations of the structural phase. In composition increases along the vertical axis, and surface As composition increases along the horizontal axis.

results of Ch. V, and we can see from the large region of the comprehensive phase diagram over which (2×4) is stable that the results of Ch. V likely have broad relevance to the thermodynamics of the InAs/GaAs(001) wetting layer.

The topology of the surface phase diagram with respect to (4×3) presents a compelling puzzle. A thorough search of the literature on the InAs/GaAs wetting layer finds that it neglects to mention the possibility of a GaAs-(4×3) starting surface, although it is evident from the broader experimental literature on GaAs(001) and from the results presented in Ch. IV that such a surface should exist at compositions intermediate to $c(4\times4)$ and $\beta2(2\times4)$. This discrepancy is likely due to the weak thermodynamic driving force to nucleate (4×3) domains when the metastable $c(4\times4)$ reconstruction is present on the surface. However, it is known that alloying greatly enhances stability of the (4×3) reconstruction phase[94, 95, 97], and the predicted free energies seems to follow this trend, since the (4×3) goes from being metastable at low μ_{In} to being stable when μ_{In} is increased.

In Ch. IV we estimated a lower bound on the error of DFT-calculated surface energies and used rigorous finite-temperature simulations to construct a GaAs(001) phase diagram and calibrate our DFT errors to the experimental phase diagram. Given the results of that calibration, it is reasonable to assume that DFT may predict an artificially high energy for the alloyed (4×3) relative to its surrounding phases. In order to explore this possibility we constructed a second surface phase diagram for the InAs/GaAs wetting layer where we assumed that the surface energy of the (4×3) is consistently predicted to be too high by a constant 8.0 meV/$A_{(1\times1)}$; this is slightly smaller than the 8.5-meV/$A_{(1\times1)}$ error calibration found in Ch. IV. The resulting phase diagram is shown in Fig. 6.2(a).

Applying an 8-meV/$A_{(1\times1)}$ shift to the (4×3) surface energies yields a significant change in the phase diagram topology. In Fig. 6.2(b), the (4×3) reconstruction phase is now stable at the un-alloyed boundary of the phase diagram, matching the result obtained in Ch. IV. Additionally, the alloy-induced stabilization is markedly enhanced. Whereas the original phase diagram exhibited only a small island of (4×3) stability, the phase diagram with shifted (4×3) surface energy shows a large incursion of (4×3) into the region of $c(4\times4)$ stability. The two predominant $c(4\times4)$ configurations that remain in the shifted phase diagram correspond to the $c(4\times4)$-γ and $c(4\times4)$-α, which were described in Ch. IV. Additionally, the region of (4×3) stability in the shifted phase diagram shares a phase boundary with the (2×4) reconstruction phase. This is a particularly notable result, as the the experimentally-measured surface phase diagram of the InAs/GaAs(001) wetting layer implies that such a phase boundary should exist[94]. The same phase diagram also indicates that the (4×3) surface phase should share a boundary with the $c(4\times4)$ surface phase, and a number of STM measurements show that the two phases can coexist on the wetting layer[92, 95].

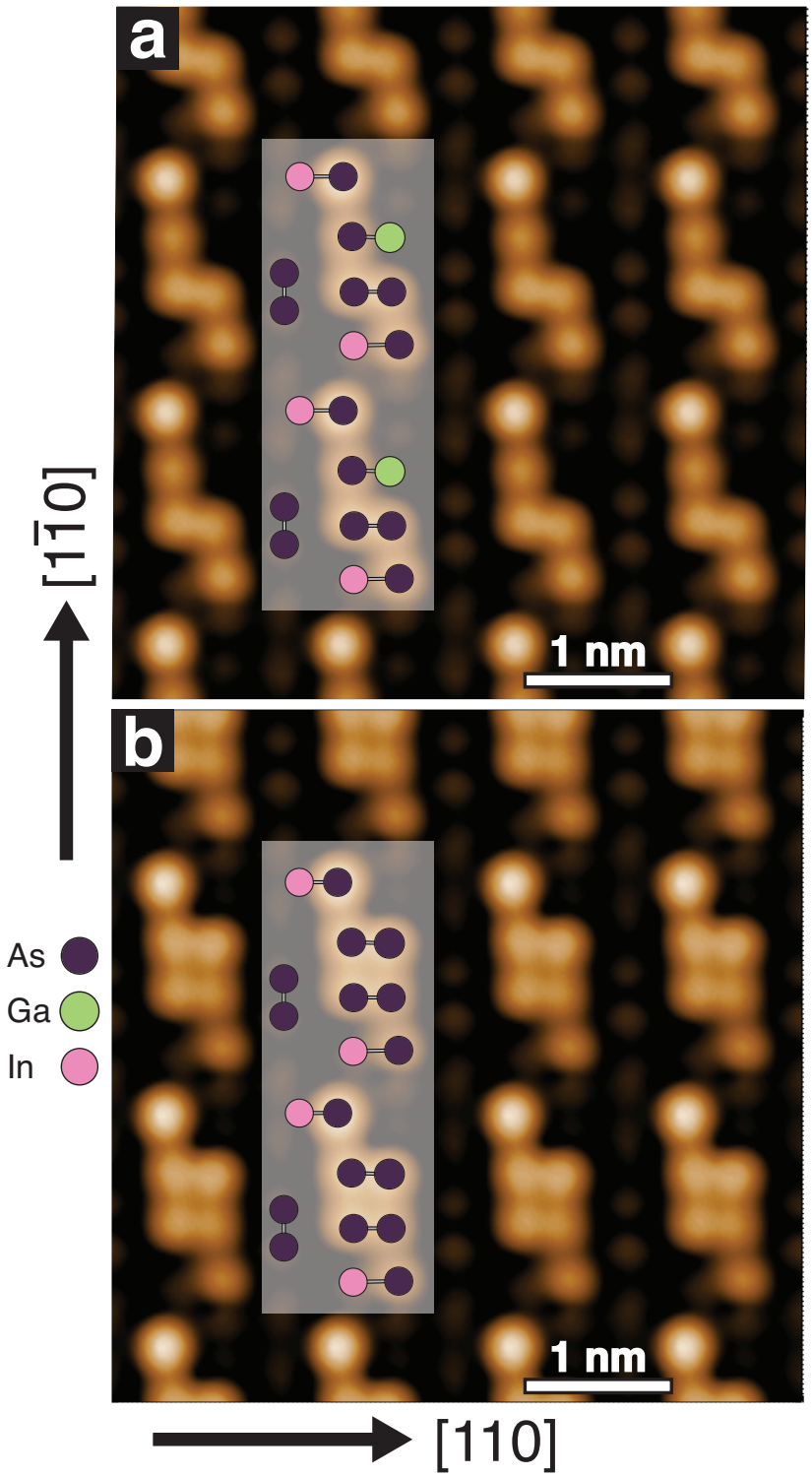


Figure 6.3. Simulated scanning tunneling micrographs of predicted (4×3) ground-state reconstructions of the InAs/GaAs(001) wetting layer. The reconstructions are configurations of the (4×3) reconstruction prototype illustrated in Fig. 6.1. (a) and (b) correspond to labeled phases in Fig. 6.2(b). Configuration (a) has two In and one Ga per unit cell, relative to the all-As $h0$(4×3) configuration; configuration (b) has two In per unit cell. Simulated micrographs correspond approximately to imaging conditions at a 1-V forward bias.

Although the (4×3) configuration that occurs at the un-alloyed boundary of the phase diagram in Fig. 6.2(b) is the $\beta(4\times3)$, which was discussed in Chs. III and IV, the configurations that comprise the two predominant alloyed configurations [labeled (a) and (b) in Fig. 6.2] of the shifted (4×3) have As–cation ordering that is distinct from $\beta(4\times3)$. The two configurations are illustrated in the insets of Figs. 6.3(a) and (b). In contrast to the $\alpha(4\times3)$ and $\beta(4\times3)$ configurations, which are energetically preferred on pure GaAs and GaSb, the alloyed (4×3) configurations exhibit a "staircase" or "chair" ordering motif, with one In atom substituting at the shifted dimer of the dimer row and another In atom substituting nearby in the "backbone" of the dimer row. This ordering motif is somewhat unexpected, given the tendency, as corroborated in Ch. IV, of sites in the backbone to have high energies for As–cation substitution in the un-alloyed systems.

The energetically preferred ordering motif of the alloyed (4×3) prototype should yield significantly different STM micrographs than either the $\alpha(4\times3)$ or $\beta(4\times3)$[59]. To demonstrate this, we have simulated STM micrographs for the two alloyed (4×3) configurations, which are shown in Figs. 6.3(a) and (b). The STM simulation uses the real-space partial charge of the relaxed surface corresponding to electronic states within a specified range of the Fermi level, calculated using DFT. The micrographs are obtained as an isocharge surface of the partial charge as the surface is approached from vacuum. This simulation scheme measures the predicted real-space distribution of states near the valence-band maximum, which, in the context of the established theory for STM[100], roughly approximates STM tunneling current at a voltage bias proportional to the specified energy-range. Partial charge densities were obtained using the implementation for band-wise charge decomposition included with VASP[38].

The simulated micrographs obtained for the stable configurations of the alloyed (4×3) differ significantly from simulated micrographs of either the $\alpha(4\times3)$ or $\beta(4\times3)$[59]. In particular, the "chair" motif of the alloyed (4×3) configurations is easily discerned, and the As atom of the shifted In–As heterodimer is visibly brighter than the other atoms. Although anion species are typically brighter in STM due to their valence-band dangling orbitals, As atoms on the shifted heterodimers in Figs. 6.3(a) and (b) are brighter than even the surrounding As atoms. The reason for the enhanced contrast of this As site remains unclear, though considering that the relaxed position of the site is not displaced farther from the surface than that of the surrounding anions, it is likely due only to differences in electronic structure.

The As–cation ordering motif, together with the enhanced contrast of the alloyed (4×3) in simulated STM has important relevance to experimental results. Although published micrographs of the $(n\times3)$ reconstruction on the InAs/GaAs(001) wetting layer are almost

never as well-ordered as the simulated micrographs in Figs. 6.3(a) and (b), they do share common traits. In particular, many published micrographs of the (n×3) surface phase on the InAs/GaAs(001) wetting layer have the appearance of narrow, meandering rows along [1$\bar{1}$0], interrupted at a periodicity of $\sim$4 times the surface lattice parameter[94, 60, 92, 95]. At short length scales many of these experimental micrographs of the (n×3) bear a striking resemblance to the simulated micrograph in Fig. 6.3(a), suggesting either that it is a fundamental unit of the (n×3) surface, or that it shares a fundamental motif with the (n×3), such as the shifted heterodimer and adjacent heterodimers in Fig. 6.3(a).

6.4 Summary

We have constructed a comprehensive 0K phase diagram for the complex InAs/GaAs(001) wetting layer from first principles. We employed the powerful combination of large-scale DFT calculations with the cluster expansion formalism to identify the ground-state reconstructions of the surface. We constructed a very large database of first principles surface energies for configurations of In, As, and Ga atoms at the under-coordinated sites of three different reconstruction prototypes and then used the cluster expansion formalism to leverage this data to screen thousands of additional configurations. The resulting phase diagram represents a particularly comprehensive prediction for the equilibrium structure of the InAs/GaAs(001) surface alloy, and the large number and variety of configurations that are predicted to be stable illustrate the necessity of conducting such a far-reaching ground state search.

The predicted phase diagram corroborates a number of trends that have been observed experimentally, including the alloy-induced stability of the (4×3) and α2(2×4) surface phases. However, better agreement with experiment can be obtained by applying an 8.0-meV/$A_{(1\times1)}$ negative shift to the (4×3) surface energy, consistent with our DFT error estimates and findings in Ch. IV. With the shift applied, the phase diagram exhibits close agreement with many of the experimental trends of the InAs/GAs wetting layer behavior. These include the existence of an un-alloyed (4×3), the shared boundary between (4×3) and (2×4) surface phases, and the significant alloy-induced enhancement in (4×3) stability. Moreover, simulated STM micrographs of stable configurations of the alloyed (4×3) differ significantly from the simulated or measured STM of the low-energy configurations on un-alloyed surfaces. The alloyed (4×3) configurations that are predicted to be stable instead exhibit a characteristic bright-spot–chair motif that at short length scales bears a strong resemblance to experimental STM of the InAs/GaAs(001) wetting layer.

CHAPTER VII

Summary and Relevance to Future Developments

In the context of historical strategies for predicting equilibrium surface structure, the work presented in this dissertation represents useful advances in systematizing the study of equilibrium surface phenomena that clear a path for future avenues of investigation that would otherwise be unwieldy. These developments follow a progression from codifying the enumeration of structural hypotheses to explain observed surface symmetries to employing energy models and simulation techniques to test these hypotheses and ultimately to exploring the effects that vibrational and configurational fluctuations have on thermodynamic stability, disorder, and alloying of the surface. Taken together, these methods permit some freedom from the trial-and-error search for low-energy arrangements and perturbations of the surface that computational investigations of surface behavior typically entail. Instead, the methods discussed here rely on formulating assumptions and guiding principles as a first step. Although the success of these methods hinges upon the proper choice of guidelines, clearly and deliberately codifying these assumptions can reveal obvious ways to improve or augment them.

The application of these methods to the (001) surface of zincblende III–V materials has revealed trends linking surface structure to electronic structure and finite temperature stability and disorder. The exhaustive nature of these methods, within the constraints of the initial assumptions, can reveal very basic relations for the III–V (001) surface, such as the relation between surface charge balance and surface unit cell geometry explored in Ch. III. Additionally, in Ch. IV these exhaustive methods revealed a situation where DFT possibly gives incorrect predictions for surface reconstruction stability on GaAs (001). The analysis in Ch. IV also revealed how vibrational and configurational disorder can alter thermodynamic stability of surface reconstructions in different directions, resulting in a finite-temperature surface phase diagram that differs significantly from zero-K stability predictions. The application of these methods to the InAs wetting layer on GaAs(001) in Ch. V revealed how size-mismatched alloying affects structural stability and enhances finite-temperature order.

This chapter summarizes the results of this work and discusses how the methods presented may be further developed and applied to other systems going forward.

7.1 Summary

Chapters I and II summarized the basic concepts that give rise to surface reconstruction and describe how theoretical and computational methods can be used to explore their effect on the thermodynamic stability of different surface structures. The overview of surface reconstructions and their relation to the III–V zincblende crystal presented in Ch. I along with the review in Ch. II of the theoretical framework within which equilibrium surface phenomena can be predicted from first principles can hopefully serve as introductory material suitable for any newcomer to first-principles equilibrium surface prediction. In particular, the electron counting rule (ECR) presented in Ch. I is an important principle that guides the choice of accessible structural and configurational degrees of freedom in later chapters. The cluster expansion formalism, described in Ch. II is an invaluable tool for exploring these degrees of freedom efficiently and predicted the surface free energies of arbitrary configuration of a surface structure. Later chapters rely heavily on the cluster expansion in particular, in addition to the thermodynamic and statistical mechanical theory overviewed in Ch. II.

Before the methods described in Ch. II can even be applied to simulations, testable hypotheses for structural models must exist. Chapter III described a simple but powerful approach that systematizes the enumeration of structural hypotheses. This approach, when applied to the GaAs (001) surface generates a large database of thousands of distinct structural hypothesis with unit cells with area up to 12 times the surface primitive cell area. An expression for the ECR excess surface charge, derived within the simplifying assumptions of the enumeration procedure, indicated that unit cells with areas that are odd multiples of the surface primitive cell area must have non-integer excess charge, according to the ECR heuristic. A first principles analysis of generated structures for experimentally-observed unit cell geometries revealed only a few low-energy candidates for the poorly understood $(n \times 3)$ surface that is observed experimentally on GaAs (001) and its alloys.

Chapter IV presented comprehensive theoretical analysis of surface reconstruction stability on GaAs (001) to compare the stability of the low-energy hypothesized structures identified in Ch. III by using more sophisticated methods that include consideration for finite-temperature phenomena. The finite-temperature surface free energies of these low-energy structures are calculated starting from first principles and account for the combined effects of finite-temperature configurational disorder and vibrational excitations. When related to the finite-temperature partial pressures of As_4, the surface free energies can be

compared to predict the GaAs (001) surface diagram. The calculated free energies were used to calibrate errors arising from approximations of DFT using experimental data, within previously established bounds, to obtain a predicted GaAs(001) surface phase diagram that exhibits good agreement with experiment. The calculated phase diagram indicates stability of a (4×3) reconstruction at low temperature, which was identified by the enumeration method of Ch. III.

On alloyed surfaces, where isovalent species substitution is possible in addition to the low-energy III/V substitution allowed at tricoordinate surface sites by ECR, changes in the alloy composition can destabilize structural stability and enhance ordering. These effects were explored in Ch. V using a cluster expansion effective Hamiltonian for alloying and As dimer chemisorption in an InAs wetting layer on GaAs (001). Ensemble averages of the global In composition and As dimer coverage demonstrated the ability of In, which has a larger ideal bond-length with As than does Ga, to destabilize the adsorption of As dimers as the surface becomes In-rich. Chapter V also included an in-depth analysis of surface order and disorder whereby topographic features of the entropy were used to identify ranges of cation and As chemical potential that are likely to exhibit order. Trends in the ensemble averaged In occupation of the symmetrically distinct surface sites indicate that site filling is highly anisotropic and strongly influenced by the local structure, including local dimer adsorption. A short-range order parameter, constructed to quantify the disorder of adsorbed dimers, provides additional evidence that In alloying enhances surface order via interaction with the adsorbed dimers. The order parameter, which measures "zig-zag" ordering of As dimers along $[1\bar{1}0]$ features a maximum near 25% surface-site In occupation that persists, at least to some degree, at synthesis temperatures.

To supply context to the complex ordering phenomena explored in Ch. V it is necessary to understand the basic trends that describe stability of the alloyed (2×4) surface phase relative to compositionally similar phases. Chapter VI described use of the cluster expansion formalism with first principles methods to perform a comprehensive search for ground-state reconstructions of the alloyed InAs/GaAs(001) surface and construct a zero-temperature phase diagram to predict the stability of the (2×4), (4×3), c(4×4) and (4×2) phases of the alloyed surface. The phase diagram revealed that the $\alpha 2$(2×4) and (4×3) both undergo alloy-induced stabilization relative to the stable phases of GaAs(001). Moreover, analysis of the sensitivity of (4×3) stability to energy shifts within the error bounds established in Ch. IV, showed that the alloyed (4×3) exhibits many of the qualitative properties of the experimentally-observed (n×3) surface phase. Additionally, simulated STM micrographs of the predicted low-energy configurations of the (4×3) demonstrate their similarity to the observed (n×3) reconstruction phase.

7.2 Future Developments

The developments described in this dissertation, the one that perhaps has the greatest potential for further development and application is the structural model enumeration described in Ch. III. This technique is distinct from most other tools used in the prediction of equilibrium surface structure in two significant respects: (i) it is exhaustive, enumerating all possibilities within the constraints posed by the codified set of assumptions regarding the most likely structural motifs that comprise the surface, and (ii) it introduces the artificial specie contrivance to abstract the in-plane bond. The first distinction suggests paths to develop improvements in *information*; the second suggests paths to develop improvements in *method*.

The structural enumeration method detailed in Ch. III has permitted the generation of an unprecedented number of structural hypotheses. Although these data have been extremely useful in the search for stable reconstructions of the GaAs (001) surface, the large number of structurally diverse hypotheses contained in this database could also enable new ways to approach standing problems in surface science besides determining equilibrium structure. Because the structural rules by which the reconstruction prototypes are generated are chosen for their tendency to describe low-energy structures, one possible application of the prototype database is to search for structures that describe intermediate states that occur during surface diffusion, epitaxial growth, or surface reconstruction transitions. These processes are all governed by kinetic rates, which depend on the path taken from the initial state to the final state over a saddle point of the potential energy surface. The rates are determined by the most probable paths, which are also those with the lowest energy at the saddle point state. To an even greater extent than in the search for thermodynamically stable surface reconstructions, hypothetical migration paths are often proposed based more on trial and error and physical intuition than on a systematic search. If a collection of low-energy intermediate states can be identified by mining our database of enumerated structures, the most likely migration path passing through one or more of these states can be calculated from first principles using, for example, the nudged-elastice band method.

In order for such an approach to give useful output it is inadequate to only identify low-energy prototypes, since a sequence of low-energy prototypes may not be sufficiently similar to one another to constitute physically meaningful path between two ground states. To avoid such complications, it would be desirable to have a quantitative measure for the degree to which two enumerated prototypes are similar. Our use of an artificial dimer "specie" to abstract the in-plane bonds that form on III–V (001) surfaces may suggest an approach for identifying such a metric. The dimer specie abstraction can be placed on a bulk lattice

position to signify relaxation of the two nearest-neighbor sites in the layer below toward each other to form an in-plane bond. It does so without breaking any point-group symmetries of the in-plane bond or the atoms which it joins. All the structural prototypes obtained from the procedure described in Ch. III can ultimately be described on a single zincblende lattice model, due to the properties of the dimer specie. The sites of a comprehensive lattice model for surface structure have four possible occupation states: cation, anion, dimer, or vacancy. Consequently, we could formally describe both the configurational *and* structural dependence of the surface energy in terms of a quaternary cluster expansion. The basis functions of such a cluster expansion provide a natural way to describe the similarity among an initial, intermediate, and final state along a migration path, assuming that they all adhere to our structural assumptions.

A generalized lattice model for III–V surfaces could be used to explore properties other than surface free energy, and even properties that are not scalar. A likely avenue for future study of III–V surface stability is to explore the effect of anisotropic strain. Although substrate materials are not deliberately strained in experiment, the equilibrium lattice parameter of the substrate in general does not give rise to the surface lattice parameters that minimize the surface free energy. The surface energy of the anisotropically strained surface can be described within linear elasticity theory and is related to the stress and strain via the surface compliance, which is a two-dimensional tensor of rank four. Tensor quantities such as the surface compliance can also be described in terms of crystal basis functions[101]. If the surface compliance and surface energies of arbitrary prototypes can be predicted using generalized cluster expansions, we can execute a truly exhaustive search for ground-state reconstructions among the set of enumerated prototypes, even if they can only exist via an elastically stabilized construction coexistence, as discussed briefly in Ch. I.

Because some of the basis functions of a generalized surface cluster expansion would describe purely structural degrees of freedom, a rigorous expansion of some surface property in terms of these structural basis functions could potentially be distilled into a phenomenological formulation like the ECR or even the "linear combination of structural motifs" proposed by Zhang and Zunger[102]. In fact, such a distillation procedure could be used in future studies, either for III–V systems of other materials, to identify an "optimal" set of structural assumptions. By assuming a very permissive set of lattice-based structural guidelines, very many structural prototypes are generated, even for small unit cell geometries. A generalized surface cluster expansion fit to first principles calculations for small super cells can be used to restrict the set of structural guidelines, either by direct inspection of the parameterized interactions, or by machine-learning techniques. The number of prototypes enumerated in larger unit cells will decrease, as will their average surface energy, as the structural guidelines

become more restrictive.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] M. Pashley, "Electron counting model and its application to island structures on molecular-beam epitaxy grown GaAs(001) and ZnSe(001)," *Phys. Rev. B*, **40**, pp. 10481–10487, (1989).

[2] W. A. Harrison, *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond*. San Francisco: WH Freeman and Company, (1980).

[3] G. A. Somorjai, "Surface reconstruction and catalysis," *Annu. Rev. Phys. Chem.*, **45**, pp. 721–751, (1994).

[4] G. Zhou and J. Yang, "Initial oxidation kinetics of Cu(100), (110), and (111) thin films investigated by in situ ultra-high-vacuum transmission electron microscopy," *J Mater. Res.*, **20**, pp. 1684–1694, (2005).

[5] U. Welp, V. K. Vlasko-Vlasov, X. Liu, J. K. Furdyna, and T. Wojtowicz, "Magnetic domain structure and magnetic anisotropy in $Ga_{1-x}Mn_xAs$," *Phys. Rev. Lett.*, **90**, p. 167206, (2003).

[6] P. Kratzer, E. Penev, and M. Scheffler, "Understanding the growth mechanisms of GaAs and InGaAs thin films by employing first-principles calculations," *Appl. Surf. Sci.*, **216**, pp. 436 – 446, (2003).

[7] B. Voigtländer, "Fundamental processes in Si/Si and Ge/Si epitaxy studied by scanning tunneling microscopy during growth," *Surf. Sci. Rep.*, **43**, p. 127, (2001).

[8] R. Osório, J. E. Bernard, S. Froyen, and A. Zunger, "Ordering thermodynamics of surface and subsurface layers in the $Ga_{1-x}In_xP$ alloy," *Phys. Rev. B*, **45**, pp. 11173–11191, (1992).

[9] T. Hofmann, V. Gottschalch, and M. Schubert, "Far-Infrared dielectric anisotropy and phonon modes in spontaneously CuPt-ordered $Ga_{0.52}In_{0.48}P$," *Phys. Rev. B*, **66**, p. 195204, (2002).

[10] S. Froyen and A. Zunger, "Surface segregation and ordering in III–V semiconductor alloys," *Phys. Rev. B*, **53**, pp. 4570–4579, (1996).

[11] Z. Kollonitsch, H.-J. Schimper, U. Seidel, K. Mller, S. Neumann, F.-J. Tegude, F. Willig, and T. Hannappel, "Improved structure and performance of the GaAsSb/InP interface in a resonant tunneling diode," *J. Cryst. Growth*, **287**, pp. 536 – 540, (2006).

[12] L. J. Whitman, P. M. Thibado, S. C. Erwin, B. R. Bennett, and B. V. Shanabrook, "Metallic III–V (001) surfaces: Violations of the electron counting model," *Phys. Rev. Lett.*, **79**, pp. 693–696, (1997).

[13] P. Laukkanen, M. P. J. Punkkinen, H.-P. Komsa, M. Ahola-Tuomi, K. Kokko, M. Kuzmin, J. Adell, J. Sadowski, R. E. Perälä, M. Ropo, T. T. Rantala, I. J. Väyrynen, M. Pessa, L. Vitos, J. Kollár, S. Mirbt, and B. Johansson, "Anomalous bismuth-stabilized (2×1) reconstructions on GaAs(100) and InP(100) surfaces," *Phys. Rev. Lett.*, **100**, p. 086101, (2008).

[14] J. Houze, S. Kim, S.-G. Kim, S. C. Erwin, and L. J. Whitman, "Structure of AlSb(001) and GaSb (001) surfaces under extreme Sb-rich conditions," *Phys. Rev. B*, **76**, p. 205303, (2007).

[15] J. Schnadt, J. Knudsen, X. L. Hu, A. Michaelides, R. T. Vang, K. Reuter, Z. Li, E. Lægsgaard, M. Scheffler, and F. Besenbacher, "Experimental and theoretical study of oxygen adsorption structures on Ag(111)," *Phys. Rev. B*, **80**, p. 075424, (2009).

[16] O. Romanyuk, F. Grosse, and W. Braun, "Stoichiometry and bravais lattice diversity: An *ab initio* study of the GaSb(001) surface," *Phys. Rev. B*, **79**, p. 235330, (2009).

[17] Y. Garreau, K. Aïd, M. Sauvage-Simkin, R. Pinchaux, C. F. McConville, T. S. Jones, J. L. Sudijono, and E. S. Tok, "Stoichiometry and discommensuration on $In_xGa_{1-x}As$/GaAs(001) reconstructed surfaces: A quantitative x-ray diffuse-scattering study," *Phys. Rev. B*, **58**, pp. 16177–16185, (1998).

[18] J. M. Millunchick, A. Riposan, B. J. Dall, C. Pearson, and B. G. Orr, "Surface reconstructions of InGaAs alloys," *Surf. Sci.*, **550**, pp. 1 – 7, (2004).

[19] L. E. Sears, J. M. Millunchick, and C. Pearson, "The coexistence of surface reconstruction domains on strained heteroepitaxial films," *J. Vac. Sci. Technol. B*, **26**, pp. 1948–1951, (2008).

[20] J. C. Thomas, N. A. Modine, J. M. Millunchick, and A. Van der Ven, "Systematic approach for determination of equilibrium atomic surface structure," *Phys. Rev. B*, **82**, p. 165434, (2010).

[21] J. C. Thomas, J. M. Millunchick, N. A. Modine, and A. Van der Ven, "Surface atomic order of compound III-V semiconductor alloys at finite temperature," *Phys. Rev. B*, **80**, p. 125315, (2009).

[22] J. Gibbs, H. Bumstead, and R. Van Name, *Scientific papers of J. Willard Gibbs*, vol. 1. Longmans, Green, (1906).

[23] J. W. Cahn, "Thermodynamics of solid and fluid surfaces," in *Interface Segregation* (W. C. Johnson and J. M. Blakely, eds.), ch. 1, p. 3, Metals Park, OH: American Society for Metals, (1979).

[24] N. Moll, A. Kley, E. Pehlke, and M. Scheffler, "GaAs equilibrium crystal shape from first principles," *Phys. Rev. B*, **54**, pp. 8844–8855, (1996).

[25] F. Larche and J. Cahn, "Thermochemical equilibrium of multiphase solids under stress," *Acta Metall. Mater.*, **26**, pp. 1579 – 1589, (1978).

[26] P. Voorhees and W. C. Johnson, "The thermodynamics of elastically stressed crystals," in *Solid State Physics* (H. Ehrenreich and F. Spaepen, eds.), vol. 59, pp. 1 – 201, Academic Press, (2004).

[27] G. Ceder, "A derivation of the ising model for the computation of phase diagrams," *Comp. Mater. Sci.*, **1**, pp. 144 – 150, (1993).

[28] K. Binder and D. Heermann, *Monte Carlo Simulation in Statistical Physics: An Introduction*. Springer Series in Solid-State Sciences, Springer, (2002).

[29] J. M. Sanchez, F. Ducastelle, and D. Gratias, "Generalized cluster description of multicomponent systems," *Physica A*, **128**, pp. 334 – 350, (1984).

[30] D. de Fontaine, "Cluster approach to order-disorder transformations in alloys," in *Solid State Physics - Advances in Research and Applications* (H. Ehrenreich and D. Trunbull, eds.), pp. 33–176, New York: Academic, (1994).

[31] A. Zunger, "1st-principles statistical-mechanics of semiconductor alloys and intermetallic compounds," in *Statics and Dynamics of Alloy Phase Transformations* (Turchi, PEA and Gonis, A, ed.), vol. 319 of *NATO ASI Series B, Physics*, pp. 361–419, NATO, Science Affairs Division, (1994).

[32] K.-C. Li, "Asymptotic optimality for Cp, CL, cross-validation and generalized cross-validation: Discrete index set," *Ann. Stat.*, **15**, pp. pp. 958–975, (1987).

[33] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Orlando: Saunders College Publishing, (1976).

[34] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*. New York: Oxford University Press, (1989).

[35] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, **136**, pp. B864–B871, (1964).

[36] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, **140**, pp. A1133–A1138, (1965).

[37] D. C. Langreth and J. P. Perdew, "Theory of nonuniform electronic systems. i. analysis of the gradient approximation and a generalization that works," *Phys. Rev. B*, **21**, pp. 5469–5493, (1980).

[38] G. Kresse and J. Furthmüller, "Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set," *Phys. Rev. B*, **54**, pp. 11169–11186, (1996).

[39] D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method," *Phys. Rev. Lett.*, **45**, pp. 566–569, (1980).

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

[41] D. Vanderbilt, "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism," *Phys. Rev. B*, **41**, pp. 7892–7895, (1990).

[42] G. Makov and M. C. Payne, "Periodic boundary conditions in *ab initio* calculations," *Phys. Rev. B*, **51**, pp. 4014–4022, (1995).

[43] C. W. Glass, A. R. Oganov, and N. Hansen, "USPEX–Evolutionary crystal structure prediction," *Comput. Phys. Commun.*, **175**, pp. 713 – 720, (2006).

[44] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, "Predicting crystal structures with data mining of quantum calculations," *Phys. Rev. Lett.*, **91**, p. 135503, (2003).

[45] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder, "Predicting crystal structure by merging data mining with quantum mechanics," *Nat. Mater.*, **5**, pp. 641–646, (2006).

[46] F. Chuang, C. Ciobanu, V. Shenoy, C. Wang, and K. Ho, "Finding the reconstructions of semiconductor surfaces via a genetic algorithm," *Surf. Sci.*, **573**, pp. L375–L381, (2004).

[47] L. Däweritz and R. Hey, "Reconstruction and defect structure of vicinal GaAs(001) and $Al_xGa_{1-x}As(001)$ surfaces during MBE growth," *Surf. Sci.*, **236**, pp. 15 – 22, (1990).

[48] J. Neugebauer, T. Zywietz, M. Scheffler, J. E. Northrup, and C. G. Van de Walle, "Clean and As-covered zinc-blende GaN (001) surfaces: Novel surface structures and surfactant behavior," *Phys. Rev. Lett.*, **80**, pp. 3097–3100, (1998).

[49] H. A. H. AL-Brithen, R. Yang, M. B. Haider, C. Constantin, E. Lu, A. R. Smith, N. Sandler, and P. Ordejón, "Scanning tunneling microscopy and surface simulation of zinc-blende GaN(001) intrinsic 4× reconstruction: Linear gallium tetramers?," *Phys. Rev. Lett.*, **95**, p. 146102, (2005).

[50] L. Ferreira, S.-H. Wei, and A. Zunger, "Stability, electronic structure, and phase diagrams of novel inter-semiconductor compounds," *Int. J. High Perform. C.*, **5**, pp. 34–56, (1991).

[51] G. L. W. Hart and R. W. Forcade, "Generating derivative structures from multilattices: Algorithm and application to hcp alloys," *Phys. Rev. B*, **80**, p. 014120, (2009).

[52] T. Arai, M. Suzuki, Y. Ueno, J. Okabayashi, and J. Yoshino, "Structure transition between two GaAs(001)-$c(4\times4)$ surface reconstructions in As_4 flux," *J. Cryst. Growth*, **301-302**, pp. 22 – 25, (2007). 14th International Conference on Molecular Beam Epitaxy - MBE XIV.

[53] J. M. Sanchez, "Cluster expansions and the configurational energy of alloys," *Phys. Rev. B*, **48**, pp. 14013–14015, (1993).

[54] A. Van der Ven, J. C. Thomas, Q. Xu, B. Swoboda, and D. Morgan, "Nondilute diffusion from first principles: Li diffusion in Li_xTiS_2," *Phys. Rev. B*, **78**, p. 104306, (2008).

[55] Q. Xue, T. Hashizume, J. M. Zhou, T. Sakata, T. Ohno, and T. Sakurai, "Structures of the Ga-rich 4×2 and 4×6 reconstructions of the GaAs(001) surface," *Phys. Rev. Lett.*, **74**, pp. 3177–3180, (1995).

[56] M. D. Pashley, K. W. Haberern, W. Friday, J. M. Woodall, and P. D. Kirchner, "Structure of GaAs(001) $(2 \times 4) - c(2 \times 8)$ determined by scanning tunneling microscopy," *Phys. Rev. Lett.*, **60**, pp. 2176–2179, (1988).

[57] H. Xu, Y. Y. Sun, Y. G. Li, Y. P. Feng, A. T. S. Wee, and A. C. H. Huan, "STM observation of Ga-dimers on a GaAs(001)-$c(8 \times 2)$-Ga surface," *Phys. Rev. B*, **70**, p. 081313, (2004).

[58] I. Chizhov, G. Lee, R. F. Willis, D. Lubyshev, and D. L. Miller, "GaAs(001)-"2×3" surface studied by scanning tunneling microscopy," *Phys. Rev. B*, **56**, pp. 1013–1016, (1997).

[59] W. Barvosa-Carter, A. S. Bracker, J. C. Culbertson, B. Z. Nosho, B. V. Shanabrook, L. J. Whitman, H. Kim, N. A. Modine, and E. Kaxiras, "Structure of III-Sb(001) growth surfaces: The role of heterodimers," *Phys. Rev. Lett.*, **84**, pp. 4649–4652, (2000).

[60] T. J. Krzyzewski, P. B. Joyce, G. R. Bell, and T. S. Jones, "Wetting layer evolution in InAs/GaAs(001) heteroepitaxy: Effects of surface reconstruction and strain," *Surf. Sci.*, **517**, pp. 8 – 16, (2002).

[61] E. Penev, P. Kratzer, and M. Scheffler, "Atomic structure of the GaAs(001)-$c(4 \times 4)$ surface: First-principles evidence for diversity of heterodimer motifs," *Phys. Rev. Lett.*, **93**, p. 146102, (2004).

[62] W. Schmidt, "III–V compound semiconductor (001) surfaces," *Appl. Phys. A-Mater.*, **75**, pp. 89–99, (2002).

[63] S.-H. Lee, W. Moritz, and M. Scheffler, "GaAs(001) surface under conditions of low As pressure: Evidence for a novel surface geometry," *Phys. Rev. Lett.*, **85**, pp. 3890–3893, (2000).

[64] M. Masnadi-Shirazi, D. Beaton, R. Lewis, X. Lu, and T. Tiedje, "Surface reconstructions during growth of $GaAs_{1-x}Bi_x$ alloys by molecular beam epitaxy," *J. Cryst. Growth*, **338**, pp. 80 – 84, (2012).

[65] B. Ferguson and X. Zhang, "Materials for terahertz science and technology," *Nat. Mater.*, **1**, pp. 26–33, (2002).

[66] M. Kaminska, Z. Liliental-Weber, E. R. Weber, T. George, J. B. Kortright, F. W. Smith, B.-Y. Tsaur, and A. R. Calawa, "Structural properties of As-rich GaAs grown by molecular beam epitaxy at low temperatures," *Appl. Phys. Lett.*, **54**, pp. 1881–1883, (1989).

[67] I. S. Gregory, C. Baker, W. R. Tribe, M. J. Evans, H. E. Beere, E. H. Linfield, A. G. Davies, and M. Missous, "High resistivity annealed low-temperature GaAs with 100 fs lifetimes," *Appl. Phys. Lett.*, **83**, p. 4199, (2003).

[68] A. C. Warren, J. M. Woodall, J. L. Freeouf, D. Grischkowsky, D. T. McInturff, M. R. Melloch, and N. Otsuka, "Arsenic precipitates and the semi-insulating properties of GaAs buffer layers grown by low-temperature molecular beam epitaxy," *Appl. Phys. Lett.*, **57**, pp. 1331–1333, (1990).

[69] E. R. Brown, F. W. Smith, and K. A. McIntosh, "Coherent millimeter-wave generation by heterodyne conversion in low-temperature-grown GaAs photoconductors," *J. Appl. Phys.*, **73**, pp. 1480–1484, (1993).

[70] H. Tanoto, J. H. Teng, Q. Y. Wu, M. Sun, Z. N. Chen, S. A. Maier, B. Wang, C. C. Chum, G. Y. Si, A. J. Danner, and S. J. Chua, "Greatly enhanced continuous-wave terahertz emission by nano-electrodes in a photoconductive photomixer," *Nature Photon.*, **6**, pp. 121–126, (2012).

[71] X. Lu, D. A. Beaton, R. B. Lewis, T. Tiedje, and M. B. Whitwick, "Effect of molecular beam epitaxy growth conditions on the Bi content of $GaAs_{1-x}Bi_x$," *Applied Physics Letters*, **92**, no. 19, p. 192110, (2008).

[72] X. Liu, Y. Sasaki, and J. K. Furdyna, "Ferromagnetic resonance in $Ga_{1-x}Mn_xAs$: Effects of magnetic anisotropy," *Phys. Rev. B*, **67**, p. 205204, (2003).

[73] A. Ohtake, "Surface reconstructions on GaAs(001)," *Surf. Sci. Rep.*, **63**, pp. 295 – 327, (2008).

[74] K. Reuter and M. Scheffler, "Composition, structure, and stability of RuO_2(110) as a function of oxygen pressure," *Phys. Rev. B*, **65**, p. 035406, (2001).

[75] J. E. Bickel, N. A. Modine, A. Van der Ven, and J. M. Millunchick, "Atomic size mismatch strain induced surface reconstructions," *Appl. Phys. Lett.*, **92**, p. 062104, (2008).

[76] W. Barvosa-Carter, R. S. Ross, C. Ratsch, F. Grosse, J. H. G. Owen, and J. J. Zinck, "Atomic scale structure of InAs(001)-(2×4) steady-state surfaces determined by scanning tunneling microscopy and density functional theory," *Surf. Sci.*, **499**, pp. L129 – L134, (2002).

[77] C. Ratsch, W. Barvosa-Carter, F. Grosse, J. H. G. Owen, and J. J. Zinck, "Surface reconstructions for InAs(001) studied with density-functional theory and STM," *Phys. Rev. B*, **62**, pp. R7719–R7722, (2000).

[78] A. Chakrabarti, P. Kratzer, and M. Scheffler, "Surface reconstructions and atomic ordering in $In_xGa_{1-x}As(001)$ films: A density-functional theory study," *Phys. Rev. B*, **74**, p. 245328, (2006).

[79] J.-H. Cho, S. B. Zhang, and A. Zunger, "Indium-indium pair correlation and surface segregation in InGaAs alloys," *Phys. Rev. Lett.*, **84**, pp. 3654–3657, (2000).

[80] S. Franchi, A. Bosacchi, F. Colonna, P. Pascarella, P. Allegri, and V. Avanzini, "Indium surface segregation in InGaAs-based structures prepared by molecular beam epitaxy and atomic layer molecular beam epitaxy," *J. Cryst. Growth*, **150**, pp. 185 – 189, (1995).

[81] C. Dorin and J. M. Millunchick, "Lateral composition modulation in AlAs/InAs and GaAs/InAs short period superlattices structures: The role of surface segregation," *J. Appl. Phys.*, **91**, pp. 237–244, (2002).

[82] V. Ozoliņš, M. Asta, and J. J. Hoyt, "Elastic relaxations in ultrathin epitaxial alloy films," *Phys. Rev. Lett.*, **88**, p. 096101, (2002).

[83] B. Yang, T. Muppidi, V. Ozoliņš, and M. Asta, "First-principles theory of nanoscale pattern formation in ultrathin alloy films: A comparative study of Fe-Ag on Ru(0001) and Mo(110) substrates," *Phys. Rev. B*, **77**, p. 205408, (2008).

[84] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, **21**, pp. 1087–1092, (1953).

[85] M. Sauvage-Simkin, R. Pinchaux, J. Massies, P. Calverie, N. Jedrecy, J. Bonnet, and I. K. Robinson, "Fractional stoichiometry of the GaAs(001) $c(4\times4)$ surface: An in-situ x-ray scattering study," *Phys. Rev. Lett.*, **62**, pp. 563–566, (1989).

[86] P. Bone, J. Ripalda, G. Bell, and T. Jones, "Surface reconstructions of InGaAs alloys," *Surf. Sci.*, **600**, pp. 973 – 982, (2006).

[87] T.-Y. Seong, A. G. Norman, G. R. Booker, and A. G. Cullis, "Atomic ordering and domain structures in metal organic chemical vapor deposition grown InGaAs (001) layers," *J. Appl. Phys.*, **75**, pp. 7852–7865, (1994).

[88] O. Ueda, Y. Nakata, and T. Fujii, "Study on microstructure of ordered InGaAs crystals grown on (110)InP substrates by transmission electron microscopy," *Appl. Phys. Lett.*, **58**, pp. 705–707, (1991).

[89] A. Y. Kim, W. S. McCullough, and E. A. Fitzgerald, "Evolution of microstructure and dislocation dynamics in $In_xGa_{1-x}P$ graded buffers grown on GaP by metalorganic vapor phase epitaxy: Engineering device-quality substrate materials," *J. Vac. Sci. Technol. B*, **17**, pp. 1485–1501, (1999).

[90] C. Pearson, C. Dorin, J. M. Millunchick, and B. G. Orr, "Imaging the evolution of lateral composition modulation in strained alloy superlattices," *Phys. Rev. Lett.*, **92**, p. 056101, (2004).

[91] J. Belk, J. Sudijono, D. Holmes, C. McConville, T. Jones, and B. Joyce, "Spatial distribution of In during the initial stages of growth of InAs on GaAs(001)-c(4×4)," *Surf. Sci.*, **365**, pp. 735–742, (1996).

[92] M. Xu, Y. Temko, T. Suzuki, and K. Jacobi, "InAs wetting layer evolution on GaAS(001)," *Surf. Sci.*, **580**, pp. 30–38, (2005).

[93] M. Kroutvar, Y. Ducommun, D. Heiss, M. Bichler, D. Schuh, G. Abstreiter, and J. Finley, "Optically programmable electron spin memory using semiconductor quantum dots," *Nature*, **432**, pp. 81–84, (2004).

[94] J. Belk, C. McConville, J. Sudijono, T. Jones, and B. Joyce, "Surface alloying at InAs–GaAs interfaces grown on (001) surfaces by molecular beam epitaxy," *Surf. Sci.*, **387**, pp. 213–226, (1997).

[95] C. Prohl, B. Höpfner, J. Grabowski, M. Dähne, and H. Eisele, "Atomic structure and strain of the InAs wetting layer growing on GaAs(001)-*c*(4×4)," *J. Vac. Sci. Technol. B*, **28**, pp. C5E13–C5E21, (2010).

[96] A. Urbańczyk and R. Nötzel, "Low-density InAs QDs with subcritical coverage obtained by conversion of In nanocrystals," *J. Cryst. Growth*, **341**, pp. 24 – 26, (2012).

[97] J. Grabowski, C. Prohl, B. Höpfner, M. Dähne, and H. Eisele, "Evolution of the InAs wetting layer on GaAs(001)-*c*(4×4) on the atomic scale," *Appl. Phys. Lett.*, **95**, p. 233118, (2009).

[98] E. Penev and P. Kratzer, "First-principles study of InAs/GaAs(001) heteroepitaxy," in *Quantum Dots: Fundamentals, Applications, and Frontiers* (Joyce, BA and Kelires, PC and Naumovets, AG and Vvedensky, DD, ed.), vol. 190 of *NATO Science Series II-Mathematics Physics and Chemistry*, pp. 27–42, NATO, Sci Affairs Div; British Assoc Crystal Growth, (2005).

[99] D. L. Feldwinn, J. B. Clemens, J. Shen, S. R. Bishop, T. J. Grassman, A. C. Kummel, R. Droopad, and M. Passlack, "Anomalous hybridization in the In-rich InAs(001) reconstruction," *Surf. Sci.*, **603**, pp. 3321 – 3328, (2009).

[100] J. Tersoff and D. R. Hamann, "Theory of the scanning tunneling microscope," *Phys. Rev. B*, **31**, pp. 805–813, (1985).

[101] A. Van De Walle, "A complete representation of structure-property relationships in crystals," *Nat. Mater.*, **7**, pp. 455–458, (2008).

[102] S. B. Zhang and A. Zunger, "Method of linear combination of structural motifs for surface and step energy calculations: Application to GaAs(001)," *Phys. Rev. B*, **53**, pp. 1343–1356, (1996).