Thermodynamic theory of epitaxial alloys: first-principles mixed-basis cluster expansion of (In, Ga)N alloy film

2009 J. Phys.: Condens. Matter 21 295402

(http://iopscience.iop.org/0953-8984/21/29/295402)

IOP Publishing Journal of Physics: Condensed Matter

J. Phys.: Condens. Matter **21** (2009) 295402 (19pp) doi:10.1088/0953-8984/21/29/295402

Thermodynamic theory of epitaxial alloys: first-principles mixed-basis cluster expansion of (In, Ga)N alloy film

Jefferson Zhe Liu and Alex Zunger

National Renewable Energy Laboratory, Golden, CO 80401, USA

Received 29 September 2008, in final form 26 May 2009
Published 29 June 2009
Online at stacks.iop.org/JPhysCM/21/295402

Abstract

Epitaxial growth of semiconductor alloys onto a fixed substrate has become the method of choice to make high quality crystals. In the coherent epitaxial growth, the lattice mismatch between the alloy film and the substrate induces a particular form of strain, adding a strain energy term into the free energy of the alloy system. Such epitaxial strain energy can alter the thermodynamics of the alloy, leading to a different phase diagram and different atomic microstructures. In this paper, we present a general-purpose mixed-basis cluster expansion method to describe the thermodynamics of an epitaxial alloy, where the formation energy of a structure is expressed in terms of pair and many-body interactions. With a finite number of first-principles calculation inputs, our method can predict the energies of various atomic structures with an accuracy comparable to that of first-principles calculations themselves. Epitaxial (In, Ga)N zinc-blende alloy grown on GaN(001) substrate is taken as an example to demonstrate the details of the method. Two (210) superlattice structures, $(InN)_2/(GaN)_2$ (at $x = 0.50$) and $(InN)_4/(GaN)_1$ (at $x = 0.80$), are identified as the ground state structures, in contrast to the phase-separation behavior of the bulk alloy.

(Some figures in this article are in colour only in the electronic version)

1. Introduction: issues surrounding a thermodynamic description of epitaxial alloys

Since the collection of individual, binary semiconductor compounds [1] offers but a limited repertoire of pertinent material properties (e.g., band-gap, effective-masses, band offset with respect to a reference etc), it has become common to mix two or three such binary 'building-blocks' to achieve intermediate values of such properties. With the advent of vapor-phase growth techniques (e.g., MBE [2] and MOCVD [3]), epitaxial evaporation onto a fixed substrate has become the method of choice for making crystalline semiconductor alloys, replacing the more traditional substrate-free melt-growth (e.g. Bridgman) approaches. It is now clear that vapor-phase epitaxial growth not only produces greater purity and provides better process-control relative to the melt-growth, but that in fact it corresponds to an altogether different thermodynamic state of the system, possibly yielding different microstructures and different phase diagrams [4–14]. The difference stems from two aspects, one existing at the top of the film (the free surface) and one at its bottom (the film–substrate interface).

Firstly, the presence of a free surface gives rise to a translational symmetry break and, consequently, surface-modified atomic interactions. This surface contribution has significant influence on the atomic microstructures of alloys near the surface, e.g., surface segregation and different disorder-order transitions from the bulk alloys [13–19]. Additionally, the existence of an exposed free surface above the film can lead to a surface reconstruction that creates strain patterning in a few near-surface layers, and leads, in turn, to an energetic driving force for selective incorporation of the smaller (larger) of two alloy atoms at high (low) strain subsurface sites, leading to atomic ordering, [5, 6, 10–12], rather than randomness, in various semiconductor alloy thin films. The surface contribution persists only within several atomic layers next to the free surface. For example, in the Al–Co and Al–Ni systems, the perturbation induced by the free surfaces is almost zero beyond the 4th atomic layer from the surface [14, 17]. In the case of surface reconstruction, for example, InGaP alloy, the strain pattern due to the surface reconstruction disappears beyond the 4th–5th atomic layer. Progress on the first-principles studies on the alloy *surface thermodynamics* have been reviewed in [6, 19] and [17].

Secondly, the existence of lattice coherence between the alloy film and the underlying substrate induces a particular form of strain, adding an energy term U^{epi} to the alloy free energy and thus potentially altering the thermodynamic state of the system. In contrast to the alloy surface thermodynamics, the coherent strain from the substrate persists throughout the whole epitaxial alloy film. In light emitting diodes, diode lasers and III–V multijunction solar cell applications [20–22], the alloy films are typically 1 μm thick, and thus the contributions from the exposed surface can be assumed negligible. The epitaxial constraint from the substrate becomes dominant in deciding the thermodynamic state of such alloy films [5, 7, 8, 23–28]. For example, the ordered chalcopyrite structures $InGaAs_2$ and Ga_2AsSb are observed in the epitaxial alloy films (away from the top free surfaces) grown on InP(001) substrates. In this paper we do not discuss the top-surface contributions (which can often be controlled and eliminated chemically via surfactants [29], addition of fast-diffusers [30] or growth temperature changes [31]), but rather present an accurate and efficient method to study the more robust thermodynamic effects of substrate-coherent strain.

The effect of substrate-coherent strain is illustrated in figure 1. Two constituents AC and BC (with respective equilibrium lattice constants $a_{\text{AC}} \neq a_{\text{BC}}$) combine to form an alloy σ (i.e., a specific assignment of atomic occupation on lattice sites) with equilibrium lattice constant a_σ. Long-standing tradition [32] suggests that in a bulk alloy with large lattice mismatch constituents, the formation energy (i.e., the energy difference between structure σ and an equivalent amount of AC + BC)

$$\Delta H_{\text{bulk}}(\sigma) = E(x,\sigma) - [(1-x)E_{\text{AC}}(a_{\text{AC}}) + xE_{\text{BC}}(a_{\text{BC}})] \quad (1)$$

is usually positive [33, 34] for both a random alloy and ordered structures, where E is the total energy. Yet, when AC + BC are grown epitaxially on a substrate that is nearly lattice-matched with the alloy $a_{\text{sub}} \approx a_\sigma$, strain could destabilize $\text{AC}(a_{\text{sub}}) + \text{BC}(a_{\text{sub}})$ sufficiently to make the alloy have a negative epitaxial formation energy:

$$\begin{aligned}\Delta H^{\text{epi}}(\sigma) &= E(x,\sigma,a_{\text{sub}}) - [(1-x)E_{\text{AC}}(a_{\text{sub}}) \\ &\quad + xE_{\text{BC}}(a_{\text{sub}})] = \Delta H_{\text{bulk}}(\sigma) - [(1-x)U^{\text{epi}}_{\text{AC}}(a_{\text{sub}}) \\ &\quad + xU^{\text{epi}}_{\text{BC}}(a_{\text{sub}}) - U^{\text{epi}}(\sigma,a_{\text{sub}})], \end{aligned} \quad (2)$$

where U^{epi} is the strain energy due to biaxially deforming a structure from its own equilibrium lattice constant to a_{sub} on the substrate plane. Epitaxy could convert, in this case, a bulk-immiscible AC + BC to an ordered alloy. Thus, in addition to kinetic factors, the constraint present during epitaxial growth could control the ensuring atomic microstructure. Since the atomic microstructure decides the electronic, optical and transport properties of the film [6], it is important to have a quantitative description of the thermodynamics of a substrate-coherent epitaxial alloy.

The difficulty with the thermodynamic description of the bulk and epitaxial alloys is the existence of numerous configurations. Indeed, a binary A/B semiconductor alloy could have $\sim 2^N$ possible atomic configurations σ in an N lattice-site unit cell (e.g., $\sim 10^6$ configurations at $N = 20$).

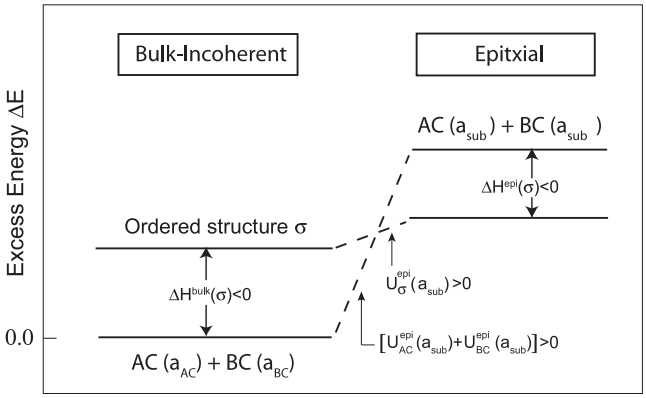

Figure 1. Schematic plot of energetic orders in bulk-incoherent and substrate-coherent epitaxial (A, B)C alloy. The lattice mismatch between AC and BC leads to a higher energy for the ordered structure σ than the incoherent phase separation $\text{AC}(a_{\text{AC}}) + \text{BC}(a_{\text{BC}})$, i.e., $\Delta E(\sigma) > 0$. In the presence of a substrate, the strain energy U^{epi} destabilizes the phase separation $\text{AC}(a_{\text{sub}}) + \text{BC}(a_{\text{sub}})$ and the formation enthalpy $\Delta H^{\text{epi}}(\sigma)$ becomes negative.

Many thermodynamic descriptions of semiconductor alloys do not attempt to resolve the energy $\Delta H(\sigma)$ of an individual configuration σ, instead modeling just the random alloy [4, 27, 35–37] (reviewed in [8, 38]). Such continuum phenomenological models work very well for many bulk semiconductor alloys. This is because a large lattice mismatch between the two constituents usually leads to a positive formation energy (of both the disordered solid solution and the ordered structures) and thus the phase diagram is characterized by the miscibility gap separating the disordered alloy from the phase separation. During the phase transition, only disordered alloy phases with different concentrations appear. In the case of a substrate-coherent epitaxial alloy, most of the works [27, 36, 39–43] have attempted to account for the epitaxy effect by simply adding the substrate strain energy term to the free energy of the disordered solid solution phase. Consequently, a large depression of miscibility gap temperature has been observed [27, 39–41, 43]. However, all these models failed to recognize that epitaxial strain could stabilize some ordered structures (i.e., negative formation energies $\Delta H^{\text{epi}}(\sigma)$) and thus convert the bulk phase separation into an epitaxially ordered one [7, 8]. It is, therefore, important to model the formation energy $\Delta H(\sigma)$ for all configurations, not just the disordered alloy.

Since direct quantum-mechanical calculations of the total energy $E_{\text{direct}}(\sigma) = \langle \Psi | H(\sigma) | \Psi \rangle$ are practical only for a few configurations $M \ll 2^N$, an efficient approach is to use M such calculated values $\{E_{\text{direct}}(\sigma)\}$, (e.g., from first-principles calculations) to fit a general mixed-basis cluster expansion [44] (CE) for the formation energy $\Delta H_{\text{CE}}(\sigma)$ in terms of the pair and many-body interactions

$$\begin{aligned}\Delta H_{\text{CE}}(\sigma) &= J_0 + \sum_i J_i s_i + \sum_{ij} J_{ij} s_i s_j + \sum_{ijk} J_{ijk} s_i s_j s_k \\ &\quad + \cdots + E_{\text{CS}}(k_\sigma) \end{aligned} \quad (3)$$

where s_i is the pseudospin variable ($s_i = -1$ or 1 if site i is occupied by atom type A or B respectively), J_0, J_i, J_{ij}, $J_{ijk} \ldots$

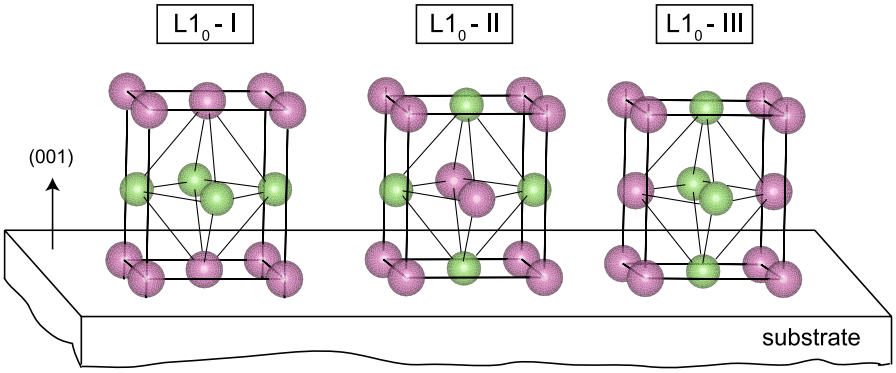

Figure 2. Three variants of the $L1_0$ structure (also called the CuAu-I type) on a (001) substrate. $L1_0$ is a superlattice structure composed of alternate atomic monolayers along the (100)-equivalent crystallographic directions: $L1_0$-I is along the (001), $L1_0$-II is along the (100) and $L1_0$-III is along the (010) crystal direction, respectively. Due to the presence of the substrate, the lattice constant on the substrate plane (i.e., (001) in this case) is fixed as $a_{\rm sub}$ while in the (001) direction, it is allowed to relax. It is obvious that the strain energy of $L1_0$-I is generally different from the strain energies of $L1_0$-II and $L1_0$-III.

are the interaction energies for the empty, point, pair, three-body ... 'figures'. The constituent strain (CS) energy $E_{\rm CS}(k_\sigma)$ is used to describe the long-range elastic energy:

$$E_{\rm CS}(k_\sigma) = \sum_{\mathbf{k}\neq 0} \frac{\Delta E_{\rm CS}(\hat{\mathbf{k}}, x)}{4x(1-x)} |S_\sigma(\mathbf{k})|^2 {\rm e}^{-(|\mathbf{k}|/k_c)^2}, \qquad (4)$$

where $S_\sigma(\mathbf{k})$ is the Fourier transform of the pseudospin variables of configuration σ, $\Delta E_{\rm CS}(\hat{\mathbf{k}}, x)$ is defined as the strain energy of the long period superlattice $(AC)_m/(BC)_n$ ($n/(m+n) = x$ and $m, n \to \infty$) along crystal direction $\hat{\mathbf{k}}$, and the last term is an attenuation function for a short concentration wave. All the quantities defining the mixed-basis cluster expansion (equations (3) and (4)) will be determined by first-principles calculations.

In equation (3), a set of lattice sites, called the 'figure', has the same interaction energies as other figures that are related to it by the space-group symmetry of the underlying lattice. Thus, geometrically, we can define the space-averaged pseudospin product $\bar{\Pi}_f(\sigma)$ corresponding to each class of symmetry-equivalent figure f of configuration σ as

$$\bar{\Pi}_f(\sigma) = \frac{1}{O_f} \sum_{f(i_1, i_2, \cdots i_m)} s_{i_1}(\sigma) s_{i_2}(\sigma) \cdots s_{i_m}(\sigma), \qquad (5)$$

where the sum runs over all O_f figures of class f. We can now express equation (3) in terms of the space-averaged pseudospin product $\bar{\Pi}_f(\sigma)$ as

$$\Delta H_{\rm CE}(\sigma) = J_0 + J_1(2x-1) + \sum_f D_f^{\rm pair} J_f^{\rm pair} \bar{\Pi}_f^{\rm pair}(\sigma) + \sum_g D_g^{\rm MB} J_g^{\rm MB} \bar{\Pi}_g^{\rm MB}(\sigma) + E_{\rm CS}(k_\sigma), \qquad (6)$$

where $D_f^{\rm pair}$ and $D_g^{\rm MB}$ (MB stands for many-body) are the number of figures of class f and g per lattice site. In this paper, we will use the term 'figure' to represent a class of symmetry-equivalent figures. In practice, the CS energy, $E_{\rm CS}(k_\sigma)$, of each configuration σ is subtracted from the formation energy calculated from the first-principles method. We only need to cluster expand the relative formation enthalpy $\Delta\tilde{H}_{\rm LDA}(\sigma) = \Delta H_{\rm LDA}(\sigma) - E_{\rm CS}(k_\sigma)$ as

$$\Delta\tilde{H}_{\rm CE}(\sigma) = J_0 + J_1(2x-1) + \sum_f D_f^{\rm pair} J_f^{\rm pair} \bar{\Pi}_f^{\rm pair}(\sigma) + \sum_g D_g^{\rm MB} J_g^{\rm MB} \bar{\Pi}_g^{\rm MB}(\sigma). \qquad (7)$$

This MBCE approach has been applied to many metallic and semiconductor bulk alloys, such as W–Ta [45], Mo–Ta [46], Au–Pd [47, 48], Ni–Pt [49], Cu–Au [49–54], Ag–Au [50, 51], Cu–Ag [50–54], Ni–Au [51–54], Al–Mg [53], Al–Zn [55], Al–Cu [55], InP–GaP [53, 56, 57], InN–GaN [58], etc.. It has also been applied to study alloy thermodynamics near the free surfaces, such as Al–Ni [14], Al–Co [17], and Pt–Rh [18]. In these studies, the surface contribution is taken as a correction to the bulk interaction and thus the CE are separated into two parts: the bulk CE (equivalent to the bulk alloy) and the surface CE. The interaction energies Js in the surface CE become site dependent and approach zero with increasing distance from the free surface. This method is developed to describe the *surface alloy thermodynamics* (persisting close to the exposed free surfaces) where the 'surface chemistry' is dominant. In this paper we will address the relative thick epitaxial thin film (1 μm) where the 'surface' effect is negligible and the epitaxial coherent strain effect (persisting throughout the whole alloy films) is dominant.

Relative to the bulk alloy, cluster expansion for the 'substrate-coherent epitaxial alloy' possesses three computational differences, which we will treat in this paper.

Firstly, whereas in a bulk alloy a certain configuration, σ, has a convenient geometry degeneracy, the presence of a substrate in the epitaxial system lifts this degeneracy, creating a few distinct $\sigma_{\rm epi}$ from one $\sigma_{\rm bulk}$. Figure 2 illustrates this situation for $\sigma_{\rm bulk} = L1_0$ (i.e., a CuAu-I structure) in an fcc alloy. $L1_0$ is a superlattice structure composed of alternate atomic monolayers along the (100)-equivalent crystallographic directions: $L1_0$-I is along the (001), $L1_0$-II is along the (100) and $L1_0$-III is along the (010) crystal direction, respectively. Due to the presence of the substrate, the lattice constant on the substrate plane (e.g., (001)) is fixed as $a_{\rm sub}$ while a free

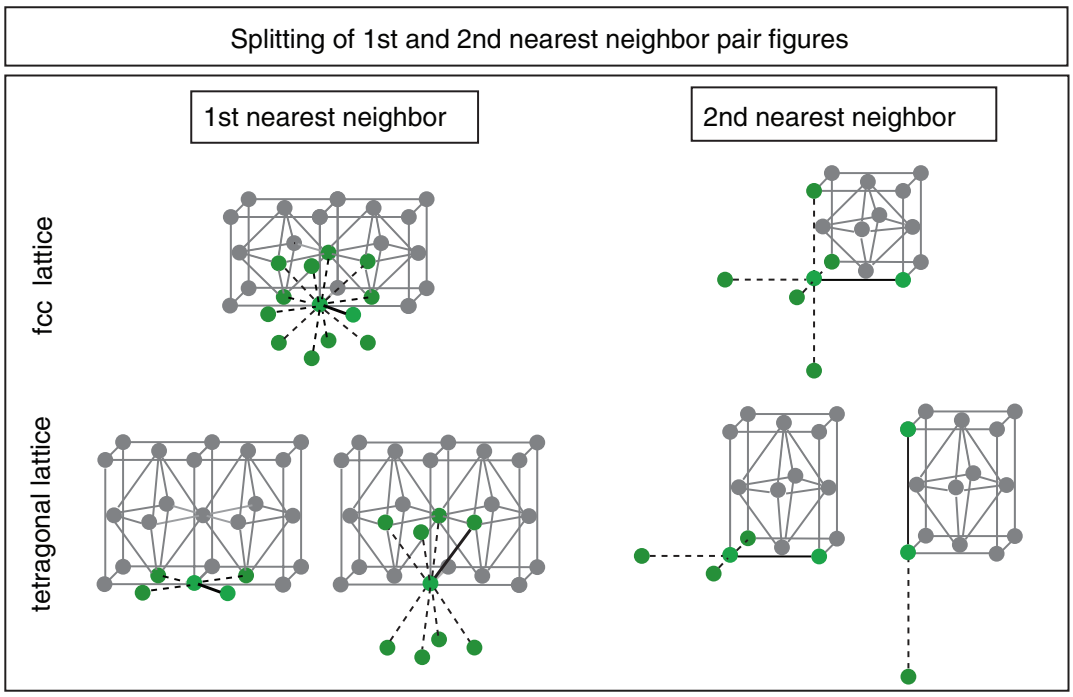

Figure 3. The splitting of the first and second pair figures on a tetragonal parent lattice (i.e., the elongated fcc parent lattice) from the cubic fcc lattice. On the fcc lattice, the nearest neighbor figures (i.e., the 12 solid and dashed bonds shown in the upper-left plot) are symmetry equivalent. On the elongated fcc lattice along the z direction, they split into two types of figures: four symmetry-equivalent pair figures on the xy plane and eight symmetry-equivalent pair figures out of the xy plane (i.e., the solid and dashed lines in the lower-left plot). Similarly, the second nearest figure on the fcc lattice (i.e., the 6 solid and dashed bonds shown on the upper-right plot) splits into two types of figures on the elongated fcc lattice along the z direction, four symmetry-equivalent figures on the xy plane and two out of the xy plane.

relaxation in the perpendicular direction is allowed. The substrate-coherent strain energy $U^{\rm epi}(\sigma)$ of $L1_0$-I is generally different from the strain energies of $L1_0$-II and $L1_0$-III. Thus we need to calculate, in the epitaxial case, more energies $E_{\rm direct}(\sigma_{\rm epi})$ than in the 'free-floating' bulk case.

Secondly, the interaction energy J_f and its associated figure f in the bulk alloy split into more members. For an fcc alloy grown on a (001) substrate, figure 3 illustrates the splitting of the first and second pair figures. On a bulk fcc lattice, the 12 nearest neighbor figures connected to a lattice site (i.e., the 12 solid and dashed bonds shown in the upper-left plot) are equivalent. As for the epitaxial fcc lattice, the (001) substrate constrains the figures on the substrate plane and thus makes them distinct from the figures out of the plane. They now split into two types: four equivalent pair figures on the (001) plane and eight equivalent pair figures out of the (001) plane (i.e., the solid and dashed lines in the lower-left corner of figure 3). Similarly, the second nearest figures on the fcc lattice (i.e., the 6 solid and dashed bonds shown on the upper-right corner of figure 3) split into two types, four equivalent figures on the (001) plane and two equivalent figures out of the plane. For the substrate-coherent epitaxial alloy, we need to reconstruct new figures and recalculate the $\{\bar{\Pi}_f(\sigma)\}$ for various atomic configurations.

Thirdly, for bulk fcc and bcc alloys, the constituent strain energy $\Delta E_{\rm CS}(\hat{\mathbf{k}}, x)$ is calculated as the strain energy to obtain lattice coherence at the interface $\hat{\mathbf{k}}$ of two semi-infinite slabs AC and BC (with different equilibrium lattice constants $a_{\rm AC} \neq a_{\rm BC}$). The two semi-infinite slabs AC and BC are biaxially strained to have the same lattice constant $a_\perp$ at the interface $\hat{\mathbf{k}}$ and freely relaxed along the perpendicular direction (to the interface $\hat{\mathbf{k}}$). In the substrate-coherent epitaxial alloy, since the interface $\hat{\mathbf{k}}$ is not necessary parallel to the substrate direction $\hat{\mathbf{Q}}$, when forming the lattice coherence at interface $\hat{\mathbf{k}}$, the free relaxation along the perpendicular direction would be constrained by the substrate. Thus, not only the coherence at interface $\hat{\mathbf{k}}$ but also the lattice coherence at the substrate-film interface $\hat{\mathbf{Q}}$ should be considered.

The three differences between the bulk and the epitaxial cluster expansions were treated in different ways in the past. Wood and Zunger [7, 8] employed the fcc bulk cluster expansion to describe the formation energy of the epitaxial alloy Cu–Au and Ga(As, Sb). The interactions J were fitted from the epitaxial formation energy $\Delta H^{\rm epi}(\sigma)$ of a few ordered structures (e.g., around five) calculated by the first-principles method or the VFF model, where only one specific variant of the ordered structures (with respect to the substrate) was considered. The splitting of atomic configurations and figures were not accounted for and the $E_{\rm CS}(k_\sigma)$ was set to zero. Teles and Ferreira *et al* [28] did first-principles calculations for different variants of an ordered structure on a given substrate. However, to avoid such splitting of σ and Js in their cluster expansion, they expanded the averaged energies and also set $E_{\rm CS}(k_\sigma) = 0$.

In this paper, we will present a method to consider fully all the three effects: splitting of σ and J_f, and the calculation of $\Delta E_{\rm CS}(\hat{\mathbf{k}}, x)$ in the presence of a substrate. We will take the epitaxial (Ga, In)N zinc-blende alloy grown

Table 1. $\bar{\Pi}_f(\sigma)$ of the three epitaxial variants of the $L1_0$ compound on the tetragonal lattice (elongated fcc along (001) crystal direction). $\bar{\Pi}_f(\sigma)$ of the bulk $L1_0$ compound on fcc lattice is shown for comparison.

Compounds	Description	1st pair		2nd pair		3rd pair		4th pair		5th pair	
		fcc lattice									
$L1_0$	{100} $(AC)_1/(BC)_1$ SL	−0.33		1.0		−0.33		1.00		−0.33	
		Tetragonal lattice									
$L1_0$-I	(001) $(AC)_1/(BC)_1$ SL	1.00	−1.00	1.00	1.00	−1.00	1.00	1.00	1.00	1.00	−1.00
$L1_0$-II	(100) $(AC)_1/(BC)_1$ SL	−1.00	0.00	1.00	1.00	0.00	−1.00	−1.00	1.00	−1.00	0.00
$L1_0$-III	(010) $(AC)_1/(BC)_1$ SL	−1.00	0.00	1.00	1.00	0.00	−1.00	−1.00	1.00	−1.00	0.00

on a GaN(001) substrate as an example to demonstrate our methodology. The thermodynamic properties of an epitaxial (Ga, In)N zinc-blende alloy on a GaN(001) substrate (e.g., the composition–temperature phase diagram and atomic microstructures) have been discussed and compared with those under other thermodynamic states (i.e., bulk-incoherent and bulk-coherent) in [59].

2. Methodology: general cluster expansion for epitaxial alloys

2.1. Formulation of cluster expansion for epitaxial formation energy $\Delta H^{epi}(\sigma)$

2.1.1. Selection of underlying lattice type. The general formulation of the MBCE has been summarized in equations (3) and (6). In principle, the MBCE is applicable to alloys on any Bravais lattice [44, 60], because for a given underlying lattice the pseudospin product functions $\bar{\Pi}_f$ of all figures f represent a complete basis set in the configuration space [60]. In the case of a semiconductor zinc-blende bulk alloy with alloyed species existing on either the cation or the anion sublattice (e.g., (In, Ga)As, Ga(As, Sb), etc.), the underlying Bravais lattice is fcc. For the zinc-blende epitaxial alloy deposit on an (001) substrate, we select the *tetragonal fcc lattice elongated along (001) direction*, because of the natural relaxation of the cubic fcc lattice caused by the constraint from the (001) substrate. Such a selection naturally describes the split of Js in comparison with bulk zinc-blende alloy (figure 3) and makes the epitaxial MBCE capable of distinguishing the different variants of an ordered structure (figure 2) (i.e., the splitting of atomic configuration). The same idea also applies to other epitaxial systems, including different lattice type (e.g., fcc, bcc, etc.) and different substrates direction $\hat{\mathbf{Q}}$ (e.g., (110), (210), (111), etc.), i.e., the underlying lattice of the epitaxial alloy can be selected as the elongated/compressed lattice of its bulk alloy along the substrate direction $\hat{\mathbf{Q}}$.

2.1.2. Generating figures and calculation of $\bar{\Pi}_f(\sigma)$. The selected underlying lattice of an epitaxial alloy, normally, has a lower symmetry than the underlying lattice of its bulk alloy. The figures, which were symmetry equivalent in bulk, now split into symmetry inequivalent figures under epitaxial conditions, as depicted in figure 3 (for an epitaxial fcc alloy on a (001) substrate). The determination of the new figure 'classes' can be done by exhaustively searching combinations of different lattice sites (on the selected underlying lattice) within a given cut-off radius and applying the group symmetry operations (of the underlying lattice) to remove the degeneracy.

For an epitaxial zinc-blende alloy grown on a (001) substrate, we construct a pool of figures (belonging to the new tetragonal fcc lattice), including 38 pair, 68 three-body, 40 four-body, 12 five-body and 6 six-body figures (in total 126 many-body figures) by gradually increasing their size in real space. The $\bar{\Pi}_f(\sigma)$ of all the atomic configurations with up to 16 cations per unit cell are also calculated by an exhaustive enumeration method [61]. Table 1 listed the calculated $\bar{\Pi}_f(\sigma)$ for the $L1_0$ compound on the tetragonal lattice (applied to the epitaxial alloy) and compares them with the $\bar{\Pi}_f(\sigma)$ on a cubic fcc lattice (applied to a bulk zinc-blende alloy). The $\bar{\Pi}_f(\sigma)$ of the $L1_0$-I variant is different from those of $L1_0$-II and $L1_0$-III.

2.1.3. Constituent strain energy for epitaxial alloy. To calculate the CS energy, $\Delta E_{CS}(\hat{\mathbf{k}}, x)$, of the substrate-coherent epitaxial alloy, not only the lattice coherence at the interface $\hat{\mathbf{k}}$ of the two semi-infinite slabs AC + BC, which is required for the bulk alloy, but also the constraint from the substrate $\hat{\mathbf{Q}}$ should be considered. A computational procedure is developed and shown in the appendix. Once the CS energies of several selected directions $\hat{\mathbf{k}}$ are calculated by the first-principles method, we use spatial harmonic functions (with the same symmetry as the underlying lattice) to fit the $\Delta E_{CS}(\hat{\mathbf{k}}, x)$ as a function of the spatial direction $\hat{\mathbf{k}}$. The coefficients of the harmonic functions are fitted as polynomials of the concentration x.

2.2. Fitting procedure

The number of terms needed to make an exact mapping of $\Delta H^{epi}_{LDA}(\sigma)$ to equation (6) is 2^N. In practice, some interactions are more important (depending on the chemical system), thus a truncated series of equation (6) may still give a good approximation. We need to select which pair and many-body figures are to kept in the truncated cluster expansion and decide their corresponding interaction Js. This is done in such a way that the resulting cluster expansion has a real prediction power, besides simply fitting well the first-principles calculations of a set of selected ordered structures. In other words, the key issue is to find the best combination of figures with the best prediction power. Intuitively, chemical and magnetic interactions become weaker as the separation of lattice sites in a figure increases. Therefore, some intuitive rules to select

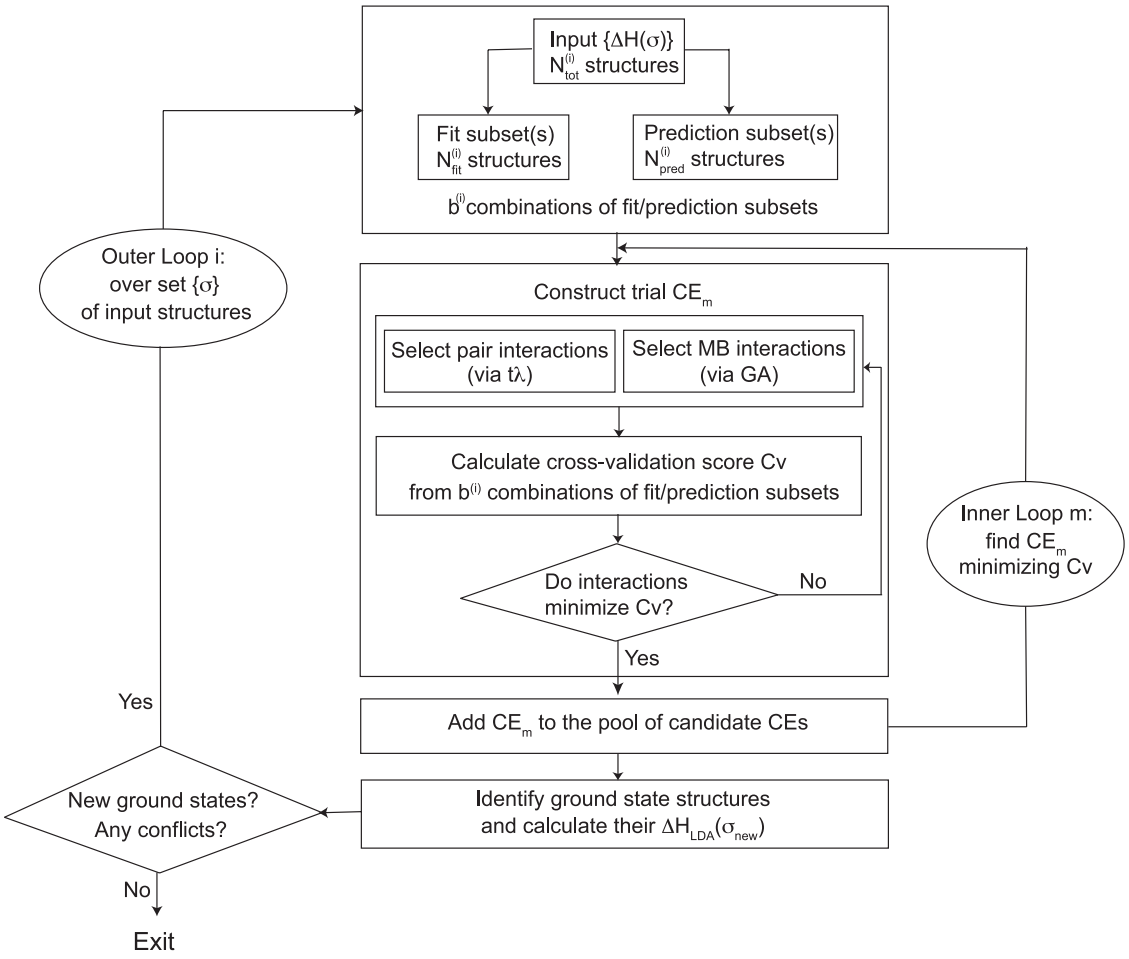

Figure 4. Flow chart of the iterative procedure used to establish the cluster expansion after figure 3, [48].

figures for a combination have been introduced [62, 63]: a figure can be included in the combination only if (1) all the figures having the same number of lattice sites but with a smaller size have been included and (2) all its sub-figures have been included. We apply this rule to our pair figure selections. But it is not feasible to apply this rule for many-body figures because of the exponential explosion of the number of many-body figures with increasing size [45]. In practice, if we adopted such intuitive selection rule, the affordable number of LDA input (e.g., ∼30–60) would make it impossible to select the high-order many-body figures (e.g., 4-body, 5-body, etc.), although there are well-established cases where the many-body interactions are important even in simple alloys [38, 64]. Instead, in our epitaxial cluster expansion, a many-body (i.e., three-body, four-body, ...) figure can be selected for the combination without interference from other figures. In summary, a pair figure can only be included in the combination if all the pair figures with smaller size have been included, while for a many-body figure it can be freely included.

Our fitting procedure follows the strategy developed in [45, 48, 65], which includes the two loops depicted in figure 4. The inner loop is intended to find the best cluster expansions, for a given set of input $\Delta\tilde{H}^{\rm epi}_{\rm LDA}(\sigma)$, with the best ability to predict the energies not included in the fitting. The outer loop iteratively adds the newly predicted ground state structures by the cluster expansions determined in the inner loop to the LDA data pool. The outer loop is repeated until the predicted ground state structures agree with the direct LDA calculations. The outer loop, here, actually acts as a feedback against spurious ground state predictions from the inner loop and/or artificially found on the basis of too few input structures. We will illustrate how this scheme works in section 3.

2.2.1. Inner loop: finding the best $\{J\}$ *from a given set of* $\{\Delta H^{epi}_{LDA}(\sigma)\}$. For each iteration in outer loop i, we provide a set of input LDA energies (the total number is $N^{(i)}_{\rm tot}$). To calculate the prediction power of different cluster expansions, we break $N^{(i)}_{\rm tot}$ into two groups: a fitting group made up of $N^{(i)}_{\rm fit}$ structures and a non-overlapping prediction group made up of $N^{(i)}_{\rm pred} = N^{(i)}_{\rm tot} - N^{(i)}_{\rm fit}$ structures. By dividing the two groups in different ways, we generate $b^{(i)}$ prediction sets. For a cluster expansion with the given figures, we firstly determine the interaction energy Js by fitting the LDA energies in the fitting group. Secondly, we compare its predictions with the direct LDA energies in the prediction group. The quality of the prediction power is measured by the averaged prediction errors over all the $b^{(i)}$ prediction sets. This gives the 'leave-many-out

cross validation score' (CV):

$$\mathrm{CV}(\mathrm{CE}_m) = \frac{1}{b^{(i)} N^{(i)}_{\mathrm{pred}}} \sum_{b^{(i)}\,\mathrm{sets}} \sum_{\sigma=1}^{N^{(i)}_{\mathrm{pred}}} |\Delta\tilde{H}^{\mathrm{epi}}_{\mathrm{CE}_m}(\sigma) - \Delta\tilde{H}^{\mathrm{epi}}_{\mathrm{LDA}}(\sigma)|^2. \tag{8}$$

In this paper, the size of the prediction set $N^{(i)}_{\mathrm{pred}}$ is set roughly to one third of the total size of the input $N^{(i)}_{\mathrm{tot}}$ [48]. While it is numerically prohibitive to use all $\binom{N^{(i)}_{\mathrm{pred}}}{N^{(i)}_{\mathrm{tot}}}$ possible prediction sets, we can choose the prediction sets (and the value of $b^{(i)}$) in such a way that each structure enters at least two prediction sets. Once determined, the prediction set is kept unchanged throughout the given outer loop iteration. The advantage of the leave-many-out cross validation score has been discussed in [46, 66].

The pair coefficients J^{pair}_f of a cluster expansion are determined using the '$t-\lambda$' constrained fit [44], which allows us to keep, in principle, an infinite number of pair terms. We rewrite the pair interaction of equation (6) in reciprocal space as

$$\sum_f D^{\mathrm{pair}}_f J^{\mathrm{pair}}_f \bar{\Pi}^{\mathrm{pair}}_f(\sigma) = \sum_{\mathbf{k}} J(\mathbf{k}) |S_\sigma(\mathbf{k})|^2, \tag{9}$$

where the sum is over a finite number of reciprocal space $\mathbf{k}$ vectors in the Brillouin zone of the underlying lattice for which $S(\mathbf{k},\sigma)$ is non-zero. We require $J(\mathbf{k})$ to be a smooth function and thus we define a 'smoothness value' M as

$$M = \frac{1}{\alpha}\sum_{\mathbf{k}} J(\mathbf{k})[-\nabla^2_{\mathbf{k}}]^{\lambda/2} J(\mathbf{k}) = \frac{N}{2\alpha}\sum_f R^\lambda_f D_f (J^{\mathrm{pair}}_f)^2, \tag{10}$$

where the exponent λ is a free parameter and α is a normalization constant:

$$\alpha = \frac{N}{2}\sum_f R^\lambda_f D_f. \tag{11}$$

The final objective function for fitting is then

$$\sum_{\sigma=1}^{N_{\mathrm{fit}}} w_\sigma |\Delta\tilde{H}^{\mathrm{epi}}_{\mathrm{CE}}(\sigma) - \Delta\tilde{H}^{\mathrm{epi}}_{\mathrm{LDA}}(\sigma)|^2 + tM, \tag{12}$$

where t is a Lagrangian multiplier. The $\{J^{\mathrm{pair}}_f\}$ and $\{J^{\mathrm{MB}}_g\}$ of the cluster expansion are determined by minimizing equation (12). The additional smooth function M naturally favors short-range pair interactions over long-range pair interactions; which is physical sensible. We access different values of t, λ and n_{pair} by means of the CV score to find the optimal $\{t, \lambda, n_{\mathrm{pair}}\}$ for a given combination of many-body figures.

We find the best combination of many-body figures in the inner loop via the genetic algorithm (GA). To do so, a large pool of many-body figures is constructed including more than 100 different types, much more than we will need (e.g., we will end up using only 8 many-body figures for the (Ga, In)N alloy on a GaN(001) substrate). Then we impose a restriction that N_{MB} many-body interactions be non-zero in the CE fit. The optimal combination of these N_{MB} non-zero figures is then explored by the GA using the procedure established in [65]. In this procedure, a 'population', consisting of N_{pop} members. evolves over a number of generations, with mating, mutation and adjustment steps [65] to replace the least-fit $(1-r_{\mathrm{s}})N_{\mathrm{pop}}$ CE individuals (i.e., those with large CV scores) with new, better individuals. In [65], it is found that the GA performance is not sensitive to the survival rate r_{s} but to the mutation rate and the 'lock-out' strategy which bring the GA out of deep local minima in the searching space. We adopted the parameters established in [65], including the average 'mutation rate' of two mutations per new individual and 'lock-out' after ∼100 locked-in generations. The GA was performed at different N_{MB} to find the optimal value $N^{\mathrm{opt}}_{\mathrm{MB}}$ by means of the CV score.

2.2.2. Outer loop: the ground state search and adding more $\{\Delta H^{epi}_{LDA}(\sigma)\}$. Although the usage of the CV score is statistically helpful to avoid over-fitting and quantify the prediction power [62], there is still a risk of finding a cluster expansion which happens to have a low CV score because of limited LDA input. This situation can be eliminated in the outer loop, where we will analyze the predictions of several cluster expansion candidates with low CV scores, instead of the one with the lowest value, to minimize the influence of over-fitting. At each iteration of the outer loop, we determine the number $N^{\mathrm{opt}}_{\mathrm{MB}}$ sufficient to obtain a low CV score and select a set of best cluster expansions with a number of many-body figures close to $N^{\mathrm{opt}}_{\mathrm{MB}}$. We then analyze the predicted energetics, especially the predicted ground state structures by this set of cluster expansions. Some of these newly predicted ground states are selected (based on some criteria described later) and calculated by the first-principles method. Their LDA energies are included in the next outer loop iteration. This ensures the redundancy of the input data used for fitting, resulting in a more reliable prediction.

Ground state structures are the phases that appear in the composition (x)—temperature (T) phase diagram as stable ordered phases at low temperature. The ground state structures characterize the thermodynamic properties and the microstructures of an alloy system and thus determine its mechanical, electronic, optical and magnetic properties. Finding the ground state structure is a classical problem in an inorganic solid state [67], metallurgy [68], and solid-state physics [38]. The outer loop, here, acts as a driving engine that helps us navigate through the configurational space, directing us to the ground state structures that are consistent with the direct LDA calculations.

At the end of each inner loop, we get a set of cluster expansion candidates from a given set of input structures. Then a ground state search is carried out with this set of CE candidates by calculating the $\Delta H^{\mathrm{epi}}_{\mathrm{CE}}(\sigma)$ of all 2^N structures up to, e.g., $N = 16$ cations per unit cell. This includes two steps: Firstly, we find the lowest energy structure σ_x at each concentration x (a finite number with a limited size of unit cell). It should be noted that not all these σ_x are ground states because they may be unstable with respect to disproportion into structures $\sigma_{x'}$ and $\sigma_{x''}$ at some other concentration x' and x''. Secondly, we build up a convex hull connecting the

Table 2. LDA epitaxial formation energies (meV/cation) of (Ga, In)N compounds grown on a GaN(001) substrate used as the input to the first iteration of the cluster expansion. For comparison, the bulk formation energies of (Ga, In)N *bulk* compounds are shown in the parenthesis following the structure name. Symmetry-equivalent variants of a given superlattice compound on an fcc lattice now split into distinct structures when grown on a (001) substrate (e.g., the $L1_0$ structure shown in figure 2). Here, $\{\cdots\}$ represents the symmetry-equivalent superlattice directions on a cubic fcc lattice, whereas $(\cdots)$ represents the distinct superlattice directions when grown on the (001) substrate. For example, {100} represents (100), (010) and (001) crystal directions, which are symmetry equivalent on the fcc lattice, while the (100) and (001) are distinct crystal directions due to the (001) substrate.

	Superlattice structures										
	{100}		{110}		{111}	{201}			{311}		
	(100) Variant-II	(001) Variant-I	(110) Variant-II	(011) Variant-I	(111)	(021) Variant-I	(210) Variant-II	(102) Variant-III	(131) Variant-I	(113) Variant-II	Others
$(GaN)_3/(InN)_1$	Z1 (66.98)		Y1 (63.46)		V1 (119.688)		$D0_{22}$ (52.10)		W1 (78.00)		$L1_2$ (84.59)
	9.10	−0.38	0.72	3.65	65.32	−8.35	−10.3	−10.3	20.27	16.72	26.87
$(GaN)_2/(InN)_1$	β_1 (80.87)		γ_1 (51.58)		α_1 (150.45)						
	18.63	0.63	−30.37	−23.54	80.33						
$(GaN)_1/(InN)_1$	$L1_0$ (93.51)				$L1_1$ (153.803)						
	29.57	−0.63			85.22						
$(GaN)_2/(InN)_2$	Z2 (90.14)		Y2 (74.01)		V2 (156.65)		CH (39.22)		W2 (68.18)		
	28.54	0.29	1.02	−6.06	91.74	−24.31	−57.53	−57.53	−6.33	−23.58	
$(GaN)_1/(InN)_2$	β_2 (83.05)		γ_2 (55.73)		α_2 (136.09)						
	35.41	−3.06	−25.69	−9.89	80.64						
$(GaN)_1/(InN)_3$	Z1 (69.81)		Y1 (59.20)		V1 (113.25)		$D0_{22}$ (36.94)		W1 (69.62)		$L1_2$ (59.36)
	32.66	−1.59	2.78	−1.87	66.18	−13.02	−25.62	−25.62	18.48	3.91	4.63

ground state structures so that all other structures have an energy higher than the connecting 'tie-line'. Such a direct enumeration approach limits the size of the unit cell we visit. In this paper, it is done up to $N = 16$ (about 10^5 structures). For practical reason, in the first few outer loop iterations, we even restricted the search to $N = 12$ cations per unit cell.

Following the strategy developed in [48], two quantities are crucial in deciding which newly predicted ground state structures will be included in the next iteration: the prediction frequency and the energy depth. The prediction frequency is defined as the frequency of occurrence of a given structure in the predictions of the different CE candidates. The energy depth, $\Delta_\sigma^{(m)}$, is the energy difference of a predicted ground state structure σ from the tie-line connecting the two neighboring ground states of the same cluster expansion candidate CE_m. It is used as a measure of the importance of a ground state structure. In this paper, we adopt the following criteria to select structures to be calculated by the first-principles method and added as input in the next outer loop iteration. (1) if a structure is predicted both frequently (>30%) and is energetically deep ($\Delta_\sigma^{(m)} > 1.5$ meV/cation), it will be included. (2) If a structure is predicted either frequently or is energetically deep, it will be included if its size is less than 8 cations per unit cell. (3) A structure will be included if it persists through several outer loop iterations, even if it is predicted neither frequently nor is energetically deep. Such criteria will help us include the deep ground states early in the outer loop iteration and at the same time, generally speaking, not miss the important but shallow states.

The outer loop iteration is terminated when the following criteria are satisfied: (1) there are no new predicted ground states whose energy are not already included, calculated by the first-principles method, and (2) there is a good agreement between the $\Delta H_{\rm LDA}^{\rm epi}(\sigma)$ and $\Delta H_{\rm CE}^{\rm epi}(\sigma)$ (several times the CV score) for all the input structures. When these criteria are satisfied, we select the one that gives the best description of the ground states of the LDA results.

2.3. Calculation of $\Delta H_{LDA}^{epi}(\sigma)$

The formation energy $\Delta H_{\rm LDA}^{\rm epi}(\sigma)$ (equation (2)) is calculated using first-principles density functional theory with the local density approximation (LDA) [69]. We use the ultrasoft pseudopotential and plane wave basis set as implemented in the VASP code [70]. The Brillouin zone is sampled with Monkhorst–Pack k-point meshes with roughly constant mesh densities corresponding to $9 \times 9 \times 9$ for the fcc unit cell. The basis set cut-off energy is set as 435 eV. Convergence tests for InN, GaN and $In_3Ga_1N_4$ $L1_2$ structures show that, with these settings, the error bound of the formation energy is about 1 meV/cation. During the calculations, the unit cell lattice constant of configuration σ on the substrate plane is fixed as $a_{\rm sub}$ while relaxation along the direction perpendicular to the substrate and atomic relaxation inside the unit cell are allowed.

The LDA inputs for the 1st outer loop iteration are listed in table 2 and compared with the formation energy $\Delta H_{\rm LDA}^{\rm bulk}(\sigma)$ of the bulk alloy (data in the parenthesis). These inputs cover almost all the 4-cation unit cell structures, most of which are well-defined layered structures along some given crystal direction, i.e., the superlattice (SL). It is clear that the symmetry-equivalent variants of a given structure σ of the bulk alloy (i.e., the fcc lattice) now split into distinct structures (shown as the multi-columns under the name of the structure in table 2). The formation energy of the epitaxial alloy (equation (2)) drastically decreases in comparison with the bulk formation energy (equation (1)) and even becomes

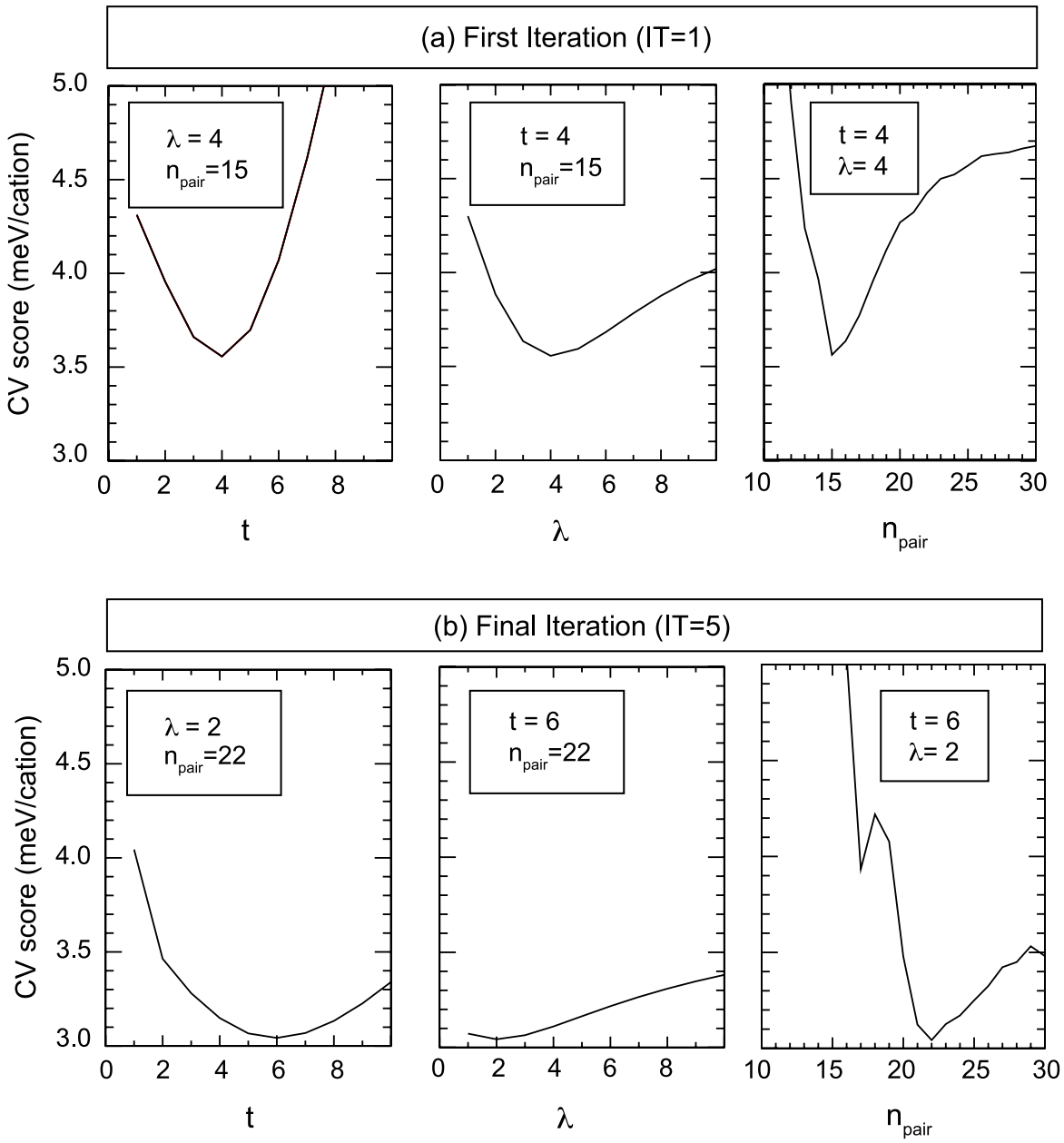

Figure 5. Selection of pair interaction by minimizing the prediction score for (a) the best CE candidate in the outer loop iteration = 1 and (b) the final CE iteration = 5.

negative. Generally, (201) superlattice structures have both lower $\Delta H^{\text{epi}}_{\text{LDA}}(\sigma)$ and $\Delta H^{\text{bulk}}_{\text{LDA}}(\sigma)$, while (111) superlattice structures have the highest values.

2.4. Calculation of Warren–Cowley short-range order

In disordered solid solutions, the occupations of lattice sites by different type of atoms are not perfectly random (especially at a temperature close to the phase boundary at a given alloy concentration). Instead, local ordering or local clustering takes place, which is referred to as short-range order (SRO). SRO can serve as the fingerprint of the ordering or phase-separation tendencies of an alloy system. One way to quantify the degree of such short-range ordering is the Warren–Cowley SRO parameters:

$$\alpha(R_{lmn})(x, T) = 1 - \frac{P^{(\text{A/B})}(R_{lmn})}{x}, \tag{13}$$

where $P^{(\text{A/B})}(R_{lmn})$ is the conditional probability that an A atom is at origin while a B atom is on site (lmn). This probability depends on concentration and temperature, thus leading to an x-dependence and T-dependence of α. If the alloy is perfectly random, the conditional probability $P^{\text{A/B}}$ is equal to x and thus $\alpha = 0$. The departure of α from 0 is, therefore, an sign that correlations between lattice sites exist in the disordered solid solution. A positive α indicates local clustering, while a negative α suggests local anti-clustering. From the definition equation (13), $\alpha(R_{000}) = 0$. All others are

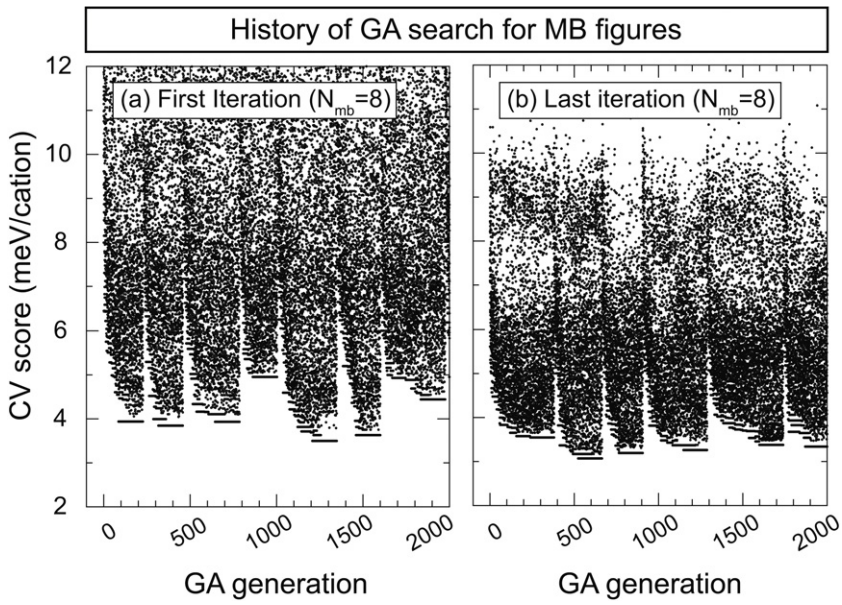

Figure 6. Genetic algorithm (GA) search for the optimal many-body interaction type. The search is repeated for different values of maximum number of many-body figures; shown are examples for (a) $N_{\mathrm{MB}} = 8$ in first iteration and (b) $N_{\mathrm{MB}} = 8$ in the last iteration.

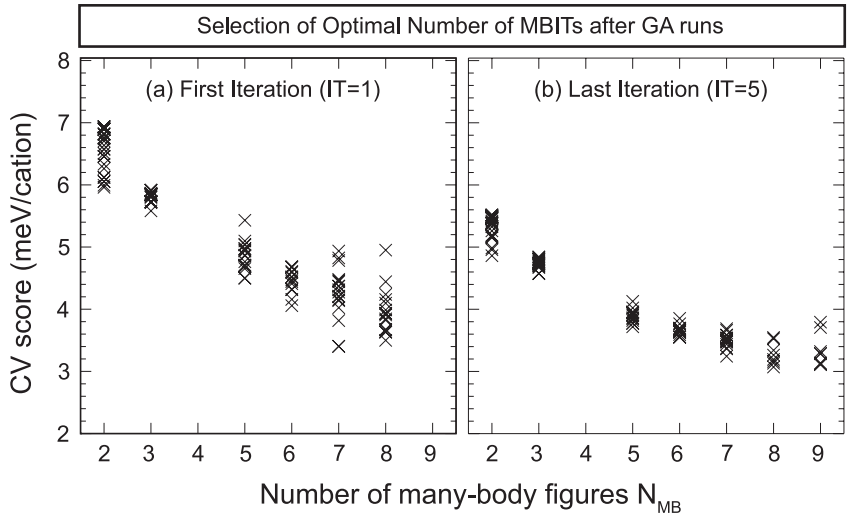

Figure 7. The prediction errors (CV score) of CE candidates found by GA (a) in first outer loop, and (b) in the final iteration, versus the maximum number of many-body figures that the GA is allowed to use.

then related to the pair $\Pi^{\mathrm{pair}}_{R_{lmn}}$ as

$$\alpha(R_{lmn}) = \frac{\langle \bar{\Pi}^{\mathrm{pair}}_{R_{lmn}} \rangle - (2x-1)^2}{1-(2x-1)^2}. \tag{14}$$

In diffraction experiments, short-range ordering does not give superstructure reflections, but gives rise to modulations in the monotonic Laue background. Using the diffuse scattering technique, one can examine these modulations between the Bragg peaks and extract the intensity portion, $I^{\mathrm{SRO}}_{\mathrm{diffuse}}$, which is proportional to the lattice Fourier transform of the Warren–Cowley SRO:

$$I^{\mathrm{SRO}}_{\mathrm{diffuse}} \propto \alpha(x, T, \mathbf{k}) = \sum_{lmn}^{n_R} \alpha_{R_{lmn}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot R_{lmn}}. \tag{15}$$

In this paper, we will calculate the SRO of (In, Ga)N epitaxial alloys at different compositions under a typical growth temperature by Monte Carlo simulations using the obtained CE as the energy functional.

3. Applications

(Ga, In)N semiconductor alloy has attracted great attention in the past decade due to its successful applications in blue and green light LEDs and laser diodes [71, 72]. It can be fabricated experimentally in both wurtzite (WZ) [72] and zinc-blende (ZB) [73–76] structures. Due to the structural similarity (e.g., WZ and ZB differ only in the relative handedness of the fourth interatomic bond along the (111) chain), the energy

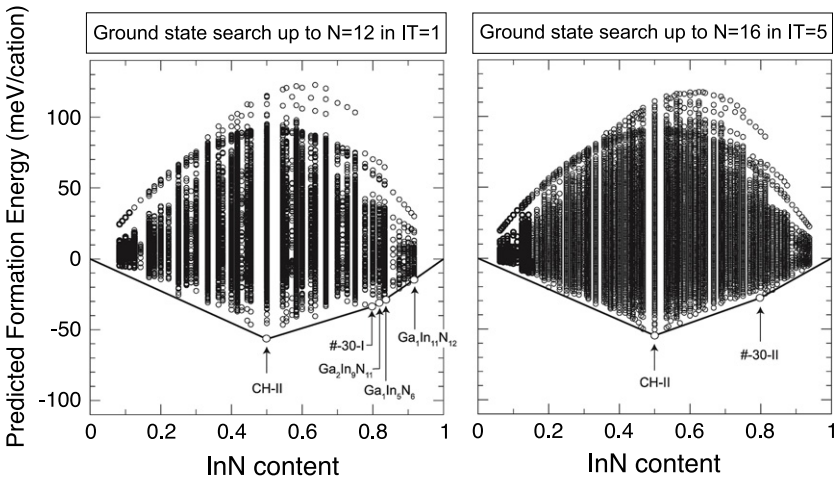

Figure 8. Results for the direct exhaustive enumeration ground state search for the cluster expansions with the lowest CV score in the first iteration (IT = 1) and the last iteration (IT = 5).

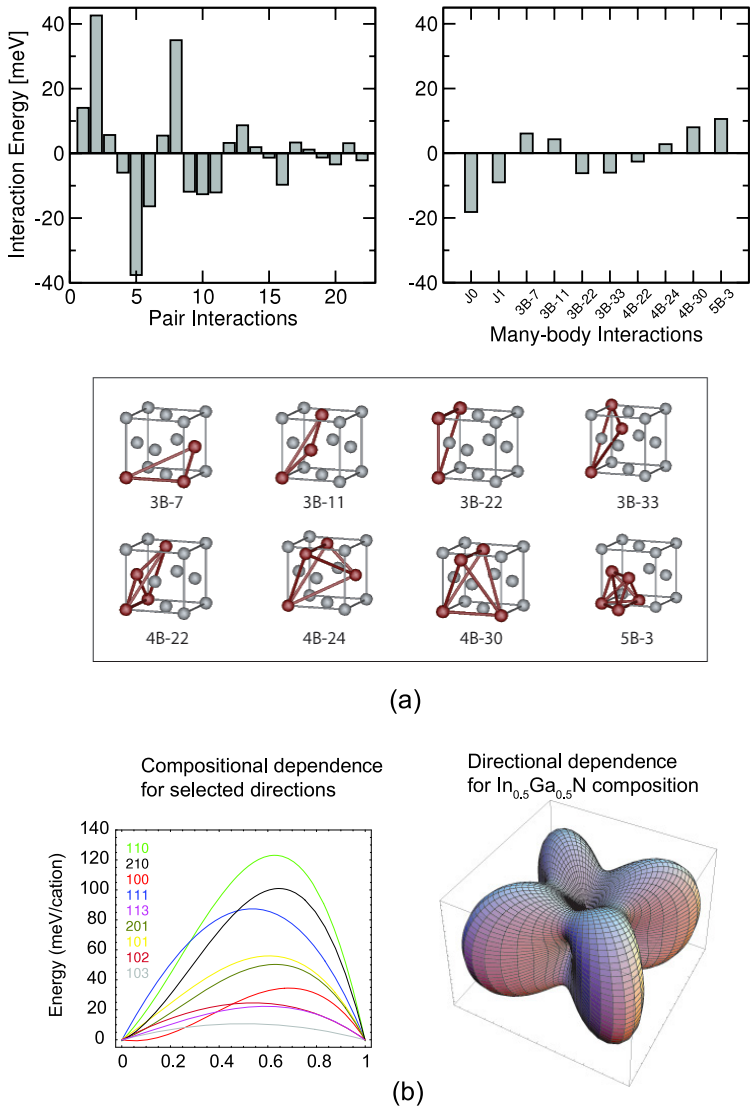

Figure 9. Energy interaction parameters defining our final cluster expansion (CE) for a (Ga, In)N epitaxial alloy on a GaN(001) substrate: (a) CE interaction parameters in real space including pair interactions (top left) and many-body interactions (top right). The many-body figures are shown in the bottom part. (b) Constituent strain (CS) energy. The left part shows the CS energy versus In concentration for some selected directions. The right part shows the spatial dependence of CS energy at In concentration $x = 0.50$.

difference between WZ and ZB structures is small. For GaN, the energy of the WZ structure is ∼12 meV/atom lower than that of the ZB structure [77]. The thermodynamic properties of the (Ga, In)N bulk alloy in WZ and ZB formats are also similar. The miscibility gap temperature calculated by the first-principles method for WZ (Ga, In)N is about 1848 K [78], while in ZB format miscibility gap temperature is 1870 K [59]. Despite these structural and thermodynamic similarities, their electronic properties can be quite different. In ZB (Ga, In)N, the piezo-electric polarization in the (001) growth direction does not exist, while in WZ (Ga, In)N there is a piezo-electric polarization field in the (0001) growth direction. Therefore, it has been suggested recently that the investigation of ZB (Ga, In)N will allow us to understand the precise mechanism of light generation in the alloy [79]. In the rest of this paper, we will take the (Ga, In)N zinc-blende epitaxial alloy grown on a GaN(001) substrate as an example to show the detailed procedures of our method.

3.1. Inner loop: identification of figures and interaction energies

We show an example of the selection of pair figures with the help of the t–λ criteria in figure 5 for the two best cluster expansion candidates in the first and last outer loop iterations. The smooth condition in the fit (equations (9)–(12)) allows one to use a large number of pair interactions without over-fitting. Figure 5 illustrates how the CV score (prediction power) varies with the parameters. From the plot, we find in the first outer loop iteration, the CV score is minimized at $n_{\text{pair}} = 15$ (with $t = 4$, $\lambda = 4$) and in the last iteration, we obtain $n_{\text{pair}} = 22$ (with $t = 6$ and $\lambda = 2$).

The selections of many-body figures in the first and last outer loop iteration are illustrated in figures 6 and 7. Figure 6 shows the selection of many-body figures with the total number fixed as $N_{\text{MB}} = 8$ via the GA. Since the LDA input is limited, generally, there will more than one possible 'family' of many-body figures which yields a low CV, i.e., a set of good cluster expansion candidates differs by a very few figures within a family while it differ by several figures between families. To find different 'families', we adopted the lock-out mechanism [45, 65]. The GA is periodically restarted when a low CV solution persists for too long (in this paper, 100 generations). This solution is then locked out and is forbidden in the remaining run. Such a lock-out and restart procedure is clearly seen in figure 6 (i.e., the abrupt change when a cluster expansion with the lowest CV persists for long time). The GA run is continued until the new locked-out candidates appear with consistently higher CV scores than those already known (typically ∼3000–5000 generations).

Figure 7 depicts the selection of the optimal $N_{\text{MB}}^{\text{opt}}$, i.e., the total number of many-body figures in a cluster expansion. Every cross symbol represents one locked-out candidate in the GA search (figure 6). The CV score first decreases with an increase of N_{MB}, and then increases, or flattens out, where we identify the optimal $N_{\text{MB}}^{\text{opt}}$. In first and last outer loop iteration, optimal $N_{\text{MB}}^{\text{opt}} = 7$ and $N_{\text{MB}}^{\text{opt}} = 8$ respectively. We use the cluster expansions around $N_{\text{MB}}^{\text{opt}}$ as candidates to predict ground state structures.

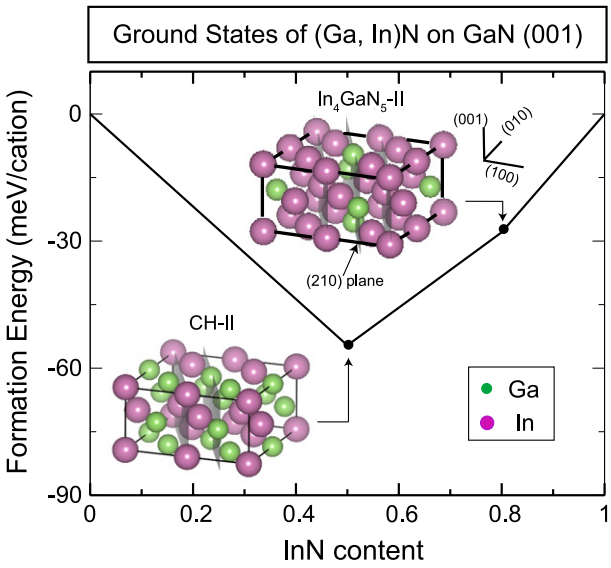

Figure 10. The final ground state structures of a (Ga, In)N zinc-blende alloy on a GaN(001) substrate. Two (210) superlattice structures: $(InN)_2/(GaN)_2$ (CH-II) at $x = 0.50$ and $(InN)_4/(GaN)_1$ at $x = 0.80$ are predicted as the ground states.

3.2. Outer loop: searching ground state structures

Figure 8 shows the global view of $\Delta H_{\text{CE}}^{\text{epi}}(\sigma)$ for a (Ga, In)N epitaxial alloy on a GaN(001) substrate spanning $\sim O(2^{16})$ configurations, which are obtained by an exhaustive enumeration method, using the best cluster expansion candidates in the first (IT = 1) and the last outer loop iteration (IT = 5). The convex hull is drawn as the solid line and the predicted ground structures are also labeled. We repeat the ground state search for several other cluster expansion candidates to overcome the over-fitting situation. For the first iteration, the predictions are summarized in table 3. Since different cluster expansions give different predictions, we list the predicted fraction for every structure predicted by the CE candidates. Also the lowest energy depth is listed. There are eight structures (highlighted in table 3) satisfying the 'importance' criteria set up in section 2.2.2. These structures will be taken as the ground state structures predicted in the first iteration. Since CH-II has been included in the first iteration, the total energies of the remaining seven structures will be calculated with the first-principles method and then their epitaxial formation energies (equation (2)) will be included in the second iteration.

The outer loop is repeated with the LDA inputs of the newly predicted ground states from the previous iteration until the final convergence. The LDA energies added at each outer loop iteration are summarized in table 4. The outer loop is continued until the fifth iteration to reach the termination criteria set up in section 2.2.2. The final CV score is 3.04 meV/cation and the maximum prediction error of all the LDA input structures is 4.19 meV/cation. Statistically, the CV score is regarded as a measure of the prediction power. However, our LDA input is limited and a large portion of the LDA inputs are low energy structures (referring to the criteria adopted to add LDA inputs during the outer loop). It is better

Table 3. First iteration: the ground state predictions of the ten CE candidates analyzed in the first iteration of the outer loop, compared with the LDA data. For each structure σ we indicate how many CE candidates predict it as the ground state (i.e., how often), the lowest prediction depth, and the range of $\Delta H_{\rm CE}^{\rm epi}(\sigma)$ given by the CE candidates. Ground state predictions satisfying our criteria (i)–(iv) are highlighted in bold. In the last two columns, 'n/a' means that no LDA calculations were performed for the indicated structure. Note that while all candidates have low CV score, their ground state predictions are not very consistent, indicating that further iterations are necessary.

Structure			CE predictions				LDA data	
					$\Delta H_{\rm CE}^{\rm epi}$ (meV)			
x	Name	Description	How often (%)	Lowest depth (meV)	Min	Max	$\Delta H_{\rm LDA}^{\rm epi}$ (meV)	Convex hull breaking point?
0.1000	**#1132-II**		30	**−2.06**	−26.72	+0.45	N/a	N/a
0.1250	**#298-II**	A_7B_1	30	**−2.91**	−30.83	+1.09	N/a	N/a
0.2500	**#313-II**	A_6B_2	**40**	**−1.60**	−31.32	−16.61	N/a	N/a
0.3333	#4430-II		10	−1.15	−44.82	−25.95	N/a	N/a
0.5000	**CH-II**	(210) A_2B_2 SL	**90**	**−36.83**	−56.61	−55.19	−57.5345	Yes
	#359-II	A_4B_4	10	**−66.68**	−88.72	−7.54	N/a	N/a
0.8000	#−**30**	(201) AB_4 SL	**60**	**−9.95**	−38.51	−30.78	N/a	N/a
0.8182	#−1873-III		**40**	−0.03	−40.78	−26.24	N/a	N/a
0.8333	#−4764-II		10	−1.44	−41.92	−21.71	N/a	N/a
	#−72-III		**40**	−0.38	−42.63	−22.42	N/a	N/a
0.8750	#−290-II	A_1B_7	10	−1.07	−38.94	−12.91	N/a	N/a
	#−**360-II**		10	**−32.65**	−46.49	−15.43	N/a	N/a
	#−**298-II**		20	**−3.30**	−43.62	−11.78	N/a	N/a

Table 4. Summary of the LDA formation energies $\Delta H_{\rm LDA}^{\rm epi}(\sigma)$ (in meV/cation) of the structures added as input over the iterations of the outer CE loop. Also shown are the range of predicted $\Delta H_{\rm CE}^{\rm epi}(\sigma)$ at the step that caused the structure to be included into the next iteration and the predicted $\Delta H_{\rm CE}^{\rm epi}(\sigma)$ of the last CE.

x	Structure	Description	$\Delta H_{\rm CE}^{\rm epi}(\sigma)$	$\Delta H_{\rm LDA}^{\rm epi}(\sigma)$	Final CE
			Iteration 2		
1/10	A_9B	(310) A_9B SL	−26.72 to 0.45	−2.03810	−1.0
1/8	A_7B	(310) A_7B SL	−30.83 to 1.09	1.077	−0.6
1/4	A_6B_2	(210) BA_2BA_4	−31.32 to −16.61	−26.79	−25.1
1/2	A_4B_4	Non-SL	−88.72 to −7.54	−9.486	−10.3
4/5	AB_4	(201) AB_4 SL	−38.51 to −30.78	−21.354	−19.4
7/8	AB_7	(312) AB_7 SL	−46.69 to −15.43	−4.81	−4.2
7/8	AB_7	(310) AB_7 SL	−43.62 to −11.78	−7.634	−8.7
			Iteration 3		
1/5	A_4B	(210) A_4B SL	−25.79 to −19.88	−21.07	−20.61
1/3	A_8B_4	(410) $(BA_3)_2B_2A_2$ SL	−43.03 to −35.30	−26.58	−28.00
2/5	A_3B_2	(210) A_3B_2 SL	−48.24 to −41.21	−34.3	−34.19
2/3	A_4B_8	(410) $(AB_3)_2A_2B_2$ SL	−49.51 to −41.56	−34.55	−35.02
4/5	AB_4	(210) AB_4 SL	−38.73 to −26.92	−27.92	−27.93
			Iteration 4		
7/15	A_8B_7	(102) $(B_2A_2)_3BA_2$ SL	−56.07 to −52.90	−48.10	−49.92
8/15	A_7B_8	(102) $(A_2B_2)_3AB_2$ SL	−56.52 to −53.01	−46.77	−49.18
			Iteration 5		
1/4	A_6B_2	(102) BA_2BA_4 SL	−30.02 to −26.34	−22.28	−24.46
9/13	A_4B_9	(123) $(AB_2)_3AB_3$ SL	−41.23 to −38.31	−25.92	−30.12

Table 5. Percentage of the cluster expansions correctly predicting the final ground state structures in each outer loop iteration.

Composition	Structure	$N = 46$ IT = 1	$N = 53$ IT = 2	$N = 58$ IT = 3	$N = 60$ IT = 4	$N = 62$ IT = 5
$In_2Ga_2N_4$	(210) A_2B_2 SL	90%	100%	83%	100%	100%
In_4GaN_5	(210) AB_4 SL	0%	91%	91%	62%	100%

to test the prediction of higher energy structures away from the convex hull. We did LDA calculations for 10 randomly selected high energy structures. The average prediction error is 3.87 meV/cation and the biggest error is 9.31 meV/cation. The details of the final cluster expansion will be discussed in next section. Table 5 shows the prediction frequency of the

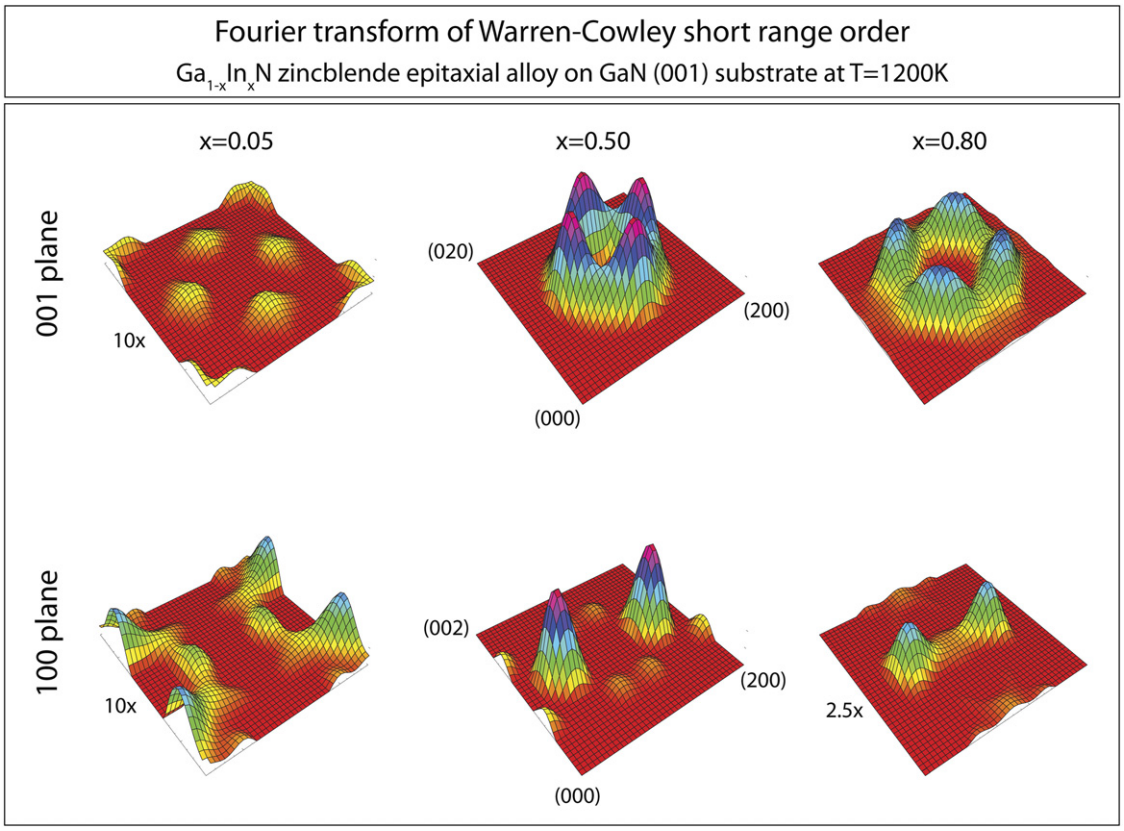

Figure 11. Fourier transform of the Warren–Cowley short-range-order parameter of the (Ga, In)N solid solution at different concentrations at $T = 1200$ K. The peaks at $(1\frac{1}{2}0)$ points indicate ordering along the (210) crystal direction. Some figures are scaled to aid visualization and the scaling factors are labeled.

two identified ground state structures in the history of outer loop iteration. From tables 2, 4 and 5, we can find the ground state structure $(InN)_2/(GaN)_2$ (CH-II) is included in the initial LDA input and always keeps its high prediction frequency throughout the outer loop iterations. The $(InN)_4/(GaN)_1$ is not included in the initial input, and even not predicted in the first iteration. Its prediction appears in the second iteration and its LDA input is thus included in the third iteration. This case justifies the necessity of the outer loop, i.e., searching for ground states via the calculations of the LDA energy of some intuitively selected structures might miss the real ground state.

4. Ensuing physics of the (Ga, In)N epitaxial alloy on a GaN(001) substrate

The interaction energies J and the ground state structures identified in the final cluster expansion are depicted in figures 9 and 10. Figure 9(a) shows the numerical values of pair and many-body interaction energies and the geometry of the selected many-body figures in real space. Figure 9(b) shows the constituent strain energy defined in equation (4). It is worth noting that the pair and many-body interactions only determine $\Delta\tilde{H}^{\rm epi}_{\rm CE}(\sigma)$, so that the constituent strain energy should be added back to get $\Delta H^{\rm epi}_{\rm CE}(\sigma)$. The left panel of figure 9 shows the concentration dependence of the $\Delta E_{\rm CS}(x, \hat{\mathbf{k}})$ for some principal directions. The right panel shows the spatial dependence of the constituent strain energy $\Delta E_{\rm CS}(x, \hat{\mathbf{k}})$ at $x = 0.50$. It is clear that $\Delta E_{\rm CS}(x, \hat{\mathbf{k}})$ shares the tetragonal symmetry of the underlying lattice (i.e., an elongated fcc lattice along the (001) direction).

Two (210) superlattice structures are identified as the ground states and illustrated in figure 10: $(InN)_2/(GaN)_2$ (CH-II) and $(InN)_4/(GaN)_1$. As we discussed previously, the existence of a substrate removes the symmetry degeneracy of a structure on the fcc lattice. The counterparts of these two structures, but along the (102) and (021) (i.e., variant-I and variant-III, respectively) crystal directions, are not ground states. It is very interesting to note that for most of the III–V zinc-blende bulk semiconductor alloys (with non-negligible lattice mismatch), it has been observed that the (210) superlattice structures are the best to accommodate strain from the lattice mismatch between the two constituents in comparison with all the other structures at the same concentration [56, 58, 80, 81]. We find that the same pattern persists in the (Ga, In)N epitaxial alloy on a GaN(001) substrate.

With the obtained cluster expansion, we calculated the Warren–Cowley SRO in the reciprocal space at $T = 1200$ K at different concentrations $x = 0.05, 0.50$ and 0.80 (figure 11). The strong peaks at $(1\frac{1}{2}0)$ indicate the ordering tendencies of the two identified (210) superlattice ground state structures. At low concentration $x = 0.05$, the non-zero Γ point is a sign of GaN clustering.

5. Summary

In this paper, we developed a cluster expansion method to describe the thermodynamic properties of a substrate-coherent epitaxial alloy and apply it to a (In, Ga)N zinc-blende alloy grown on a GaN(001) substrate. Our developed cluster expansion approach is fully capable of considering the three new features of epitaxial alloys in comparison with the bulk alloys: (1) more than one distinct epitaxial structure σ_{epi} from one bulk configuration σ_{bulk}, (2) splitting of the figures, and (3) the constituent strain energy calculation in the presence of a substrate. The underlying lattice of the epitaxial alloy is selected as the elongated/compressed lattice of its bulk alloy along the substrate directions $\hat{\mathbf{Q}}$. Such a selection will make the cluster expansion (equations (3) and (6)) naturally distinguish the distinct epitaxial structures from one atomic bulk configuration. The algorithm for the constituent strain energy calculation described in the appendix considers the lattice coherence at both the $(AC)_m/(BC)_n$ interface $\hat{\mathbf{k}}$ and the substrate–film interface $\hat{\mathbf{Q}}$.

We take the (In, Ga)N epitaxial alloy grown on a GaN(001) substrate as an example to demonstrate this method. We follow the strategy developed in [45, 48] to determine the parameters defining our epitaxial cluster expansion from the first-principles calculations. The fitting consists of two loops. The inner loop selects the best combination of figures with the best prediction power based on a set of $\{\Delta H^{\text{epi}}_{\text{LDA}}(\sigma)\}$. The outer loop searches the ground state structures from the cluster expansion determined in the inner loop and adds the newly predicted ground state structures into the next iteration. This procedure is repeated until our cluster expansion predictions agree with the ground states calculated by the first-principles method.

Using the obtained cluster expansion, two ground state structures are identified: the $(InN)_2/(GaN)_2$ (210) superlattice (CH) and the $(InN)_4/(GaN)_1$ (210) superlattice. Our results show that the presence of a substrate indeed radically alters the thermodynamics of the (In, Ga)N alloy, converting it from a phase separation to an ordered system. To facilitate the detection of the ordering tendency in this substrate-coherent epitaxial alloy, we also calculated the Warren–Cowley short-range order parameter of the disordered solid solution above the phase transition temperature. The strong ordering peaks are consistent with the superlattice directions of the identified ground states.

Acknowledgments

This research was funded by the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering, under contract No. DE-AC36-99GO10337 to NREL, and the LDRD project from NREL.

Appendix

The cluster expansion approach has been successfully employed to describe the total energy (as a function of atomic configurations) of binary or quasi-binary alloys. In many cases, it is found that keeping only a set of finite size figures (in three-dimensional real space) in equations (3) and (6) can yield good accuracy. This is consistent to the intuition that the inter-atom chemical and magnetic interactions degrade rapidly with increasing atomic distances. However, for the inherent long-range interactions, such as the elastic energy due to the size mismatch of the two alloying constituents AC and BC, such a finite-ranged cluster expansion fails. A well-known case is the energy of a long period superlattice structure $(AC)_m/(BC)_n$ ($\frac{m}{m+n} = x$ and $m, n = \infty$). The superlattice energy is the combined epitaxial strain energy of both the AC and BC region, both of which are distorted to fit a common lattice constant $a_\perp$ at the interface $\hat{k}$, but are free to relax in the perpendicular direction. A finite-ranged CE is completely incapable of capturing such a long period superlattice energy. The reason for this failure is simple: the CE sees all A atoms that are far from interface as if they were in bulk AC crystal, since the figures of the finite-ranged CE connect them exclusively to other A atoms. In the same way, the CE treat almost all the B atoms as in bulk BC. Therefore, the final CE prediction for formation energy ΔH is zero (the formation energy of bulk AC and BC). A mixed-basis cluster expansion (equation (3)) is thus introduced to remedy the failure of such a finite-ranged CE, in which the pair interactions are expanded in reciprocal space (for details, please refer to [44]) to take the elastic energy into consideration. The key quantity in equation (3) is the so called constituent strain energy $\Delta E_{\text{CS}}(\hat{\mathbf{k}}, x)$, which is defined as the strain energy of the long period limit superlattice $(AC)_m/(BC)_n$ ($x = \frac{n}{m+n}$) along $\hat{\mathbf{k}}$.

Calculating the constituent strain energy $\Delta E_{\text{CS}}(\hat{k}, x)$ is crucial in the mixed-basis cluster expansion. In valence force field (VFF) model calculations, the strain energy of the superlattice $(AC)_m/(BC)_n$ converges fairly slowly with respect to periodicity m and n, often up to 50–100. Direct large supercell calculations to obtain the CS energy $\Delta E_{\text{CS}}(\hat{k}, x)$ are, therefore, not feasible, especially for first-principles calculations. In fact, the AC and BC region inside the superlattice are homogeneously deformed from the two constituents AC and BC at their own equilibrium lattice constant. The deformations inside AC and BC are different, but they are correlated in such a way that the AC and BC slabs form a lattice coherent superlattice $(AC)_m/(BC)_n$ in the crystal direction $\hat{\mathbf{k}}$. If we can derive such correlated deformations, the superlattice strain energy is just the product of the volume with the strain energy of the deformed unit cells of AC and BC. For the bulk fcc and bcc alloys, such correlated deformations have been obtained in [44], where the AC and BC unit cell are biaxially deformed to share the same lattice constant $a_\perp$ at interface $\hat{\mathbf{k}}$ and a free relaxation is allowed in the perpendicular directions (to $\hat{\mathbf{k}}$). In the epitaxial alloy, substrate coherence requires the superlattice $(AC)_m/(BC)_n$ to have the unit cell lattice constant as substrate a_{sub} on the substrate plane $\hat{\mathbf{Q}}$, whereas atomic relaxation is allowed inside the unit cell. Thus, in addition to forming a lattice coherent at $\hat{\mathbf{k}}$, the correlated deformation of AC and BC should also consider the constraint from the substrate. This appendix will address this issue.

A.1. Derivation for the (001) substrate

In this section, we will restrict our derivation to the case of the (001) substrate. The extension to other substrates will be

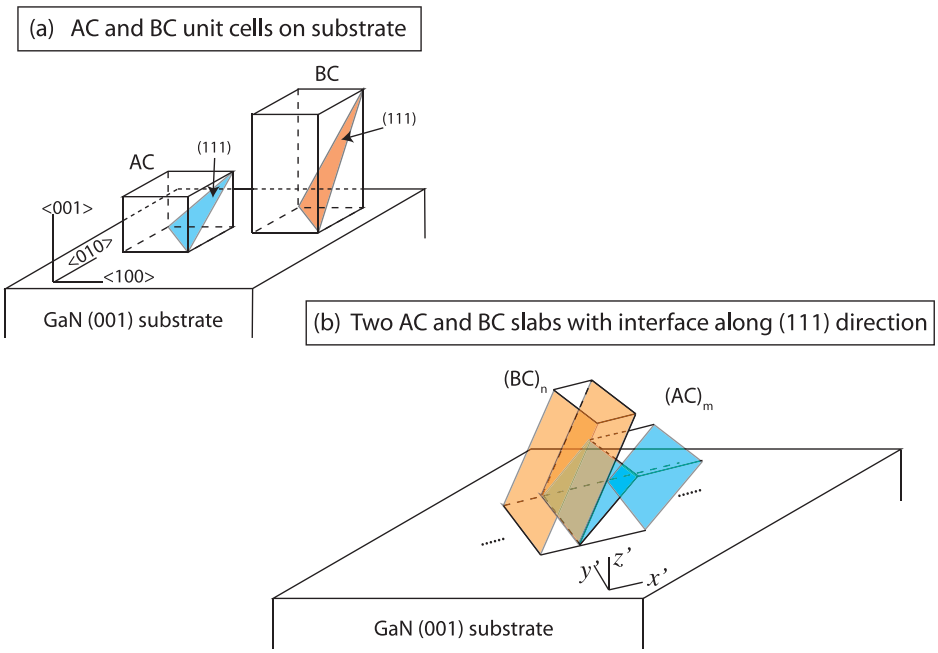

Figure A.1. (a) Sketch of the two constituents AC and BC grown on the substrate. The lattice constants on the substrate plane are fixed as the substrate lattice constant, whereas a structural relaxation in the perpendicular direction is allowed. Consequently, crystal planes with the same index in these two constituents do not match each other. (b) The two as-cut slabs of $(AC)_n$ and $(BC)_m$. The energy cost (per atom) to obtain lattice coherence at the interface is defined as the constituent strain energy when n and m approach infinity.

discussed at the end of this appendix. The starting point is the AC and BC deposited on a (001) substrate, respectively, as illustrated in figure A.1(a), where a relaxation in the $\hat{\mathbf{Q}}$ direction is allowed so that the c/a ratio of the AC and BC unit cells are generally different. The same index (l, m, n) interface of AC and BC will then have different lattice constants in the plane and different angles to the substrate (as the (111) interface shown in figure A.1(a)). The sketch of the two as-cut $(AC)_m$ and $(BC)_n$ slabs with the (l, m, n) interface are shown in figure A.1(b), where they extend infinitely along the y' direction and are periodically repeated in the x' direction. Now we need to deform these two slabs separately to form a lattice coherence at both the $\hat{\mathbf{k}}$ and $\hat{\mathbf{Q}}$ interfaces as explained earlier.

Deformation of a solid can be described by a deformation tensor $\mathbf{F}$. For a infinitesimal vector $\mathrm{d}\mathbf{x}$ under deformation $\mathbf{F}$, it becomes $\mathrm{d}\mathbf{x}' = \mathbf{F} \cdot \mathrm{d}\mathbf{x}$. There are nine independent elements in this tensor. For the $(AC)_m$ and $(BC)_n$ slabs, there will generally be 18 independent elements. To describe the deformation more easily, we build up two Cartesian coordinate systems. The global xyz coordinates are built up with the z axis along the substrate plane direction (001) and the x and y axes along the cubic crystal direction (100) and (010). The local coordinates $x'y'z'$ are set up with the y' axis along the intersection line of the interface $\hat{\mathbf{k}}$ and $\hat{\mathbf{Q}}$ and the z' axis along the substrate plane direction (001). The elements of the tensor $\mathbf{F}$ in the global and local coordinates will be expressed by symbols F_{ij} and f_{ij}, respectively.

In the local coordinates $x'y'z'$, the deformation along the y' direction is fixed, i.e., $f_{22} = 0$ due to the constraint from the substrate. When the superlattice orientation $\hat{\mathbf{k}}$ is along a high symmetry direction, such as (110), (111) or (100) on the (001) substrate, the $x'z'$ plane becomes a mirror plane, and thus the $f_{12} = f_{21} = f_{23} = f_{32} = 0$. Only the deformations inside the $x'z'$ plane (f_{11}, f_{33}, f_{13} and f_{31}) are non-zero. This drastically reduces the complexity of the deformation. Here we *assume* $f_{12} = f_{21} = f_{23} = f_{32} = 0$ hold for any other orientations (lmn). We will show later that this 'plane-deformation approximation' only introduces a small error (e.g., 2 meV/cation).

To form lattice coherence at the interface $\hat{\mathbf{k}}$ in the presence of a substrate, we deform the $(AC)_m$ and $(BC)_n$ slabs in three steps as schematically shown in figure A.2. The first step is to match the interfaces $\hat{\mathbf{k}}$ by uniformly stretching/depressing AC/BC in the z' direction, i.e., $f_{33} \neq 1$ (equation (16)). The second step is to relax the structure by moving the interface k (i.e., $f_{11} \neq 1$) and preserving the unit cell vector length (i.e., $|\vec{ab}|$ unchanged, due to the constraint from the substrate) (equation (17)). The third step is to relax the structure by shearing (i.e., $f_{13} \neq 0$ and $f_{31} \neq 0$). We also need to remove the rigid rotation along the y' axis because such a rotation does not make a contribution to the strain energy (equation (18)).

$$\mathbf{F}^{\mathrm{A}} \cdot (\vec{Oh}_{\mathrm{A}}) = \mathbf{F}^{\mathrm{B}} \cdot (\vec{Oh}_{\mathrm{B}}) \tag{16}$$

$$|\mathbf{F}^{\mathrm{A}} \cdot (\vec{Oa}) \cdot \vec{e}'_x - \mathbf{F}^{\mathrm{B}} \cdot (\vec{Ob}) \cdot \vec{e}'_x| = |\vec{ab}| \tag{17}$$

$$\mathbf{F}^{\mathrm{A}} \cdot (\vec{Oa}) \cdot \vec{e}'_z = \mathbf{F}^{\mathrm{B}} \cdot (\vec{Ob}) \cdot \vec{e}'_z \tag{18}$$

In terms of the deformation tensor, we have

$$f^{\mathrm{B}}_{33} = \frac{-f^{\mathrm{A}}_{31} \cos\phi + x c_{\mathrm{A}} f^{\mathrm{A}}_{33} \sin\phi}{x c_{\mathrm{B}} \sin\phi} \tag{19}$$

$$f^{\mathrm{B}}_{11} = \frac{1 - (1 - x) f^{\mathrm{A}}_{11}}{x} \tag{20}$$

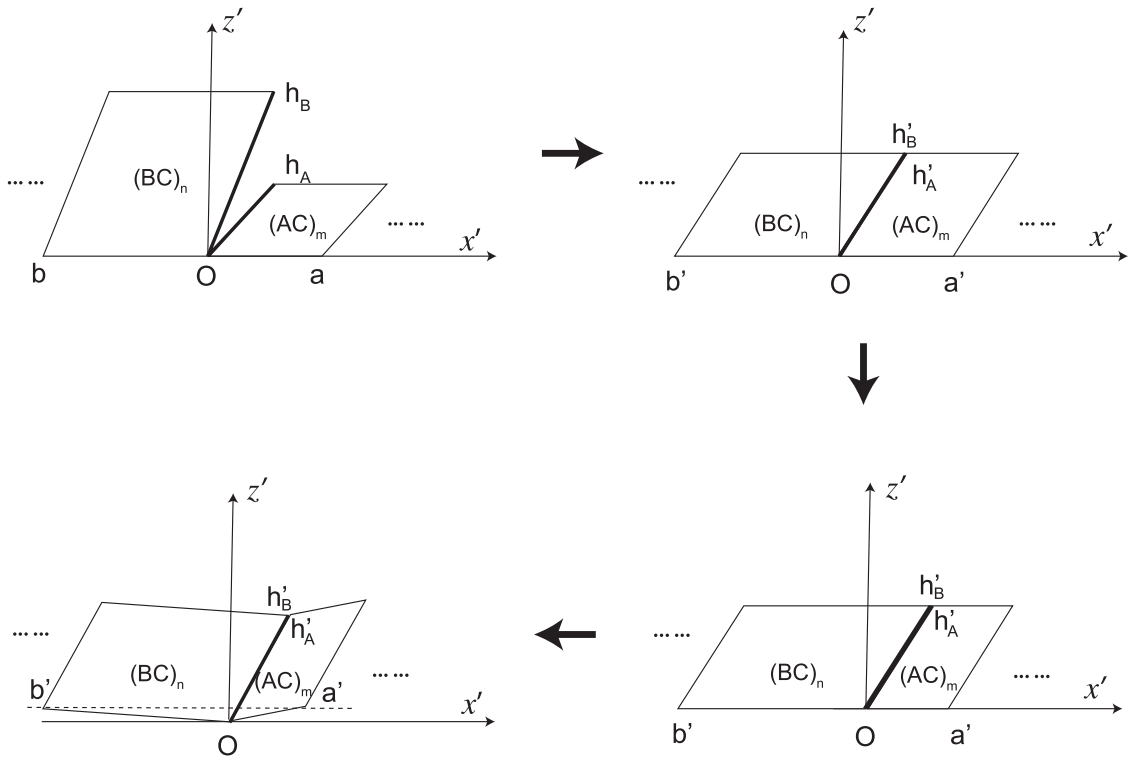

Figure A.2. The three steps of the 'plane-deformation' of the two slabs to obtain the lattice coherence at the interface $\hat{\mathbf{k}} = (lmn)$ and at the substrate plane $\hat{\mathbf{Q}}$. Cartesian coordinates $x'y'z'$ represent the local coordinates $x'y'z'$ in figure A.1. The thick bold lines represent the interface (lmn) of each slab. Plane $x'y'$ represents the substrate plane $\hat{\mathbf{Q}}$.

$$f_{13}^{\mathrm{B}} = \frac{(1 - f_{11}^{\mathrm{A}})\cos\phi + xc_{\mathrm{A}} f_{13}^{\mathrm{A}} \sin\phi}{xc_{\mathrm{B}} \sin\phi} \tag{21}$$

$$f_{31}^{\mathrm{B}} = -\frac{(1-x)}{x} f_{31}^{\mathrm{A}} \tag{22}$$

where c_{A} and c_{B} are the c/a ratio of the AC and BC unit cells respectively and $x = n/(m+n)$. ϕ and the θ are the Euler angles of the interface (lmn) in the global xyz coordinates: $\phi = \arccos(\frac{l}{\sqrt{l^2+m^2}})$ and $\theta = \arccos(\frac{n}{\sqrt{l^2+m^2+n^2}})$. In the global coordinates xyz, the deformation tensors can be expressed as:

$$\mathbf{F}^{\mathrm{A}} = \begin{bmatrix} \frac{1}{2}[1 + f_{11}^{\mathrm{A}} + (f_{11}^{\mathrm{A}} - 1)\cos 2\theta] & \frac{1}{2}(f_{11}^{\mathrm{A}} - 1)\sin 2\theta & f_{13}^{\mathrm{A}}\cos\theta \\ \frac{1}{2}(f_{11}^{\mathrm{A}} - 1)\sin 2\theta & \frac{1}{2}[1 + f_{11}^{\mathrm{A}} + (1 - f_{11}^{\mathrm{A}})\cos 2\theta] & f_{13}^{\mathrm{A}}\sin\theta \\ f_{31}^{\mathrm{A}}\cos\theta & f_{31}^{\mathrm{A}}\sin\theta & f_{33}^{\mathrm{A}} \end{bmatrix} \tag{23}$$

$$\mathbf{F}^{\mathrm{B}} = \begin{bmatrix} 1 + \frac{1-x}{x}(1 - f_{11}^{\mathrm{A}})\cos^2\theta & \frac{1-x}{2x}(1 - f_{11}^{\mathrm{A}})\sin 2\theta & \left(\frac{(1-f_{11}^{\mathrm{A}})\cos\phi + xc_{\mathrm{A}} f_{13}^{\mathrm{A}}\sin\phi}{xc_{\mathrm{B}}\sin\phi}\right)\cos\theta \\ \frac{1-x}{2x}(1 - f_{11}^{\mathrm{A}})\sin 2\theta & 1 + \frac{1-x}{x}(1 - f_{11}^{\mathrm{A}})\sin^2\theta & \left(\frac{(1-f_{11}^{\mathrm{A}})\cos\phi + xc_{\mathrm{A}} f_{13}^{\mathrm{A}}\sin\phi}{xc_{\mathrm{B}}\sin\phi}\right)\sin\theta \\ -\frac{1-x}{x} f_{31}^{\mathrm{A}}\cos\theta & -\frac{1-x}{x} f_{31}^{\mathrm{A}}\sin\theta & \frac{xc_{\mathrm{A}} f_{33}^{\mathrm{A}}\sin\phi - f_{31}^{\mathrm{A}}\cos\phi}{xc_{\mathrm{B}}\sin\phi} \end{bmatrix}. \tag{24}$$

Under such deformations, the $(\mathrm{AC})_m$ and $(\mathrm{BC})_n$ slabs will form lattice coherence at interfaces $\hat{\mathbf{k}}$ and $\hat{\mathbf{Q}}$. There are only four free elements in $\mathbf{F}^{\mathrm{A}}$ and $\mathbf{F}^{\mathrm{B}}$, i.e., f_{11}^{A}, f_{33}^{A}, f_{13}^{A} and f_{31}^{A}.

Strain energies can be calculated for the AC and BC unit cells with deformation $\mathbf{F}^{\mathrm{A}}$ and $\mathbf{F}^{\mathrm{B}}$ in equation (23) with a total energy functional, such as, VFF, the first-principles method etc. The constituent strain energy is the minimal value of the concentration weighted average of the strain energies of the AC and BC unit cells with respect to the four free elements of the deformation tensor. In this paper, we obtain the CS energy by exploring a discrete grid mesh in the parameter space composed of the four free deformation tensor elements.

The directly calculated energies of the superlattice $(\mathrm{InN})_n/(\mathrm{GaN})_n$ ($x = 0.50$) epitaxially grown on the GaN(001) substrate versus the periodicity n are carried out with the VFF model and compared with the CS energy calculated with the algorithm described in this appendix. The results are shown in figure A.3. At the asymptotic limit of n, the directly calculated superlattice energies approach the calculated CS energy. There is about 2–3 meV/cation difference. It is worth noting that, in the VFF calculations, the structure is fully relaxed except the fixed lattice constant at the substrate interface. This suggests that the plane-deformation approximation we made for interfaces $\hat{k}$ other than (111), (100), (010), etc introduces only minor errors.

A.2. Extension to other substrates

The derived equations (23) and (24) can be applied to other epitaxial alloy systems, including different lattice types (such as bcc, simple cubic (sc), etc) and different substrates (such as (101), (201), (111), etc). To use the results, one needs to

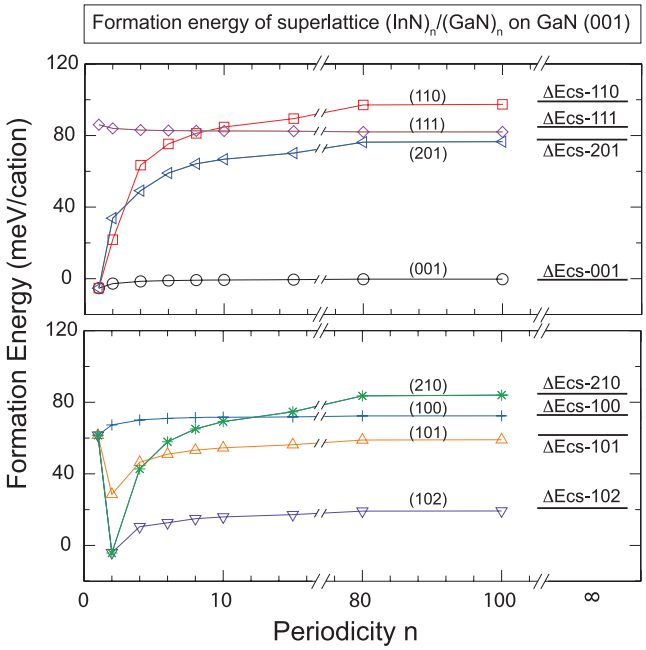

Figure A.3. Formation energy of the superlattice structure $(InN)_n/(GaN)_n$ grown epitaxially on a GaN(001) substrate directly calculated by the VFF model versus the periodicity n. The constituent strain energies $\Delta E_{CS}(\hat{\mathbf{k}}, x)$ calculated by the algorithm (equations (23) and (24)) described in the appendix, using the VFF model as the energy functional, are plotted as the solid lines on the right.

construct the unit cells of AC and BC on the substrates and relax them in the substrate direction $\hat{\mathbf{Q}}$. However, one should be aware of the approximation adopted in the derivation of equations (23) and (24): the deformations are only in the $x'z'$ plane. This assumption is strictly true only for an interface $\hat{\mathbf{k}}$ with mirror symmetry about the $x'z'$ plane (in other words, $\vec{y}' = \vec{Q} \times \vec{k}$ is a mirror plane). For fcc, bcc and sc alloys grown on the (001) substrate, these $(AC)_m/(BC)_n$ superlattice directions include (100), (010), (110), (102), (103), (113) and so on. For a $\vec{Q}$ = (101) substrate, the directions are (001), (111), (112), (113), (100), (010), (210) and so on. For a (201) substrate, they are (100), (001), (101), (102), (103) and so on. For other $\vec{\mathbf{k}}$ directions, to check the accuracy, comparison with the direct long period superlattice calculations should be carried out, as we have done in figure A.3.

Here we would like to provide some further discussions of the numerical errors introduced by the plane-deformation approximation. It is obvious that further relaxation steps out of the $x'z'$ plane would occur if the $x'z'$ plane does not have a mirror symmetry. The magnitude of such an additional relaxation strain should be proportional to the lattice mismatch of the two constituents AC and BC. For the very large lattice mismatch alloy InGaN (e.g., ∼11%) studied in this paper (figure A.3), it introduced a 2–3 meV/cation error for the calculated CS energy in the $\vec{k}$ = (210) and (311) directions. This suggests that the relaxation strain out of the $x'z'$ plane is indeed very small for such a large lattice mismatched alloy. It is known that the elastic strain energy is a parabolic function of deformation strain. The 'plane-deformation-approximation' should, therefore, work very well for other chemical systems with a similar or smaller lattice mismatch, such as (InGa)As, Ga(AsSb), Cd(SeTe), etc.

References

[1] *Landolt-Börnstein—Group III Condensed Matter* 2002 vol 41A1b (Berlin: Spinger)
[2] Tsao J Y 1993 *Materials Fundamentals of Molecular Beam Epitaxy* (San Diego, CA: Academic)
Farrow R F C 1995 *Molecular Beam Epitaxy: Applications to Key Materials* (Norwich, NY: William Andrew)
[3] Stringfellow G-B 1999 *Organometallic Vapor-Phase Epitaxy: Theory and Practice* (San Diego, CA: Academic)
[4] Stringfellow G B 1974 *J. Cryst. Growth* **27** 21
[5] Zunger A and Mahajan S 1992 *Handbook on Semiconductors* vol 3b, ed T S Moss (Amsterdam: North-Holland) p 1399
[6] Zunger A 1994 *Handbook of Crystal Growth* vol 3, ed D T J Hurle (Amsterdam: Elsevier) p 999
[7] Wood D M and Zunger A 1988 *Phys. Rev. Lett.* **61** 1501
[8] Wood D M and Zunger A 1989 *Phys. Rev.* B **40** 4062
[9] Zunger A and Wood D M 1989 *J. Cryst. Growth* **98** 1
[10] Kelires P C and Tersoff J 1989 *Phys. Rev. Lett.* **63** 1164
[11] Bernard J E, Froyen S and Zunger A 1991 *Phys. Rev.* B **44** 11178
[12] Gomyo A, Makita K, Hino I and Suzuki T 1994 *Phys. Rev. Lett.* **72** 673
[13] Tersoff J 1995 *Phys. Rev. Lett.* **74** 434
[14] Drautz R, Reichert H, Fahnle M, Dosch H and Sanchez J M 2001 *Phys. Rev. Lett.* **87** 236102
[15] Diehl H W 1998 *Phase Transitions and Critical Phenomena* vol 10, ed C Domb and J L Lebowitz (London: Academic)
[16] Dosch H 1992 *Critical Phenomena at Surface and Interfaces* (*Springer Tracts in Modern Physics* vol 126) (Berlin: Springer)
[17] Muller S 2003 *J. Phys.: Condens. Matter* **15** R1429
[18] Muller S 2006 *Surf. Interface Anal.* **38** 1158
[19] Reuter K, Stampfl C and Scheffler M 2005 *Handbook of Materials Modelling, Part A. Methods* ed S Yip (Berlin: Springer) p 149
[20] http://www.philipslumileds.com/products/luxeon/luxeonrebel
[21] Nakamura S *et al* 1998 *Appl. Phys. Lett.* **72** 211
[22] Geisz J F *et al* 2007 *Appl. Phys. Lett.* **91** 023502
[23] Stringfellow G B 1972 *J. Appl. Phys.* **43** 3455
Stringfellow G B 1982 *J. Electron. Mater.* **11** 903
[24] DeCremoux B 1982 *J. Physique Coll.* **43** C5–19
[25] Glas F 1987 *J. Appl. Phys.* **62** 3201
[26] Johnson W C and Chiang C S 1988 *J. Appl. Phys.* **64** 1155
Larche F C, Johnson W C, Chiang C S and Martin G 1988 *J. Appl. Phys.* **64** 5251
[27] Karpov S Y 1998 *Int. J. Nitride Semicond.* **3** 1
[28] Teles L K, Ferreira L G, Scolfaro L M R and Leite J R 2004 *Phys. Rev.* B **69** 245317
[29] Shurtleff J K, Lee R T, Fetzer C M and Stringfellow G B 1999 *Appl. Phys. Lett.* **75** 1914
[30] Gomyo A, Suzuki T, Kobayashi K, Kawata S, Hino I and Yuasa T 1987 *Appl. Phys. Lett.* **50** 673
[31] Gomyo A, Suzuki T and Iijima S 1988 *Phys. Rev. Lett.* **60** 2645
[32] Panish M B and Ilegems M 1972 *Progress in Solid State Chemistry* (New York: Pergamon) p 39
[33] Dandrea R G, Bernard J E, Wei S-H and Zunger A 1990 *Phys. Rev. Lett.* **64** 36
[34] Ferreira L G, Wei S-H and Zunger A 1989 *Phys. Rev.* B **40** 3197
[35] Ho I and Stringfellow G B 1996 *Appl. Phys. Lett.* **69** 2701
[36] Stringfellow G B 1972 *J. Appl. Phys.* **43** 3455
[37] Gan C K, Feng Y P and Srolovitz D J 2006 *Phys. Rev.* B **73** 235214

[38] Zunger A 1994 *Statistics and Dynamics of Alloy Phase Transformations* ed P Turchi and A Gonis (New York: Plenum) p 361
[39] Quillec M, Launois H and Joncour M C 1983 *J. Vac. Sci. Technol.* B **1** 238
[40] Stringfellow G B 1982 *J. Electron. Mater.* **11** 903
[41] Hirth J P and Stringfellow G B 1977 *J. Appl. Phys.* **48** 1813
[42] Quillec M, Daguet C, Benchimol J L and Launois H 1982 *Appl. Phys. Lett.* **40** 325
[43] Ishikawa M and Ito R 1984 *Japan. J. Appl. Phys.* **23** L21
[44] Laks D B, Ferreira L G, Froyen S and Zunger A 1992 *Phys. Rev.* B **46** 12587
[45] Hart G L W, Blum V, Waloski M J and Zunger A 2005 *Nat. Mater.* **4** 391
[46] Blum V and Zunger A 2004 *Phys. Rev.* B **70** 155108
[47] Muller S and Zunger A 2001 *Phys. Rev. Lett.* **87** 165502
[48] Barabash S V, Blum V, Muller S and Zunger A 2006 *Phys. Rev.* B **74** 035108
[49] Sanati M, Wang L G and Zunger A 2003 *Phys. Rev. Lett.* **90** 045502
[50] Wei S-H, Mbaye A A, Ferreira L G and Zunger A 1987 *Phys. Rev.* B **36** 4163
[51] Ozolins V, Wolverton C and Zunger A 1998 *Phys. Rev.* B **57** 6427
[52] Ozolins V, Wolverton C and Zunger A 1998 *Phys. Rev.* B **57** 4816
[53] Wolverton C, Ozolins V and Zunger A 2000 *J. Phys.: Condens. Matter* **12** 2749
[54] Wolverton C, Ozolins V and Zunger A 1998 *Phys. Rev.* B **57** 4332
[55] Muller S, Wang L-W, Zunger A and Wolverton C 1999 *Phys. Rev.* B **60** 16448
[56] Lu Z-W, Laks D B, Wei S-H and Zunger A 1994 *Phys. Rev.* B **50** 6642
[57] Wolverton C and Zunger A 1995 *Phys. Rev. Lett.* **75** 3162
[58] Liu J Z, Trimarchi G and Zunger A 2007 *Phys. Rev. Lett.* **99** 145501
[59] Liu J Z and Zunger A 2008 *Phys. Rev.* B **77** 205201
[60] Sanchez J M, Ducastelle F and Gratias D 1984 *Physica* A **128** 334
[61] Ferreira L G, Wei S-H and Zunger A 1991 *Int. J. Supercomput. Appl.* **5** 34
[62] van de Walle A and Ceder G 2002 *J. Phase Equilib.* **23** 348
[63] Zarkevich N A and Johnson D D 2004 *Phys. Rev. Lett.* **92** 255702
[64] de Gironcoli S and Baroni S 1992 *Phys. Rev. Lett.* **69** 1959
[65] Blum V, Hart G L W, Waloski M J W and Zunger A 2005 *Phys. Rev.* B **72** 165113
[66] Shao J 1993 *J. Am. Stat. Assoc.* **88** 486
[67] Wells A F 1984 *Inorganic Chemistry* 5th edn (Oxford: Clarendon)
[68] Ducastelle F 1991 *Order and Phase Stability in Alloys* (Amsterdam: North-Holland)
[69] Perdew J P and Zunger A 1981 *Phys. Rev.* B **23** 5048
[70] Kresse G and Furthmüller J 1996 *Comput. Mater. Sci.* **6** 15
[71] Nakamura S 1998 *Science* **281** 956
[72] Chichibu S F *et al* 2006 *Nat. Mater.* **5** 810
[73] Abernathy C R, MacKenzie J D, Bharatan S R, Jones K S and Pearton S J 1995 *Appl. Phys. Lett.* **66** 1632
[74] Mullhauser J R, Jenichen B, Wassermeier M, Brandt O and Ploog K H 1997 *Appl. Phys. Lett.* **71** 909
[75] Tabata A, Leite J R, Lima A P, Silveira E, Lemos V, Frey T, As D J, Schikora D and Lischka K 1999 *Appl. Phys. Lett.* **75** 1095
[76] Lemos V, Silveira E, Leite J R, Tabata A, Trentin R, Scolfaro L M R, Frey T, As D J, Schikora D and Lischka K 2000 *Phys. Rev. Lett.* **84** 3666
[77] Yeh C-Y, Lu Z W, Froyen S and Zunger A 1992 *Phys. Rev.* B **45** 12130
[78] Burton B P, van de Walle A and Kattner U 2006 *J. Appl. Phys.* **100** 113528
[79] Chichibu S F *et al* 1999 *Mater. Sci. Eng.* B **59** 298
[80] Wei S-H, Ferreira L G and Zunger A 1990 *Phys. Rev.* B **41** 8240
[81] Chen S, Gong X G and Wei S-H 2008 *Phys. Rev.* B **77** 073305