PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

A Generalized Method for the Inversion of Cohesive Energy Curves from Isotropic and Anisotropic Lattice Expansions

Kevin M. Schmidt[a] and Victor R. Vasquez[a‡]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Cohesive energy curves contain important information about energetics of atomic interactions in crystalline materials, and these are more often obtained using *ab initio* methods such as density functional theory. Decomposing these curves into the different interatomic contributions is of great value to evaluate and characterize the energetics of specific types of atom-atom interactions. In this work, we present and discuss a generalized method for the inversion of cohesive energy curves of crystalline materials for pairwise interatomic potentials extraction using detailed geometrical descriptions of the atomic interactions to construct a list of atomic displacements and degeneracies, which is modified using a Gaussian elimination process to isolate the pairwise interactions. The proposed method provides a more general framework for cohesive energy inversions that is robust and accurate for systems well-described by pairwise potential interactions. Results show very good reproduction of cohesive energies with the same or better accuracy than current approaches with the advantage that the method has broader applications.

1 Introduction

Cohesive energy curves describe the energetics of crystalline materials as a function of the size of the crystal, usually characterized in units of the lattice constant. They contain all the contributions of the interatomic potential energy interactions in the crystal and are commonly used to determine equilibrium structures and estimate various thermophysical properties. Decomposing the cohesive energy into the different interatomic contributions is of great value to evaluate and characterize the energetics of specific types of atom-atom interactions, which can then be used to extract parameter-free interatomic potentials suitable for use in molecular dynamics (MD) and Monte Carlo (MC) methods, among several other applications, or to parametrize specific interatomic potential energy models. Significant efforts are reported in the literature for the inversion of cohesive energy curves for the extraction of atomic pairwise interactions[1–5], ternary interaction[6], and density-dependent[7,8] potentials for a variety of systems. Unfortunately, the inversion methods currently available restrict significantly the type of geometry and atomic structural complexity that can be studied. These are mostly constrained to single atom-type systems or lattice geometries that vary isotropically upon expansion.

In this work, we introduce a generalized method for the inversion of cohesive energy curves of crystalline materials for pairwise potentials extraction. Details of the lattice geometry are used to construct a list of atomic displacements and degeneracies, which is subsequently modified using a Gaussian elimination process until single pairwise interactions are isolated. The method can be applied to a variety of systems and geometries that cannot be analyzed with current techniques. For example, interatomic pairwise potentials can be produced for atoms within rigid clusters interacting with *(i)* other atoms not constrained by any bonds (*i.e.*, free to move) or *(ii)* atoms located in separate clusters within the system. Additionally, the inter-cluster interactions can be supplemented with bonded intramolecular potentials to describe interactions at many length-scales. We compare the accuracy of the proposed method with current techniques such as the Chen-Möbius inversion[9] in crystals where both methods are applicable. Results show increased accuracy of the proposed technique in the reproduction of cohesive energy curves.

Ab initio methods are quite suitable for the calculation of cohesive energy curves. A significantly larger variety of systems and length-scales can be studied with the increasing computational power and codes optimization. However, calculating the dynamics of most systems for any time-scale of interest is still computationally expensive with electronic structure methods, and the study of time-dependent phenomena of atomic systems is still preferred using classical frameworks such as MD and MC-based methods which require detailed description of the energetics of

[a] *Chemical and Materials Engineering Department, University of Nevada–Reno, Reno, NV 89557 USA. Fax: +1 775 784 4764; Tel: +1 775 784 6060; E-mail: victor.vasquez@unr.edu*

‡ To whom correspondence should be addressed.

the different atomic interactions in the system. Additionally, it is appropriate and convenient in many situations to model a collection of atoms within a lattice as a single unit or rigid body with fixed intramolecular distances and angles. This type of analysis is well-suited for MD simulations of systems with very rapid intramolecular bond vibrations—where the use of classical methods is questionable[10], relative to the vibrations of neighboring units or atoms. For example, in metal hexaborides the internal modes of the boron octahedra sub-lattice structures are of frequencies in the range of 20-38 THz[11]. To capture the dynamics of these high-frequency vibrations with MD simulations, very short integration time-steps of less than 2.5 fs are required, and probably significantly smaller for practical reasons. In addition, the external modes of vibration are of far smaller frequencies; therefore, the octahedral units effectively act as rigid bodies in comparison to their surroundings. This type of model has recently been used by Schmidt *et al.*[12] to study the energetic and dynamic trends in various metal hexaboride systems. The use of rigid bonds and bodies has the benefit of reducing the number of degrees of freedom of the system, allowing for molecular simulations of greater time and length-scales. Many systems contain structures that can be appropriately described by rigid body dynamics, including metal nanoparticles, polymers, and fullerenes among others. Dynamic or equilibrium analysis in these type of systems requires accurate interatomic potentials for atoms interacting among fixed clusters or rigid units. These interaction atomic potentials are typically obtained from either experimental data or quality *ab initio* calculations of the energetics in the system.

Extraction of interatomic potentials from energy landscapes is very costly from a computational standpoint. A fine grid of energy evaluations is required for accurate sampling of the landscape. Expanding the application of cohesive energy inversion techniques could provide important advantages in simplifying, systematizing, and speeding up the task of obtaining pairwise potentials using *ab initio* methods. The method proposed in this work provides a more general framework for cohesive energy inversions that is robust and accurate for systems well-described by pairwise potential interactions. Our results show very good reproduction of cohesive energies with more accuracy than using current approaches such as Chen-Möbius[9] with the additional advantage of broader applications.

2 Background

2.1 Crystal Lattices

Crystals are defined by an atomic basis, or unit cell, containing any number of atoms in a precise arrangement which infinitely repeats on a lattice geometry[13]. Three lattice vectors, $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, specify the span of the unit cell in $\mathbb{R}^3$. Defining the origin to be any coordinate in $\mathbb{R}^3$, each atom "i" described by the basis will be located at a point $\mathbf{r}_i$ from the origin within the unit cell, in addition to appearing in every replica cell in the lattice. The set $\mathbf{R}_i$ of coordinates for each atom "i" satisfies

$$\mathbf{R}_i = \{\mathbf{r}_i + m\mathbf{a}_1 + n\mathbf{a}_2 + l\mathbf{a}_3 : \{m,n,l\} \subseteq \mathbb{Z}\}, \quad (1)$$

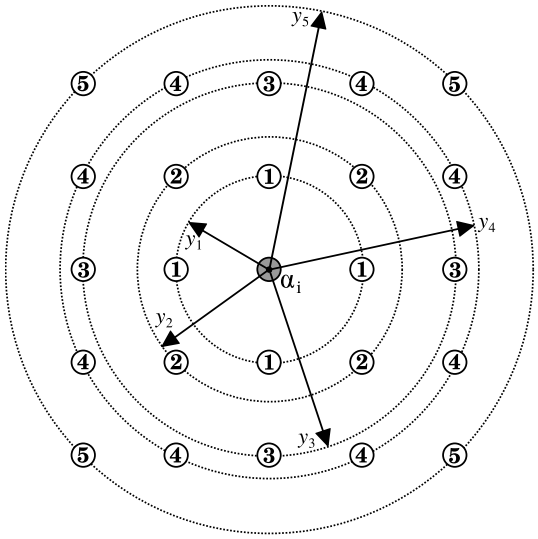

Fig. 1 Two-dimensional lattice of a single atom α_i surrounded by 24 β atoms. β atoms labeled according to normalized shell distances y_k for five shells.

where $\mathbb{Z}$ is the set of all integers, $\mathbf{R}_i \subseteq \mathbf{R}$, and $\mathbf{R}$ is the set of all atomic coordinates within the crystal displaced from the origin. The lattice geometries can be of the familiar isometric types, such as body-centered cubic (BCC), face-centered cubic (FCC), and simple cubic (SC), or any of the other eleven unique Bravais lattices. For a specified lattice geometry, atomic basis, and lattice constant, the distance between two arbitrary α-type and β-type atoms is given by the separation vector $\mathbf{r}_{ij}^{\langle\alpha\beta\rangle} = \mathbf{r}_i^{\langle\alpha\rangle} - \mathbf{r}_j^{\langle\beta\rangle}$, with $\{\mathbf{r}_i^{\langle\alpha\rangle}, \mathbf{r}_j^{\langle\beta\rangle}\} \in \mathbf{R}$. Note that the α-β pair can result from either homatomic or hetero-atomic interactions.

Figure 1 shows a two-dimensional square lattice with a single atom α_i surrounded by its β-type neighbors. We find that multiple atoms rest on the boundary of circles or shells having radius $y_k^{\langle\alpha\beta\rangle}$ as we move outward from α_i, where we have defined

$$y_k^{\langle\alpha\beta\rangle} = \frac{\left|\mathbf{r}_k^{\langle\alpha\beta\rangle}\right|}{a} \quad \forall \quad \mathbf{r}_k^{\langle\alpha\beta\rangle} \in \left\{\mathbf{r}_{ij}^{\langle\alpha\beta\rangle}\right\}, \quad (2)$$

as the dimensionless distance, and the subscript "k" denotes the shell where the atoms are located. Each element in the ordered set $y_k^{\langle\alpha\beta\rangle} \in \mathbf{Y}^{\langle\alpha\beta\rangle}$ refers to atoms located on the surface of the k^{th} nearest shell of radius $r_k^{\langle\alpha\beta\rangle} = (y_k^{\langle\alpha\beta\rangle} \cdot a)$ in the lattice, where "a" is the lattice constant. We can enumerate the β_j atoms which have degenerate displacements from α_i using the set of pairwise distances $\{\mathbf{r}_{ij}^{\langle\alpha\beta\rangle}\}$ to create the radial number distribution function

$$\mathbf{n}^{\langle\alpha\beta\rangle}(y) = \sum_j \delta\left[y - \frac{\left|\mathbf{r}_{ij}^{\langle\alpha\beta\rangle}\right|}{a}\right], \quad (3)$$

where $\delta[\cdot]$ is the unit impulse function. The set $\mathbf{Y} \supseteq \mathbf{Y}^{\langle\alpha\beta\rangle}$ and distribution $\mathbf{n}^{\langle\alpha\beta\rangle}(\cdot)$ are each uniquely determined by the basis and lattice-type for each lattice constant. This relationship is depicted in Fig. 2 for the square lattice of Fig. 1.

A common procedure used to generate cohesive energy data

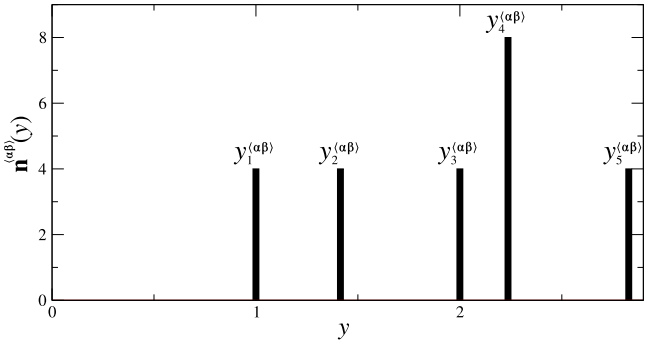

Fig. 2 Radial number distribution function $\mathbf{n}^{\langle\alpha\beta\rangle}$ for the two-dimensional lattice of Fig. 1. Each non-zero value represents the number of β atoms located at the normalized distance $y_k^{\langle\alpha\beta\rangle}$.

as a function of interatomic distance from *ab initio* methods is to perform *in silico* expansions of the lattice and compute the energy of the system as a function of distance—typically in units of lattice constant. In isotropic expansions of a lattice, all the atoms move uniformly from a given origin to new positions proportional to the increase or decrease in size. In such cases, both the lattice constant and the interatomic pairwise distances increase by the same factor during the cell volume expansions. Lattices with isotropic expansions will therefore have a fixed set $\mathbf{Y}$ regardless of the lattice constant and produce a single radial number distribution function independent of the lattice constant and α-β pairs. These expansions are typical for calculations involving isometric lattices lacking rigid bodies, where each atom is prescribed a location within the unit cell that is linearly dependent on one or all of the lattice vectors. When rigid bodies are used to characterize a system, the distances between atoms within the rigid unit are constant and no longer proportional to the lattice size throughout the expansion process. Atoms constrained by a fixed distance or bond will produce a set $\mathbf{Y}$ and number distribution $\mathbf{n}(\cdot)$ which are dependent on the lattice constant, and we refer to these cases as non-uniform, or *anisotropic*, expansions.

2.2 Cohesion

The total energy of a crystal can be decomposed into a series of summations of interatomic potentials as a function of the atomic coordinates, $\{\mathbf{r}\}$, plus the contributions from isolated atomic energies, $\phi^\circ(\cdot)$. The interatomic potentials include interactions between pairs, triplets, and so forth, and the total energy is given by

$$\begin{aligned}\mathbb{E}_{tot} &= \sum_i \phi^\circ(\mathbf{r}_i) + \frac{1}{2!}\sum_i\sum_j \phi_2(\mathbf{r}_i,\mathbf{r}_j) \\ &+ \frac{1}{3!}\sum_i\sum_j\sum_k \phi_3(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) + \cdots,\end{aligned} \tag{4}$$

where $\phi_N(\cdot)$ refers to an interaction potential involving N atoms. For the purpose of this work, we approximate the total energy per atom as

$$\mathbb{E}_{tot,i} = \phi_i^\circ + \frac{1}{2}\sum_j \phi(\mathbf{r}_{ij}), \tag{5}$$

where $\mathbf{r}_{ij}$ is the separation vector between atoms "i" and "j", and the contribution of the higher-order interactions (ternary and above) is implicitly included or added in the binary potentials. For many applications, Eqn (5) is a valid approximation as most of the interatomic contributions to the total energy are due to binary interactions. The cohesive energy per atom is then defined as

$$\mathbb{E}_{coh,i} = \mathbb{E}_{tot,i} - \phi_i^\circ = \frac{1}{2}\sum_j \phi(\mathbf{r}_{ij}). \tag{6}$$

Using the relationships given by Eqn (2) and Eqn (3), we can rewrite the cohesive energy per α_i atom as

$$\mathbb{E}_{coh,i}^{\langle\alpha\rangle}(a) = \frac{1}{2}\sum_\beta\sum_k \mathbf{n}^{\langle\alpha\beta\rangle}(y_k^{\langle\alpha\beta\rangle})\cdot\phi^{\langle\alpha\beta\rangle}(y_k^{\langle\alpha\beta\rangle}\cdot a), \tag{7}$$

in which, for a given α-β lattice structure, the only unknown is the pairwise interaction potential, $\phi^{\langle\alpha\beta\rangle}(\cdot)$.

2.3 Cohesive Energy Inversion Methods

Methods to extract interaction potentials from cohesive energy data typically follow a Gaussian elimination process to yield an "*exact*" potential with the ability to reproduce the correct cohesive energy through a lattice summation using Eqn (7). For a one-component system, though the method can be easily extended to multicomponent systems, the cohesive energy per atom is given by

$$\mathbb{E}_{coh,i}(a) = \sum_{k=1}^{N_s}\frac{\mathbf{n}(y_k)}{2}\phi(y_k\cdot a), \tag{8}$$

where the subscript "k" indicates the spherical shell in which neighboring atoms are located, ordered monotonically from nearest to furthest, and N_s is the total number of shells included in the calculation. To solve for the value of the interaction potential at the nearest-neighbor distance, $y_1\cdot a$, we can readily rearrange Eqn (8) to obtain

$$\phi(y_1\cdot a) = \frac{2\cdot\mathbb{E}_{coh,i}(a)}{\mathbf{n}(y_1)} - \sum_{k=2}^{N_s}\frac{\mathbf{n}(y_k)}{\mathbf{n}(y_1)}\phi(y_k\cdot a). \tag{9}$$

The complete pair-potential is then constructed by solving some variation of Eqn (9) for a range of nearest-neighbor distances, and the specific procedure varies depending on which method is used for the inversion process. The method proposed by Carlsson, Gelatt, and Ehrenreich[14] (CGE) defines a linear transformation, θ, which produces the cohesive energy when applied to the pair-potential, such that $\mathbb{E}_{coh,i} = \theta\phi(r)$. Inversion of this relationship isolates the pair-potential at the nearest-neighbor distance as an explicit function of the cohesive energy. The expression for θ is found through use of a weighted-scale transformation operator, $\mathbf{T}_k$,

$$\mathbf{T}_k\cdot\phi(y_1\cdot a) = \frac{1}{2}\mathbf{n}(y_k)\phi(y_k\cdot a), \tag{10}$$

which, when applied to the pair-potential at the first nearest-neighbor distance, gives the contribution of the k^{th} shell atoms to the cohesive energy using the set of shell distances, y_k, and their respective multiplicities, $\mathbf{n}(y_k)$. This relationship assumes that all

distances vary linearly with lattice constant sizes, since the set of values of $\{y\}$ and $\mathbf{n}(\cdot)$ are derived from a single geometrical layout of the crystal. The cohesive energy can then be written in terms of the linear transformation θ as

$$\mathbb{E}_{coh,i}(a) = \sum_k \mathbf{T}_k \phi(y_1 \cdot a) = \theta\phi(y_1 \cdot a), \tag{11}$$

with

$$\theta = \mathbf{T}_1 \left(1 + \sum_{k=2}^{\infty} \mathbf{T}_1^{-1}\mathbf{T}_k\right). \tag{12}$$

Taking the inverse of θ, we get

$$\phi(y_1 \cdot a) = \theta^{-1}\mathbb{E}_{coh,i}(a), \tag{13}$$

and substitution of the expressions for the transformation operator $\mathbf{T}_k$ produces the desired expression for the pairwise potential at y_1.

The procedure by Bazant[15] uses recursive substitution of Eqn (9) in a manner similar to the CGE method, with one caveat—the cohesive energy of crystals at low density, or equivalently at large lattice expansions, is assumed to correspond to the energy of unscreened atoms interacting in a gaseous phase, which is not the same as the energy of screened long-range atomic interactions encountered in bulk crystals. Bazant[15] therefore modifies the cohesive energy curve to achieve a smooth convergence to zero at an optimized separation distance. The pair-potential is then obtained in reverse order by using the largest lattice constant of the expansions first, which require a minimal number of shells, and iteratively works backward to produce the inversion.

The Chen-Möbius inversion method[9] is an alternative for highly symmetric lattices, where the elements $y_k \in \mathbf{Y}$ closely form a multiplicative semi-group. In lattices with a high degree of symmetry, the set of distances $\mathbf{Y}$ contains many elements that obey the relationship $y_m \cdot y_n = y_p$ for $m \neq n \neq p$. The Möbius function can be used with this, or a modified semi-group, to obtain inversion coefficients such that

$$\phi(y_1 \cdot a) = 2\sum_{k=1}^{N_s} J(k) \cdot \mathbb{E}_{coh,i}\big(B(k) \cdot a\big), \tag{14}$$

where $J(k)$ are the inversion coefficients, N_s is the number of lattice points used to invert the potential at $y_1 \cdot a$, and $B(k)$ refers to the multiplicative semi-group describing the atomic lattice points in the crystal. When applicable, the method provides fast convergence and ease of implementation.

All of these methods require lattice expansions to be performed isotropically, where the set $\mathbf{Y}$ is fixed and the radial number distribution function $\mathbf{n}(\cdot)$ is identical for any lattice constant value. Systems containing fixed or rigid intramolecular bonds cannot be described by uniform lattice expansions. The set $\mathbf{Y}$ and $\mathbf{n}(\cdot)$ are no longer constant, precluding usage of the aforementioned methods for extraction of pair-potentials from direct inversion of cohesive energy curves. Atoms connected by rigid bonds are required to remain at fixed distances relative to each other. Although the centroids of the rigid units have displacements varying equally proportional to the lattice expansion, each absolute atomic distance to the origin will not. A new set of distances $\mathbf{Y}_d$ and number distribution function $\mathbf{n}_d(\cdot)$ have to be defined for these type of systems as a function of the separation distance, a_d, typically in units of lattice constant.

3 Rigid Cluster Systems

3.1 Isotropic vs Anisotropic Expansions

To illustrate the difference between the term "isotropic" versus "anisotropic" for *in silico* expansions of lattice structures, we start with a one-component simple cubic (SC) lattice of an atom at the origin. In this structure, atoms are located at the vertices of an infinitely repeating cubic box. The locations of each atom "j" from some atom "i" are described by the set $\mathbf{Y}^{\langle SC\rangle}$, with elements

$$y_k^{\langle SC\rangle} \subseteq \left\{\frac{|\mathbf{r}_{ij}|}{a} = \sqrt{m^2+n^2+l^2}\right\}$$
$$\forall \quad \{\{m,n,l\} \subseteq \mathbb{Z} : (m,n,l) \neq (0,0,0)\}. \tag{15}$$

In isotropic expansions of the lattice structure, the distance of the k^{th} shell surrounding an atom is determined by product $y_k \cdot a$. The displacements of these shells as a function of the lattice constant are shown in Fig. 3 for this SC lattice structure. Each line represents an occupied spherical shell as the lattice expands. In addition to the set $\mathbf{Y}^{\langle SC\rangle}$ being identical for all lattice sizes, the distribution $\mathbf{n}^{\langle SC\rangle}$ (shown in Fig. 4) is also constant with respect to a.

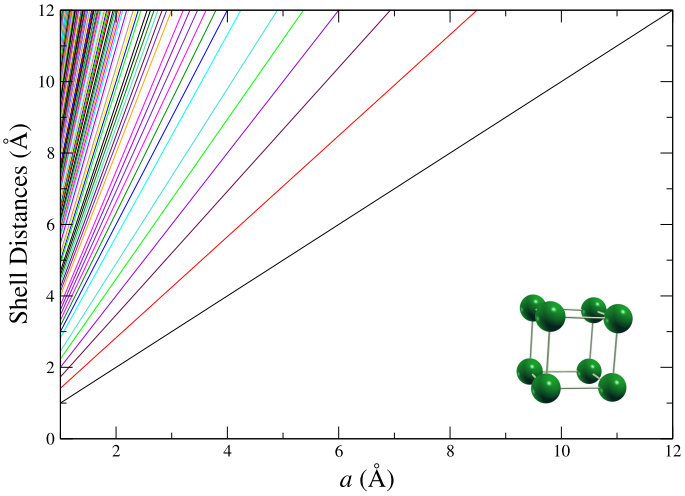

Fig. 3 Locations of radial shells containing neighboring atoms as a function of lattice constant for a single atom in a SC lattice with isotropic expansions of the crystal structure. Lowest line refers to the 1^{st} nearest-neighbor distance.

However, the simple and straight-line patterns of Fig. 3 cannot be obtained for a lattice structure with fixed sub-lattice units. For example, Fig. 5 shows the structure of a SC lattice with octahedral units of space group $Pm\bar{3}m$ symmetry, typical of the boron framework in metal hexaborides and other compounds. The octahedral unit has side-lengths L^{O_h} with atoms located on the vertices of the polyhedron at the $6f$ crystallographic sites. The barycenters of the octahedra are placed at the central sites of a cube, as depicted in Fig. 5. In the case of metal hexaborides, the boron octahedral sub-structures are fairly rigid[16]. Density functional

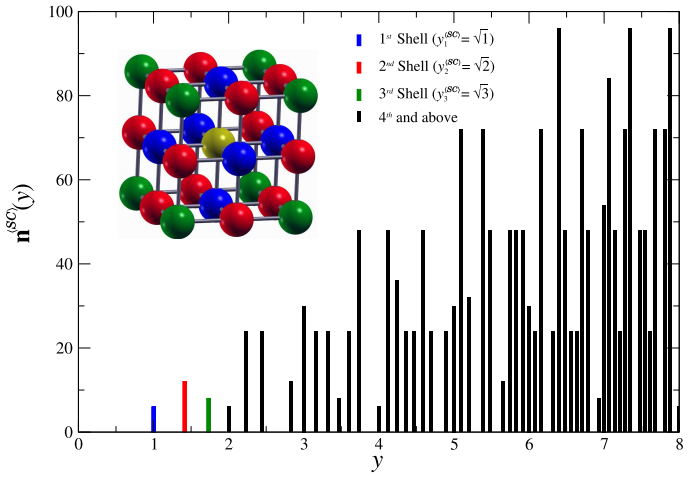

Fig. 4 Radial number distribution function of neighboring atoms in a simple cubic lattice for pairwise distances up to $\mathbf{r}_k^{\langle SC\rangle} = 8 \cdot a$. Each bar corresponds to the number of atoms located at the normalized distance, $y_k^{\langle SC\rangle} = |\mathbf{r}_k^{\langle SC\rangle}|/a$. **Inset:** 27 unit cells showing spatial locations of first three radial distances—yellow atom placed at the origin.

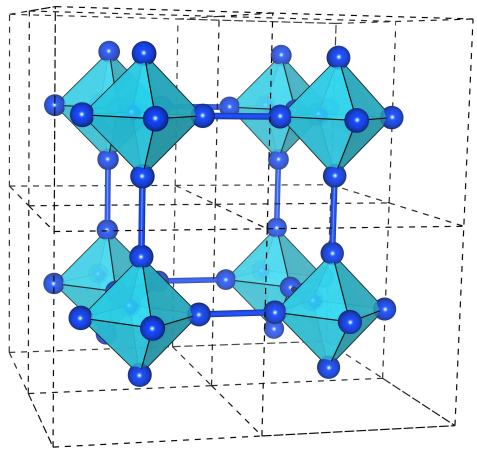

Fig. 5 Octahedral cluster arrangement typical of the boron structure in metal hexaborides. The first nearest-neighbor inter-octahedral bonds at $y_1^{\langle O_h\rangle}$ are shown as blue rods connecting the octahedra.

theory (DFT) methods can be used to obtain energetics of the boron-boron interaction between atoms in separate octahedral units as a function of the separation distances by performing energy calculations on lattice expansions that fix the relative location of the octahedral barycenters in the unit cell while increasing the inter-octahedral distance. In this case, the lattice expansions are anisotropic, as opposed to isotropic, in order to keep the intramolecular bonds and volumes of the octahedral units constant, and the set of normalized distances $y_k^{\langle O_h\rangle} \in \mathbf{Y}^{\langle O_h\rangle}$ varies with the lattice constant due to the anisotropic expansion of the lattice.

To distinguish from the isotropic expansion case, let the set $y_{k,d} \in \mathbf{Y}_d$ and $\mathbf{n}_d(\cdot)$ denote the normalized displacements and atom number distribution as function of the lattice constant a_d, respectively, where the subscript "d" represents the number position in the array of lattice constants used for calculations. The set

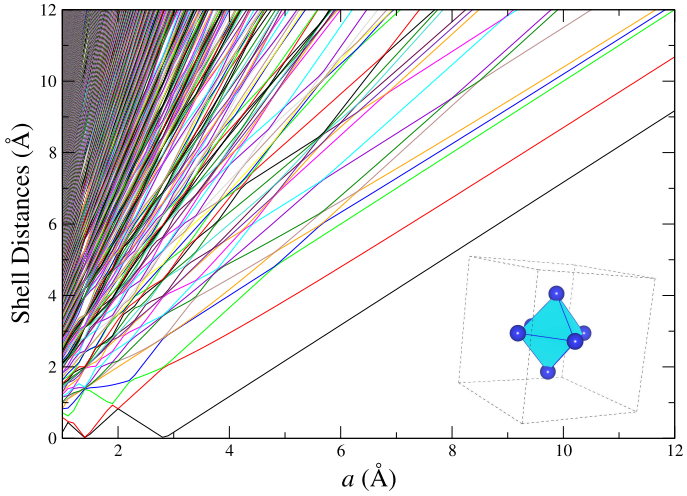

Fig. 6 Locations of radial shells containing neighboring atoms as a function of lattice constant for a single atom within a rigid octahedral unit having fixed side lengths of 2 Å. The octahedra are placed at the central site of a SC lattice . Lowest line refers to the 1st nearest-neighbor distance.

$\mathbf{Y}_d^{\langle O_h\rangle}$ for the octahedral system is calculated using

$$y_{k,d}^{\langle O_h\rangle} \subseteq \begin{cases} \sqrt{m^2+n^2+l^2}, \\ \sqrt{(m+2\gamma_d^{O_h})^2+n^2+l^2}, \\ \sqrt{(m+\gamma_d^{O_h})^2+(n+\gamma_d^{O_h})^2+l^2} \end{cases} \tag{16}$$

$$\forall \quad \{\{m,n,l\} \subseteq \mathbb{Z} : (m,n,l) \neq (0,0,0)\}$$

for arbitrary side-lengths L^{O_h}, where $\gamma_d^{O_h}$ is defined as

$$\gamma_d^{O_h} = \frac{\sqrt{2}\cdot L^{O_h}}{2\cdot a_d}. \tag{17}$$

The locations of each spherically distributed atomic shell are calculated using Eqn (16) and shown in Fig. 6 for a system with $L^{O_h} = 2$ Å. In this case, we are only interested in the inter-octahedral atomic interactions such that the first nearest-neighbor distance is located *between* the octahedra—blue rods in Fig. 5; therefore, the pairwise distances between neighbors located within the same rigid structure are omitted in Fig. 6. Including these will show two additional horizontal lines located at L^{O_h} and $\sqrt{2}\cdot L^{O_h}$ on the ordinate axis, as these refer to the rigid intra-octahedral distances and would remain constant for every lattice constant. The interactions from these two intra-octahedral shells are then effectively absorbed into the isolated atomic energies using this potential model since they are invariant with respect to lattice expansions.

There are notable differences between the shell locations of the two systems in Fig. 3 and Fig. 6, and we will highlight two of these in reference to the octahedral system. Figure 6 shows unusual dips in the curves at the lower values of the lattice constant where the first few nearest-neighbor distances have both increasing and decreasing tendencies. These are caused by the expansion of interpenetrating octahedra until the vertices of separate units meet at $a = 2\sqrt{2}$ Å. This effect, shown for illustration

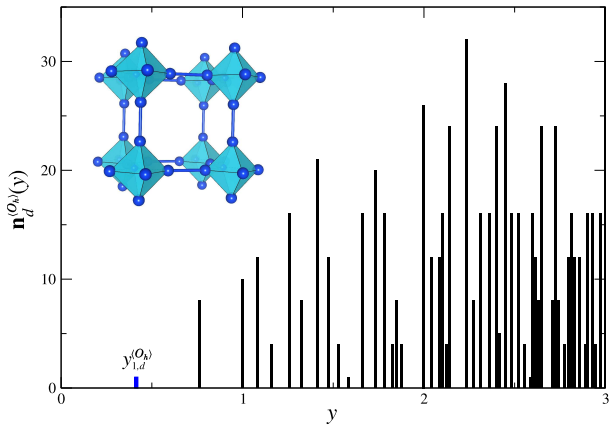

Fig. 7 Distribution function for a single atom within the octahedral unit for $r_{k,d}^{\langle O_h\rangle} \leq 3 \cdot a_d$. Octahedral units have a side length is 2 Å with the barycenters located at the centers of a cubic lattice with $a_d = (2+2\sqrt{2})$ Å. Note that $y_{1,d}^{\langle O_h\rangle}$ refers to the nearest-neighbor shell containing one atom. **Inset:** Eight unit cells of the octahedra with nearest-neighbor distances connected by blue rods.

purposes, highlights the precaution one should take when generating these *in silico* lattice structures to avoid overlapping nuclei. A related characteristic exhibited by the octahedral shell loci is the occurrence of intersections between the shell curves after $a = 2\sqrt{2}$ Å, corresponding to expansions where the radial number distribution function $\mathbf{n}_d^{\langle O_h\rangle}(\cdot)$ temporarily experiences increased values due to the overlap of shell locations. When using fixed lengths within the basis, certain lattice constants will allow the radial distance of two separate shells to approach the same value and eventually intersect, at which point their multiplicities are combined and spikes occur in the radial number distribution. For reference, the octahedral distribution $\mathbf{n}_d^{\langle O_h\rangle}(y)$ is plotted in Fig. 7 using $L^{O_h} = 2$ Å and $a_d = (2+2\sqrt{2})$ Å for a range of interatomic distances.

The inversion methods described earlier work quite well for systems whose sets $\mathbf{Y}^{\langle\alpha\beta\rangle}$ and distribution functions $\mathbf{n}^{\langle\alpha\beta\rangle}(\cdot)$ remain unchanged during the course of *in silico* lattice expansions necessary to produce the cohesive energy as a function of distance with DFT methods. These systems will have shell distance functions similar to that observed for the SC system in Fig. 3, characterized by straight lines with a single intersection point. For the case of the rigid sub-structures model, fixed sets $\mathbf{Y}^{\langle\alpha\beta\rangle}$ and distribution functions $\mathbf{n}^{\langle\alpha\beta\rangle}(\cdot)$ no longer accurately describe the geometry of the system upon lattice expansions. Changing the values of a_d or L^{O_h} describing the octahedral structures will alter both $\mathbf{Y}_d^{\langle O_h\rangle}$ and $\mathbf{n}_d^{\langle O_h\rangle}(\cdot)$, such that these must be calculated for each lattice constant.

3.2 Inter-cluster Atomic Pair-Potentials

It is common practice to assume transferablity of pair-potentials in different atomic arrangements or molecular structures. For the case described in Fig. 5, the intra-cluster or intra-octahedral environment is heavily dependent on the specific bonding motifs within the cluster, and spatial separations between clusters play a large role in determining the electrostatic distribution around the atoms. This causes difficulties for atomic pair-potentials developed from this type of geometry to be transferable due to the difference between the electron localization of atoms in this geometry versus that of more classical bulk materials. Therefore, alternative methods are typically employed to describe these interactions.

Molecular dynamics studies of cluster systems[17,18] generally use one of two approaches to perform bulk simulations. They either *(i)* model the entire cluster as a single coarse-grained entity or *(ii)* model the cluster as a rigid unit composed of atoms, interacting through general exponential or power-law potentials to describe the energetics of cluster-atom and atomic cluster-cluster pairs. The coarse-graining method provides the advantage of significantly reducing the number of interactions required and has been used to study gold nanoparticles[17], spherical silica nanoparticles[19], and generic spherical nanoparticles[18]. These methods essentially replace the cluster with a single point mass interacting through an effective pair-potential and seem to be solely applicable to spherical particles. The all-atom methods have been used to investigate aggregation of fullerenes[20] and rheology of arbitrarily shaped nanoparticles[21] with interactions described by basic Lennard-Jones potentials. These types of models perform well in describing basic physical behavior, but their accuracy decreases in capturing atomistic details related to the cluster configuration. High-quality *ab initio* calculations, applied to a specific cluster atomic model (*e.g.,* rigid bodies, pairwise interactions, among others), can be used to calculate detailed energetics of the different atomic interactions, which is then used for parametrization of potential energy models for MD simulations.

The approach proposed in this work for inversion of cohesive energy curves generalizes traditional methods such that a whole new set of systems or atomic configurations can be used for the extraction of both *homo-* and *hetero*-atomic pair-potentials. The method is very well-suited to deal with systems containing rigid atomic units such as the one depicted in Fig. 5, where lattice expansions are anisotropic in nature. In order to use a recursive relationship such as that of Eqn (9), the expressions $\mathbf{Y}$ and $\mathbf{n}(\cdot)$ must be re-defined for the case of anisotropic expansions where relative locations within the unit cell are dependent upon the lattice constant as shown, for example, in Eqn (16).

4 Generalized Inversion Approach

4.1 Governing Equations

Rather than assuming a specified set of normalized distances, $\tilde{\mathbf{Y}}_d$ contains all of the *absolute* pairwise distances

$$\tilde{y}_{k,d} = |\mathbf{r}_{k,d}| \quad \forall \quad \mathbf{r}_{k,d} \in \{\mathbf{r}_{ij,d}\}, \tag{18}$$

for a lattice constant a_d with the first nearest-neighbor distance denoted as $\tilde{y}_{1,d}$. We define r_{cut} as the radial cutoff distance for the interatomic potential whose value should coincide with the largest lattice constant used to determine the cohesive energy, a_f,

such that the following conditions are satisfied:

$$\begin{aligned} \tilde{y}_{1,f} &\leq r_{cut}, \\ \mathbb{E}_{coh}(a_f) &= 0, \\ \lim_{a\to a_f} \frac{\partial \mathbb{E}_{coh}(a)}{\partial a} &\approx 0. \end{aligned} \tag{19}$$

The relationships in Eqn (19) guarantee that the range of the inverted potential allows at least one shell to contribute to the cohesive energy for all lattice constants with $\mathbb{E}_{coh} \neq 0$. If these conditions are not met, there may be a lattice constant giving a non-zero energy with all neighbors lying outside of r_{cut} and thus no interactions to contribute to the cohesion.

Each distance $\tilde{y}_{k,d}^{\langle\alpha\beta\rangle} \in \tilde{\mathbf{Y}}_d^{\langle\alpha\beta\rangle}$ has a corresponding multiplicity determined by the radial number distribution function for α_i in an α–β pair, given by

$$\tilde{\mathbf{n}}_d^{\langle\alpha\beta\rangle}(\tilde{y}) = \sum_j \delta\left[\tilde{y} - \left|\mathbf{r}_{ij,d}^{\langle\alpha\beta\rangle}\right|\right], \tag{20}$$

where $\left|\mathbf{r}_{k,d}^{\langle\alpha\beta\rangle}\right| \leq r_{cut}$. The cohesive energy per α_i atom for any lattice constant a_d is then given by

$$\mathbb{E}_{coh,i}^{\langle\alpha\rangle}(a_d) = \frac{1}{2}\sum_\beta \sum_{k=1}^{N_{s,d}} \tilde{\mathbf{n}}_d^{\langle\alpha\beta\rangle}\left(\tilde{y}_{k,d}^{\langle\alpha\beta\rangle}\right)\phi^{\langle\alpha\beta\rangle}\left(\tilde{y}_{k,d}^{\langle\alpha\beta\rangle}\right), \tag{21}$$

where $N_{s,d}$ is the number of shells included up to r_{cut} for lattice constant a_d. For an isolated α–β interaction, Eqn (21) can be rearranged to give the value of the interaction potential at the first nearest-neighbor distance, $\tilde{y}_{1,d}^{\langle\alpha\beta\rangle}$, as a function of the cohesive energy and a linear combination of potential energy values at $\tilde{y}_{k,d}$ for $k \geq 2$. Simplifying, we obtain

$$\phi(\tilde{y}_{1,d}) = \frac{2\cdot\mathbb{E}_{coh,i}^{\langle\alpha\beta\rangle}(a_d)}{\tilde{\mathbf{n}}_d(\tilde{y}_{1,d})} - \sum_{k=2}^{N_{s,d}} \frac{\tilde{\mathbf{n}}_d(\tilde{y}_{k,d})}{\tilde{\mathbf{n}}_d(\tilde{y}_{1,d})}\phi(\tilde{y}_{k,d}), \tag{22}$$

and with this expression we can isolate the pairwise potential for the first nearest-neighbor distance $\tilde{y}_{1,d}$ of any lattice constant using a Gaussian elimination process—see the following section for details.

4.2 Gaussian Elimination

Extracting the value of the potential located at $\tilde{y}_{1,d}$ for any arbitrary lattice constant, a_d, requires an expression for $\phi\left(\tilde{y}_{1,d}\right)$ to be dependent only on known values. However, the expression in Eqn (22) gives the pair-potential at the nearest-neighbor distance as a function of a single cohesive energy point and values of the potential energy function at increasing distances. Since the set $\mathbf{Y}_d$ only contains distances less than r_{cut}, fewer potential energy function evaluations are required as the lattice size increases. Recursive substitution of Eqn (22) allows for the elimination of all potential energy terms after $\tilde{y}_{1,d}$ up to r_{cut} to produce an expression for $\phi(\tilde{y}_{1,d})$ which depends only on calculated cohesive energy curves and radial number distribution values. Using the recursive relationship solves for a single point of the interaction potential, and the process must be performed for a range of nearest-neighbor distances to populate the potential energy curve as a function of the interatomic distance. For example, when solving the pair-potential for the first nearest-neighbor distance $\tilde{y}_{1,1}$ with the lattice constant a_1, the subscript "$d = 1$" refers to the smallest lattice constant. Each time a new lattice size is analyzed with Eqn (22), this counter increases by one such that the next value of "d", in increasing order, refers to larger values of the lattice constants used. Additionally, relationships between lattice constants and corresponding 1^{st} and 2^{nd} nearest-neighbor distances are helpful in the derivation. To this end, we define

$$f_{\tilde{y},1}(a_d) = \tilde{y}_{1,d}, \tag{23}$$

$$f_{\tilde{y},2}(a_d) = \tilde{y}_{2,d}, \tag{24}$$

$$f_{a,1}(\tilde{y}_{1,d}) = a_d, \tag{25}$$

$$f_{a,2}(\tilde{y}_{2,d}) = a_d, \tag{26}$$

where $f_{\tilde{y},n}(\cdot)$ and $f_{a,n}(\cdot)$ are mapping functions between $\tilde{y}_{k,d}$ and a_d for the n^{th} nearest-neighbor.

In order to invert the cohesive energy and yield $\phi(\tilde{y}_{1,1})$, the unknown values of the interaction potential at $\phi(\tilde{y}_{k,1})$ for $k \geq 2$ must be replaced with functions dependent on cohesive energies through the recursive use of Eqn (22). Special care must be taken to ensure that the values of $\tilde{\mathbf{Y}}_d$ and $\tilde{\mathbf{n}}_d(\cdot)$ are accurately defined as a function of the lattice constant a_d during the DFT configuration energy calculations of the lattice expansions. The proposed inversion method for cohesive energy curves relies on the generalization of the distance and radial number distribution functions $\tilde{\mathbf{Y}}_d$ and $\tilde{\mathbf{n}}_d(\cdot)$.

The elimination process begins by solving for $\phi(\tilde{y}_{1,1})$ using Eqn (22) to obtain

$$\phi(\tilde{y}_{1,1}) = \frac{2\cdot\mathbb{E}_{coh,i}^{\langle\alpha\beta\rangle}(a_1)}{\tilde{\mathbf{n}}_1(\tilde{y}_{1,1})} - \sum_{k=2}^{N_{s,1}} \frac{\tilde{\mathbf{n}}_1(\tilde{y}_{k,1})}{\tilde{\mathbf{n}}_1(\tilde{y}_{1,1})}\phi(\tilde{y}_{k,1}). \tag{27}$$

Next, we consider the potential at the next nearest-neighbor (NNN) after $\tilde{y}_{1,1}$ in the summation of Eqn (27). This happens to be at the second nearest-neighbor distance, $\tilde{y}_{2,1}$, of a_1 for the first iteration, though this will not always be the case. To replace $\phi(\tilde{y}_{2,1})$, we recognize that there exists a lattice constant a_2 described by $\tilde{\mathbf{Y}}_2$ and $\tilde{\mathbf{n}}_2(\cdot)$ whose first nearest-neighbor is located at $\tilde{y}_{1,2} = \tilde{y}_{2,1}$, and substitution into Eqn (22) gives

$$\phi(\tilde{y}_{1,2}) = \frac{2\cdot\mathbb{E}_{coh,i}^{\langle\alpha\beta\rangle}(a_2)}{\tilde{\mathbf{n}}_2(\tilde{y}_{1,2})} - \sum_{k=2}^{N_{s,2}} \frac{\tilde{\mathbf{n}}_2(\tilde{y}_{k,2})}{\tilde{\mathbf{n}}_2(\tilde{y}_{1,2})}\phi(\tilde{y}_{k,2}), \tag{28}$$

where the condition

$$f_{\tilde{y},1}(a_2) = f_{\tilde{y},2}(a_1), \tag{29}$$

is met, such that

$$\phi(\tilde{y}_{1,2}) = \phi(\tilde{y}_{2,1}). \tag{30}$$

Replacing $\phi(\tilde{y}_{2,1})$ in Eqn (27) with the expression from Eqn (28),

we obtain

$$\phi(\tilde{y}_{1,1}) = \frac{2}{\tilde{\mathbf{n}}_1(\tilde{y}_{1,1})}\left(\mathbb{E}^{\langle\alpha\beta\rangle}_{coh,i}(a_1) - \frac{\tilde{\mathbf{n}}_1(\tilde{y}_{2,1})}{\tilde{\mathbf{n}}_2(\tilde{y}_{2,1})}\mathbb{E}^{\langle\alpha\beta\rangle}_{coh,i}(a_2)\right) \quad (31)$$

$$+ \frac{\tilde{\mathbf{n}}_1(\tilde{y}_{2,1})}{\tilde{\mathbf{n}}_1(\tilde{y}_{1,1})}\sum_{k=2}^{N_{s,2}}\frac{\tilde{\mathbf{n}}_2(\tilde{y}_{k,2})}{\tilde{\mathbf{n}}_2(\tilde{y}_{2,1})}\phi(\tilde{y}_{k,2}) - \sum_{k=3}^{N_{s,1}}\frac{\tilde{\mathbf{n}}_1(\tilde{y}_{k,1})}{\tilde{\mathbf{n}}_1(\tilde{y}_{1,1})}\phi(\tilde{y}_{k,1}).$$

The first iteration shown in Eqn (31) replaces the contribution of $\phi(\tilde{y}_{2,1})$ with $\mathbb{E}^{\langle\alpha\beta\rangle}_{coh,i}(a_2)$ and another summation of terms involving the pair-potential and particle number distribution for the set $\tilde{\mathbf{Y}}_2$, defined by the lattice constant $a_2 = f_{a,1}(\tilde{y}_{2,1})$. In this step, we generate a system with a cohesive energy, by rearranging Eqn (31), which can be described by the form

$$\mathbb{E}^{res}_{coh} = \mathbb{E}^{\langle\alpha\beta\rangle}_{coh,i}(a_1) - \frac{\tilde{\mathbf{n}}_1(\tilde{y}_{2,1})}{\tilde{\mathbf{n}}_2(\tilde{y}_{2,1})}\mathbb{E}^{\langle\alpha\beta\rangle}_{coh,i}(a_2)$$

$$= \frac{\tilde{\mathbf{n}}_1(\tilde{y}_{1,1})\phi(\tilde{y}_{1,1})}{2} + \frac{1}{2}\sum_{k=3}^{N_{s,1}}\tilde{\mathbf{n}}_1(\tilde{y}_{k,1})\phi(\tilde{y}_{k,1}) \quad (32)$$

$$- \frac{\tilde{\mathbf{n}}_1(\tilde{y}_{2,1})}{2}\sum_{k=2}^{N_{s,2}}\frac{\tilde{\mathbf{n}}_2(\tilde{y}_{k,2})\phi(\tilde{y}_{k,2})}{\tilde{\mathbf{n}}_2(\tilde{y}_{2,1})},$$

with a new set of distances, $\tilde{\mathbf{Y}}_{res}$, and number distribution function, $\tilde{\mathbf{n}}_{res}(\cdot)$. Since we have removed the interaction from atoms in the second shell at $\tilde{y}_{1,2} = \tilde{y}_{2,1}$, the set $\tilde{\mathbf{Y}}_{res}$ contains all distances in $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Y}}_2$ except $\tilde{y}_{1,2}$, and the resulting distribution can be determined for any distance $\tilde{y}$ using

$$\tilde{\mathbf{n}}_{res}(\tilde{y}) = \tilde{\mathbf{n}}_1(\tilde{y}) - \frac{\tilde{\mathbf{n}}_1(\tilde{y}_{2,1})}{\tilde{\mathbf{n}}_2(\tilde{y}_{1,2})}\tilde{\mathbf{n}}_2(\tilde{y}). \quad (33)$$

This same process outlined for the cancellation of the second shell in a_1, given by Eqn (31), is continued until the potential energy values for all distances greater than $\tilde{y}_{1,1}$ have been replaced. Each iteration takes the NNN distance found in the summation terms from the previous iteration located at $\tilde{y}_{2,res}$ as an input to Eqn (22) to remove the value of the potential $\phi(\cdot)$ at that location, again modifying the set $\tilde{\mathbf{Y}}_{res}$ and distribution $\tilde{\mathbf{n}}_{res}(\cdot)$. Note that the number of sampled values from the cohesive energy curve increases exponentially with r_{cut}; therefore, care should be exercised to avoid expensive calculations.

4.3 Iterative Algorithm

A simple computer program can be developed to handle this inversion procedure in a linear fashion for any type of lattice, as long as the cohesive energy *(i)* has been calculated as a function of the lattice constant and *(ii)* has been modified to reflect a single α-β interaction between the atom α_i and its surrounding β-type neighbors. These modifications are necessary since the inversion relies on the usage of cohesive energies for a specific interaction. If the cohesive energy is assumed to be a sum of separate contributions from specific interactions between atoms, the non-α-β contributions can be removed by using, for example, lattice summations with known pairwise interaction potentials or some other method. In the simplest case, the total cohesive energy for a one-component system of A-type atoms corresponds to the cohesion

```
G ← [a_1, 1];  H ← F(a_1)
while length(H) > 1 do
    a_d ← f_{a,1}(H_{2,1})
    H^d ← F(a_d)
    m_d ← −H_{2,2} / H^d_{1,2}
    G.append([a_d, m_d])
    for i ∈ 1 to length(H^d) do
        H^d_{i,2} ← m_d · H^d_{i,2}
    H.append(H^d)
    H ← sort(H)
    H** ← H_1;  n ← 1
    for i ∈ 2 to length(H) do
        if |H**_{n,1} − H_{i,1}| > TOL then
            H**.append(H_i);  n ← n + 1
        else
            H**_{n,2} ← H**_{n,2} + H_{i,2}
            if H**_{n,2} = 0 then
                Remove H**_n;  n ← n − 1
    H ← H**
return G
```

Fig. 8 Generalized inversion method algorithm for a single point extraction of a pairwise interatomic potential. $\mathbf{H}^d$ and $\mathbf{H}^{**}$ are dummy arrays used in the **while** loop.

due only to homatomic A-A interactions in the lattice. The total cohesive energy is then identical to the α-β interaction, such that $\mathbb{E}^{tot}_{coh} = \mathbb{E}^{\langle A\cdot A\rangle}_{coh}$, and no modifications are necessary. Some DFT simulations provide the total energy *per unit cell*, and appropriate calculations to obtain the cohesive energy per atom have to be performed taking into account the structure of the lattice and the nature of the interaction. For instance, for a one-component FCC lattice, the total energy has to be divided by four to produce the cohesion *per atom*. Methods to isolate specific energy interactions from multicomponent systems require additional analysis of the cohesive energy and Sec. 6.2 provides additional examples.

The first lattice constant used in the inversion process is a_1, which corresponds to the location of the potential at $\tilde{y}_{1,1}$. Recalling that the subscript "d" accounts for increasing values of the lattice constant, *i.e.* $a_{d-1} < a_d < a_{d+1}$, the method first solves for a set of coefficients $\{(a_d, m_d)\}$ to be used with

$$\phi(\tilde{y}_{1,1}) = \frac{2}{\tilde{\mathbf{n}}_1(\tilde{y}_{1,1})}\sum_{d=1}^{N_d} m_d \cdot \mathbb{E}^{\langle\alpha\beta\rangle}_{coh,i}(a_d), \quad (34)$$

where N_d is the number of calls to the cohesive energy curve and the constants m_d are defined as the ratio of multiplicities assigned to each lattice constant a_d. This relationship in Eqn (34) should be clear from inspection of Eqn (31), as it is a direct result of replacing the potential values at any location.

Figure 8 describes the algorithm for the inversion process proposed in this work. We introduce additional functions and data structures to facilitate its implementation. The function $\mathbf{F}(a_d)$ returns an array, sorted by increasing displacements, containing all pairs of distances and multiplicities for a particular lattice constant a_d, such that

$$\mathbf{F}(a_d) = \left([\,\tilde{y}_{1,d}, \tilde{\mathbf{n}}_d(\tilde{y}_{1,d})\,],\ [\,\tilde{y}_{2,d}, \tilde{\mathbf{n}}_d(\tilde{y}_{2,d})\,], \ldots\right)$$
$$\forall \quad \tilde{y}_{k,d} \le r_{cut}. \quad (35)$$

All pairs of lattice constants and factors, $[\,a_d, m_d\,]$ used in Eqn (34) are stored in array **G**. Another array, **H**, is initially populated with $\mathbf{F}(a_1)$ and keeps track of the resulting distribution function, $\tilde{\mathbf{n}}_{res}(\cdot)$ and its corresponding set of values $\tilde{\mathbf{Y}}_{res}$ through each iteration.

We find the NNN after $\tilde{y}_{1,1}$ for each iteration of the algorithm in $\tilde{\mathbf{Y}}_{res}$ at location $\tilde{y}_{2,res}$ and use it create an new *in silico* lattice of size a_d with a first nearest-neighbor at $f_{\tilde{y},1}(a_d) = \tilde{y}_{2,res}$. We then multiply the distribution function $\tilde{\mathbf{n}}_d(\cdot)$ of the new lattice by the factor m_d such that the values of $\tilde{\mathbf{n}}_d(\tilde{y}_{1,d})$ and $\tilde{\mathbf{n}}_{res}(\tilde{y}_{2,res})$ are equal and opposite. Adding these two distributions removes the contribution of the shell at $\tilde{y}_{2,res}$ in the resulting lattice, modifying $\tilde{\mathbf{n}}_{res}(\cdot)$ and $\tilde{\mathbf{Y}}_{res}$. The array **H** stores the remaining locations of shells up to r_{cut} and their respective multiplicities generated after each iteration, and the elimination process is considered complete once **H** contains only the single pair $[\,\tilde{y}_{1,1}, \tilde{\mathbf{n}}_1(\tilde{y}_{1,1})\,]$. In Fig. 8, the notation $\mathbf{H}_1$ refers to the first index within the list, and $\mathbf{H}_{1,1}$ refers to the first element in the first index (*e.g.,* $\mathbf{H}_{1,1}(a_d)$ will output $\tilde{y}_{1,d}$). We also employ Eqn (25) to provide a relationship between the first nearest-neighbor distance and its corresponding lattice constant.

The basic algorithm of Fig. 8 allows for the solution of the array **G** containing all pairs of lattice constants and factors, $[\,a_d, m_d\,]$, generated in the process. Assuming the cohesive energy has been calculated for a range of values $a = [\,a_1, f_{a,1}(r_{cut})\,]$—see Eqn (19), the value of the pair-potential at $\tilde{y}_{1,1}$ is obtained by performing the following operation:

$$\phi(\tilde{y}_{1,1}) = \frac{2}{\tilde{\mathbf{n}}_1(\tilde{y}_{1,1})} \sum_{j=1}^{length(\mathbf{G})} \mathbf{G}_{j,2} \cdot \mathbb{E}_{coh,i}^{\langle\alpha\beta\rangle}(\mathbf{G}_{j,1}). \tag{36}$$

4.4 Graphical Interpretation

We further illustrate the application of the proposed inversion approach using a simple cubic lattice structure and the rigid octahedral system described in Fig. 5 to remove the second nearest-neighbor shells in each lattice structure.

4.4.1 Simple Cubic Lattice

The first and second nearest-neighbor distances in a simple cubic lattice are given by

$$f_{\tilde{y},1}^{\langle SC\rangle}(a_d) = a_d \qquad 1^{st}\ \text{Shell} \tag{37}$$

$$f_{\tilde{y},2}^{\langle SC\rangle}(a_d) = \sqrt{2}\cdot a_d \qquad 2^{nd}\ \text{Shell}. \tag{38}$$

To solve for the pair-potential at $\tilde{y}_{1,1}^{\langle SC\rangle} = 1$ Å with $r_{cut} = 12$ Å, the first lattice constant used to populate **H** is $a_1 = 1$ Å, and the elimination of the second shell requires the use of a lattice characterized by

$$a_2 = f_{a,1}^{\langle SC\rangle}\left(f_{\tilde{y},2}^{\langle SC\rangle}(a_1)\right) = \sqrt{2}\cdot a_1. \tag{39}$$

Figure 9(a) shows the $\tilde{\mathbf{n}}_1^{\langle SC\rangle}(\tilde{y}_{k,1})$ and $\tilde{\mathbf{n}}_2^{\langle SC\rangle}(\tilde{y}_{k,2})$ distributions along with their overlap in shell locations. This overlapping gives an indication of the symmetry of the lattice and how well the set $\tilde{\mathbf{Y}}^{\langle SC\rangle}$ forms a multiplicative semi-group. These are characterized by distances $y_m \cdot y_n = y_p$ for $m \neq n \neq p$, and it is this property of highly-symmetrical lattices which is exploited in the Chen-Möbius

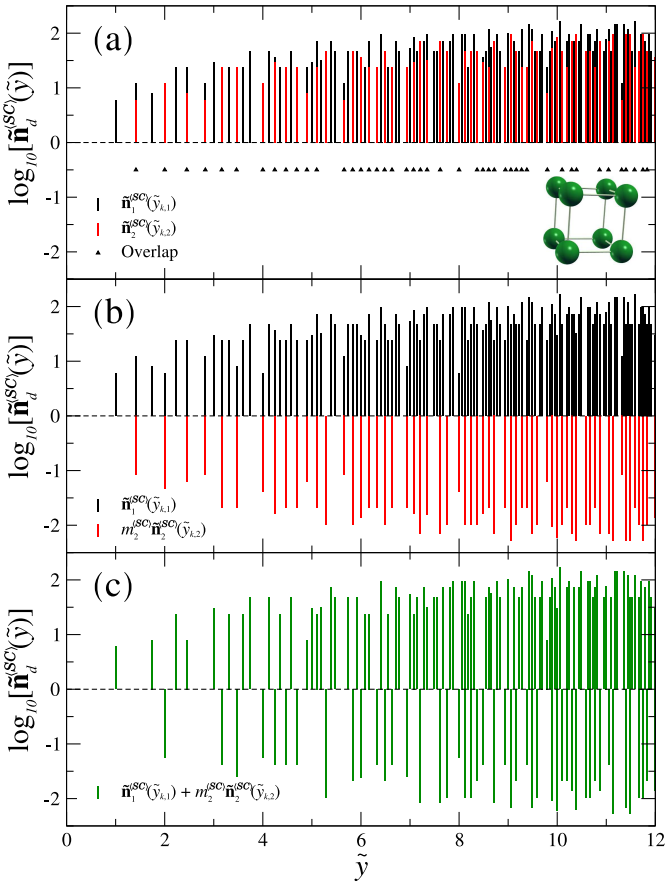

Fig. 9 (a) First two distributions for $\tilde{\mathbf{n}}_d^{\langle SC\rangle}$. Black and red bars represent atoms in spherical shells for systems with lattice constants $a_1 = 1$ Å and $a_2 = \sqrt{2}$ Å, respectively. Overlap between the locations of the two distributions are shown as triangles (▲). **(b)** The distribution $\tilde{\mathbf{n}}_2^{\langle SC\rangle}(\tilde{y}_{k,2}^{\langle SC\rangle})$ is multiplied by the factor $m_2^{\langle SC\rangle}$ to allow for the removal of the second shell in $\tilde{\mathbf{n}}_1^{\langle SC\rangle}(\tilde{y}_{k,1}^{\langle SC\rangle})$. **(c)** Resulting distribution after removal of the 2^{nd} nearest-neighbor shell obtained by summing the two distributions, $\tilde{\mathbf{n}}_{res}^{\langle SC\rangle} = \tilde{\mathbf{n}}_1^{\langle SC\rangle}(\tilde{y}_{k,1}^{\langle SC\rangle}) + m_2^{\langle SC\rangle}\cdot\tilde{\mathbf{n}}_2^{\langle SC\rangle}(\tilde{y}_{k,2}^{\langle SC\rangle})$.

inversion method. The overlap points loosely signify the number of cohesive energy values used to produce the pair-potential, such that a greater overlap requires fewer cohesive energy evaluations. As the lattice is expanded to create the new set $\tilde{\mathbf{Y}}_d$, if a majority of the points overlap in the distributions $\tilde{\mathbf{n}}_d(\cdot)$ and $\tilde{\mathbf{n}}_{res}(\cdot)$, then there is a greater likelihood that distant shells will also cancel when removing the contribution from the NNN at current lattice constant a_d.

The next step is to multiply the distribution $\tilde{\mathbf{n}}_2^{\langle SC\rangle}(\tilde{y}_{k,2}^{\langle SC\rangle})$ by the factor

$$m_2^{\langle SC\rangle} = -\frac{\tilde{\mathbf{n}}_1^{\langle SC\rangle}(\tilde{y}_{2,1}^{\langle SC\rangle})}{\tilde{\mathbf{n}}_2^{\langle SC\rangle}(\tilde{y}_{1,2}^{\langle SC\rangle})} \tag{40}$$

such that the two distributions have equal and opposite values at $\tilde{y}_{2,1}^{\langle SC\rangle} = \tilde{y}_{1,2}^{\langle SC\rangle}$, as shown in Fig. 9(b). Adding the two distributions removes the contribution from the second shell in a_1 and produces the resulting radial number distribution, illustrated in Fig. 9(c). The process continues iteratively by *(i)* creating a lat-

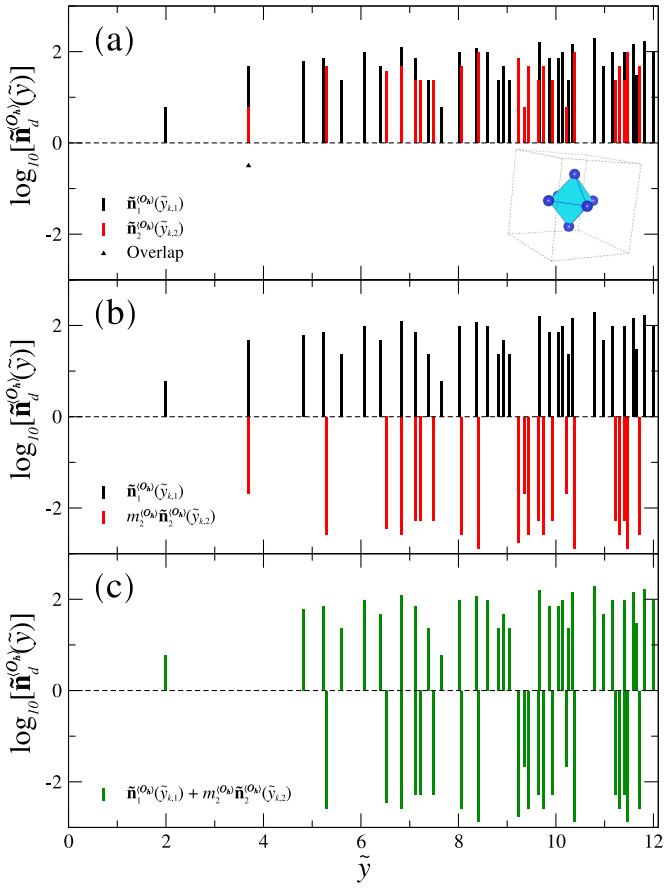

Fig. 10 (a) First two radial number distributions, $\tilde{\mathbf{n}}_d^{\langle O_h\rangle}$, in the unit cell. Black and red bars represent atoms in spherical shells for systems with lattice constants $a_1 = (2+2\sqrt{2})$ Å and $a_2 = f_{a,1}^{\langle O_h\rangle}(f_{\tilde{y},2}^{\langle O_h\rangle}(a_1))$, respectively. Only one point of overlap (▲) within r_{cut} is observed at $\tilde{y}_{2,1}(a_1)$. **(b)** The distribution $\tilde{\mathbf{n}}_2^{\langle O_h\rangle}(\tilde{y}_{k,2}^{\langle O_h\rangle})$ is multiplied by the factor $m_2^{\langle O_h\rangle}$ to allow for the removal of the second shell in $\tilde{\mathbf{n}}_1^{\langle O_h\rangle}(\tilde{y}_{k,1}^{\langle O_h\rangle})$. **(c)** The resulting distribution $\tilde{\mathbf{n}}_{res}^{\langle O_h\rangle}$ after the summation $\tilde{\mathbf{n}}_1^{\langle O_h\rangle}(\tilde{y}_{k,1}^{\langle O_h\rangle}) + m_2^{\langle O_h\rangle} \cdot \tilde{\mathbf{n}}_2^{\langle O_h\rangle}(\tilde{y}_{k,2}^{\langle O_h\rangle})$ and removal of the 2^{nd} nearest-neighbor shell.

tice with $a_d = f_{a,1}(\tilde{y}_{1,d})$, where $\tilde{y}_{1,d} = \tilde{y}_{2,res}$ is the NNN after $\tilde{y}_{1,1}$ in $\tilde{\mathbf{Y}}_{res}$, *(ii)* determining the factor m_d to cancel the contribution $\tilde{\mathbf{n}}_{res}(\tilde{y}_{1,d})$, and *(iii)* adding $m_d \cdot \tilde{\mathbf{n}}_d(\cdot)$ to $\tilde{\mathbf{n}}_{res}(\cdot)$, until all locations in the resulting distribution up to r_{cut} have been removed except the first nearest-neighbor $\tilde{y}_{1,1}^{\langle SC\rangle}$.

4.4.2 Rigid Octahedral Structure

For the octahedral system described in Fig. 5, the first two nearest-neighbors between separate octahedra depend upon the lattice constant and octahedral volumes since the lattice structure expands anisotropically during the calculation of the cohesive energy curve. The locations of the first two shells are

$$f_{\tilde{y},1}^{\langle O_h\rangle}(a_d) = a_d - \sqrt{2}\cdot L^{O_h} \qquad 1^{st}\ \text{Shell} \tag{41}$$

$$f_{\tilde{y},2}^{\langle O_h\rangle}(a_d) = \sqrt{a_d^2 - \sqrt{2}\cdot L^{O_h}\cdot a_d + (L^{O_h})^2} \qquad 2^{nd}\ \text{Shell} \tag{42}$$

for all $a_d > \sqrt{2}\cdot L^{O_h}$, since smaller lattice constants will produce interpenetrating octahedra. Setting $L^{O_h} = 2$ Å and calculating the cohesive energy in the range $a_d = [\,2+2\sqrt{2}, 12+2\sqrt{2}\,]$ Å, the inversion process produces inter-octahedral potentials for interatomic separation distances $\mathbf{r}_{ij} = [\,2,\ 12\,]$ Å, assuming the conditions given in Eqn (19) are met. Using $a_1 = (2+2\sqrt{2})$ Å as the first lattice constant allows for the solution of the interaction atomic potential at $\tilde{y}_{1,1}^{\langle O_h\rangle} = 2$ Å.

The radial number distributions for the first two shells in this system, $\tilde{\mathbf{n}}_1^{\langle O_h\rangle}(\cdot)$ and $\tilde{\mathbf{n}}_2^{\langle O_h\rangle}(\cdot)$, evaluated at $a_1 = (2+2\sqrt{2})$ Å and

$$a_2 = f_{a,1}^{\langle O_h\rangle}\left(f_{\tilde{y},2}^{\langle O_h\rangle}(a_1)\right), \tag{43}$$

are shown in Fig. 10(a). Removal of the contribution to the potential energy from atoms located at $\tilde{y}_{2,1}^{\langle O_h\rangle}$ with multiplicity $\tilde{\mathbf{n}}_1^{\langle O_h\rangle}(\tilde{y}_{2,1}^{\langle O_h\rangle})$ requires the next distribution $\tilde{\mathbf{n}}_2^{\langle O_h\rangle}(\cdot)$ to be multiplied by the factor

$$m_2^{\langle O_h\rangle} = -\frac{\tilde{\mathbf{n}}_1^{\langle O_h\rangle}(\tilde{y}_{2,1}^{\langle O_h\rangle})}{\tilde{\mathbf{n}}_2^{\langle O_h\rangle}(\tilde{y}_{1,2}^{\langle O_h\rangle})}, \tag{44}$$

to produce distributions of equal and opposite values at the second position of the current lattice distance set, $\tilde{\mathbf{Y}}_1^{\langle O_h\rangle}$. Figure 10(b) shows the distributions for $\tilde{\mathbf{n}}_1^{O_h}(\cdot)$ and $m_2^{O_h}\cdot\tilde{\mathbf{n}}_2^{O_h}(\cdot)$ which must be summed to eliminate the contributions from the second nearest-neighbors in $\tilde{\mathbf{Y}}_{res}^{\langle O_h\rangle}$. The result of adding the two distributions in Fig. 10(b) are shown in Fig. 10(c), characterized by a modified set $\tilde{\mathbf{Y}}_{res}^{\langle O_h\rangle}$ containing new distances $\tilde{y}_{k,res}^{\langle O_h\rangle}$ with positive and negative values for $\tilde{\mathbf{n}}_{res}^{\langle O_h\rangle}(\cdot)$. The iterative process continues to remove the points at NNN distances using the aforementioned steps until the only remaining terms, within a spherical radius of r_{cut}, are located at $\tilde{y}_{1,1}^{\langle O_h\rangle}$.

5 Test Cases

We first apply the proposed approach for cohesive energy inversions using in typical isometric lattice systems, which expand isotropically during the cohesive energy curve calculations. We also show results of pair-potentials extracted for systems that expand anisotropically, specifically the inter-octahedral boron-boron pair-potential of metal hexaboride frameworks.

Lattice energies for the hexaboride system are calculated with the integrated suite of open-source computer codes, QUANTUM ESPRESSO[22]. The code, based upon DFT, planewaves, and pseudopotentials, allows for the self-consistent solution of the Kohn-Sham equations for electron density distributions localized around atomic nuclei. We use ultrasoft pseudopotentials with Perdew-Burke-Ernzerhof exchange-correlation functionals and the generalized-gradient approximation. Kinetic energy and charge density cutoffs are set to 80 *Ry* and 960 *Ry*, respectively, and Marzari-Vanderbilt smearing functions have a width of 0.02 *Ry*. Integration over the Brillouin zone uses an $8\times8\times8$ Monkhorst-Pack grid of **k**-points. The convergence threshold for self-consistent field calculations is 10^{-6}.

5.1 Common Crystal Structures

We focus on the three cubic Bravais lattices: SC, FCC, and BCC. For the SC lattice, $y_k^{\langle SC\rangle}$ is described by Eqn (15). For a single atom-type system with origin at (0, 0, 0), the cohesive energy per

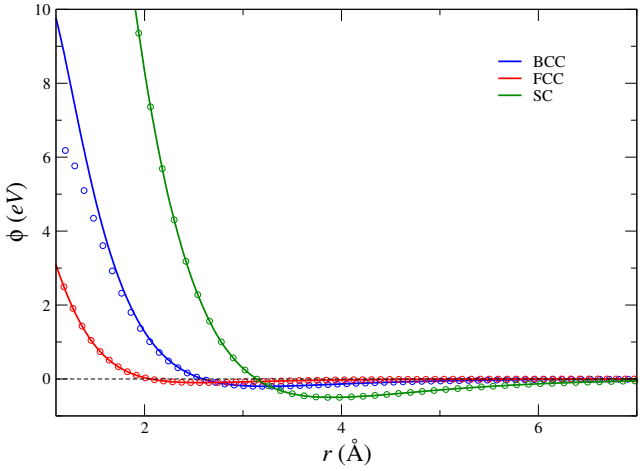

Fig. 11 Pair-potentials extracted from the inversion of cohesive energy described by Eqn (48) for the BCC, FCC, and SC lattice structures. Circles show the results produced by the Chen-Möbius inversion method[23].

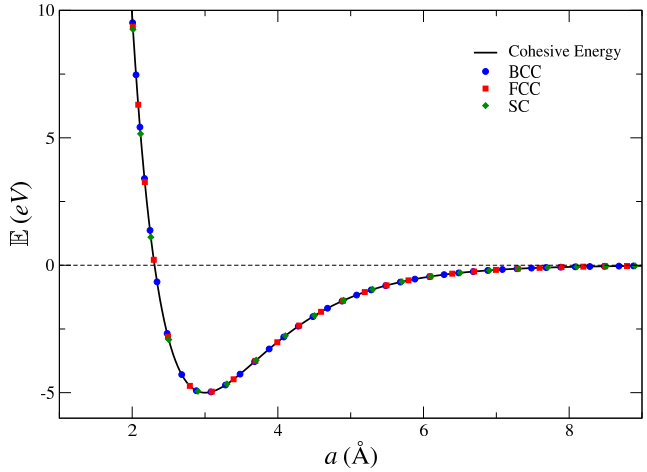

Fig. 12 Cohesive energy recalculation from extracted pair-potentials for the BCC, FCC, and SC lattice structures. Cohesive energy described by Eqn (48).

atom is given by

$$\mathbb{E}^{\langle SC\rangle}_{coh,i}(a) = \frac{1}{2}\sum_{(m,n,l)\neq(0,0,0)} \phi^{\langle SC\rangle}\left(\sqrt{m^2+n^2+l^2}\cdot a\right). \qquad (45)$$

For the FCC lattice, with normalized lattice vectors (0, 0, 0), (1/2, 1/2, 0), (1/2, 0, 1/2), and (0, 1/2, 1/2), the cohesive energy per atom is given by

$$\mathbb{E}^{\langle FCC\rangle}_{coh,i}(a) = \frac{1}{2}\sum_{(m,n,l)\neq(0,0,0)} \phi^{\langle FCC\rangle}\left(\sqrt{m^2+n^2+l^2}\cdot a\right) + \frac{3}{2}\sum_{(m,n,l)} \phi^{\langle FCC\rangle}\left(\sqrt{(m-1/2)^2+(n-1/2)^2+l^2}\cdot a\right). \qquad (46)$$

Similarly, for the BCC lattice with atoms at positions (0, 0, 0) and (1/2, 1/2, 1/2), the cohesive energy per atom is given by

$$\mathbb{E}^{\langle BCC\rangle}_{coh,i}(a) = \frac{1}{2}\sum_{(m,n,l)\neq(0,0,0)} \phi^{\langle BCC\rangle}\left(\sqrt{m^2+n^2+l^2}\cdot a\right) + \frac{1}{2}\sum_{(m,n,l)} \phi^{\langle BCC\rangle}\left(\sqrt{(m-1/2)^2+(n-1/2)^2+(l-1/2)^2}\cdot a\right). \qquad (47)$$

For illustration purposes, we assume that the cohesive energy *per unit cell* in a one-component isometric lattice can be described or fitted to a function of the form

$$\mathbb{E}^{\langle * \rangle}_{coh}(a) = \varepsilon\left\{\left(1-e^{-\kappa\cdot(a-a_{eq})}\right)^2 - 1\right\} \qquad (48)$$

where ε, a_{eq}, and κ are adjustable parameters, and a is lattice constant. The parameters used for the test cases are $\varepsilon = 5\ eV$, $\kappa = 1\ \text{Å}^{-1}$, and $a_{eq} = 3$ Å.

We use a radial cutoff of $r_{cut} = 12$ Å to generate the inverted potentials, after which atoms no longer interact. Using the algorithm of Fig. 8, with a convergence tolerance of 10^{-14}, Fig. 11 and Fig. 12 show the inverted potentials and the recalculated cohesive energies, respectively. For comparison, the Chen-Möbius

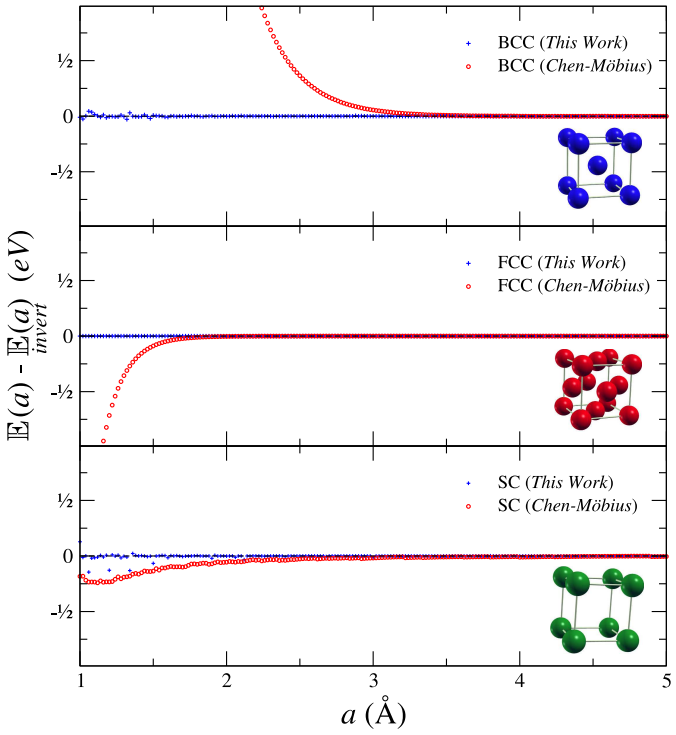

Fig. 13 Accuracy comparison of recalculated cohesive energies from lattice summations of pairwise potentials with $r_{cut} = 12$ Å for both inversions from this work and the Chen-Möbius method.

inversion method results are presented for the inverted potentials of the FCC, BCC, and SC lattice structures and denoted as circles in Fig. 11. The required inversion coefficients are from Chen[23]. Overall the pair-potentials produced by both methods are in very good agreement with the exception of the BCC system, where the the Chen-Möbius method diverges at lower values of the interatomic separation distance.

Calculation of the cohesive energy curve using the pair-potentials extracted for these systems allows further validation and comparison of the proposed approach. Figure 13 shows the

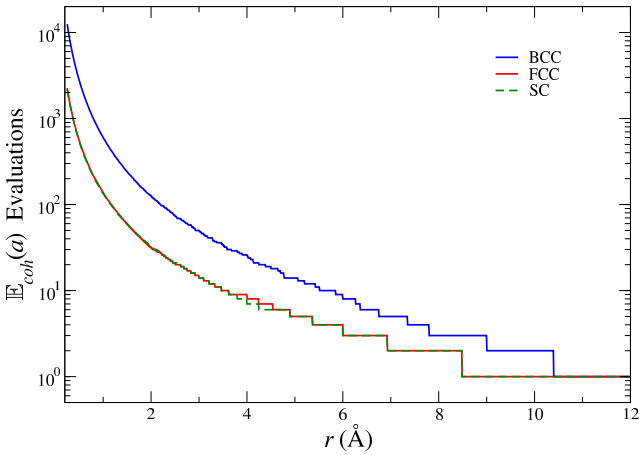

Fig. 14 Number of cohesive energy evaluations required for a convergence tolerance of 10^{-14} in the inversion of the cohesive energies. Radial cutoff set to $r_{cut} = 12$ Å for pairwise potentials.

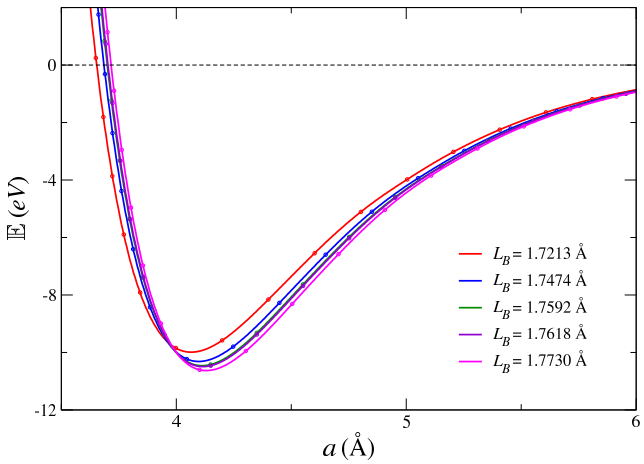

Fig. 15 Cohesive energies from DFT (solid lines) and lattice summations (circles) for rigid hexaboride units interacting inter-octahedrally. Octahedra side length denoted as L_B.

absolute error in computing the cohesive energies using both of the inversion methods. Again, overall both methods work quite well, although the proposed approach performs significantly better than the Chen-Möbius inversion for all three isometric lattices used. The reproduction of the cohesion is almost exact for the FCC lattice with the proposed method. This is expected, as the set of distances within the FCC lattice most-closely form a multiplicative semi-group. This means that the number of overlaps in shell locations are maximal in the FCC lattice in comparison to the other two, and fewer "new" points are added each time a new lattice is subtracted from the resultant distribution.

Inaccuracies from the Chen-Möbius inversion may have been due to an insufficient number of inversion coefficients. Small errors at lower separation distances using the proposed method may be due to computational inaccuracies, as many lattice structures are generated during the inversion process. Figure 14 shows the required number of evaluations of the cohesive energy function $\mathbb{E}_{coh,i}(a)$ for obtaining the pairwise potential from inversion of the cohesive energy as a function of the interatomic distance. These values represent the size of the array $\mathbf{G}$ generated in the inversion process—see Fig. 8 for details. For example, for $\mathbf{r}_{ij} = 0.25$ Å, the number of cohesive energy function evaluations range from close to 2200 for SC lattices to over 12,000 for the BCC system to generate the value of the pair-potential, although in practice one would rarely need such short interaction distance. Note that these values all converge to one when the pairwise distance represents a single nearest-neighbor for all the lattice structures.

5.2 Inter-octahedral Interatomic Boron Potential

The boron clusters shown in Fig. 5 are typical of metal hexaboride materials, which we use as an *in silico* structure to evaluate the inter-octahedral boron-boron interaction potential. For a lattice constructed of discrete hexaboride units with side lengths $L^{O_h} = L_B$, the six boron atoms are located at $(1/2 \pm \gamma_d^{O_h},\ 1/2,\ 1/2)$, $(1/2,\ 1/2 \pm \gamma_d^{O_h},\ 1/2)$, and $(1/2,\ 1/2,\ 1/2 \pm \gamma_d^{O_h})$, where $\gamma_d^{O_h}$ is defined by Eqn (17). The set $\tilde{\mathbf{Y}}_d^{\langle B\cdot B\rangle}$ is given by Eqn (16) and the cohesive energy per atom within this system interacting inter-octahedrally is given by

$$\begin{aligned}\mathbb{E}_{coh,i}^{\langle B\cdot B\rangle}(a) &= \frac{1}{2}\sum_{(m,n,l)\neq(0,0,0)} \phi^{\langle B\cdot B\rangle}\left(\sqrt{m^2+n^2+l^2}\cdot a\right)\\ &+ \frac{1}{2}\sum_{(m,n,l)\neq(0,0,0)} \phi^{\langle B\cdot B\rangle}\left(\sqrt{(m+2\gamma_d^{O_h})^2+n^2+l^2}\cdot a\right)\\ &+ 2\times\sum_{(m,n,l)\neq(0,0,0)} \phi^{\langle B\cdot B\rangle}\left(\sqrt{(m+\gamma_d^{O_h})^2+(n+\gamma_d^{O_h})^2+l^2}\cdot a\right).\end{aligned} \tag{49}$$

DFT is used to calculate the cohesive energy of the hexaboride framework in the range $L_B = [\,1.7213, 1.7730\,]$ Å, which contains dimensions typical of metal hexaborides crystalline structures. The anisotropic expansions of the boron framework are in the range of $a_d = [\,3.2, 11.0\,]$ Å, for which the energy variations beyond the upper limit are $\partial\mathbb{E}/\partial a < \pm 0.0002\ eV/$Å. To estimate the isolated octahedral energy, $\phi^{\circ}_{B_6}$, we use the energy evaluation at the largest lattice constant, which corresponds to the energy of a single octahedral unit interacting locally via intra-octahedral bonds described by the same two nearest shells at $\tilde{y}^{\langle B\cdot B\rangle} = L_B$ and $\tilde{y}^{\langle B\cdot B\rangle} = \sqrt{2}\cdot L_B$. Figure 15 shows the cohesive energies for various values L_B after subtraction of $\phi^{\circ}_{B_6}$ along with the predictions from lattice summations of pair-potentials obtained from the inversion process, which are in excellent agreement with the DFT calculations. Figure 16 contains the resulting pair-potentials obtained using a radial cutoff of 9 Å and a convergence tolerance of 10^{-14}.

The cohesive energy curves of Fig. 15 show the effect of octahedral volumes on the potential energy of the system. Void of cation interactions, the octahedra seem to favor larger volumes when placed in a simple cubic lattice. Additionally, the shifting of the minima observed in Fig. 15 shows how the equilibrium of the simple cubic lattice is affected upon octahedral volume expansions, a known phenomenon that occurs[11] at very fast time-scales. As the octahedral units expand, Fig. 16 shows deepening of the attractive wells of the interatomic pair-potentials making these bonds to become energetically favorable. These octahedral units typically

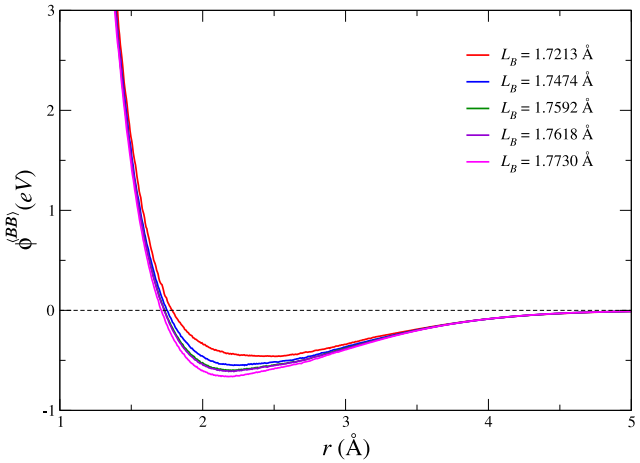

Fig. 16 Inverted inter-octahedral boron-boron potential for hexaboride clusters characterized by L_B.

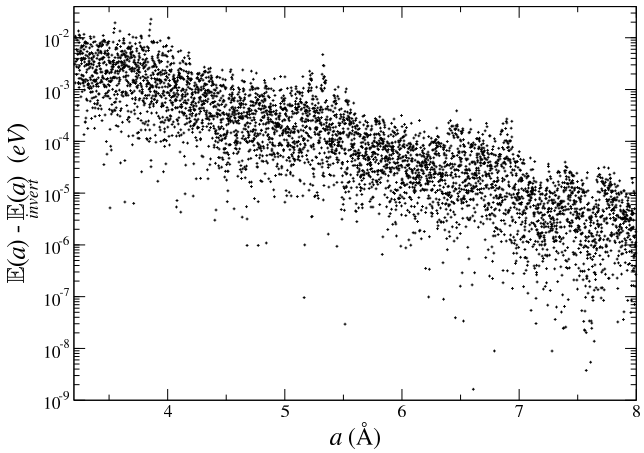

Fig. 17 Error absolute value between predicted cohesive energies from inverted potentials from this work and DFT calculations for hexaboride systems.

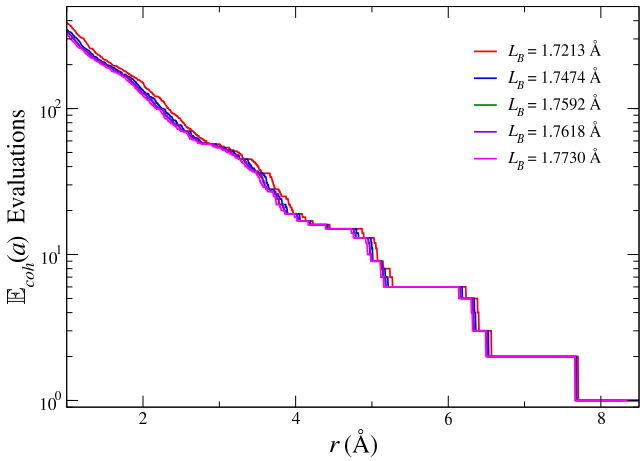

Fig. 18 Number of cohesive energy evaluations required for a convergence tolerance of 10^{-14} in the inversion of the cohesive energies of inter-octahedral boron-boron interactions, $\phi^{\langle B \cdot B \rangle}(r)$. Radial cutoff set to $r_{cut} = 9$ Å for pairwise potentials.

form binary compounds with metals capable of electron donation to stabilize polarized regions, effect that is not taken into account in these calculations. Compression of the octahedra may require a larger electron density to be centrally localized in the hexaboride to balance the net positive charge of the cluster, pulling the bonding potentials from inter-octahedral sites.

Figure 17 presents further analysis on the ability of the inverted pair-potentials reproducing the correct cohesive energy. The figure contains the absolute value of the energy differences between the cohesive energy obtained from DFT and that of lattice summations of pair-potentials for all the L_B values. Errors are the largest at the lower lattice constants where the greatest number of cohesive energy evaluations is required. For example, for a lattice constant of $a = 3.2$ Å, there is an error of approximately 0.005 eV, but the magnitude of the cohesive energy at this lattice constant is on the order of 100 eV, making the relative error rather small. The error increases with increasing number of cohesive energy evaluations, suggesting these may be due to computational error. Overall, the prediction of the cohesive energy from the inverted pair-potentials is very good for the region sampled. Further, the error becomes zero when the lattice constant reaches a value where only a single shell is required to generate the potential and no inversion is necessary. For reference, this value is located at

$$a_{SS}^{\langle B \cdot B \rangle} = \frac{\sqrt{2} \cdot L_B + \sqrt{4 \cdot r_{cut}^2 - 2 \cdot L_B^2}}{2}, \tag{50}$$

where $a_{SS}^{\langle B \cdot B \rangle}$ is the lattice constant having the pair-potential specified by a single shell.

The number of cohesive energy evaluations necessary for the accurate evaluation of the pair-potentials is an important parameter that affects the computational cost of the inversion method. Figure 18 shows this number as a function of the interatomic separation distance for all the values of L_B used in this case study. All lattices require similar numbers of cohesive energy values to produce the potential $\phi^{\langle B \cdot B \rangle}$ for a given interatomic distance, r. The location r where the number of shells or cohesive energy values required becomes one can be determined from a_{SS} by subtracting the longest length across an octahedron to determine the nearest-neighbor distance, given by

$$r_{SS}^{\langle B \cdot B \rangle} = a_{SS}^{\langle B \cdot B \rangle} - \sqrt{2} \cdot L_B. \tag{51}$$

The small deviations in the number of cohesive energy evaluations among the lattices in Fig. 18 are due to the differences in the first nearest-neighbor distances and the level of symmetry of each of the lattices, which affects the efficiency of the Gaussian elimination in the inversion process.

6 Additional Remarks

6.1 Computational Considerations

Several factors can potentially affect the convergence of the inversion process. These include the quality of the DFT calculations, which should be well-converged with proper basis sets and pseudopotentials. The number of values used to populate the cohesive energy should be sufficient to generate a smooth curve, with a finer mesh generally required near equilibrium configurations

where changes in the derivative of cohesive energy with distance are more pronounced. Smoothing of the cohesive energy may be desirable, as converged solutions of electronic structure calculations may not approach zero for the cohesive energies at lattice constants corresponding to r_{cut} and to avoid discontinuities in the first derivative arising from oscillating values, which can also be reflected in the interatomic potential calculations.

Cutoff values play an important role in the inversion process. The range of applicability of the pair-potential is constrained to the range of lattice constants sampled on the cohesive energy curve. Since the value of the pair-potential at $\tilde{y}_{1,1}$ always requires a minimum $\mathbb{E}_{coh,i}$ value located at $f_{a,1}(\tilde{y}_{1,1}) = a_1$, the range of applicability of the pair-potential is limited to $[\, f_{\tilde{y},1}(a_i), f_{\tilde{y},1}(a_f)\,]$ for the range of lattice constants $[\, a_i,\, a_f\,]$, where typically $r_{cut} = f_{\tilde{y},1}(a_f)$. Additionally, inversions of long-tailed cohesive energy curves also produce long-tailed pair-potentials. For materials with screened interactions where only the first few "*bonded*" neighbors are included, modifications to the cohesive energies as proposed by Bazant[15] may be appropriate to avoid the incorporation of non-crystalline behavior into the cohesive energy curve prior to inversion.

As a final note, pairwise potentials are appropriate for systems composed of atoms with spherically symmetric interactions, and the interatomic potentials extracted with the proposed approach should work very well for these types of systems. Departure from spherical bonding symmetry requires the use of different interatomic potential energy models more inline with the nature of the bonding. For the case of metal hexaborides, we treat the octahedral units as rigid bodies constraining the interactions such that these have an asymmetrical force distribution, while at the same time reducing the complexity of the molecular model significantly.

6.2 Extension to Multicomponent Systems

Extension of the proposed inversion method to multicomponent systems is very similar to mono-atomic systems as long as the geometrical relationships of the different interatomic interactions can be established and their respective energy contributions isolated. Additionally, electrostatic contributions have to be handled separately. To illustrate the application of method, we use a binary system of atom-types *A* and *B* with a known lattice geometry. In this case, we are interested in extracting the *A*-*B* interatomic potential by inversion of cohesive energies. The total energy of the system is described by

$$\mathbb{E}_{tot} = \phi^{A\circ} + \phi^{B\circ} + \mathbb{E}_{coh}^{\langle A\cdot A\rangle} + \mathbb{E}_{coh}^{\langle A\cdot B\rangle} + \mathbb{E}_{coh}^{\langle B\cdot B\rangle} + \mathbb{E}_{elec}, \tag{52}$$

where $\mathbb{E}_{elec}$ is the energy due to electrostatics, $\phi^{A\circ}$ and $\phi^{B\circ}$ are atomic background energies for species *A* and *B*, and the terms $\mathbb{E}_{coh}^{\langle\cdot\rangle}$ are the contributions of the atomic species and their interactions to the total energy. To extract the $\phi^{\langle A\cdot B\rangle}(r)$ interaction potential, we need to isolate $\mathbb{E}_{coh}^{\langle A\cdot B\rangle}$ from the cohesive energy of the system

$$\mathbb{E}_{coh}^{\langle A\cdot B\rangle} = \mathbb{E}_{coh} - \left(\mathbb{E}_{coh}^{\langle A\cdot A\rangle} + \mathbb{E}_{coh}^{\langle B\cdot B\rangle} + \mathbb{E}_{elec}\right) \tag{53}$$

for a range of lattice constants before performing the inversion process. A possible way to achieve this is to determine the energy contributions of the three terms in parentheses in Eqn (53). For example, the two homatomic energies can be calculated through lattice summations using inversions derived from bulk phases of *A* and *B*, if appropriate. The calculation of the electrostatics contribution requires *a priori* knowledge of charge and dielectrics in the material or inferred from chemical theory or other approaches. Then, a simple Coulomb potential or Ewald summation can be used to determine the $\mathbb{E}_{elec}$ contribution.

Once $\mathbb{E}_{coh}^{\langle A\cdot B\rangle}(a)$ is determined, an expression for this cohesive energy contribution in terms of heteroatomic pair-potentials is needed to proceed with the inversion process. The cohesive energy of atom "*i*" from the *A*–*B* interactions is given by

$$\mathbb{E}_{coh,i}^{\langle A\cdot B\rangle}(a_d) = \sum_k \tilde{\mathbf{n}}_d^{\langle A\cdot B\rangle}(\tilde{y}_{k,d}^{\langle A\cdot B\rangle}) \cdot \boldsymbol{\phi}^{\langle A\cdot B\rangle}(\tilde{y}_{k,d}^{\langle A\cdot B\rangle}), \tag{54}$$

where the one half factor is omitted since no overcounting occurs when adding up the heteroatomic interactions. The multiplicity and distance sets for the system are easily calculated for any lattice constant a_d using the lattice structure, and the inversion of the heteroatomic cohesive energy curve is performed using the algorithm of Fig. 8.

An alternative approach is to isolate the energy contribution of the *A*-*B* interactions by making use of multiple *in silico* lattice geometries or structures. Recent work by Chen *et al.*[24] demonstrates that combining the energetics of different *in silico* crystals can be used to isolate the contribution to the cohesive energy from an *A*-*B* type of interaction. Although the lattice structures are artificial, they are constructed in such a way that the radial particle number distributions for all the pairs different from *A*-*B* are identical. In this way, when subtracting the energies of two or more of these *in silico* structures, the resulting energy is only a function of the *A*-*B* interactions. For example, for two lattice geometries λ_1 and λ_2 characterized by

$$\tilde{\mathbf{n}}_{d,\lambda_1}^{\langle A\cdot A\rangle}(\cdot) = \tilde{\mathbf{n}}_{d,\lambda_2}^{\langle A\cdot A\rangle}(\cdot), \tag{55}$$

and

$$\tilde{\mathbf{n}}_{d,\lambda_1}^{\langle B\cdot B\rangle}(\cdot) = \tilde{\mathbf{n}}_{d,\lambda_2}^{\langle B\cdot B\rangle}(\cdot), \tag{56}$$

each of these homatomic interactions will produce the same terms in Eqn (53). Taking the difference of the cohesive energies for λ_1 and λ_2 removes the contributions from the *A*-*A* and *B*-*B* interactions. A modified set of distances $\tilde{\mathbf{Y}}_{d,\lambda_1\cdot\lambda_2}^{\langle A\cdot B\rangle}$ containing *A*-*B* separations from combining the λ_1 and λ_2 structures produces a modified radial number distribution function $\tilde{\mathbf{n}}_{d,\lambda_1\cdot\lambda_2}^{\langle A\cdot B\rangle}(\cdot)$ that can be used in the inversion algorithm for the $\phi^{\langle A\cdot B\rangle}(r)$ extraction. Note that if electrostatics are prevalent in the system, these will also have to either be included in the inverted potential or calculated prior to the inversion of the cohesive energies for each structure.

7 Conclusions

The method proposed for inversion of cohesive energy curves and extraction of interatomic pair-potentials adds on significantly to current approaches. A larger set of crystalline systems and structures can be studied using this approach. As illustrated in the test

cases, the method produces pair-potentials capable of reproducing cohesive energies very accurately with the ability to handle multiple types of systems and interactions. For crystalline materials where current methods work well, the proposed inversion framework produces results with the same level accuracy or better. Depending on the level of accuracy, the method can be computationally demanding due to the cohesive energy evaluations required from DFT calculations. The method has potential for automated generation of pair-potentials in combination with DFT codes and automatic lattice structures generation.

8 Acknowledgments

We gratefully acknowledge funding support from NSF Award No. 1246792 ("SNM: Scalable Manufacturing of Unique Hexaboride Nanomaterials for Advanced Energy Generation and Gas Storage Applications"). A portion of the figures were created in part using VESTA[25].

References

1 J. Cai, X. Hu and N. Chen, *J. Phys. Chem. Solids*, 2005, **66**, 1256–1263.

2 S. Liu, S. Shi, H. Huang and C. Woo, *J. Alloys Compd.*, 2002, **330–332**, 64–69.

3 J. Guo, J. Shen and N. Chen, *Chem. Phys.*, 2006, **324**, 314–322.

4 W. qing Zhang, Q. Xie and X.-J. Ge, *J. Appl. Phys.*, 1997, **82**, 578–582.

5 W. Jain, Z. Kaiing and X. Xide, *J. Phys.: Condens. Matter*, 1994, **6**, 989–996.

6 Y. Liu, Y. Kang and N. Chen, *J. Alloys Compd.*, 2003, **349**, 17–22.

7 X. Duan, B. Zhou, R. Chen, H. Zhou, Y. Wen and B. Shan, *Curr. Appl. Phys.*, 2014, **14**, 1794–1802.

8 X. Duan, B. Zhou, Y. Wen, R. Chen, H. Zhou and B. Shan, *Comput. Mater. Sci.*, 2015, **98**, 417–423.

9 N.-X. Chen and G. bao Ren, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1992, **45**, 8177–8180.

10 M. Allen and D. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York, 1987.

11 H. Smith, G. Dolling, S. Kunii, M. Kasaya, B. Liu, K. Tekegahara, T. Kasuya and T. Goto, *Solid State Commun.*, 1985, **53**, 15–19.

12 K. Schmidt, O. Graeve and V. Vasquez, *J. Phys. Chem. C*, 2015, **119**, 14288–14296.

13 C. Kittel, *Introduction to Solid State Physics*, John Wiley & Sons, New York, 2005.

14 A. Carlsson, C. Gelatt and H. Ehrenreich, *Philos. Mag. A*, 1980, **41**, 241–250.

15 M. Bazant and E. Kaxiras, *Phys. Rev. Lett.*, 1996, **77**, 4370–4373.

16 C.-H. Chen, T. Aizawa, N. Iyi, A. Sato and S. Otani, *J. Alloys Compd.*, 2004, **366**, L6–L8.

17 B. S. Jabes, H. Yadav, S. Kumar and C. Chakravarty, *J. Chem. Phys.*, 2014, **141**, 154904.

18 R. van Zon, *J. Stat. Mech.: Theory Exp.*, 2009, **2009**, P02008.

19 C. Lee and C. Hua, *J. Chem. Phys.*, 2010, **132**, 224904.

20 H. Su-Zhen, M. Holger and W. Chen-Xu, *Chin. Phys. B*, 2014, **23**, 048201.

21 D. Heine, M. Petersen and G. Grest, *J. Chem. Phys.*, 2010, **132**, 184509.

22 P. Giannozzini, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. Chiarotti, M. Cococcioni, I. Dabo, A. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. Seitsonen, A. Smogunov, P. Umari and R. Wentzcovitch, *J. Phys.: Condens. Matter*, 395502, **21**, 2009.

23 N.-X. Chen, *Möbius Inversion in Physics*, World Scientific, New Jersey, 2010.

24 S. Zhang and N. Chen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2002, **66**, 064106.

25 K. Momma and F. Izumi, *J Appl Crystallogr*, 2011, **44**, 1272–1276.