Weighted multibody expansions for computing stable structures of multiatom systems

Veera Sundararaghavan* and Nicholas Zabaras†

Materials Process Design and Control Laboratory, Sibley School of Mechanical and Aerospace Engineering, 188 Frank H. T. Rhodes Hall, Cornell University, Ithaca, New York 14853-3801, USA

(Received 5 April 2007; published 5 February 2008)

The effect of structural relaxations in alloys is described using a multibody energy expansion formalism. *N*-body potentials in the multibody expansion are computed from energies of isolated clusters, which, in turn, are calculated from empirical potentials or self-consistent quantum mechanical calculations. Convergence characteristics of multibody expansions (MBEs) are improved by weighting energies obtained from various truncations of many-body expansions in a method called weighted MBE (WMBE). It is shown that multibody expansions of many-atom systems can be efficiently constructed using interpolation of isolated cluster energies from databases. In contrast to the method of cluster expansion, WMBE focuses on positional degrees of freedom and, hence, explicitly handles structural relaxations during computations of stable atom clusters and periodic or amorphous phase structures.

DOI: 10.1103/PhysRevB.77.064101 PACS number(s): 64.60.Cn, 65.40.−b, 61.66.Dk, 05.50.+q

I. INTRODUCTION

The calculation of stable structures of alloys, clusters, surfaces, and molecules from first principles is an important step towards the design of materials with exceptional properties. Identification of stable alloy phases aids in the construction of phase diagrams from first principles. Because of the immense variety of phase structures, identification of stable structures at different combinations of the alloying elements is a nontrivial problem. While a first-principles approach based on density functional theory (DFT) provides a rigorous way for calculating the formation energies of phase structures, the computational complexity of performing fully relaxed calculations over the entire set of possible phase structures makes this method prohibitive. Techniques such as cluster expansion[1–7] and, more recently, data mining techniques[8,9] allow one to accelerate the search for stable phase structures.

In cluster expansion methods (CEMs),[1–4] the relaxed energy of an atomic structure is represented as a linear combination of the characteristic energies of clusters of atoms over a fixed lattice. The coefficients in the cluster expansion are computed using relaxed DFT energy calculations of a few prototype structures.[1] This method includes only ordering degrees of freedom as provided by different possible arrangements of atom types on a fixed parent lattice. Consequently, CEMs fail in cases where the alloy phases have complex structures that are different from the superstructures of the underlying parent lattice (for example, fcc or bcc lattices) and exhibit convergence issues in cases where structural relaxation effects are dominant[5,6] (for example, in alloys involving constituents with large size differences).

In another technique called multibody expansion (MBE), *N*-body potentials (or, otherwise, cluster potentials[10]) constructed from *ab initio* calculations are used to describe the energies of arbitrary atomic structures as a function of atom positions. The total energy is represented as a summation over potentials of underlying isolated atom clusters in the structure, with series terms involving pair, three-body, four-body,…, *N*-body potentials. Up to third-order truncations of multibody expansions have been previously used in related empirically derived potentials: namely, the Gupta[11] and Murrell-Mottram[12–15] (MM) potentials. Multibody potentials focus on positional degrees of freedom and, hence, explicitly handle structural relaxations during computations of stable phase structures. Structural relaxation effects can be treated in a cluster expansion approach by combining it with position-dependent potentials in the form of a hybrid cluster expansion.[6] Another method combining CEMs and multibody potentials was proposed recently for introducing positional degrees of freedom in a more generalized cluster expansion.[16,17] However, building the *N*-body potentials from atomistic calculations is quite a challenging problem. First, the number of clusters (and thus the number of cluster energies that need to be calculated) in an *N*-atom system increases geometrically with the order of expansion (Fig. 1). Second, although the approach provides good convergence for rare-gas crystals, convergence of MBEs is not smooth for metallic crystals.[18,21] The absence of smooth convergence does not allow the establishment of a hard cutoff for the series terms. Consequently, there have been no published reports of a multibody expansion constructed directly from first-principle calculations.

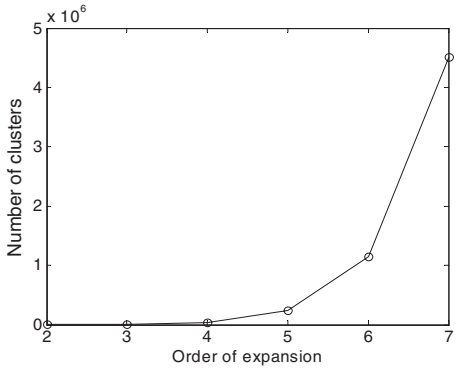

FIG. 1. Increase in the total number of clusters involved in a multibody expansion for a 32-atom system with the order of expansion.

 ©2008 The American Physical Society

This paper addresses these drawbacks by proposing a multibody expansion with weighted terms. Using this technique, it becomes possible to accurately compute the energies of N-atom systems from knowledge of the small cluster energies computed from first principles. The efficiency of this method, called weighted multibody expansion (WMBE), is emphasized through examples in this work. Another contribution of this paper is a formal technique to rapidly calculate multibody expansions using linear interpolation over tessellated cluster configurational spaces. The MBE using interpolated energies is several orders of magnitude faster than using DFT calculations, since cluster energies are computed beforehand and are directly sampled from the database when computing the multibody expansion.

II. MULTIBODY EXPANSION METHODOLOGY

Consider, for instance, a configuration of M atoms (possibly all different), whose energy we intend to compute. We denote the total energy of this M-particle system using $E_P = E_P(X_1, X_2, \ldots, X_M)$, where P is the order of the expansion used and the position $\mathbf{R}_n$ of atom n is grouped together with the species of atom n denoted by an integer σ_n, $X_n = (\mathbf{R}_n, \sigma_n)$. As the order of labeling the M atoms is arbitrary, the form of $E_P(X_1, \ldots, X_i, \ldots, X_j, \ldots, X_M)$ must be symmetric with respect to an interchange of X_i and X_j.

From here on, we denote M as the total number of atoms in the system; $N = 1, 2, \ldots, P$ denotes an N-atom cluster within the M-atom system. Further, $L = 1, 2, \ldots, N$ denotes an arbitrary L-atom cluster within an N-atom cluster. The energy E_P of an M-particle system is represented as a summation over a series of N-body interaction potentials $V^{(N)}$ via

$$E_P(X_1, X_2, \ldots, X_M) = \sum_{N=1}^{P} E^{(N)}(X_1, X_2, \ldots, X_M),$$

$$E^{(N)} = \sum_{m_1=1}^{M} \sum_{m_2=m_1+1}^{M} \cdots \sum_{m_N=m_{N-1}+1}^{M} V^{(N)}(X_{m_1}, X_{m_2}, \ldots, X_{m_N}). \tag{1}$$

The potentials can be inverted via the Mobius inversion approach from number theory. Mobius inversion has been used previously for the extraction of potentials from energy data by Chen and Ren[19,20] although in a different context. In the case of multibody potentials $V^{(N)}$, the Mobius inversion is given as[16]

$$V^{(N)}(X_1, X_2, \ldots, X_N) = \sum_{L=1}^{N} (-1)^{N-L} \sum_{m_1=1}^{N} \sum_{m_2=m_1+1}^{N} \cdots \sum_{m_L=m_{L-1}+1}^{N} E^*(X_{m_1}, X_{m_2}, \ldots, X_{m_L}). \tag{2}$$

Here, we denote the energies of L-atom clusters within the N-atom cluster as E^*. The above equation constitutes a unique definition of N-body potentials $V^{(N)}$ which are structure independent because this equation does not carry any information about the environment of the atom clusters.[16] $V^{(2)}(X_i, X_j)$ can be understood as the excess energy attributed to pair interactions in an isolated atom pair i, j—i.e., $V^{(2)}(X_i, X_j) = E^*(X_i, X_j) - E^*(X_i) - E^*(X_j)$. Similarly, $V^{(3)} \times (X_i, X_j, X_k)$ can be understood as the excess energy attributed to three-body interactions in an isolated trimer (i, j, k):

$$V^{(3)}(X_i, X_j, X_k) = E^*(X_i, X_j, X_k) - [V^{(2)}(X_i, X_j) + V^{(2)}(X_j, X_k) + V^{(2)}(X_i, X_k)] - [E^*(X_i) + E^*(X_j) + E^*(X_k)]. \tag{3}$$

Once the potentials $V^{(N)}$ have been constructed, they can be used to calculate the energy $E_P(X_1, X_2, \ldots, X_M)$ for an M-atom system using Eq. (1). The first critical requirement of the technique is knowledge of the complete energy surface (cluster energies versus atom positions and types) of small isolated clusters of atoms $(E^*, L = 1, \ldots, 5)$ for building the potentials in Eq. (2). Second, it is essential that the expansion converge within a small order of expansion (i.e., $P \leqslant 5$) for computational efficiency. These two aspects are addressed in the next two sections. A complete energy surface for small isolated clusters is created by mathematically defining the configurational space of clusters, tessellation of the space, and computation of cluster energies on nodal points, followed by interpolation of cluster energies as described in the next section. Computational efficiency is improved through the use of weighted multibody expansions as explained in Sec. IV. Efficiency can be further improved by performing computations in parallel by distributing the M atoms involved in the loop index m_1 in Eq. (1) to different processors.

III. CONSTRUCTION OF CLUSTER ENERGY SURFACES

The basic idea of the approach to rapidly compute multibody expansions of arbitrary systems is to build an interpolation function for the isolated cluster energies E^* from the precomputed database. Given a set of n m-atom clusters represented as $\Theta = \{\boldsymbol{\xi}_d^i\}_{i=1}^n$ in d-dimensional configurational space, we try to build a smooth function that maps clusters to *ab initio* energies, $f: \mathbb{R}^d \rightarrow \mathbb{R}$. In particular, we use an interpolant $\mathcal{I}f$ such that $\mathcal{I}f(\boldsymbol{\xi}_d^i) = f(\boldsymbol{\xi}_d^i)$, $\forall i = 1, \ldots, n$.

The first step in this procedure will be to define the d-dimensional configurational space of an m-atom cluster. The positions of the atoms in the cluster are represented by

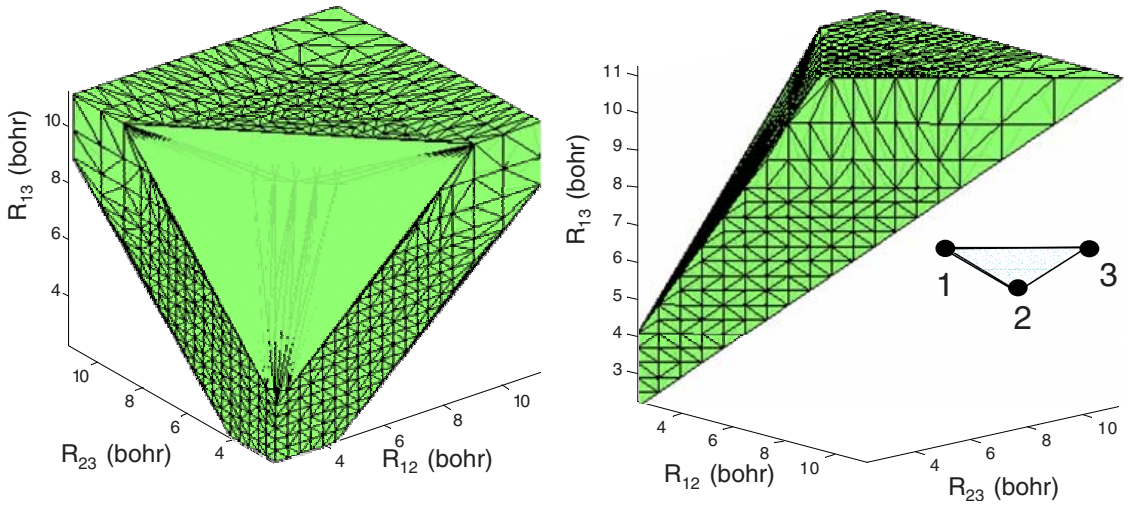

FIG. 2. (Color online) Left: the space of all possible three-atom clusters within an upper and lower cutoff cluster size. This space represents a convex hull in three dimensions (3D). Right: use of symmetries (in the case where all three atoms are of one type—e.g., Pt-Pt-Pt clusters) can further reduce the space. The simplices used to perform local linear interpolation of energies are also shown. In 3D, the simplex is a tetrahedron.

the distance between atoms, $R_{ij}>0$. For two-atom clusters ($m=2$), the configurational space is one dimensional, with each point x in the space representing a two-atom cluster with interatomic distance of $R_{12}=x$. As the number of atoms, m, in the cluster increases, the number of distances, R_{ij}, necessary to completely and uniquely describe the cluster increases rapidly. Up to $m\leqslant 4$, clusters are uniquely represented by $\frac{1}{2}m(m-1)$ independent variables.

For example, the space of all possible three-atom clusters is three-dimensional as shown in Fig. 2(a). This space is a convex hull with nine planes (symmetries not included) due to a linear set of constraints arising from three triangle inequalities of the form $R_{ij}+R_{jk}\geqslant R_{ik}$ that constrain the location of atoms in the three-atom cluster and the upper and lower cutoffs used for possible cluster sizes in the database: $R_{ij}>l$ and $R_{ij}<u$ with $i,j,k=1,\ldots,3$. Cluster symmetries can be used to further reduce the space and, consequently, reduce the number of energy calculations required. Figure 2(b) shows the reduced space accounting for symmetries ($R_{12}\leqslant R_{23}\leqslant R_{13}$) in the case where all three atoms are of the same type. Also shown in Fig. 2(a) is the tessellation of the configurational space of clusters. The energy of a cluster corresponding to each nodal point in the space is calculated and stored in the database. The plot of energy versus interatomic distance for a two-atom Pt cluster is shown in Fig. 3 with the location of nodal points for two-atom clusters. Higher-dimensional spaces are adaptively tessellated as shown in Fig. 2(a) with a finer discretization of regions involving small clusters. The tessellation of the configurational space is carried out using n-dimensional Delaunay triangulation, as implemented in the QHULL program.[22] Tessellation generates elements (known as a simplex) over which local linear interpolation is carried out to find the energy of any other three-atom cluster within the space. The discretization and interpolation techniques are, in essence, the same as those used in the popular finite-element techniques for partial differential equations (PDEs). The energy (E) of an arbitrary cluster with cluster specifier $\boldsymbol{\xi}_d^*=[\xi_1^*,\xi_2^*,\ldots,\xi_d^*]$ in a tessellated d-dimensional configurational space is given as

$$E=\boldsymbol{\alpha}^T\boldsymbol{E}^e, \tag{4}$$

where $\boldsymbol{E}^e$ is the vector containing energies at the nodes of the simplex within which $\boldsymbol{\xi}_d^*$ is located. $\boldsymbol{\alpha}$ is obtained as $\boldsymbol{\alpha}=\boldsymbol{A}^{-1}\boldsymbol{b}$, where $\boldsymbol{A}=[\mathbf{1},\boldsymbol{\xi}_1^e,\boldsymbol{\xi}_2^e,\ldots,\boldsymbol{\xi}_d^e]^T$ and $b=[1,\xi_1^*,\xi_2^*,\ldots,\xi_d^*]^T$. Here, $\boldsymbol{\xi}_i^e$ denotes a vector containing the ith coordinate value of all nodes in an element e. The element e is located by calculating $\boldsymbol{\alpha}$ for every element in sequence and selecting the element e where all elements of $\boldsymbol{\alpha}>0$. This step becomes more time consuming as the dimensionality of configurational space (hence, the number of elements) increases. Further, the geometry of the configurational space becomes more complex as the dimensionality of

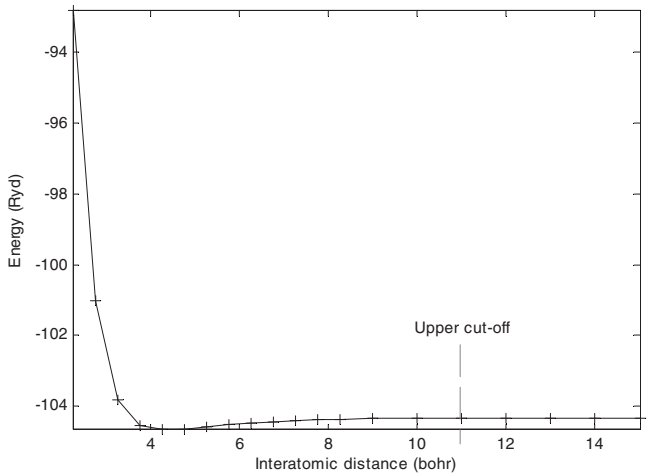

FIG. 3. The plot of energy versus interatomic distance for a two-atom Pt cluster. The location of nodal points in this one-dimensional configurational space and the upper cutoff used for calculations are indicated.

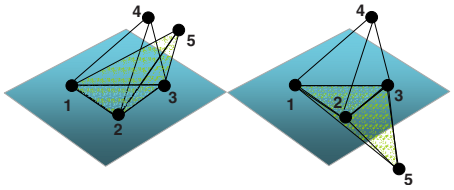

FIG. 4. (Color online) In the case of a five-atom cluster, the locations of the fourth and fifth atoms can be fixed with respect to the plane formed by atoms 1-2-3 using the cluster specifiers $[R_{12},R_{23},R_{13},R_{14},R_{24},R_{34},R_{15},R_{25},R_{35}]$. However, these specifiers do not completely represent the cluster. The dependent variable in this case is R_{45} which can take one of possible two values based on the location of atom 5 either above or below the plane formed by atoms 1-2-3.

the configurational space increases. For example, the configurational space of a fourth-order cluster (excluding symmetries) involves 24 linear constraints and a quartic constraint.

The number of independent variables specifying an $m>4$ atom cluster is given by $d=3m-6$ although $\frac{1}{2}m(m-1)$ variables are needed to uniquely define a cluster.[23] An example of how cluster specifiers are determined for a five-atom cluster is illustrated in Fig. 4. In this example, there are nine independent variables and one dependent variable (R_{45}) that can take one of two values based on the location of the fifth atom. Thus, $m>4$ cases present special difficulties associated with dependent variables. We address this issue by creating different configuration spaces corresponding to the values that each dependent variable takes. For the case of a five-atom cluster, this means that two potentials need to be created, one for the case where atom 5 is above the plane formed by atoms 1-2-3 and another when it is below that plane.

For a binary AB system, all possible cluster configurational spaces are created for a given cluster size; e.g., for L-atom clusters, $L+1$ energy databases [e.g., for $L=2,3$ databases containing $E^*(X_A,X_A),E^*(X_A,X_B),E^*(X_B,X_B)$] need to be generated. The upper and lower cutoffs were selected by carefully analyzing the energies of two-atom clusters over a large range of R_{12} to locate an upper cutoff beyond which the interaction between atoms were not significant and a lower cutoff where the interaction energy is positive.

For platinum with lattice parameter of $a=7.5$ bohrs, the lower cutoff of atom spacing in a cluster within the database was fixed as $R_{ij}>0.3a$ and the upper cutoff was fixed as $R_{ij}<1.5a$. The plot of energy versus interatomic distance for a two-atom Pt cluster, from which the cutoffs were identified, is shown in Fig. 3. The cutoffs signify that clusters with $R_{ij}<0.3a$ and $R_{ij}>1.5a$ are not available in the database. During MBE calculations, the energies of clusters containing such interatomic distances are approximated using the following means. For $R_{ij}<0.3a$, cluster energies were given an artificial high value to signify that such configurations are not energetically feasible. For N-atom clusters with $R_{ij}>1.5a$, the excess energy attributed to N-body interactions is assumed to be zero (i.e., $V^{(N)}\approx 0$, for N-atom clusters with an $R_{ij}>1.5a$). This is mathematically equivalent to approximating the energies of large clusters using energies of smaller subclusters. For example, Fig. 5(a) shows the energy surface of three-atom platinum clusters up to an upper cutoff size of $1.5a$ and Fig. 5(b) shows the complete energy surface when the energies beyond the upper cutoff are approximated using two- and one-atom energies.

IV. WEIGHTED MULTIBODY EXPANSION

Multibody expansion has been shown to work very well for rare gases where the expansion is dominated by pair interactions making higher terms in the expansion negligible. The total energy of metallic systems, however, has significant contributions from higher-order interactions and the ex-

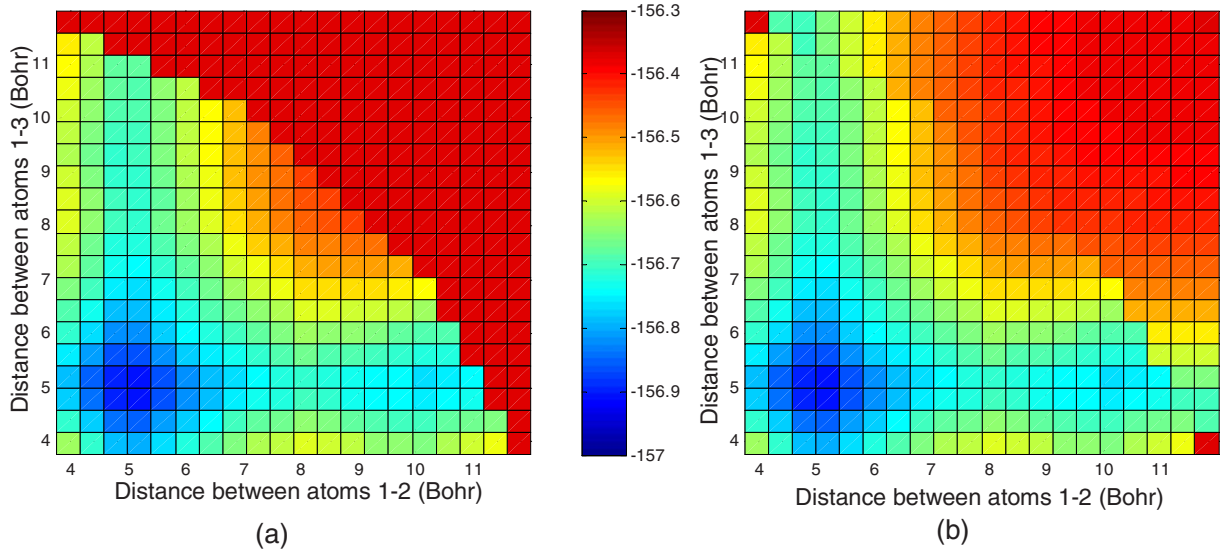

FIG. 5. (Color online) Energy surface $[E^*(X_1,X_2,X_3)]$ for three-atom Pt clusters whose atoms are positioned at the vertices of a right-angled triangle with line joining atoms X_2 and X_3 forming the hypotenuse. (a) shows computed platinum three-atom cluster energies, while (b) shows the extension of energies beyond the cutoff using energies of smaller clusters $[E^*(X_i,X_j)$ and $E^*(X_i)]$.

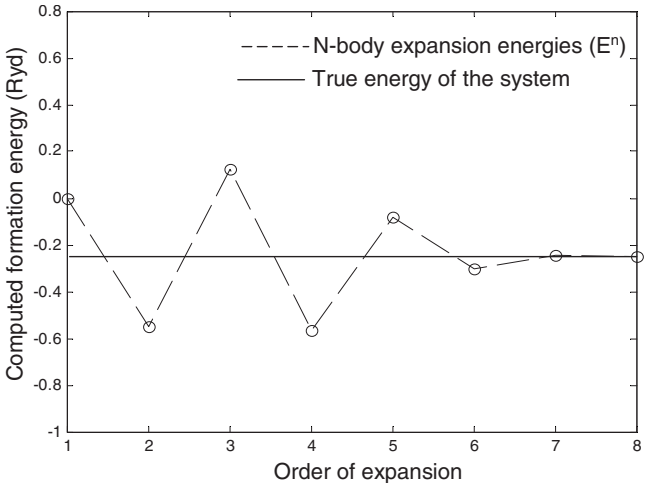

FIG. 6. Convergence of the many-body energy expansion of an eight-atom fcc platinum cluster requires at least a seventh-order expansion to reasonably capture the true energy.

pansion has nonsmooth convergence behavior.[18] Figure 6 shows the behavior of multibody expansion for an eight-atom (two unit cell) fcc platinum cluster that requires at least a seventh-order expansion to capture the true energy. It is observed here that energies computed by including successively higher orders of interaction, in fact, oscillate around the true energy. An *ad hoc* numerical approach for estimating the true energies for the case in Fig. 6 will be to appropriately weight the energies obtained at different orders of the multibody expansion, which is akin to smoothing (or filtering) the energy oscillations in Fig. 6. Numerical experiments presented in the next section indicate that WMBE calculations lead to a dramatic improvement in the convergence behavior of the multibody expansion. In the WMBE approach, the energies up to a cutoff order of expansion P are weighted so that we reach as close to the true energy (E_M) of an M-atom system as follows:

$$E_M(X_1,X_2,\ldots,X_M) = \alpha_1 E_1(X_1,X_2,\ldots,X_M) + \alpha_2 E_2(X_1,\ldots,X_M) + \ldots + \alpha_P E_P(X_1,\ldots,X_M). \quad (5)$$

The coefficients $\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\ldots,\alpha_P]^T$ are computed by solving the equation

$$\boldsymbol{\alpha} = \boldsymbol{C}^{+}\boldsymbol{E}, \quad (6)$$

where $\boldsymbol{E}$ are the true energies of q M-atom clusters $(\boldsymbol{X}^i, i=1,\ldots,q)$ computed with self-consistent DFT calculations. Each row of $\boldsymbol{C}$ contains the energies $[E_1,E_2,\ldots,E_P]^{(i)}$ obtained from multibody expansion of each of these clusters [where $E_p^{(i)}=E_p(\boldsymbol{X}^i)$]. $\boldsymbol{C}^{+}$ is the pseudoinverse of matrix $\boldsymbol{C}$. The technique to obtain coefficients $\boldsymbol{\alpha}$ is thus similar to the method of Connolly and Williams[1] used for cluster expansions. In their technique, truncation of the expansion is based on which clusters are important; for example, in fcc crystals where only clusters containing nearest neighbors are important, the series is truncated at fourth order. In the case of the WMBE, however, the cutoff of the order of interactions needs to be identified through numerical experiments as will be demonstrated in the next section.

Before proceeding to examples, we summarize the steps involved in the overall algorithm as follows.

(1) *Offline calculations (steps 1–4): constructing a database*. Generate coordinates $(\{\boldsymbol{\xi}_d^i\}_{i=1}^n)$ for sampling the configurational space of all cluster sizes involved. For example, each three-atom cluster (1-2-3) corresponds to the coordinate $\xi_3^i=(R_{12},R_{23},R_{13})$ in the configurational space where, e.g., R_{12} is the interatomic distance between atoms 1 and 2. During this step, various constraints based on geometry or symmetry are used to reduce the number of nodes in the configurational space.

(2) For an L-atom cluster of a binary system, coordinates for $L+1$ configurational spaces need to be created during step 1. Each configurational space corresponds to a different atom type list; e.g., for $L=2$ atom clusters of a binary alloy A-B, configurational spaces for clusters of types A-A, A-B, and B-B need to be generated.

(3) Perform tessellation of coordinates in all configurational spaces and store nodal coordinates and element-node lists in the database.

(4) Generate input files and perform self-consistent DFT calculations to compute energies $[E^*(\boldsymbol{\xi}_d^i)]$ at nodal locations of all configurational spaces. Energies from the DFT calculation are read and stored in databases, one corresponding to each configurational space.

(5) *Calculation of MBE coefficients*. Computed self-consistent *ab initio* calculations to compute energies [to obtain $\boldsymbol{E}$ in Eq. (6)] of a few (three or four) different N-atom configurations.

(6) Compute energies using MBE with increasing orders of expansion and obtain $[E_1,E_2,\ldots,E_P]^{(i)}$ for each N-atom configuration used in step 5. During multibody expansion, potentials $(V^{(N)})$ are created using Eq. (2) on the fly, using cluster energies E^* obtained by interpolating from the database constructed in steps (1–4). The steps involved to compute cluster energy, E^*, of an arbitrary cluster are

(a) Locate the cluster in the corresponding configurational space. For example, a three-atom cluster of type A-A-B is located at the coordinate $\xi_i^3=(R_{12},R_{23},R_{13})$ in the three-dimensional configurational space of A-A-B type.

(b) Identify the element in which the cluster is located and perform linear interpolation using known energies at nodal values in that element using Eq. (4).

(c) Energies of clusters that are not available in the database are approximated using the methods detailed at the end of Sec. III.

(7) Compute the coefficients $\boldsymbol{\alpha}$ of weighted MBE using Eq. (6). Perform tests for convergence by comparing the energies predicted by WMBE with *ab initio* calculation for a few other configurations of N atoms.

(8) *MBE calculations for arbitrary N-atom configurations*. The converged weighted expansion can be now be employed for computing energies of other N-atom systems

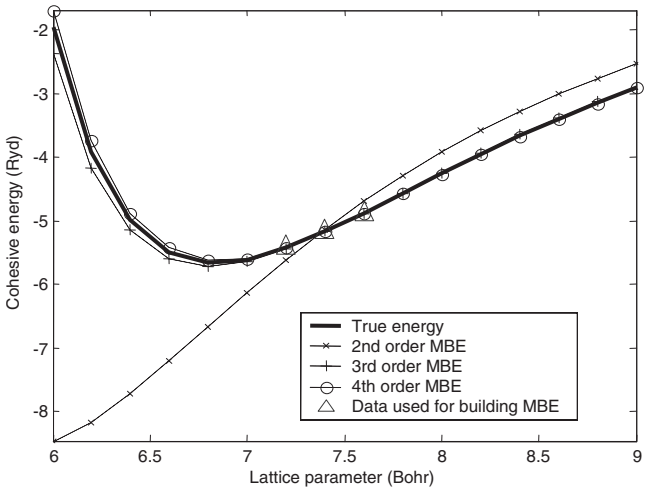

FIG. 7. Comparison of true energies of the a $4\times1\times1$ (16-atom) fcc platinum cluster with that predicted by multibody expansion. Weights in the multibody expansion were computed using three energies at lattice parameters of 7.2, 7.4, and 7.6 bohrs.

using Eqs. (5) and (1). During calculations, cluster energies E^* are again interpolated from the database as in step 6.

V. RESULTS FOR METALLIC SYSTEMS

A. Extrapolatory performance of the WMBE approach

In the first test case, energies predicted by multibody expansion are compared with true energies obtained using the embedded-atom potential of Sutton and Chen[24] for platinum atom clusters. Convergence of the expansion is tested using exact cluster energies (without performing interpolation). Atom configurations used in these cases correspond to $n_x \times n_y \times n_z$ clusters with n_i unit cells located in the ith direction.

Figure 7 shows the energies obtained for an isolated 4

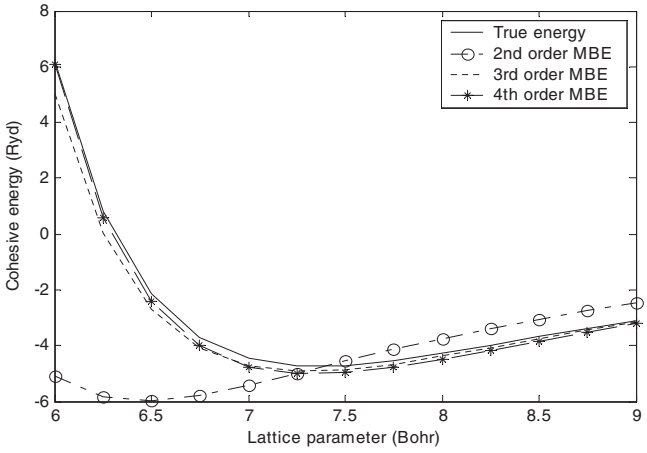

FIG. 8. Comparison of the true energies of a $4\times1\times1$ (16-atom) fcc platinum cluster with MBE results for an extrapolatory case of a fcc lattice where the face-centered atom in the x-y plane and y-z plane in the fcc basis are translated by $(-0.1,-0.1,0)$ and (0, 0.1, 0.1), respectively, in crystal coordinates.

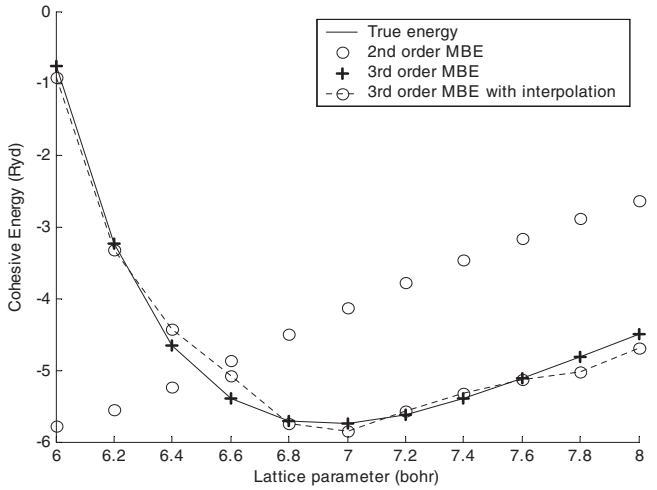

FIG. 9. Comparison of the true energies obtained for a $2\times2\times1$ (16-atom) fcc platinum cluster with energies computed using second- and third-order WMBEs.

$\times1\times1$ (16-atom) cluster of platinum computed using second-, third-, and fourth-order WMBEs and the true energies. In all cases, the parameters $\boldsymbol{\alpha}$ were computed using $4\times1\times1$ Pt clusters using just three energies at lattice parameters of 7.2, 7.4, and 7.6 bohr as indicated in Fig. 7. The energies increase linearly within this range of lattice parameters. In spite of this, the predicted energies from the third- and fourth-order multibody expansions exactly capture the parabolic nature of the true energy profile. As a test of the extrapolatory performance of wMBE, we perturb the face centered atoms in the x-y plane and y-z planes of the fcc basis by $(-0.1,-0.1,0)$ and (0, 0.1, 0.1), respectively, in crystal coordinates. In spite of the large changes in energy resulting from this perturbation, the expansion built previously for a fcc cluster is able to reproduce the energy profile of this distorted cluster accurately (Fig. 8).

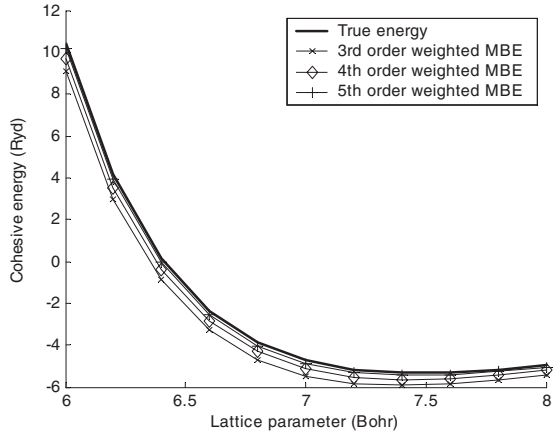

FIG. 10. Comparison of the true energies obtained for a $2\times2\times1$ (16-atom) fcc platinum cluster with energies computed using third-, fourth-, and fifth-order WMBEs for an extrapolatory case of a fcc lattice with the face-centered atom in the x-y plane of the fcc basis is translated by $(-0.1,-0.1,0)$ in crystal coordinates.

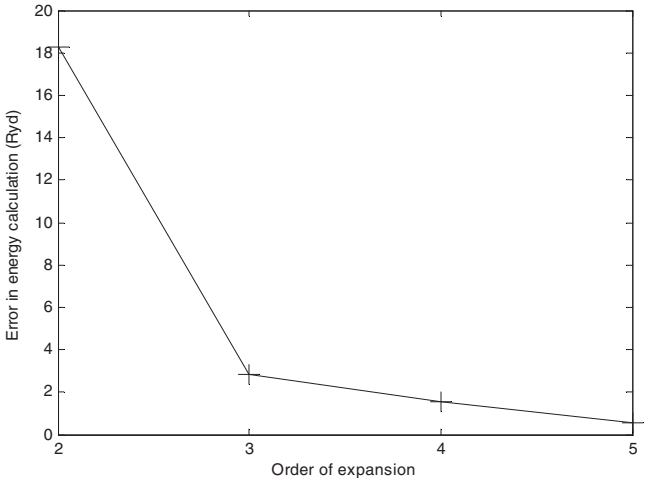

FIG. 11. Decrease in the l^2 norm error in energies with increasing order of multibody expansion for the extrapolatory case in Fig. 10.

B. Convergence of WMBE in extrapolatory cases

Figure 9 shows the energies predicted at various lattice parameters using second- and third-order WMBEs for an isolated $2\times2\times1$ fcc platinum cluster. In this case, the parameters $\boldsymbol{\alpha}$ were originally computed using $2\times2\times1$ fcc clusters of Pt using 11 lattice parameters between 6 and 8 bohrs in increments of 0.2 bohr. Although Fig. 9 shows that the third-order expansion is adequate to capture the true energy profile, use of higher orders of expansion improves the performance in extrapolatory cases. Figure 10 depicts the performance of third-, fourth-, and fifth-order MBEs in an extrapolatory case where the face-centered atom in the x-y plane of the fcc basis is translated by $(-0.1,-0.1,0)$ in crystal coordinates. Figure 11 shows the decrease in the l^2 norm error in energies predicted with increasing order of the multibody expansion. Several other numerical experiments of this kind indicate that the WMBE approach captures the energy profile for any random configurations of N-atom Pt clusters and, thus, has potential applications in N-V-E or N-V-T atomistic simulations. The weighting procedure aims to average out the extraneous energy contributions (e.g., surface energies) arising due to lack of environment in isolated clusters. The limitation in the procedure is that a change in the number of atoms (N) simulated necessitates recalibration of the MBE coefficients. Figure 9 shows the energy variation with lattice parameter obtained from an MBE expansion calculated using cluster energies interpolated from a database. For interpolation, the second-order configurational space is discretized into 20 linear elements (21 nodes) and the third-order configurational space (including symmetries) was approximated using 16 374 tetrahedral elements on which energies were calculated at 3191 nodal locations. Although discretization and linear interpolation introduce errors in the calculation of energies, it is seen that the technique still reasonably captures the energy profile and the energy minima. The advantage of the interpolation approach is that it is an order of magnitude faster since cluster energies are computed beforehand and are directly sampled from the database during simulations.

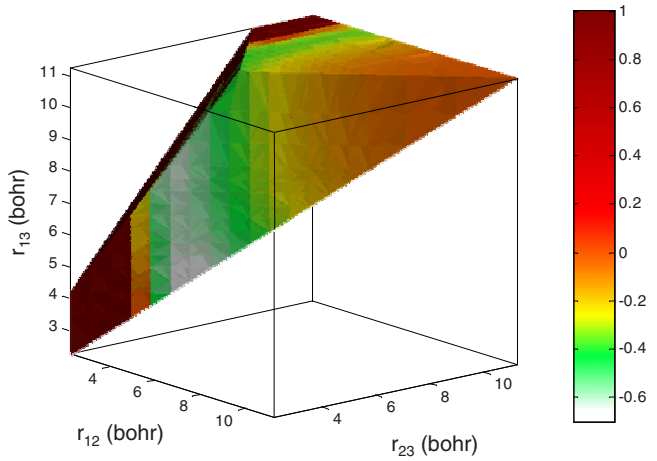

FIG. 12. (Color online) The three-body cluster configurational space for platinum colored by the cohesive energies $[E_c(X_1,X_2,X_3)=E^*(X_1,X_2,X_3)-\Sigma_{i=1}^{3}E^*(X_i)]$ computed from *ab initio* simulations.

C. WMBE using interpolated energies from *ab initio* calculations

Figure 12 shows structure optimization to find the lattice constants for the fcc platinum system using *interpolated* energies of clusters computed from first-principles DFT calculations. Since MBE inherently uses nonperiodic configurations, the energies of periodic structures are computed by considering supercells with true energies $\boldsymbol{E}$ used for fitting Eq. (5) obtained from self-consistent DFT calculations of a periodic unit cell. In this example, a $5\times5\times5$ (500-atom) fcc cluster is considered. The variation of cohesive energy $[E_c(X_1,\ldots,X_m)=E^*(X_1,\ldots,X_m)-\Sigma_{i=1}^{m}E^*(X_i)]$ of three-atom clusters with interatomic distances (R_{12},R_{23},R_{13}) is shown in the configurational space (accounting for symmetry) in Fig. 12. The configurational space is discretized into 4609 tetrahedral elements on which linear interpolation is carried out.

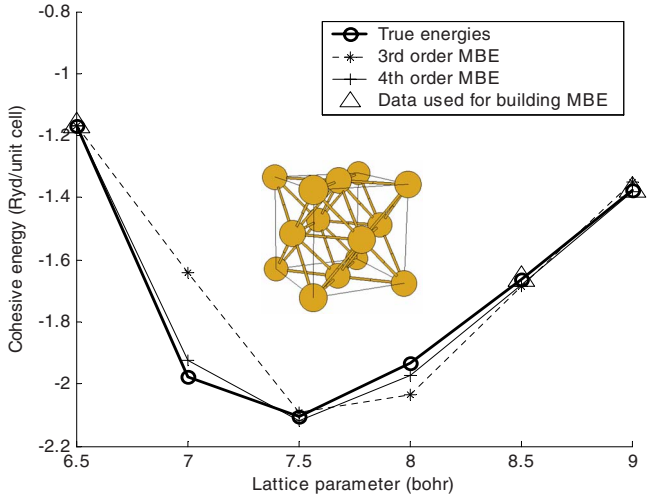

FIG. 13. (Color online) Comparison of the variation of energies with lattice parameter for a periodic fcc pt lattice with WMBE calculations involving cluster energy interpolation.

Ab initio energy data were computed on 1027 nodal locations. Figure 13 shows a comparison of the energies computed using third- and fourth-order WMBEs with the true energies. The coefficients in the multibody expansion were generated using three *ab initio* energy calculations of a periodic fcc platinum lattice with lattice parameters of 6.5, 8.5, and 9.0 bohrs. It is seen from Fig. 13 that the energy profile is well captured using the technique within expected error bounds as discussed later in this section. The significant advantage of using interpolated energies is that they do not utilize any significant computational resource. This is due to the fact that all heavy *ab initio* calculations are performed beforehand and the data are stored for interpolation.

D. Analysis of the accuracy of WMBE with interpolated *ab initio* energies

The main sources of error in the WMBE procedure are the errors involved in the interpolation of energies from the database, fitting weighting coefficients and convergence accuracy for the *ab initio* energy data. The maximum interpolation error over any element (including higher-dimensional elements) is tightly bounded by $c_t r_{mc}^2$, where the absolute curvature of the true *ab initio* energy surface is bounded in each element t by a constant $2c_t$ and r_{mc} is the minimum containment radius of an element. For the two-atom Pt cluster energy surface, the maximum interpolation error was 0.03 mRy. Although this error cannot be completely eliminated, we use smaller element sizes in the regions where large energy variations are expected in order to reduce the interpolation error for larger clusters. The convergence accuracy of self-consistent DFT calculations of small clusters was within 0.01 mRy in all cases. In order to study the error in fitting MBE weights, we carried out a leave-one-out cross-validation (CV) procedure. Here, the error in the reproduction of energies is studied by fitting the energy with $N-1$ clusters and computing the error in reproduction of energy of the left-out cluster (E_i). The process is repeated with every single cluster used once as a left-out cluster. The CV error is computed as the mean error $\frac{1}{N}\Sigma_1^N|E_{true}-E_i|$. Compared to statistical estimates such as variance, the CV error provides a more reliable estimate of the future performance of WMBE when the energies of new clusters need to be predicted.

The *ab initio* energies of $N=300$ randomly generated 24-atom Pt clusters were used for testing the accuracy of the WMBE procedure. The 24 atoms were randomly placed at grid points spaced 7 bohrs apart in each direction over a cube of 105 bohr length and *ab initio* energies ($\boldsymbol{E}$) for use in Eq. (6) are computed. The mean CV error for third-order MBE was found to be 0.381 Ry (15.9 mRy per atom). The mean CV error during cross validation for fourth-order expansion reduces to 0.121 Ry (5.04 mRy per atom). This demonstrates convergence towards *ab initio* energies, although the error may still be significant for modeling phenomena such as phase transformations where accuracy of the order of mRy may be required.

E. Convergence of WMBEs for a binary system (α-alumina Al_2O_3)

A multibody expansion is constructed for the α-alumina (Al_2O_3) system using cluster energies computed using the Streitz-Mintmire (SM) model.[25] Streitz-Mintmire potential is a many-body functional that merges electrostatic potential with an embedded-atom potential to describe metal-oxide energies. α-alumina (Al_2O_3) has a rhombohedral primitive unit cell and is described in space group $R\bar{3}c$ (No. 167) with two lattice parameters a and b. The lattice parameter a is varied while b is fixed at 0.4856 bohr. Figure 14 plots the variation of energies, computed using WMBE, as a function of lattice parameter a for a $2\times2\times1$ cluster of α-alumina. The true energies as computed by the SM model at each lattice parameter are also shown. Four energies at lattice parameters $a=7.0$, 7.2, 7.4, and 7.6 bohrs were used to compute the MBE coefficients. Within this range of lattice parameters, energies increase linearly as indicated in Fig. 14. In spite of this, a fourth-order expansion is able to represent the curvature of the α-alumina energy profile predicted by the SM model. In contrast to the fcc Pt case in Fig. 7, the predicted energies from the third-order multibody expansion are not able to predict the energy minima, while the fourth-order predicts the lattice parameter $a=6.6$ bohrs accurately. Instead of the SM model, *ab initio* calculations of isolated cluster energies (E^*) could have been used. Since we compute energies of isolated clusters by approximating a periodic lattice, care must be taken to avoid the influence of lattice Coulomb potentials on the ionic Al-O cluster (due to finite-size effects) by using a large enough unit cell. DFT calculations were avoided in this example due to the computational complexity of handling a large number of plane waves because of the sharply peaked valence states in oxygen and the requirement of a large unit cell. WMBE of a binary metallic system that uses *ab initio* calculations is reported in the next example.

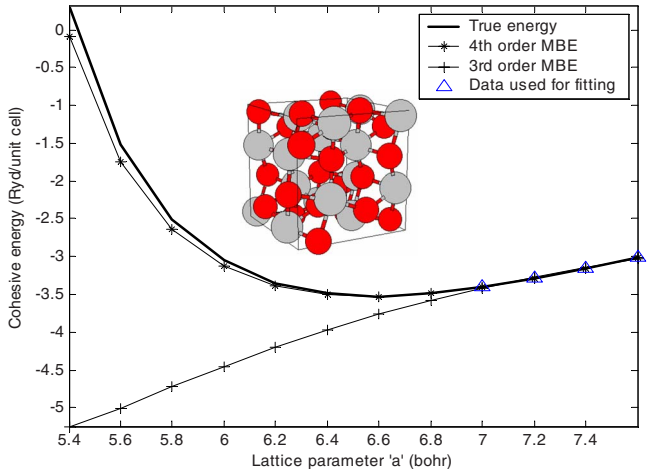

FIG. 14. (Color online) Comparison of the variation of energies with lattice parameter for a $2\times2\times1$ supercell of α-alumina (space group $R\bar{3}c$) using third- and fourth-order WMBEs. The true energies and the energies used for computing MBE coefficients are indicated.

F. WMBE of a binary (Au-Cu) system using interpolated *ab initio* energies

This example demonstrates structure optimization to find the lattice constants for the fcc $CuAu_3$ system (space group

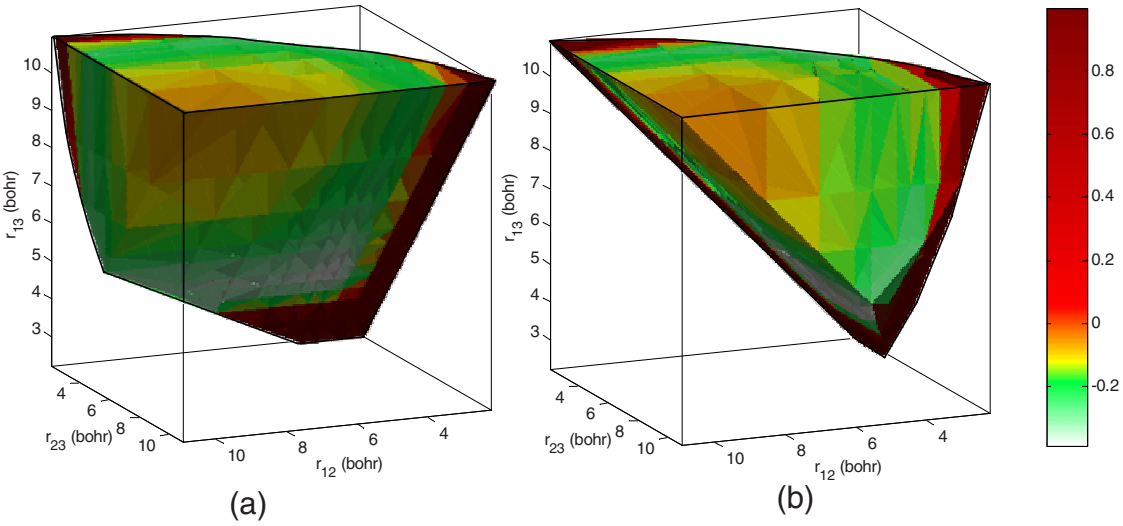

FIG. 15. (Color online) The three-body cluster configurational space for (a) Cu-Cu-Au and (b) Cu-Au-Au colored by the cohesive energies (as defined in Fig. 12) computed from *ab initio* simulations. The two spaces have different geometries due to different underlying symmetries.

$Pm\bar{3}m$, No. 221) using interpolated energies of clusters computed from first-principles DFT calculations. As in the case of fcc Pt, the energies $\boldsymbol{E}$ used for fitting Eq. (5) are obtained from self-consistent DFT calculations of a periodic unit cell. A $6\times6\times6$ (864-atom) fcc cluster is considered to approximate the periodic lattice, and MBE is constructed using energies interpolated from the tessellated configurational space. As an example, the cohesive energy (E_c) variation with cluster specifiers (R_{12},R_{23},R_{13}) in the configurational space for three-atom Cu-Cu-Au and Cu-Au-Au clusters is shown in Figs. 15(a) and 15(b), respectively. Apart from the nine constraints discussed in Sec. III, the inequalities $R_{23}<R_{13}$ and $R_{12}<R_{13}$, respectively, are additionally used to account for cluster symmetries in the space shown in Figs. 15(a) and 15(b). The lower and upper cutoffs used for constructing these spaces were 2.19 and 10.95 bohrs, respectively. For single atom-type clusters of copper or gold, the upper and lower cutoffs were fixed at 0.3 and 1.5 times the lattice parameters of pure fcc Cu and Au lattices.

Figure 16 shows a comparison of the energies computed using third- and fourth-order WMBEs with the true energies. The coefficients in the multibody expansion were generated using three *ab initio* energy calculations of a periodic fcc $CuAu_3$ lattice with lattice parameters of 8.6, 8.7, and 8.8 bohrs. Similar to the Al_2O_3 case, the third-order multibody expansion is not able to capture the energy profile of fcc $CuAu_3$ whereas a fourth-order expansion provides a reasonable approximation of the energy profile. The WMBE approach allows computation of the energy of large systems with accuracy subject to the errors discussed previously. Cross-validation accuracy for this system using a similar procedure as described before was also carried out. We employed 300 random clusters of 24-atom $CuAu_3$ clusters for testing the accuracy of the WMBE procedure. The 24 atoms were randomly placed at grid points spaced 7.5 bohrs apart in each direction over a cube of 112.5 bohr length. A cross-validation error of 0.187 Ry (7.8 mRy per atom) was achieved when a fourth-order expansion was used. This error may be significant for modeling phenomena such as phase transformations, but WMBE is a good replacement for empirical potentials in several other multiscale modeling applications where reasonable accuracy is required. Although use of higher (5+) body interactions is expected to improve the fit, it greatly increases computational overhead in tessellation and data generation. We are currently working on the use of data-adaptive hierarchical interpolation to address this issue.

VI. CONCLUSIONS

The developments presented here advance the existing state of the art in multibody expansion technique for the representation of energies of alloy systems through the following new contributions.

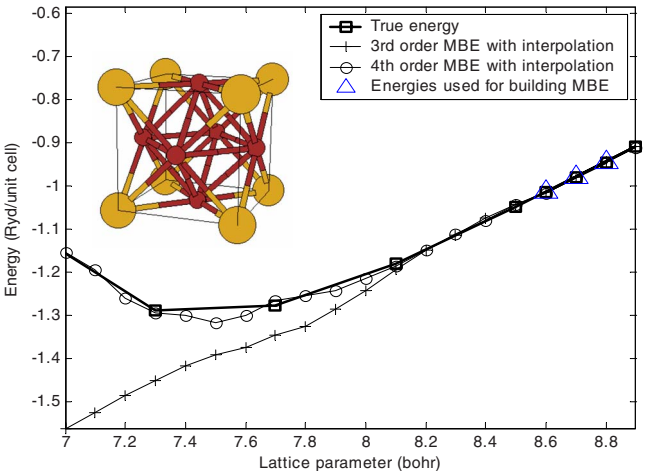

FIG. 16. (Color online) Comparison of the variation of energies with lattice parameter for a periodic fcc $CuAu_3$ lattice with WMBE calculations involving cluster energy interpolation.

(i) The convergence characteristics of multibody expansions were improved by weighting energies obtained from various orders of atom interactions in a method called weighted multibody expansion.

(ii) In contrast to methods such as cluster expansion that involve the ordering degrees of freedom, WMBE focuses on the positional degrees of freedom. This allows one to explicitly model structural relaxations.

(iii) Database interpolation techniques are demonstrated for accelerating computation of energies using multibody expansions. Multibody expansions were computed directly from the *ab initio* energies of small clusters to model energies of platinum and a binary-alloy (Au-Cu) system. The quality of the expansion was quantified using the leave-one-out cross-validation technique.

(iv) The technique involves considerably lesser computational cost, with no requirement of periodicity, and hence, could be used to perform more accurate *N-V-E* or *N-V-T* molecular simulations of metallic clusters and complex phase structures compared to other commonly used position-dependent potential approximations. We are currently working on data-adaptive hierarchical interpolation which would allow us to build higher (5+) order potentials that would lead to improved accuracy.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Materials Science Division of the Army Research Office and of the Computational Mathematics Program of the Air Force Office of Scientific Research.

APPENDIX: *AB INITIO* CALCULATIONS

Ab initio electronic-structure calculations were carried out using density functional theory in the local density approximation, as implemented in the PWSCF package, using Perdew-Zunger parametrization of the exchange correlation energy and Rabe-Rappe-Kaxiras-Joannopoulos[26] (ultrasoft) pseudopotential. Kohn-Sham orbitals were expanded in a plane-wave basis up to an energy cutoff calculated to ensure convergence. Brillouin zone integrations were carried out using single *k*-point calculation and Methfessel-Paxton first-order spreading.[27] The cell size is taken to be sufficiently large to effectively simulate an isolated cluster. For platinum, the energy cutoff was 244.8 eV and the cell size was taken as 4 times the maximum size of the cluster.

*vs85@cornell.edu
†njz1@cornell.edu

[1] J. W. D. Connolly and A. R. Williams, Phys. Rev. B **27**, 5169 (1983).
[2] J. M. Sanchez, F. Ducastelle, and D. Gratias, Physica A **128**, 334 (1984).
[3] A. Zunger, in *Statics and Dynamics of Alloy Phase Transformations*, Vol. 319 of NATO Advanced Study Institute, Ser. B, edited by P. Turchi and A. Gonis (Plenum, New York, 1994).
[4] D. de Fontaine, in *Solid State Physics*, edited by H. Ehrenreich and D. Turnbull (Academic Press, New York, 1994), Vol. 47, p. 33.
[5] Z. W. Lu, S. H. Wei, A. Zunger, S. Frota-Pessoa, and L. G. Ferreira, Phys. Rev. B **44**, 512 (1991).
[6] H. Y. Geng, M. H. F. Sluiter, and N. X. Chen, Phys. Rev. B **73**, 012202 (2006).
[7] A. van de Walle and G. Ceder, J. Phase Equilib. **23**(4), 348 (2002).
[8] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder, Nat. Phys. **5**, 641 (2006).
[9] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, Phys. Rev. Lett. **91**, 135503 (2003).
[10] A. E. Carlsson, in *Solid State Physics* edited by H. Ehrenreich and D. Turnbull (Academic, Boston, 1990), Vol. 43.
[11] R. P. Gupta, Phys. Rev. B **23**, 6265 (1981).
[12] Y. Li, E. Blaisten-Barojas, and D. A. Papaconstantopoulos, Phys. Rev. B **57**, 15519 (1998).
[13] J. N. Murrell and R. E. Mottram, Mol. Phys. **69**, 571 (1990).
[14] J.-Y. Fang, R. L. Johnston, and J. N. Murrell, Mol. Phys. **78**, 1405 (1993).
[15] H. Cox, R. L. Johnston, and J. M. Murrell, J. Solid State Chem. **145**, 517 (1999).
[16] R. Drautz, M. Fahnle, and J. M. Sanchez, J. Phys.: Condens. Matter **16**, 3843 (2004).
[17] M. Fahnle, R. Drautz, F. Lechermann, R. Singer, A. Diaz-Ortiz, and H. Dosch, Phys. Status Solidi B **242**, 1159 (2005).
[18] B. Paulus, K. Rosciszewski, N. Gaston, P. Schwerdtfeger, and H. Stoll, Phys. Rev. B **70**, 165106 (2004).
[19] N. X. Chen, Phys. Rev. Lett. **64**, 1193 (1990).
[20] N. X. Chen and G. B. Ren, Phys. Rev. B **45**, 8177 (1992).
[21] B. Paulus, Phys. Rep. **428**, 1 (2006).
[22] Computer code QHULL, http://www.qhull.org/
[23] J. W. Martin, J. Phys. C **8**, 2837 (1975).
[24] A. P. Sutton and J. Chen, Philos. Mag. Lett. **61**, 139 (1990).
[25] F. H. Streitz and J. W. Mintmire, Phys. Rev. B **50**, 11996 (1994).
[26] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, 1227 (1990).
[27] M. Methfessel and A. T. Paxton, Phys. Rev. B **40**, 3616 (1989).