JOM, Vol. 73, No. 11, 2021
https://doi.org/10.1007/s11837-021-04840-6

Cluster Expansion of Alloy Theory: A Review of Historical Development and Modern Innovations

SARA KADKHODAEI [1,3] and JORGE A. MUÑOZ [2,4]

1.—Department of Civil and Materials Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA. 2.—Department of Physics, The University of Texas at El Paso, El Paso, TX 79968, USA. 3.—e-mail: sarakad@uic.edu. 4.—e-mail: jamunoz@utep.edu

The parameterization of a physical or empirical model from a set of highly accurate but expensive calculations or measurements to generate less precise but cheaper predictions is common in many disciplines. In computational materials science and informatics-enabled design of materials, the cluster expansion (CE) method provides a direct approximation of the free energy of a lattice, or any other thermodynamic variable, in terms of a discrete cluster function, making it one of the most widely used approaches for phase diagram calculations, including order–disorder phase transitions. In this article, we review the theoretical developments that culminated in the formulation of the CE method, numerous statistical techniques currently used to fit and optimize the parameters of the CE model, the convergence of the CE method with modern machine learning and data science techniques, and recent developments that push the field beyond the conventional CE, including for structural alloy design.

INTRODUCTION

The cluster expansion (CE) method is a widely used and valuable technique to describe the thermodynamics and cooperative phenomena, e.g., the order–disorder transition, of complex, multicomponent systems.[1–4] The method expands a lattice model free energy as a linear combination of the free energies of individual finite-size clusters of sites within the lattice structure. The CE formalism sets up a mathematical framework in which any scalar property of the crystal, e.g., formation energy, electronic bandgap, and elastic strain energy, can be represented in terms of multisite correlation functions.[5,6] The conventional cluster expansion in its compact form expands a property of a lattice per unit cell, $P(\boldsymbol{\sigma})$, as a linear combination of correlation functions $\bar{\phi}_\alpha$,

$$P(\boldsymbol{\sigma}) = \sum_{\alpha} m_\alpha J_\alpha \bar{\phi}_\alpha(\boldsymbol{\sigma}), \tag{1}$$

where the sum is over all the sets (also called "orbits") of symmetrically equivalent clusters of sites, denoted by α; i.e., the sum in Eq. 1 is over symmetrically distinct clusters α. The multiplicity m_α is the number of clusters in set α per unit cell (or the number of symmetrically equivalent clusters in an orbit), and J_α represents the effective cluster interaction (ECI) coefficients. The cluster of sites α can be any subset of sites in the lattice. Symmetrically equivalent clusters are those that can be mapped onto each other as a result of the application of an intrinsic symmetry operation of the lattice, and belong to the same orbit (Fig. 1). The cluster correlation function is defined as the average of the product of the spin variables σ_i, defined for each site i in the cluster,

$$\bar{\phi}_\alpha(\boldsymbol{\sigma}) = \frac{1}{N_\alpha} \sum_{\beta \subset \alpha} \left(\prod_{i \in \beta} \sigma_i \right) \tag{2}$$

where the sum is over all subclusters β contained within cluster α, and N_α is the number of subclusters in cluster α (Fig. 2). The spin variables are also known as occupation variables since they describe the type of atom that "occupies" a given site. The correlation functions are identical for symmetrically equivalent clusters. Much of the utility of the CE method stems from the fact that the cluster

(Received March 17, 2021; accepted August 5, 2021; published online September 22, 2021)

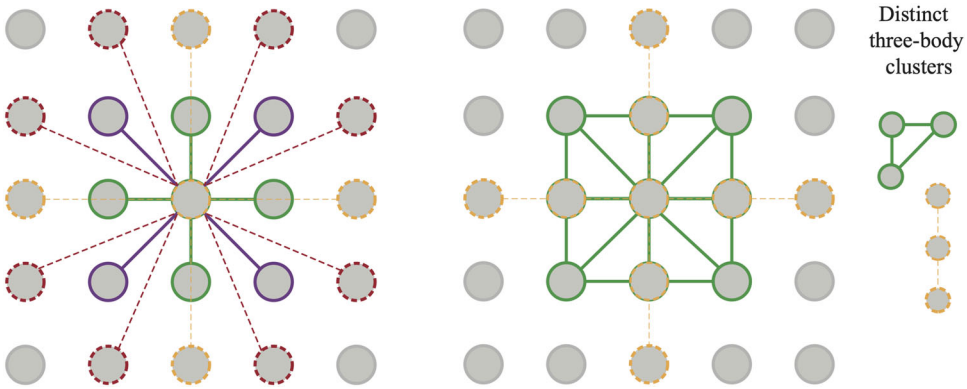

Fig. 1. (Left) Four distinct sets (or orbits) of pair clusters highlighted in a prototype square lattice. Symmetrically equivalent pair clusters belong to the same orbit and are marked with the same color. Green, purple, red, and yellow correspond to nearest, second nearest, third nearest, and fourth nearest orbits of pair clusters. The multiplicity of a cluster m_α is defined as the number of clusters in the same orbit, e.g., four for the nearest and second nearest orbits in the square lattice. (Right) Two distinct set of three-body clusters highlighted in a prototype square lattice. Symmetrically equivalent three-body clusters are marked in the same color. Note that in all cases the pair or three-body cluster includes the central site (Color figure online).

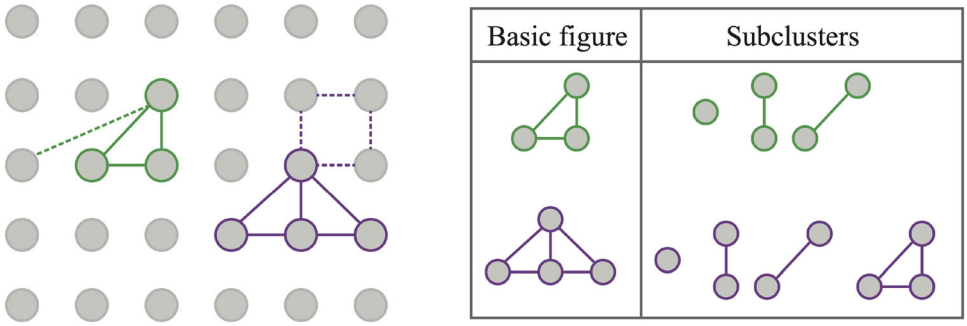

Fig. 2. (Left) Two different basic figures (or maximum clusters) shown in different colors on a prototype square lattice. The subclusters included in each basic figure are connected by solid lines, whereas those not included in the maximum cluster are connected by dashed lines. (Right) Representation of all the subclusters that are included in each basic figure.

correlation functions encompass the symmetry of the space group of the underlying lattice and form a complete orthogonal basis set.[2,3] Accordingly, any function of configuration can be expanded into this complete orthogonal basis set.

The CE formalism is a restricted variational principle that consists of minimizing a free energy functional where the exact entropy is approximated by a linear combination of entropies of finite clusters included in a given maximum cluster (or basic figure).[4,7,8] The accuracy of a particular cluster expansion depends on the degree of convergence from the truncation of the infinite sum in Eq. 1 to a sum over a finite number of clusters. Successively larger clusters lead to better approximations, but the number of independent (model) variational parameters, i.e., number of distinct correlation functions, increases sharply with the maximum cluster size, leading to practical limitations for variational optimization.

The popularity of the CE method stems from the major advantage of dealing with localized effects that cannot be properly encapsulated with a mean-field theory approach; For example, the method produces flexible models that can include short-range correlation effects with arbitrary complexity. However, the method falls short near critical points, where the correlations become infinitely long. Accordingly, the cluster expansion is best suited to study first-order transitions and to explore the critical behavior of phases with substitutional disorder.

With the progress of accurate electronic structure calculations, the CE method has made a significant contribution to the first-principles computation of alloy properties.[9–17] Density functional theory (DFT)[18,19] is the workhorse of modern computational materials research, but it is computationally expensive and in practice limited to a relatively small number of atoms. The CE has enabled DFT-based simulations and understanding of multicomponent systems and systems with chemical disorder that at any given point of computational maturity lie beyond what is possible using direct DFT calculations. The standard approach is to determine the set of ECI coefficients by minimizing the difference between the total energies of a set of crystal structures with different ionic

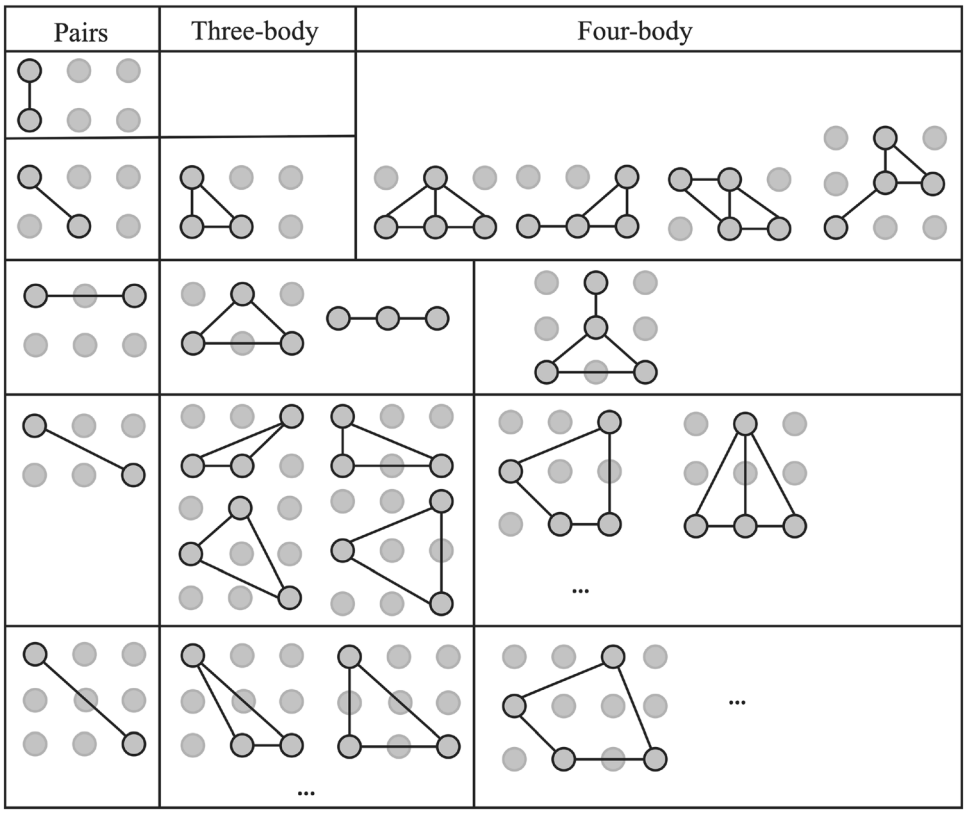

Fig. 3. Clusters in a prototype square lattice, ranked in importance according to heuristic cluster selection rules. Cluster diameter, i.e., the maximum distance between any two sites in the cluster, increases along the vertical axis from top to bottom. Cluster size, i.e., the number of sites in the cluster, increases horizontally from left to right. A given cluster can be included in the expansion of Eq. 2 only if more important clusters have already been included, which extends to clusters with smaller diameters and all the subclusters of the given cluster. Adapted from Ref.26.

configurations calculated using a first-principles method and the prediction of the truncated cluster expansion,[9,12] but many other approaches have been developed, and several are reviewed herein.

Many of the developments in the field since the formalization of the CE method in 1984[3] have been on the intertwined decisions of where to truncate the infinite sum in Eq. 1, which extends to the number of sites to include as well as to their relative distances, and which clusters of the same size but different relative distances to include (Fig. 3). Although initially the selection of these clusters was by trial and error, constraints and heuristics were introduced rapidly[12] and formal frequentist statistics with techniques that nowadays we would recognize as "data science" were established in the early 2000s.[20] More recent developments include Bayesian methods that can implement knowledge or assumptions about the system via physics-informed priors,[21,22] which can substantially reduce the amount of data required to train CE models. The convergence of modern machine learning and related fields with the CE method has brought powerful additions to the methodology; For example, compressed sensing[23] has been used to optimize the size and configuration of clusters in a Bayesian framework,[24] and nonlinear interpolation tools to be used with neural networks and other machine learning algorithms have been developed, providing more resolution by implementing symmetry-adapted cluster functions.[25]

In this review, we present the basic formalism underlying the CE method and provide an overview of its development starting from the early model of Bragg and Williams and through the generalized CE. We discuss the formulation of the original cluster variation method (CVM), where the variational parameters are based on probability distributions, and compare it with the formulation of the general CE in which the cluster correlation functions are the variational parameters. The chronological organization provides insight into the practical issues of improving the accuracy of a cluster expansion by selecting the best set of clusters, and how the search for accuracy resulted in the application of techniques from frequentist and Bayesian statistics and the development of special quasirandom structures (SQS). Moreover, we review various extensions to the standard CE formalism to account for lattice distortions and relaxations, thermodynamics of point defects, and diffusion calculations. The seminal paper by Kikuchi was published seven decades ago, and during that time the vocabulary and notation have changed

and even now are not completely homogenized. We make our best effort to show all the formulae in consistent notation to emphasize relationships and facilitate comparisons, and we include compendia of abbreviations and variables. The literature on both methodological advances and applications of the CE is extensive, hence it is impossible for any review to be complete. Much has been omitted, particularly regarding applications, but the formalism development and implementations reviewed here are illustrative and we hope will serve as a starting point for the interested reader.

This article is organized into four sections following this "Introduction:" a chronological review of the CVM and its development into the CE is given in "Cluster Variation Method" section, the application of the CE method to predict thermodynamic properties and order–disorder phase transitions is outlined in "Development and Applications of the Cluster Expansion Method" section, a review of extended formulations of the CE that go beyond the standard formulation is described in "Beyond the Conventional Cluster Expansion Method" section, and a discussion of current trends and future directions is presented in "Outlook and Conclusions" section. In "Original Development" section, we review various developments that led to the formulation of the CVM and its evolution towards the CE method, and in "Cluster Variation Method within the Variational Principle" section, we describe the CVM from the perspective of the variational principle while briefly discussing different optimization approaches to be adopted for different variational parameters in the CVM and in the CE. In "Cluster Inversion Technique" section, we trace the origin of the cluster inversion technique that allowed the interpolation in cluster space of thermodynamic quantities, in "Optimal Clusters" section we discuss the development and formalization of goodness of fit metrics for CE models using frequentist and Bayesian frameworks, in "Chemical Disorder and Special Quasirandom Structures" section, we review the development and use of SQSs, and in "Code Implementations" section, we list several implementations of the cluster expansion formalism in standalone or integrated computer codes such as the Alloy Theoretic Automated Toolkit (ATAT)[27] and the Atomic Simulation Environment (ASE).[28] In "Long-Range Order" section, we review a CE formalism that integrates the long-range ordering into the cluster function expansion, in "Atomic Relaxations" section, we review the reciprocal representation of the CE, which facilitates incorporation of atomic relaxations into the cluster correlation functions, in "Lattice Vibrations and Distortions" section, we review various CE-based frameworks that take into account atomic vibration and lattice distortion effects in the free energy expansion, and in "Defects" section, we review a local CE formalism that can take into account vacancy formation and expand a kinetically resolved energy barrier for vacancy-mediated diffusive hops.

CLUSTER VARIATION METHOD

Original Development

The original concept of a cluster expansion was built upon similarities with the theory of ferromagnetism[7,29–31] to describe the order–disorder transition in an Ising model.[32,33] Bragg and Williams[32] revealed the thermal statistical nature of atomic arrangements in an alloy by considering the long-range ordering in a lattice and the thermal agitations that counter it. They described the potential energy in terms of a degree of order η, i.e., long-range order parameter, to understand the order–disorder transition in alloys (see Table I for the detailed formalism). Subsequent work by Bethe[33] and Kramers and Wannier[30,31] included the short-range order effects on dynamic equilibrium of atomic configuration (or spin configuration) in a lattice by introducing a "neighbor order" parameter ς (see Table I for the detailed formalism). Those authors provided the first model that described the variation of potential energy in a lattice in terms of neighbor pair correlations, although the model is limited to one- and two-dimensional lattices. Bragg and Williams described the order–disorder transition according to the average state of order throughout the lattice, i.e., long-range order, but the model by Bethe could describe the local fluctuations of atomic arrangements for a given long-range order parameter. Subsequent work such as the "generalized quasichemical method" focused on improving the approximation of the configurational energy, i.e., successive approximations of the free energy of the crystal, by including larger clusters as basic figures.[34–37]

Kikuchi generalized and extended the Bethe model so as to describe order–disorder phenomena with better accuracy for three-dimensional lattices.1 This effort led to the formulation of the CVM by Kikuchi and coworkers[38–41] based on the variational principle for the free energy, in which the free energy is minimized with respect to variations of basic figures of sites (clusters). In this framework, the free energy F_{cvm} can be expressed in terms of the cluster distribution ρ_α,

$$F_{\mathrm{cvm}} = \sum_{\alpha \subset \beta} \left\{ \sum_{\alpha} \rho_\alpha V_\alpha + kTa_\alpha \sum_{\alpha} \rho_\alpha \ln \rho_\alpha \right\}, \qquad (3)$$

where V_α is the contribution of cluster α to the internal energy and a_α is the coefficient of cluster α, which weighs its contribution to the configurational entropy (see Table I for the detailed formalism). The outer summation is over all subclusters α included in the basic figure β, while the inner summations are over all configurations of cluster α. Within the

Table I. Chronological development of the CVM formalism

Bragg and Williams's model

$\eta(T,V) = 1 - \frac{\{4r(1-r)(\exp(\frac{V}{kT})-1)+1)\}^{1/2}-1}{2r(1-r)(\exp(\frac{V}{kT})-1)}$

$V = V_0\eta, \qquad \frac{dV_0}{dT} = 0$

η: order parameter ($\eta = 1$ complete order and $\eta = 0$ complete disorder)
r: fraction of positions of order in the crystal block
T: temperature; k: Boltzmann constant
V: increase of potential energy due to one atomic replacement moving from an ordered to a disordered position.
V_0: maximum potential energy change due to atomic replacement moving from order to disorder, corresponding to $\eta = 1$
This model considers an effective average Hamiltonian variation $V(\eta, T)$ and disregards local fluctuations in degree of order η at different points.
Ref. 32.

Bethe's model

$E = E_0 + \frac{1}{4}NzV(1-\varsigma), \qquad V = \frac{1}{2}(V_{aa} + V_{bb}) - V_{ab}$

$\eta = \tanh z\delta$ [a], $\qquad 1 - \varsigma = \frac{2\sinh(z-2)\delta}{\sinh(2z-2)\delta\cosh z\delta}$ [a]

η: long-range order parameter
ς: order of neighbors (ς=1 perfect order among neighbors and ς=0 perfect disorder)
z: number of neighbors of each atom; N: total number of atoms
V: difference of interaction energies between two equal and two unequal neighbors; E_0: arbitrary constant
V_{aa}, V_{bb}: interaction energies between two equal neighbor atoms
V_{ab}: interaction energy between two unequal neighbor atoms
δ: effective interaction factor with the outer shell (this factor accounts for the pair interactions beyond the first shell)
This model distinguished between two different types of order; the order of neighbors, i.e., short-range order, and the order of a crystal as a whole entity, i.e., long-range order. Including the neighbor order parameter in the energy expansion enables capturing the local fluctuations in atomic ordering.
Ref. 33.

Kikuchi's original CVM

Simple cubic lattice:
$S = k[9\sum_{i=1}^{3}\beta_i y_i \ln y_i - 7\sum_{i=1}^{2} x_i \ln x_i - 3\sum_{i=1}^{6}\gamma_i z_i \ln z_i],$
$E = 3\sum_{i=1}^{3}\beta_i \epsilon_i y_i$
Face-centered cubic lattice:
$S = k[6\sum_{i=1}^{3}\gamma_i z_i \ln z_i - 2\sum_{i=1}^{5}\alpha_i x_i \ln x_i - 5\sum_{i=1}^{2}\omega_i \ln \omega_i],$
$E = 6\sum_{i=1}^{3}\gamma_i \epsilon_i z_i$
S: entropy per atom; E: energy per atom; ϵ_i: nearest neighbor bond energy
$\alpha_i, \beta_i, \gamma_i$: number of different configurations having the same probability,
For Simple cubic lattice:
x_i, y_i, z_i: probability of point, bond, and square configurations (basic figure: square)
For Face-centered cubic lattice:
ω_i, z_i, x_i: probability of point, bond, and tetrahedron configurations (basic figure: tetrahedron)
This model extends Bethe's model to three-dimensional lattices and uses a combinatorial technique for counting the number of atomic arrangements.
Original approach of Ref. 1.
Extension of CVM to other basic figures and lattices can be found in Refs. 38–41[b]

Generalized CVM

$S_{\text{cvm}} = \sum_{\alpha\subset\beta} a_\alpha S_\alpha, \quad S_\alpha = -k\sum_\alpha \rho_\alpha \ln \rho_\alpha$

S_α: entropy of cluster α; a_α: geometric coefficient of cluster α
$\sum_{\alpha\subset\beta}$: only sums over clusters α that are included in the maximum cluster β
ρ_α: probability density function of cluster α; k: Boltzmann constant
$\sum_\alpha$: sum over all configurations of clusters α
$a_\alpha = \sum_{\alpha\subset\beta}(-1)^{n_\beta - n_\alpha}, \quad \rho_{\text{cvm}} = \prod_{\alpha\subset\beta}\rho_\alpha^{a_\alpha}$
n_β, n_α: number of sites in cluster β and α
$\prod_{\alpha\subset\beta}$: only multiply over clusters α that are included in the maximum cluster β

Table I. continued

Generalized CVM

Within the generalized CVM framework, the exact entropy is approximated by a linear combination of entropies of finite clusters (S_α) derived from the basic cluster, i.e., the largest cluster, denoted by β. The underlying assumption is that, for any cluster not included in the basic cluster, $\rho = 1$ at any temperature. The CVM formalism ensures that the probability density ρ is exact at $T \to \infty$.
Ref. 55[b]

Sanchez, Ducastelle, and Gratias's CE formalism

$$F = NkT\left[-\sum_{\alpha}\sum_{s} m_\alpha J_{\alpha s} < \phi_{\alpha' s} >_\alpha + \sum_{\alpha} b_\alpha \sum_{\sigma \subset \alpha} \rho_\alpha \ln \rho_\alpha\right],$$

$$b_\alpha = m_\alpha a_\alpha$$

$$< \phi_{\alpha s} > = \sum_{\sigma \in \alpha} \rho_\alpha \phi_{\alpha s}$$

F: Helmholtz free energy
N: number of lattice points in the crystal; k: Boltzmann's constant; T: temperature
a_α: geometric coefficient of cluster α (defined in the generalized CVM formalism above)
m_α: number of α-type clusters per unit cell
$J_{\alpha s}$: the interaction energy coefficients
$< \phi_{\alpha s} >$: multisite correlation functions (see Eq. 4)
ρ_α: probability distribution function for symmetrically equivalent clusters of type α (see Eq. 6)
The CE formalism introduced a set of multisite correlation functions that forms an orthogonal and complete basis set to represent the thermodynamic functions, e.g., entropy, free energy, and the Hamiltonian. The correlation function description within the CE formalism provides two major advantages compared with the CVM: (1) it provides a systematic approach to determine the number of independent variational parameters in the CVM approximation by a subcluster decomposition of the maximum cluster (or the basic cluster), and (2) it leads to a simpler computational approach for free energy minimization in terms of *independent* correlation functions instead of Lagrange multipliers minimization of free energy with respect to *dependent* cluster distribution functions with intricate symmetry constraints among them.
Ref. 3.

a. Valid for the first approximation of Bethe's model, where the neighbor order parameter is determined by the first shell only.b. An infinitely large basic figure in CVM coincides with the long-range order parameter η defined in the Bragg and Williams formulation. Similarly, the infinitely large multisite correlation factor in the CE is the same as the long-range order parameter.** Models of ferromagnetism and the quasichemical models are not summarized in this table. Refer to Refs. 7,30,31,34–37

CVM, basic clusters are defined as the largest clusters, and subclusters are distinct clusters into which a basic cluster can be decomposed (Fig. 2).

In principle, the CVM is a mean-field variational method for which the free energy functional is derived in terms of a truncated number of clusters, while their interactions with the system are considered via a mean field. The major improvement of the CVM with respect to the generalized quasichemical model was the establishment of a general rule for selection of the basic figures or clusters of sites that improves the quality of the free energy expansion.[38] Clusters of the same number of sites can have different dimensionality; for example, four sites can be in a line (one dimension), a square (two dimensions), or a tetrahedron (three dimensions). A question persisted of whether the dimensionality for a given number of lattice sites played a role in the quality of the expansion, but this was settled when several studies indicated that it is better in general to increase the number of sites in the cluster while keeping its dimensionality low; e.g., it is better to include a three-point basic cluster rather than a square basic cluster in addition to pair clusters[40,41] (Fig. 3).

The CVM can expand the free energy in terms of basic figures of arbitrary sizes, but the methodology is cumbersome for large cluster sizes in three-dimensional lattices, which led to the development of a heuristic known as the "combinatory factor" approach.[1,38] Within this approach, the free energy functional is minimized with respect to the probability distribution of the basic clusters, subject to proper constraints (Eq. 10). The underlying constraint is that the distribution of clusters is to remain unchanged after adding a lattice point to the system. This heuristic approach formulates the configuration distribution in terms of combinatorial factors, which ensures the correct distribution of the basic cluster and its subclusters, e.g., both spins and pairs.

A reformulation of the CVM by Barker[42] and later by Morita[8,43] derived from the exact variational principle of statistical mechanics[7] provides a simpler approach to calculate the entropy of an arbitrary cluster α of size n_α. This approach reduces the task of calculating the entropy based on combinatorial factors to essentially counting the number of clusters in the crystal unit cell. These developments

provided a general understanding that successive inclusion of larger clusters would result in an improved approximation of the free energy, but the central problem of how to truncate the expansion, i.e., how to choose the best basic figures in the CVM, remained unsolved.

Two decades later, Kikuchi gave several tutorials to de Fontaine, Sanchez, and collaborators about the CVM methodology (see, e.g., Ref. 44), and this eventually resulted in the restructuring of the CVM by Sanchez and collaborators into what is now known as the CE method.[2,3,45–48] The CE formalism has two major advantages with respect to the generalized CVM. First, Sanchez and de Fontaine discovered that the number of independent variables for a given CVM approximation is the total number of subclusters into which the basic cluster can be decomposed, i.e., all subclusters plus the basic cluster.[2] This treatment resolved the previous ambiguities about the relationship between basic clusters and convergence of CVM raised earlier by Barker and Guggenheim.[42] In other words, they found that increasing the basic cluster size does not necessarily result in a more accurate CVM approximation unless it adds to the number of distinct subclusters. Second, Sanchez, Ducastelle, and Gratias discovered that the cluster distribution functions can be written as a linear combination of newly introduced "correlation functions."[3] They showed that the correlation functions form an orthogonal basis in the multidimensional space of discrete spin or occupation variables. Accordingly, instead of minimizing the free energy with respect to the probability distribution of clusters (Eq. 10) while applying intricate constraints among different cluster configurations through the Lagrange multipliers approach, the free energy functional is minimized with respect to independent correlation functions (Fig. 4). As a result, the CE formalism enabled the handling of large clusters, which seemed formidable using the heuristic combinatorial factor technique, and paved the way to incorporate second and third nearest neighbor pairs and multisite clusters into the free energy expansion.

Within the CE formalism, any configuration of spin (or occupation) variables on a crystalline system with N lattice points is fully specified by the N-dimensional vector $\boldsymbol{\sigma} = [\boldsymbol{\sigma_1}, \boldsymbol{\sigma_2}, \ldots, \boldsymbol{\sigma_N}]$. The occupation variable that characterizes lattice site i, σ_i, can take values $\pm m$, $\pm(m-1)$, ..., ± 1, (0) for an M-component system with $M = 2m$ (or $2m+1$). Within this framework, a complete set of orthogonal cluster functions $\phi_{\alpha s}$ is constructed in terms of the first M discrete Chebyshev polynomials, $\Theta(\sigma_i)$, as follows:

$$\phi_{\alpha s}(\sigma_i, \sigma_{i'}, \ldots, \sigma_{i''}) = \Theta_n(\sigma_i)\Theta_{n'}(\sigma_{i'}) \ldots \Theta_{n''}(\sigma_{i''}), \quad (4)$$

where cluster α is defined by the set of lattice points $\{i, i', \ldots, i''\}$ and $s = \{n, n', \ldots, n''\}$ represents the set of indices for the order of the polynomial (more details can be found in Ref. 3). The Ising model Hamiltonian H takes a particularly simple form when expressed on the basis $\{\phi_{\alpha s}\}$:

$$-(kT)^{-1}H = \sum_{\alpha} \sum_{s} J_{\alpha s} \phi_{\alpha s}, \quad (5)$$

where $J_{\alpha s}$ is an interaction energy coefficient. The multisite correlation functions are defined as the expectation values of the orthogonal cluster functions (or basis functions) defined in Eq. 4, $< \phi_{\alpha' s} >_{\alpha} \equiv \bar{\phi}_{\alpha s}$. $< \cdots >_{\alpha}$ indicates the average over the basis functions defined in Eq. 4 for all clusters α' that are symmetrically equivalent to cluster α, i.e., clusters that belong to the same orbit. As opposed to the original formulation of the CVM where a particular space group symmetry is imposed through cumbersome linear constraints on cluster distribution functions $\{\rho_{\alpha}\}$, the symmetry can be expressed straightforwardly in terms of the multisite correlation functions. This is because any two clusters related by symmetry operations are characterized by a unique set of correlation functions $\bar{\phi}_{\alpha s}$. Accordingly, the cluster probability distribution function ρ_{β} for a given cluster of type β (where symmetrically equivalent clusters β' belong to the same type or orbit) consisting of n_{β} lattice points is given by

$$\rho_{\beta} = M^{-n_{\beta}} \left[1 + \sum_{\alpha \subset \beta} \sum_{s} < \phi_{\alpha' s} >_{\alpha} \phi_{\alpha s} \right], \quad (6)$$

where the sum is carried over all nonempty clusters α contained within cluster β.

Cluster Variation Method Within the Variational Principle

The free energy functional can be expressed in terms of the probability distribution function as follows:

$$\begin{aligned} F(\rho) &= U(\rho) - TS(\rho) = -kT < \log \frac{\exp\left(\frac{-H(C)}{kT}\right)}{\rho} > \\ U(\rho) &= \sum_{C} \rho H(C) \\ S(\rho) &= -k \sum_{C} \rho \log \rho, \end{aligned} \quad (7)$$

where ρ is the probability distribution function of configuration C, $H(C)$ is the Hamiltonian of configuration C, $U(\rho)$ and $S(\rho)$ are the energy and entropy functionals in terms of the probability distribution function, k is the Boltzmann's constant, and T is the temperature. The equilibrium distribution ρ_{ex} is the Boltzmann distribution that results in the exact free energy of the system

$$\begin{aligned} \rho_{\text{ex}} &= \frac{\exp\left(-\frac{H(C)}{kT}\right)}{Z} \\ Z &= \sum_{C} \exp(-\frac{H(C)}{kT}), \end{aligned} \quad (8)$$

where Z is the partition function. Due to the convexity of the logarithm function, i.e.,

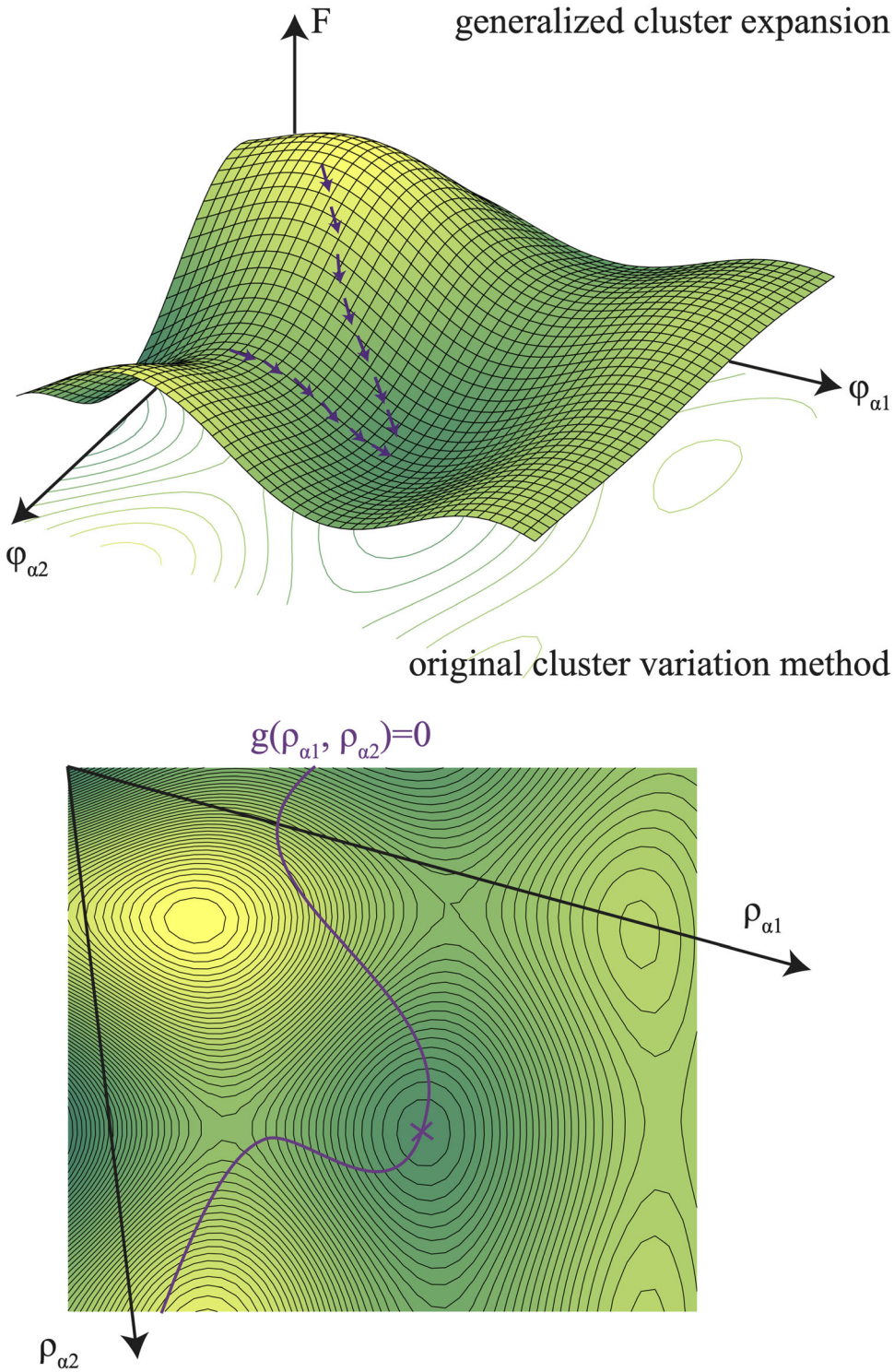

Fig. 4. Comparison of variational parameters in the CE and CVM. (Top) In the CE framework, the cluster functions $\{\phi_\alpha\}$ form a complete orthogonal basis for the free energy expansion, allowing unconstrained variational optimization. (Bottom) In the CVM, the cluster (probability) distribution functions $\{\rho_\alpha\}$ form a nonorthogonal basis with intricate constraints among distribution functions (indicated as $g(\rho_{\alpha_1}, \rho_{\alpha_2})$), which arise from retaining the correct distribution of clusters by adding lattice points to the system. The constrained variational optimization in the CVM becomes formidable due to the existence of intricate relations among the basis vectors.

$< \log X > \leq \log < X >$, the free energy functional for any arbitrary distribution ρ is larger than the exact free energy,

$$F(\rho) \geq -kT \log \left\langle \frac{\exp\left(-\frac{H(C)}{kT}\right)}{\rho} \right\rangle = -kT \log Z. \quad (9)$$

Therefore, the variational principle can be written as

$$F_{\text{ex}} = \min_{\rho} \left\{ \sum_{C} \rho H + kT \sum_{C} \rho \log \rho; \sum_{C} \rho = 1 \right\}. \quad (10)$$

Equation 10 shows that the exact free energy is the minimum of $F(\rho)$ over all *admissible* states, with normalization $\sum_C \rho = 1$. Within any approximation to minimize Eq. 10, only a subset of all admissible states is considered, which are characterized by a finite number of variational parameters; For example, the Bragg–Williams method[32] is the earliest approximation in which the probability density function factorizes over site probabilities, $\rho_{\text{BW}} = \prod_i \rho_i$, where ρ_i is the probability function of different occupations of site i. Since $\sum_C \rho_{\text{BW}} = 1$, the Bragg–Williams approximation represents a subset of admissible states. Bethe's model[33] and the Bethe–Guggenheim model[49] are next-level approximations that account for the correlations between pairs of sites. The probability density function is approximated as $\rho_{\text{BG}} = \frac{\prod_{<ii'>} \rho_{ii'}}{\prod_i \rho_i^{p-1}}$, where p is the first neighbor connectivity, and i and i' label pairs of sites. This is the simplest combinatorial factor representation of the pair distribution function, which was generalized and extended in the CVM.

The CVM approximation extends the factorization of the probability density function to a set of larger clusters and considers the correlations among all subclusters that are included in a given maximum cluster or basic figure. For a basic cluster α, $\rho_\alpha = \prod_{\beta \subset \alpha} \tilde{\rho}_\beta$, where $\tilde{\rho}_\beta$ represents the contribution of subcluster β contained within α. Therefore, $\rho_{\text{CVM}} = \prod_\alpha \rho_\alpha^{a_\alpha}$, with the geometric coefficient of cluster a_α defined in Table I. The approximation in the CVM stems from assuming that $\tilde{\rho}_\gamma = 1$ for any cluster γ that is not contained in (or is not a subcluster of) the basic cluster α. Within this framework, earlier developments by Bragg and Williams and by Bethe are in fact special cases of the generalized variational method of the CVM.

The CE formalism provides an alternative approach for variational minimization of free energy in terms of correlation functions instead of cluster probability distribution (see Table I for the free energy equation in terms of correlation functions). Therefore, the free energy is obtained according to the following equations:

$$F = NkT[- \sum_{\alpha} \sum_{s} m_\alpha J_{\alpha s} < \phi_{\alpha' s} > _\alpha + \sum_{\alpha} b_\alpha \sum_{\sigma \subset \alpha} \rho_\alpha \ln \rho_\alpha],$$

$$\rho_\alpha = M^{-n_\alpha} \left[1 + \sum_{\beta \subset \alpha} \sum_{s} < \phi_{\beta' s} > _\beta \phi_{\beta s} \right], \quad (11)$$

$$0 = \frac{\partial F}{\partial < \phi_{\alpha s} >}.$$

DEVELOPMENT AND APPLICATIONS OF THE CLUSTER EXPANSION METHOD

Cluster Inversion Technique

The CVM was originally developed to approximate the entropy as a linear combination of multisite cluster entropies. Because it provides a direct approximation for the free energy, many early studies used the CVM to compute phase diagrams,[50–54] where the energy parameters entering the CVM free energy were determined by reproducing available experimental thermodynamic data and equilibrium concentrations in the experimental phase diagrams.

The advent of accurate electronic structure calculations enabled the determination of the ECIs from first principles, for example using DFT.[18,19] These calculations, coupled with the availability of increasingly cheaper and faster computation over the past several decades as well as robust and widely available codes, have enabled a revolution in computational materials science. Nevertheless, they are constrained by the size of the unit cell and the number of chemical species. The cluster inversion technique (also known as structure inversion) of Connolly and Williams9 provided a way to simulate the effect of chemical disorder on thermodynamic quantities of interest by interpolating from calculations (or even measurements) of ordered structures in the orthogonal space of the CE formalism.

In this technique, the ECIs are obtained by directly inverting the following linear algebra equation (rewritten from Eq. 1),

$$\Pi \mathbf{J} = \mathbf{P}, \quad (12)$$

where the scalar property $\mathbf{P}$ is calculated for as many structures as the number of truncated cluster coefficients $\mathbf{J}$, and Π is a matrix containing the cluster correlation functions. The structure inversion method is a determined linear problem and has been used for phase diagram calculations, for example, in Ref. 56. Later studies showed that solving the overdetermined linear problem of Eq. 12, where the number of data points, e.g., formation energies of

various atomic configurations in a lattice, is larger than the model parameters (the ECIs), through a least-squares procedure or a similar method, is a more efficient and accurate approach for determining the ECIs.[10,12] Within the aforementioned approach, the central task of constructing a cluster expansion consists of generating an initial set of training data and calculating a fit to ECIs. The predictive capability of the CE can be assessed in different ways, some of which are discussed in "Optimal Clusters" section. Several studies utilized the CE method alongside first-principles calculations for computing phase diagrams; for example, see Refs. 9,11–16. Due to the generality of the CE method to describe any scalar lattice property, some studies have used it to approximate other crystal properties such as the short-range diffuse intensity distribution for binary ordered alloys,[5] the interplay between magnetic properties and chemical ordering,[57] and the electronic bandgap.[6] Some studies utilized the cluster expansions for low-symmetry systems such as thin films, surfaces, and nanoparticles.[21,58–61]

Optimal Clusters

The CE is in principle exact if all possible clusters are included in the expansion, but its success is due to the fast convergence with a finite and relatively small number of compact clusters. Initially there was little understanding of the artifacts induced by the truncation of the expansion at a maximum cluster, and the selection of clusters was often done ad hoc for each system. Other sources of error were not well characterized either, such as the uncertainty in the first-principles data used in the fit. Additionally, the effects of atomic relaxations due to atomic size mismatch or more subtle phenomena sometimes resulted in the model's being unable to correctly predict the quantity being expanded for structures that were used to fit the model, such as in the case of the Cu–Pd system.[62] Sometimes this is called the "parrot test" since any acceptable model should accurately predict data points that were used in its training (repeat them back). In one of the first attempts to remove human judgment from the CE, Garbulsky and Ceder introduced additional constraints on the ECI[12] that optimized the CE orthogonal space itself. Two of these constraints are (cf. Eq. 1)

$$\sum_{\alpha} m_{\alpha} J_{\alpha} \bar{\phi}_{\alpha}(\boldsymbol{\sigma}) \leq \boldsymbol{P_j}(\boldsymbol{\sigma}) + \Delta_j \,, \tag{13a}$$

$$\sum_{\alpha} m_{\alpha} J_{\alpha} \bar{\phi}_{\alpha}(\boldsymbol{\sigma}) \geq \boldsymbol{P_j}(\boldsymbol{\sigma}) - \Delta_j \,, \tag{13b}$$

where the subscript j denotes a particular atomic configuration that is used to fit the model and Δ_j is the error bar of data point $P_j(\boldsymbol{\sigma})$, thus the model is optimized with respect to the number and order of the clusters used in the expansion. When the constraints are implemented in the CE, convergence can be obtained with between 10 and 20 ECIs obtained from between 30 and 50 atomic configurations.

Traditional techniques from frequentist statistics were first applied in a systematic way to the optimization of the CE by van de Walle and Ceder.[20] It can be seen in Eq. 13 that two distinct challenges exist: selection of optimal clusters and selection of optimal atomic configurations. Cross-validation,[63,64] along with certain heuristics such as the inclusion of clusters of lower order before those of higher order, was used to determine the optimal clusters with minimum human intervention, although some input, such as the size and the order of the largest cluster to consider, is still required since the number of ECIs is in principle infinite (Fig. 3). Cross-validation consists of leaving one (or more) data points at a time out of the dataset used to fit the model parameters (typically called the "training set" in machine learning settings), then using the model to sequentially predict the values of the data point(s) that was (were) left out (typically called the "validation set"), and using the errors of the predictions to determine how well a model "generalizes," that is, how well it predicts values that are not in its training set. In the case of the clusters to include in the CE, the cross-validation score is used to find the best combination of clusters that satisfies the heuristics and user input. The energies of the atomic configurations are not known a priori, so cross-validation cannot be used to select the best configurations, but it is shown in Ref. 20 that the trace of the covariance matrix of the ECI can be used to estimate the variance of the energies predicted from a CE and hence identify the configuration that maximizes the reduction of the trace. These ideas are implemented in ATAT.[27]

Bayesian statistics[65] have also been used in conjunction with the CE method and are particularly appropriate for systems with low symmetry such as nanoparticles and surfaces that with the traditional methodology would require a large number of ECIs and hence a large training dataset. The main difference between the Bayesian and frequentist interpretations of probability is philosophical, but it has real-life consequences: while frequentism assumes that model parameters have *true* values that can be determined from data (even if there is uncertainty or noise), Bayesianism holds that there is a probability distribution for the range of values of the model parameters determined by the data that has been observed. In practice, the key difference between these two approaches is that in Bayesian inference the probability distribution of the values of a parameter is updated each time a new observation is made or new data become available, and consequently a probability distribution (called the "prior") is required even before any data has been observed. Following the methodology developed by Mueller and Ceder,[21] consider a training dataset consisting of a vector of output

values $\mathbf{y}$ and a matrix of input values X where the jth element of $\mathbf{y}$ is the quantity of interest with atomic configuration j and the elements of X are given by

$$X_{j\alpha} = m_\alpha \bar{\phi}_{j\alpha}(\boldsymbol{\sigma}). \quad (14)$$

The probability density for the optimal ECI is $Q(\mathbf{v}|X,\mathbf{y})$, where the variable $\mathbf{v}$ is defined over all the possible values of the ECI, and $Q(\mathbf{v}|X,\mathbf{y})$ can be estimated from Bayes' theorem[65] if a prior $Q(\mathbf{v}|X)$ is known or can be estimated. The key insight in Ref. 21 is that property predictions are close to those that can be predicted by a simple model and the prediction of the simple model can be used as the prior; For example, the energy of an alloy can in principle have any real value, but in reality it will be close to the average of the energy of the pure substances, and the implementation of this knowledge in the model can drastically reduce the amount of data required to converge the CE. More recently, the Bayesian approach has been used to study nanostructures[66,67] and has been combined with compressed sensing for high-throughput computational thermodynamics.[24]

Advanced machine learning techniques have been employed in an attempt to transcend the linear interpolation embedded in the standard CE; For example, Natarajan and Van der Ven utilized a neural network, leveraging its nonlinearity to describe distinct local orderings in a lattice.[25] While neural networks are widely used for describing atomic potentials, Natarajan and Van der Ven implemented them within the CE formalism by using the correlation functions as inputs to the neural network. Instead of describing the crystal energy as a linear expansion of correlation functions, the neural network can represent the lattice site energy with a nonlinear function of multisite correlation functions. Their framework obviates the truncation error in the linear cluster expansion and can describe complex, multibody interactions by involving smaller clusters.

Chemical Disorder and Special Quasirandom Structures

One of the most useful advantages of the cluster inversion technique is that the properties of alloys with chemical disorder (which are challenging to study using DFT) can be predicted using Eq. 12 even when only ordered structures (which are readily studied using DFT) are used in the system of linear equations. It can be proved by induction that, for a random binary alloy,

$$< \bar{\phi}_{\alpha s} >_{\mathrm{D}} = (x_A - x_B)^{n_\alpha}, \quad (15)$$

where x_A and x_B are the normalized concentrations of A and B atoms, n_α is the number of sites in cluster α, and "D" denotes that the structure is chemically disordered. This derivation reveals that the first few interaction coefficients of the CE have a natural interpretation for binary alloys. The zeroth order is the value of the expanded quantity (energy, etc.) for a disordered system with an equiatomic composition, the first-order coefficients carry the effect of composition (+1 for $x_A = 1$ and −1 for $x_A = 0$), and the second-order coefficients carry the effect of ordering (from +1 for probabilistic certainty that atoms of the same chemical species will be found in a given orbit to −1 for probabilistic certainty that atoms of the same chemical species will not be found in a given orbit). The cluster inversion technique has been influential in the development of computational methods to study chemically disordered systems,[68] several of which are discussed next, but has also been used in experiments, for example, for the interpolation of phonon density-of-states curves of several binary alloys.[69–72]

Consider Eq. 15 for the case $x_A = x_B = 0.5$. In a state of perfect disorder, all the correlation functions vanish except for the zeroth order, but this is difficult to achieve in periodic systems, particularly if the repeating cell (supercell) has a small number of atoms. Given an arbitrary configuration of atoms, how well the configuration approaches randomness can be quantified by $\sqrt{< \phi_{\alpha' s} >_\alpha^2}$. Zunger et al. generated structures in which the positions of the atoms in a finite repeating cell were selected randomly and noticed that, although the standard deviations of the correlations decreased as the size of the supercell increased, they did not decrease rapidly.[73] They introduced the concept of the SQS, which is an atomic configuration that minimizes the correlation functions for a given structure and size. In practice, a structure for which all the correlation functions vanish might not exist, but minimizing the correlation functions of smaller clusters (in terms of both the number of sites in the cluster and its physical size) even at the expense of bigger clusters results in better approximations of randomness because smaller clusters have a greater impact on the quantity being expanded if the CE is to converge.

An SQS can be designed for any composition and degree of order, so the work of Zunger et al. opened new research directions in computational materials science, both in the development of methods to generate SQS structures and in applications for predicting the properties of alloys with chemical disorder, particularly from first principles.[74] Several SQSs developed early on are likely overrepresented in the scientific literature, e.g. body-centered cubic,[75] hexagonal close-packed,[76] and face-centered cubic.[77] Historically, SQSs have been generated by producing an exhaustive enumeration of structures and correlation function values that are then used to select the best one, but modern methods of designing SQSs take advantage of probabilistic global optimum search algorithms, e.g., Ref. 78. The thermodynamic and electronic structure

properties of many classes of disordered materials have been calculated via SQSs, such as chemical disorder in semiconductors,[79] metals,[80] perovskites,[81] and high-entropy alloys (HEAs).[82] Magnetic disorder has been simulated using magnetic SQSs in transition-metal nitrides[83] and magnetic unaries.[84] A comparison between the electronic structure of isotropic equiatomic FeV calculated using the mean-field theory coherent potential approximation (CPA) and a small SQS is presented in Ref. 85.

Structural Alloy Design

For many structural applications, the alloy design requires a profound knowledge of the role of defects in the mechanisms of deformation, precipitate hardening, and solid-solution strengthening. The inherent capacity of the cluster expansion framework to describe an arbitrary degree of configurational order (or disorder) makes it a useful model for studying the interactions of alloying elements, alloy composition, and structural defects. In particular, cluster expansions can be straightforwardly employed to represent the energy of a system in the presence of antiphase boundaries (APBs), because APBs, unlike many other defects, do not break the underlying lattice symmetry. An APB is a planar defect that can be characterized as a collection of antisite point defects on the same parent lattice as the perfect system. These planar defects are generated by dislocation slips during order–disorder transitions in many structural alloys. The combination of retained lattice symmetry in the presence of APBs and the role of (partial) ordering in their formation makes CE a useful means to study the thermodynamics and energy of formation of APBs. This is because the CE formalism explicitly considers the configurational disorder and requires a distinct underlying lattice. Of critical importance is to study the planar defect energy of an APB, because it is a measure of resistance to dislocation motion, with a low-energy APB associated with a high likelihood of dislocation slip. Thereby, many studies focus on enhancing APB energy, achieved via impurity solute addition, for example, as a strengthening strategy.

A number of studies have used CE as a high-fidelity surrogate model to perform Monte Carlo simulations of APB equilibrium, where direct quantum-mechanical-based Monte Carlo simulations become intractable. van de Walle and Asta pioneered the use of CE to parameterize the configurational dependence of energy in hcp-based Ti-Al alloy.[86] They obtained short-range order parameters from Monte Carlo simulations performed on the CE-parameterized alloy energy, which were then used to obtain the diffuse antiphase boundary (DAPBs) energies.[86] In another study, Sun et al. employed CE to directly calculate the energy of Ni_3Al in the presence of an APB and study the effect of nondilute Ti and Hf impurities on APB energies.[87,88] Dodaran et al. studied the effect of multiple alloying elements and temperature on the APB energies of the γ' precipitates ($L1_2$-ordered structures) in Ni-based superalloys.[89] The ATAT toolkit has an implementation of the DFT-based CE that can be combined with Monte Carlo sampling to study APBs in arbitrary crystal structures and alloy compositions.[88]

Natarajan and Van der Ven recently introduced an extended CE formalism that can be utilized to quantify a generalized stacking fault energy in alloys.[90] This new approach can significantly benefit the design of many structural alloys, where understanding the interrelation of chemical disorder in the alloy and the stacking fault energy can elucidate crystal growth and plastic deformation mechanisms. A more detailed description of this framework is presented in "Defects" section.

Unlike APBs, other types of planar defects break the underlying symmetry of a perfect lattice. Accordingly, to employ the CE in the presence of planar defects, additional considerations are needed to define the cluster expansion framework. In what follows, we review a number of studies that have employed the CE to study surface problems, including adsorbates ordering, surface segregation and ordering, and surface phase diagrams,[58,91–95] as well as precipitation with coherent and incoherent interfaces.[96–106] Detailed chemical and structural characterization of surfaces is essential in structural alloys for a microscopic understanding of the stability of surface phases. Aside from structural applications, a detailed understanding of surface behavior is needed in many applications such as substrates for thin-film growth, surface reactions, and catalysis. Understanding the thermodynamics and kinetics of precipitation in structural alloys is essential to fully understand precipitate hardening and its mechanisms. Precipitate–solid solution interfaces play a crucial role in the mechanical properties of an alloy.

To employ the CE framework to study surfaces, a CE for the surface (or surface-cluster expansion) is constructed in addition to the CE for the bulk. As expected by the broken symmetry at the surface, larger numbers of ECIs are needed to describe the surface energy with the same level of approximation (or truncation) as the bulk CE; For example, Drautz et al.[58] used a CE framework to study the competition between the tendency for Al segregation and ordering in the Ni–Al system, where ordering is induced by the symmetry breakage at the surface. They showed that the 24 ECIs used to describe the bulk Al–Ni system split into 400 inequivalent ECIs at the surface, although many have negligible ECIs and can be disregarded (see details in Ref. 58). In another study, Han et al. employed a CE framework to simultaneously study the segregation and surface ordering tendency due to the presence of oxygen adsorbates on the (111) surface of dilute Ru in Pt.

They used a coupled cluster expansion framework, in which the energy of the system is defined by two coupled binary disorder subsystems, namely the alloy surface (Pt or Ru occupation of sites) and the adsorbate layer (oxygen adsorbed or not).[92] Another study by Lerch et al. extended the CE framework to include an arbitrary number of H adsorbate sites on Ir(100) surface.[93]

For precipitation study in alloys via the CE framework, one should consider the misfit strain energy (if any) aside from the interface energy arising from the broken symmetry at the interphase interface; For example, studies in Refs. 98,99,101,102 employed a mixed-space CE to study the thermodynamics and time evolution of the distribution of precipitate shapes and sizes in Al-Zn alloy. The mixed-space CE combines a pair cluster expansion transformed to the reciprocal space with a many-body cluster (i.e., three-body or higher) expansion in the real space. Pair cluster representation in the reciprocal space enables a converged description of even long-range pair interactions. More details are described in "Atomic Relaxations" section. The mixed-space CE is used in Monte Carlo or kinetic Monte Carlo simulations, respectively, to understand the effect of temperature on the size and shape of precipitates or their kinetics. In these studies, the mixed-space CE includes an additional constituent strain energy term, which is defined as the strain energy of bulk Al and Zn needed to maintain the coherency along the interface. The strain energy term is calculated by deforming the bulk fcc Zn and Al from their equilibrium lattice constants to a common lattice constant. This term describes the anisotropic long-range coherency strain energy. Studies in Refs. 100 and 105 employ the standard CE formalism to account for the chemical disorder effects without explicitly including the interface and strain energies in the CE framework. Instead, the interface and misfit strain effects are included via other means, such as the parameters of a phase-field model[105] or direct DFT calculations.[100]

Code Implementations

There are various code implementations of the cluster expansion formalism, which differ mainly in the fitting procedure for obtaining the ECIs. ATAT,[27] which implements the CE fitting procedure according to the approach in Ref. 12 as part of an automated phase diagram calculation framework from first-principles data, was one of the first open-access implementations of the CE for phase diagram calculations, and is constantly being extended.[107,108] The Cluster Approach to Statistical Mechanics (CASM) code implements a parameterization of the CE that incorporates the lattice-dynamical degrees of freedom in addition to the configurational variation in the standard CE[109,110] (see Eq. 19 in "Lattice Vibrations and Distortions" section). The Piecewise Polynomial Potential Partitioning code (P4)[111] implements the coarse-grained cluster expansion as part of ATAT (see Eq. 21 in "Beyond the Conventional Cluster Expansion Method" section). More recent implementations of the CE, such as the Integrated Cluster Expansion Toolkit (ICET)[112] and the CLuster Expansion in Atomic Simulation Environment (CLEASE),[113] which is integrated with ASE,[28] enable advanced statistical techniques for the ECI training and validation procedures, such as linear regression algorithms (with and without regularization), Bayesian regression, feature selection, and cross-validation (Table II).

Integration with CALPHAD

The calculation of phase diagrams (CALPHAD) method[114] is widely used in industry and academia to calculate phase equilibria in multicomponent systems. The method is mature and robust, with both proprietary and open-source codes and databases. Its mathematical formalism[115] introduces terms to account for interactions between chemical species that depend on each of the components (unaries) of the system, and this increases the sensitivity of the model dramatically with the number of components, limiting the practical applicability of CALPHAD to binaries and ternaries. The databases consist of experimental or computational data that can be used to describe the Gibbs free energies of each individual phase in a compound. Experimental data are typically too sparse to accurately characterize the free energies of the relevant phases, including metastable ones, so must be augmented with results from first-principles calculations, empirical models, or machine learning models.[116] While CALPHAD is not as inherently compatible with DFT calculations as the CE method (see "Cluster Inversion Technique," "Optimal Clusters," and "Chemical Disorder and Special Quasirandom Structures" sections), it can integrate the predictions of any model, including DFT.[78,117,118]

BEYOND THE CONVENTIONAL CLUSTER EXPANSION METHOD

Long-Range Order

One major practical issue with the CE method is the truncation of cluster variables to a finite-size cluster. While in principle an infinitely large cluster represents the long-range order in the CVM approximation, in practice a truncated cluster expansion is used. Smaller clusters such as the pair or three-body clusters are a good representation of short-range order in the solid. However, by restricting the cluster variation parameters to an arbitrary finite-size cluster, the long-range ordering parameter is overlooked, which is the major variational parameter in earlier developments such as the Bragg–Williams model (Table I). To overcome this issue, Laks et al.[6] presented a general model that combines the long-range order parameter with a finite-

Table II. Comparison of available implementations of the CE formalism and its extensions

Software	CE optimization scheme	Characteristics
ATAT	1. Bayesian regression and cross-validation; 2. Least-squares fitting	1. Reproduces energy values included in the fit; mitigates truncation problem; 2. Minimizes the mean-squared error
ICET	1. Least squares, LASSO, Bayesian ridge regression; 2. Recursive feature elimination	1. Applies different l_1 regularization schemes; 2. Eliminates less important ECIs starting from a full set of parameters
CLEASE	1. l_1 and l_2 regularization, Bayesian compressive sensing; 2. Genetic algorithm	1. Applies different l_1 regularization schemes to promote sparsity; 2. Cluster selection based on a genetic algorithm.
CASM	Combines the vibrational degrees of freedom involved in structural phase transitions with configurational disorder in multicomponent solids to approximate the Hamiltonian.	Goes beyond conventional CE by including deformation in clusters.
P4	Builds a coarse-grained CE to integrate the vibrational degrees of freedom involved in structural phase transitions with configurational disorder in multicomponent solids.	Goes beyond conventional CE by coarse-grained parameterization of free energy on an augmented lattice.

size cluster expansion to describe any lattice property. To this end, a general correlation function is defined in terms of multisite correlation functions $\bar{\phi}_f(\boldsymbol{\sigma})$ and the long-range order parameter η; For example, the general correlation function for the pair basic figure f is defined as

$$\bar{\Pi}_f(x,\eta) = (2x-1)^2 + \eta^2\left[\bar{\phi}_f(\boldsymbol{\sigma}) - (2\boldsymbol{X}_{\boldsymbol{\sigma}} - \boldsymbol{I})^2\right], \quad (16)$$

where $\bar{\Pi}_f(x,\eta)$ is the generalized correlation function, $\bar{\phi}_f(\boldsymbol{\sigma})$ is the multisite correlation function (or the average of spin products of all subclusters in figure f), x is the normalized concentration of the binary alloy, and $2X_\sigma - 1$ is the Fourier transform of the spin variables σ_i measured at zero (more details can be found in Ref. 6).

Atomic Relaxations

Another practical limitation of the conventional CE technique is that it converges slowly for relaxed crystal structures, and this reduces its utility in the study of atom size-mismatched alloys (alloys with chemical species that have very different atomic radii). However, the effects of atomic relaxation are too important to neglect, particularly for alloys with large size-mismatch effects. To incorporate relaxations into the CE formalism, Laks et al.[119] introduced a reciprocal-space representation of the CE, in which the individual real-space interaction energy coefficients J_f (where f represents the cluster type) are replaced by a single reciprocal-space function $J(\mathbf{k})$. While in real space the ECIs of different figures are not related, a "smoothness" condition imposed upon the single reciprocal-space function allows to minimize the gradient of $J(\mathbf{k})$, which results in a requirement that pair interactions decrease rapidly for large distances and a requirement that pair figures that are not rigorously necessary to improve the fit are assigned an ECI of zero. This results in a major reduction in the number of coefficients to be fit within the CE formalism and the consequent rapid convergence of the reciprocal-space CE. In addition, it can correctly predict the energies of arbitrary relaxed structures. The following equations represent the recast of the CE for pair interactions into reciprocal space. The conventional CE:

$$E_{\mathrm{CE}}(\sigma) = N\sum_f m_f J_f \bar{\phi}_f(\sigma)\,, \quad (17)$$

where m_f is the degeneracy of cluster f per site and the sum is over f, all the pair type interactions. The reciprocal CE:

$$E_{\mathrm{CE}}(\sigma) = N\sum_{\mathbf{k}} J(\mathbf{k})|\hat{\sigma}(\mathbf{k},\sigma)|^2\,,$$

$$\hat{\sigma}(\mathbf{k},\sigma) = \frac{1}{N}\sum_l^N \sigma_l(\sigma)e^{i\mathbf{k}.\mathbf{R_l}}\,, \quad (18)$$

$$J(\mathbf{k}) = \frac{1}{2}\sum_n^N J_{0,n}e^{i\mathbf{k}.\mathbf{R_n}}\,,$$

where $\mathbf{k}$ runs over the first Brillouin zone, and $J(\mathbf{k})$ and $\hat{\sigma}(\mathbf{k},\sigma)$ are the Fourier transforms of the interaction energy coefficients (ECIs) and spin products, respectively. Later, a mixed-based cluster expansion in which a deterministic and direct map from a first-principles configuration Hamiltonian to an Ising-like Hamiltonian was developed.[120]

Lattice Vibrations and Distortions

The variational parameters in the standard CE framework are rigid cluster figures, which can only describe the Hamiltonian dependence on the occupation variation of an ideal lattice structure. On the other hand, prediction of finite-temperature thermodynamic properties, thermal conductivity, and structural phase transitions depends on describing the variation of free energy in terms of atomic displacements in addition to configuration variation. Thomas and Van der Ven[110] developed a cluster-based model in which individual clusters are free to undergo deformation and the potential energy is expanded in terms of deformed clusters as follows:

$$E = E_0 + \sum_{\alpha} E^{\alpha}(\tilde{U}^{(\alpha)}), \tag{19}$$

where E_0 is the constant energy term, $E^{\alpha}(\tilde{U}^{(\alpha)})$ is the contribution of cluster α to the total crystal energy, and $U^{(\alpha)}$ is the deformation of cluster α. An arbitrary deformation of cluster α is described as a linear combination of $3N_{\alpha} - 6$ terms (invariant under rigid translations or rotations of the crystal and similar to normal modes of vibrations) that are measured with reference to the ideal lattice structure. An earlier development by Geng et al.[121] provided a hybrid cluster expansion for local structural relaxation by combining pair potential models with the CE method. To expand upon the conventional CE based on a fixed ideal lattice, each site is characterized not only by its occupation variable but also by its displacement vector. This model was used to expand the pair interactions based on hybrid cluster expansion, while multibody terms were expanded using a standard CE formalism.[121]

Kadkhodaei, Hong, and van de Walle[122] provided an alternative approach to encompass finite-temperature atomic displacement effects in the free energy expansion. They introduced a coarse-grained cluster expansion framework which is built upon an "augmented lattice," L_{aug} (Fig. 5). In addition to the high-symmetry ideal sites, the augmented lattice includes sites that correspond to local distortions of the ideal lattice, which implicitly permits cluster deformations. The local vibrations in the vicinity of lattice distortions (where the vicinity of distortions is defined by the Wigner–Seitz cells of the augmented lattice and denoted as ζ_{σ}) are continuously sampled using a coarse-grained formulation of the free energy (consisting of a constrained vibrational free energy F^*_{σ} for each configuration σ) as follows:

$$F = -kT \ln \sum_{\sigma \in L_{\text{aug}}} e^{-F^*_{\sigma}/kT},$$

$$F^*_{\sigma} = -kT \ln \int_{\mathbf{r} \in \zeta_{\sigma}} e^{-V(\mathbf{r})/kT}, \tag{20}$$

where $V(\mathbf{r})$ is the potential energy in terms of atomic position $\mathbf{r}$. Within this nested framework, the constrained vibrational free energy is expanded using cluster figures on the augmented lattice as follows:

$$F^*_{\sigma} = \sum_{\alpha} m_{\alpha} J_{\alpha}(T) < \prod_{i \in \alpha} \sigma_i >_{\alpha'}, \tag{21}$$

where m_{α} is the number of symmetrically equivalent clusters α, $J_{\alpha}(T)$ are cluster interaction coefficients that are temperature dependent, and the average $< \prod_{i \in \alpha} \sigma_i >$ is the correlation function of cluster α on L_{aug}. The sum in Eq. 21 is over symmetrically distinct clusters α, while the average is over clusters α' that are symmetrically equivalent to α.

Defects

The original development of the CE is based on a perfect lattice structure, which neglected the presence of point defects such as vacancies. Van der Ven and Ceder[123] developed a perturbative approach to model the equilibrium vacancy concentration and the short-range order among vacancies and other chemical species in alloys. They invoked a local cluster expansion as a perturbation to the standard CE of the energy, assuming that vacancy concentration is sufficiently dilute to neglect the interactions among vacancies. They parameterized the effective vacancy formation energy ΔE_i^{eff} with a local CE according to

$$\underbrace{\Delta E_i^{eff}}_{\text{perturbativelocalCE}} = E_i^V(\sigma) - \frac{1}{2} \underbrace{\left[E_i^A(\sigma) - E_i^B(\sigma)\right]}_{\text{standardCEw/ovacancies}},$$

where $E_i^V(\sigma)$ is the energy of the crystal with configuration σ but with a vacancy occupying site i, and $E_i^{A/B}(\sigma)$ is the energy of the crystal with configuration σ but with site i occupied by A (or B). Van der Ven et al.[124] also utilized the concept of local cluster expansion to extend the standard CE to describe the activation energy barriers for ionic migrations in configurationally disordered solids. They introduced a kinetically resolved activation (KRA) barrier, ΔE_{KRA}, which describes the average of activation barriers for all possible diffusive hops of an ion at a given site. As the ΔE_{KRA} depends on the local configuration around the migrating ion but is independent of hop direction as it averages over all ionic hops, it can be described using a cluster expansion in terms of kinetic effective cluster interactions (KECI) K_{α} and cluster correlation functions $\bar{\phi}_{\alpha}$:

$$\Delta E_{KRA} = K_0 + \sum_{\alpha} K_{\alpha} \bar{\phi}_{\alpha}. \tag{22}$$

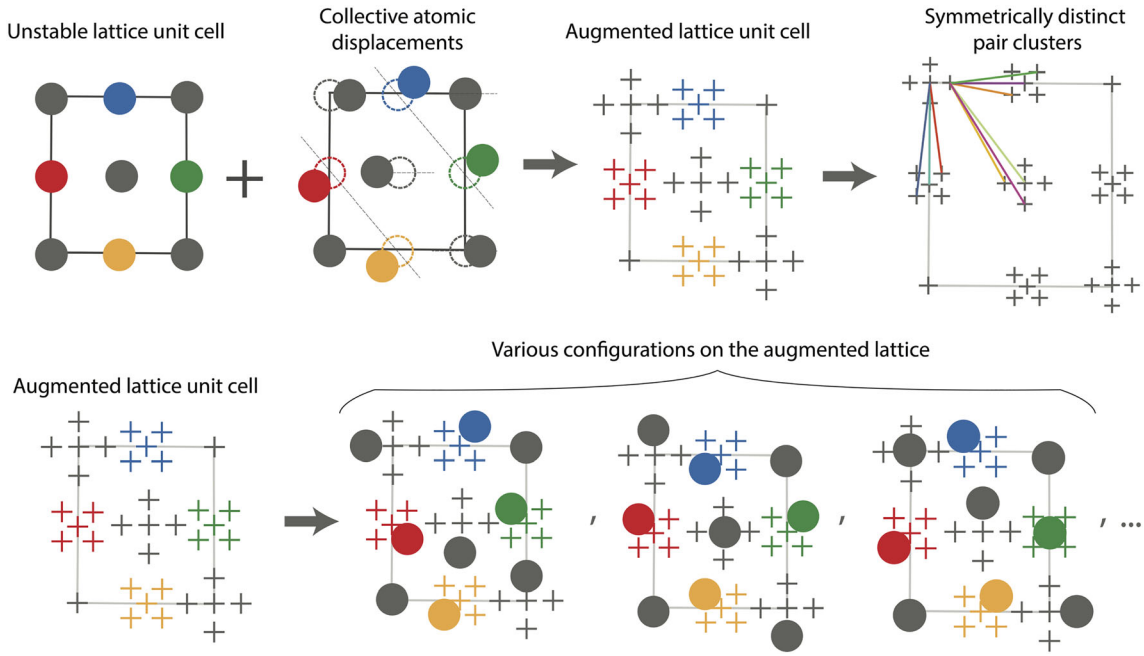

Fig. 5. Illustration of the Piecewise Polynomial Potential Partitioning (P^4) method to embed finite-temperature lattice distortions in a coarse-grained CE formulation. (Top row) The augmented lattice consists of atomic positions related to the collective distortions of the unstable lattice in addition to the ordinary basis points. Some of the symmetrically distinct pair clusters on the augmented lattice are marked with different colors. (Bottom row) Sampling of various configurations on the augmented lattice describes occupation variation as well as vibrational distortions (Color figure online).

A later study by Zhang and Sluiter[125] presented a method that ensures that strictly local CEs, e.g., for KRA energy barriers, yield only positive energy barriers in vacancy-mediated diffusion in substitutional alloys.

Natarajan and Van der Ven introduced the configurationally resolved planar fault (CRPF) energy notion, which enables extension of the standard CE formalism to quantify the dependence of generalized stacking fault (GSF) energies (a class of planar defects) on chemical disorder in alloys.[90] Within this framework, an arbitrary GSF with a glide vector of $\mathbf{r}$ is obtained upon application of a combination of three glide vectors, $\mathbf{r}_1$, $\mathbf{r}_2$, and r_3, which coincide with the elementary transnational vectors of the underlying lattice, with $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_3$ representing the configurations associated with each glide vector, respectively. Accordingly, the energy of the GSF, $E(\mathbf{r}, \delta, \boldsymbol{\sigma})$, is decomposed into an average configurational energy, $E^{ave}(\mathbf{r}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3)$, and a CRPF energy, $E^{CRPF}(\mathbf{r} - \mathbf{r}_1, \delta, \boldsymbol{\sigma}_1)$, which characterizes the local excess energy arising from the planar defect:

$$E(\mathbf{r}, \delta, \boldsymbol{\sigma}) = E^{CRPF}(\mathbf{r} - \mathbf{r}_1, \delta, \boldsymbol{\sigma}_1) + E^{ave}(\mathbf{r}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3) \,. \tag{23}$$

Here, δ represents the magnitude of the separation perpendicular to the glide plane by which the crystal is cleaved at the GSF. E^{ave} relates the defected system energy to the energies of the three nearest orderings $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_3$ relative to the perfect crystal configuration $\boldsymbol{\sigma}$, and E^{CRPF} is the excess energy arising from a translation of $\mathbf{r} - \mathbf{r}_1$ and a separation of δ that map the defected configuration to the $\boldsymbol{\sigma}_1$ configuration. This energy decomposition is valid under the assumption that the local ordering of atoms far away from the fault plane is unaffected by the glide.

OUTLOOK AND CONCLUSION

Various trends are apparent in the historical development of the CE formalism, and we expect several of them to continue. One of them is the movement towards generalization: the improved modeling of more phenomena based on more rigorous mathematics. An example of this is the evolution from nonorthogonal bases computed from cluster distribution functions in the CVM,[1] to full orthogonal bases originating from discrete Chebyshev polynomials that describe distinct clusters in the CE,[3] to the tensorial cluster expansion developed in Ref. 126. The modeling of phenomena that break crystal symmetries, such as relaxations from atom size mismatch or thermal displacements, remains a challenge, but these broken symmetries are crucial for understanding and predicting the finite-temperature properties of many functional and structural materials. The cluster expansion of an anharmonic Born–Oppenheimer potential energy surface such as that developed in Ref. 110 and the coarse-grained cluster expansion of Ref. 122, which incorporates an augmented lattice that

includes lattice distortions, are distinct steps in this direction. Future variations might take advantage of recent progress in the group-theoretic representation of phonon modes in crystalline materials[127] or the graph-theoretic use of marginalized graph kernels to measure similarity among atomic configurations.[128] Several systems are prone to benefit from advances in this area, among which magnetic systems[129,130] and HEAs[131] are perhaps at the forefront.

Another trend is the formalization and optimization of penalty functions, which results in increasingly better predictions for a given number of clusters and computations, starting with the inclusion of algebraic constraints[12] and moving on to statistical techniques to improve predictions such as cross-validation[20] and Bayesian priors,[21,22] which can significantly reduce the amount of data required to train a CE model. Much of the theory is well established, but code implementations are not always user-friendly. ATAT[27] is one of the most mature codes and provides a consistent framework to construct cluster expansions from first principles as well as SQS support, tensorial cluster expansions, etc. and it interfaces with popular DFT codes such as the the Vienna Ab Initio Simulation Package (VASP) and Quantum Espresso (QE). Other codes are in earlier stages of development, but we expect some consolidation and code that is more easily extendable to a variety of more complicated systems.

The third trend is the proliferation of data, particularly computational data generated via DFT, which is highly compatible with CE approaches. Most of these data are of high quality and the datasets are "small," often generated by individual research groups to investigate targeted systems or phenomena, and ontologically different from what has been termed "Big Data" (see, e.g., Ref. 132 for an enumeration of the ontological differences). There is currently an important effort and investment by governments and companies to develop computational pipelines, collaboration infrastructure and logistics, data standards, etc. to encourage and allow the community to share and reuse computationally expensive calculations. One example is the Materials Project,[133] which is part of the Materials Genome Initiative, a major program of the US government to accelerate the pace of materials discovery.[134] These projects aim to create large datasets of "small" data that can take advantage of recent developments in computation and data science precipitated by "Big Data," and a transparent example is the high-throughput computational design of materials,[135] which is quickly approaching maturity. The CE method has already been integrated into several pipelines and is used in conjunction with data science techniques such as data mining,[136] and we expect even closer integration in the future.

The fourth trend is the convergence of cluster techniques and machine learning, which can go in both directions. Technologies that have recently become widespread such as neural networks have been combined with the CE, learning for example the correlation functions produced by arbitrary local atomic orderings,[25] or as part of active learning pipelines.[137] The CVM has found uses in areas such as computer vision and the theory of error-correcting codes,[138–140] which share with CVM in crystal lattices the requirement that marginal probabilities need to be computed. These probabilities can be represented by probabilistic graphical models such as Bayesian networks and Markov random fields (see, e.g., Ref. 141 for a description of both). At a fundamental level, the ideas behind CE are not too different from those of methodologies that are used for machine learning: to map inputs into a high-dimensional feature space that can more clearly separate the points and perform an interpolation. This is true of support vector machines (SVMs),[142] neural networks,[143] and others. While the CE methodology has been preferred in the past several decades over the CVM due to its mathematical simplicity, new developments such as physics-informed neural networks (PINNs)[144] which respect the laws of physics stated as time-dependent and nonlinear partial differential equations might find implementations in CVM. It is easy to predict that cluster-based methods will continue to find applications in computational materials science and will continue their coevolution with machine learning and data science.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ABBREVIATIONS AND VARIABLES

Abbreviations

APB	Antiphase boundaries
ASE	Atomic Simulation Environment (see Ref. 28)
ATAT	Alloy Theoretic Automated Toolkit (see Refs. 27,107,108)
CASM	Cluster Approach to Statistical Mechanics (see Refs. 109,110)
CLEASE	CLuster Expansion in Atomic Simulation Environment (see Ref. 113)
CE	Cluster expansion
CVM	Cluster variation method
DAPB	Diffuse antiphase boundary
DFT	Density functional theory
ECI	Effective cluster interaction
HEA	High-entropy alloy
ICET	Integrated Cluster Expansion Toolkit (see Ref. 112)

KECI	Kinetic effective cluster interaction
KRA	Kinetically resolved activation
PINN	Physics-informed neural network
P^4	Piecewise Polynomial Potential Partitioning (see Ref. 111)
QE	Quantum Espresso
SQS	Special quasirandom structure
SVM	Support vector machine
VASP	Vienna Ab Initio Simulation Package

Variables

C	A cluster configuration
F_{cvm}	Free energy in the cluster variation method (CVM) (defined in Eq. 3)
$F(\rho)$	Free energy functional in terms of the probability distribution function ρ
F^*_σ	Constrained vibrational free energy (see Eq. 21)
H	Ising model Hamiltonian
J_α	Effective cluster interaction (ECI) coefficient for cluster α
$J(\mathbf{k})$	Fourier transform of J (see Eq. 18)
$\mathbf{J}_\alpha$	A vector of ECIs collected for different configurations
K_α	Kinetic effective cluster interaction (see Eq. 22)
L_{aug}	An augmented lattice that includes local distortions of the ideal lattice in addition to the ideal (high-symmetry) lattice (see Eq. 20)
M	Number of components (typically distinct chemical species) in a crystal lattice
N	Number of sites (lattice points) in a crystal lattice
N_α	Number of subclusters in cluster α (see Fig. 2)
$P(\boldsymbol{\sigma})$	A property of the lattice expanded as a function of the collection of spin variables $\boldsymbol{\sigma}$ (defined in Eq. 1)
$\mathbf{P}$	A vector of the scalar property of a lattice collected for different configurations
$Q(\mathbf{v}\|X,\mathbf{y})$	Bayesian probability density for an ECI over its possible values $\mathbf{v}$ trained with dataset $X,\mathbf{y}$
$S(\rho)$	Entropy functional in terms of the probability distribution function ρ
T	Temperature
$U(\rho)$	Internal energy functional in terms of the probability distribution function ρ
V_α	Contribution of cluster α to the internal energy in the cluster variation method (CVM)
X	A matrix of dependent variables ("input values") used in a training or test dataset (each row corresponds to an output value in $\mathbf{y}$, see Eq. 14)
Z	Partition function
a_α	Weight of the cluster α towards the total configurational entropy in the cluster variation method (CVM)
f	A basic figure
$i, i', i'', \dot{s}$	An individual lattice site
k	Boltzmann constant
m	A possible value for σ_i; it can take the value of any integer between $-M/2$ and $M/2$ except for 0 if M is even, and any integer between $-(M-1)/2$ and $(M-1) \div 2$ including 0 if M is odd
m_α	Multiplicity of cluster α, i.e., the number of symmetrically equivalent clusters α
$n, n', n'', \dot{s}$	Individual indices used in a Chebyshev polynomial Θ_n
n_α	The number of lattice sites in cluster α
n_β	The number of lattice sites in cluster β
s	The set of indices $\{n, n', n'', \dot{s}\}$ for the order of the Chebyshev polynomial Θ_n used in building a cluster function $\phi_{\alpha s}$ (see Eq. 4)
x_A, x_B	Normalized concentration of atom of type A (B) in an alloy
$\mathbf{y}$	A vector of dependent variables ("output values") used in a training or test dataset (see also X)
$\bar{\Pi}_f(x,\eta)$	Generalized correlation function for basic figure f in terms of concentration x and long-range order parameter η, (see, e.g., Eq. 16)
Π	Matrix of cluster correlation functions
$\Theta_n(\sigma_i)$	Discrete Chebyshev polynomial of order n as a function of the spin variable σ_i
α, β, γ	Clusters of lattice sites, relationships between them are explained for each model in the text (see Figs. 1 and 2)
α'	A set of clusters that are symmetrically equivalent to cluster α, i.e., clusters that belong to the same orbit as cluster α
ζ_σ	Vicinity of lattice distortions in an augmented lattice L_{aug} (see Eq. 20)
η	Long-range order parameter
ρ_α	Cluster (probability) distribution function in the cluster variation method (CVM)
ρ_{ex}	Equilibrium probability distribution function
σ_i	The spin or occupation variable of site i, associated with the type of atom at site i in a cluster, e.g., in a binary alloy, +1 if site i is occupied by atom A and -1 if occupied by atom B
$\boldsymbol{\sigma}$	The vector of spin variables for all the N sites in the crystal
$\hat{\sigma}(\mathbf{k},\sigma)$	Fourier transform of σ (see Eq. 18)
ς	The neighbor order parameter in the Bethe model; $\varsigma = 1$ for perfect neighbor order and $\varsigma = 0$ for perfect neighbor disorder

$\phi_{\alpha s}$ The cluster function corresponding to cluster α and polynomial order s (defined in Eq. 4)

$\bar{\phi}_{\alpha}(\boldsymbol{\sigma})$ The multisite or cluster correlation function for the collection of spin variables $\boldsymbol{\sigma}$ corresponding to cluster α (defined in Eq. 2)

REFERENCES

1. R. Kikuchi, *Phys. Rev.* 81, 988 (1951a).
2. J.M. Sanchez and D. de Fontaine, *Phys. Rev. B* 17, 2926 (1978). https://doi.org/10.1103/PhysRevB.17.2926.
3. J.M. Sanchez, F. Ducastelle and D. Gratias, *Physica A Stat. Mech. Appl.* 128, 334 (1984).
4. R. Kikuchi, *Prog. Theor. Phys. Suppl.* 115, 1 (1994). https://doi.org/10.1143/PTPS.115.1.
5. T. Mohri, J.M. Sanchez, and D. De Fontaine, *Acta Metall.* 33, 1463 (1985).
6. D.B. Laks, S.-H. Wei, and A. Zunger, *Phys. Rev. Lett.* 69, 3766 (1992a). https://doi.org/10.1103/PhysRevLett.69.3766.
7. G.H. Wannier, *Rev. Mod. Phys.* 17, 50 (1945). https://doi.org/10.1103/RevModPhys.17.50.
8. T. Morita, *J. Phys. Soc. Jpn.* 12, 1060 (1957). https://doi.org/10.1143/JPSJ.12.1060.
9. J.W.D. Connolly and A.R. Williams, *Phys. Rev. B* 27, 5169 (1983). https://doi.org/10.1103/PhysRevB.27.5169.
10. Z.W. Lu, S.-H. Wei, A. Zunger, S. Frota-Pessoa, and L.G. Ferreira, *Phys. Rev. B* 44, 512 (1991). https://doi.org/10.1103/PhysRevB.44.512.
11. J.M. Sanchez, J.P. Stark, and V.L. Moruzzi, *Phys. Rev. B* 44, 5411 (1991). https://doi.org/10.1103/PhysRevB.44.5411.
12. G.D. Garbulsky and G. Ceder, *Phys. Rev. B* 51, 67 (1995). https://doi.org/10.1103/PhysRevB.51.67.
13. V. Ozoliņš., C. Wolverton, and A. Zunger. *Phys. Rev. B* 57, 6427 (1998). https://doi.org/10.1103/PhysRevB.57.6427.
14. M.H.F. Sluiter, Y. Watanabe, D.D. Fontaine, and Y. Kawazoe, *Phys. Rev. B* 53, 6137 (1996). https://doi.org/10.1103/PhysRevB.53.6137.
15. A. Van der Ven, M.K. Aydinol, G. Ceder, G. Kresse, and J. Hafner, *Phys. Rev. B* 58, 2975 (1998). https://doi.org/10.1103/PhysRevB.58.2975.
16. N.A. Zarkevich, T.L. Tan, and D.D. Johnson, *Phys. Rev. B* 75, 104203 (2007). https://doi.org/10.1103/PhysRevB.75.104203.
17. N.A. Zarkevich and D.D. Johnson, *Phys. Rev. Lett.* 92, 255702 (2004). https://doi.org/10.1103/PhysRevLett.92.255702.
18. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, B864 (1964). https://doi.org/10.1103/PhysRev.136.B864.
19. W. Kohn and L.J. Sham, *Phys. Rev.* 140, A1133 (1965). https://doi.org/10.1103/PhysRev.140.A1133.
20. A. van de Walle and G. Ceder, *J. Phase Equilibria* 23, 348 (2002). https://doi.org/10.1361/105497102770331596.
21. T. Mueller and G. Ceder, *Phys. Rev. B* 80, 024103 (2009). https://doi.org/10.1103/PhysRevB.80.024103.
22. E. Cockayne and A. van de Walle, *Phys. Rev. B* 81, 012104 (2010). https://doi.org/10.1103/PhysRevB.81.012104.
23. D.L. Donoho, *IEEE Trans. Inf. Theory* 52, 1289 (2006).
24. L. J. Nelson, V. Ozoliņš, C. S. Reese, F. Zhou, and G. L. W. Hart, *Phys. Rev. B* 88, 155105 (2013). https://doi.org/10.1103/PhysRevB.88.155105.
25. A. R. Natarajan and A. Van der Ven, *npj Comput. Mater.* 4, 1 (2018). https://doi.org/10.1038/s41524-018-0110-y.
26. A. van de Walle, *JOM* 65, 1523 (2013). https://doi.org/10.1007/s11837-013-0764-3.
27. A. van de Walle, M.D. Asta, and G. Ceder, *Calphad* 26, 539 (2002).
28. A.H. Larsen, J.J. Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dułak, J. Friis, M.N. Groves, B. Hammer, and C. Hargus et al., *J. Phys. Condens. Matter* 29, 273002 (2017).
29. F. Bloch, *Z. Phys.* 61, 206 (1930). https://doi.org/10.1007/BF01339661.
30. H. A. Kramers and G. H. Wannier, *Phys. Rev.* 60, 263 (1941a). https://doi.org/10.1103/PhysRev.60.263.
31. H.A. Kramers and G.H. Wannier, *Phys. Rev.* 60, 252 (1941b). https://doi.org/10.1103/PhysRev.60.252.
32. W. L. Bragg and E. J. Williams, *Proc. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character* 145, 699 (1934), https://doi.org/10.1098/rspa.1934.0132.
33. H. A. Bethe and W. L. Bragg, *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* 150, 552 (1935). https://doi.org/10.1098/rspa.1935.0122.
34. C.N. Yang, *Variation Method J. Chem. Phys.* 13, 195 (1945). https://doi.org/10.1063/1.1698926.
35. C.N. Yang and Y.Y. Li, *Acta Phys. Sin.* 11, 59 (1947).
36. Y.-Y. Li and Y.-Y. Lr, *J. Chem. Phys.* 17, 1071 (1949).
37. T.L. Hill, *J. Chem. Phys.* 18, 441 (1950). https://doi.org/10.1063/1.1747615.
38. M. Kurata, R. Kikuchi and T. Watari, *J. Chem. Phys.* 21, 434 (1953). https://doi.org/10.1063/1.1698926.
39. R. Kikuchi, *J. Chem. Phys.* 19, 1230 (1951b). https://doi.org/10.1063/1.1748002.
40. R. Kikuchi and S.G. Brush, *J. Chem. Phys.* 47, 195 (1967). https://doi.org/10.1063/1.1711845.
41. R. Kikuchi, *J. Chem. Phys.* 47, 1664 (1967). https://doi.org/10.1063/1.1712147.
42. J. A. Barker and E. A. Guggenheim, *Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci.* 216, 45 (1953). https://doi.org/10.1098/rspa.1953.0005.
43. T. Morita ,*J. Math. Phys.* 13, 115 (1972). https://doi.org/10.1063/1.1665840.
44. D. de Fontaine, in *Theory and Applications of the Cluster Variation and Path Probability Methods, edited by J. L. Moran-Lopez and J. M. Sanchez (Springer US, Boston, MA, 1996), pp. 125–144.*.
45. J.M. Sanchez, *Phys. Rev. B* 48, 14013 (1993). https://doi.org/10.1103/PhysRevB.48.14013.
46. J.M. Sanchez, *Phys. Rev. B* 81, 224202 (2010). https://doi.org/10.1103/PhysRevB.81.224202.
47. J.M. Sanchez, *J. Phase Equilibria Diffusion* 38, 238 (2017). https://doi.org/10.1007/s11669-017-0521-3.
48. J.M. Sanchez, *Phys. Rev. B* 99, 134206 (2019). https://doi.org/10.1103/PhysRevB.99.134206.
49. R. Fowler and E. Guggenheim, *Statistical Thermodynamics* (Cambridge University Press, Cambridge, 1952).
50. R. Kikuchi, J. Sanchez, D. De Fontaine, and H. Yamauchi, *Acta Metall.* 28, 651 (1980).
51. J.M. Sanchez and D. de Fontaine, *Phys. Rev. B* 21, 216 (1980). https://doi.org/10.1103/PhysRevB.21.216.
52. J.M. Sanchez and D. de Fontaine, *Phys. Rev. B* 25, 1759 (1982). https://doi.org/10.1103/PhysRevB.25.1759.
53. C. Sigli and J. Sanchez, *Acta Metall.* 34, 1021 (1986).
54. B.P. Burton, *Phys. Rev. B* 59, 6087 (1999). https://doi.org/10.1103/PhysRevB.59.6087.
55. A. Finel, in *Statics and Dynamics of Alloy Phase Transformations* , edited by P. E. A. Turchi and A. Gonis (Springer, Boston, MA, 1994), pp. 495–540. https://doi.org/10.1007/978-1-4615-2476-2_33.
56. M. Asta, R. McCormack, and D. de Fontaine, *Phys. Rev. B* 48, 748 (1993). https://doi.org/10.1103/PhysRevB.48.748.
57. J.M. Sanchez, J.L. Moran-Lopez, C. Leroux, and M.C. Cadeville, *J. Phys. Condens. Matter* 1, 491 (1989).
58. R. Drautz, H. Reichert, M. Fähnle, H. Dosch, and J. M. Sanchez, *Phys. Rev. Lett.* 87, 236102 (2001), https://doi.org/10.1103/PhysRevLett.87.236102.
59. T. Mueller, *Phys. Rev. B* 86, 144201 (2012). https://doi.org/10.1103/PhysRevB.86.144201.
60. L.-L. Wang, T.L. Tan, and D.D. Johnson, *Nano Lett.* 14, 7077 (2014). https://doi.org/10.1021/nl503519m.
61. L. Cao and T. Mueller, *Nano Lett.* 16, 7748 (2016). https://doi.org/10.1021/acs.nanolett.6b03867.
62. Z.W. Lu, D.B. Laks, S.-H. Wei, and A. Zunger, *Phys. Rev. B* 50, 6642 (1994). https://doi.org/10.1103/PhysRevB.50.6642.

63. M. Stone, *J. R. Stat. Soc. Ser. B (Methodological)* 36, 111 (1974). https://doi.org/10.1111/j.2517-6161.1974.tb00994.x.
64. M. Stone, *J. R. Stat. Soc. Ser. B (Methodological)*39, 44 (1977).
65. T. Bayes, *Phil. Trans. R. Soc.a* 53, 370 (1763), https://doi.org/10.1098/rstl.1763.0053.
66. X. Huang, Z. Zhao, L. Cao, Y. Chen, E. Zhu, Z. Lin, M. Li, A. Yan, A. Zettl, Y.M. Wang et al., *Science* 348, 1230 (2015).
67. C. Li, D. Raciti, T. Pu, L. Cao, C. He, C. Wang, and T. Mueller, *J. Phys. Chem. C* 122, 18040 (2018). https://doi.org/10.1021/acs.jpcc.8b03868.
68. D.D. Fontaine, in *Solid State Physics, vol. 47, ed. by H. Ehrenreich, D. Turnbull (Academic, Cambridge, 1994), pp. 33–176*.
69. M.S. Lucas, A. Papandrew, B. Fultz, and M.Y. Hu, *Phys. Rev. B* 75, 054307 (2007). https://doi.org/10.1103/PhysRevB.75.054307.
70. M.S. Lucas, M. Kresch, R. Stevens, and B. Fultz, *Phys. Rev. B* 77, 184303 (2008). https://doi.org/10.1103/PhysRevB.77.184303.
71. M. S. Lucas, J. A. Mu noz, L. Mauger, L. C. W., A. O. Sheets, Z. Turgut, J. Horwath, D. L. Abernathy, M. B. Stone, O. Delaire, et al., *J. Appl. Phys.* 108, 023519 (2010). https://doi.org/10.1063/1.34565003.
72. M.S. Lucas, L. Mauger, J.A. Muñoz, I. Halevy, J. Horwath, S.L. Semiatin, S.O. Leontsev, M.B. Stone, D.L. Abernathy, Y. Xiao et al., *J. Appl. Phys.* 113, 17A308 (2013). https://doi.org/10.1063/1.4794354.
73. A. Zunger, S.-H. Wei, L.G. Ferreira, and J.E. Bernard, *Phys. Rev. Lett.* 65, 353 (1990). https://doi.org/10.1103/PhysRevLett.65.353.
74. S.-H. Wei, L.G. Ferreira, J.E. Bernard, and A. Zunger, *Phys. Rev. B* 42, 9622 (1990). https://doi.org/10.1103/PhysRevB.42.9622.
75. C. Jiang, C. Wolverton, J. Sofo, L.-Q. Chen, and Z.-K. Liu, *Phys. Rev. B* 69, 214202 (2004). https://doi.org/10.1103/PhysRevB.69.214202.
76. D. Shin, R. Arróyave, Z.-K. Liu, and A. Van de Walle, *Phys. Rev. B* 74, 024204 (2006). https://doi.org/10.1103/PhysRevB.74.024204.
77. J. von Pezold, A. Dick, M. Friák, and J. Neugebauer, *Phys. Rev. B* 81, 094203 (2010). https://doi.org/10.1103/PhysRevB.81.094203.
78. A. van de Walle, P. Tiwary, M. de Jong, D. Olmsted, M. Asta, A. Dick, D. Shin, Y. Wang, L.-Q. Chen, and Z.-K. Liu, *Calphad* 42, 13 (2013).
79. K.C. Hass, L.C. Davis, and A. Zunger, *Phys. Rev. B* 42, 3757 (1990). https://doi.org/10.1103/PhysRevB.42.3757.
80. C. Jiang, C.R. Stanek, K.E. Sickafus,and B.P. Uberuaga, *Phys. Rev. B* 79, 104203 (2009). https://doi.org/10.1103/PhysRevB.79.104203.
81. Z. Jiang, Y. Nahas, B. Xu, S. Prosandeev, D. Wang, and L. Bellaiche, *J. Phys. Condens. Matter* 28, 475901 (2016).
82. M.C. Gao, C. Niu, C. Jiang, and D.L. Irving, in *High-Entropy Alloys: Fundamentals and Applications*. ed. by M.C. Gao, J.-W. Yeh, P.K. Liaw, Y. Zhang (Springer, Cham, 2016), pp. 333–368. https://doi.org/10.1007/978-3-319-27013-5_10.
83. B. Alling, T. Marten, and I.A. Abrikosov, *Phys. Rev. B* 82, 184430 (2010). https://doi.org/10.1103/PhysRevB.82.184430.
84. F. Körmann, P.-W. Ma, S.L. Dudarev, and J. Neugebauer, *J. Phys. Condens. Matter* 28, 076002 (2016).
85. J.A. Munoz, M.S. Lucas, O. Delaire, M.L. Winterrose, L. Mauger, C.W. Li, A.O. Sheets, M.B. Stone, D.L. Abernathy, Y. Xiao et al., *Phys. Rev. Lett.* 107, 115501 (2011). https://doi.org/10.1103/PhysRevLett.107.115501.
86. A. Van De Walle and M. Asta, *Metall. Mater. Trans. A* 33, 735 (2002). https://doi.org/10.1007/s11661-002-1002-8.
87. R. Sun, C. Woodward, and A. van de Walle, *Phys. Rev. B* 95, 214121 (2017). https://doi.org/10.1103/PhysRevB.95.214121.
88. R. Sun and A. van de Walle, *Calphad* 53, 20 (2016).
89. M. Dodaran, A.H. Ettefagh, S. Guo, M. Khonsari, W. Meng, N. Shamsaei, and S. Shao, *Intermetallics* 117, 106670 (2020).
90. A.R. Natarajan and A. Van der Ven, *npj Comput. Mater.* 6, 80 (2020). https://doi.org/10.1038/s41524-020-0348-z.
91. S. Muller, *J. Phys. Condens. Matter* 15, R1429 (2003). https://doi.org/10.1088/0953-8984/15/34/201.
92. B.C. Han, A. Van der Ven, G. Ceder, and B.-J. Hwang, *Phys. Rev. B* 72, 205409 (2005). https://doi.org/10.1103/PhysRevB.72.205409.
93. D. Lerch, O. Wieckhorst, L. Hammer, K. Heinz, and S. Müller, *Phys. Rev. B* 78, 121405 (2008). https://doi.org/10.1103/PhysRevB.78.121405.
94. C. Lazo, and F.J. Keil, *Phys. Rev. B* 79, 245418 (2009). https://doi.org/10.1103/PhysRevB.79.245418.
95. W. Chen, D. Schmidt, W.F. Schneider, and C. Wolverton, *Phys. Rev. B* 83, 075415 (2011). https://doi.org/10.1103/PhysRevB.83.075415.
96. C. Wolverton,*Philos. Mag. Lett.* 79, 683 (1999). https://doi.org/10.1080/095008399176724.
97. C. Wolverton, *Modell. Simul. Mater. Sci. Eng.* 8, 323 (2000). https://doi.org/10.1088/0965-0393/8/3/312.
98. S. Muller, C. Wolverton, L.-W. Wang, and A. Zunger, *Acta Mater.* 48, 4007 (2000).
99. S. Muller, C. Wolverton, L.-W. Wang,and A. Zunger, *Europhys. Lett. (EPL)* 55, 33 (2001). https://doi.org/10.1209/epl/i2001-00377-0.
100. N. Zarkevich, D. Johnson, and A. Smirnov, *Acta Mater.* 50, 2443 (2002).
101. S. Müller, L.-W. Wang, and A. Zunger, *Modell. Simul. Mater. Sci. Eng.* 10, 131 (2002). https://doi.org/10.1088/0965-0393/10/2/303.
102. V. Vaithyanathan, C. Wolverton, and L.Q. Chen, *Phys. Rev. Lett.* 88, 125503 (2002). https://doi.org/10.1103/PhysRevLett.88.125503.
103. S. Muller, *Surface Interface Anal.* 38, 1158 (2006).
104. M. Y. Lavrentiev, D. Nguyen Manh, and S. L. Dudarev, in *Solid-Solid Phase Transformations in Inorganic Materials* (Trans Tech Publications Ltd, 2011), vol. 172 of *Solid State Phenomena, pp. 1002–1007.*.
105. D. Kleiven and J. Akola, *Acta Mater.* 195, 123 (2020).
106. A. Redermeier and E. Kozeschnik, *Model. Simul. Mater. Sci. Eng.* 29, 035014 (2021). https://doi.org/10.1088/1361-651x/abe5b2.
107. A. van de Walle, *Calphad* 33, 266 (2009).
108. A. van de Walle, P. Tiwary, M.M. de Jong, D.L. Olmsted, M.D. Asta, A. Dick, D. Shin, Y. Wang, L.-Q. Chen, and Z.-K. Liu, *Calphad* 42, 13 (2013).
109. A. Van der Ven, J. Thomas, Q. Xu, J. Bhattacharya, Math. Comput. Simul. 80, 1393 (2010).
110. J.C. Thomas and A.V.D. Ven, *Phys. Rev. B* 88, 214111 (2013). https://doi.org/10.1103/PhysRevB.88.214111.
111. S. Kadkhodaei and A. van de Walle, *Comput. Phys. Commun.* 246, 106712 (2020).
112. M. Ångqvist, W.A. Muñoz, J.M. Rahm, E. Fransson, C. Durniak, P. Rozyczko, T.H. Rod, and P. Erhart, *Adv. Theory Simul.* 2, 1900015 (2019). https://doi.org/10.1002/adts.201900015.
113. J.H. Chang, D. Kleiven, M. Melander, J. Akola, J.M. Garcia-Lastra, and T. Vegge, *J. Phys. Condens. Matter* 31, 325901 (2019).
114. L. Kaufman and H. Bernstein (1970).
115. U.R. Kattner, *Tecnol Metal Mater Min* 13, 3 (2016).
116. B. Bocklund, R. Otis, A. Egorov, A. Obaied, I. Roslyakova, Z.-K. Liu, MRS Commun. 9, 618 (2010).
117. Z.-K. Liu, *J. Phase Equilib. Diffus.* 30, 517 (2009).
118. A. van de Walle, R. Sun, Q.-J. Hong, and S. Kadkhodaei, *Calphad* 58, 70 (2017).
119. D.B. Laks, L.G. Ferreira, S. Froyen, and A. Zunger, *Phys. Rev. B* 46, 12587 (1992b). https://doi.org/10.1103/PhysRevB.46.12587.
120. V. Blum and A. Zunger, *Phys. Rev. B* 70, 155108 (2004). https://doi.org/10.1103/PhysRevB.70.155108.

121. H.Y. Geng, M.H.F. Sluiter, and N.X. Chen, *Phys. Rev. B* 73, 012202 (2006). https://doi.org/10.1103/PhysRevB.73.012202.
122. S. Kadkhodaei, Q.-J. Hong, and A. van de Walle, *Phys. Rev. B* 95, 064101 (2017), https://doi.org/10.1103/PhysRevB.95.064101.
123. A. Van der Ven and G. Ceder, *Phys. Rev. B* 71, 054102 (2005). https://doi.org/10.1103/PhysRevB.71.054102.
124. A. Van der Ven, G. Ceder, M. Asta, and P.D. Tepesch, *Phys. Rev. B* 64, 184307 (2001). https://doi.org/10.1103/PhysRevB.64.184307.
125. X. Zhang and M.H.F. Sluiter, *J. Phase Equilibria Diffusion* 37, 44 (2016). https://doi.org/10.1007/s11669-015-0427-x.
126. A. van de Walle, *Nat. Mater.* 7, 455 (2008).
127. Y. Ikeda, A. Carreras, A. Seko, A. Togo, and I. Tanaka, *Phys. Rev. B* 95, 024305 (2017). https://doi.org/10.1103/PhysRevB.95.024305.
128. Y.-H. Tang and W.A. de Jong, *J. Chem. Phys.* 150, 044107 (2019). https://doi.org/10.1063/1.5078640.
129. M.Y. Lavrentiev, D. Nguyen-Manh, and S.L. Dudarev, *Phys. Rev. B* 81, 184202 (2010). https://doi.org/10.1103/PhysRevB.81.184202.
130. F. Körmann, A. A. H. Breidi, S. L. Dudarev, N. Dupin, G. Ghosh, T. Hickel, P. Korzhavyi, J. A. Munoz, and I. Ohnuma, *Phys. Stat. Sol. (b)* 251, 53 (2014). https://doi.org/10.1002/pssb.201350136.
131. C. Jiang and B.P. Uberuaga, *Phys. Rev. Lett.* 116, 105501 (2016). https://doi.org/10.1103/PhysRevLett.116.105501.
132. R. Kitchin and T.P. Lauriault, *GeoJournal* 80, 463 (2015).
133. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder et al., *APL Mater.* 1, 011002 (2013). https://doi.org/10.1063/1.4812323.
134. J.J. de Pablo, N.E. Jackson, M.A. Webb, L.-Q. Chen, J.E. Moore, D. Morgan, R. Jacobs, T. Pollock, D.G. Schlom, E.S. Toberer et al., *npj Comput. Mater.* 5, 41 (2019). https://doi.org/10.1038/s41524-019-0173-4.
135. S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, *Nat. Mater.* 12, 191 (2013).
136. A.R. Oganov, C.J. Pickard, Q. Zhu, and R.J. Needs, *Nat. Rev. Mater.* 4, 331 (2019).
137. G. Teichert, A. Natarajan, A. Van der Ven, and K. Garikipati, Comput. *Methods Appl. Mech. Eng.* 371, 113281 (2020), ISSN 0045-7825, http://www.sciencedirect.com/science/article/pii/S0045782520304667.
138. J. S. Yedidia, W. T. Freeman, and Y. Weiss, in *Proceedings of the 13th International Conference on Neural Information Processing Systems* (MIT Press, Cambridge, MA, USA, 2000), NIPS'00, pp. 668–674.
139. H.J. Kappen and W. Wiegerinck, in Advances, in *Neural Information Processing Systems*. ed. by T.G. Dietterich, S. Becker, Z. Ghahramani (MIT Press, Boston, MA, 2002), pp. 415–422.
140. J.S. Yedidia, W.T. Freeman, and Y. Weiss, *IEEE Trans. Inf. Theory* 51, 2282 (2005).
141. D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, Cambridge, 2012).
142. C. Cortes and V. Vapnik, *Mach. Learn.* 20, 273 (1995).
143. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, 2016).
144. M. Raissi, P. Perdikaris, and G. Karniadakis, *J. Comput. Phys.* 378, 686 (2019).