Relative entropy as model selection tool in cluster expansions

Jesper Kristensen,[1,2] Ilias Bilionis,[1] and Nicholas Zabaras[1,*]
[1]*Materials Process Design and Control Laboratory, Sibley School of Mechanical and Aerospace Engineering, 101 Frank H. T. Rhodes Hall, Cornell University, Ithaca, New York 14853-3801, USA*
[2]*School of Applied and Engineering Physics, 271 Clark Hall, Cornell University, Ithaca, New York 14853-3501, USA*
(Received 16 March 2013; revised manuscript received 18 April 2013; published 28 May 2013)

Cluster expansions are simplified, Ising-like models for binary alloys in which vibrational and electronic degrees of freedom are coarse grained. The usual practice is to learn the parameters of the cluster expansion by fitting the energy they predict to a finite set of *ab initio* calculations. In some cases, experiments suggest that such approaches may lead to overestimation of the phase transition temperature. In this work, we present a novel approach to fitting the parameters based on the relative entropy framework which, instead of energies, attempts to fit the Boltzmann distribution of the configurational degrees of freedom. We show how this leads to T-dependent parameters.

DOI: 10.1103/PhysRevB.87.174112 PACS number(s): 71.20.−b, 81.30.Bx

I. INTRODUCTION

Construction of binary alloy phase diagrams generally relies on a computationally tractable parametrized surrogate model for the quantum mechanical energy surface. The *cluster expansion* is a commonly used model which has been very successful in describing configurational properties of alloys.[1] The model parameters are referred to as *effective cluster interactions* (ECI) and for a typical system, they range in number from 20 to 80. The ECI are often fitted to 50–100 observed energies, e.g., using least squares with cross validation,[2] potentially coupled with genetic algorithms,[3] or compressive sensing.[4] These observations are made at a high computational cost, involving *ab initio* software, e.g., the Vienna *ab initio* simulation package (VASP).[5,6] Given the ECI, the energy of any configuration can be computed and subsequently used for thermodynamic simulations.[7–13]

Let us provide an intuitive argument as to why fitting the ECI to equally weighted observed energies does not necessarily lead to an optimal description of the thermodynamics of the system. Consider the case of a canonical ensemble. The Boltzmann factor, in equilibrium, dictates that more energetic states are exponentially less likely to be observed. So, at low temperatures, the partition function is mostly influenced by the few low energy states. The Boltzmann factor for all other states is essentially zero. Therefore, the *thermodynamic importance* of a state is quantified by its Boltzmann factor. We conclude that as the temperature is increased, so is the importance of any state. Furthermore, at infinite temperature all states are equally important.

Motivated by the previous discussion it is desirable to investigate techniques which obtain the ECI based on thermodynamic arguments. All the necessary information is encapsulated in the probability distribution over states (PDS). The idea is to bring the PDS induced by the cluster expansion (candidate PDS) as close as possible to the true one. We propose to measure this distance in terms of *relative entropy* (also known as the *Kullback-Leibler divergence*).[14] Thus, we obtain a *variational* problem, namely, the minimization of the relative entropy functional with respect to the candidate PDS. Though theoretically sound this problem is computationally intractable. To cope with this, we show that the relative entropy functional can be approximated by the variance (with respect to the true PDS) of the difference between the true and the candidate cluster expansion energy. Restricting this approximation on the observed data leads to a weighted least-squares problem making the proposed approach computationally attractive.

We test the performance of our method in a study of canonical phase transformations in Si-Ge (two-phase coexistence to disorder) and Mg-Li (order to disorder) alloys at various compositions. We compare the relative entropy results to least squares with leave-one-out-cross-validation (least-squares LOOCV) where we, for Mg-Li, observe noticeable differences in the transition temperatures. Our results are found to be in better agreement with guiding experimental data.

The paper is organized as follows. The theoretical framework is discussed in Sec. II, where we first introduce the cluster expansion (Sec. II A), followed by the relative entropy framework (Sec. II B). In Sec. III, we explain how the *ab initio* observations were made along with details on the thermodynamic calculations (Sec. III A). Then we present numerical results starting with a toy model in Sec. III B. This is followed by a study of a Si-Ge diamond alloy (Sec. III C) predicting the transition from two-phase coexistence to disorder at fixed 50% composition. The same transition has been simulated in Refs. 15–17 and by using cluster expansions with least-squares LOOCV in Refs. 7 and 18. As a final example we then turn to a Mg-Li alloy on a body-centered cubic (bcc) lattice (Sec. III D) which has recently[19] been simulated at low temperature, predicting order-to-disorder phase transformations at 33%, 50%, and 66% Mg. Finally we compare the relative entropy results with both those in Ref. 19 and the guiding experimental data in Ref. 20.

II. THEORY

A. Cluster expansions

Consider a binary alloy of a fixed Bravais lattice (e.g., face-centered cubic, bcc, etc.) augmented with a crystal basis collectively defining N crystal sites. Any such site can be occupied by one of two distinct atoms each assigned a unique integer value among ± 1. Specifying which atoms sit where on each crystal site defines a configuration $\boldsymbol{\sigma}$, represented

as an N-dimensional vector. The collection of all possible configurations $\{\boldsymbol{\sigma}\}$ forms the configuration space of size 2^N.

It is possible to construct an orthonormal and complete basis on the space of functions of $\boldsymbol{\sigma}$.[21] Physically, each basis function is associated with a set of crystal sites. Each possible group of sites is called a *cluster*. An expansion coefficient is associated with each cluster. Symmetry considerations can greatly reduce the number of such independent coefficients.[22] This symmetry reduced set of coefficients is what we termed ECI in Sec. I.

In this study we ignore nonconfigurational degrees of freedom such as phonons and electrons. Under this assumption a state of the alloy is uniquely defined by $\boldsymbol{\sigma}$. The cluster expansion approximates its energy as

$$E(\boldsymbol{\sigma}|\boldsymbol{\gamma}) = \sum_{i=1}^{M} \gamma_i \phi_i(\boldsymbol{\sigma}), \tag{1}$$

where the sum is over all M basis functions $\phi_i(\cdot)$ chosen for this system, and $\boldsymbol{\gamma} = \{\gamma_i\}$ are the ECI to be learned. This is analogous to an extended Ising model, as justified in Ref. 23. The expansion is *exact* when including all possible basis functions.[21] In practice, a truncation to M terms is necessary. This is achieved by fixing the maximum number of crystal sites present in any cluster as well as its maximum spatial extent (largest distance between any two crystal sites in the cluster).[7]

Nonconfigurational contributions are readily accounted for if Eq. (1) is interpreted as the free energy of the system.

B. Relative entropy

Let the binary alloy, as introduced in Sec. II A, be at a fixed composition and in a heat bath at temperature T. The PDS is

$$p(\boldsymbol{\sigma}|\beta) = \frac{\exp[-\beta E(\boldsymbol{\sigma})]}{Z}, \tag{2}$$

where $\beta = 1/k_B T$, with k_B the Boltzman factor, $E(\boldsymbol{\sigma})$ is the *ab initio* energy of state $\boldsymbol{\sigma}$, and $Z = \sum_{\sigma} \exp[-\beta E(\boldsymbol{\sigma})]$ is the partition function. The candidate PDS, given some ECI $\boldsymbol{\gamma}$, is

$$p(\boldsymbol{\sigma}|\boldsymbol{\gamma},\beta) = \frac{\exp[-\beta E(\boldsymbol{\sigma}|\boldsymbol{\gamma})]}{Z(\boldsymbol{\gamma})}. \tag{3}$$

As discussed in Sec. I we seek to bring the candidate PDS [Eq. (3)] as close as possible to the true PDS [Eq. (2)]. The measure of "distance" we choose is the relative entropy, defined by

$$\mathcal{S}_{\mathrm{rel}}[\boldsymbol{\gamma}] := \sum_{\sigma} p(\boldsymbol{\sigma}|\beta) \ln\left[\frac{p(\boldsymbol{\sigma}|\beta)}{p(\boldsymbol{\sigma}|\boldsymbol{\gamma},\beta)}\right]. \tag{4}$$

The relative entropy quantifies the information loss when using the candidate PDS instead of the true one. It has been used in numerous applications ranging from machine learning (variational methods)[24] and coarse graining,[25–28] to free energy calculations.[29] We postulate that the optimal ECI should be selected by minimizing Eq. (4).

Notice that Eq. (4) involves a sum over the entire phase space, rendering its minimization practically impossible. Several approximations are essential in order to bring the problem to a soluble form. Towards this goal, let us first rewrite Eq. (4) as

$$\begin{aligned}\mathcal{S}_{\mathrm{rel}}[\boldsymbol{\gamma}] &= \beta\langle E(\boldsymbol{\sigma}|\boldsymbol{\gamma}) - E(\boldsymbol{\sigma})\rangle + \ln[Z(\boldsymbol{\gamma})] + \ln(Z)\\ &= -\langle\Delta\rangle + \ln\langle\exp(\Delta)\rangle\\ &= \ln\langle\exp[\Delta - \langle\Delta\rangle]\rangle,\end{aligned}$$

where $\langle\cdot\rangle$ denotes the expectation with respect to the true PDS and $\Delta := \beta[E(\boldsymbol{\sigma}) - E(\boldsymbol{\sigma}|\boldsymbol{\gamma})]$ can be thought of as the error of the approximation, in units of the thermal energy. We now derive an approximation to the relative entropy whose minimization is equivalent to the solution of a linear system. To leading order in Δ we have

$$\begin{aligned}\mathcal{S}_{\mathrm{rel}}[\boldsymbol{\gamma}] &= \ln\left\langle 1 + (\Delta - \langle\Delta\rangle) + \frac{1}{2}(\Delta - \langle\Delta\rangle)^2 + \cdots\right\rangle\\ &\approx \ln\left(1 + \frac{1}{2}\mathrm{Var}[\Delta]\right) \approx \frac{1}{2}\mathrm{Var}[\Delta]\\ &= \frac{1}{2}\left[\sum_{\sigma} p(\boldsymbol{\sigma}|\beta)\Delta^2 - \left(\sum_{\sigma} p(\boldsymbol{\sigma}|\beta)\Delta\right)^2\right],\end{aligned} \tag{5}$$

where, in the last equality, we recall that Δ depends on $\boldsymbol{\sigma}$. Equating the derivative of Eq. (5) with respect to $\boldsymbol{\gamma}$ to zero,

$$\begin{aligned}0 = \frac{\partial\mathcal{S}_{\mathrm{rel}}[\boldsymbol{\gamma}]}{\partial\gamma_i} &= -\beta[\langle\phi_i(\boldsymbol{\sigma})\Delta\rangle - \langle\Delta\rangle\langle\phi_i(\boldsymbol{\sigma})\rangle]\\ &= -\beta^2\mathrm{cov}\{[E(\boldsymbol{\sigma}) - \hat{E}(\boldsymbol{\sigma}|\boldsymbol{\gamma})],\phi_i(\boldsymbol{\sigma})\},\end{aligned}$$

and solving for $\boldsymbol{\gamma}$, yields the following linear system:

$$\mathrm{cov}[\boldsymbol{\phi}(\boldsymbol{\sigma}),\boldsymbol{\phi}(\boldsymbol{\sigma})]\boldsymbol{\gamma} = \mathrm{cov}[\boldsymbol{\phi}(\boldsymbol{\sigma}),E(\boldsymbol{\sigma})], \tag{6}$$

where

$$\mathrm{cov}[\mathbf{X},\mathbf{Y}] := \langle(\mathbf{X} - \langle\mathbf{X}\rangle)(\mathbf{Y} - \langle\mathbf{Y}\rangle)^t\rangle,$$

with $\mathbf{X}^t$ being the transpose of $\mathbf{X}$, denotes the covariance of the true PDS between two vector functions $\mathbf{X}$ and $\mathbf{Y}$ of configuration space. However, Eq. (6) is still unmanageable. The next step is to approximate it using the observed data.

Assume we have collected N observations $\{[\boldsymbol{\sigma}^{(i)},E(\boldsymbol{\sigma}^{(i)})]\}_{i=1}^{N}$, where $\boldsymbol{\sigma}^{(i)}$ is the configuration and $E(\boldsymbol{\sigma}^{(i)})$ is the corresponding *ab initio* energy. To approximate Eq. (6), we replace all averages with respect to the true PDS with averages over the observed data. In other words, we approximate the average of any quantity $\mathbf{X}$ by

$$\langle\mathbf{X}\rangle \approx \sum_{i=1}^{N} p_{N,i}\,\mathbf{X}(\boldsymbol{\sigma}^{(i)}),$$

where

$$\mathbf{p}_N := \left(\frac{\exp[-\beta E(\boldsymbol{\sigma}^{(1)})]}{Z_N}, \ldots, \frac{\exp[-\beta E(\boldsymbol{\sigma}^{(N)})]}{Z_N}\right),$$

with $Z_N := \sum_{i=1}^{N} \exp[-\beta E(\boldsymbol{\sigma}^{(i)})]$. Thus, Eq. (6) becomes a weighted least-squares problem:

$$\boldsymbol{\Phi}^t\mathbf{W}\boldsymbol{\Phi}\boldsymbol{\gamma} = \boldsymbol{\Phi}^t\mathbf{W}\mathbf{E}, \tag{7}$$

where $\mathbf{E}$ is a vector of length N containing the observed energies, and

$$\mathbf{W} := (\mathbf{I}_N - \mathbf{p}_N\mathbf{1}_N)\mathrm{diag}(\mathbf{p}_N)(\mathbf{I}_N - \mathbf{p}_N\mathbf{1}_N)^t,$$

where $\mathbf{I}_N$ is the $(N \times N)$ identity matrix, $\mathbf{1}_N$ a $(1 \times N)$ vector of all elements equal to 1, and diag($\mathbf{p}_N$) an $(N \times N)$ matrix with the vector $\mathbf{p}_N$ as its diagonal.

The least-squares problem [Eq. (7)] is what makes the relative entropy method computationally tractable. The extent to which it is valid requires further mathematical analysis that goes beyond the scope of this work. For the cases considered in Sec. III it was observed that the ECI obtained by solving Eq. (7) converged as the number of observations N increased in a way similar to least squares. Notice that the ECI are, generally, T dependent through the weight matrix $\mathbf{W}$ in Eq. (7).

Before concluding this section, it is instructive to compare Eq. (5) with the least-squares loss function

$$\mathcal{L}[\boldsymbol{\gamma}] = \sum_{\sigma}[E(\boldsymbol{\sigma}) - E(\boldsymbol{\sigma}|\boldsymbol{\gamma})]^2 = \beta^{-2}\sum_{\sigma}\Delta^2. \qquad (8)$$

Notice that the first summand in Eq. (5) is similar to that of Eq. (8), albeit weighted by the true PDS. The second summand does not have an analog in least squares. The important observation is that learning efforts are concentrated on the relevant states mostly, at a given temperature. On the contrary, least squares always considers all states equally important. It is therefore anticipated that the relative entropy ECI would perform better when used in thermodynamic calculations.

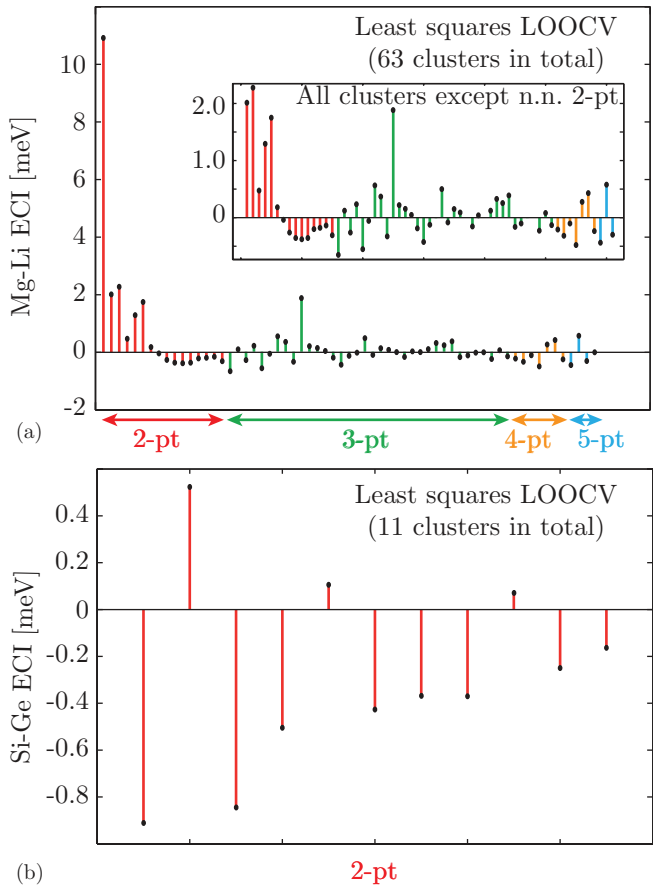

FIG. 1. (Color online) ECI obtained for (a) the Mg-Li and (b) the Si-Ge system using least-squares LOOCV as implemented in ATAT. Left-right arrows, with colors and text, group the number of sites in the clusters associated with the ECI. "i-pt" is shorthand for i-point clusters. Within each group, spatially larger clusters are found farther to the right. The total number of clusters reported excludes the empty and 1-pt cluster. The inset in (a) shows the ECI for all clusters except the nearest-neighbor 2-pt cluster (n.n. 2-pt) for a more detailed variation. Notice the difference in scale in energies between the two systems.

III. NUMERICAL RESULTS

A. General setup

In this section, we explain how the *ab initio* data for the two alloys were obtained. Then, we briefly discuss the details of the underlying Monte Carlo scheme used to extract the thermodynamic information needed to detect the phase transitions.

The first-principles configurational energies were generated using VASP.[5,6] For the Mg-Li body-centered cubic system we used Perdew-Burke-Ernzerhof (PBE)[30,31] projector-augmented-wave[32] pseudopotentials. We treated the $2p$ and $3s$ orbitals of Mg and the $1s$ orbital of Li as valences. Convergence to within 1 meV/atom was achieved around a wave cutoff of 340 eV, in agreement with Ref. 19 using the same technique, and was checked for both the pure structures and for a couple of randomly chosen mixed structures. Spin polarization was not included. We used a dense gamma-centered Monkhorst-Pack[33] k grid for the Brillouin zone integration with just over 7500 k points per reciprocal atom and scaled the grid according to the size of each supercell being computed. In this way, a training set of 82 structures was generated. Then, a cluster expansion with up to and including five-point clusters was constructed using ATAT. The least-squares LOOCV ECI were obtained using ATAT. We obtained a LOOCV score of 4.2 meV. The same training set was used to compute the ECI for the relative entropy method at each temperature using Eq. (7).

For the Si-Ge diamond system, we also used PBE pseudopotentials and the system converged to within 1 meV/atom around a wave cutoff of 340 eV using the same k-point grid type, density, and scaling as for the Mg-Li system. A training set of 38 structures was constructed and a cluster expansion

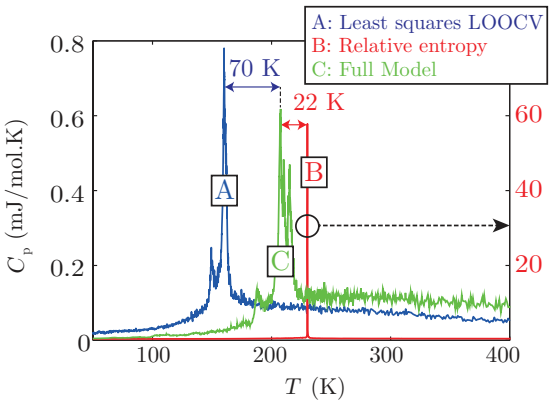

FIG. 2. (Color online) Canonical ordered phase transition from $C11_b$ to disorder in a toy model based on the Mg-Li system at 66% Mg. The left-right arrows with associated temperatures identify the peak-to-peak distance between predicted transitions of the two considered methods and the full model, see discussion in Sec. III B. Note that the relative entropy result B is measured on the right scale as indicated by the dashed arrow.

containing up to and including three-point clusters, as in Ref. 7, computed with ATAT. The least-squares LOOCV ECI were obtained with ATAT. A LOOCV score of <1 meV was achieved. The training data were used to compute the ECI for the relative entropy method at each temperature using Eq. (7).

The clusters obtained with ATAT are used in the relative entropy method as basis functions, i.e., as columns in $\mathbf{\Phi}$ in Eq. (7). The ECI obtained for both systems using the least-squares LOOCV method are shown in Fig. 1. The empty cluster is left out as it is just an overall scale of the energy which is irrelevant when sampling the Boltzmann factor. By a similar argument, which is valid because the system is in a canonical ensemble, the one-point cluster ECI does not play a role either.

The ground states found for Mg-Li, at the compositions studied in this work, were B_f, B2, and $C11_b$ at 33%, 50%, and 66% Mg, respectively. Reference 19 finds the same result except for obtaining a $C11_b$ structure at 33% Mg. We find that only the two pure diamond structures are ground states for the Si-Ge system at zero kelvin.

Thermodynamic properties were calculated on a $30 \times 30 \times 30$ simulation cell with periodic boundary conditions. Sampling of the PDS was performed using an adaptive sequential Monte Carlo (ASMC) technique[34,35] coupled with an underlying Metropolis-Hastings algorithm using double-spin-flip dynamics to conserve composition. ASMC approximates the PDS with a weighted finite set of δ functions, called *particles*, at some convenient initial temperature. The particles are then propagated from this starting point to any other temperature as desired, and the weights are updated accordingly. Steps in temperature are taken adaptively to ensure that the distribution does not undergo large changes, and thermodynamic information is recorded at each step.

The ASMC sampler was initialized with 512 particles, each on its own computational core, at 2000 K. Each particle was initialized randomly. At this high temperature, random configurations are distributed (almost) correctly according to the Boltzmann factor offering a very convenient starting point. Nevertheless, each particle was thermalized with 100 sweeps.[36] A target temperature of 50 K was specified and to improve the chances that we did not miss a transition, we split the interval 50−2000 in bins of size 10 K. ASMC drove the ensemble of particles from the right corner of this interval to the left.

The heat capacity at constant pressure was obtained from the standard statistical relation to the variance in energy

$$C_{\mathrm{p}} = \beta^2 k_{\mathrm{B}}(\langle E^2\rangle - \langle E\rangle^2).$$

For an infinite system a divergence in this quantity signals a phase transition. For a finite system we should expect a distinct peak.

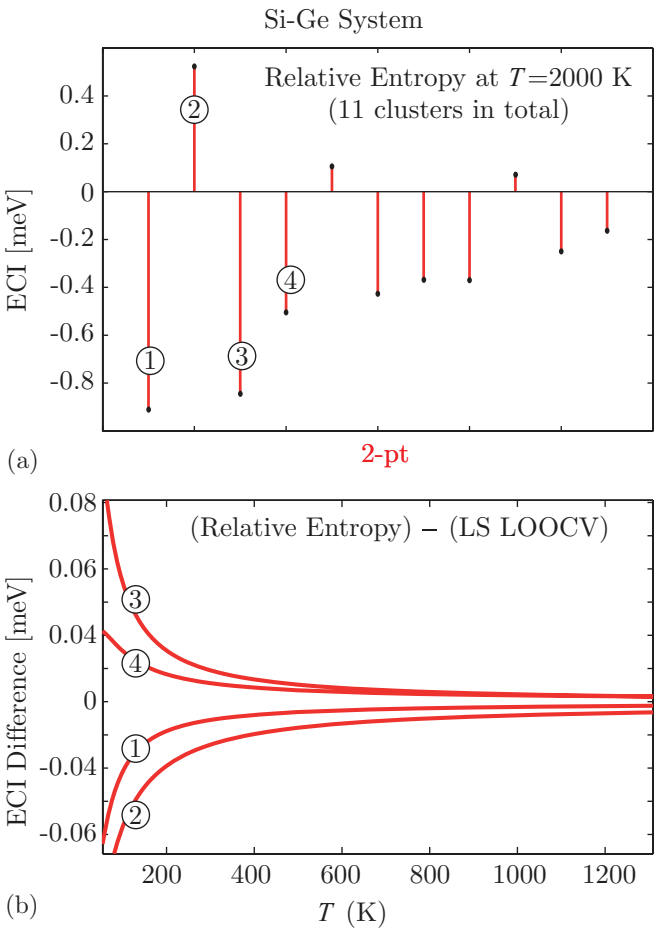

FIG. 3. (Color online) Si-Ge system. (a) ECI obtained with the relative entropy method at 2000 K. Compare to Fig. 1(b). (b) Difference in ECI vs T between least squares with LOOCV (LS LOOCV) and relative entropy. Circled numbers in (b) identify the clusters in (a) selected for plotting.

B. Magnesium lithium toy model

The purpose of this section is to assess the predictive power of the relative entropy method compared to least-squares LOOCV in the context of canonical order-to-disorder phase

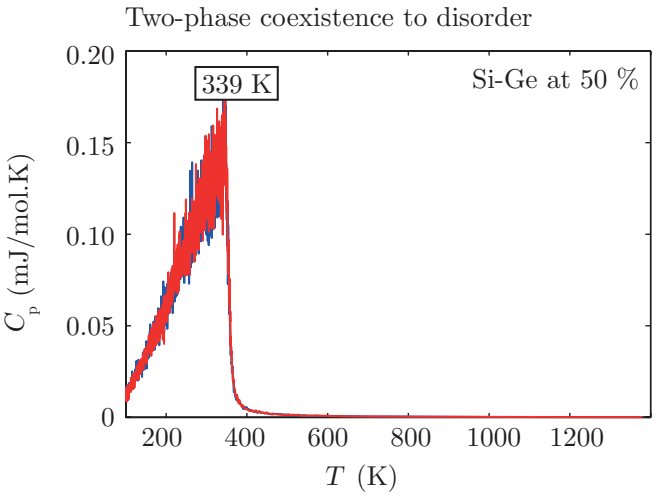

FIG. 4. (Color online) Prediction of two-phase coexistence to disorder phase transition as indicated by a "divergence" in constant pressure heat capacity C_{p} for Si-Ge on a diamond lattice at 50% composition. The least-squares LOOCV method is compared to relative entropy. Both methods give (almost) identical curves for the heat capacity. The rectangular box near the peak reports the temperature at which the maximum value of the heat capacity was obtained, which, to within 1 K, was the same for both methods.

transitions. To do this we devise an artificial problem in which the phase transition is exactly known. We consider the Mg-Li system and assume that the true energy surface is the one obtained by the set of clusters and ECI determined by ATAT, see Fig. 1(a). For this cluster expansion, we compute the phase transition temperature from order to disorder (true transition temperature). We create an artificial data set by evaluating this cluster expansion on the 82 observed configurations and assume that the energies predicted are the true energies. Then, we reduce the cluster expansion to include only the two-point clusters (of which there are 16) and train it on the artificial data using both least-squares LOOCV and relative entropy. Finally, for each method we compute the order-to-disorder phase transition temperature for the 66% Mg composition and compare them with the true one. It must be noted that special attention was paid to ensure that the reduced cluster expansion could predict the same ordered state, which was $C11_b$, as the artificial model. This was verified by monitoring the pair correlation values versus temperature. The results are shown in Fig. 2. The relative entropy method differs by 22 K of the true transition temperature compared to 70 K for least-squares LOOCV. This is evidence of the anticipated, enhanced, predictive capabilities of the relative entropy method.

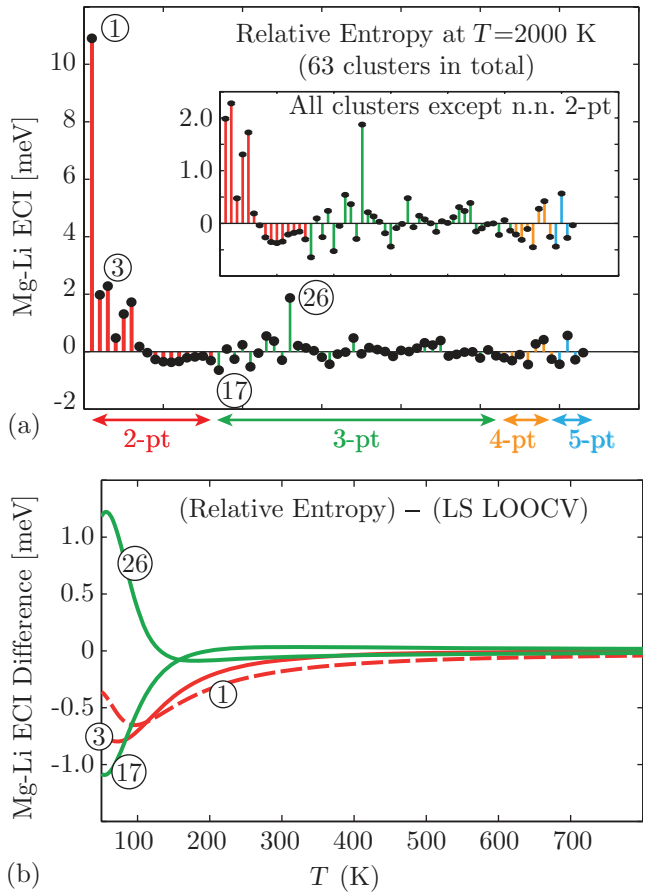

FIG. 5. (Color online) Mg-Li system. (a) ECI obtained with the relative entropy method at 2000 K. Compare with Fig. 1(a). (b) Difference in ECI vs T between least squares with LOOCV (LS LOOCV) and relative entropy. Circled numbers in (b) identify the clusters in (a) selected for plotting.

C. Transition from two-phase coexistence to disorder in diamond silicon germanium

The ECI obtained with relative entropy at the highest temperature 2000 K are shown in Fig. 3(a). At high temperatures relative entropy considers all states thermodynamically important. Therefore, the ECI should be comparable to least

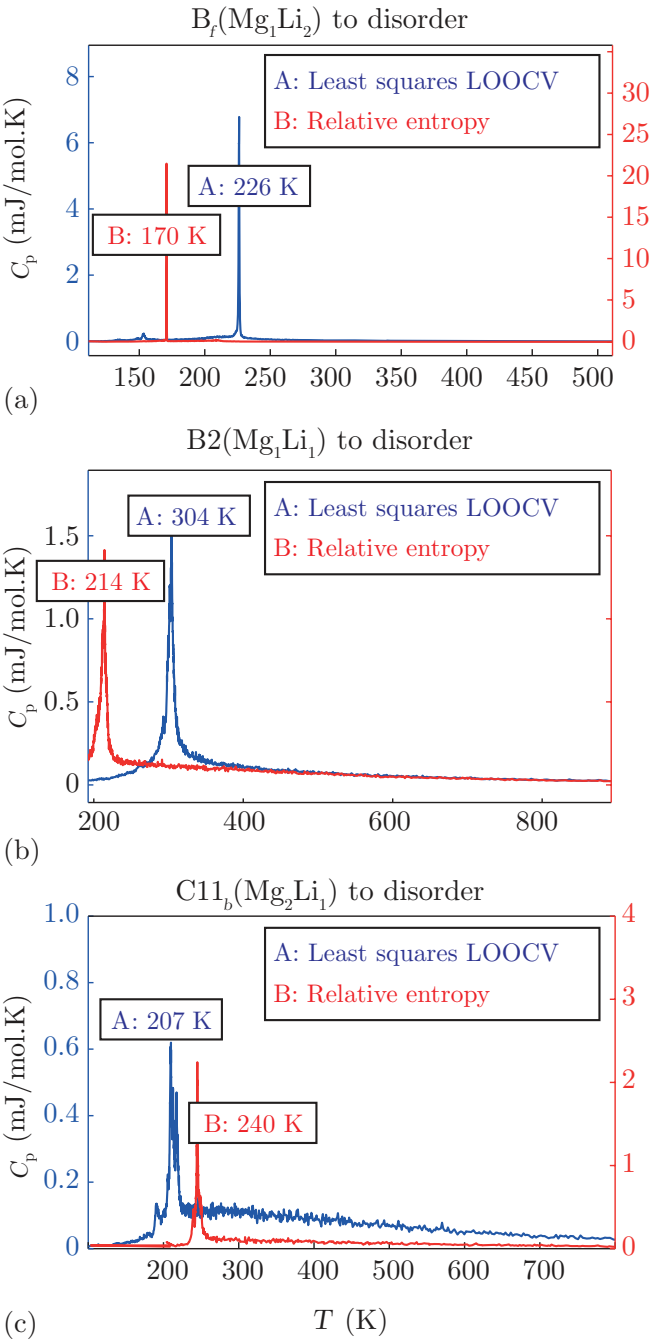

FIG. 6. (Color online) Prediction of the order-to-disorder phase transitions for Mg-Li bcc as indicated by "divergence" in the constant pressure heat capacity C_p at compositions (a) 33%, (b) 50%, and (c) 66% Mg. In each plot least-squares LOOCV (A: blue, associated with the left y axis) are compared to relative entropy (B: red, associated with the right y axis). The rectangular boxes on each peak report the temperature at which the maximum value of the heat capacity was obtained.

squares in Fig. 1(b), which is seen to be the case. The difference between the methods becomes increasingly pronounced as the temperature is lowered. This claim is supported by the difference in selected ECI versus T as shown in Fig. 3(b).

The prediction of the Si-Ge two-phase coexistence to disorder transition temperature at 50% composition is seen for ECI obtained using both least-squares LOOCV and relative entropy in Fig. 4. The least-squares result, yielding a transition temperature around 339 K, is in good agreement with the 325 K found in Ref. 7, using the same method but different sets of clusters and Monte Carlo sampling methods.

Relative entropy agrees with least-squares LOOCV. This can be explained from the small variation in the ECI between the two methods for this system on the order of 0.01 meV as seen in Fig. 3(b).

D. Order-to-disorder phase transition in bcc magnesium lithium

The ECI obtained with relative entropy at the highest temperature 2000 K are shown in Fig. 5(a). Similar to Sec. III C, at high temperatures the ECI should be comparable to least-squares LOOCV, shown in Fig. 1(a), which is found to be the case. The difference between selected ECI is shown in Fig. 5(b).

In Fig. 6 we show the predicted order-to-disorder transition temperatures for Mg-Li bcc at 33%, 50%, and 66% Mg compositions. Using least-squares LOOCV we obtain transition temperatures of 226, 304, and 207 K, respectively. This is in general agreement with Ref. 19 obtaining 190 and 210 K for 33% and 66% Mg, respectively. For 50% composition Ref. 19 reports a transition between 300 and 450 K. Overall, discrepancies are due to the different particular fitting methods used and different Monte Carlo techniques.

With the relative entropy method we predict 170, 214, and 240 K for 33%, 50%, and 66% Mg, respectively. The difference between the two methods can be explained from the changing ECI around the transition temperatures as shown in Fig. 5(b). The ECI differences are an order of magnitude larger than for the Si-Ge system.

Interestingly, the relative entropy method is in good agreement with the indicative experimental results in Ref. 20 according to which the order-to-disorder transitions across all compositions commence between 140 and 200 K. For the highest composition, i.e., 66% Mg, we note that the experiments had the highest error. For this largest composition, the relative entropy predicts a larger transition temperature than least-squares LOOCV.

IV. CONCLUSION

We have proposed a new paradigm based on a variational principle for obtaining thermodynamically relevant ECI. The principle is built upon the relative entropy which measures the information loss induced by replacing the true PDS with the candidate one. Through a series of suitable approximations we managed to bring the variational problem to a weighted least-squares form enabling a practical solution.

We observed differences in predicted order-to-disorder transition temperatures of the Mg-Li alloy between the proposed method and least-squares LOOCV. For the Si-Ge system, where a transition from two-phase coexistence to disorder was studied, we found the two methods to agree.

The main drawback of the method is that it requires a good set of clusters to start with. In this work, we relied on the methodologies implemented in ATAT (namely, least-squares LOOCV to capture the ground states). Alternatively, we could have used the state-of-the-art method of compressive sensing. An interesting research question is to investigate the performance of compressive sensing when the error function is replaced by relative entropy. Under the approximations developed in this paper this corresponds to changing from a Euclidean norm to a weighted norm when accounting for the error term.

ACKNOWLEDGMENTS

We appreciate the computing resources kindly provided by Pacific Northwest National Laboratory (PNNL). We further thank the PNNL Institutional Computing team for technical support on this project. Additional computing resources were provided by the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. This research was supported by the AFOSR Computational Mathematics Program, an OSD/AFOSR MURI09 award on uncertainty quantification, the US Department of Energy, Office of Science, Advanced Scientific Computing Research, and the National Science Foundation (DMS-1214282).

*Corresponding author: nzabaras@gmail.com; URL: http://mpdc.mae.cornell.edu/

[1] J. M. Sanchez, Phys. Rev. B **81**, 224202 (2010).
[2] A. van de Walle, M. Asta, and G. Ceder, Calphad **26**, 539 (2002).
[3] V. Blum, G. L. W. Hart, M. J. Walorski, and A. Zunger, Phys. Rev. B **72**, 165113 (2005).
[4] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozolins, Phys. Rev. B **87**, 035125 (2013).
[5] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[6] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[7] A. Walle and G. Ceder, J. Phase Equilib. **23**, 348 (2002).
[8] A. Kohan, P. Tepesch, G. Ceder, and C. Wolverton, Comput. Mater. Sci. **9**, 389 (1998).
[9] P. R. Alonso, P. H. Gargano, and G. H. Rubiolo, Calphad **38**, 117 (2012).
[10] R. Arróyave and M. Radovic, Phys. Rev. B **84**, 134112 (2011).
[11] C. Ravi, B. K. Panigrahi, M. C. Valsakumar, and A. van de Walle, Phys. Rev. B **85**, 054202 (2012).
[12] Z. W. Lu, D. B. Laks, S.-H. Wei, and A. Zunger, Phys. Rev. B **50**, 6642 (1994).
[13] A. S. Dalton, A. A. Belak, and A. Van der Ven, Chem. Mater. **24**, 1568 (2012).

[14]S. Kullback and R. A. Leibler, The Annals of Mathematical Statistics **22**, 79 (1951).
[15]P. C. Kelires and J. Tersoff, Phys. Rev. Lett. **63**, 1164 (1989).
[16]S. de Gironcoli, P. Giannozzi, and S. Baroni, Phys. Rev. Lett. **66**, 2116 (1991).
[17]M. Laradji, D. P. Landau, and B. Dünweg, Phys. Rev. B **51**, 4894 (1995).
[18]A. Qteish and R. Resta, Phys. Rev. B **37**, 6983 (1988).
[19]R. H. Taylor, S. Curtarolo, and G. L. W. Hart, Phys. Rev. B **81**, 024112 (2010).
[20]C. Barrett and O. Trautz, Trans. Metall. Soc. AIME **175**, 579 (1948).
[21]J. Sanchez, F. Ducastelle, and D. Gratias, Phys. Stat. Mech. Appl. **128**, 334 (1984).
[22]L. Ferreira, S.-H. Wei, and A. Zunger, Int. J. High Perform. Comput. Appl. **5**, 34 (1991).
[23]G. Ceder, Comput. Mater. Sci. **1**, 144 (1993).
[24]C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
[25]M. S. Shell, J. Chem. Phys. **129**, 144108 (2008).
[26]A. Chaimovich and M. S. Shell, Phys. Rev. E **81**, 060104 (2010).
[27]A. Chaimovich and M. S. Shell, J. Chem. Phys. **134**, 094112 (2011).
[28]M. S. Shell, J. Chem. Phys. **137**, 084503 (2012).
[29]I. Bilionis and N. Zabaras, J. Chem. Phys. **138**, 044313 (2013).
[30]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[31]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396 (1997).
[32]G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[33]H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[34]P. Del Moral, A. Doucet, and A. Jasra, J. Roy. Stat. Soc. B Stat. Meth. **68**, 411 (2006).
[35]I. Bilionis and P. Koutsourelakis, J. Comput. Phys. **231**, 3849 (2012).
[36]M. E. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University, New York, 1999).