Automating First-Principles Phase Diagram Calculations

A. van de Walle and G. Ceder

(Submitted 24 September 2001)

Devising a computational tool that assesses the thermodynamic stability of materials is among the most important steps required to build a "virtual laboratory," where materials could be designed from first principles without relying on experimental input. Although the formalism that allows the calculation of solid-state phase diagrams from first principles is well established, its practical implementation remains a tedious process. The development of a fully automated algorithm to perform such calculations serves two purposes. First, it will make this powerful tool available to a large number of researchers. Second, it frees the calculation process from arbitrary parameters, guaranteeing that the results obtained are truly derived from the underlying first-principles calculations. The proposed algorithm formalizes the most difficult step of phase diagram calculations, namely the determination of the "cluster expansion," which is a compact representation of the configurational dependence of the alloy's energy. This is traditionally achieved by a fit of the unknown interaction parameters of the cluster expansion to a set of structural energies calculated from first principles. We present a formal statistical basis for the selection of both the interaction parameters to include in the cluster expansion and the structures to use to determine them. The proposed method relies on the concepts of cross-validation and variance minimization. An application to the calculation of the phase diagram of the Si-Ge, CaO-MgO, Ti-Al, and Cu-Au systems is presented.

1. Introduction

Steadily growing computer power and improvements in numerical algorithms are making more and more materials problems approachable by computer simulations. First-principles computations, in which properties of materials are derived from quantum mechanics, are particularly interesting because they allow for the exploration of new materials even before a procedure to synthesize them has been devised [1990Ced, 1998Ven]. For any new material, a determination of its stable structure is of utmost importance. Hence, the determination of phase diagrams from first principles is among the most important steps required to build a "virtual laboratory."

Over the last 20 years, the formalism enabling the calculation of a solid-state phase diagram from quantum mechanical energy calculations has been carefully laid out [1994Fon, 1994Zun, 1991Duc]. This formalism establishes that the thermodynamic properties of an alloy can, in principle, be computed as accurately as desired through a technique known as the *cluster expansion*. In practice, however, the construction of this expansion can be tedious, and relies on the researcher's physical intuition to guide the construction process. These difficulties have so far limited the use of phase diagram calculations from first principles to members of the alloy theory community. To make this powerful tool available to a large number of people outside of the alloy theory community, we have developed a fully automated algorithm to perform such calculations and implemented it in an easy-to-use software package [2001Wal]. Our automated algorithm also offers the advantage of removing arbitrary parameters from the computations, ensuring that the results obtained are truly derived from the underlying first-principles calculations, rather than being based on the physical intuition of the user.

This paper is organized as follows: First, a brief description of the formalism to calculate phase diagrams from first principles is presented. We then propose and motivate an algorithm for the automatic determination of phase diagrams. Finally, various examples of the application of our method are shown.

2. The Cluster Expansion Formalism

The computation of an alloy phase diagram from first principles typically consists of three steps. First, the partition function of the system is *coarse grained* to that of a lattice model representing the possible configurational disorder of the alloy [1993Ced, 2002Wal]. If needed, this process can be accomplished without neglecting the additional entropy arising from the electronic and vibrational degrees of freedom, even though these degrees of freedom are no longer explicitly present in the lattice model. The resulting coarse-grained partition function consists of a sum over every possible way to place the atoms on a given *parent lattice*.[1] Because of the large number of terms in the sum, it is inconceivable to attempt to compute the energy of every configuration from first principles. Hence, in the second step, the dependence of energy on alloy configuration is

A. van de Walle, Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208; and **G. Ceder,** Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139. Contact e-mail: avdw@alum.mit.edu.

[1]Here the term "parent lattice" is used to distinguish it from the crystallographic lattice. The parent lattice may have more than one site per unit cell.

parametrized with a simpler model using one of alloy theory's most powerful tools: the cluster expansion formalism [1984San]. The cluster expansion provides a compact representation of the configurational dependence of the energy of an alloy, the accuracy of which can be systematically improved by adding a sufficient number of terms in the expansion. In the third step, the system is thermally equilibrated and free energies are obtained from Monte Carlo (MC) simulations. In this paper, we focus on the cluster expansion, because it is the most difficult step.

The cluster expansion is a generalization of the well-known Ising Hamiltonian. In the common case of a binary alloy system, the Ising model consists of assigning a spin-like occupation variable σ_i to each site i of the parent lattice, which takes the value -1 or $+1$ depending on the type of atom occupying the site. A particular arrangement of spins of the parent lattice is called a *configuration* and can be represented by a vector σ containing the value of the occupation variable for each site in the parent lattice. Although we focus here on the case of binary alloys, this framework can be extended to arbitrary multicomponent alloys (the appropriate formalism is presented in [1984San], whereas examples of applications can be found in [1994Ced, 1996McC, 1994Fon]).

The cluster expansion parametrizes the energy E (per atom) as a polynomial in the occupation variables:

$$E(\sigma) = \Sigma_\alpha m_\alpha J_\alpha \langle \Pi_{i\in\beta}\sigma_i \rangle \qquad \text{(Eq 1)}$$

where α is a cluster (a set of sites i). The sum is taken over all clusters α that are not equivalent by a symmetry operation of the space group of the parent lattice, whereas the average is taken over all clusters β that are equivalent to α by symmetry. The coefficients J_α in this expansion embody the information regarding the energetics of the alloy and are called the effective cluster interaction (ECI). The multiplicities m_α indicate the number of clusters that are equivalent by symmetry to α (divided by the number of lattice sites).

It can be shown that when *all* clusters α are considered in the sum, the cluster expansion is able to represent any function $E(\sigma)$ of configuration σ by an appropriate selection of the values of J_α. However, the real advantage of the cluster expansion is that, in practice, it is found to converge rapidly. An accuracy that is sufficient for phase diagram calculations can be achieved by keeping only clusters α that are relatively compact (e.g., short-range pairs or small triplets). The unknown parameters of the cluster expansion (the ECI) can then determined by fitting them to the energy of a relatively small number of configurations obtained, for instance, through first-principles computations. This approach is known as the structure inversion method (SIM) or the Connolly-Williams [1983Con] method.

The cluster expansion thus presents an extremely concise and practical way to model the configurational dependence of an alloy's energy. How many ECI and structures are needed in practice? A typical well-converged cluster expansion of the energy of an alloy consists of about 10-20 ECI and necessitates the calculation of the energy of around 30-50 ordered structures (see, for instance, [1998Ven, 1995Gar, 1998Ozo1]). Once the cluster expansion is known, various statistical mechanical techniques such as MC simulations [1988Bin], the low-temperature expansion (LTE) [1998Koh], or the cluster variation method (CVM) [1951Kik, 1991Duc] can be used to obtain phase diagrams.

3. Optimal Cluster Expansion Construction

Deciding which ECI are retained in the cluster expansion and which structural energies are used in the fit is largely a process of trial and error based on the experience of the researcher, making an automated procedure difficult. In response to this, we have devised an algorithm that constructs an optimal cluster expansion by alternatively answering the following two questions. Given the information obtained at one point in the calculations regarding the alloy system:

- Which cluster α should be included in the cluster expansion (nearest neighbor pair, second nearest neighbor pair, triplets, quadruplets, etc.)?
- Which atomic arrangements should be used in order to determine the unknown coefficients J_α?

Each question will be treated consecutively.

3.1 ECI Selection

3.1.1 Cross-Validation (CV). The determination of a cluster expansion differs from standard fitting procedures because the number of unknown parameters is theoretically infinite; the true physical system cannot be described exactly with a finite number of nonzero ECI. The problem is that we only have access to a finite number of structural energies, implying that we can never determine the exact cluster expansion. The question is then: What is the best we can do? We obviously need to truncate the series to a finite number of terms and our focus is to determine the optimal number of terms to keep, given that a certain number of structural energies are known. If we keep too few terms, the predicted energies may be imprecise because the truncated cluster expansion cannot account for all sources of energy fluctuations. If too many terms are kept, the more insidious problem of overfitting manifests itself. The mean squared error of the fit may appear very small, but the true predictive power of the cluster expansion for data that was not included in the fit is in fact much lower. The source of this problem is that energy variations caused by one ECI J_β not included in the fit may be incorrectly attributed to another ECI J_α that *is* included in the fit, simply because $\langle \Pi_{i\in\alpha}\sigma_i \rangle$ happens to be correlated with $\langle \Pi_{i\in\beta}\sigma_i \rangle$ in the finite set of structures, the energies of which are known. We thus need to find the choice of ECI that represents the best compromise between those two unwanted effects. Although it is not widely known, a formal solution to this problem exists (see for instance, [1974Sto]). It consists of evaluating the predictive power of a cluster expansion using the CV score, defined as

$$CV = n^{-1}\Sigma_{i=1}^{n}(E_i - \hat{E}_{(i)})^2$$

where E_i is the calculated energy of structure i, whereas $\hat{E}_{(i)}$ is the predicted value of the energy of structure i obtained from a least-squares fit to the $(n - 1)$ other structural energies. The choice of the number of terms that minimizes the CV score has been shown to be an asymptotically optimal [1987Li] selection rule. Although the formal proof of this result is rather technical, we provide a heuristic proof of its validity in the Appendix for the convenience of the reader. In contrast to the well-known mean squared error, the CV score is not monotonically decreasing. As the number of parameters to be fitted increases, the CV score first decreases because an increasing number of degrees of freedom are available to explain the variations in energy. The CV score then goes through a minimum before increasing, because of a decrease in predictive power caused by an increase in the noise in the fitted ECI. The best compromise between these two effects can then be found.

The idea that a correct measure of the accuracy of the model should involve attempting to predict data points that are not included in the fit is quite intuitive and has been employed in previous first-principles phase-diagram calculations (see, for instance, [1991Fer]). The optimality property of CV, however, establishes the important fact that there is no need to partition the data in various ways excluding 1, 2, 3, etc., structures at a time. Simply removing one point at a time is all that is needed.[2] It is also important to note that, although the calculation of the CV score involves removing points from the fit, the CV score nevertheless gives a measure of the predictive power of the fit obtained with *all* structures included [1974Sto, 1987Li]. The CV criterion thus enables the assessment of the predictive power of the cluster expansion while still using all of the structural energies available to fit the ECI. There is no need to exclude perfectly valid data points from the fit when obtaining the ECI.

It would appear that removing one point at a time involves recomputing n distinct regressions, a process exhibiting a computational complexity of order n^2. However, the CV score can actually be computed in order n operations, using the formula

$$(\mathrm{CV})^2 = n^{-1}\Sigma_{i=1}^{n}\{[E_i - \hat{E}_i]/[1 - X_{i\cdot}(X^TX)^{-1}X_{i\cdot}^T]\}^2$$

which involves only the predicted values $\hat{E}_i$ obtained from the standard "all points included" least-squares regression and the matrix $X_{i\alpha}$ containing the values of $\langle\Pi_{j\in\alpha}\sigma_i\rangle$ for each structure i. We let $X_{i\cdot}$ denote row i of the matrix X. The order n complexity is achieved by noting that the matrix $(X^TX)^{-1}$ can be computed once and for all.

3.1.2 Narrowing the Search for ECI. Applying the CV criterion to the problem of ECI selection requires a few additional steps. CV tends to perform better in practice when the number of choices is kept small, not only because this minimizes the computational requirements of the search, but also for a more fundamental reason. The CV score is a statistical quantity and thus provides an error-contaminated measure of the "true" quantity in which we are interested. Although the statistical noise vanishes in the large sample limit, it may still affect finite-sample performances. As the number of alternative cluster choices increases, the likelihood that *one* suboptimal choice happens to give a smaller CV score than the true optimal choice increases. By restricting the search only to "physically meaningful" candidate cluster expansions, we minimize this unwanted phenomenon.

Let us now propose a formal definition of what a "physically meaningful" cluster choice is. Define the diameter of a cluster as the maximum distance between two sites in the cluster.

- A cluster can be included only if all its subclusters have already been included.
- An m-point cluster can be included only if all m-point clusters of a smaller diameter have already been included.

The first rule formalizes the observation that an m-point cluster can be interpreted as describing the coupling between an m'-point cluster and an $(m\text{-}m')$-point cluster. Under the reasonable assumption that coupling terms are less important than each term taken separately, we expect a given cluster to be associated with a smaller ECI than any of its subclusters, implying that subclusters should always be included in the expansion first. The second rule summarizes the intuition that large and extended clusters tend to describe weaker interactions than small and compact clusters. The diameter is used as a measure of spatial extent, and it is assumed that a cluster is associated with an interaction that is weaker than the weakest pair it contains, the weakest being presumably the longest.

Although it is possible that the optimal cluster expansion does not satisfy these requirements, the only consequence of imposing them is that the algorithm might choose a cluster expansion containing an unnecessarily large number of terms, with some ECI being set close to zero. These two rules are sufficient to keep the number of possibilities to a finite and reasonable number, and lead to a particularly simple algorithm to systematically go through all "physically meaningful" cluster choices.

First consider pair clusters in increasing order of length. For each choice of pair clusters, consider triplets, in increasing order of diameter, up to the diameter of the longest pair. For each choice of triplets, consider quadruplets, in increasing order of diameter, up to the diameter of the largest triplet, etc. Note that when ties between the diameters of different clusters occur, all clusters of the same diameter are added at the same time. The fact that the total number of cluster must be less than the number of structures guarantees that the number of cluster choices to be considered is finite. For instance, consider the artificially simple case where the pairs have length 1, 2, 3, . . . and similarly for larger clusters. If one has access to the energy of seven structures, the following cluster choices would be considered (each choice is described by the maximum diameter of the included pairs, triplets, quadruplets, etc.): (1), (1,1),

[2]Note that this property, however, relies on the assumption that the errors in the calculated energies (because of, for instance, numerical noise) are statistically independent, an assumption that should be justified in the context of cluster expansion construction.

(1,1,1), (1,1,1,1), (2), (2,1), (2,1,1), (2,2), (3), (3,1), (4). The number of clusters included is at most 4, because the empty cluster and the point cluster (assuming there is only one) are always included, and because CV requires at least one more structure than there are clusters.

It is possible that for a particular class of systems, more specific rules can be derived (e.g., fcc-based transition metal alloys, the fourth nearest-neighbor pair is often associated with a larger ECI than the third nearest-neighbor pair [1991Duc]), but as a general system-independent rule, this hierarchy of clusters seems sensible.

3.1.3 Ground State Prediction. The accuracy of a cluster expansion is not solely measured by the error in the predicted energies. It is particularly important that a cluster expansion is able to predict the correct ground states. Because the mean squared error focuses on optimizing the absolute energy values, whereas the ground states are determined by the ranking of energies, the lowest mean squared error does not necessarily lead to the most accurate prediction of the ground state line, as discussed in detail in [1995Gar]. For this reason, the search for the best cluster expansion in our algorithm gives an absolute preference to choices of clusters which yield the same ground states as the calculated energies. More specifically, a candidate cluster expansion that has a higher CV score but predicts the correct ground states will be preferred to a cluster expansion that has a lower CV score but predicts incorrect ground states.

For a given set of structures, it is possible that no candidate cluster expansion predicts the correct ground states. In these cases, a simple way to ensure that the ground states are correctly predicted is to give extra "weight" to specifically chosen structures to obtain the correct ground states. We now have to also define an appropriate CV score for the purpose of ranking the quality of weighted least-squares fit. Because a linear regression of E_i on $X_{i\alpha}$ with weights w_i is equivalent to a regression of w_iE_i on $w_iX_{i\alpha}$, the natural extension of the CV score to weighted fits is

$$(\mathrm{WCV})^2 = n^{-1} \Sigma_{i=1}^{n} [w_i(E_i - \hat{E}_{-i})]^2,$$

which, for the purpose of calculations, can written as

$$(\mathrm{WCV})^2 = n^{-1}\Sigma_{i=1}^{n} \{w_i[E_i - \hat{E}_i]/ [1 - w_iX_{i\cdot} (X^TW^2X)^{-1}X_{i\alpha}{}^Tw_i]\}^2$$

where $W_{ij} = \delta_{ij}w_i$. The following heuristic rule is used to set the weights. First, assign a unit weight to all structures. If the ground states are not correctly predicted, flag the problematic structures as follows. Let c_i be the concentration of structure i. Let g_j be the indices of the true ground states among the calculated structures, sorted in increasing order of concentration. If a structure i is such that $(\hat{E}_i,c_i)$ lies below the line joining $(\hat{E}_{g_j},c_{g_j})$ and $(\hat{E}_{g_{j+1}},c_{g_{j+1}})$ for some j, then mark structures i, g_j, g_{j+1} as problematic. Increase the weights of all problematic structures by one and repeat the process until the ground states are correctly predicted or until a specified number of unsuccessful trials has been reached. The rationale behind this algorithm is that, whenever possible, weights should be avoided. By adjusting weights, one can get a wide variety of results. When only the weights that are strictly needed to get a qualitatively correct ground state line are adjusted, we minimize the level of arbitrariness of the fitting procedure.

3.2 Structure Selection

How can we select which structural energy to add to an existing fit to improve its accuracy at the least computational cost? Ideally, one would again select the structure that yields, for a given amount of computational time, the largest reduction in the prediction error of least-squares fit, as estimated by the CV score. Unfortunately, the CV score requires the knowledge of the energy of the new candidate structure to be added to the fit, which makes this approach unfeasible. However, as we will now describe, it is still possible to construct a useful estimate of the improvement in the accuracy of a cluster expansion even without the knowledge of the energy of a candidate new structure to be added to the fit.

3.2.1 Variance Reduction. As discussed in more details in the Appendix, the prediction error of a least-squares fit can be separated into two components: the bias and the variance. These quantities are best described in terms of the following thought experiment. If you were to choose at random many samples of n structures and were to compute a clusters expansion for each sample, each fit would give slightly different values for the ECI. The ECI, averaged over different samples, may give a slightly biased estimate of the true value of the ECI (because a truncated cluster expansion does account for all possible sources of energy fluctuations). In addition to this systematic bias component, the ECI fitted to each sample will exhibit fluctuations around the mean, the magnitude of which can be characterized by a variance component. Of course, because the ECI are a multidimensional quantity, a matrix of covariances is needed to fully characterize the fluctuations.

Ideally, we would want to add to the fit the structure that most reduces the sum of the bias squared and of the variance. Unfortunately, the reduction in the bias is impossible to predict without the knowledge of the energy of the new candidate structure to be added.[3] We thus focus solely on the variance component, which can be estimated without the knowledge of the energy of the new structure.

Is it well known from the theory of least-squares estimation [1991Gol] that the covariance matrix of the ECI is given by

$$V = (X^TX)^{-1}e^2 \qquad \text{(Eq 2)}$$

where e^2 is the mean squared error of the fit. Note that X, which is the matrix of the $\langle \Pi_{j_\epsilon\alpha}\sigma_j \rangle$ for each structure, indeed does not depend on the structural energies. (In principle, the e^2 term depends on the structural energy, but it is a second-

[3]One would need to known the value of the ECI obtained with the new structure included in the fit, which is impossible without the knowledge of its energy.

order effect. To see this, consider e^2 as a function of the ECI. Because a least-squares fit is obtained by minimizing e^2, it follows that e^2 is a quadratic function of the ECI in the vicinity of the minimum, implying that a first-order change in the estimated ECI yields a second-order change in the estimated e^2.)

We can now use Eq (2) to derive the variance of the predicted energies. The predicted energy $\hat{E}_i$ of structure i is a linear function of the ECI

$$\hat{E}_i = \Sigma_\alpha X_{i\alpha} J_\alpha \equiv X_{i\cdot} J.$$

and the variance of a linear function of a vector $J_\cdot$ with known covariance matrix V is given by

$$\text{Var}\,[\hat{E}_i] = X_{i\cdot} V X_{i\cdot}^T.$$

This expression gives the variance of the predicted energy of only *one* structure. To quantify the predictive power of the cluster expansion, we then need to average this quantity over all structures (even those not yet included in the fit). The set of the values of the correlations for every possible structure is hard to characterize. Although it is known to be a convex polytope embedded in a cube of side two, constructing this polytope is computationally intensive [1991Duc]. To keep the calculations tractable, we make the simplifying assumption that the correlations $X_{i\cdot}$ of every possible structure are distributed isotropically in a sphere. The expected variance of a structure picked at random is then given by an integral over a sphere S weighted by some spherically symmetric density $f(|u|)$:

$$\begin{aligned} E\,[\text{Var}\,(\hat{E}_i)] &= \int_S u^T V u f(|u|)\, du \\ &= \int_0^\infty \left(\int_{|u|=r} u^T V u du\right) f(r)\, dr \\ &= \left(\int_{|u|=r} u^T V u du\right) \int_0^\infty f(r)\, dr \\ &\propto \left(\int_{|u|=1} u^T V u du\right) \\ &= \text{tr}\,(V). \end{aligned}$$

The last step follows from the fact that for any set of orthogonal vectors u_i, tr $(V) = \Sigma_i u_i^T V u_i$. In summary, the trace of the covariance matrix of the ECI

$$\text{tr}\,[e^2 (X^T X)^{-1}]$$

provides us with a criterion to estimate the expected variance of the energies predicted from a cluster expansion.

The question becomes: What happens to the expected variance when a new structure is added to the fit? Changes in e^2 can be shown to be of a second order as follows. The least squares procedure minimizes e^2, implying any first-order change in the ECI (as a result of adding a new point to the fit) yields no first-order change in e^2. We can thus focus solely on finding the new structure i that maximizes the reduction r of the trace of $(X^T X)^{-1}$:

$$r = \text{tr}\,[(X^T X)^{-1}] - \text{tr}\,[(X^T X + X_{i\cdot}^T X_{i\cdot})^{-1}]$$

where $X_{i\cdot}$ values are the correlations of the new structure to be added. Because the length of the vector $X_{i\cdot}$ is bounded, there is a limit to how much a single structure can reduce the variance of the fit. The maximum variance reduction $\Delta V_{\max}$ is reached when $X_{i\cdot}^T$ is the longest possible column vector v parallel to the eigenvector of $X^T X$ associated with the smallest eigenvalue:

$$\Delta V_{max} = \text{tr}\,[(X^T X)^{-1} - (X^T X + v v^T)^{-1}] \qquad \text{(Eq 3)}$$

where v is normalized so that $\max_j |v_j| = 1$, because correlations cannot exceed 1 in magnitude.

This result can be shown by maximizing tr $[(X^T X)^{-1} - (X^T X + v v^T)^{-1}]$ with respect to v while introducing the constraints that $v^T v = c$ through a Lagrange multiplier λ. The first-order condition for this maximization problem can be found by first calculating the first-order change in tr $[(X^T X)^{-1} - (X^T X + v v^T)^{-1}]$ with respect to changes in $v v^T$:

$$\begin{aligned} \text{tr}[(X^T X)^{-1} - (X^T X + v v^T)^{-1}] &= \text{tr}\,\{[X^T X]^{-1} \\ &\quad - (I + (X^T X)^{-1} v v^T)^{-1} (X^T X)^{-1}) \\ &\approx \text{tr}\,\{(X^T X)^{-1} \\ &\quad - (I - (X^T X)^{-1} v v^T)\,(X^T X)^{-1}\} \\ &= \text{tr}\,[(X^T X)^{-1} v v^T\,(X^T X)^{-1}] \\ &= \text{tr}\,[(X^T X)^{-2} v v^T] \\ &= \text{tr}\,[v^T (X^T X)^{-2} v] \\ &= v^T\,(X^T X)^{-2} v \qquad \text{(Eq 4)} \end{aligned}$$

Note that the elements of the matrix $X^T X$ increases in magnitude linearly with the number n of structures included in the fit while the elements of the matrix $v v^T$ remain bounded. As the number of structure increases, $v v^T$ becomes small relative to $X^T X$, thus justifying the use of the above Taylor expansion.

The first-order condition is then given by

$$\nabla_v\,[v^T (X^T X)^{-2} v - \lambda\,(v^T v - c)] = 0,$$

which is just an eigenvalue problem:

$$(X^T X)^{-2} v = \lambda v.$$

Because we are looking for a maximum, we seek the largest eigenvalue of $(X^T X)^{-2}$ or, equivalently, the smallest eigenvalue of $X^T X$. This establishes that v must be parallel with the eigenvector associated with the smallest eigenvalue of $X^T X$.

3.2.2 Trade-Off Between Computational Requirements and Variance Reduction. We now derive an optimality criterion for selecting a new structure to include in the fit that takes both variance reduction and computational requirements into account. Consider the change in variance when one structure is added to a fit already containing n structures. Then $X^T X$ is of order n, while the change in $X^T X$ due to the new structures, $X_{i\cdot}^T X_{i\cdot}$, is of the order 1. Assuming that $n \gg 1$, a Taylor series argument similar to that of Eq 4 yields:

$$\text{tr}\,[(X^T X)^{-1}] - \text{tr}\,[(X^T X + X_{i\cdot}^T X_{i\cdot})^{-1}] \approx \text{tr}\,[(X^T X)^{-2} X_{i\cdot}^T X_{i\cdot}]$$

This shows that when the number n of structures included is large, the variance-reducing effect of adding a few structures to the fit is linear. Hence, when faced with the decision to add a structure reducing the variance by ΔV_1 at a com-

putational cost C_1 or a structure reducing the variance by ΔV_2 at a computational cost C_2, we should add the structure that maximizes $\Delta V_i/C_i$. Indeed, if $C_1 < C_2$ one can compute about C_2/C_1 structures similar to structure 1 in the time it takes to compute structure 2. The resulting variance reduction using structure 2 is ΔV_2, whereas the variance reduction using C_2/C_1 structures similar to structure 1 is $C_2/C_1\Delta V_1$. Structure 2 will be preferable if $\Delta V_2 > \Delta V_1 C_2/C_1$ or $\Delta V_2/C_2 > \Delta V_1/C_1$.

The computational cost of obtaining the energy of a structure can be estimated in advance from the number of atoms in its unit cell and the known scaling law of the first-principles method used.[4]

Therefore, the algorithm to choose the "best" structure is the following. We scan through structures in increasing order of computational cost, that is, in increasing order of unit cell size. This task can be accomplished using well-known structure enumeration algorithms [1991Fer]. For each structure, we compute the potential gains for adding it to the fit

$$G_i = \mathrm{tr}\,[(X^TX)^{-1} - (X^TX + X_{i\cdot}{}^TX_{i\cdot})^{-1}]/C_i$$

where $X_{i\cdot}$ and C_i are, respectively, the correlations and the computational cost of structure i. The search can be aborted when

$$\Delta V_{\max}/C_i < G_{I,\max}$$

where $G_{I,\max}$ is the best gain G_i found so far, and $\Delta V_{\max}$ is the maximum possible variance reduction, given by Eq 3.

3.2.3 Ground State Prediction. We have so far solely focused on reducing the variance of the fit. But a successful cluster expansion must be able to predict the correct ground states and a special effort must be made to reduce the possibility of incorrect ground states. To address this issue, whenever a cluster expansion is constructed, we generate a large number of structures that were not included in the fit, and check whether there are new predicted ground states among these structures. If there are none, then the minimum variance criterion for selecting new structures just presented is appropriate, because the cluster expansion we have is likely to be qualitatively correct and we want only to obtain more precise ECI. If new ground states *are* predicted, verifying their validity has priority over minimizing the variance. In this case, the search for the most variance-reducing structure is thus restricted to the set of newly predicted ground states.

3.3 Minimal Cluster Expansion

Before the refinement procedure described above can be used, a "minimal" start-up cluster expansion must be constructed and its ECI determined. The minimal cluster expansion consists of the empty cluster, all the point clusters, and all the nearest-neighbor pairs surrounding all atoms.[5] To determine the ECI, we simply add structures to the fit in increasing order of computational requirements[6] until we have one more structure than there are ECI (so that the CV score can be calculated). At this point, refinement with the procedures outlined above can be started.

4. Implementation

4.1 Design goals

After having described the theoretical principles supporting our automated algorithm, we are now ready to discuss its practical implementation. While designing our automated phase diagram calculation system, we sought to fulfill a number of important requirements.

The system should be able to run over a long period of time without requiring user intervention while still allowing the user to intervene at any time, if desired. Such a feature is useful in the context of ab initio calculations because the first few structural energy calculations are typically set up manually to determine the values of the various input parameters of the ab initio method that offer the best trade-off between accuracy and computational requirements. Once these parameters are chosen, they can be kept constant for the subsequent runs and the remainder of the process can be entirely automated.

The system should keep the user informed of the best result obtained so far, as the calculation proceeds. In this fashion, the user can interrupt the process at any time if the currently available results are suitably accurate. In addition, the user has constant access to preliminary results, which can be used to try out subsequent processing steps (e.g., MC simulations), while the code keeps running, continuously improving the precision of the cluster expansion.

Another important requirement is fault tolerance. As good as modern first-principles codes are, they sometimes fail and our system should gracefully handle these instances. The user may also decide at a later time to recalculate some energies more accurately, and the code should keep track of such updates.

We wanted our system to be easily adaptable to any computing environment or energy methods. Many first-principles codes and many computer architectures are currently available. We cannot simply assume that one code or one system will eventually replace all others. Finally, we wanted our system to make maximum use of parallelization opportunities.

4.2 Organization of the Package

We achieve the goals described in the previous section through the setup illustrated in Fig. 1. The process of fitting a cluster expansion using the algorithm presented in the previous section is performed by a subsystem called the

[4]The symmetry of the structure also has some effect on computational costs, although we neglect this fact to reduce the computational burden of finding the best structure to add.

[5]Consider a site centered at the origin and sort all other sites (centered at r_i) in increasing order of distance from the origin. The nearest-neighbor shell can be defined as the sites $r_1, \ldots, r_n$ where n is the largest n such that $r_i \cdot r_j < 0$ for all $i,j \leq n$. When there are more than one atom per unit cell, we consider the radius of each nearest neighbor shell, take the maximum $r_{\max}$, and then take all the pairs shorter $r_{\max}$.

[6]Structures for which correlations are coplanar are skipped.

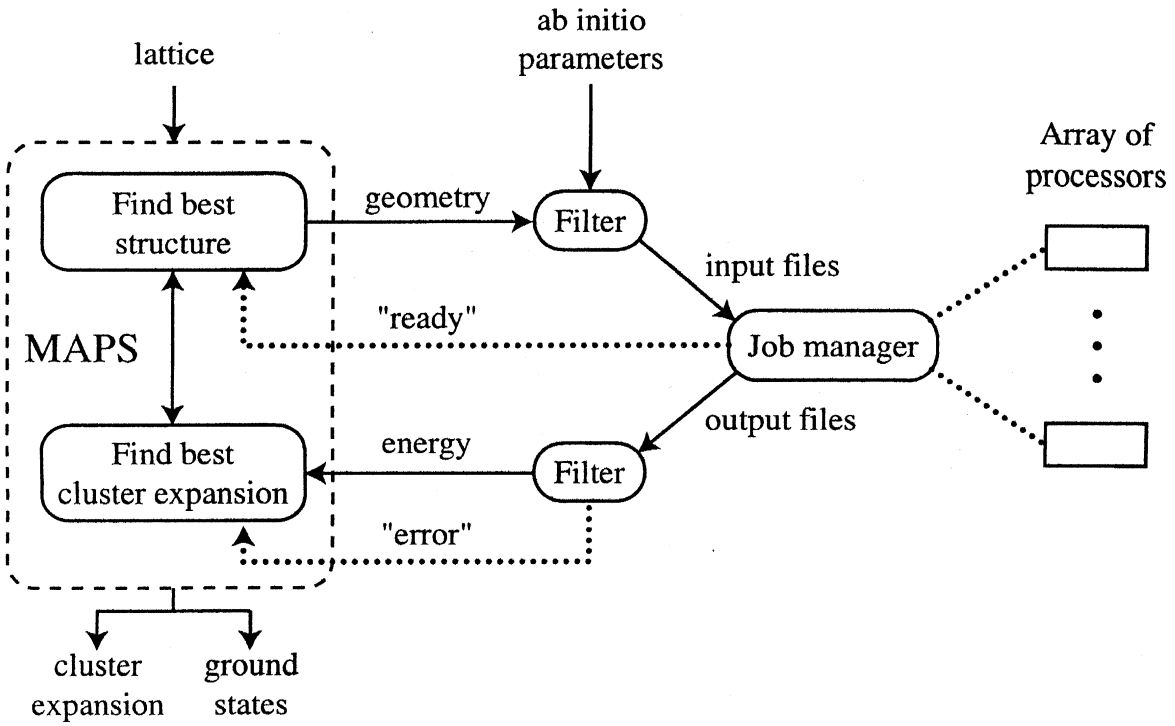

Fig. 1 Organization of the automated cluster expansion package. The MAPS module alternatively solves two optimization problems: (i) given a cluster expansion, find the best structure to add to the fit to improve the accuracy of the cluster expansion (ii) given a set of structures of known energy, find the cluster expansion that provides the best predictive power. The MAPS module communicates with the first-principles code and a *job manager* via a set of messages that (i) provide the geometry of the structure whose energy needs to be calculated ("geometry") (ii) provide the calculated energy of the requested structure ("energy") (iii) indicate the availability of processor cycles ("ready") (iv) indicate the presence of an error ("error").

MIT ab-initio phase stability (MAPS) code. This subsystem is completely independent of both the ab initio code and the computing environment used. It implements the algorithms which determine, at any given time, (1) the best structure to add to a fit, and (2) the best clusters to add to a fit. The only input this subsystem requires is the geometry of the lattice, which can typically be described by a file less than 10 lines long. All other information needed for the construction of a cluster expansion, such as the space group of the lattice or a list of candidate clusters and structures, is automatically constructed from the geometrical input.

The remaining subsystems *are* dependent on the ab initio code and the computing environment used. They consist of various simple "scripts" that can be easily adapted to any environment. A first filter combines a structure's geometric information, generated by MAPS, with various ab initio code-specific input parameters to produce input files for the first-principles code. In the case of pseudopotential-based codes, the code-specific parameters include k-point mesh density, energy cutoff, and various switches that indicate the type of calculation to be performed. Less than 10 lines of input are typically needed. Another filter extracts the structure energies from the output files of the first-principles calculations.

A "job manager" serves as the interface between MAPS and an array of processors dedicated to structural energy calculations. Communication takes place through a set of simple signals:

- Ready: Processors are available to run first-principles calculations and MAPS should generate new structures the energy of which is to be determined.
- Energy: One first-principles code has successfully terminated and the energy of the structure is available.
- Error: An error has occurred and the energy of a given structure is not available at this point.

A specific signal is given by simply creating a file of the corresponding name. In contrast to more sophisticated communication schemes, such a simple approach is easy to implement and to customize by the end user. It is also very portable across different computer systems. These files can either be generated by the user or by simple automated scripts, thus allowing the code to be either controlled manually or automatically. MAPS simply monitors the creation of these files and reacts accordingly. If a new structural energy becomes available (or a new error is reported), a new cluster expansion is fitted. If a new processor becomes available, a new structure is generated.

Because most structural energy calculations can comfortably run on a single processor, it is natural to parallelize the process at the level of the energy calculation. At this level, there is very little need for interprocess communication, and the parallelization efficiency essentially reaches 100%. Any group of workstations can thus be directly set up to cooperatively construct a cluster expansion without requiring the installation of sophisticated networks or libraries.

This software package can be obtained by contacting the authors at avdw@alum.mit.edu.

5. Applications

To illustrate the usefulness of our algorithms and associated software package, we computed the phase diagrams of a wide variety of alloy systems: Si-Ge, CaO-MgO, Ti-Al, and Cu-Au. Our examples include insulating (CaO-MgO), semiconducting (Si-Ge), and metallic systems (Ti-Al and Cu-Au). The crystal structure of our example systems also exhibits a wide variety, including systems in which the parent lattice has more than one atom per primitive unit cell (Si-Ge), systems that have spectator ions (O in the CaO-MgO system), or systems with multiple parent lattices (hcp and fcc in the Ti-Al system).

With a few exceptions discussed below, our predicted phase diagrams exhibit the same stable phase as the experimentally determined phase diagram. However, as is commonly the case in phase diagrams calculated from first-principles, the temperature scale of the phase transitions is often overestimated. The physical origin of this overestimation has not been unambiguously established, but likely candidates are (1) the omission of the effect of lattice vibrations [1994Gar, 1996Gar, 1998Wal, 2002Wal] (2) the omission of weak but very long-range elastic interactions due to atomic size mismatch [1992Lak, 1998Ozo3, 1998Wol, 1998Ozo1] and (3) the limited precision of first-principles calculations based on the local density approximation (LDA). Let us discuss each point in turn. First, the formalism that account for lattice vibrations and electronic excitations have been developed and would represent a very important addition to our package. These effects can be included by replacing the cluster expansion of the energy of each alloy configuration σ by a cluster expansion of the (local) vibrational and electronic free energy associated with each alloy configuration σ [1994Gar, 1996Gar, 1998Wal, 2002Wal]. Second, although the present calculations accounts for elastic interactions due to atomic relaxations, they are truncated to a finite range in order to reduce the computational requirements of the Monte Carlo simulations. For systems where this approximation may not be appropriate, our package also includes a module enabling the inclusion of *long-range* elastic interactions using the constituent strain formalism [19992Lak]. Finally, it is important to note that our package can easily be adapted to make use of improvement upon the LDA first principles energy method, as they become available.

5.1 Methodology

Our first-principles calculations are performed within the LDA using the VASP [1996Kre2, 1996Kre1] package, which implements ultrasoft [1990Van] pseudopotentials [1959Phi]. In all calculations, we used the default settings implied by the high-precision option of the code. The *k*-point mesh was constructed using the Monkhorst-Pack scheme and was chosen such that the number of *k*-points times the number of atoms in the unit cell was at least 3500 for metallic systems and 1000 for insulating systems. This choice keeps the *k*-point density constant, despite changes in the unit cell size. To ensure that the *k*-point density is as isotropic as possible, the number of mesh points along a given reciprocal lattice vector a_1 was set proportional to $|a_1 \cdot (a_2 \times a_3)|/|a_2 \times a_3|$, where a_2 and a_3 are the two remaining reciprocal lattice vectors. In all structural energy calculations, both the atom positions and the cell shape are allowed to relax to their equilibrium values. In all cases, we neglect the potential presence of vacancies, an assumption that is clearly justified in the case of the CaO-MgO and Si-Ge systems but that may deserve further consideration in the cases of the metallic Ti-Al and Cu-Au alloy systems. In all cases, the thermodynamic contributions arising from vibrational and electronic excitations were neglected.

Table 1 Characteristics of the Calculated Cluster Expansions

Characteristic	Si-Ge	CaO-MgO	Ti-Al hcp	Ti-Al fcc	CuAu
Number of structures	27	20	55	23	33
Number of clusters	2 + 8 + 3	2 + 3 + 7 + 1	2 + 11 + 6	2 + 3 + 2	2 + 6
CV score, meV/atom	1	18	35	49	23

The number of clusters is given as the number of each type of multiplet: empty and point clusters + pairs + triplets + quadruplets

Our calculated phase diagrams were obtained through LTE [1998Koh] and (MC) simulations [1988Bin]. Our MC simulation cells ranged from 10 × 10 × 10 supercells to 20 × 20 × 20 supercells. The number of equilibration MC passes and the number of averaging MC passes that were used to obtain thermodynamic quantities were chosen so that the error on the average concentration of the alloy is less than 0.1%. This precision typically required from 2000 to 50,000 equilibration MC passes and from 2000 to 100,000 averaging MC passes, depending on temperature.

The thermodynamic integration method was used to determine the grand canonical potential ϕ^α of each phase α as a function of chemical potential μ. The grand canonical potential at the starting point of the integration path was obtained from the LTE. Phase boundaries between two phases α and β are found by locating the intersection of the curves $\phi^\alpha(\mu)$ and $\phi^\beta(\mu)$. The concentrations of each phase in equilibrium at the phase transition are given by the slope of the curves of the point of intersection. Additional details regarding these calculations will be presented in a separate communication [2002Wal].

5.2 Examples

While a summary of the characteristics of the five cluster expansions we constructed can be found in Table 1, the details specific to each system are discussed below.

5.2.1 The Si-Ge System. First-principles investigation of phase stability in the Si-Ge system were performed ear-

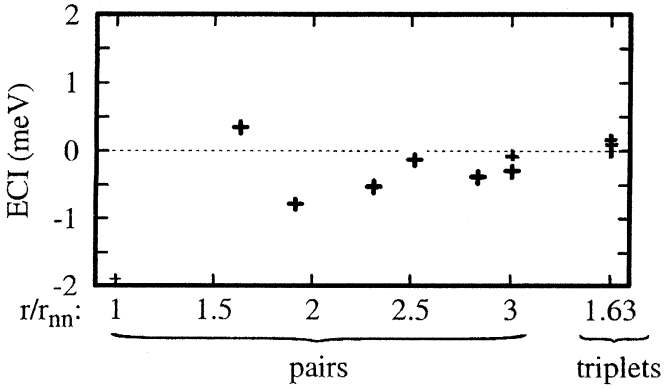

Fig. 2 Calculated ECI as a function of cluster diameter for the Si-Ge system

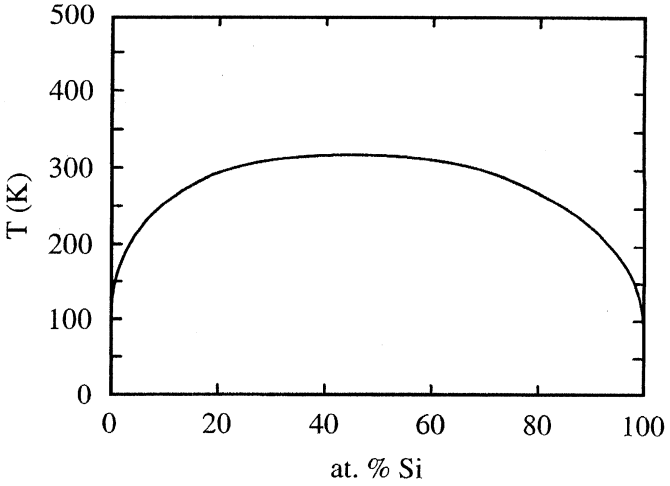

Fig. 3 Calculated phase diagram of Si-Ge system

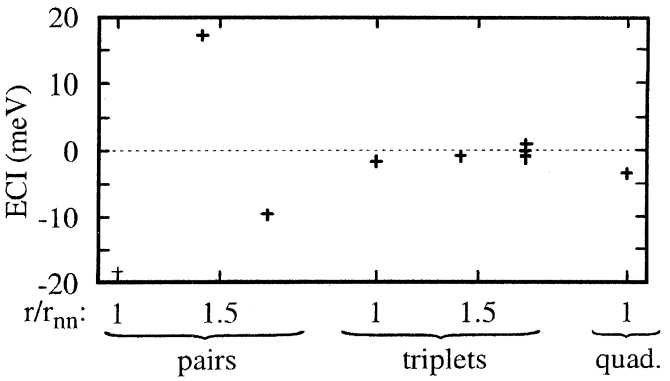

Fig. 4 Calculated ECI as a function of cluster diameter for the CaO-MgO system

lier [1988Qte, 1991Gir]. One of the most thorough studies [1991Gir] relied on the so-called computational alchemy method. Although this ingenious method enables the direct determination of long-range pair interactions, clusters larger than pairs are not included in their full generality. (Computational alchemy does allow the pair interactions to depend on concentration, thereby implicitly accounting, to a limited extent, for multiplet interactions.) It is thus of interest to verify that such a simplification is appropriate by employing our automated tool, which allows the inclusion of multiplets. Figure 2 shows that the energetic contribution of pairs indeed dominates the contribution of multiplets. Not only are the included multiplets associated with very small ECI, but the CV algorithm automatically determines that very few triplets should be included in the cluster expansion. Figure 3 shows the calculated phase diagram. The predicted critical temperature (325 K) lies between the predictions of [1988Qte] (375 K) and [1991Gir] (170 K).

5.2.2 The CaO-MgO Pseudo-Binary System. A very complete first-principles analysis of the CaO-MgO system was undertaken in [1996Tep]. Although a substantial amount of labor was needed to complete this analysis, our software package enabled us to reproduce this study[7] in a fully automatic fashion. In contrast to the Si-Ge system, multiplet interactions are important in the CaO-MgO system (Fig. 4) and are responsible for the pronounced asymmetry in the miscibility gap observed both in our calculation and in the experimental measurements (Fig. 5). This feature was not given as an input to the code and it is interesting that the code was able to identify it independently.

5.2.3 The Ti-Al System. Our third example seeks to reproduce the very detailed first-principles calculation of the Ti-Al phase diagram found in [1993Ast]. We focus on the Ti-rich portion of the phase diagram because the Al-rich half of the phase diagram exhibits order-order transitions that would prompt the inclusion of the thermodynamic effects of lattice vibrations, which is beyond the scope of the present study. For the same reason, we do not investigate the Ti (hcp) → Ti (bcc) transition.

However, we do model both hcp-based phases (the Ti solid solution and the DO_{19} Ti_3Al phase) and the fcc-based $L1_0$ TiAl phase. As shown in Fig. 6, the cluster expansion on the hcp lattice could easily be pushed to a higher accuracy than in the earlier study [1993Ast], which only included up to second-nearest-neighbor pair interactions and a few multiplets. The cluster expansion on the fcc lattice (Fig. 7) does not need to be as accurate because the location of the phase boundaries was found to be insensitive to further improvement in the fcc cluster expansion.

Figure 8 shows our calculated phase diagram. As discussed at the beginning of this section, our calculations overestimate the experimental order-disorder transition temperature of the DO_{19} Ti_3Al phase (1450 K). However, our calculated order-disorder transition temperature of 1850 K corroborates the first-principles prediction of 1800 K found by [1993Ast].

5.2.4 The Cu-Au System. The most accurate first-principles study of the Cu-Au system is described in a series of articles [1998Ozo1, 1998Wol, 1998Ozo2]. Although our calculations include neither infinite range interactions nor the contribution of lattice vibrations, we were able to obtain good quantitative agreement with these earlier calculations as well as with experimental measurements, as seen by the transition temperatures reported in Table 2. We focus on the

[7]This applies except for the inclusion of lattice vibrations.

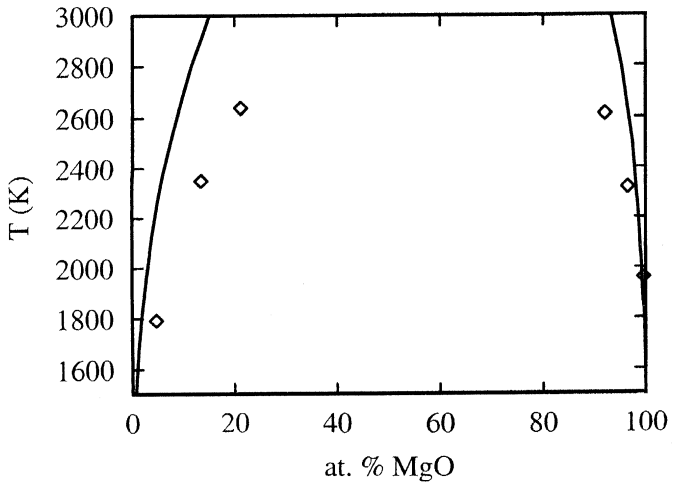

Fig. 5 Calculated (lines) and experimental (diamonds) phase diagrams of the CaO-MgO system

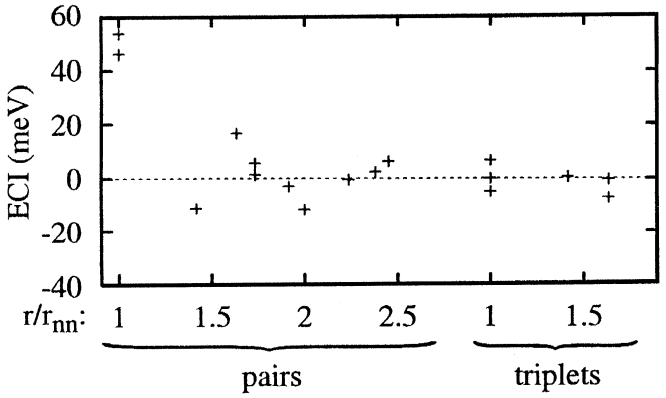

Fig. 6 Calculated ECI as a function of cluster diameter for the hcp phases of the Ti-Al system

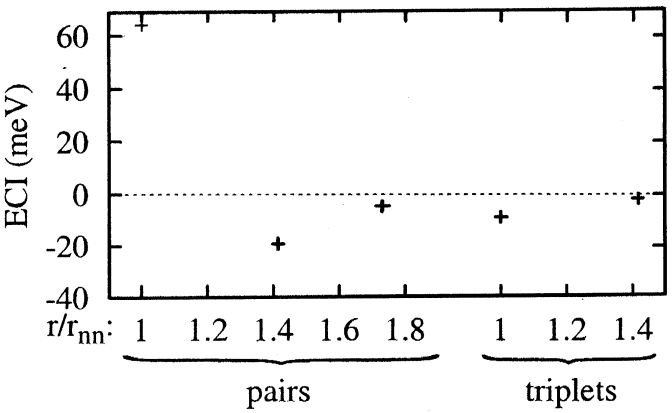

Fig. 7 Calculated ECI as a function of cluster diameter for the fcc phases of the Ti-Al system

Cu-rich half of the phase diagram, because there is well-documented disagreement between the experimentally observed ground state at the $CuAu_3$ composition and that predicted from LDA calculations [1998Ozo1].

As in earlier LDA calculations, we were not able to model the CuAu-I → CuAu-II transition, which is driven by very small free energy differences that are beyond the accuracy of current LDA calculations [1998Ozo1]. Our calculated ECI are plotted in Fig. 9, whereas the resulting phase diagram is shown in Fig. 10.

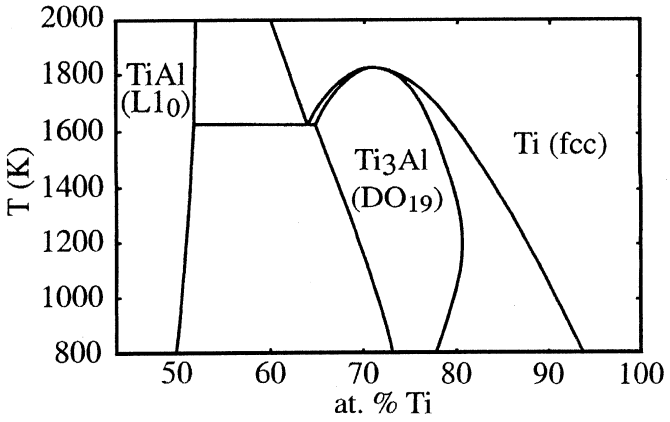

Fig. 8 Calculated phase diagram of the Ti-Al system (Ti-rich portion)

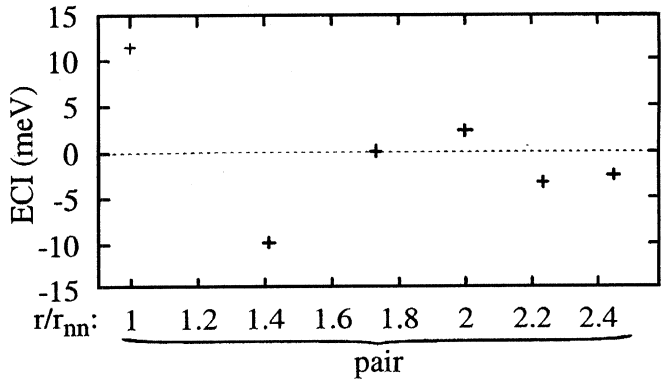

Fig. 9 Calculated ECI as a function of cluster diameter for the Cu-Au system

Table 2 Calculated and Experimental Transition Temperatures of the Cu-Au System

Transition	Present Work	Ref. [1998Ozo2]	Expt.
Cu_3Au: $L1_2 \rightarrow$ fcc	460 K	455 K	663 K
CuAu: $L1_0 \rightarrow$ fcc	430 K	560 K	683 K

6. Conclusion

We formalized the decision rules that allow the automatic construction of a cluster expansion. The development of a fully automated procedure to compute phase diagrams from first principles allows the powerful cluster expansion formalism to be used by anyone who possesses basic material science knowledge. Through a variety of examples, we have shown that earlier first-principles phase diagram calculations, which represented significant contribution to

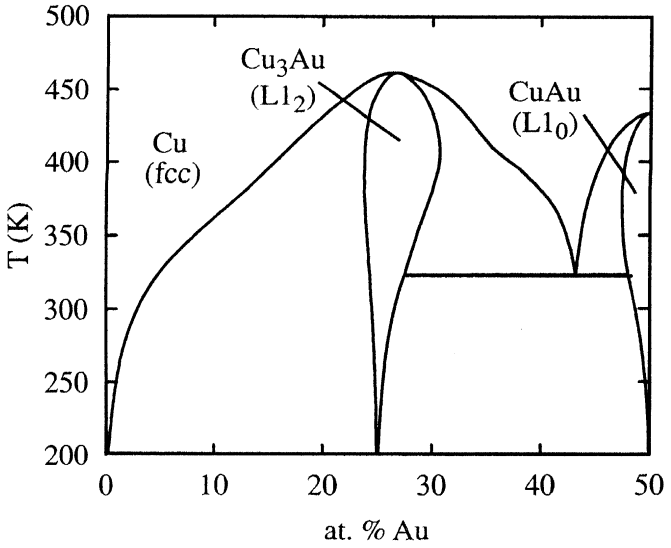

Fig. 10 Calculated phase diagram of the Cu-Au system (Cu-rich portion)

the field of alloy theory at the time of their publication, can now be reproduced in an automated fashion.

A variety of extensions to our work would prove extremely useful. A module that computes the vibrational free energies at a reasonable computational cost (for instance, through the methods presented in [1988Mor] or [2000Wal, 2002Wal]) would allow the modelization of allotropic transitions and improve the precision of the calculated transition temperatures. The inclusion of electronic entropy should be straightforward [1995Wol]. The explicit inclusion of elastic contributions to the energy through a reciprocal space cluster expansion [1992Lak, 1998Ozo1] would also be very helpful.

7. Appendix

This section provides a heuristic proof of the validity of the CV selection rule. A formal proof, under the assumptions relevant to our application, can be found in [1987Li]. The concepts used are described in more detail in [1974Sto, 1991Gol].

Let us first describe the conventions underlying our discussion. For a given choice of clusters $\alpha_1, \ldots, \alpha_k$, consider a sample of n correlation vectors $X_{i\cdot} = (X_{i\alpha 1}, \ldots, X_{i\alpha k})$, for $i = 1, \ldots, n$. Each correlation vector $X_{i\cdot}$ is associated with a large set of structures that have the same correlation vector $X_{i\cdot}$ (although their other correlations, not included in $X_{i\cdot}$, differ). We then consider the structural energy E_i obtained for a structure with correlation $X_{i\cdot}$ to be a random variable, because fixing $X_{i\cdot}$ does not entirely determine E_i. To handle this randomness, we will be concerned with calculating expectation values, denoted $\langle\cdot\rangle$, averaged over every possible sample of n structures having the given correlation vectors $X_{i\cdot}$, $i = 1, \ldots, n$. For instance, $\langle E_i\rangle$ is the average energy of all structures with correlation $X_{i\cdot}$, which is equal to the energy that would be obtained from a cluster expansion that uses the true exact ECI for all included clusters $\alpha_1, \ldots, \alpha_k$, but a zero ECI for all the other clusters.

Let us now define two important quantities, adopting the convention that $i = 0$ denotes a structure (with correlation $X_{0\cdot}$) that is not included in the least-squares fit.

- Let $\hat{E}_i$ (for $i = 0, \ldots, n$) be the energy of structure i predicted using a cluster expansion fitted to the n structural energies available.
- Let $\hat{E}_{(i)}$ (for $i = 1, \ldots, n$) be the energy of structure i predicted using a cluster expansion fitted to all known structural energies except the one of structure i, for a total of $n - 1$ structures.

We are now ready to compute the predictive power P of the cluster expansion, that is, the expected squared error of the prediction of a structural energy not included in the fit:

$$P = \langle(\hat{E}_0 - E_0)^2\rangle.$$

In general, for any structure i, the expected squared error can be decomposed as

$$\langle(\hat{E}_i - E_i)^2\rangle = \langle(E_i - \langle E_i\rangle)^2\rangle + \langle(\hat{E}_i - \langle E_i\rangle)^2\rangle - 2\,\langle(E_i - \langle E_i\rangle)\,(\hat{E}_i - \langle E_i\rangle)\rangle, \quad \text{(Eq 5)}$$

where $\langle(E_i - \langle E_i\rangle)^2\rangle$ is the so-called bias term, and $\langle(\hat{E}_i - \langle E_i\rangle)^2\rangle$ is the so-called variance term. The covariance term $\langle(E_i - \langle E_i\rangle)\,(\hat{E}_i - \langle E_i\rangle)\rangle$ vanishes for a structure not included in the fit ($i = 0$) because E_0 has no effect on $\hat{E}_0$. However, for a structure included in the fit ($i > 0$), this covariance term is positive. Whenever E_i is large, $\hat{E}_i$ will tend to be large as well, because the least-squares procedure attempts to minimize the distance between $\hat{E}_i$ and E_i. This effect worsens as the ratio k/n increases. For this reason, attempting to estimate the predictive power using the mean squared error of the least-squares fit leads to a biased estimate of the true predictive power. This problem is corrected by using the CV score, which focuses on the quantity $\langle(\hat{E}_{(i)} - E_i)^2\rangle$ instead of $\langle(\hat{E}_i - E_i)^2\rangle$. In Eq 5, the expected value of the bias term is unaffected by this change, whereas the variance term is changed by an amount that goes to zero as sample size n grows. More importantly, the covariance term $\langle(E_i - \langle E_i\rangle)\,(\hat{E}_{(i)} - \langle E_i\rangle)\rangle$ now vanishes for $i = 1, \ldots, n$ because the value of $\hat{E}_{(i)}$ does not depend on E_i. It follows that $\langle(\hat{E}_{(i)} - E_i)^2\rangle$ is an unbiased measure of the predictive power. By the law of large numbers, the abstract quantity $P = \langle(\hat{E}_{(i)} - E_i)^2\rangle$ can be estimated by the corresponding sample average $n^{-1}\sum_{i=1}^{n}(\hat{E}_{(i)} - E_i)^2$, the CV score.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under contract no. DE-F502-96ER 45571. Gerbrand Ceder acknowledges support of Union Minière through a Faculty Development Chair. Axel van de Walle acknowledges support of the National Science Foundation under program DMR-0080766 during his stay at Northwestern University

References

1951Kik: R. Kikuchi: *Phys. Rev.*, 1951, *81*, pp. 988-1003.

1959Phi: J.C. Phillips and L. Kleinman: *Phys. Rev.*, 1959, *116*, pp. 287-94.

1974Sto: M. Stone: *J. R. Stat. Soc. B Met.*, 1974, *36*, pp. 111-47.

1983Con: J.W. Connolly and A.R. Williams: *Phys. Rev. B*, 1983, *27*, pp. 5169-172.

1984San: J.M. Sanchez, F. Ducastelle, and D. Gratias: *Physica*, 1984, *128A*, pp. 334-50.

1987Li: K.-C. Li: *Ann. Stat.*, 1987, *15*, pp. 958-75.

1988Bin: K. Binder and D.W. Heermann: *Monte Carlo Simulation in Statistical Physics*. Springer-Verlag, New York, 1988.

1988Mor: V.L. Moruzzi, J.F. Janak, and K. Schwarz: *Phys. Rev. B*, 1988, *37*, pp. 790-99.

1988Qte: A. Qteish and R. Resta: *Phys. Rev. B*, 1988, *37*, pp. 6983-990.

1990Ced: G. Ceder, M. Asta, W.C. Carter, M. Sluiter, M.E. Mann, M. Kraitchman, and D. de Fontaine: *Phys. Rev. B*, 1990, *41*, pp. 8698-9701.

1990Van: D. Vanderbilt: *Phys. Rev. B*, 1990, *41*, pp. 7892-895.

1991Gir: S. de Gironcoli and P. Giannozzi: *Phys. Rev. Lett.*, 1991, *66*, pp. 2116-119.

1991Duc: F. Ducastelle: *Order and Phase Stability in Alloys*, Elsevier Science, New York, 1991.

1991Fer: L.G. Ferreira, S.-H. Wei, and A. Zunger: *Int. J. Supercomput.*, 1991, *5*, pp. 34-55.

1991Gol: A.S. Goldberger: *A Course in Econometrics*, Harvard University Press, Cambridge, MA, 1991.

1992Lak: D.B. Laks, L.G. Ferreira, S. Froyen, and A. Zunger: *Phys. Rev. B*, 1992, *46*, pp. 12587-2605.

1993Ast: M. Asta, D. de Fontaine, and M. van Schilfgaarde: *J. Mater. Res.*, 1993, *8*, pp. 2554-568.

1993Ced: G. Ceder: *Comp. Mater. Sci.*, 1993, *1*, pp. 144-49.

1994Ced: G. Ceder, G.D. Garbulsky, D. Avis, and K. Fukuda: *Phys. Rev. B*, 1994, *49*, pp. 1-7.

1994Fon: D. de Fontaine: *Solid State Phys.*, 1994, *47*, pp. 33-176.

1994Gar: G.D. Garbulsky and G. Ceder: *Phys. Rev. B*, 1994, *49*, pp. 6327-330.

1994Zun: A. Zunger: First Principles Statistical Mechanics of Semiconductor Alloys and Intermetallic Compounds, in *NATO ASI on Statics and Dynamics of Alloy Phase Transformation*, Vol. 319, P.E. Turchi and A. Gonis, ed., Plenum Press, New York, 1994, pp. 361-93.

1995Gar: G.D. Garbulksy and G. Ceder: *Phys. Rev. B*, 1995, *51*, pp. 67-72.

1995Wol: C. Wolverton and A. Zunger: *Phys. Rev. B*, 1995, *52*, pp. 8813-828.

1996Gar: G.D. Garbulsky and G. Ceder: *Phys. Rev. B*, 1996, *53*, pp. 8993-9001.

1996Kre1: G. Kresse and J. Furthmüller: *Comp. Mater. Sci.*, 1996, *6*, pp. 15-50.

1996Kre2: G. Kresse and J. Furthmüller: *Phys. Rev. B*, 1996, *54*, pp. 11169-1186.

1996McC: R. McCormack and D. de Fontaine: *Phys. Rev. B*, 1996, *54*, pp. 9746-755.

1996Tep: P.D. Tepesch, A.F. Kohan, G.D. Garbulsky, and G. Ceder, C. Coley, H.T. Stokes, L.L. Boyer, M.J. Mehl, B.P. Burton, R.J. Cho, and J. Joannopoulos: *J. Am. Ceram.*, 1996, *49*, pp. 2033-40.

1998Koh: A.F. Kohan, P.D. Tepesch, G. Ceder, and C. Wolverton: *Comp. Mater. Sci.*, 1998, *9*, pp. 389-96.

1998Ozo1: V. Ozoliņš, C. Wolverton, and A. Zunger: *Phys. Rev. B*, 1998, *57*, pp. 6427-443.

1998Ozo2: V. Ozoliņš, C. Wolverton, and A. Zunger: *Phys. Rev. B*, 1998, *58*, pp. R5897-R5900.

1998Ozo3: V. Ozoliņš, C. Wolverton, and Alex Zunger: *Phys. Rev. B*, 1998, *57*, pp. 4816-828.

1998Ven: A. van der Ven, M.K. Aydinol, G. Ceder, G. Kresse, and J. Hafner: *Phys. Rev. B*, 1998, *58*, pp. 2975-987.

1998Wal: A. van de Walle, G. Ceder, and U.V. Waghmare: *Phys. Rev. Lett.*, 1998, *80*, pp. 4911-914.

1998Wol: C. Wolverton, V. Ozoliņš, and A. Zunger: *Phys. Rev. B*, 1998, *57*, pp. 4332-348.

2000Wal: A. van de Walle and G. Ceder: *Phys. Rev. B*, 2000, *61*, pp. 5972-978.

2001Wal: A. van de Walle: The MIT Ab initio Phase Stability (MAPS) Code, http://www.mit.edu/navdw/maps.

2002Wal: A. van de Walle and M. Asta: *Model. Simul. Mater. Sci.* 2002, *10*, in press.