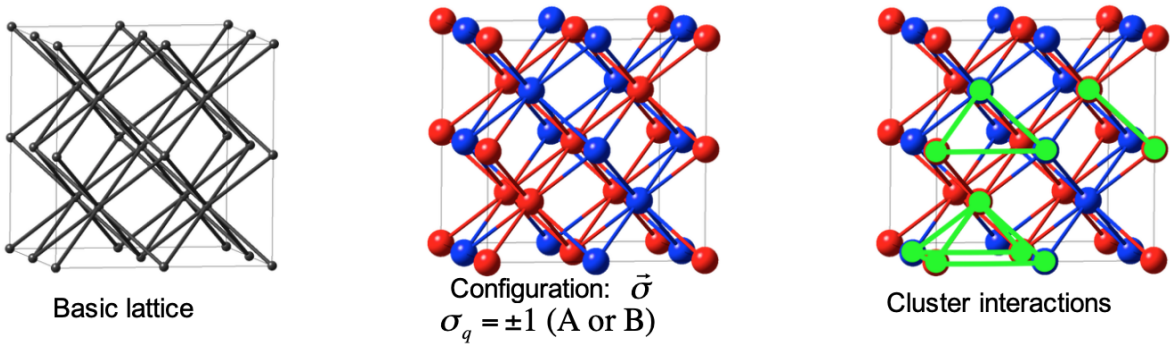
Cluster expansion theory

Contents

Cluster expansion [1], [2], [3] is a method describing the energy (or a similar scalar property) of a system as a function of occupation variables for each lattice position. On such a lattice the atom configuration, that is the distribution of the atomic species (including vacancies), is varied and the energies of the resulting configurations are swiftly predicted.

From an optimized cluster expansion, a set of effective cluster interactions can be extracted and used in large-scale Monte Carlo simulations to explore order-disorder phenomena and phase segregation processes as a function of temperature.

On a basic lattice various atoms, for example of type *A* and *B*, are distributed to define structure $\overrightarrow{\sigma}$, a periodic configuration of *A* and *B* atoms. This configuration is described by the pseudo spin operator $\sigma_q = \pm 1$, which has the value +1 if atom *A* sits on site q or -1 if that atom is *B*.



energy $E(\overrightarrow{\sigma})$ associated with structure $\overrightarrow{\sigma}$ can be described by an expansion of cluster interactions and their respective interaction energies J using equation:

$$E(\overline{\sigma}) = J_0 + J_1 \sum_i \sigma_i + \sum_{i>j} J_{ij}\sigma_i\sigma_j + \sum_{i>j>k} J_{ijk}\sigma_i\sigma_j\sigma_k + \ \ldots \tag{1}$$

In this equation, J_0, the first term, describes a constant, configuration independent contribution. The second term is concentration-dependent and is a sum over all N sites of structure $\overrightarrow{\sigma}$ with onsite energy J_i times the pseudo spin operator σ_i at each site i. Further terms describe the cluster interactions between multiple sites, for example, two-body interactions J_{ij} or three-body interaction J_{ijk}. They contain spin products $\sigma_i\sigma_j\ldots$ overall $\overrightarrow{f}$ vertices of a cluster times its effective cluster interaction energy J_{ij_o} summed up over all the possible ways that the cluster can be placed on the lattice of structure $\overrightarrow{\sigma}$.

[1] J. Sanchez, F. Ducastelle, and D. Gratias, “Generalized cluster description of multicomponent systems”, Physica A: Statistical Mechanics and its Applications 128, no. 1-2 (November 1984): 334-350

[2] D Lerch, O Wieckhorst, G L W Hart, R W Forcade, and S Müller, “UNCLE: a Code for Constructing Cluster Expansions for Arbitrary Lattices with Minimal User-Input”, Modelling and Simulation in Materials Science and Engineering 17, no. 5 (June 4, 2009): 055003.

[3] Stefan Müller, “Bulk and Surface Ordering Phenomena in Binary Metal Alloys”, Journal of Physics: Condensed Matter 15, no. 34 (August 15, 2003): R1429-R1500.

In other words, the energy $E(\overrightarrow{\sigma})$ of structure $\overrightarrow{\sigma}$ is broken down into clusters with their associated effective interaction energies. The core issue of cluster expansion is to identify a universal set of interactions J best-suited to describe a given model.

To accomplish this it is useful to reformulate the above equation into the more compact form

$$E(\overrightarrow{\sigma}) = \sum_{C \varepsilon \overrightarrow{C}} J_C \prod_C (\overrightarrow{\sigma}) \tag{2}$$

The cluster expansion equation sums up the product of cluster C's interaction energy J_C with its correlation function,

$$\prod_C (\overrightarrow{\sigma}) = N^{-1} \sum_{i=1}^{N} \sum_{k \in C} \prod_{\nu \in f} \sigma_\nu \tag{3}$$

a sum over all the possible ways a cluster C with $\overrightarrow{f}$ vertices can be placed on the N sites of the structure. In the correlation function the spin product $\sigma_1 ... \sigma_f$ goes over all $\overrightarrow{f}$ vertices of the cluster. Only symmetry inequivalent clusters are now considered and clusters included in an expansion can be collected by the vector $\overrightarrow{C} = \{C_1, ..., C_n\}$.

1 Inversion and Least Square Fitting

In order to find the solution for the cluster expansion, that is to identify appropriate effective cluster interactions, a training set of structures is needed.

Such a training set contains m structures $\{\overrightarrow{\sigma}_1, ..., \overrightarrow{\sigma}_m\}$ with the corresponding energies $\overrightarrow{E} = (E_1, ..., E_m)^T$. Together with the cluster vector $\overrightarrow{C} = (C_1, ..., C_n)^T$, containing all n inequivalent clusters, and the corresponding effective cluster interactions $\overrightarrow{J} = (J_1, ..., J_n)^T$ the $m \times n$ correlation matrix

$$\overline{\prod} = \begin{pmatrix} \prod_{C_1}(\overrightarrow{\sigma}_1) & \cdots & \prod_{C_n}(\overrightarrow{\sigma}_1) \\ \vdots & \ddots & \vdots \\ \prod_{C_1}(\overrightarrow{\sigma}_m) & \cdots & \prod_{C_n}(\overrightarrow{\sigma}_m) \end{pmatrix} \tag{4}$$

is defined.

Using this correlation matrix the cluster expansion problem can be formulated as

$$\overrightarrow{E} = \overline{\prod} \overrightarrow{J} \tag{5}$$

A straightforward way to define the values of the effective cluster interactions $\overrightarrow{J}$ is inversion

$$\overrightarrow{J} = \overline{\prod}^{-1} \overrightarrow{E} \tag{6}$$

However, this approach has serious drawbacks. The number of cluster n has to equal the number of structures m in the training set. Therefore, such an approach comes without redundancy. If the energy of a new structure is evaluated, a new cluster has to be added to the cluster expansion. This makes an efficient ground-state search or a Monte Carlo simulation unfeasible. Another issue is that the values of the energies of the training set structures usually have an uncertainty attached to them, they are noisy. A perfect fit with inversion will also fit the effective cluster interactions to that noise and not to actual physics.

A better approach to identify the effective cluster interactions is to use a least-square fit on a larger training set with *m* structures in the set and $m > n$:

$$\left(\overrightarrow{E} - \overline{\prod} \overrightarrow{J}\right)^T \left(\overrightarrow{E} - \overline{\prod} \overrightarrow{J}\right) \rightarrow min \tag{7}$$

2 Truncating Structures and Clusters

Employing a least-square fit, an approach has been identified by which n effective cluster interactions $\overrightarrow{J}$ are obtained from a training set containing m structures. However, the core issue remains unanswered. How to select the optimum training set and the optimum clusters for a given model from an infinite number of possibilities?

At first, before selecting optimum structures and clusters, a pool of possible candidates to be considered by the cluster expansion needs to be identified. For such an "enumeration" of all possible structures of a given model, *MedeA* UNCLE uses an integer-based algorithm based on the Hermite normal form and the Smith normal form [4], [5]. The algorithm scales linearly with the number of structures and enumerates structures up to a user-defined size. The maximum size of a structure is limited by a maximum number of unit cells it may contain. The configuration space scales by N^k, with k defining the number of atomic species and N the number of sites (that is the number of unit cells times the number of active sites in a unit cell). Therefore, a large value of N can result in too many structures to be handled efficiently within the cluster expansion.

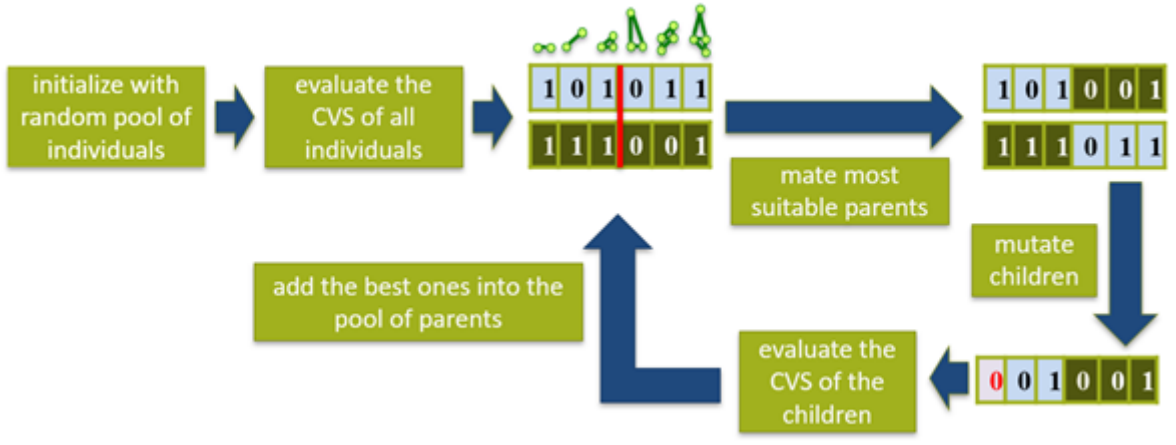
Similar to structure enumeration all possible clusters on the lattice of the model are enumerated up to hard-coded cut off values. These include 2-body, 3-body, 4-body, 5-body, and 6-body interactions.

Once the pool of possible structures and clusters has been defined an optimum set can be selected from that pool. To find the best cluster interactions for a pre-defined training set genetic algorithm is used.

3 Selecting Cluster Interactions Using the Genetic Algorithm

Cluster selection is performed using the genetic algorithm on a single training set $\{\overrightarrow{\sigma}_1, ..., \overrightarrow{\sigma}_m\}$ of structures with the corresponding energy values $\overrightarrow{E} = (\overrightarrow{E}_1, ..., \overrightarrow{E}_m)^T$. The process is stochastic in nature meaning that an identical training set yields slightly different results if the genetic algorithm is repeated. However, if the cluster expansion has converged such fluctuations are negligible.

As a fitness criterion the leave one out cross validation score (CVS) is used. It is evaluated by removing each structure once from the training set and using the remaining $N - 1$ structures in a least-square fit to determine $\overrightarrow{J}$. The thus obtained effective cluster interaction energies $\overrightarrow{J}$ are used to predict the energy of the one structure removed from the training set which is then compared with its actual value. This is done for all N structures in the training set and evaluated as a standard deviation of the predicted energies from the actual energies.



[4] Gus L. W. Hart and Rodney W. Forcade, "Algorithm for generating derivative structures", Physical Review B 77 (June 26, 2008): 224115.

[5] Gus L. W. Hart and Rodney W. Forcade, "Generating derivative structures from multi-lattices: Algorithm and application to hcp alloys", Physical Review B 80 (July 31, 2009): 014120.

4 Iterative Optimization of the Training Set

The iterative procedure by which the structures for the training set are selected differs if the model has a miscibility gap or miscible constituents. To distinguish these two types of models, the heats of formation $\Delta H_f(\overrightarrow{\sigma})$ are evaluated for all structures in the training set. It is defined by

$$\Delta H_f(\overrightarrow{\sigma}) = \frac{E_{DFT}(\overrightarrow{\sigma}) - \sum n_i(\overrightarrow{\sigma}) E^i_{DFT}}{\sum n_i(\overrightarrow{\sigma})} \tag{8}$$

wherein $E_{DFT}(\overrightarrow{\sigma})$ describes the DFT total energy of structure $\overrightarrow{\sigma}$, $n_i(\overrightarrow{\sigma})$ is the number of atoms of atomic species i contained in $\overrightarrow{\sigma}$, and E^i_{DFT} denotes the DFT total energy of the pure phase of atomic species i. The sums go over all type of atoms contained in structure $\overrightarrow{\sigma}$.

Models with miscible constituents have structures with negative ΔH_f, (thermodynamically stable, ordered structures) while models with a positive ΔH_f, where none of the ordered structures is thermodynamically stable and phase separation occurs, have a miscibility gap.

MedeA UNCLE initializes the iterative procedure by adding a user-defined number of independent and identically distributed structures (given by the parameter Number of structures to initialize the first iteration in the configuration panel) to the training set. As a default behavior, after the first iteration the type of model is automatically identified and the procedure switches to the suitable miscibility mode.

4.1 Miscible constituents

If a model has miscible constituents, the structures with energies close to the ground states, that is those structures with the lowest ΔH_f at a given concentration (see below figure), are the most important ones and the cluster expansion should be most accurate for those. To accomplish this, those structures predicted by the cluster expansion to be more favorable (with a lower ΔH_f) and are not yet part of the training set are added to the training set. This is done iteratively until no new structures are predicted by cluster expansion to be more favorable than those alcreated included in the training set. At this point, the cluster expansion has converged and from all structures considered by the cluster expansion the thermodynamically stable ones have been identified.

4.2 Miscibility Gap

If the model is phase separating, no stable ordered structures exist apart from the two pure phases and all structures are of equal importance to the cluster expansion. Therefore, the selection process of structures to be added to the training set has to improve the quality of the cluster expansion for all structures considered, irrespective of their formation energies ΔH_f.

To determine how good (or bad) the energies of the structures are predicted by the cluster expansion the stochastic nature of the genetic algorithm is used. Multiple cluster expansions are performed using an identical training set. The energies of all considered structures are then predicted by these multiple J's and a standard deviation of the predicted energies is evaluated. Structures with the highest standard deviation are those whose description by the cluster expansion is the worst. Therefore, these are added iteratively to the training set.