

MedeA UNCLE: Explore Phase Stability and Bridge the Length Scales with Universal Cluster Expansion

Contents

- [System Setup](#)
- [MedeA UNCLE Flowchart](#)
- [Optimize Cluster Expansion Stage](#)
- [Ground State Search Stage](#)
- [Random Mixing Stage](#)
- [Monte Carlo Simulation Stage](#)

The *MedeA* **UN**iversal **CL**uster **E**xpansion (UNCLE) package allows you to set up, construct and automatically converge a cluster expansion for bulk or surface systems with a partial disorder on some or all sublattices. Accurate (ab-initio) calculations of a relatively small number of supercells are the basis on which the effective cluster interactions are determined. *MedeA* UNCLE can be applied to arbitrary lattice types and multi-component systems. It automatically performs ground-state searches and applies genetic algorithms or compressive sensing to explore configuration space. The as such derived effective cluster interactions map the ab-initio results onto a much simpler Hamiltonian which allows large-scale Monte Carlo simulations on systems with up to millions of atoms. Refer to the section on Cluster Expansion Theory for more information.

The UNCLE software was developed by Stefan Müller, currently at the Technical University of Hamburg, and Gus Hart at the Brigham-Young University in Salt Lake City, Utah, and their co-workers. Users of the *MedeA* UNCLE package are required to acknowledge the authorship of the UNCLE Software in any articles they submit for publication by citing: [1]

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.

This paper outlines the algorithms implemented in the UNCLE software. Furthermore, a review article of Stefan Müller on the cluster expansion technique and its applications is recommended. [2]

Articles using compressive sensing should cite: [3], [4]

1 System Setup

The main structural input information for *MedeA* UNCLE, i.e. on which positions multiple specified elements are to be considered for an occupational variation, is conveyed by the system. The cluster expansion is carried out on all sites of a system that are occupied by multiple elements or vacancies. The total number of

[1] 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.

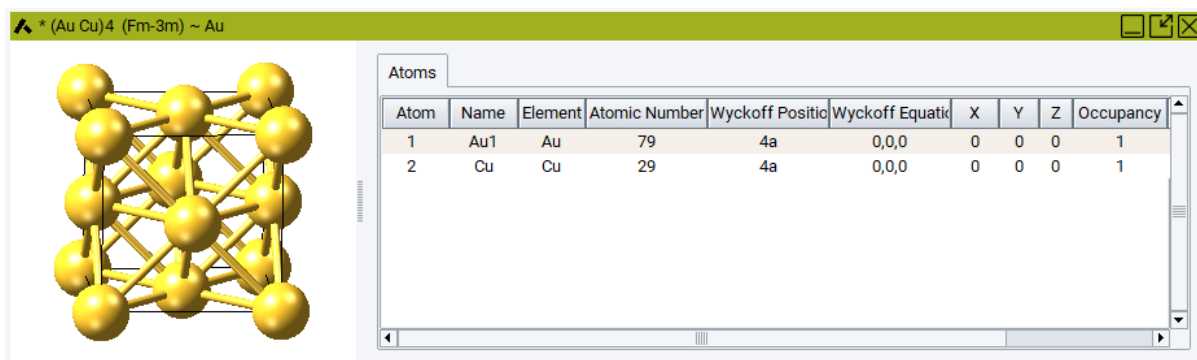
[2] Stefan Müller, "Bulk and Surface Ordering Phenomena in Binary Metal Alloys", *Journal of Physics: Condensed Matter* 15, no. 34 (August 15, 2003): R1429-R1500.

[3] Lance J Nelson, Vidvuds Ozolins, C Shane Reese, Fei Zhou, and Gus L W Hart, "Cluster Expansion Made Easy with Bayesian Compressive Sensing", *Physical Review B* 88, no. 15 (October 2013): 155105.

[4] Lance Nelson, Gus Hart, Fei Zhou, and Vidvuds Ozolins, "Compressive Sensing as a Paradigm for Building Physics Models", *Physical Review B* 87, no. 3 (January 18, 2013): 035125.

elements including vacancies occupying such sites defines the rank of the cluster expansion (binary, ternary, ...). If the occupation of any element is set to a value smaller than one, then the cluster expansion will be performed in a limited concentration range going from zero to a value that does not exceed the set occupation value. However, if the sum of all occupancies is less than one on any of the active sites *MedeA* UNCLE will add vacancies to the cluster expansion thereby increasing its rank by one. Summarizing, atoms of all elements involved including vacancies will populate all positions, on which multiple occupancies have been defined, even if not explicitly defined to occupy some of these positions.

The occupation of sites with multiple atom types can be set up by accessing the menu item **Edit** >> **Edit structure...** and selecting the **Add Atom** tab in the resulting Structure Editor window. Alternatively, right-clicking on an atom and selecting **Add Atom** from the context-sensitive menu provides access to the same panel. A good overview of the site occupation set up for the active system can be obtained from the molecular spreadsheet, accessible by the spreadsheet button at the top left of the *MedeA* window, or from the **Analysis** >> **Geometry Analysis** menu item. Through a right-click and selecting **Atom labels** >> **atomic symbol** the occupation can be studied in the Structure Viewer window.



The screenshot shows the 'Atoms' table in the Structure Editor window. The table lists the following data:

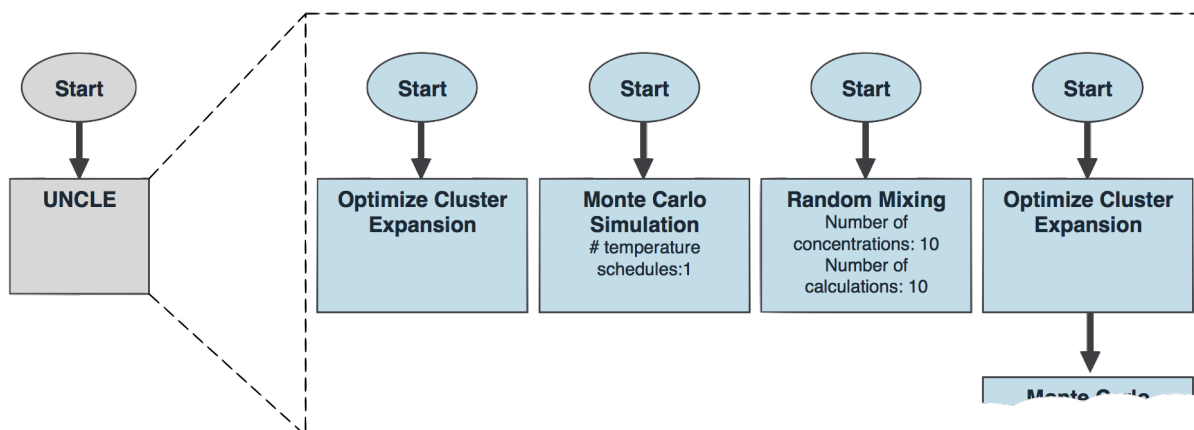
Atom	Name	Element	Atomic Number	Wyckoff Positio	Wyckoff Equatic	X	Y	Z	Occupancy
1	Au1	Au	79	4a	0,0,0	0	0	0	1
2	Cu	Cu	29	4a	0,0,0	0	0	0	1

Note: Configuration space and computational demands increase exponentially with the rank of the cluster expansion. In general, configuration space scales by N^k , with N for the maximum number of lattice sites and k for the cluster expansion rank.

2 MedeA UNCLE Flowchart

The *MedeA* UNCLE user interface is based on flowcharts. The UNCLE flowchart can be set up independently of an active system in *MedeA*. A blank flowchart is opened by selecting **Jobs** >> **New Job**, and can be edited to include an UNCLE stage by adding the UNCLE: Universal Cluster Expansion stage from the list of stages in the right pane. The blank UNCLE flowchart stage can then be edited by including further stages as discussed below. Alternatively, pre-prepared *MedeA* UNCLE flowcharts can be loaded by selecting **Open Library...**, **Open User...** or **From Job...** The flowchart library provides *ClusterExpansion_GGA.flow*, *ClusterExpansion_isotropic_GGA.flow*, or *ClusterExpansion_quick_GGA.flow* for optimizing cluster expansions by means of Genetic Algorithm as well as a flowchart to run Monte Carlo simulations *MonteCarlo_Simulation.flow*.

A custom *MedeA* UNCLE flowchart can be constructed by combining stages, as shown in the examples:



3 Optimize Cluster Expansion Stage

The **New Optimization** stage performs a cluster expansion by automatic optimization of effective cluster interactions. Double-click on the **Optimize Cluster Expansion** stage to open the control panel.

3.1 The Optimize Cluster Expansion Control Panel

Type of Cluster Expansion switches between a **bulk** and a **surface** cluster expansion optimization. When set to **surface** *MedeA* UNCLE assumes that the symmetry is broken along the x-axis requiring that the surface slab is orientated perpendicular to the x-axis.

The **Maximum number of unit cells** field defines the maximum cell size considered by *MedeA* UNCLE, and thereby also the maximum cell size for geometry optimizations and energy calculations (e.g. by VASP). This parameter limits the size of the configuration space considered in the structure search, and strongly affects the computational demands. The number of atoms of the largest considered cell is the number of atoms of the initial system multiplied by this parameter.

The algorithm for fitting the cluster expansion can either be set to **Genetic Algorithm** or **Compressive Sensing**. The contents of the subsequent sections depend on this setting. With the option **Run the different structures simultaneously** the user can toggle if the physical property (e.g. energy) of multiple structures are evaluated at the same time or not.

Genetic Algorithm

An iterative procedure is applied to optimize the cluster expansion to find the best set of effective cluster interactions. The procedure initializes a set of independent and identically distributed structures in the training set. At each iteration, new structures are automatically added to it.

Due to the nature of such an iterative procedure a comparatively small number of DFT calculations is needed to define the optimum training set most suitable for the system.

Edit stage: Optimize Cluster Expansion

Controls

Type of Cluster Expansion:

Maximum number of unit cells:

Algorithm for fitting the Cluster Expansion:

Iterative Optimization Parameters

Maximum number of structures added in each iteration:

Number of structures to initialize the first iteration:

Maximum number of iterations for the optimization:

Miscibility of constituents:

Convergence criterion for cluster expansion optimization: meV/pos.

Number of cluster expansions used to assess quality:

Genetic Algorithm Parameters

Maximum number of clusters in a fit:

Training Set Flowchart

Run the different structures simultaneously

The iterative procedure is controlled by the settings provided by the section Iterative Optimization Parameters:

Maximum number of structures added in each iteration : Defines the maximum number of structures to be calculated at each iteration step and to be added to the training-set used by the cluster expansion.

Number of structures to initialize the first iteration : Defines the number of structures with which to initialize the cluster expansion. These structures are selected to be independent and identically distributed, i.e., to be the most distinctive structures within the configuration space limited by the maximum unit cell size.

Maximum number of iterations for the optimization : Even if the cluster expansion has not converged it will be terminated when reaching this value.

Miscibility of constituents : Possible options are, **Automatic mode** , **Miscible constituents** and **Miscibility gap** . If the **Automatic mode** is selected *MedeA* UNCLE will automatically switch, as required, between the miscible constituents and miscibility gap mode during optimization. The convergence criteria for **Miscible constituents** is a stable ground state line, i.e., that no new structures are predicted by the cluster expansion to be energetically more favorable (lower enthalpy/heat of formation) than the ones already included in the training set. This mode will only work for systems with miscible constituents. Addressing systems with **Miscibility gap** the convergence criteria is defined such that the standard deviation of 95% of all enumerated structures in the defined configuration space (limited by **Maximum number of unit cells**) using several separate CE's, with the identical training set, falls below the given value. The parameters controlling this miscibility gap mode is discussed by the next two entries.

Convergence criterion for cluster expansion optimization : This option is only available in the automatic or miscibility gap mode and is one of two parameters controlling the convergence criteria of the miscibility gap mode. If the standard deviation of 95% of the cluster expansion predictions of all structures within the considered configuration space falls below this energy value in meV per active position the cluster expansion is considered to have converged.

Number of cluster expansions used to assess quality : This is the second parameter that controls the convergence criteria of the miscibility gap mode. It defines the number of parallel cluster expansion with identical training set used to assess the standard deviation of the cluster expansion predictions in the considered configuration space.

The Genetic Algorithm is controlled by the entries of the Genetic Algorithm Parameters section:

Maximum number of clusters in a fit : This variable defines the maximum number of possible clusters.

Additional Settings under Advanced Settings

Genetic Algorithm Parameters

Maximum number of clusters in a fit:	20	
		Hide Advanced Settings
Number of generations (simulation time):	200	
Number of different populations running in parallel:	2	
Number of individuals in each population:	80	
Maximum number of generations an individual can live:	40	
Number of children per generation:	40	
Number of individuals replaced by children each generation:	20	
Accuracy of the fit for the pure elements:	0.0000001	eV/pos.
Cluster definition:	Custom	▼
Number of 2-body 3-body ... n-body clusters:		

Number of generations (simulation time) : Number of generations in which the Genetic Algorithm improves upon the cluster expansion using the current training set. A higher value can slightly increase the quality of a cluster expansion while also increasing the computational cost.

Number of different populations running in parallel : The genetic algorithm in *MedeA-UNCLE* allows for different “island” populations, the number of which is defined by this variable, to run isolated from each other. The population between these “islands” can only cross once a predefined diversity limit is exceeded otherwise the populations remain isolated.

Number of individuals in each population : This variable defines the number of individuals in each population. A larger value ensures a slightly better cluster expansion at a higher computational cost when using the same **Number of generations** .

Maximum number of generations an individual can live : Defines the maximum age an individual can achieve before dying. This value should be adapted when changing **Number of generations** .

Number of children per generation : Number of children in each population generated at each genetic algorithm generation. It is recommended that this value is about smaller or equal to half the value of **Number of individuals in each population** .

Number of individuals replaced by children each generation : Number of individuals in each population that are replaced by children at each iteration/generation. A value smaller or equal than half the **Number of children per generation** is recommended.

Accuracy of the fit of pure elements : The accuracy in eV per active position with which the pure elements are fitted. This value should be increased by orders of magnitude if UNCLE has problems optimizing a cluster expansion.

Cluster definition can be set to either **Automatic** or **Custom** . If set to **Automatic** default cluster definitions specific for the genetic algorithm are used, which depend also on the rank of the cluster expansion. If set to **Custom** the number of 2-body, 3-body ... n-body clusters, considered by the genetic algorithm, can be defined by the user. A field, **Number of 2-body 3-body ... n-body clusters:** , appears in which the user can enter the number of 2-body, 3-body ... n-body clusters. The entry for each cluster type is separated by a space. For example, *150 100 50 25*, defines a cluster set with 150 2-body clusters, 100 3-body clusters, 50 4-body clusters, and 25 5-body clusters. In automatic mode, the current default values are *50 100 50 20 20* up to ternary, and *100 50 20* for quaternary and higher ranked cluster expansions.

Note: MedeA UNCLE only includes clusters up to 4-body when optimizing a quaternary or higher ranked cluster expansion.

Compressive Sensing

Compressive sensing is non-iterative and is applied to a single set of structures. It is a novel method and was originally developed for signal processing. The method is well suited to find a set of effective cluster interactions universally applicable to the entire configuration space. However, this method requires a much larger training set than the genetic algorithm, and all structures in the set have to be independent and identically distributed. A holdout set is needed to assess the quality of a compressive sensing optimization. See Refs. [5], [6] for further information on this method.

Controls

Type of Cluster Expansion:

Maximum number of unit cells:

Algorithm for fitting the Cluster Expansion:

Compressive Sensing Parameters

Number of structures used for fitting the cluster expansion:	<input type="text" value="200"/>
Number of holdout structures used to verify the fit:	<input type="text" value="30"/>
Number of structures in a subset:	<input type="text" value="195"/>
Number of subsets:	<input type="text" value="100"/>

[Show Advanced Settings](#)

Number of structures used for fitting the Cluster Expansion : This parameter defines the number of independent and identically distributed structures used in the training set. A value above 200 and a comparatively high value for the variable **Maximum number of unit cells** are recommended for a good compressive sensing optimization.

Number of holdout structures used to verify the fit : To assess the quality of a compressive sensing optimization a holdout set containing structures excluded from the training set is used. The size of this holdout set are defined by this parameter.

Number of structures in a subset : The training set is subdivided into subsets each containing as many structures as defined by this variable. The specific structures within these subsets are selected randomly.

Number of subsets : How many subsets should be created from the training set? Each subset contains a number of randomly selected structures from the training set as defined by **Number of structures in a subset** .

Additional Settings under **Advanced Settings**

[5] Lance J Nelson, Vidvuds Ozolins, C Shane Reese, Fei Zhou, and Gus L W Hart, "Cluster Expansion Made Easy with Bayesian Compressive Sensing," Physical Review B 88, no. 15 (October 2013): 155105.

[6] Lance Nelson, Gus Hart, Fei Zhou, and Vidvuds Ozolins, "Compressive Sensing as a Paradigm for Building Physics Models," Physical Review B 87, no. 3 (January 18, 2013): 035125.

Compressive Sensing Parameters

Number of structures used for fitting the cluster expansion:

Number of holdout structures used to verify the fit:

Number of structures in a subset:

Number of subsets:

[Hide Advanced Settings](#)

Type of penalty function: ▼

Cluster definition: ▼

Type of penalty function : Different type of penalty functions can be set, namely `logsum` , `arctan` , `quartic` , `hexic` , `octic` or `logsig` . The default is not to use a penalty function.

Cluster definition can be set to either `Automatic` or `Custom` . If set to `Automatic` default cluster definitions specific for the compressive sensing algorithm is used, which depend also on the rank of the cluster expansion. If set to `Custom` the number of 2-body, 3-body ... n-body clusters, considered by compressive sensing, is defined by the user. A field, `Number of 2-body 3-body ... n-body clusters:` , appears in which the user can enter the number of 2-body, 3-body ... n-body clusters. The entry for each cluster type is separated by a space. For example, `150 100 50 25`, defines a cluster set with 150 2-body clusters, 100 3-body clusters, 50 4-body clusters and 25 5-body clusters. In automatic mode, the current default values are `500 500 500 500 500` up to ternary, and `500 500 500` for quaternary and higher ranked cluster expansions.

Note: The iterative genetic algorithm approach and the process of deriving independent and identically distributed structures in compressive sensing is stochastic. As a consequence, two identically set up calculations may involve different structures for deriving the effective cluster interactions, in general. However, the stochastic uncertainties should be smaller than the confidence level of the obtained results, e.g. the obtained ground state line should be identical for the same system and parameters.

3.2 The Cluster Expansion Optimization Flowchart

To access the training set flowchart click the `Flowchart` button on the bottom of the `Optimize Cluster Expansion` control panel. This flowchart defines the workflow on which the cluster expansion operates on to establish the physical properties of the training-set structures.

Currently, the choice for the variable is the electronic energy as calculated in the `VASP 5.4 High Throughput` , `VASP 6: High Throughput` , `VASP 5.4: Plane Wave DFT` , `VASP 6: Plane Wave DFT` , or in the `LAMMPS: Molecular Dynamics and Statics` stage.

Note: Both type of VASP stages, i.e., the `High Throughput` and the `Plane Wave DFT` , can be used with UNCLE. The `High Throughput` type stages do not support many of the post-DFT methods that can be accessed from `Plane Wave DFT` type ones, however, it does not require that the PAW potentials of all the elements considered in the cluster expansion is initialized before running the calculation as is required in the `Plane Wave DFT` type VASP stages.

Add `UNCLE: Save Property` after the VASP or LAMMPS stage(s). The `Save Property` stage defines the physical variable to be extracted from the previous stages, which is currently the electronic energy obtained from VASP or the energy obtained from LAMMPS.

The VASP High Throughput Flowchart for the Electronic Energy

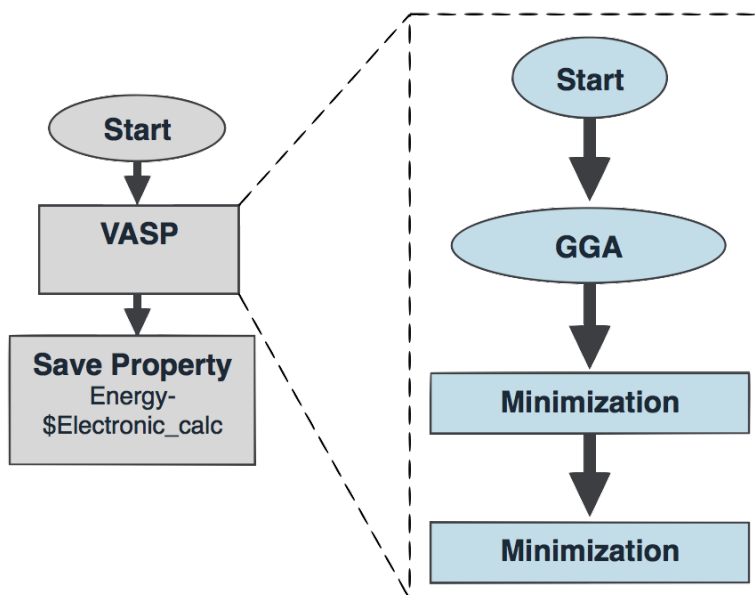
Double-clicking on the **VASP** stage opens the VASP flowchart controlling the VASP calculations.

Depending on which method is the best for your system either add **Initialize VASP for LDA** or **Initialize VASP for GGA**. Double-click on the **LDA / GGA** stage to adapt the **Precision** and **Cutoff** parameters accordingly.

It is recommended to increase the cutoff manually by at least 30% since full geometry optimizations including cell parameters are involved (see below). A safe option to achieve suitable accuracy is to set **Precision** to **Standard 500**. If desired modify any other parameters, e.g., potentials, type of systems, etc.

Now add stages to calculate DFT total energy of the structures. Use **Structure Optimization** stages to optimize the crystal structures, i.e. full optimization of the cell and all atom positions, and or a **Single Point** stage as a final stage.

Cluster Expansion Optimization flowchart



Note: For accurate cell shape optimization use two separate steps to enable adaption of the plane-wave basis to the optimized cell parameters. To reduce the number of unnecessary relaxation steps it is recommended to reduce the value **Maximum number of steps** to a low value, for example, 20, in all the structure optimization steps stages.

When configuring a VASP flowchart for your system test it first for a smaller configuration space by lowering the maximum number of unit cells.

If the system is a metal and accurate energy values are required, the last VASP stage should have a k-mesh spacing below **0.3** 1/Ång.

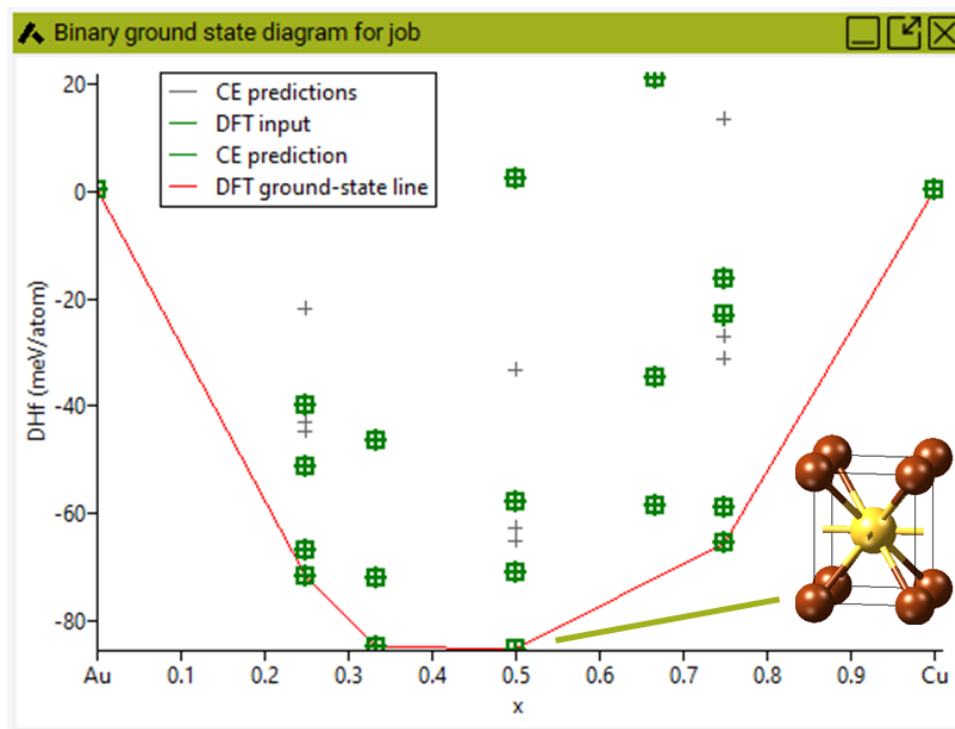
3.3 Results

Interactive Binary Ground State Diagram

The binary ground state diagrams of a cluster expansion are accessible in *MedeA* by clicking on **Analysis** >> **UNCLE Binary Ground State Diagram**. Select the desired diagrams and click on **OK**. This opens zoom-able binary ground state diagrams. You can zoom into a part of the diagram by holding the left mouse

button and selecting a region. Un-zoom by right-clicking and selecting **Unzoom**. Any DFT input structures plotted within the figure, marked by green rectangles, or the CE predictions of these structures, marked by green crosses, can be visualized by right-clicking and selecting **View ceXX.sci**. This opens that structure in a new window allowing you to study its physical properties by further calculations.

Note: For systems with miscible constituents a converged cluster expansion has no predictions below the DFT based ground state line.



Job.out

A summary of the cluster expansion parameters can be found in the first section of the *Job.out* file on the JobServer.

Stage 1.1: This calculation optimizing a bulk Cluster Expansion, running different structures simultaneously, has 2 stages:

For exploring the configuration space, 4 unit cells are taken into account. Effective cluster interactions are fitted using a Genetic Algorithm. This fitting scheme will run for a maximum of 20 iterations, adding a maximum of 5 structures in each iteration, and starting from an initial training set of 5 structures. The iterations will continue until the energies of all structures as predicted by the cluster expansion are higher than the energy calculated for the structure of the ground state line at each sampled concentration.

If it turns out that the constituents are not miscible and all energies are higher than for the pure element structures, a different scheme is pursued: 5 different cluster expansion fits are performed, and the standard deviation of the energy predictions is evaluated. Iterations are continued until the standard deviation of 95% of the structures is within 5 meV per active atom position.

The parameters for the Genetic Algorithm are:

```

Maximum number of generations (simulation time): 200
Maximum number of clusters in a fit: 20
Number of different populations running in parallel: 2
Number of individuals in each population: 80
Maximum number of generations an individual can live: 40
Number of children per generation: 40
Number of individuals replaced by children each generation: 20
Accuracy of the fit for the pure elements: 0.0000001
  
```

The description of the cluster expansion parameters is followed by an overview of the flowchart configuration (not shown).

You will also find a summary on the iterative optimization progress (when using the genetic algorithm method) informing you on the number of structures used in the training set at each iteration, the number of new structures added, the cross-validation score CVS and the percentage of structures with a standard deviation below the value defined by the `Convergence criterion for Cluster Expansion optimization`. The second last column (% struc. with SD below 5 meV) will only be filled when the genetic algorithm optimization uses the miscibility gap procedure. The cross-validation score itself indicates how well the observables (e.g. the energies) of the structures in the training set compare with each other. However, systematic errors will not show up. The last column lists the name of the structures added at each iteration to the training set.

Note: An indication of a good cluster expansion is usually a cross-validation score (CVS) value below 5 meV per active position.

It.	no. struc.	no. new struc.	CVS [meV/pos.]	% str. with SD below 5 meV	new struc.
0	0	2	-	-	ce1 ce2
0	0	2	-	-	ce20 ce18
1	2	2	20	-	ce21 ce8
2	4	2	0.14	-	ce24 ce10
3	5	1	0.071	-	ce13
4	6	1	0.35	-	ce7
5	8	2	0.26	-	ce22 ce9
6	9	1	0.069	-	ce4
7	10	1	0.38	-	ce14
8	11	1	0.25	-	ce29
9	12	0	0.059	-	

The Cluster Expansion optimization finished successfully!

A section provides an overview of the fitting errors of the final cluster expansion follows this summary. It can be found for the other iterations in the file `fittingErrors.1.out` located in the appropriate directories `Iteration_` depending on the flowchart configuration. The overview lists for each structure in the training set the concentrations of all elements/vacancies $x(1:k, \dots)$, the DFT input value E_{DFT} in eV per active position, the prediction error $D(E_{DFT}, E_{CE})$ which is the difference between DFT input and cluster expansion prediction, the cluster expansion prediction value E_{CE} , the heat of formation of the DFT input DHf_{DFT} and the CE prediction DHf_{CE} , respectively, the distance of the structure from the ground state line GSL and the title of the structure. Comparatively large prediction errors, $D(E_{DFT}, E_{CE})$, might indicate problems with that structure (e.g., when relaxing the structure into an incompatible lattice).

At the end of the `Job.out` file you will find a summary of the stable ground-state structures listed with their respective concentrations, the heats of formation in eV per active position, structure title, cell formula and space group.

The structures constituting the DFT ground state line, their compositions x and formation energies DHf are:

```

# x(1:k,1:nC) | DHf / atom [eV] | structure title | cell formula | space group
#-----
  
```

0.0000000000	1.0000000000	0.0000000000	ce2	Cu	Fm-3m
0.6666666667	0.3333333333	-0.0484193333	ce7	Au2Cu	I4/mmm
0.7500000000	0.2500000000	-0.0467052500	ce14	Au3Cu	P4/mmm
1.0000000000	0.0000000000	0.0000000000	ce1	Au	Fm-3m

The above table is followed by a summary of the effective cluster interactions.

Other output files

Text-based summaries can be found, depending on your flowchart, in directories `*/Iteration_*`, i.e. fitting errors in files `fittingErrors.1.out`, ground state lines from DFT input and cluster expansion predictions in files `gsl.dft.out` and `gsl.ce.out`, respectively, and effective cluster interactions in files `J.1.summary.out`. In addition to these also graphical depictions of the results are given in the files `BinaryGroundstateDiagram.png`, `FittingErrors.png` and `Jsummary.png`. Click on [as plain text](#) to open these png files.

For binary cluster expansions, `BinaryGroundstateDiagram.png` plots heat of formation (DHf) in meV/atom against the concentration of the first element. The following data sets are distinguished in this plot: the cluster expansion predictions from `gss.out`, the DFT input values from `fittingErrors.1.out` and the DFT based ground state line from `gsl.dft.out`. In systems with miscible constituents a converged cluster expansion has no predictions below the DFT based ground state line. As with the text-based summaries the plots can be found for each iteration in the directories `*/Iteration_*`.

An overview on the difference of the cluster expansion predictions to the DFT input values is given in `FittingErrors.png` which plots $D(E_{DFT}, E_{CE})$ from `fittingErrors.1.out` in meV/atom. Comparatively large values can be indicative of erroneous DFT calculations. Again, the plot can be found for each iteration step in `*/Iteration_*`.

The optimized effective cluster interactions are summarized in `Jsummary.png` in `*/Iteration_*`, plotting the effective cluster interaction values (meV) against weighted vertices distance of 2-body, 3-body, 4-body, 5-body, and 6-body interactions. Positive values of the two-body effective cluster interactions indicate ordering while negative values indicate phase separating tendencies

4 Ground State Search Stage

This stage can be used to calculate the energies of structures using a different configuration space than the one used to establish the effective cluster interactions. The configuration space can differ by the maximum number of unit cells or the concentration range as defined in the model. It will produce a binary ground state diagram in which it compares the predicted properties with those in the training set obtained by means of a previous [Optimize Cluster Expansion](#) stage.

Note: You can set up an UNCLE flowchart that only contains the [Ground State Search](#) stage or add the [Ground State Search](#) stage after a [New Optimization stage](#).

4.1 Configure the Ground State Search Stage

Double-click on [Ground State Search](#) to access its configuration panel.

Cluster Expansion

Apply cluster expansion from previous stage ▼

Apply cluster expansion of stage: last

Maximum number of unit cells: 6

Apply Cluster Expansion from previous stage or from previous job : if you have configured your flowchart with a New Optimization stage followed by a Ground State Search stage you may choose the first option, i.e. from previous stage . The entry last applies to the last applicable stage, which is appropriate for most cases. Otherwise select from previous job and click on This opens the window “Get files from previous calculation” where you can then select the desired *J.1.out* file from a job and subdirectory.

Define with Maximum number of unit cells the cell size used in the ground state search. This defines the maximum size of structures used for the configuration search. The number of atoms of the largest unit cell of the structures is the number of atoms in the primitive unit cell of the initial model multiplied by this parameter. For all of these structures, the property is predicted using the effective cluster interactions from the previous Optimize Cluster Expansion stage.

4.2 Results

Interactive Binary Ground State Diagram

The binary ground state diagram of the ground state search is accessible in *MedeA* by clicking on Analysis >> UNCLE Binary Ground State Diagram . Select the desired diagrams and click on OK . This opens zoomable binary ground state diagrams. You can zoom into a part of the diagram by holding the left mouse button and selecting a region. Unzoom by right-clicking and selecting Unzoom . Any DFT input structures plotted within the figure, marked by green rectangles, or the CE predictions of these structures, marked by green crosses, can be visualized by right-clicking and selecting View ceXX.sci . This opens that structure in a new window allowing you to study its physical properties by further calculations.

5 Random Mixing Stage

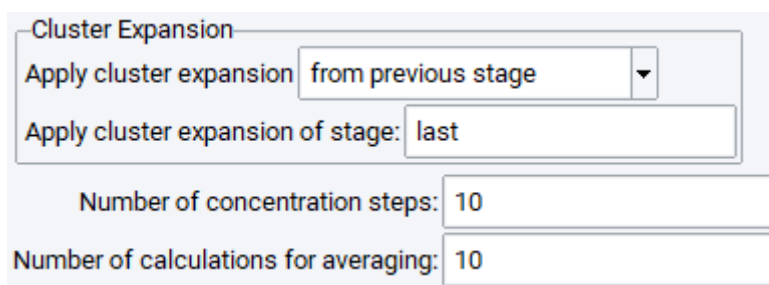
The random mixing energy describes a concentration dependent energy of the system at maximum entropy, i.e., the energy of totally disordered configurations.

Note: You can set up an independent UNCLE flowchart that only contains the Random Mixing stage or add the Random Mixing stage after a New Optimization stage.

5.1 Configure the Random Mixing Stage

Double-click on Random Mixing to access its configuration panel.

Apply Cluster Expansion from previous stage or from previous job : If you have configured your flowchart with a New Optimization stage followed by a Random Mixing stage you may choose the first option, i.e. from previous stage . The entry last applies to the last applicable stage, which is appropriate for most cases. Otherwise select from previous job and click on This opens the window “Get files from previous calculation” where you can then select the desired *J.1.out* file from a job and subdirectory.



Define the **Number of concentrations steps** between $x_A = 0$ and 1 and specify the quality of the statistical evaluation of the random mixing energy at each concentration step by **Number of calculations for averaging**.

5.2 Results

The top of the *Job.out* file summarizes the parameters used for the random mixing calculations. The calculated random mixing energy is provided by listing at each concentration step the average energy per atom (in eV) as well as the average heat of formation with their associated standard deviations.

#	c1	c2	avg. E [eV]	avg. Hf [meV]
#	0.0000	1.0000	-3.745132 +/- 0.000000	0.000000 +/- 0.000000
	0.0100	0.9900	-3.737488 +/- 0.000007	2.134452 +/- 0.006997
	0.0200	0.9800	-3.729908 +/- 0.000018	4.205184 +/- 0.017675
	0.0300	0.9700	-3.722395 +/- 0.000015	6.209066 +/- 0.014959
	0.0400	0.9600	-3.714942 +/- 0.000031	8.153518 +/- 0.030758
	0.0500	0.9500	-3.707574 +/- 0.000030	10.011950 +/- 0.029634

The file *RandomMixingDiagram.png* contains a figure plotting concentration c1 against the average heat of formation in meV/atom.

6 Monte Carlo Simulation Stage

Monte Carlo simulations based on cluster expansion can be used to analyze phase stabilities of systems by sampling the configuration space at given temperatures using a canonical ensemble.

Note: You can set up an independent UNCLE flowchart that only contains the **Monte Carlo** stage or add the **Monte Carlo** stage after a **New Optimization stage**.

Cluster Expansion

Apply cluster expansion: from previous stage ▼

Apply cluster expansion of stage: last

Simulation Cell

Supercell in x: 10 y: 10 z: 10

Concentrations:

Statistical Ensemble

Ensemble: canonical ▼

Temperature Schedules

1: Initial: 1000 Increment: +/- ▼ 50 Final: 500

Add schedule

Number of steps to average over: 10000

Convergence

Convergence criterion: Accuracy of the energy ▼

Accuracy of energy: 0.0001 eV

Run simulation serial ▼

Configure the Monte Carlo Simulation stage

Double-click on the **Monte Carlo Simulation** stage to access its configuration panel.

Apply Cluster Expansion **from previous stage** or **from previous job** : If you have configured your flowchart with a **New Optimization** stage followed by a **Monte Carlo** stage you may choose the first option, i.e. **from previous stage** . The entry **last** applies to the last applicable stage, which is appropriate for most cases. Otherwise select **from previous job** and click on **...** . This opens the window “Get files from previous calculation” where you can then select the desired *J.1.out* file from a job and subdirectory.

The size of the simulation cell can be configured by the entry fields **Supercell in x** : , **y** : and **z** : defining the x, y, and z multiples of the initial system’s primitive unit cell.

The concentrations of each species and vacancies can be configured by the entry field for **Concentrations** : Enter the fraction of each atomic species (and vacancy) separated by a blank. Note, if the fractions don’t sum up to one *MedeA* UNCLE will rescale them such that the sum equals one. The sequence of atomic species follows the sequence of the model. That is, the sequence shown in the spreadsheet of the *MedeA* structure viewer. The sequence is also shown when right-clicking on the model >> **Atom labels** >> **atomic symbol** .

The temperature schedules can be configured in the Temperature Schedules subpanel. For each temperature schedule define the **Initial** : , the **Increment** : and the **Final** : temperatures in Kelvin. Clicking on **+/-** at the **Increment** : opens a pop-up window that allows you to change the increment from a fixed additive/subtractive to a multiplicative one (***/:**). The incremental temperature change is applied several times until the final temperature is reached. Further schedules can be added by clicking on **Add** schedule, or removed by clicking on **Remove** schedule. Upon adding schedules, the final temperature of the last schedule becomes the initial temperature of the added schedule, and vice versa upon removing schedules.

Number of steps to average over : defines the number of Monte Carlo steps over which the energy is averaged.

Note: A good integer value for **Number of steps to average over** is the number of lattice-sites multiplied by 10, e.g., if you have a 10 x 10 x 10 supercell created using a unit cell containing two lattice-sites use a value of 10000.

The **Convergence criterion** can be defined either via **Accuracy of the energy** or via **Number of steps at each temperature**. If the **Accuracy of the energy** defines convergence, the energy to be set is the averaged energy over the number of steps as defined by **Number of steps to average over**. If both variables are too small the Monte Carlo calculation might not converge. However, if you set the value for **Accuracy of energy** too high the Monte Carlo simulation might stop before convergence is reached. If the **Number of steps at each temperature** defines convergence, again make sure that the **Number of steps** chosen is sufficient to reach convergence at each temperature value.

The Monte Carlo procedure can be run **serial** or **parallel** from the **Run simulation** pulldown menu. In **parallel** mode there is an additional selection, whether the simulation is intended as a **production run** or a **speed test only**, the latter enabling short test simulation to explore parallel performance. In **parallel** mode an additional subpanel appears which allows you to specify the **Dependency cell in x:**, **y:** and **z:** and the **Number of dependency lists**. Good parallel performance is expected for very large simulations cells running on a large number of processors.

Click on **OK** to close the configuration window and again on **OK** to close the UNCLE flowchart. Define the job name and click on **Run**.

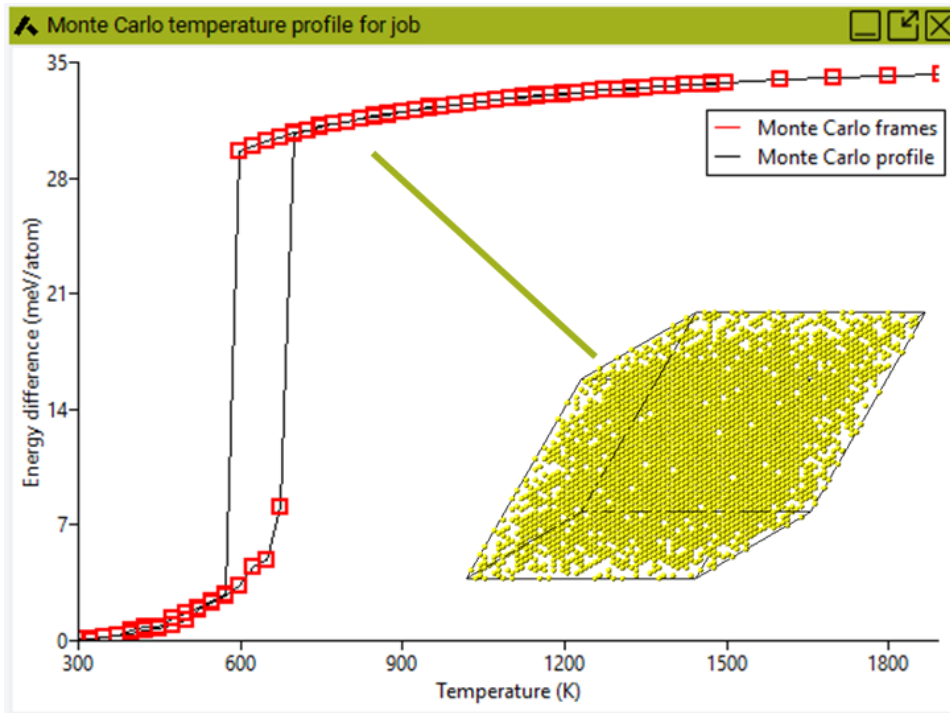
6.1 Results

Interactive Monte Carlo Temperature Profile

The interactive temperature profiles of a Monte Carlo simulation can be accessed in *MedeA* over **Analysis** >> **UNCLE Monte Carlo Temperature Profile**. In this temperature vs. energy difference plot, you can zoom in to a region of interest by holding the left mouse button and selecting that region. Unzoom by right-clicking and selecting in the pop-up window the option **Unzoom**. The red squares mark the temperature and energy at each temperature frame. When hovering with the mouse pointer over a frame the energy difference, the exact energy, temperature, and frame number are shown. By a **right-click** and selecting the option **View Monte Carlo frame XXX** the structure of that frame is shown in a new window.

Note: When viewing structure containing many atoms it is necessary to first modify the display quality in preferences (>> **File** >> **Preferences...** >> **Display Quality**) by increasing by a factor of 10 the number of atoms at which display quality is reduced.

Structures containing a large number of atoms can be manipulated by creating subsets for each atom type (>> **right-click** >> **Subsets** >> **Create**) which then can be deleted (>> **Subsets** >> **Delete atoms in subset**) or hidden (>> **right-click** >> **Subsets** >> **Select atoms in subset**). It is also possible to view a slice of the structure (>> **View** >> **Options...** >> **View** and modify the view limits).



Job.out and Other Output Files

At the top of the *Job.out* file the parameters used in the Monte Carlo simulation is stated. This is followed by a line by line summary of each temperature step with the associated energy per atom (in eV/atom), Monte Carlo steps, the number of rejected Monte Carlo attempts as well as the total concentrations of each element or vacancy.

#	T#	T / K	Energy	...	MC steps	rejected	total conc. (1:k,1:nC)
0	1500.0	-3.5188567E+00	...	40056	3651	0.5000	0.5000
1	1400.0	-3.5190442E+00	...	40056	3626	0.5000	0.5000
2	1300.0	-3.5212924E+00	...	40056	3799	0.5000	0.5000
3	1200.0	-3.5199532E+00	...	40056	3770	0.5000	0.5000
4	1100.0	-3.5230142E+00	...	40056	3963	0.5000	0.5000
5	1000.0	-3.5266042E+00	...	40056	4169	0.5000	0.5000

The file *MonteCarlo_TemperatureProfile.png* on the JobServer provides a simplified figure of the Monte Carlo temperature profile plotting the normalized energy curve against temperature.

Another file of interest is *MCoutput.out* which gives a more detailed view on the energies per averaging interval at each temperature, with $\langle E \rangle$ describing the average energy over the interval, $d\langle E \rangle$ the change of this average energy with regard to the previous interval and E the energy at the end of the interval.

MC steps	$\langle E \rangle$ /eV	$d\langle E \rangle$ /eV	E/eV	...
5007	-3.517550	-3.517550	-3.51975588957	...
10014	-3.520344	-0.002794	-3.51917705458	...