CE Screen: An energy-based structure screening automatic workflow

Zongguo Wang a, Xiaoyu Yang a,b,e, Ligen Wang c, Juan Wang a,b, Mingming Zhang a,b, Xushan Zhao a, Jie Ren a,b, Zhi Zeng d,e

a Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
b University of Chinese Academy of Sciences, Beijing 100049, China
c General Research Institute for Nonferrous Metals, Beijing 100088, China
d Key Laboratory of Materials Physics, Institute of Solid State Physics, Chinese Academy of Sciences, Hefei 230031, China
e University of Science and Technology of China, Hefei 230026, China

A R T I C L E   I N F O

Article history:
Received 23 June 2017
Received in revised form 25 September 2017
Accepted 20 October 2017

Keywords:
Cluster expansion
High-throughput screening
First principles calculation
MatCloud

A B S T R A C T

The high-throughput computational techniques coupled with the exponential growth of computational power can accelerate the simulation of material properties from first principles. Cluster expansion is a method that calculates the effective cluster interactions from partial first principles total energies of different supercells, and then predicts energies of all the structures derived from the same parent lattice. However, using cluster expansion requires people to understand the principles of cluster expansion. In this paper, centering on the well known cluster expansion method, we developed an automatic workflow, namely, CE Screen, within a high-throughput computational material framework, to facilitate the doping/solid solution calculation that involves high-throughput screening. The stable crystal structures can be found out much quicker in an automatic way using the CE Screen. It also makes cluster expansion simple and easy to use whether you are familiar with this method or not. Two doping calculations are investigated to evaluate the results from this workflow and credible results have been obtained by comparing with direct first principles calculations.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

More and more materials problems are made approachable by computation simulation with the growing of computer power and improvements of calculation methods. Since the last century, first principles simulations are powerful tools, because they allow for the exploration of new materials before an experiment devises them. Doping, as an effective way to improve the performance of known materials, is also considered as an important way to construct structures for new materials. In order to screen the stable structures from many known doped crystal structures with different doping concentrations, a large number of first principles calculations need to be carried out. The high-throughput (HT) method, which involves setting up and performing ab initio calculations, reorganizing and analyzing the results with minimal intervention by the user, has become an effective and efficient tool for materials development and prediction[1]. Since the Materials Genome Initiative (MGI) was launched in 2011 in the United States [2], many high-throughput calculation platforms and codes have been developed. For example, a software framework named Automatic Flow (AFLOW) [3] for HT calculation of crystal structures developed by Curtarolo group, a core program named Materials Project [4] developed by Ceder group, and a platform named MatCloud (http://matcloud.cnic.cn) developed by Yang group. Although these HT methods and their implementations accelerate the process of materials exploration, some sort of automatic optimization technique still needed to be developed and integrated. As we known, finding the most stable crystal structures of compounds is one of the classical problems in inorganic materials, because knowing the stable structure holds the key to material properties. To make sure the completeness of doping calculations, different concentrations with series of atomic sites should be considered. Using direct quantum mechanical calculations to search for the most stable structure from all doping configurations is not computationally feasible. Some platform (e.g. Material Project) has supplied good intuitive interface for users to carry out structure prediction, using the FireWorks which they have developed to automate and manage all computational steps at large supercomputing centers. But a graphical user interface for users to create the workflow on-demand for high-throughput screening is still lacking. Therefore, to improve the efficiency of
doping calculation, a method/workflow that can quickly and automatically screen stable structures from enormous known structures should be developed.

Since cluster expansion (CE) was proposed by Mayer in 1941, it has been employed as an approximate computation method to express the partition function as a power series expansion [5]. CE method has been a very compact and efficiency way to represent the alloy energetics [6] and it can be constructed using the alloy theoretic automated toolkit (ATAT) [7]. Besides alloys, CE method has recently been invoked to explore the stability of various two-dimensional and three-dimensional materials [8–11]. These applications have approved that CE is a good method to predict new structures and their energies. However, using CE method to predict new structures, users should be deeply understand its principles. In additions, existing codes implementing CE only can predict structures from some compositions, and the function of screening structures from known doped configurations by CE method is lacking. These aforementioned problems bring inconvenience to some researchers.

In this paper, an automatic workflow, named as CE Screen, is developed, within a high-throughput computational material framework called MatCloud. It can quickly and automatically screen the stable structures from lots of doped configurations which have the same parent lattice through partial first-principle calculations. The theoretical core of CE Screen is cluster expansion which has been implemented into ATAT, and related codes from ATAT are modified and wrapped into CE Screen. We provide a convenient and efficient way to use this workflow on MatCloud (http://matcloud.cnic.cn). Using CE Screen, to get stable structures for specific doping systems, only four inputs need to be supplied: initial structure, doping compositions, doping concentrations, and cross validation score.

The organization of this paper is as follows. First, we briefly introduce supercell approach, CE method, and parameters of the first principles calculations. Second, we depict how the CE Screen works, and describe how to use CE Screen on MatCloud platform in details. Finally, we apply CE Screen for two systems, which are fcc Al-Ti and bcc Fe-Al, these two cases doping Ti into fcc Al and doping Al into bcc Fe serve as the test cases to evaluate the workflow.

2. Theories and methods

The workflow of CE Screen is used to quickly find out stable structures from large number of doped structures. Applying CE Screen to doped system, the following theories and methods are involved: (i) Supercell construction, (ii) Cluster expansion, and (iii) First principles simulation.

2.1. Supercell construction and doped structures generation

Creating a supercell is the common method used to generate doped structures. Usually, structure compositions supplied by experiments are non-integer atomic site occupancies. But in practical study, a structure is usually described as a regular periodic structure with well-defined parameters, a space group and a set of crystallographic sites, where each crystallographic site is strictly occupied by a single type of atom. Therefore, in order to investigate the energy and other physical properties of structures who have non-integer atomic site occupancies, we need to construct one or more supercells (In the case of more supercells, we take the one whose calculation value is closer to the experiment value). In the supercell, atomic configurations can correspond with the partial occupancies and mixed compositions of structures with non-integer atomic site occupancies.

To obtain some key stoichiometric compounds, and balance the calculation loads, both integral and non-integral supercell should to be considered. The non-integral supercell is much more complex than integral one. In order to meet the needs of both integral and non-integral supercell construction, a $3 \times 3$ coefficient matrix $A$ is supplied to directly character the multiples for each direction. Once the matrix $A$ is identified, the supercell will be constructed.

The lattice vectors of unit cell is $(a, b, c)$, coefficient matrix $A$ is

$$
\begin{pmatrix}
a' \\
b' \\
c'
\end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix}
$$

(1)

Setting coefficient matrix $A$, we can obtain any multiples of unit cell, and all the atomic sites will be multiplied according to invariance of translation. This method has been encoded into MatCloud as a modeling tool, named as Build Supercell.

After supercells are constructed, doped structures with different concentrations can be generated by changing the elemental species on crystallographic site based on a supercell. Although the crystallographic equivalence is operated to all these doped structures, the number of doped structures is still very large, and it is impossible to select the stable structure for each concentration by direct first principles calculation. Therefore, the Cluster Expansion method is used to initially select the stable structures from those doped structures (i.e. prescreening).

2.2. Cluster expansion

Cluster expansion is a generalization of the well-known Ising Hamiltonian [12–16]. The formation energy of a particular configuration, $E$, is expanded exactly in polynomials of the spin variables weighted by multisite interaction parameters $J$ as:

$$
\Delta E(\sigma) = J_0 + \sum_{i} J_1 \sigma_i + \sum_{\langle i, j \rangle} J_2 \sigma_i \sigma_j + \sum_{\langle i, j, k \rangle} J_3 \sigma_i \sigma_j \sigma_k + \cdots
$$

(2)

It can approximately expressed as a series expansion of “clusters” which has been implemented in ATAT

$$
\Delta E_{CE}(\sigma) = \sum_{\alpha} m_{\alpha} \sum_{i \in \sigma_{\alpha}} \prod_{i \in \sigma_{\alpha}} \sigma_i
$$

(3)

where $\alpha$ is a cluster (a set of sites $i$). The summation takes over all clusters $\alpha$ that not equivalent by a symmetry operation of the space group of parent lattice, the average is taken over all the clusters $\alpha'$ that are equivalent to $\alpha$ by symmetry. The coefficients $J_n$ in this expansion embody the information regarding the energetics of the system and are called the Effective Cluster Interaction (ECI). The multiplicities $m_{\alpha}$ indicate the number of clusters that are equivalent to $\alpha$. Using cluster expansion method, some structural energies should be supplied to fit the ECIs, therefore, the selection of fitting structures which need to be calculated is important. The method how to select fitting structures will be explained in Section 3.1.3 in details.

One cluster $\alpha$ is a particular lattice sites combination. If all possible types of clusters of a particular configuration are included, its formation energy can be exactly mapped. The clusters can be found among the lattice sites of structure: all inequivalent pairs, triplet, quadruplets, etc., up to the m-body interaction which include the whole lattice. While, contributions from larger clusters often become negligible, a radius cutoff is needed to determine the size of maximal cluster. In CE Screen, the radius cutoff is the 8th nearest neighbor of parent structure referring to $N_dR_0$ criterion which has been employed in ATAT [7,17]. All the cluster information is determined by the parent lattice data and its symmetry operations.
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات