Grouping of structures in cluster expansion of multicomponent systems

Atsuto Seko[1], Isao Tanaka[1,2]

[1]Department of Materials Science and Engineering, Kyoto University, Kyoto 606-8501, Japan
[2]Nanostructures Research Laboratory, Japan Fine Ceramics Center, Nagoya, 456-8587, Japan

Control of errors over the whole range of structures is essential when we combine a large set of density functional theory calculations and the cluster expansion method[1-3] for predicting the ground-state structures and configurational thermodynamics of multicomponent systems. An optimal CE is generally constructed from the DFT results of many ordered structures that are sampled from the total population of structures. Figure 1 shows a distribution function of the structure population in the configurational space. In the schematic map, group 4 includes the largest number of structures. They correspond to structures near a random structure. On the other hand, group 1 includes the smallest number of structures, which are far from a random structure. They will be hereafter called "minority structures." It should be emphasized that ground-state structures are usually included in minority structures. Therefore, the accuracy for predicting minority structures should be very important. In this study, we propose a procedure based on the cluster analysis of the structure population, which can adequately take into account the errors of minority structures as well as those of random structures.

The cross validation (CV) score has been widely accepted as a quantity for controlling the accuracy of the CE[4,5]. For the accurate evaluation of the CV score, an optimal set of many DFT structures is necessary[6,7]. In such a case, however, the errors of minority structures are only a minor part of the CV score. When the errors of minority structures are much larger than the average error, they tend to be underestimated in the CV score. In order to estimate the accuracy of a wide range of structures including minority structures, the cluster analysis of the structure population (CASP) is introduced[8]. CASP enables us to classify structures of similar correlation functions into the same group, as illustrated in Fig. 1. Here CASP is performed by the model-based cluster analysis[9,10].

The usefulness of the procedure is demonstrated by applying it to configurational behaviors of $MgAl_2O_4$ spinel. The uniform sampling of DFT structures can be achieved by evenly selecting structures from all the groups divided by CASP. Here, CEs are constructed from DFT structures sampled evenly and randomly from all the groups. The number of clusters is fixed at 17. To examine the quality of the CEs constructed by the proposed procedure, CEs are constructed from two kinds of DFT structures prepared by different sampling procedures. One is composed of high-symmetry structures (HSs), which have multiple symmetry operations. The other is composed of randomly selected structures (RAs). The dependence of CE error on the DFT energy is shown in Fig. 2. The CEs are constructed from 120 DFT structures. The CE error is approximately estimated from the RMS difference between the DFT and CE energies for all structures in the structure population. In the CE with the CASP sampling, structures with a wide range of energies can be most precisely predicted among three sampling. On the other hand, the CE with the RA sampling has a low accuracy for structures with a low energy. It can reconstruct structures only in the energy range of 50–125 meV.

This study was supported by a Grant-in-Aid for Young Scientists (A) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. I.T. acknowledges support in the form of both a Grant-in-Aid for Scientific Research (A)

and a Grant-in-Aid for Scientific Research on Priority Areas “Nano Materials Science for Atomic Scale Modification 474” from MEXT, Japan.

References

[1] J. M. Sanchez, F. Ducastelle, and D. Gratias, Physica A 128, 334 (1984).
[2] D. de Fontaine, Solid State Phys. 47, 33 (1994).
[3] F. Ducastelle, Order and Phase Stability in Alloys (North-Holland, Amsterdam, 1994).
[4] A. van de Walle and G. Ceder, J. Phase Equilib. 23, 348 (2002).
[5] G. L. W. Hart, V. Blum, M. J. Walorski, and A. Zunger, Nat. Mater. 4, 391 (2005).
[6] A. Seko, Y. Koyama, and I. Tanaka, Phys. Rev. B 80, 165122 (2009).
[7] B. Arnold, A. Díaz Ortiz, G. L. W. Hart, and H. Dosch, Phys. Rev. B 81, 094116 (2010).
[8] A. Seko and I. Tanaka, Phys. Rev. B 83, 224111 (2011).
[9] C. Fraley and A. E. Raftery, J. Am. Stat. Assoc. 97, 611 (2002).
[10] C. Fraley and A. E. Raftery, MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering, Technical Report 504 (University of Washington, Department of Statistics, 2006) (revised 2009).

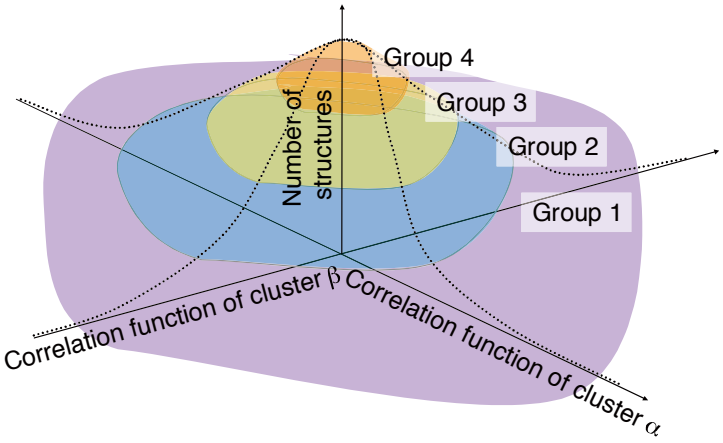

FIG. 1. Schematic illustration of distribution function of structure population in space of correlation functions.

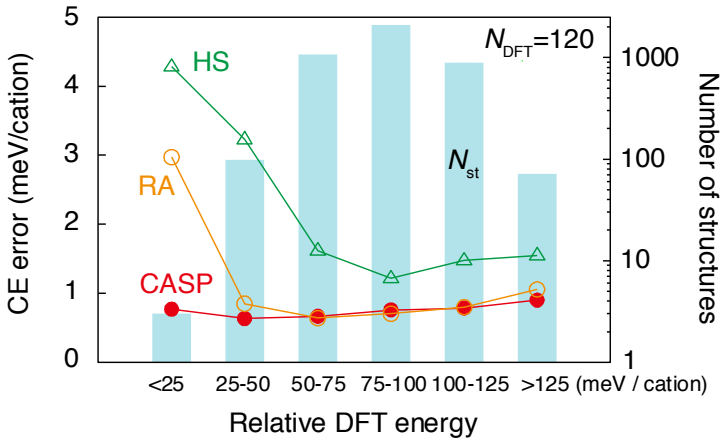

FIG. 2. DFT energy dependencies of CE errors made from three types of 120 DFT structure. The number of structures belonging to a group classified on the basis of the DFT energy is also shown. The relative energy is measured from the energy of the normal spinel.