A Bayesian Implementation of Cluster Expansion in Multidimensional Discrete Wavelet Transform Representation

Jin Xu, Nicholas Zabaras and Guang Lin

1Material Process Design and Control Laboratory, Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14850 USA

2Pacific Northwest National Laboratory, Richland, WA 99352 USA

Email: jx77@cornell.edu, zabaras@cornell.edu, guang.lin@pnl.gov
URL: http://mpdc.mae.cornell.edu/

Developments towards an automated first-principles based virtual materials laboratory that minimizes the need for experimental input during the process of materials design is discussed. The cornerstone of the methodology is a cluster expansion (CE) algorithm which distinguishes itself from available CE approaches in the following ways: First, an adaptive basis set from the multidimensional discrete wavelet transform (MDWT) representation of the CE is employed. Secondly, a Bayesian scheme is adopted in determining the correlation functions (coefficients) of the expansion. With the configurational energy thus calculated, and interfacing with thermodynamic databases and other existing software one has the ability to accurately and efficiently compute various thermodynamic properties of solid systems.

The main advantage in adopting the MDWT basis set lies in its improved convergence rate thus reduced dimensionality for the fitting problem. This improvement sidesteps the necessity to impose non-physical properties on the function being expanded. The Bayesian scheme greatly reduces the computational intensity during the regression procedure by incorporating physical insights and empirical data into the prior probability distribution. This approach is of particular significance to low-symmetry systems. While most current CE implementations use the least square method or cross-validation to fit the correlation functions deterministically, the implementation of such Bayesian scheme also allows us to conduct quantitative uncertainty analysis of the fitted parameters and provides us an insight to the propagation of these uncertainties to the various derived thermodynamic properties.