Bayesian cluster expansion with lattice parameter dependence for studying surface alloys

by
Le Niu

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Engineering

Baltimore, Maryland

May, 2016

Abstract

DFT+Cluster Expansion+Monte Carlo (DFT-CE-MC) process of studying thermodynamic properties of lattice structure is of great use in materials design for a variety of applications. Among these applications, rational design of alloy surface for catalyst usage is a promising field with great importance. Unlike the traditional experimental approach, the DFT-CE-MC process provides a foundation for a "virtual laboratory" in which all surface structures can be studies computationally with little aids of experimental data. This kind of "virtual laboratory" will save a lot of effort in experimental tries and thus speed up the design of high catalyst structures.

In details, thermodynamic stable ground states of surface alloy under different conditions and the relationship between atomic structure, formation enthalpy, chemical potential and catalyst properties can be gathered through this process. With these information, structures optimizing catalyst properties could be proposed and a detailed analyze of influential factors of catalyst properties is available.

However, a problem is raised if considering the composition mismatch between surface and bulk region. Since the lattice of surface region should always match that of bulk region, the lattice parameter of the surface, and hence the interactions among near-surface atoms, varies with the composition of the underlying bulk material.

Collecting the training data of surface structures with varies underlying bulk composition through DFT demands a large amount of calculation, and should be wasteful when the strain-energy relationship can be predicted accurately through known model, such as quadratic rule.

Here, in this thesis, we demonstrate that quadratic rule model can be included in cluster expansion framework, especially for Bayesian cluster expansion approach, so that surfaces under a variety of strains can be used to train a single cluster expansion that predicts properties as a function of atomic order and surface strain. We implemented this idea on Bayesian cluster expansion and tested our method in Au-Pt and Ni-Pt alloy surface with oxygen adsorption. The result shows the validity of the idea and suggests a great potential of this method in application of catalyst research.

We will firstly introduce the basic concepts of DFT-CE-MC process, and then briefly review the cluster expansion approach and its recent development. After that, we will introduce the lattice parameter dependent cluster expansion and the validity test of this newly developed method. Finally, we will go through some of the possible applications of this method.

ACKNOWLEDGMENT

I gratefully acknowledge my advisor, Professor Tim Mueller, for his valuable guidance, feedback in support of my research and preparing me for the future. Prof. Mueller's intelligence, enthusiasm and patience are excellent characters not only guide his students to success, but also provide a perfect example of a respectable scholar. The time working with him is truly enjoyable. I particularly thank Dr. Liang Cao for his valuable help and our fruitful conversations about my project. I sincerely thank Pandu Wisesa for his help and enthusiasm. I also acknowledge Thomas Nilson, André Botelho, Chenyang Lee, Fenglin Yuan, Peter Lile, Alberto Hernandez-Valle and all other people I have been in contact with in Prof. Mueller's group, for fruitful conversations and comradery.

I should also mention Department of Materials Science and Engineering and Johns Hopkins University, for providing me such a excellent environment.

Finally, I would like to thank my family: AnCai Niu(dad), LuQin Zhao(mom) for their overflowing love and support.

Contents

1. Introduction to DFT-CE-MC technique

We use a number of computational methods to study surface alloy structure, with the ultimate goal of building our knowledge of this thermodynamic systems all from first principle. Thanks to the steady growing computer power and improvement in numerical algorithm, first principle methods, such as Density functional theory (DFT) which aims at solving many electron $schödinger$ equation in a single electron way [1, 2], are currently applicable method in evaluating formation energies of alloy structures. However, accurate DFT computations are still relatively expensive and the scale are typically restricted to structures containing less than 100 atoms . While to compute thermodynamic quantities for typical Monte Carlo simulation, Hamiltonian of supercells with $\sim 10^3$to 10^6 atoms are required, inaccessible by direct DFT calculations. Thus, cluster expansion method are used to bridge the gap between needs of Monte Carlo and supply from DFT through constructing Hamiltonian of large structures from information gathered through small size DFT calculation(training sets).

1.1 Density functional theory

As an easy approach to first principle method, DFT has been widely used for calculations in solid-state physics since 1970s. The name density functional theory came from the Hohenberg-Kohn theorem where it makes an assumption that physical properties of solid with periodic symmetry is a functional of electron density of the system. By accepting this, it is possible to transform a many body $Schrödinger$ equation into a single body equation, thus makes it great easier to solve complex $Schrödinger$ equation and a calculation of massive electronic system

becomes possible. For this solving process, the only parameters needed to be fit is the exchange-correlation functional which corresponds to the interaction between electrons. Conventionally, LDA or GGA approximation [1, 3] can be used to simulate this interaction accurately and thus the total solving process is based on solid theories. As an easy access to first principle method, DFT finds increasingly broad application in the chemical and materials science. For example, for the needs of this study, formation energy of given atomic structure was calculated through DFT. Under 40 years development, DFT has become a mature technique with many commercial software developed to implement it, such as the Vienna Ab initio Simulation Package(VASP), SIESTA [4, 5, 6, 7].

Below we provide a brief outline of DFT, whose details can be easily found in various literature.

For a given structure consisting of N ions and M electrons:

$$S(ion_1, ion_2, \dots ion_N, e_1, e_2, \dots e_M) \quad (1)$$

the time independent $Schrödinger$ equation is

$$\hat{H}\psi(r_1, r_2, \dots, r_N) = E\psi(r_1, r_2, \dots, r_N) \quad (2)$$

$$With\ \hat{\mathrm{H}} = -\frac{1}{2}\sum_{\mathrm{i}}^{\mathrm{M}} \nabla^2 - \sum_{\mathrm{I}}^{\mathrm{N}} \sum_{\mathrm{i}}^{\mathrm{M}} \frac{\mathrm{Z_I}}{|\mathrm{r_i} - \mathrm{R_I}|} + \sum_{\mathrm{i<j}}^{\mathrm{M}} \frac{1}{|\mathrm{r_i} - \mathrm{r_j}|} \quad (3)$$

Where the first term on the right is the kinetic energy of electrons, second term is the ion-electron attraction, and the third term is the electron-electron repulsion. R_I and r_i are coordinates of ions and electrons respectively, Z_I is the number of positive charges in an ion.

Following the Hohenburg and Kohn theorem, we get the electron density $\rho(r)$ uniquely determines ground states $\widehat{H}$ and thus the ground state energy of the structure is a functional of electron density alone:

$$E[\rho(r)] = T[\rho(r)] + \frac{1}{2}\int\int \frac{\rho(r)\rho(r\prime)}{|r-r\prime|} drdr' - \sum_I^N \int \frac{2I}{|r-R_I|}\rho(r)dr + E_{xc}[\rho(r)] \quad (4)$$

Where first three functional are kinetic energy of non-interacting electron gas, the classical Coulombic electron-electron interactions energy due to ion-electron interaction which are all known exactly. The last term $E_{xc}[\rho(r)]$ is the so called exchange-correlation functional, represent the interaction contribution to the energy, whose exact form is unknown.

Different technique could be applied to deal with exchange correlation term. The first technique used widely is local density approximation (LDA), which assume $E_{xc}[\rho(r)]$ is a function of the local electron density

$$\mathrm{E_{xc}[\rho(r)] \approx \int dr\rho(r)\epsilon_{xc}[\rho(r)]} \quad (5)$$

where $\epsilon_{xc}[\rho(r)]$ is the exchange and correlation energy density of a small region of local electron gas, which we assume can be viewed as homogeneous. Another option which based on LDA is generalized gradient approximation (GGA), which is to assume the $\epsilon_{xc}[\rho(r)]$ can be trucated within first order in its Tayler expansion

$$E_{xc}[\rho(r)] \approx \int dr\rho(r)\epsilon_{xc}[\rho(r), \nabla\rho(r)] \quad (6)$$

knowing the expression of $E_{xc}[\rho(r)]$, we then are able to solve the equation into

$$\left[-\frac{1}{2}\nabla^2+\int\frac{\rho(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{d}\mathrm{r}'-\sum_{\mathrm{I}}^{\mathrm{N}}\frac{\mathrm{Z}_{\mathrm{I}}}{|\mathrm{r}-\mathrm{R}_{\mathrm{I}}|}+\frac{\delta\mathrm{E}_{\mathrm{xc}}[\rho(\mathrm{r})]}{\delta\rho(\mathrm{r})}\right]\phi_{\mathrm{i}}=\epsilon_{\mathrm{i}}\phi_{\mathrm{i}} \quad (7)$$

Simplified into

$$\widehat{\mathrm{h}_{\mathrm{ks}}}\phi_{\mathrm{i}}=\epsilon_{\mathrm{i}}\phi_{\mathrm{i}},\ \widehat{\mathrm{h}_{\mathrm{ks}}}=-\frac{1}{2}\nabla^2+\int\frac{\rho(\mathrm{r}')}{|\mathrm{r}-\mathrm{r}'|}\mathrm{d}\mathrm{r}'-\sum_{\mathrm{I}}^{\mathrm{N}}\frac{\mathrm{Z}_{\mathrm{I}}}{|\mathrm{r}-\mathrm{R}_{\mathrm{I}}|}+\frac{\delta\mathrm{E}_{\mathrm{xc}}[\rho(\mathrm{r})]}{\delta\rho(\mathrm{r})} \quad (8)$$

Where ϕ_i is the orbitals of non-interacting particles and $\rho(r)=\sum_i|\phi_i(r)|^2$

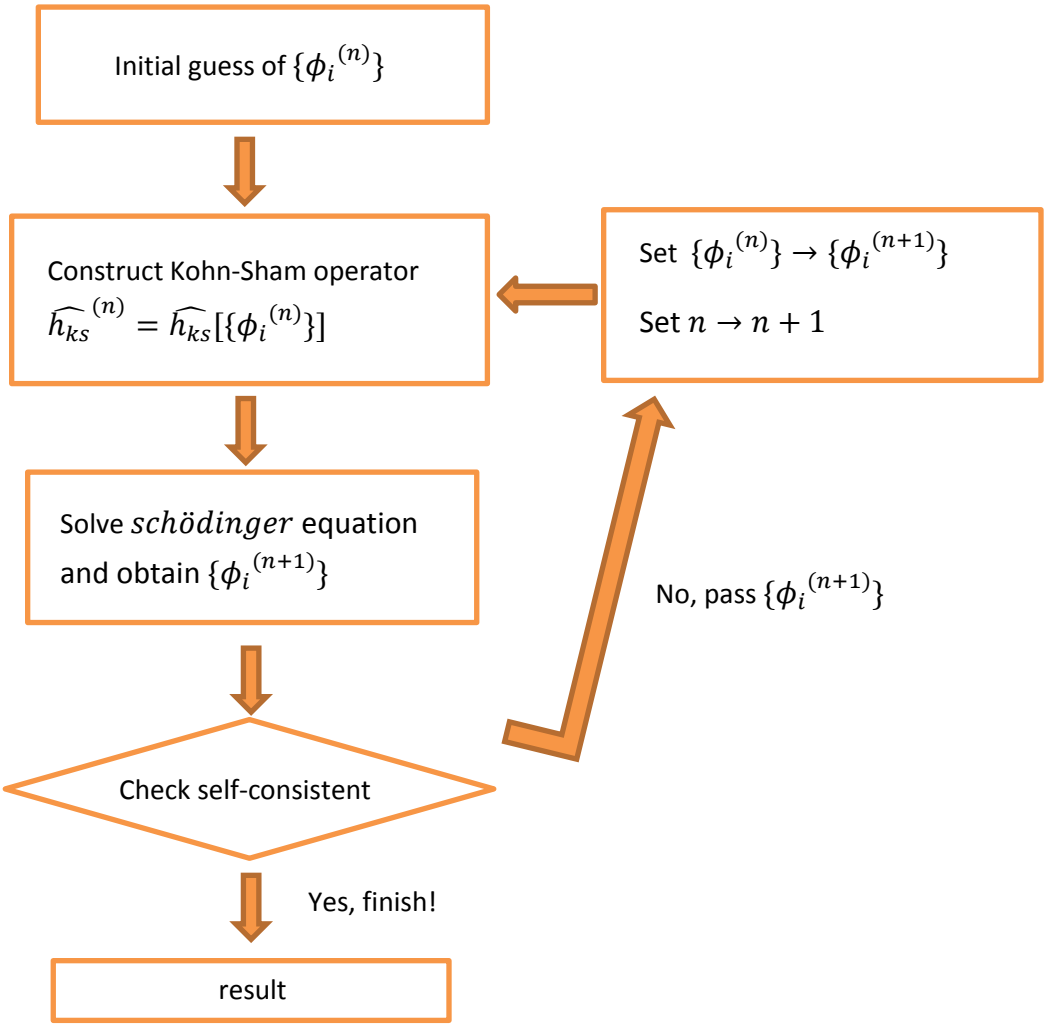
Equation above shows the process to transform a many body $schödinger$ equation into solution of several single body equations.

Still, there is a problem that our Kohn-Sham operator $\widehat{h_{ks}}$ is actually the function of molecular orbitals

$$\widehat{\mathrm{h}_{\mathrm{ks}}}^{(\mathrm{n})}=\widehat{\mathrm{h}_{\mathrm{ks}}}[\{\phi_{\mathrm{i}}{}^{(\mathrm{n})}\}] \quad (9)$$

and a analytical solution of [equation number] becomes inaccessible. However, using a self-consistent procedure, it is numerically solvable.

The self-consistent procedure can be described as following plot:



Practically, different basis in representing the electronic states is used in different implementation and correspondingly have different advantage and disadvantage. For example, the VASP [6, 5, 4] use a plane wave basis which assume the orbital is plan wave orthogonal between one another . This treatment is especially suitable for lattice structures where electron orbital can be safely treated as plane wave in periodic potentials. On the contrary, SIESTA [7] used a local orbital basis, which is more suitable for large molecules.

1.2 Monte Carlo Simulation

Knowing the expression of Hamiltonian of a system, the convention approach to study the thermodynamic average quantities in equilibrium is Monte Carlo simulation, which can be simply understood as a massive stochastic sampling method. Based on statistical mechanical knowledge, we know the thermodynamically average value is given by

$$< A >= \frac{\sum_{\vec{\sigma}} A(\vec{\sigma}) exp(-\beta H(\vec{\sigma}))}{\sum_{\vec{\sigma}} exp(-\beta H(\vec{\sigma}))} \quad (10)$$

Where the denominator $\sum_{\vec{\sigma}} \exp(-\beta H(\vec{\sigma}))$ is the partition function of the system, $\vec{\sigma}$ represent the structure which contains N sites, $\beta = 1/k_B T$ is the inverse temperature.

As the configuration number $M(\vec{\sigma}) = n^M$ (where n is the number of possible elements can be decorated on a site) increase exponentially with the increase of the cluster size M, it is actually impossible to mapping over all possible configurations in order to realize a true thermodynamic average as in **eq. 10**. A possible approach is to sample important configurations that have the major probability of occurrence. This idea then results in the Metropolis algorithm.

The basic idea of metropolis algorithm is to generate random state according to normalized Boltzman factor $p(x) = \frac{1}{z}\exp(-H(x)/k_B T)$, so that a simple sampling of this random sequence will become an importance sampling we need and thus represent thermodynamic average. In practical, a sequence of random walk step is generated (Markov Chain [8]) instead of generating independent states according to their probability of occurrence. For realizing the effect that each step represents a random state occurring with a probability of $p(x)$, a transition probability T_{ij} for the current state moving to a new state is defined and detailed

balance which ensure that such a process can be equilibrated by a reverse process for the demand of thermodynamic equilibrium is satisfied.

In details, a Markov chain is a sequence of random states $\{x_i\}$, for which the value x_{i+1} depends only on the previous value x_i. Let us assume the probability of the state moving from x_i to x_{i+1} is

$$\mathrm{T}(\mathrm{x_i} \rightarrow \mathrm{x_{i+1}}) = \alpha(\mathrm{x_i} \rightarrow \mathrm{x_{i+1}})\mathrm{acc}(\mathrm{x_i} \rightarrow \mathrm{x_{i+1}}) \quad (11)$$

where $\alpha(x_i \rightarrow x_{i+1})$ is the probability of selecting a new state x_{i+1} from all possible neighbor states of x_i for next move. $acc(x_i \rightarrow x_{i+1})$ is the probability of accepting this new state x_{i+1}.

Then to realize detailed balance:

$$\mathrm{P}(\mathrm{x_i})\mathrm{T}(\mathrm{x_i} \rightarrow \mathrm{x_{i+1}}) = \mathrm{P}(\mathrm{x_{i+1}})\mathrm{T}(\mathrm{x_{i+1}} \rightarrow \mathrm{x_i}) \quad (12)$$

Which means the flow of states from x_i to x_{i+1} exactly cancle the flow from x_{i+1} to x_i, which realize equilibrium.

We have:

$$\mathrm{P}(\mathrm{x_i})\alpha(\mathrm{x_i} \rightarrow \mathrm{x_{i+1}})\mathrm{acc}(\mathrm{x_i} \rightarrow \mathrm{x_{i+1}}) = \mathrm{P}(\mathrm{x_{i+1}})\alpha(\mathrm{x_{i+1}} \rightarrow \mathrm{x_i})\mathrm{acc}(\mathrm{x_{i+1}} \rightarrow \mathrm{x_i}) \quad (13)$$

if we just randomly select new state from the nearby states then

$$\alpha(\mathrm{x_i} \rightarrow \mathrm{x_{i+1}}) = \alpha(\mathrm{x_{i+1}} \rightarrow \mathrm{x_i}) \quad (14)$$

We have:

$$\frac{\mathrm{acc}(x_i \to x_{i+1})}{\mathrm{acc}(x_{i+1} \to x_i)} = \frac{P(x_{i+1})}{P(x_i)} \quad (15)$$

The metropolis algorithm makes the explicit choice that if $\frac{P(x_{i+1})}{P(x_i)} > 1$, then always accept the new states. Then it follows:

$$\mathrm{acc}(x_i \to x_{i+1}) = \frac{P(x_{i+1})}{P(x_i)} = f(x) = \begin{cases} exp\left(\frac{-\left(E_{x_{i+1}} - E_{x_i}\right)}{k_B T}\right), & E_{x_{i+1}} > E_{x_i} \\ 1, & E_{x_{i+1}} \leq E_{x_i} \end{cases} \quad (16)$$

Following such process, we can realize the thermodynamic average through a rather simple accepting rule without any cost of generality. A code implementing basic Monte Carlo simulation is quite simple that within handful of lines:

```
FTotal= 0
For(step=0 to NumOfStep){
    Pick a trial move
    Compute the change in energy for the trial move(DE)
    If(DE<0 or randf()<exp(-DE/kT))
        then accept the move
        Else reject the move
    FTotal=NumTotal+FCurrentConfiguration
}
FAverage=FTotal/NumOfStep
```

The power of metropolis algorithm is that with this simple step, it is possible to undertake massive steps of run and with such large number of trials:

1. The average will reach thermodynamic average
2. After the Monte Carlo system reaches equilibrium, the new accepted states will become the thermodynamic real states which would much likely to be found in a real system, unlike the initial arbitrary guess states which may not be thermodynamically favored.

Bearing such a fact, it is possible to pick up some states during equilibrium runs for detailed analyze, and simulated annealing process can be applied for finding ground states where every energy step is a Monte Carlo step.

2. A brief review of cluster expansion method and its recent development

2.1 cluster expansion method

As we have mentioned before, there is a big gap between the scale of the Hamiltonian needs for Monte Carlo simulation and the scale of the Hamiltonian provided by DFT. A direct choice is to produce a large scale Hamiltonian using the information collected from DFT Hamiltonian. Cluster expansion method is an implementation of such choice [9]. While its foundation is laid on solid theory of statistical mechanism as an approximate technique for the treatment of cooperative phenomenal in periodic system which give very accurate results outside the critical region, it can also be viewed as a machine learning process since it learns from DFT Hamiltonian to get correct ECIs(effective cluster interaction).

Theoretically, the basic idea of cluster expansion method is that the extensive properties of periodic systems(like lattice structure) is a linear combination of cluster functions, where cluster functions are tenser product of site basis belonging to a single cluster.

$$\mathrm{F}(\vec{s}) = \sum_{\vec{b}} \mathrm{V}_{\vec{b}} \prod_{i} \Theta_{bj}(\mathrm{S}_i) \qquad (17)$$

In details, we assign variable to single sites according to their decorated elements and then define site basis to be orthogonal polynomials of site variables assigned to each site, a tensor product of site basis mapping over all the sites in a cluster(a collection of sites which have interaction between one another) then forms a cluster function

$$\Phi_{\vec{b}}(\vec{S}) = \prod_{i} \Theta_{bj}(\mathrm{S}_i) \qquad (18)$$

and a linear combination of all possible cluster functions can estimate the extensive properties of lattice structure accurately when the system is outside of critical region. The process is firstly proposed by J.M. Sanchez et al [9] where they used discrete Chebyshev polynomials as site basis. Later, it became standard procedure for constructing cluster expansion.

Practically, once we have constructed cluster functions through known decorated lattice structure, and get the coefficients of the linear combinations ($V_{\vec{b}}$ in eq.), also known as ECIs. We can predict the extensive properties accurately. The ECIs are unknown parameters for the system and not available for an exact expression. Conventionally, a least square fit for several structures with their known value of extensive properties is undertaken to get the ECIs. The process is known as training process, and these structures are training set. The expression of

least-square fit can be written as $\vec{V} = (X^T X)^{-1} X^T \vec{y}$, where X is known as input matrix where each line contains the value of cluster functions belonging to the same structure and each column contains the value of cluster functions belonging to the same cluster [10]. While the number of all possible clusters in a structure expand exponentially with the number of site, $N = M!$. and since for least square fitting, the number of training structure should not be smaller the number of cluster functions of choice, for a non-truncated cluster expansion, the training set will be extremely large where practically inaccessible. That is to say, we should truncate the choice of clusters and try to achieve accuracy within limited number of clusters. Fortunately, since clusters with larger number of sites and larger distance between sites tend to have smaller ECIs and the curve drops quickly that ECIs for clusters larger than certain size could be ignored, a sufficient accuracy can be achieved by keeping only clusters that are relatively compact.

The choice of clusters and training structures is an important topic in truncated cluster expansion procedure. Many different techniques have been developed in order to optimize the training system. Before long, this process much relies on the researchers' personal physical intuitions, however, the automated algorithm proposed by Axel van de Walle et al [11] has overcome this problem and became widely used approach for cluster expansion. In this algorithm, the concepts of cross-validation and variance minimization is applied so that removing arbitrary parameters from the computation and guaranteed that the results obtained are truly derived from the underlying first principle calculation.

Cross-validation:

Although the well-known mean square error is a choice of the estimation of the goodness of a fitting. There are specific needs for cluster expansion fitting process: a well estimation of how the terms of clusters influence the fitting results. In details, if too few terms are kept in CE, predict result will be imprecise, however, on the other hand, if too many terms are kept, given fixed number of training set, the problem of overfitting may appear. The score of the fitting thus should represent the compromise between these two unwanted effects. Mean Square Error fails in this aspect, since even if overfitting happens, the MSE value still can be very small. van de Walle et al [11] rediscovered a formal solution to this problem: the cross-validation score.

The cross validation score is defined as:

$$(\mathrm{CV})^2 = \mathrm{n}^{-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{E_i} - \overline{\mathrm{E}}_{(\mathrm{i})})^2 \qquad (19)$$

Where E_i is the calculated energy of structure i, $\overline{E}_{(i)}$ is the predicted value of the energy of structure i obtained from a least square fit to the n-1 other structural energies. In contrast to MSE, the CV score is not monotonically decreasing, instead, as the number of clusters included increase, it firstly decreases because of an increasing number of degree of freedom, but then increase because of an increase in the noise of the DFT energies.

Cross-validation score is a reliable measurement of the goodness of fitting, and the problem of selecting clusters according to one's physical intuitions then become an problem of selecting clusters to minimize the CV scores, thus allow minimizing algorithm such as simulated annealing method to take place.

Variance reduction:

As Alex van de Walle et al states [11], the prediction error of a cluster expansion fit can be separated into two components: the bais and the variance.

Here the bias is the accumulate result of the bias of ECIs caused by fit of each training structure and the variance is caused by fluctuations around the mean ECIs of each ECIs fitted by each training structure. Although we can not know the bias without the knowledge of the energies of training structures, we can estimate variance with only the information of input Matrix X.

The variance can be represented in a form of covariance matrix of the ECIs through least square fitting theory:

$$\mathrm{M} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{e}^{2} \qquad (20)$$

Where e^2 is the mean square error of the fit.

Then the variance of predicted energy can be written as:

$$\mathrm{Var}\left[\overline{\mathrm{E}}_{\mathrm{i}}\right] = \mathrm{x}_{\mathrm{i}}\mathrm{M}\mathrm{x}_{\mathrm{i}}^{\mathrm{T}} \qquad (21)$$

Where $\overline{E}_i$ is the predict energy of structure i, x_i is the ith row of input matrix X which represents structure i.

We average this quantity over all structures in training set, then comes the predictive power of the cluster expansion:

$$\mathrm{E}\left[\mathrm{Var}\left[\overline{\mathrm{E}}_{\mathrm{i}}\right]\right] \propto \int_{||\mathrm{u}||=1} \mathrm{u}^{\mathrm{T}}\mathrm{M}\mathrm{u}\mathrm{d}\mathrm{u} = \mathrm{Tr}(\mathrm{M}) \qquad (22)$$

For details of this conduction, see [11]. Here an assumption was made that the cluster functions x_i of every possible structures are distributed isotropicaly in a sphere, so that the density function $f(||u||)$ of the randomly picked structures in phase space is spherically symmetric.

Thus we can get the expected variance of the energies predicted from a cluster expansion. This expression is then used to provide guidance for training structure selection. Since the least square fitting procedure minimizes MSE e^2, indicating that adding new point to the fit yields no first order change in e^2. A direct consideration is finding new structure i which maximizes the reduction r of the trace of $(X^T X)^{-1}$:

$$\mathrm{r} = \mathrm{Tr}((\mathrm{X^T X})^{-1}) - \mathrm{Tr}\left(\left(\mathrm{X^T X} + \mathrm{X_i}^{\mathrm{T}}\mathrm{X_i}\right)^{-1}\right) \quad (23)$$

The maximum variance reduction ΔV_{max} is then reached when ${X_i}^T$ is the longest possible column vector v parallel to the eigenvector of $X^T X$ associated with the smallest eigenvalue:

$$\Delta\mathrm{V_{max}} = \mathrm{Tr}((\mathrm{X^T X})^{-1} - \left(\mathrm{X^T X} + \mathrm{X_i}^{\mathrm{T}}\mathrm{X_i}\right)^{-1}) \quad (24)$$

This condition turned out to be an eigenvalue problem

$$(\mathrm{X^T X})^{-2}\mathrm{v} = \lambda\mathrm{v} \quad (25)$$

Solving this for the largest eigenvalue of $(X^T X)^{-2}$, we can find the new structure v which minimize the variance of cluster expansion fitting.

Axel van de Walle's process provides an analytic way of select clusters and training structures out of physical intuitions of researchers, and laid the foundation of widely accessible of CE

method to large number of researchers. However, this is no ending of the development of this method. Next, we will explore other developments based on this foundation.

2.2 Bayesian cluster expansion

As we indicated in previous section, Axel van de Walle et al [11] proposed an process which can automatically select clusters and training structures without help of physical intuitions of researchers. However, a question raised that if there is a general physical intuition which can help automatically selecting training structure and clusters thus speed up the process of Alex et al. The answer is yes.

Actually, early in 1992, Lakes David B. et al [12] have showed this possibility in his proposal of reciprocal cluster expansion which can automatically give weight to occurrence probabilities of clusters according to their size (number of atoms and distance between sites). This turns out to be the early embryo of Bayesian cluster expansion.

The Bayesian cluster expansion, proposed by Tim Mueller et al [13], based on the idea that physical intuition of choosing clusters can provide great help in optimize cluster selection process, and this idea became possible through a Bayesian approach, which transform a problem of minimizing prediction error of ECIs into a problem of maximizing the distribution of ECIs given training data using Bayes' theorem.

The physical intuitions considered in Bayesian cluster expansion mainly comes in form of 3 types:

1. Properties prediction should be close to those predicted by theoretical or empirical

models.

2. The greater the number of sites of a cluster and the greater the distance between sites, the smaller the ECIs should be.

3. Similar cluster functions should have similar ECIs.

These 3 physical intuitions then result in 3 treatment. For intuition 1, we can transform our target function of property value (like energy value) into a predict error of theoretical or empirical models. For example, we expected energy of binary alloy should be $E_{expect} = C_A E_A + C_B E_B$. where E_A E_B is the chemical potential of pure A and B, C_A C_B is the composition of A and B. Then instead of studying energy of alloy E, we can study the error between energy E and our expected energy E_{expect}: $E_{error} = E - E_{expect}$.

Since prior expected value for E_{error} is zero and the prior expected value for each ECIs in the cluster expansion of E_{error} must also be zero, it is safe to say the mean of the prior distribution for the ECIs is zero.

For intuition 2, since we have assume the mean of prior distribution of ECI is zero, we can further assume this distribution is a Gaussian distribution. Then the physical intuition 2 can be transformed into:

$$\mathrm{P}(\vec{\mathrm{V}}|\mathrm{X}) \propto \prod_{\alpha} \mathrm{e}^{-\mathrm{V}_{\alpha}^2/2\sigma_{\alpha}^2} \quad (26)$$

Where V_α is the αth element of $\vec{V}$ and σ_α^2 is the variance of V_α.

For intuition 3, we can further assume the distribution of difference of ECIs between similar clusters is also Gaussian like:

$$P(\vec{V}|X) \propto \prod_{\alpha,\beta\neq\alpha} e^{-(V_\alpha - V_\beta)^2/2\sigma_{\alpha\beta}^2} \quad (27)$$

Where $2\sigma_{\alpha\beta}^2$ indicates the degree of expected similarity between the two ECIs. Thus for equation

$$P(\vec{V}|X,\vec{y}) = \frac{P(\vec{y}|\vec{V},X)P(\vec{V}|X)}{P(\vec{y}|X)} \quad (28)$$

We have

$$P(\vec{V}|X) \propto \prod_{\alpha} e^{-V_\alpha^2/2\sigma_\alpha^2} \prod_{\alpha,\beta\neq\alpha} e^{-(V_\alpha - V_\beta)^2/2\sigma_{\alpha\beta}^2} \quad (29)$$

$P(\vec{y}|X)$ is a constant with respect to $\vec{V}$, and $P(\vec{y}|\vec{V},X)$ is also a Gaussian distribution with zero mean as we analyzed in intuition 1.

$$P(\vec{y}|\vec{V},X) \propto \prod_{i} e^{-(y_i - \vec{x}_i\cdot\vec{v})^2/2\sigma_i^2} \quad (30)$$

We maximize $P(\vec{V}|X,\vec{y})$ and thus got the ECIs:

$$\vec{V} = argmin\left[-\ln\left(P(\vec{y}|\vec{V},X)\right) - \ln\left(P(\vec{V}|X)\right)\right]$$

$$= \text{argmin}\left[\sum_i \frac{(y_i - \vec{x}_i\cdot\vec{v})^2}{2\sigma_i^2} + \sum_\alpha \frac{V_\alpha^2}{2\sigma_\alpha^2} + \sum_{\alpha,\beta\neq\alpha} \frac{(V_\alpha - V_\beta)^2}{2\sigma_{\alpha\beta}^2}\right] \quad (31)$$

The final result would be:

$$\vec{V} = (X^T W X + \Lambda)^{-1} X^T W \vec{y} \quad (32)$$

Where W is a diagonal weight matrix, Λ is the regularization matrix with elements given by

$$\Lambda_{\alpha\alpha} = \frac{\sigma^2}{\sigma_\alpha{}^2} + \sum_{\beta|\beta\neq\alpha} \frac{\sigma^2}{\sigma_{\alpha\beta}{}^2} \quad (33)$$

$$\Lambda_{\alpha\beta} = \Lambda_\alpha = \Lambda_\beta = -\frac{\sigma^2}{\sigma_{\alpha\beta}{}^2} \quad (34)$$

Where σ^2 is an unknown constant.

We can better understand the function of Λ through comparison with least square fit:

$$\vec{V} = (X^T W X)^{-1} X^T W \vec{y} \quad (35)$$

We see the regularization matrix input information of physical intuition into fitting process which guides the system into the direction of its preference, thus realizes our goal of automatically select clusters according to physical intuitions

For convenience, we define the orbit regularization parameter $\lambda_\alpha = \frac{\sigma^2}{\sigma_\alpha{}^2}$, and coupled regularization parameter $\lambda_{\alpha\beta} = \lambda_{\beta\alpha} = \frac{\sigma^2}{\sigma_{\alpha\beta}{}^2}$, and Λ becomes:

$$\Lambda = \begin{pmatrix} \lambda_\alpha + \sum_{\beta|\beta\neq\alpha} \lambda_{\alpha\beta} & -\lambda_{\alpha\beta} & \cdots \\ -\lambda_{\beta\alpha} & \lambda_\beta + \sum_{\beta|\beta\neq\alpha} \lambda_{\alpha\beta} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (36)$$

Actually, if we restrict $\lambda_{\alpha\beta} = 0$ and $\lambda_\alpha \in \{0, \infty\}$ the process returned to a normal cluster expansion, and CV technique can be used to optimize the cluster selection. However, this is a process lack of physical insights. In order to have physical intuition helping the cluster selection process, we need λ_α and $\lambda_{\alpha\beta}$ as a function of configuration of clusters. This comes the generator function of λ_α and $\lambda_{\alpha\beta}$.

Here we introduce the generator function used for this study.

As we have indicated, the λ_α corresponds to intuition 2 and $\lambda_{\alpha\beta}$ corresponds to intuition 3. We then find the λ_α should be an increase function of n and r (number of sites and maximum distance between sites) $\lambda_\alpha(n, r)$.

Below we will see, just follow the general convergence condition of cluster expansion, physical intuition 2 is naturally satisfied.

The convergence of the cluster expansion can be expressed as:

For every $\varepsilon > 0$, there exist $\{n_{cut}, r_{cut}\}$ such that the expectation value, which mapping over all the square of predicted value of very training structure, is smaller than ε

$$\langle \mathrm{E}\left\{\left[\sum_{\vec{\mathrm{b}}\notin \mathrm{B}_{\mathrm{cut}}} \mathrm{V}_{\vec{\mathrm{b}}} \Phi_{\vec{\mathrm{b}}}(\vec{\mathrm{S}})\right]^2\right\}\rangle < \varepsilon \qquad (37)$$

Where B_{cut} is the set of all cluster functions dependent on clusters of no more than n_{cut} sites and with a distance of no more than r_{cut} from each other.

Given the number of training structures, this is a general condition that the cluster expansion can always reach acceptable accuracy with large enough cutoff $\{n_{cut}, r_{cut}\}$.

Since the clusters are orthonormal (as in **section 2.1**) to each other and the mean value of ECIs should be 0: $E(V_{\vec{b}}) = 0$

Adding to the condition that the site variable may take discrete values, the convergence of cluster expansion can be simplified as:

For every $\varepsilon > 0$, there exist $\{n_{cut}, r_{cut}\}$, such that:

$$\langle \sum_{\vec{b}\notin B_{cut}} \delta_{\vec{b}}{}^2 \rangle < \varepsilon \qquad (38)$$

Where $\delta_{\vec{b}}{}^2$ is the variance of the prior distribution of $V_{\vec{b}}$.

In the limit of $n_{cut} \to \infty$ and $r_{cut} \to \infty$, if $\delta_{\vec{b}}{}^2$ decrease more rapidly than the number of clusters in B_{cut} increase. The convergence condition stand, and as $r_{cut} \to \infty$, the number of clusters in B_{cut} per formular unit is approximately proportional to $(\gamma_1 r_{cut})^{\gamma_2 n_{cut}}$

Where γ_1 is a scale factor and γ_2 depends on the number of periodic dimensions.

So that: $\delta_{\vec{b}}{}^2 \cdot (\gamma_1 r_{cut})^{\gamma_2 n_{cut}} < 1$

Thus $\delta(n, r)$ have the form of $\delta(n, r) = (\gamma_1 r)^{\gamma_2 n}$

More generally

$$\lambda_\alpha(\mathrm{n}, \mathrm{r}) = \gamma_1(\gamma_2 \mathrm{r} + \gamma_3 + 1)^{\gamma_4 \mathrm{n} + \gamma_5} \qquad (39)$$

Or we can write:

$$\lambda_\alpha(\mathrm{n}, \mathrm{r}) = \begin{cases} \gamma_1, & \text{for one point cluster} \\ \gamma_2 \mathrm{e}^{\gamma_3 \mathrm{r} + \gamma_4 \mathrm{n}}, & \text{for clusters larger than one point} \end{cases} \qquad (40)$$

We can see here, just following the natural convergence condition of cluster expansion, the physical intuition 2 is satisfied.

For generating $\lambda_{\alpha\beta}$, we basically come with the idea that $\lambda_{\alpha\beta}$ represents physical intuition 3 and we should decide the degree to which we treated the different cluster functions to be

similar. For simplicity of this research, we simply separate the relationship of cluster functions into two classes: congruent clusters and non-congruent clusters.

The congruent clusters are cluster functions for which the non-constant site functions are related by an isometry. In details, there must be a way to map the sites of one cluster onto the sites of another that preserves all distances and angles between sites.

If two clusters are not congruent, we simply define them as non-congruent clusters.

Then just because $\lambda_{\alpha\beta}$ represent the relationship between cluster α and cluster β, we simply set:

$\lambda_{\alpha\beta} = 0$, if α and β are non-congruent

$\lambda_{\alpha\beta} = \gamma_s\lambda_\alpha = \gamma_s\lambda_\beta$, if α and β are congruent.

So far, to generate λ_α and $\lambda_{\alpha\beta}$, we need to have 6 parameters $\{\gamma_1, \gamma_2, \gamma_3 \dots, \gamma_s\}$

And we apply a CV minimization scheme to fit them, that is, using minimization algorithm such as congruent conjugate method to find the set of parameters $\{\gamma_1, \gamma_2, \gamma_3 \dots, \gamma_s\}$. Which minimize the CV score of cluster expansions. We can see, compared with normal cluster expansion under CV scheme, the searching process is greatly simplified, from a randomly tiral of adding or removing clusters into a fitting process of parameters $\{\gamma_1, \gamma_2, \gamma_3 \dots, \gamma_s\}$.

2.3 Exact expression for training structure selection

In **section 2.1**, we see that the variance of the prediction error for a structure S is given by sMs^T where M is the convariance matrix and covaiance reduction is $Tr(M)$.

For normal cluster expansion with least square fitting $M = (X^T X)^{-1} e^2$,

They then applied an assumption of the distribution of randomly selected structure and then reach the expression of

$$\mathrm{E}\left[\mathrm{Var}\left[\overline{\mathrm{E}}_{\mathrm{i}}\right]\right] \propto \mathrm{Tr}(\mathrm{M}) \qquad (41)$$

However, since we may not likely to select structure randomly, especially for Bayesian cluster expansion, some structures are favored, here, a new process of selecting training structure is needed. Tim Muller et all proposed an exact expression of estimate the expected variance of prediction error which allows different assumption of distribution of training structure taken into account (prior distribution) [14].

The main idea of this treatment is that the expected variance for the predicted property value for a given population of structure, instead of integral in continues space and comes in the form of $Tr(M)$., can be expressed as:

$$\langle \mathrm{xMx^T} \rangle_{\mathrm{pop}} = \mathrm{M}: \langle \mathrm{xx^T} \rangle_{\mathrm{pop}} \qquad (42)$$

Where $\langle \quad \rangle_{pop}$ indicates the average value over all structures in the population and : represent Frobenius innder product $A: B = \sum_i \sum_j A_{ij} B_{ij}$ And we define $\langle x x^T \rangle_{pop}$ to be domain Matrix D.

Removing the mean square error e^2, we reach a factor $\tau = (X^T X)^{-1}: D$ for normal cluster expansion.

Also, following the fact $M = (X^T X)^{-1} e^2$ for normal cluster expansion , we can easily comes to $M = (X^T X + \Lambda)^{-1} e^2$ for Bayesian cluster expansion where $\Lambda^{-1} e^2$ is the covariance matrix for a multivariate Gaussian prior distribution of ECI values. Then $\tau = (X^T X + \Lambda)^{-1} : D$ for Bayesian cluster expansion.

This is a good measurement of how well a given set of training structures reduces prediction error.

If we get the expression for D, we then can undertake a process of structure selection which minimize the variance error of predicted energy, just as the same goal in **section 2.1**.

Fortunately, in some conditions, with certain prior distribution of structures, specific expression for D exist.

Since $D = \langle x x^T \rangle_{pop}$, it is natural to write D in the form

$$\mathrm{D}_{\alpha\beta} = \langle \mathrm{x}_\alpha \mathrm{x}_\beta \rangle_{\mathrm{pop}} = \langle \varphi_\alpha(\mathrm{s}) \varphi_\beta(\mathrm{s}) \rangle_{\mathrm{pop}} = \frac{\sum_{\mathrm{b}\in\alpha} \sum_{\mathrm{b}\prime\epsilon\beta} \langle \varphi_{\mathrm{b}}(\mathrm{s}) \varphi_{\mathrm{b}\prime}(\mathrm{s}) \rangle_{\mathrm{pop}}}{\mathrm{N}_\alpha \mathrm{N}_\beta} \quad (43)$$

Where N_α and N_β are the number of cluster functions in orbits α and β.

Given prior distribution of structures, such form can be transformed into simplified form. For example, for the interest of the study, where not all structures are of equal interest, we make an assumption that all included cluster functions are dependent on a finite number of sites and the crystal has infinite periodicity. Since the overlap between clusters are vanishingly small relative to the size of $N_\alpha N_\beta$, in the crystal limit with infinite periodicity, we can in effect regard

all clusters are independent with each other. Thus

$$D_{\alpha\beta} = \frac{\sum_{b\in\alpha}\sum_{b\prime\epsilon\beta}\langle\varphi_b(s)\varphi_{b\prime}(s)\rangle_{pop}}{N_\alpha N_\beta} = \frac{\sum_{b\in\alpha}\sum_{b\prime\epsilon\beta}\langle\varphi_b(s)\rangle_{pop}\langle\varphi_{b\prime}(s)\rangle_{pop}}{N_\alpha N_\beta}$$

$$= \frac{N_\alpha N_\beta\langle\varphi_b(s)\rangle_{pop}\langle\varphi_{b\prime}(s)\rangle_{pop}}{N_\alpha N_\beta} = \langle x_\alpha\rangle_{pop}\langle x_\beta\rangle_{pop} \quad (44)$$

For the sought of binary alloy systems, where only allowed value for site variable are +1 and -1, we can see $\langle x_\alpha\rangle_{pop} = (2c-1)^{N_\alpha}$, where c is the concentration thus

$$D_{\alpha\beta} = (2c-1)^{N_\alpha+N_\beta} \quad (45)$$

Practically, knowing Domain Matrix of a system, we can estimate expectation error τ of any known training set. With such target parameter, optimized training set can be found using minimization algorithm such as simulated annealing.

3. Proposal of lattice parameter dependent cluster expansion

3.1 Problem statement

Above we review the basic context and current progress of cluster expansion method, applied in DFT-CE-MC process. As we introduced in **section 1**, a process of studying thermodynamic properties of large supercell structures is established. Compared with high-throughput DFT computation which also showed ability in predict new structures, DFT-CE-MC process brought an advantage of inexpensive computation effort and the ability to study system with long scale periodic. For example, for studying high entropy alloy, a large supercell is required in order to

allow enough degree of freedom for configuration change [15]. High through-put DFT can not deal with this since its ability in studying large scale structures is limited. Another example is in low symmetry (low dimensional) systems such as surface or nanoparticles. Since the expense of DFT increase greatly when periodic in some dimensions are broken, the computation expense in calculating large scale structure of low dimensional system is unaffordable.

However, DFT-CE-MC process also has its own shortages. One of them, which we will mainly cover in this thesis is the limitation of this process in studying strain effects. Especially for the case where strain effect can be accurately predicted through known rules (such as quadratic rules), but since parameters of the rule is different for different structure, cluster expansion can not give Hamiltonian with different strains. Thus a MC simulation with consideration of strain effect is impossible.

Current solution to this problem is collect another collection of training data(same amount as with non-strain training data) using DFT if new strain is taken into account. As DFT calculation is the most expensive step of the whole DFT-CE-MC process. Calculation of structures with strain effect in consideration is a rather expensive process.

The bad news is that, current progress in alloy catalyst demand this calculation for a reason of mismatch of composition between surface and bulk region of catalyst. The reason of this composition mismatch is due to the chemical environment of catalyst and specific properties difference between surface region and bulk region as we have indicated in **section 1**. This effect may provide promising new catalyst structures experimentally, however, computational it result in severe problems.

For DFT calculation, different composition will result in different lattice parameters of DFT ground state structures. However, in reality, the surface area will experience a strain compared with their ground state lattice parameters in order to match the lattice parameter of bulk region, and resulting an energy shift compared with DFT result. For the purpose of our study, we basically use slab structures to model surface and the lattice parameter of those structures should be fixed to the bulk lattice parameters, instead of using their own ground state lattice parameters directly from by DFT molecular dynamic result. Since there are a range of bulk lattice we are interested in, DFT result for different lattice parameters for the same slab structure should be provide, which greatly expand our calculation amount and is practical.

As the convention DFT-CE-MC approach with strain effect in consideration is awkward and expensive. Aside that high throughput DFT calculation meet difficulty with computation expense for low dimensional system. Computational study of composition shift of catalyst surface became a difficult problem.

A possible approach is to only consider a small region of composition shift where the strain effect is so small that can be ignored. This comes the work of Liang Cao et al [16]. In their study, with experimental result that Ni3Pt skeleton structure is favored, they scanned a small region of composition shifted around Ni3Pt and using DFT-CE-MC process studied the ground states and catalyst activity map under a grand canonical ensemble, indicating the power of DFT-CE-MC approach.

Here, in this thesis, we propose that, with a small modification of cluster expansion method, information of prediction rule such as quadratic rule can become input data and a lattice

parameter dependent cluster expansion method is available for DFT-CE-MC approach, which solve the current difficulty we meet in surface alloy catalyst study with consideration of composition shift.

3.2 Method Proposal

The basic idea of lattice parameter dependent cluster expansion is that ECIs can not only contain information of energies related with configuration, but also energies related with strains.

A hint is that in small region of parameter change (in-plane strain), the formation energy of alloy surface follows a quadratic rule. Then, if we force our ECIs to obey such a rule with respect to strains, it is possible to predict energies of different strain sets through cluster expansion method.

Below we will firstly look over this quadratic strain rule in DFT result, and then proposed a scheme to set this rule as input information into cluster expansion method.

3.2.1 Quadratic Strain rule of DFT energy

In the theory of linear elasticity, the elastic stress-strain response in solid is normally described with up to second-order elastic constant, which means the Taylor expansion expression of energy-strain relationship is truncated up to the second order, namely, the quadratic rule [17]. The quadratic rule stands when strains are small enough so that effect of higher-order elastic constants can be ignored. Generally, when strain is larger than 5%, for alloy, third-order elastic

constant should be taken into account. Thus it is necessary to test the goodness of quadratic rule for the system we are interested in.

For the purpose of testing quadratic strain rule of DFT energy of surface alloy, we calculated the energy of a slab model of AuPd2 alloy (with FCC lattice structure) under different strain sets.

The DFT calculation is implemented with VASP package and PBE projector-augmented wave(PAW) exchange-correlation functional [18, 19]. The Brillouin zone was sampled with a $12 \times 12 \times 1$ Monkhorst-Pack k-point mesh for 4×4 (111) cell calculations. The calculation is undertaken with normal precise and the convergence criteria for the electronic self-consistent iteration and the ionic relaxation loop were set to be 10^{-4} eV and 10^{-3} eV eV, respectively.

The model used in this calculation is 7 layers of slab of (111) face. For the supercell used in calculation, the slab has periodic boundary condition along (111) face and is separated by a $20\dot{A}$ vacuum along $[111]$ direction so that the interaction between the two surfaces of the slab is small enough during the calculation.

The strain set for of choice in this test is as follows:

Base Structure	Pd	Pd3Au	PdAu	PdAu3	Au
Lattice Parameter($\dot{A}$)	3.9115	3.9616	4.0116	4.0617	4.1118
Strain Compared with Original Lattice Parameter	-0.01678	-0.00419	0.00839	0.02097	0.03356

Where we select 5 values of composition as base structure composition, and force the Pd-Au slab we have to match the equilibrium lattice parameters (also calculated by DFT) of these base

structure. As showed in above table, the strain sets are within 5%(small strain sets) where the quadratic rule is expected to stand.

The fitting result is showed in the figure below, the p-value for the quadratic fitting is 5.08×10^{-6}, which is very small, indicating the quadratic model is well accepted.

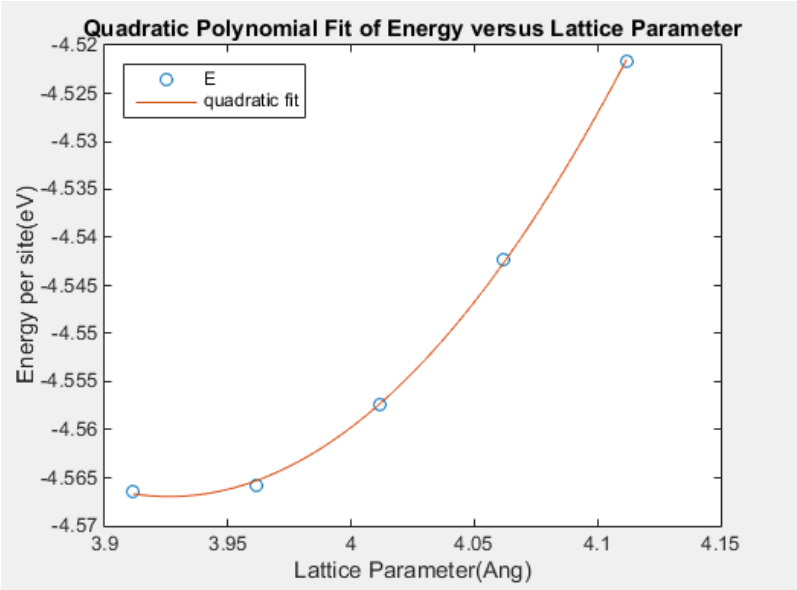


Fig 3.1 Quadratic fitting test of energy-strain relationship

This result suggests, for small strain sets, the truncation to second order term of Taylor expansion is accurate enough. Thus it is safe to consider the strain-stress response only with second-order elastic constant.

3.2.2 Lattice parameter dependent cluster expansion

Above, we have check that the quadratic rule works well in the surface system we are interested in. We need to input this information into cluster expansion approach.

Considering the idea of forcing the ECIs following the quadratic rule, basically, we have 2 choice:

1. Force every cluster following the same rule as in total energy

2. Fitting ECIs of clusters independently so that each cluster obeys the quadratic rule with its own parameters

The first choice is quite straightforward and easy to undertake, but without much value. Since different structures obey different quadratic rules, if we apply the same rule to all ECIs, every time we change our structures, the rule of ECIs will also change. This means we can not use the same set of ECIs for all possible strains within our region of interest.

What we actually need is to store the quadratic rule information in the ECIs, so that expanding our space of configuration with a new dimension with respect to strains, which comes the second choice. Now we makes an assumption that if certain sets of ECIs and their specific quadratic rule with respect to strain was found, the total system of configuration space(with strain axis) can be estimated accurately.

Below, we will do two things to verify this assumption:

1. Theoretically analyze
2. Implementing this idea and see if it works.

3.2.2.1 Theoretically analyze

What we want to do is to expand input matrix X 3 times

$$\mathrm{X_{new}} = (\mathrm{X}|\mathrm{SX}|\mathrm{S^2X}) \quad (46)$$

Where S is the in-plane strain ϵ we applied isotropically on the slab structure. Practically, we can also use the expression of $S = \epsilon + 1$, which will guarantee that the expanded term is not ignorable.

$$X_{new} = \left[\begin{array}{cccc|cccc|cccc} x_{i\alpha} & x_{i\beta} & x_{i\gamma} & \cdots & Sx_{i\alpha} & Sx_{i\beta} & Sx_{i\gamma} & \cdots & S^2x_{i\alpha} & S^2x_{i\beta} & S^2x_{i\gamma} & \cdots \\ x_{j\alpha} & x_{j\beta} & x_{j\gamma} & \cdots & Sx_{j\alpha} & Sx_{j\beta} & Sx_{j\gamma} & \cdots & S^2x_{j\alpha} & S^2x_{j\beta} & S^2x_{j\gamma} & \cdots \\ x_{k\alpha} & x_{k\beta} & x_{k\gamma} & \cdots & Sx_{k\alpha} & Sx_{k\beta} & Sx_{k\gamma} & \cdots & S^2x_{k\alpha} & S^2x_{k\beta} & S^2x_{k\gamma} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \end{array}\right] \quad (47)$$

Then we use this new matrix as input matrix and applied it into least square fitting process.

$$\vec{V} = (X^T X)^{-1}\vec{y} \quad (48)$$

Assume the total number of cluster functions of a convention cluster expansion is N, then we can view the configuration space of cluster functions to be a N dimensional Dirac space where each coordinate is orthogonal to one another.

As indicated in **section 2.1**, we see the cluster functions are orthogonal between each other, so that the least square fitting of ECIs is actually a fitting of a line in a Dirac space with N orthogonal axis.

We can consider our expansion of input matrix equal to adding up 2N more freedom where each index of new freedom is Sx, or S^2x.

Here is the question that after adding new 2N axis, if the configurational space is still Dirac (coordinates are orthogonal to each other) so that the least square fitting process is still the same as before?

Actually this is easy to analyze, since

$$\langle \phi_{\alpha s}, \phi_{\beta s'} \rangle = \delta_{\alpha\beta}\delta_{ss'} \tag{49}$$

We have:

$$\langle \phi_{\alpha s}, S\phi_{\beta s'} \rangle = \int_{-a}^{a} S dS = 0 \tag{50}$$

$$\langle \phi_{\alpha s}, S^2\phi_{\beta s'} \rangle = \int_{-a}^{a} S^2 dS = k\,\delta_{\alpha\beta}\delta_{ss'} \tag{51}$$

$$\langle S\phi_{\alpha s}, S^2\phi_{\beta s'} \rangle = \int_{-a}^{a} S^3 dS = 0 \tag{52}$$

Note that the expressions above stand only when the strain sets are choose mirror symmetrically around 0.

Since $\langle \phi_{\alpha s}, S^2\phi_{\alpha s} \rangle = k\delta_{\alpha\alpha}\delta_{ss} = k$, where k is a constant known as overlap integrals, $\phi_{\alpha s}, S^2\phi_{\beta s'}$ are not orthogonal to each other, so that our new basis is non orthogonal. In this case, although the basic least-square fitting process works, we should be careful when applying some specific technique based on orthogonal basis. For example, if the amount of our training set is very large, we can see our matrix of $X^T X$ can be view as diagonal matrix:

$$X^T X = \begin{bmatrix} \sum_i^N x_{i\alpha}x_{i\alpha} & \sum_i^N x_{i\alpha}x_{i\beta} & \cdots \\ \sum_i^N x_{i\beta}x_{i\alpha} & \sum_i^N x_{i\beta}x_{i\beta} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \xrightarrow{N\to\infty} \begin{bmatrix} 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{53}$$

However, after we expand the input matrix of X, since some of the basis function is not orthogonal with each other, the situation is different:

$$X^T X =$$

$$\begin{bmatrix}
\sum_i^N x_{i\alpha}x_{i\alpha} & \sum_i^N x_{i\alpha}x_{i\beta} & \cdots & \sum_i^N x_{i\alpha}Sx_{i\alpha} & \sum_i^N x_{i\alpha}Sx_{i\beta} & \cdots & \sum_i^N x_{i\alpha}S^2x_{i\alpha} & \sum_i^N x_{i\alpha}S^2x_{i\beta} & \cdots \\
\sum_i^N x_{i\beta}x_{i\alpha} & \sum_i^N x_{i\beta}x_{i\beta} & \cdots & \sum_i^N x_{i\beta}Sx_{i\alpha} & \sum_i^N x_{i\beta}Sx_{i\beta} & \cdots & \sum_i^N x_{i\beta}S^2x_{i\alpha} & \sum_i^N x_{i\beta}S^2x_{i\beta} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\
\sum_i^N Sx_{i\alpha}x_{i\alpha} & \sum_i^N Sx_{i\alpha}x_{i\beta} & \cdots & \sum_i^N S^2x_{i\alpha}x_{i\alpha} & \sum_i^N S^2x_{i\alpha}x_{i\beta} & \cdots & \sum_i^N S^3x_{i\alpha}x_{i\alpha} & \sum_i^N S^3x_{i\alpha}x_{i\beta} & \cdots \\
\sum_i^N Sx_{i\beta}x_{i\alpha} & \sum_i^N Sx_{i\beta}x_{i\beta} & \cdots & \sum_i^N S^2x_{i\beta}x_{i\alpha} & \sum_i^N S^2x_{i\beta}x_{i\beta} & \cdots & \sum_i^N S^3x_{i\beta}x_{i\alpha} & \sum_i^N S^3x_{i\beta}x_{i\beta} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\
\sum_i^N S^2x_{i\alpha}x_{i\alpha} & \sum_i^N S^2x_{i\alpha}x_{i\beta} & \cdots & \sum_i^N S^3x_{i\alpha}x_{i\alpha} & \sum_i^N S^3x_{i\alpha}x_{i\beta} & \cdots & \sum_i^N S^4x_{i\alpha}x_{i\alpha} & \sum_i^N S^4x_{i\alpha}x_{i\beta} & \cdots \\
\sum_i^N S^2x_{i\beta}x_{i\alpha} & \sum_i^N S^2x_{i\beta}x_{i\beta} & \cdots & \sum_i^N S^3x_{i\beta}x_{i\alpha} & \sum_i^N S^3x_{i\beta}x_{i\beta} & \cdots & \sum_i^N S^4x_{i\beta}x_{i\alpha} & \sum_i^N S^4x_{i\beta}x_{i\beta} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots
\end{bmatrix}$$

$$\xrightarrow{N\to\infty} \begin{bmatrix}
1 & 0 & \cdots & 0 & 0 & \cdots & k_1 & 0 & \cdots \\
0 & 1 & \cdots & 0 & 0 & \cdots & 0 & k_1 & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\
0 & 0 & \cdots & k_1 & 0 & \cdots & 0 & 0 & \cdots \\
0 & 0 & \cdots & 0 & k_1 & \cdots & 0 & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\
k_1 & 0 & \cdots & 0 & 0 & \cdots & k_2 & 0 & \cdots \\
0 & k_1 & \cdots & 0 & 0 & \cdots & 0 & k_2 & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots
\end{bmatrix} \quad (54)$$

Thus the $X^T X$ is no longer diagonal matrix. Note that since $X^T X$ matrix expand 3 times in each direction, the $\vec{y}$ vector will also expand 3 times, i.e. the size of our training data need to expand 3 times for normal cluster expansion. For Bayesian cluster expansion, since the number of training structures do not need to be equal or larger than the number of clusters included, it is not necessary to expand the size of training data accordingly.

After we get ECIs through fitting process, the final expression for cluster expansion will just remain the same:

$$F(\vec{s}) = \sum_{\vec{b}} \left(V_{\vec{b}}^{(0)} \Phi_{\vec{b}} + V_{\vec{b}}^{(1)} S \Phi_{\vec{b}} + V_{\vec{b}}^{(0)} S^2 \Phi_{\vec{b}} \right) \quad (55)$$

Where we see every cluster is fitted to its unique quadratic rule, and the ECIs can be viewed as fitting parameters.

For Bayesian approach

$$\vec{\mathrm{V}} = (\mathrm{X}^{\mathrm{T}}\mathrm{X} + \Lambda)^{-1}\vec{\mathrm{y}} \tag{56}$$

We also need to know the expression of Λ after we expands the input matrix, the method to do this is quite straightforward:

Since $X^T X =$

$$\begin{bmatrix}
\sum_i^N x_{i\alpha}x_{i\alpha} & \sum_i^N x_{i\alpha}x_{i\beta} & \cdots & \sum_i^N x_{i\alpha}Sx_{i\alpha} & \sum_i^N x_{i\alpha}Sx_{i\beta} & \cdots & \sum_i^N x_{i\alpha}S^2x_{i\alpha} & \sum_i^N x_{i\alpha}S^2x_{i\beta} & \cdots \\
\sum_i^N x_{i\beta}x_{i\alpha} & \sum_i^N x_{i\beta}x_{i\beta} & \cdots & \sum_i^N x_{i\beta}Sx_{i\alpha} & \sum_i^N x_{i\beta}Sx_{i\beta} & \cdots & \sum_i^N x_{i\beta}S^2x_{i\alpha} & \sum_i^N x_{i\beta}S^2x_{i\beta} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\
\sum_i^N Sx_{i\alpha}x_{i\alpha} & \sum_i^N Sx_{i\alpha}x_{i\beta} & \cdots & \sum_i^N S^2x_{i\alpha}x_{i\alpha} & \sum_i^N S^2x_{i\alpha}x_{i\beta} & \cdots & \sum_i^N S^3x_{i\alpha}x_{i\alpha} & \sum_i^N S^3x_{i\alpha}x_{i\beta} & \cdots \\
\sum_i^N Sx_{i\beta}x_{i\alpha} & \sum_i^N Sx_{i\beta}x_{i\beta} & \cdots & \sum_i^N S^2x_{i\beta}x_{i\alpha} & \sum_i^N S^2x_{i\beta}x_{i\beta} & \cdots & \sum_i^N S^3x_{i\beta}x_{i\alpha} & \sum_i^N S^3x_{i\beta}x_{i\beta} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\
\sum_i^N S^2x_{i\alpha}x_{i\alpha} & \sum_i^N S^2x_{i\alpha}x_{i\beta} & \cdots & \sum_i^N S^3x_{i\alpha}x_{i\alpha} & \sum_i^N S^3x_{i\alpha}x_{i\beta} & \cdots & \sum_i^N S^4x_{i\alpha}x_{i\alpha} & \sum_i^N S^4x_{i\alpha}x_{i\beta} & \cdots \\
\sum_i^N S^2x_{i\beta}x_{i\alpha} & \sum_i^N S^2x_{i\beta}x_{i\beta} & \cdots & \sum_i^N S^3x_{i\beta}x_{i\alpha} & \sum_i^N S^3x_{i\beta}x_{i\beta} & \cdots & \sum_i^N S^4x_{i\beta}x_{i\alpha} & \sum_i^N S^4x_{i\beta}x_{i\beta} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots
\end{bmatrix}$$

If we assume $A = \begin{bmatrix} \sum_i^N x_{i\alpha}x_{i\alpha} & \sum_i^N x_{i\alpha}x_{i\beta} & \cdots \\ \sum_i^N x_{i\beta}x_{i\alpha} & \sum_i^N x_{i\beta}x_{i\beta} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$

We can simplified the expression into

$$\mathrm{X}^{\mathrm{T}}\mathrm{X} = \begin{bmatrix} \mathrm{A} & \mathrm{SA} & \mathrm{S}^2\mathrm{A} \\ \mathrm{SA} & \mathrm{S}^2\mathrm{A} & \mathrm{S}^3\mathrm{A} \\ \mathrm{S}^2\mathrm{A} & \mathrm{S}^3\mathrm{A} & \mathrm{S}^4\mathrm{A} \end{bmatrix} \tag{57}$$

We can see the matrix of $X^T X$ can be divided into 9 block and each block can be expressed as $S^x A$, where $x = \{0,1,2,3,4\}$. Since the regularization matrix Λ contains the information of prior distribution which gives each element in $X^T X$ a weight during the the accumulate process of $(X^T X + \Lambda)^{-1}\vec{y}$, it is safe to assume that for every block of $X^T X$, the regularization matrix Λ should be approximately the same:

$$\Lambda_{\text{new}} = \begin{bmatrix} \Lambda & \Lambda & \Lambda \\ \Lambda & \Lambda & \Lambda \\ \Lambda & \Lambda & \Lambda \end{bmatrix} \tag{58}$$

Another option is that, since Sx and $S^2 x$ are also orthogonal axis in configuration space, we may treat the prior distribution of Sx and $S^2 x$ different from x, but not too much different. We can assign order of stains with a new parameter times the initial elements in regularization matrix then:

$$\Lambda_{\text{new}} = \begin{bmatrix} \kappa_0\kappa_0\Lambda & \kappa_0\kappa_1\Lambda & \kappa_0\kappa_2\Lambda \\ \kappa_0\kappa_1\Lambda & \kappa_1\kappa_1\Lambda & \kappa_2\kappa_1\Lambda \\ \kappa_0\kappa_2\Lambda & \kappa_1\kappa_2\Lambda & \kappa_2\kappa_2\Lambda \end{bmatrix} \tag{59}$$

Then we will add 3 new parameters together with the parameters $\{\gamma_1, \gamma_2, \gamma_3 \dots, \gamma_5, \gamma_s\}$ for generating regularization matrix Λ. Since we basically use conjugate gradient method searching for these parameters minimizing CV score, we do not need to worry about the influence it may cause to our searching process when adding 3 new parameters.

3.2.2.2 Implementing of the method

Above we have theoretically analyzed the validity of this treatment, now we will implement this method into practical study and see if it works.

We basically implement this idea in Bayesian cluster expansion based on the source code <matsci> written by Dr. Tim Mueller. Below, we firstly explain the structure of <matsci> and then show how we add our new method into it.

Since the process of building cluster functions are quite standard and we do not need to modify it in this research, here we basically explain the fitting process of <matsci>, excluding the cluster expansion building process.

After the cluster expansion is built for the training set, we have input matrix X where each line contains the value of cluster functions belonging to the same structure and each column contains value of same cluster function in different structures.

Also we have array of formation energy of each structure in the training set, calculated by DFT, $\vec{y}$.

Then, the regularization matrix Λ is generated using the generating function:

$$\lambda_\alpha(n,r) = \begin{cases} \gamma_1, & for\ one\ point\ cluster \\ \gamma_2 e^{\gamma_3 r + \gamma_4 n}, & for\ clusters\ larger\ than\ one\ point \end{cases}$$

$\lambda_{\alpha\beta} = 0$, if α and β are non-congruent

$\lambda_{\alpha\beta} = \gamma_s \lambda_\alpha = \gamma_s \lambda_\beta$, if α and β are congruent.

Where n is the number of atoms in a cluster, r is the scale of cluster. $\{\gamma_1, \gamma_2, \gamma_3 \dots, \gamma_4, \gamma_s\}$ are unknown parameters we need to find.

We give an initial guess of 5 parameters and the ECIs then can be calculated through

$$\vec{\mathrm{V}} = (\mathrm{X^T W X})^{-1} \mathrm{X^T W} \vec{\mathrm{y}} \tag{60}$$

and the estimated energy should be

$$\overline{\mathrm{E}} = \sum_{\vec{\mathrm{b}}} \mathrm{V}_{\vec{\mathrm{b}}} \Phi_{\vec{\mathrm{b}}} \tag{61}$$

The CV score then can be calculated through

$$(\mathrm{CV})^2 = \mathrm{n}^{-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{E_i} - \overline{\mathrm{E}}_{(\mathrm{i})})^2 \tag{62}$$

This whole process can be viewed as a large function

$$(\mathrm{CV})^2 = \mathrm{f}(\gamma_1, \gamma_2, \gamma_3 \dots, \gamma_4, \gamma_\mathrm{s}) \tag{63}$$

And with a conjugate gradient method in order to find the exact value of the parameters which minimize the CV score. The satisfactory ECIs can be calculated.

For the implementation of lattice parameter dependent cluster expansion, we just need to change the generate function of X and Λ so that allow them expand in the form of

$$\mathrm{X_{new}} = (\mathrm{X|SX|S^2X}) \tag{64}$$

$$\Lambda_{\text{new}} = \begin{bmatrix} \kappa_0\kappa_0\Lambda & \kappa_0\kappa_1\Lambda & \kappa_0\kappa_2\Lambda \\ \kappa_0\kappa_1\Lambda & \kappa_1\kappa_1\Lambda & \kappa_2\kappa_1\Lambda \\ \kappa_0\kappa_2\Lambda & \kappa_1\kappa_2\Lambda & \kappa_2\kappa_2\Lambda \end{bmatrix} \quad (65)$$

And put into function of

$$(\text{CV})^2 = \text{f}(\gamma_1, \gamma_2, \gamma_3 \dots, \gamma_5, \gamma_s) \quad (66)$$

Which yields a new function of

$$(\text{CV})^2 = \text{f}(\gamma_1, \gamma_2, \gamma_3 \dots, \gamma_5, \gamma_s, \kappa_0, \kappa_1, \kappa_2) \quad (67)$$

With this single treatment, a normal cluster expansion process will become cluster expansion with lattice parameter dependents.

3.3 Method test

3.3.1 generating training structures

The training structures are generated following the context we mentioned in **section 2.3**, where a minimization algorithm was applied in order to find training structures minimize the parameter τ.

In details, firstly a prim cell of the supercell we are interested was built manually. Such cell is basically a 1×1 cell of an FCC slab structure with surface along (111). Then the prim cell is expanded into supercells with randomly selected miller index along (111) plane. Finally, each site in the supercell is decorated with elements of the binary alloy. For the purpose of this study, sites are decorated following a mirror symmetry along the middle layer of the slab, so that the

middle layer will be in a symmetrical environment and behaves much similar with bulk region. When each structure is generated this way, it will then experience a test controlled by a minimization algorithm, to see if adding this new structure to training set can help minimize the parameter τ. If the structure is rejected, then the above process will be repeat, until enough training structures are collected. At last, the training structures were then applied with an in-plane strain selected from our strain sets of choice. (see Fig 3.2 for an example of structures).

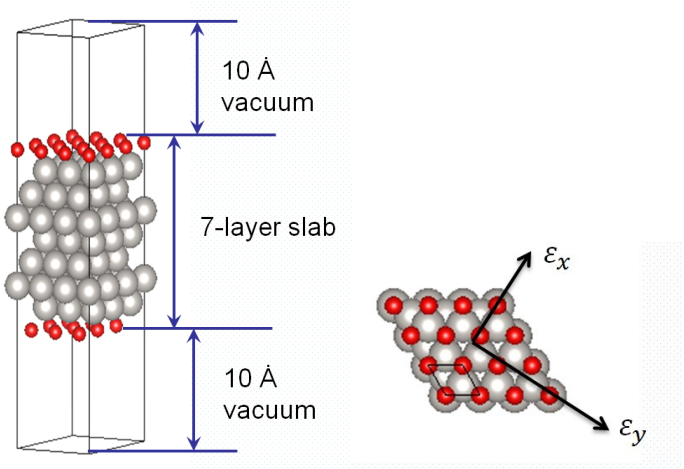


Fig 3.2 Left: side view of an example structure in training set.

Right: top view of the example structure and the strains applied

In practice, we mainly tested our method with 2 systems:

1. Au-Pd (111) slab(7 layers) with up to monolayer of oxygen adsorption
2. Ni-Pt (111) slab(7 layers) with up to monolayer of oxygen adsorption

Au-Pd system with oxygen adsorption:

Since DFT calculation of Au-Pd (111) slab have the most accurate result among these 3 systems, we firstly test this system. Only if the test for this system passed, can we go through the remaining systems. A problem here is that when oxygen coverage is high, surface

reconstruction may happen resulting from the interaction between oxygen atoms and alloy slab. In order to avoid this effect, the coverage of oxygen is limited to lower than 1/3 of all FCC sites on surface, through which the results passed our test of quadratic rule.

The model we used for this system is 7 layers of slab of (111) face. For the supercell used in calculation, the slab has periodic boundary condition along (111) face and is separated by a $20\dot{A}$ vacuum along $[111]$ direction so that the interaction between the two surfaces of the slab is small enough during the calculation.

The composition range of our study is from 1:1 to 0:1 for the ratio of Au:Pd, which is to say, we scan through all Pd rich compositions.

The strain set we used for this test is the equal distance strain set as in **section 3.2.1**

Base Structure	Pd	Pd3Au	PdAu	PdAu3	Au
Lattice Parameter($\dot{A}$)	3.9115	3.9616	4.0116	4.0617	4.1118
Strain Compared with Original Lattice Parameter	-0.01678	-0.00419	0.00839	0.02097	0.03356

The DFT calculation is implemented with VASP package and PBE projector-augmented wave(PAW) exchange-correlation functional [19, 20, 18]. The Brillouin zone was sampled with a $12 \times 12 \times 1$ Monkhorst-Pack k-point mes. The calculation is undertaken with normal precise and the convergence criteria for the electronic self-consistent iteration and the ionic relaxation loop were set to be $10^{-4}\ eV$ and $10^{-3}\ eV$ eV, respectively. For the molecular dynamic part, we fixed the strained lattice parameters of the structure and relaxed the sites within lattice.

We firstly calculated training data of 700 structures with 140 different supercells. For the same supercell, 5 different in-plane strains as listed above were applied to resize the supercell along (111) face. After cluster expansion fitting, we have successfully reached an accuracy of LOOCV score (leave one out cross-validation score) per atom to be: 5.973×10^{-4}eV/atom, which is quite small and meet our requirement, where if we do not use lattice parameter dependent cluster expansion but use a normal cluster expansion, the LOOCV score per atom will reach a value as high as 0.0127eV/atom, which is not acceptable.

Next, we randomly selected 100 training data from the pool of 700 training data and with the help of lattice parameter dependent cluster expansion, the LOOCV score per atom will be 0.001011eV/atom, which is still acceptable.

Ni-Pt system with oxygen adsorption:

After successfully passed the test for system 1, we then undertake the test of the second systems. The difficulty of calculating this system is that since Ni-Pt is magnetic materials, we should take spin effect into account during our DFT calculation, which results in a decrease of accuracy for DFT calculation and the quadratic rule may be hindered. For the purpose of increasing accuracy, we used RPBE functional here instead of PBE functional we used for Au-Pd system [19, 20].

The model we used for this system is just the same as in system 1, but with sites decorated by Ni and Pt. And the composition we scan through should be Pt rich.

Here, we changed the strain set to the one we truly want to apply: just randomly generate strains in the range of (-0.03 0.03). This strain set was then applied to Ni-Pt with oxygen adsorption.

We calculated the training data of 100 different training structures applied with random in-plane strains. After cluster expansion fitting, LOOCV score per atom is 0.006961eV/atom, still acceptable for us.

4. Application of lattice parameter dependent cluster expansion(LPDCE)

4.1 LPDCE in studying Ni-Pt catalyst

As we have seen, LPDCE works perfect in Ni-Pt-oxygen system under small stain range. Here we introduced a process of studying catalyst properties through DFT-LPDCE-MC approach.

The most important thing we want to know for analyzing equilibrium state of a system in given temperature is the relationship of chemical potential, formation energy, composition and configuration. With this information, all other thermodynamic properties can be easily derived. For example, the Gibbs energy-composition graph can be derived since chemical potential difference of alloy elements such as Ni Pt at certain composition is the slope of Gibbs energy-composition curve at that composition point. And once we have the Gibbs energy graph, phase diagram can be generated using a convex-hull technique [21]. Another example is that the adsorption energy of oxygen ΔE_O can be obtained from the difference of formation energy between Ni-Pt slab and Ni-Pt slab with oxygen adsorption. The catalyst activity can then be analyzed using a volcano plot technique [16]:

$$\text{activity} = min(-0.297 + 0.5 \times (\Delta E_O - 1.806), -0.297 + 0.53 \times (1.806 - \Delta E_O))(\text{eV}) \quad (68)$$

Thus we can build a connection of chemical potential activity, composition and configuration. Undertaking a searching for equilibrium states for optimized activity would be possible.

Good news is that this relationship of chemical potential, formation energy and configuration can be built through a Grand Canonical Monte Carlo simulation [22].

Grand Canonical Monte Carlo simulation is Monte Carlo simulation based on Grand Canonical ensemble which allows composition of the system to change. Since adding or removing atoms will result in energy shift with a value of the chemical potential of that atom. The chemical potential is essential in building partition function and probability of a state:

$$\Xi = \sum_{i,N} exp[-\beta E_i + \beta \mu N] \quad (69)$$

$$P(A_i) = \frac{1}{\Xi}(A_i exp[-\beta E_i + \beta \mu N]) \quad (70)$$

Practically, we can manually set chemical potential window and study the equilibrium composition and configuration with this chemical potential window. Then, we undertake a series of Grand Canonical Monte Carlo simulation with chemical potential mapped through the range we are interested and finally get the map of chemical potential-composition-formation energy-configuration of interest.

The Grand Canonical Monte Carlo simulation requires the formation energy E_i should allow composition shift. Normal cluster expansion fails accurately estimating this, and LPDCE works well in this, thus allows the whole system to be accessible.

4.2 Future extension of LPDCE

We still haven't fully explored the potential of the exiting idea of LPDCE in this thesis, there are 2 possible choices of extending this method in the future:

1. Consider not only quadratic rules, but also the effect of higher order elastic constant
2. Consider not only in-plane strain, but also a 3D strain tensor

Extending method in the first way will enable elastic analyze of large strain sets. Note that the strain should still be within elastic limits, since in inelastic region, lattice structure may not remain the same.

Extending method in the second way will enable analyzing of structures under more complex mechanical environment, such as twist and bend. This would be helpful for studying the behavior of low dimensional materials (nanotube, surface) in nano-machine or piezoelectric devices.

Reference

[1] W. Kohn et L. J. Sham, «Self-Consistent Equations Including Exchange and Correlation Effects,» *Phys. Rev.,* vol. 140, pp. A1133--A1138, Nov 1965.

[2] P. Hohenberg et W. Kohn, «Inhomogeneous Electron Gas,» *Phys. Rev.,* vol. 136, pp. B864--B871, Nov 1964.

[3] J. P. Perdew, K. Burke et Y. Wang, «Generalized gradient approximation for the exchange-correlation hole of a many-electron system,» *Phys. Rev. B,* vol. 54, pp. 16533-16539, Dec 1996.

[4] G. Kresse et J. Hafner, «ab initio molecular dynamic for liquid metals,» *physical review b,* vol. 47, p. 558(R), 1993.

[5] G. Kresse et J. Furthmuller, «efficiency of ab-initial total energy calculations for metals and semiconductors using a plane-wave basis set,» *computational materials science,* vol. 6, n° %11, pp. 15-51, 1996.

[6] G. K. a. J. Furthmuller, «efficient interative schemes for ab initio total-energy calculations using a plane-wave basis set,» *physical review b,* vol. 54, p. 11169, 1996.

[7] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón et D. Sánchez-Portal, «The SIESTA method for ab initio order- N materials simulation,» *Journal of Physics: Condensed Matter,* vol. 14, n° %111, p. 2745, 2002.

[8] A. E. Gelfand et A. F. M. Smith, «Sampling-Based Approaches to Calculating Marginal Densities,» *Journal of the American Statistical Association,* vol. 85, n° %1410, pp. 398-409, 1990.

[9] J. M. Sanchez, F. Ducastelle et D. Gratias, «Generalized cluster description of multicomponent systems,» *Physica A: Statistical Mechanics and its Applications,* vol. 128, n° %11, pp. 334-350, 1984.

[10] C. V. Sheth, A. Ngwengwe et P. H. Borcherds, «Least squares fitting of a straight line to a set of data points: II. Parameter variances,» *European Journal of Physics,* vol. 17, n° %16, p. 322, 1996.

[11] A. van de Walle et G. Ceder, «Automating first-principles phase diagram calculations,» *Journal of Phase Equilibria,* vol. 23, n° %14, pp. 348-359.

[12] D. B. Laks, L. G. Ferreira, S. Froyen et A. Zunger, «Efficient cluster expansion for substitutional systems,» *Phys. Rev. B,* vol. 46, pp. 12587-12605, Nov 1992.

[13] T. Mueller et G. Ceder, «Bayesian approach to cluster expansions,» *Phys. Rev. B,* vol. 80, p. 024103, Jul 2009.

[14] T. Mueller et G. Ceder, «Exact expressions for structure selection in cluster expansions,» *Phys. Rev. B,* vol. 82, p. 184107, Nov 2010.

[15] Y. F. Ye, Q. Wang, J. Lu, C. T. Liu et Y. Yang, «High-entropy alloy: challenges and prospects,» *Materials Today ,* pp. - , 2015.

[16] L. Cao et T. Mueller, «Ratioanl Design of Pt3Ni surface structures for the oxygen reduction reaction,» *the journal of physical chemistry,* pp. 17735-17747, 2015.

[17] J. Zhao, J. M. Winey et Y. M. Gupta, «First-principles calculations of second- and third-order elastic constants for single crystals of arbitrary symmetry,» *Phys. Rev. B,* vol. 75, p. 094105, Mar 2007.

[18] G. Kresse et D. Joubert, «From ultrasoft pseudopotentials to the projector augmented-wave method,» *Phys. Rev. B,* vol. 59, pp. 1758-1775, Jan 1999.

[19] J. P. Perdew, K. Burke et M. Ernzerhof, «Generalized Gradient Approximation Made Simple,» *Phys. Rev. Lett.,* vol. 77, pp. 3865-3868, Oct 1996.

[20] Y. Zhang et W. Yang, «Comment on ``Generalized Gradient Approximation Made Simple'',» *Phys. Rev. Lett.,* vol. 80, pp. 890-890, Jan 1998.

[21] A. N. Varchenko, «Evolutions of convex hulls and phase transitions in thermodynamics,» *Journal of Soviet Mathematics,* vol. 52, n° %14, pp. 3305-3325.

[22] D. Adams, «Molecular Physics: An International Journal at the Interface Between Chemistry and Physics,» *Molecular Physics,* vol. 29, n° %11, pp. 307-311, 1975.

[23] V. Stamenkovic, B. S. Mun et etc, «Changing the activity of electrocatalysts for oxygen reduction by tuning the surface electronic structure,» *Chem., Int. Ed.,* vol. 45, pp. 2897-901, 2006.

[24] C. Wang, M. Chi et etc, «correlation between surface chemistry and electrocatalytic properties of monodisperse PtxNi(1-x) nanoparticles,» *adv. funct. mater.,* vol. 21, pp. 147-152, 2011.

Biographical Statement

Le Niu is a Master student in Department of Materials Science and Engineering in Johns Hopkins University. He received a bachelor's degree in materials physics from University of Science and Technology Beijing. His current research interests are in development of cluster expansion method and its implementation in design of PtNi alloy catalysts for oxygen reduction.