Doctoral Dissertation
博士論文

A Noise-Robust Data Assimilation Method for Crystal Structure Prediction Using Powder Diffraction Intensity
(結晶構造予測のための粉末回折強度を用いたノイズロバストデータ同化法)

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2020
令和 2 年 12 月博士（理学）申請
Department of Physics, Graduate School of Science,
The University of Tokyo
東京大学大学院理学系研究科
物理学専攻

Seiji Yoshikawa
吉川　誠司

Abstract

It is a formidable problem to search stable crystal structures with only the information of their compositions. Various methods to systematically and efficiently find the global minimum of the potential energy have been researched and developed. We are developing the structure prediction method to incorporate the experimental data into the theoretical structure search based on data assimilation.

In this method, we define the penalty function as the "difference" between the reference experimental data and the theoretical value for the crystal structure, and perform the multi-objective optimization of the potential energy and the penalty function. In the previous researches [1, 2], they introduce the "crystallinity"-type penalty function of the powder X-ray diffraction (XRD). This penalty function uses only the diffraction angles in the reference data, and works well even with unreliable experimental intensity ratio. On the other hand, this is not suitable for the target structure with low symmetry or the reference diffraction pattern with noise.

In this study, we adopt new "correlation-coefficient"-type penalty function that explicitly includes the diffraction peak intensities, and work on improving the success rate and robustness of the structure search. In test calculations of the structure search for coesite, the new penalty function significantly improves the search efficiency compared to the crystallinity-type one. We verify the noise-robustness of this method by adding artificial noise to the reference diffraction pattern. In test calculations for ϵ-Zn(OH)$_2$, we confirm the effectiveness of the hybrid penalty function of XRD and the neutron diffraction (ND), which can detect hydrogen positions. In a practical application to high-pressure synthesized Al-Ca-H system, we discover new $\mathrm{Al}_{12}\mathrm{Ca}_{20}\mathrm{H}_{76}$ structure based on data assimilation.

Contents

Chapter 1 Introduction 1

1.1 Crystal structure determination problem 1

1.2 Purpose of the study 2

1.3 Organization of the thesis 2

Chapter 2 Methods 4

2.1 Structure prediction method based on data assimilation 4

2.2 Potential energy calculation 9

2.3 Introduction of the penalty function 12

2.4 Structure optimization 14

2.5 Judging success or failure using fingerprint 17

Chapter 3 Results and Discussion 20

3.1 Comparison of the correlation-coefficient-type and crystallinity-type penalty functions 20

3.2 System containing hydrogen 31

3.3 New hydride in high-pressure synthesized Al-Ca-H system 43

Chapter 4 Conclusion and Outlook 50

4.1 Outcomes of the correlation-coefficient-type penalty function 50

4.2 Expandability of the structure prediction method based on data assimilation 51

Acknowledgments 52

Bibliography 53

Appendix A Calculation of the powder diffraction pattern 56

A.1 Formulation of the powder diffraction pattern 56

A.2 Diffraction peak smearing 58

A.3 Range of the Miller indices 58

A.4 Derivative with respect to atomic coordinates 60

Appendix B Crystallographic data in this study 61

B.1 Coesite 61

B.2 ϵ-Zn(OH)$_2$ 62

B.3 Al-Ca-H 62

Chapter 1

Introduction

1.1 Crystal structure determination problem

Crystal structure determination is one of the most fundamental problems in materials science. This is an inverse problem for finding the atomic arrangement from its physical properties. Given the crystal structure, we can theoretically predict physical properties by first-principles calculations (cf. Section 2.2.1).

The powder diffraction is such a physical property that is generally measured in the experiment for the crystal structure determination. The basic formula of the powder diffraction is explained in Appendix A.1, and directly reflects the atomic arrangement. The Rietveld method [3] is one of the crystal structure analysis methods, in which various parameters are refined by fitting the powder diffraction pattern. This method requires an appropriate structure model, and is not able to search for unknown structures. In the Reverse Monte Carlo (RMC) method [4], we solve the inverse problem from the powder diffraction pattern to the crystal structure in the Monte Carlo simulations. It is difficult for this method to determine a complicated structure because there are many atomic arrangements with similar powder diffraction patterns.

Due to experimental limitations such as noise, it is often not possible to determine the structure by only the experiments. In such cases, the theoretical structure prediction is useful. The stable structure corresponds to the global minimum of the potential energy, and we minimize the potential energy by the structure optimization (see Section 2.4). Since first-principles calculations of the potential energy require high computational costs, it is impossible to explore the whole search space in practice. We search for locally stable structures in the search space with limited cell parameters and compositions, and theoretically propose the most stable and experimentally consistent one as the correct structure. Structure prediction methods have been researched and developed in order to find it systematically and efficiently. Some of them are designed to overcome the energy barriers between global or local minima, such as Simulated Annealing (SA) [5], basin-hopping [6], minima hopping [7] and metadynamics [8]. Moreover, some of them are designed to generate new candidate structures, such as random sampling [9], genetic algorithm [10] and Particle Swarm Optimization (PSO) [11].

1.2 Purpose of the study

In the previous researches [1, 2] and this study, we are developing the structure prediction method based on data assimilation. This method improves the search efficiency by incorporating the experimental data into the theoretical structure search. The details of this method is described in Chapter 2, and we explain the improvements of this method made in this study below.

In this method, we define the penalty function $D(R)$ as the "difference" between the reference experimental data and the theoretical value for the crystal structure R, and minimize the cost function $F = E + \alpha ND$ instead of the potential energy E. Both E and D are minimum at the target structure of the structure search, and D restricts the search space to match the experimental data. We can improve the search efficiency by performing the multi-objective optimization of E and D.

We implement the penalty functions of the powder diffraction. In the previous researches [1, 2], they define the "crystallinity"-type penalty function that uses only the diffraction angles without the intensity information in the reference diffraction pattern. This penalty function works well even with unreliable experimental intensity ratio, but not when the target structure has low symmetry without the extinction law, or when it is difficult to determine peak positions due to noise.

In this study, by adopting the "correlation-coefficient"-type penalty function that explicitly includes the diffraction peak intensities, we are working on improving the success rate and robustness of the structure search. This new penalty function is expected to improve the search efficiency of our method by using more experimental information than the crystallinity-type one. Instead, if the experimental and calculated diffraction patterns are different due to experimental limitations such as noise, this new penalty function does not become zero for the correct structure. We consider that the penalty function method can support the optimization of the potential energy even if the correct structure does not correspond to the global minimum of the penalty function.

1.3 Organization of the thesis

The organization of this thesis is as follows.

In Chapter 2, we explain the structure prediction method based on data assimilation developed in the previous researches [1,2] and this study. In general, the structure prediction consists of the potential energy calculation for examining the stability of the structure, and the structure optimization for minimizing the potential energy. In addition, our method includes the penalty function calculation for examining the consistency of the structure with the reference experimental data. We also explain the "fingerprint" [12] used for judging success or failure of the structure search in test calculations for known structures.

In Chapter 3, we show the results of the structure search in this study. For comparison with the crystallinity-type penalty function of the previous study [1], we perform test calculations for coesite with the correlation-coefficient-type one. We also perform test calculations for ϵ-Zn(OH)$_2$ as the system containing hydrogen. As a practical application, we search for new hydride in Al-Ca-H system with actual experimental data, and discover new $Al_{12}Ca_{20}H_{76}$ structure based on data assimilation.

In Chapter 4, we summarize the outcomes in this study, and describe future improvements of our structure prediction method. Through test calculations for known structures, we confirm that the correlation-coefficient-type penalty function improves the success rate and noise-robustness of the structure search. The hybrid penalty function of XRD and ND is effective for the system containing hydrogen. It is considered that the application range of our method can be expanded by assimilating other experimental data than the powder diffraction. We expect that the data assimilation scheme can be incorporated into existing structure prediction methods by replacing the potential energy with the cost function.

Chapter 2

Methods

The theoretical structure prediction is the multidimentional global optimization problem of the potential energy. Due to the high computational costs of the potential energy and the complexity of the multidimentional search space, we need methods to efficiently search for stable structures. In the previous researches [1, 2] and this study, we are developing the structure prediction method based on data assimilation.

First, Section 2.1 describes how to improve the search efficiency by assimilating the experimental data in this method. Next, Sections 2.2, 2.3, and 2.4 explain the potential energy, the penalty function and the structure optimization used in our simulations, respectively. Section 2.5 is about “Fingerprint” [12] used in test calculations for judging success or failure of structure search.

2.1 Structure prediction method based on data assimilation

In the structure prediction without the experimental data, the most stable or metastable structures are searched by the potential energy optimization (see Section 2.4). Since it is impossible to explore the whole search space in practice, a stable structure consistent with experimental results is theoretically proposed as the correct one. If there are many metastable structures, the structure found in the experiment cannot be uniquely determined from only the potential energy.

Our central idea is to minimize the potential energy and reproduce the experimental data at the same time. From the experimental point of view, such a joint optimization approach is to optimize the atomic configuration that reproduces the experimental data (Refs. [13,14]). On the other hand, in the theoretical structure prediction, this approach is to support the potential energy optimization using the experimental data (Refs. [15–20]).

In the structure prediction method based on data assimilation [1], the search space itself is restricted to match the experimental data during simulations. We can efficiently search for theoretically stable and experimentally consistent structures by excluding structures with low energy but inconsistent with the experiments from the search space.

2.1.1 Assimilating experimental data

In this method, we define the difference between the experimental data and the calculated value for the structure as the penalty function. The data need to reflect the atomic arrangements and be computable at low cost. In this study, we use the powder diffraction pattern as the experimental data. Instead of the potential energy, we minimize the cost function F defined as follows,

$$F(R; I_{\mathrm{ref}}) = E(R) + \alpha N D[I_{\mathrm{ref}}, I_{\mathrm{calc}}(R)], \tag{2.1}$$

where R is the crystal structure, I_{ref} is the reference experimental data, I_{calc} is the calculated one, E is the potential energy, D is the penalty function, N is the number of atoms, and α is the control parameter. That is, we solve the multi-objective optimization problem for the potential energy and the penalty function.

Such a combined cost function was originally proposed in the previous study by Putz et al. [15]. They use a simple model as the potential energy and search for a structure that matches the powder X-ray diffraction pattern by optimizing the penalty function. However, when there are many unstable structures that almost match the experimental data, it is difficult to find the correct structure with this approach. In our approach, the potential energy is calculated accurately, and the penalty function supports searching for the target structure which is theoretically stable and experimentally consistent.

Our method is explained with a schematic diagram as shown in Fig. 2.1. The correct structure is ideally the global minimum of both the potential energy and the penalty function. Since the local minima of them are different, the correct structure is more emphasized in the cost function than in the potential energy. If the target material is metastable or the experimental data is incomplete, the correct structure is not the common global minimum of the potential energy and penalty function, but is most suitable as the global minimum of the cost function. Therefore, by optimizing the cost function, we can find the correct structure faster than by optimizing only the potential energy.

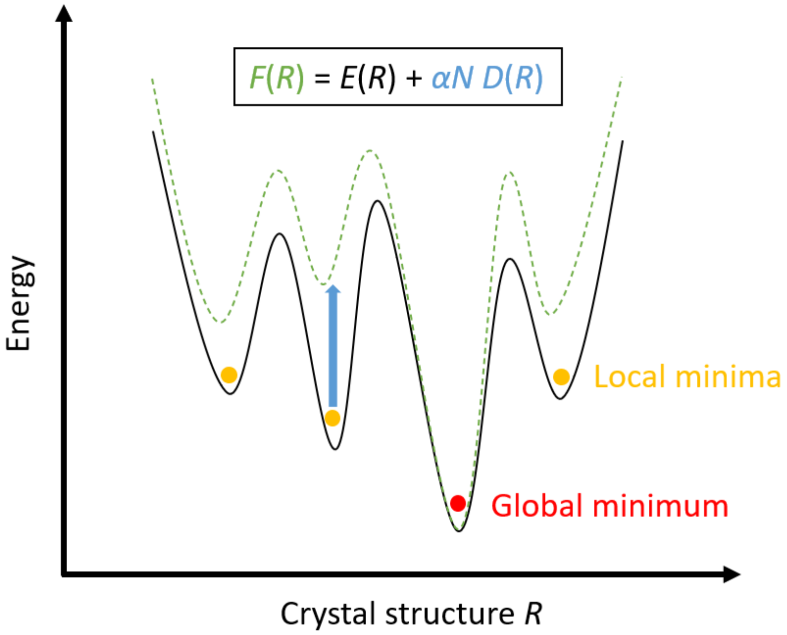

Fig. 2.1. Schematic diagram of the cost function. The horizontal axis represents the crystal structure R, which is actually a multidimensional space, such as cell parameters and atomic coordinates. The vertical axis represents the energy, which is optimized in structure search. The black line represents the potential energy E, and the red and yellow circles are its global minimum and local minima, respectively. The blue arrow represents the penalty function αND added to the potential energy, and the green dotted line represents the cost function F.

The control parameter α adjusts how strongly the search space is restricted by the penalty function to match the experimental data. The role of the penalty function is to fill the local minima of the potential energy. In general, for too large α, the structure change is restricted by the penalty function, and the search efficiency deteriorates. In our test calculations, we estimate the appropriate range of α by examining the magnitude of the penalty function and structure change in short pre-simulations for different values of α.

The advantage of the penalty function method is that it can be directly applied to existing structure prediction methods by simply replacing the potential energy with the cost function. In this study, we adopt Simulated Annealing using Molecular Dynamics because of its simplicity of implementation (see Section 2.4.2). It is easy to extend this method for multiple experimental data by adding penalty functions of them to the cost function. In Section 3.2.3, we simultaneously optimize the penalty functions of the X-ray and neutron diffraction patterns in Zn-O-H system.

This method can be interpreted by Bayes' theorem for the conditional probability of the crystal structure given the experimental data,

$$\rho(R|I_{\mathrm{ref}}) = \frac{\rho(R) \times \rho(I_{\mathrm{ref}}|R)}{\rho(I_{\mathrm{ref}})}. \tag{2.2}$$

In the case of the optimization by Simulated Annealing, the probability of the crystal structure follows the Boltzmann distribution,

$$\rho(R) \propto \exp\{-\beta E(R)\}, \tag{2.3}$$

where β is the inverse temperature. On the other hand, the conditional probability of the reference data given the crystal structure can be expressed as follows using the penalty function,

$$\rho(I_{\mathrm{ref}}|R) \propto \exp\{-\beta\alpha N D[I_{\mathrm{ref}}, I_{\mathrm{calc}}(R)]\}. \tag{2.4}$$

Finally, it turns out that the optimization of the cost function corresponds to the conditional structure search given the experimental data, as shown in the following equation,

$$\rho(R|I_{\mathrm{ref}}) \propto \exp\{-\beta F(R; I_{\mathrm{ref}})\}. \tag{2.5}$$

2.1.2 Flowchart of the method

In this study, we perform the structure prediction based on data assimilation according to the flowchart shown in Fig. 2.2.

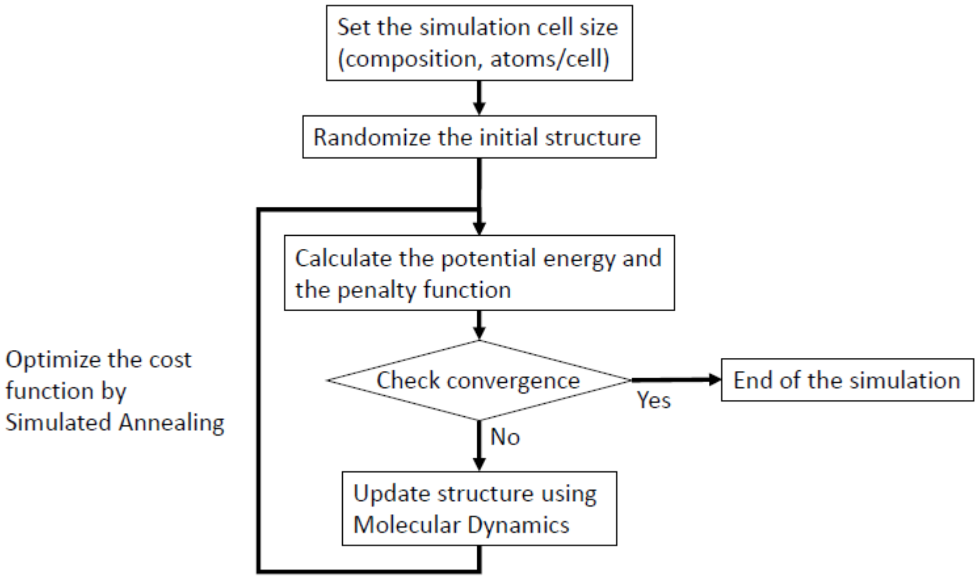
Set the simulation cell size (composition, atoms/cell)
Randomize the initial structure
Calculate the potential energy and the penalty function
Check convergence
Yes
End of the simulation
No
Update structure using Molecular Dynamics
Optimize the cost function by Simulated Annealing

Fig. 2.2. Flowchart of the structure prediction method based on data assimilation in our simulations.

First, we set the appropriate simulation cell size to reproduce the experimental data in the target system. If the composition and the number of atoms per cell cannot be determined from experiments, it is necessary to perform simulations for each candidate cell size.

After setting the cell size, we generate random structures as initial atomic coordinates for each simulation. A structure with too small interatomic distance is inappropriate as the initial arrangement because the potential energy calculation is difficult and the force is too large. In our simulations, we limit the minimum interatomic distance of the initial structure.

Then, we minimize the cost function by Simulated Annealing (SA) using Molecular Dynamics (MD). When the structure converges to a local minimum on the cost function, the MD simulation is finished. The details of those calculations are described in the following Sections 2.2, 2.3, and 2.4.

2.2 Potential energy calculation

The theoretical structure prediction is the optimization problem to search for the most stable structure that minimizes the potential energy. In this section, we explain the potential energy calculations used in our study.

2.2.1 Potential by first-principles calculation

For the theoretical prediction of unknown structures, it is necessary to calculate the potential energy only from atomic arrangements. Such calculation according to the established laws of physics without using any empirical parameters is called "first-principles" or "ab initio" calculation. In the case of the potential energy calculation, we perform the first-principles electronic structure calculation according to the Schrödinger equation.

Density functional theory (DFT) [21] is one of the most popular approaches for the first-principles calculation to solve the Schrödinger equation in many-body systems. Simply, the ground state of non-relativistic many-body systems is expressed as the following Schrödinger equation in atomic units,

$$\hat{H}\left|\Phi\right\rangle = (\hat{T}+\hat{U}+\hat{V})\left|\Phi\right\rangle = E_0\left|\Phi\right\rangle, \tag{2.6}$$

where E_0 is the potential energy of the ground state required for the structure search. $\hat{T}$ is the term of the kinetic energy,

$$\hat{T} = -\frac{1}{2}\sum_i \nabla_i^2, \tag{2.7}$$

where subscript i is the index of electrons. $\hat{U}$ is the term of the electron-electron Coulomb interaction,

$$\hat{U} = \frac{1}{2}\sum_{i\neq j}\frac{1}{|\boldsymbol{r}_i-\boldsymbol{r}_j|}, \tag{2.8}$$

where $\boldsymbol{r}$ is the electronic coordinate. $\hat{V}$ is the term of electron-nucleus Coulomb interaction,

$$\hat{V} = \sum_{i,I}\frac{Z_I}{|\boldsymbol{r}_i-\boldsymbol{R}_I|}, \tag{2.9}$$

where subscript I is the index of nuclei, R is the nuclear coordinate, and Z is the nuclear charge. In general, even with a supercomputer, it is almost impossible to solve such equation straightforwardly because of huge degrees of freedom. In order to deal with the many-body Schrödinger equation, it is necessary to reduce the number of degrees of freedom.

Density functional theory overcomes this difficulty by expressing the Schrödinger equation in terms of the electronic density n, whose degrees of freedom are independent of the number of electrons. According to the first Hohenberg-Kohn (H-K) theorem [22], the external potential $\hat{V}$ is a unique functional of the electronic density n. Furthermore, according to the second H-K theorem [22], the ground state density n_0 minimizes the ground state energy functional, that is,

$$E_0[n] = \langle\Phi_0[n]|\, \hat{H}\, |\Phi_0[n]\rangle \geq E_0[n_0] = E_0. \tag{2.10}$$

Therefore, we can obtain the ground state density n_0 by searching for n which minimizes $E_0[n]$.

The Kohn-Sham (K-S) theory [21] enables us to calculate the ground state based on the H-K theorems. In the K-S theory, the fictitious non-interacting system is introduced,

$$\left(-\frac{1}{2}\nabla^2 + V_{\mathrm{eff}}\right)|\psi_i\rangle = \epsilon_i\, |\psi_i\rangle\,, \tag{2.11}$$

where V_{eff} is chosen to make the ground state density of the fictitious system equal to that of the real system n_0. By using such a fictitious system, the energy functional is expressed as

$$E_0[n] = T_s[n] + \frac{1}{2}\int \mathrm{d}\boldsymbol{r} \int \mathrm{d}\boldsymbol{r}' \frac{n(\boldsymbol{r})n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + \int \mathrm{d}\boldsymbol{r} n(\boldsymbol{r})V(\boldsymbol{r}) + E_{\mathrm{XC}}[n]. \tag{2.12}$$

The first term on the right side represents the kinetic energy in the fictitious system,

$$T_s[n] = \sum_i \langle\psi_i| - \frac{1}{2}\nabla^2\, |\psi_i\rangle\,. \tag{2.13}$$

The second and third term represents the electron-electron and electron-nucleus Coulomb interaction expressed in the terms of n, respectively. The fourth term E_{XC} is called the exchange correlation energy, and is an unknown functional that connects the real and fictitious system.

According to the variational method with respect to n, the terms of the ground state energy vanish, and V_{eff} is expressed as

$$V_{\mathrm{eff}}(\boldsymbol{r}) = \int \mathrm{d}\boldsymbol{r}' \frac{n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + V(\boldsymbol{r}) + \frac{\delta E_{\mathrm{XC}}}{\delta n(\boldsymbol{r})}. \tag{2.14}$$

Finally, the ground state energy of the real system is expressed as follows in terms of n using the computable ground state energies of the fictitious system,

$$E_0[n] = \sum_i \epsilon_i - \frac{1}{2}\int \mathrm{d}\boldsymbol{r} \int \mathrm{d}\boldsymbol{r}' \frac{n(\boldsymbol{r})n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + E_{\mathrm{XC}}[n] - \int \mathrm{d}\boldsymbol{r} n(\boldsymbol{r}) \frac{\delta E_{\mathrm{XC}}}{\delta n(\boldsymbol{r})}. \tag{2.15}$$

The exchange-correlation energy E_{XC} is unknown, and approximated in several ways. The local density approximation (LDA) [21] is the simplest and remarkably successful. In LDA method, E_{XC} is approximated by that of a homogeneous electron gas (HEG) with the corresponding density n. For beyond LDA, the generalized gradient approximation (GGA) [23] takes the density gradient into account.

In our simulations, we use Vienna Ab initio Simulation Package (VASP) [24, 25] to calculate the potential energy based on DFT. VASP is a famous package for performing ab initio quantum mechanical molecular dynamics simulations using pseudopotentials and a plane wave basis set. The pseudopotential is a method of incorporating inner-shell electrons near nuclei into the potential in order to reduce the computational costs. The plane wave basis set is a basis set of the wave function in DFT calculation.

We also use some code in PIMD [26, 27] to make the potential energy calculation in VASP into a subroutine. PIMD is an open-source software for parallel molecular simulations.

2.2.2 Empirical potential

There are various force fields based on appropriate models for target systems. Empirical parameters in a model are adjusted to reproduce physical properties based on experiments or first-principles calculations. From the viewpoint of structure search, it is important that the global or local minima on empirical potentials correspond to actually stable structures.

Due to the high computational costs, it is difficult to perform first-principles simulations in large systems. In this study, in order to test the structure prediction method based on data assimilation, we perform the structure search for known structures, such as coesite in Section 3.1 and ϵ-Zn(OH)$_2$ in Section 3.2. In those calculations, we use model potentials as the potential energy, which are much faster to be calculated than the DFT potential.

We use Tsuneyuki potential [28] and Tersoff potential [29] for coesite, and the reactive force field (ReaxFF) potential [30] for ϵ-Zn(OH)$_2$. Tsuneyuki potential is the interatomic potential extracted from first-principles cluster calculations, and a simple model expressed by the following formula,

$$V_{ij}(r_{ij}) = V_{ij}^{\text{coulomb}}(r_{ij}) + f_0(b_i + b_j)\exp\left(\frac{a_i + a_j - r_{ij}}{b_i + b_j}\right) - \frac{c_i c_j}{r_{ij}^6}, \tag{2.16}$$

where subscripts i and j are the indices of atoms and r is the atomic distance. It consists of Coulomb interaction with some corrections, Born-Mayer-type repulsion, and dispersive interaction. This potential succeed in the structure simulations for polymorphs of SiO$_2$ such as low-quartz, low-cristobalite, coesite, and stishovite. Both Tersoff and ReaxFF potential are based on a bond order scheme as expressed by the following formula,

$$V_{ij}(r_{ij}) = A\exp(-\lambda_A r_{ij}) - b_{ij} B\exp(-\lambda_B r_{ij}), \tag{2.17}$$

where b_{ij} is the bond order. In Tersoff potential, the bond order is taken to have the following form,

$$b_{ij} = \left(1 + \zeta_{ij}^{\eta}\right)^{-\delta}, \tag{2.18}$$

$$\zeta_{ij} = \sum_k g(\theta_{ijk})\exp\left(p(r_{ij} - r_{ik})\right)^q, \tag{2.19}$$

$$g(\theta) = 1 + \frac{c^2}{d^2} - \frac{c^2}{d^2 + (h - \cos\theta)^2}, \tag{2.20}$$

where θ_{ijk} is the bond angle between bonds ij and ik.

In the structure prediction, if there is a model potential that reproduces first-principles calculations well, we can reduce the computational cost and expand the search range. The remarkable progress of machine learning makes it possible to optimize empirical parameters in general-purpose models such as the neural network [31].

In our simulations, we use Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [32, 33] to perform empirical potential calculations. LAMMPS is a famous classical molecular dynamics code with a focus on materials modeling.

2.3 Introduction of the penalty function

In the structure prediction method based on data assimilation, we define the difference between the experimental data and the calculated value as the penalty function. In this study, we use the powder diffraction pattern as the experimental data, and implement artificial forms of penalty functions.

2.3.1 Penalty function of the powder diffraction pattern

The powder diffraction is one of the most popular experimental methods for the crystal structure analysis. We implement the calculations of the powder X-ray diffraction (XRD) and neutron diffraction (ND) in our code (see Appendix A). Since X-rays are scattered at the electrons of the atomic shell, it is difficult for XRD to detect the positions of light atoms such as hydrogen. ND is theoretically almost the same as XRD except for being scattered at the nucleus, and useful to detect the positions of light atoms.

As Eqs. A.1-A.4 show, the diffraction pattern directly reflects the atomic arrangement. Since the diffraction pattern is easy to calculate and can be differentiated with respect to the atomic coordinates (see A.4), it is suitable as the reference experimental data used for data assimilation. The penalty function of the diffraction pattern restricts the search space so as to reproduce the translational periodicity of the crystal structure. On the other hand, the potential energy restricts the partial structure at a short distance. It is expected that we can improve the search efficiency by optimizing the cost function that combines them.

The penalty function needs to take the minimum value with the correct structure, and not change significantly with respect to the small displacement of structures. As Eq. A.5 shows, the peak position does not depend on the atomic fractional coordinates, only on the cell parameters. A penalty function that compares the intensities for each diffraction angle does not work well if the peak positions shift due to changes in the cell parameters. Therefore, it is difficult to optimize the cell parameters by the penalty function of the diffraction pattern. In this study, we estimate the cell parameters from the experiment, and fix them during the simulations. If the cell parameters cannot be uniquely determined from the experiment, we perform the simulations in each candidate.

We can deal with the small shifts in the peak positions by increasing the peak width in the calculated diffraction pattern (see Appendix A.2). If the peak width is too large compared to that in the experiment, the penalty function changes small with respect to the atomic displacement and the effect of emphasizing the correct structure becomes weak. On the other hand, if the peak width is too small, overfitting will occur in an attempt to reproduce one peak in the experiment with multiple peaks. In order to improve the search efficiency, it is desirable to set the same peak width as that in the experiment.

We can also deal with the polyphasic diffraction pattern by removing the range with peaks other than these of the target material from the reference range of the diffraction angle. It is inappropriate to consider the intensities in this range as zero because the peaks of the target material may overlap. In Section 3.3, we perform the structure search in Al-Ca-H system using the actual experimental data including peaks of pure Al.

2.3.2 Form of the penalty function

The form of the penalty function changes the properties of the structure prediction method, such as the search efficiency and applicable systems. Since it is difficult to directly compare the peak intensities between the experimental data and the calculated value, we formulate the difference in the peak positions and the intensity ratio as the penalty function. In the following, we explain two types of penalty functions implemented in our code.

In the previous research [1], the "crystallinity"-type penalty function is defined as follows,

$$D(R) = 1 - \frac{\int_{\theta=\theta_{\mathrm{obs}}} I_{\mathrm{calc}}(R)\mathrm{d}\theta}{\int_{\theta_{\min}}^{\theta_{\max}} I_{\mathrm{calc}}(R)\mathrm{d}\theta}, \tag{2.21}$$

where R is the crystal structure, I_{calc} is the calculated diffraction intensity, θ is the diffraction angle, $[\theta_{\min}, \theta_{\max}]$ is the reference angle range, and θ_{obs} is the peak position observed in the reference experimental diffraction pattern I_{ref}. This penalty function does not depend on the experimental intensity information, and represents the coincidence ratio of peak positions between experimental and calculated diffraction patterns. By minimizing this penalty function, we can restrict the search space to satisfy the same extinction rule as in the experiment. Due to experimental errors such as the effect of preferred orientation, the peak intensity ratio in experiments deviates from the ideal calculated one. This penalty function works well even with unreliable experimental intensity ratio. On the other hand, it does not work well when the target structure has low symmetry without the extinction law, or when it is difficult to determine peak positions due to noise.

In this study, we introduce a new penalty function of the correlation coefficient between I_{calc} and I_{ref} as follows,

$$D(R) = 1 - \frac{\int_{\theta_{\min}}^{\theta_{\max}} (I_{\mathrm{calc}}(R) - \overline{I}_{\mathrm{calc}}(R))(I_{\mathrm{ref}} - \overline{I}_{\mathrm{ref}})\mathrm{d}\theta}{\sqrt{\int_{\theta_{\min}}^{\theta_{\max}} (I_{\mathrm{calc}}(R) - \overline{I}_{\mathrm{calc}}(R))^2 \mathrm{d}\theta}\sqrt{\int_{\theta_{\min}}^{\theta_{\max}} (I_{\mathrm{ref}} - \overline{I}_{\mathrm{ref}})^2 \mathrm{d}\theta}}, \tag{2.22}$$

where $\overline{I}$ is the average intensity among the reference angle range $[\theta_{\min}, \theta_{\max}]$. The crystallinity-type penalty function emphasizes peaks not found in the experiment, whereas this one emphasizes large peaks. This penalty function explicitly includes the experimental intensity information, and restricts the search space more to match the experimental data. Therefore, it is expected that this penalty function can improve the search efficiency more. Instead, due to the difference between experimental and calculated peak intensities, this penalty function does not become zero for the correct structure. Even if the correct structure does not correspond to the global minimum of the penalty function, the penalty function method can support the optimization of the potential energy. We can find stable structures by performing the structure relaxation with only the potential energy (see Section 2.4.1) after the optimization of the cost function.

In this study, we mainly perform the structure prediction based on data assimilation using the correlation-coefficient-type penalty function, and investigate how this new penalty function improves the efficiency and robustness of structure search. In Section 3.1.1, we compare the crystallinity-type and correlation-coefficient-type penalty functions by test calculations for coesite. In Section 3.1.3, we examine the noise-robustness of the correlation-coefficient-type one. We implement white Gaussian noise as the artificial noise on the calculated diffraction pattern, and generate different noises for each simulation.

2.4 Structure optimization

In general, “structure optimization” is a simulation method that searches for the structure that minimizes some cost function. In order to theoretically determine the crystal structure, we search for the structure that minimizes the potential energy. The ab initio crystal structure prediction requires the first-principles calculations of the potential energy for each structure in the simulation. In that case, the potential energy calculation accounts for most of the computational costs, and we need efficient optimization methods.

In this section, we explain the structure search method used in our study.

2.4.1 Structure relaxation

“Structure relaxation” is one of the simplest optimization methods. In this method, we calculate the force applied to each atom, and move atoms a little along its direction. This is repeated until the force or the energy change in each step become sufficiently small. By gradually displacing the structure in this way, we can finally find a locally stable structure (see Fig. 2.3). The cell parameters can also be optimized by calculating the stress tensor instead of the force.

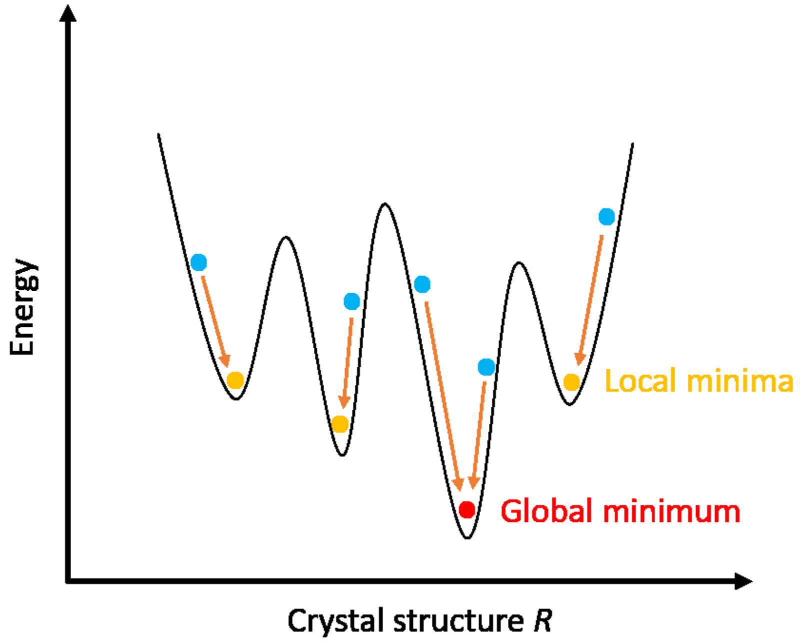

Fig. 2.3. Schematic diagram of structure relaxation. The horizontal axis represents the crystal structure R, and the vertical axis represents the energy. The black line represents the potential energy, and the red and yellow circles are its global minimum and local minima, respectively. The blue circles are the initial structures in the simulations, and structure relaxation allows us to find the nearest globally or locally stable structure as shown by the orange arrows.

The locally stable structure means the local minimum on the cost function and includes the metastable structure. As the number of atoms in the simulation increases, that is, the number of parameters to be optimized increases, the number of local minima on the cost function also increases. It is difficult to find the desired most stable structure because the structure is trapped in the locally stable structure. In addition, the crystal structure simulation requires the cell with translational symmetry, and the structure may be trapped at the saddle point.

In both VASP [24, 25] and LAMMPS [32, 33] which we use to calculate the potential energy (see Section 2.2), the structure relaxation is implemented in some algorithms, such as the simplest gradient descent method and the more efficient conjugate gradient method.

2.4.2 Simulated Annealing

In order to efficiently search for the most stable structure, the global optimization methods are required. Simulated Annealing (SA) [5] is one of metaheuristic global optimization algorithms. As the name implies, in SA, the barrier of the cost function is overcome depending on the temperature, and the temperature is gradually lowered during the simulation. When the temperature is high, the state can move back and forth between the barriers, but when the temperature decreases, the state shifts to the side with the higher barrier, that is, with the lower energy (see Fig. 2.4). At the end of the simulation, the temperature becomes zero, and the state move only in the direction in which the cost function decreases. As in the case of structure relaxation, the state converges to some local minimum.

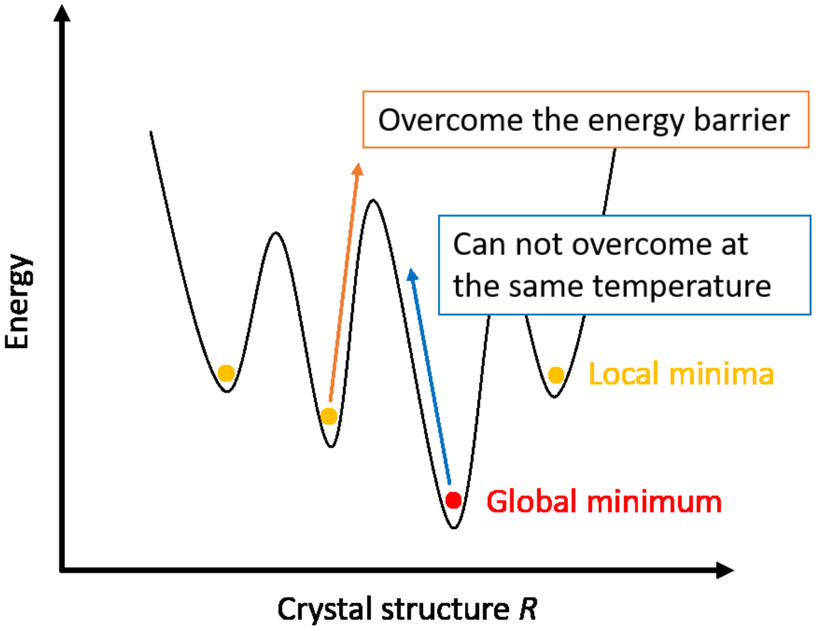
Overcome the energy barrier
Can not overcome at the same temperature
Local minima
Global minimum
Energy
Crystal structure *R*

Fig. 2.4. Schematic diagram of Simulated Annealing. The horizontal axis represents the crystal structure R, and the vertical axis represents the energy. The black line represents the potential energy, and the red and yellow circles are its global minimum and local minima, respectively. As the temperature decreases during the simulation, as indicated by the orrange and blue arrows, the state transition becomes one-way from the local minima to the global minimum at some temperature.

It is known that we can always find the global minimum by cooling from a sufficiently high temperature slowly enough [34]. However, in order to reduce the computational cost, it is necessary to set the temperature and the step size to a finite value. The computational cost and the success rate to find the global minimum are in the trade-off relationship.

2.4.3 Molecular Dynamics

To apply Simulated Annealing to the structure optimization, we set the method of the state transition, typically using classical Molecular Dynamics (MD) or Monte Carlo (MC). In our study, we implement SA using MD.

The classical MD is a method of simulating atomic movements from the calculated forces according to classical mechanics. In the classical MD simulation, we repeatedly calculate the force applied to each atom and update the atomic velocity and position based on it as follows,

$$\boldsymbol{p}_i \leftarrow \boldsymbol{p}_i + \boldsymbol{F}_i \Delta t, \tag{2.23}$$

$$\boldsymbol{r}_i \leftarrow \boldsymbol{r}_i + \frac{\boldsymbol{p}_i}{m_i} \Delta t, \tag{2.24}$$

where subscript i is the index of atoms, $\boldsymbol{p}$ is the atomic momentum, $\boldsymbol{r}$ is the atomic position, m is the atomic mass, $\boldsymbol{F}$ is the force, and Δt is the simulation timestep.

The temperature corresponds to the average kinetic energy of atoms, as shown in the following formula,

$$T = \frac{1}{\frac{3}{2} N k_B} \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m_i}, \tag{2.25}$$

where N is the number of atoms, and k_B is Boltzmann constant. There are some methods for controlling the temperature, such as the simplest velocity scaling method [35] and Nosé-Hoover thermostat [36, 37] which gives the canonical distribution. In our simulations, we use velocity scaling method for the structure optimization based on SA.

In the velocity scaling method, as the name implies, the temperature parameter is controlled by scaling the atomic velocities as follows,

$$\boldsymbol{p}_i \leftarrow \boldsymbol{p}_i \sqrt{\frac{T_{\text{ext}}}{T}}, \tag{2.26}$$

where T_{ext} is the bath temperature. In SA, the bath temperature gradually decreases to zero. At zero temperature, the atomic velocities become zero once from Eq. 2.26, and the atomic positions are updated in the direction of the forces from Eq. 2.23 and 2.24. Therefore, the velocity scaling method naturally leads to the structure relaxation using the gradient descent method at the end of SA.

2.5 Judging success or failure using fingerprint

In this study, in order to test the structure prediction method based on data assimilation, we perform the structure search for known structures, such as coesite in Section 3.1 and ϵ-Zn(OH)$_2$ in Section 3.2. In the test calculations, we can investigate the success rate of structure search by checking whether the structure obtained in each simulation matches the correct structure. Since the structure is invariant to translation and rotation, we cannot directly compare the atomic coordinates.

We use "fingerprint" [12] for judging success or failure of the structure search. Fingerprint function is a crystal structure descriptor expressed by the following formula,

$$F_{AB}(l) = \sum_{i_A}^{N_A} \sum_{j_B}^{l_{\min}<l_{ij}<l_{\max}} \frac{V}{4\pi N_A N_B \Delta} \frac{\delta(l - l_{ij})}{l_{ij}^2} - 1, \tag{2.27}$$

where l is the interatomic distance, $[l_{\min}, l_{\max}]$ is the reference range of l, A and B are the atomic species, subscript i and j are the indices of atoms, N is the number of atoms, V is the cell volume, and Δ is the step width of l. The delta function is smoothed by Gaussian smearing function (cf. Appendix A.2). Figure 2.5 illustrates the fingerprint for coesite.

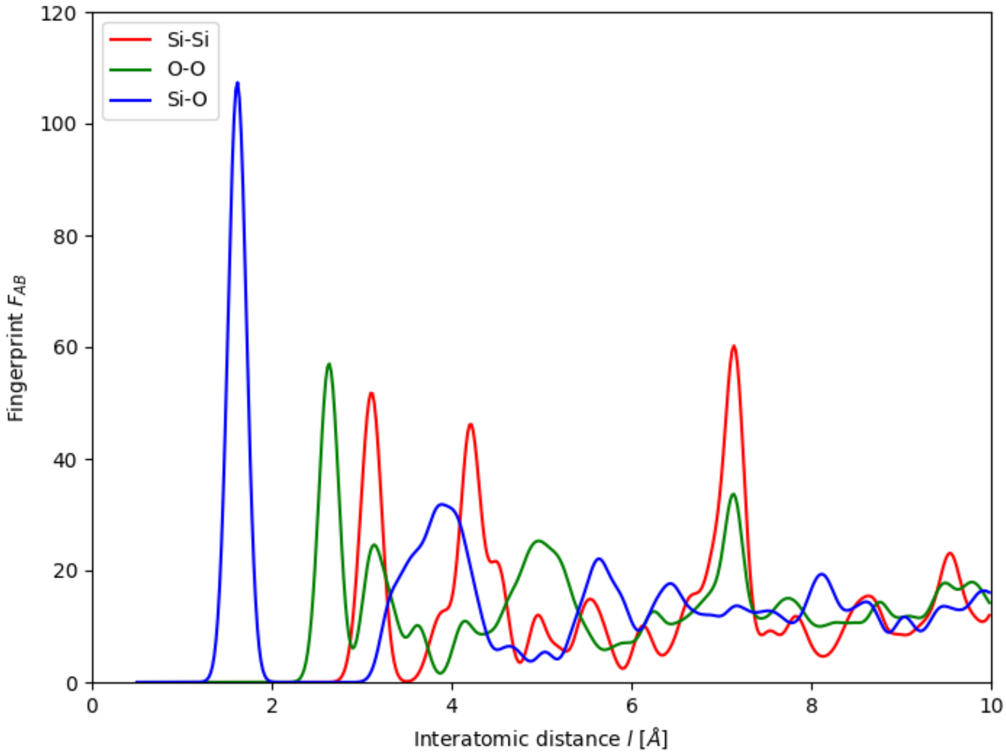

Fig. 2.5. Fingerprint for coesite. Each line represents the fingerprint of the combination of the atomic species A-B. The origin of the fingerprint is shifted by -1 to make it easier to see. We set $[l_{\min}, l_{\max}] = [0.05, 10]\,\AA$, $\Delta = 0.02\,\AA$, and the standard deviation of Gaussian smearing function $\sigma = 0.1\,\AA$ (see Eq. A.8).

We evaluate the difference between crystal structures by the residual sum of squares (RSS) of the fingerprints as follows,

$$\mathrm{RSS} = \frac{\int (F_{AB}(l;R) - F_{AB}(l;R_0))^2 \mathrm{d}l}{\int (F_{AB}(l;R) + F_{AB}(l;R_0))^2 \mathrm{d}l}, \tag{2.28}$$

where R is the crystal structure in the simulation and R_0 is the correct structure. The denominator is for normalization. Figure 2.6 shows the RSS of the fingerprints and the potential energy of obtained structures in a certain structure search simulation for coesite in Section 3.1.1. Although the potential energy is not converged sufficiently at the end of simulations, it can be seen that about 40 % of obtained structures have almost the same atomic arrangements as the correct structure by judging based on the RSS of the fingerprints. In order to finally predict the correct structure, it is necessary to relax obtained structure again with only the potential energy without the artificial penalty function.

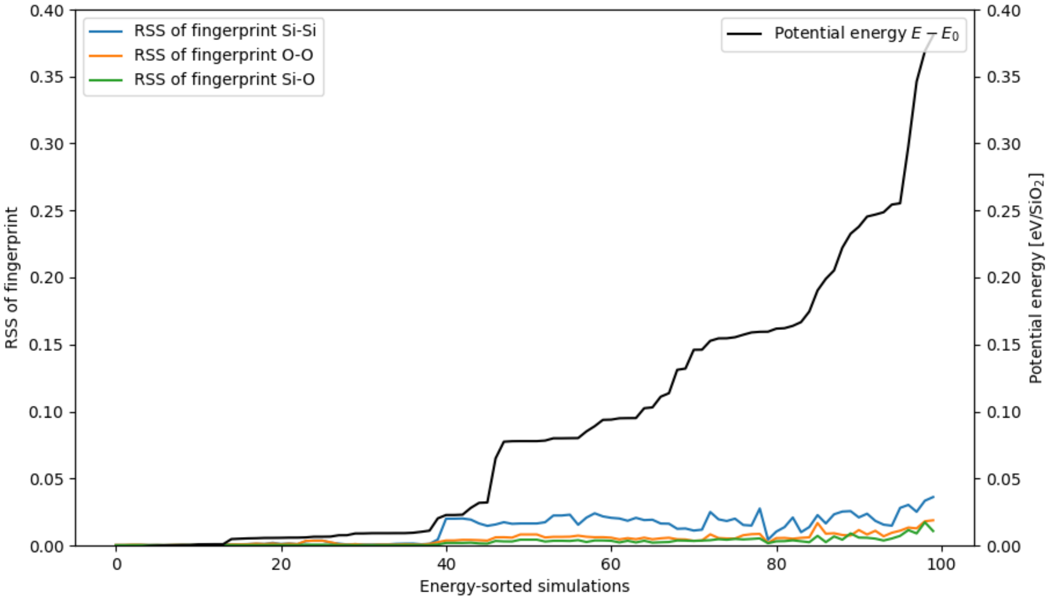

Fig. 2.6. Comparison between RSS of fingerprints and potential energy for coesite. The structures obtained in certain simulations are arranged in ascending order of the potential energy. The black line represents the potential energy with that of the correct structure as the origin. Other lines are the RSS of the fingerprints.

Chapter 3

Results and Discussion

We work on improving the structure prediction method based on data assimilation by applying it to various systems. Sections 3.1 and 3.2 explain test calculations of the structure search for known structures, coesite and ϵ-Zn(OH)$_2$, respectively. In test calculations for SiO$_2$ coesite, we investigate the improvements made by the new correlation-coefficient-type penalty function compared to the crystallinity-type penalty function in the previous study [1]. In test calculations for ϵ-Zn(OH)$_2$, we examine the effectiveness of our structure prediction method for the system containing hydrogen which contributes little to the XRD pattern.

Section 3.3 describes the results of the structure search for new hydride in Al-Ca-H system. This new hydride is synthesized under high temperature and high pressure in the experiment, and the composition ratio has not yet been determined. We discover new $Al_{12}Ca_{20}H_{76}$ structure by data assimilation and first-principles calculations. In addition, we try the structure search with different compositions using the neural network potential.

3.1 Comparison of the correlation-coefficient-type and crystallinity-type penalty functions

For comparison of the correlation-coefficient-type and crystallinity-type penalty functions, we perform the structure prediction based on data assimilation for coesite. Coesite is suitable for this comparison due to its large primitive cell and relatively low symmetry. In these test calculations, we use model potentials as the potential energy in order to reduce the computational costs.

The model potential for Si-O system is Tsuneyuki potential [28] in Sections 3.1.1 and 3.1.3, and Tersoff potential [29] in Section 3.1.2 (cf. Section 2.2.2). In Section 3.1.3, we verify the noise-robustness of our structure prediction method by adding artificial noise to the reference XRD pattern.

3.1.1 Test calculations for coesite

As in the previous study using the crystallinity-type penalty function [1], we perform test calculations of the structure prediction based on data assimilation using the correlation-coefficient-type penalty function in Si-O system. We use Tunesyuki potential as the potential energy, and set coesite in the primitive cell as the target material (see Fig. 3.1).

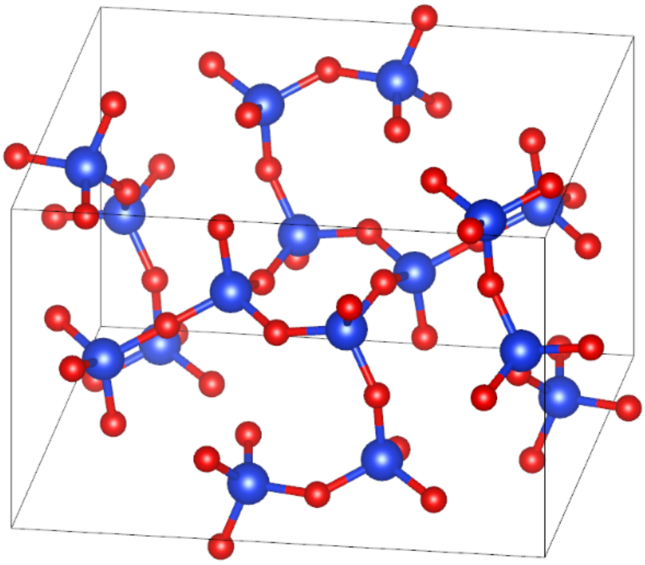

Fig. 3.1. Crystal structure of coesite. The blue and red spheres represent silicon and oxygen atoms, respectively. The primitive cell contains 48 atoms. The crystallographic data is shown in Appendix B.1.

We perform the structure relaxation from the reference structure of coesite [38] using LAMMPS with Tsuneyuki potential, and set this relaxed structure as the correct one in the structure search. In these test calculations, instead of the actual experimental data, we use the calculated XRD pattern for the correct structure as the reference data in the penalty functions (see Eqs. 2.21 and 2.22). Figure 3.2 shows the XRD pattern for coesite used in these test calculations. If the correct structure is obtained in the end of structure search simulations, the calculated XRD pattern I_{calc} is exactly the same as the reference one I_{ref}, and the penalty function ideally becomes zero.

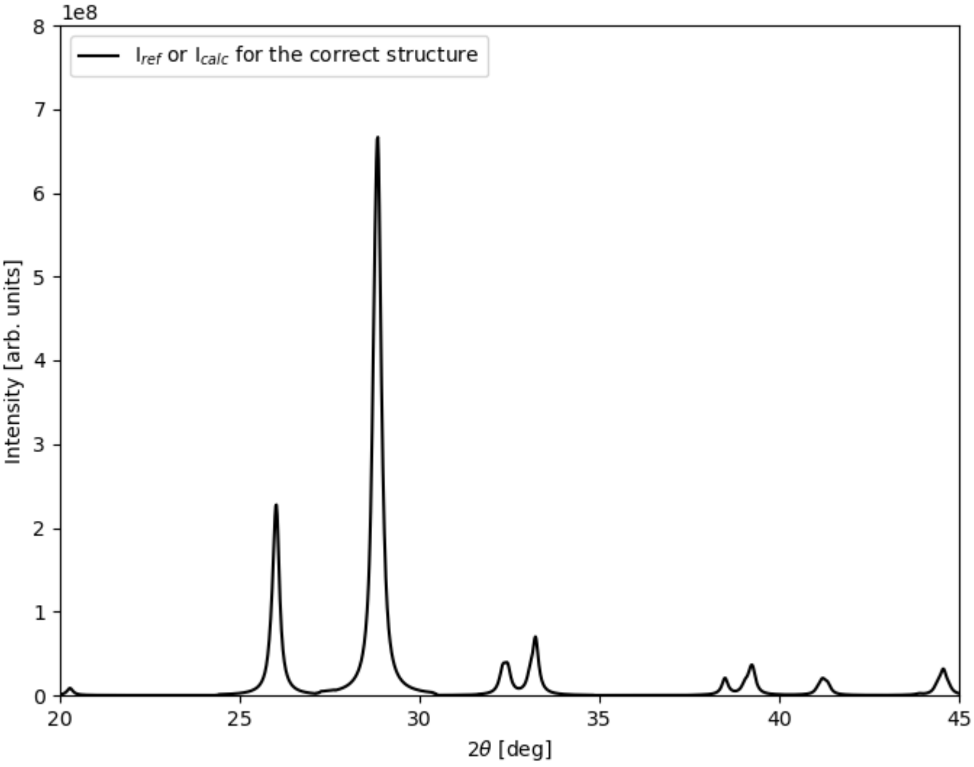

Fig. 3.2. Reference XRD pattern I_{ref} and the calculated one I_{calc} for coesite in the test structure search simulations. The wavelength λ is set to 1.54 $\AA$, which corresponds to CuKα-radiation. The diffraction peaks are smoothed with Lorentzian smearing function, where the scale parameter γ is 0.1 degree in Eq. A.9. The reference range of the diffraction angle $[2\theta_{\mathrm{min}}, 2\theta_{\mathrm{max}}]$ in Eqs. 2.21 and 2.22 is set to $[20, 45]$ degrees.

Then, we perform the structure optimization by simulated annealing using molecular dynamics to minimize the cost function (see Eq. 2.1). The initial temperature is set to 10,000 K, the temperature step is -1 K/step, and the time step is 1 fs/step. The cell parameters are fixed to those of the primitive cell for the correct structure, and the initial structure of each simulation is randomly generated under the constraint that the interatomic distances are 0.5 $\AA$ or more.

Figure 3.3 shows the results of the success rate to find the correct structure in our test calculations. We investigate the success rate by the residual sum of squares (RSS) of the fingerprints between the obtained and correct structure (see Section 2.5). We perform 100 simulations for each value of the control parameter α in Eq. 2.1. The larger α, the more the penalty function restricts the search space to match the experimental data. If α is set to zero and the structure optimization is performed using only the potential energy without the experimental data, the correct structure cannot be found in this condition. When α is set to 10 eV or more, the success rate is almost 100 %. In the case of the crystallinity-type penalty function, the success rate is about 5 % under the same condition. Therefore, the search efficiency is significantly improved by using the penalty function that explicitly includes the peak intensity of the reference XRD pattern.

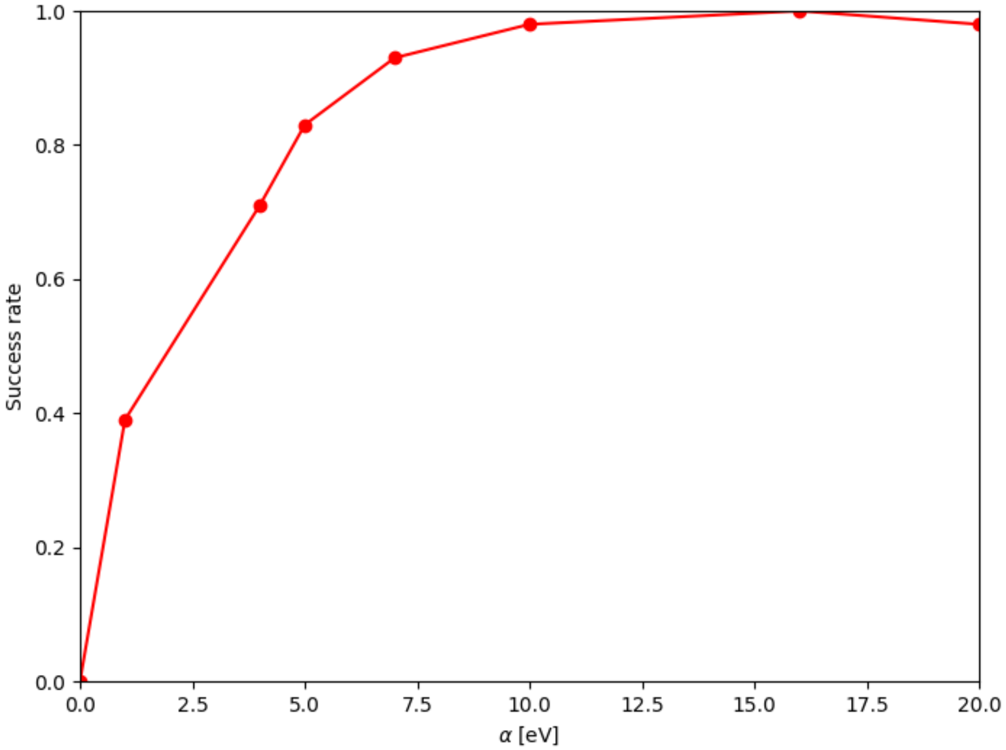

Fig. 3.3. Success rate of the structure search for coesite using Tsuneyuki potential as the potential energy and the correlation-coefficient-type penalty function of XRD. The horizontal axis α is the control parameter in Eq. 2.1.

However, there is a problem that the success rate becomes almost 100 % even if α is very large such as 100 eV. In this case, due to the large forces from the penalty function, the atoms reach the correct positions directly from the random initial structures without the process of Simulated Annealing. It is considered that the cost function in this case has almost no local minimum. One of the reasons is that Tsuneyuki potential is a simple model. Another reason is that the penalty function is defined with the ideal reference data I_{ref} that exactly matches the calculated one I_{calc} for the correct structure. In Section 3.1.2, we perform test calculations using Tersoff potential as the potential energy, which is more complex than Tsuneyuki potential. In Section 3.1.3, we also perform test calculations using the incomplete reference data with artificial noise.

3.1.2 Using Tersoff potential as the potential energy

In order to investigate the effects of different potential energy, we perform similar simulations using Tersoff potential for SiO_2, which is a more complex model than Tsuneyuki potential. As shown in Eq. 2.16, Tsuneyuki potential is an ionic two-body force field depending only on the atomic distances. On the other hand, as shown in Eqs. 2.17-2.20, Tersoff potential is a three-body force field depending on the bond angles. Other computational conditions for the structure optimization are the same as those in Section 3.1.1. We use the correlation-coefficient-type penalty function with the ideal reference data.

Figure 3.4 shows the results of the success rate in our simulations using Tersoff potential. We perform 100 simulations for each value of the control parameter α. The maximum success rate is about 6 %, which is very small compared to the case of Tsuneyuki potential (see Fig. 3.3). The reason why the search efficiency deteriorate is considered that the number of local minima on the cost function increases due to the more complicated potential energy. Especially for a very large α, the success rate drops to zero. From this result, it can be seen that not only the metastable structures but also the locally stable structures with relatively high energies affect the success rate of the structure prediction based on data assimilation. The reason why the success rate drops with α around 15 eV is considered that the results have large error due to the low success rate and the number of local minima on the cost function changes depending on α.

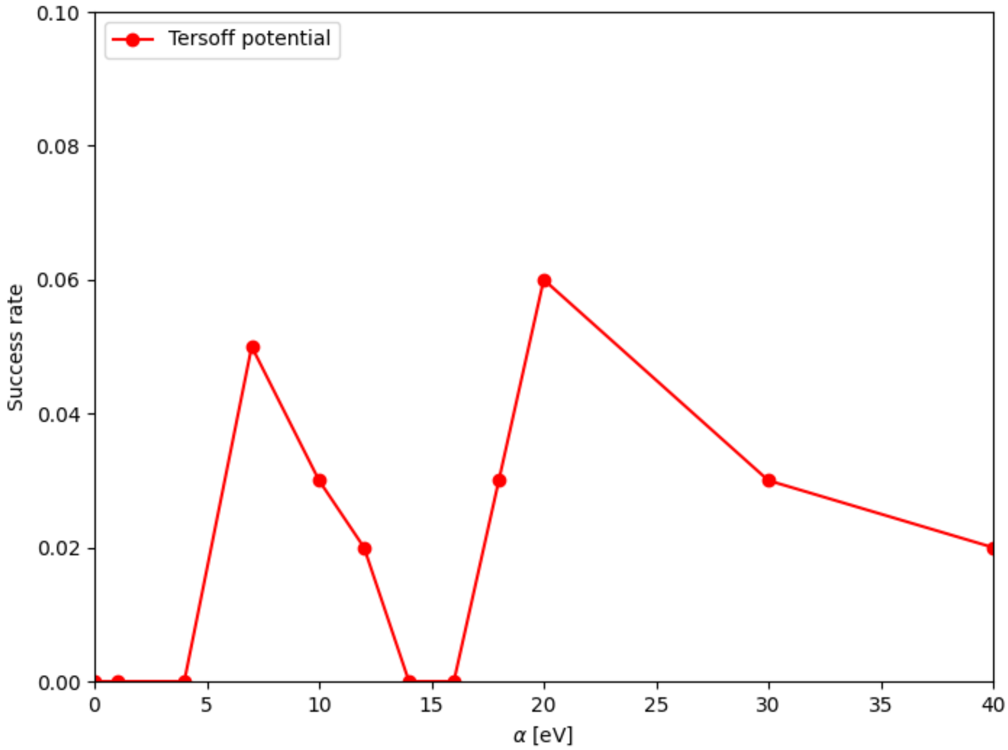

Fig. 3.4. Success rate of the structure search for coesite using Tersoff potential as the potential energy and the correlation-coefficient-type penalty function of XRD.

Figure 3.5 shows the potential energy histograms for obtained structures at the end of simulations, and Figure 3.6 shows the relationship between the potential energy and the penalty function. Compared with the case where $\alpha = 0$ eV, the distribution of the potential energy is widened by adding the penalty function to the cost function. In the multi-objective optimization of the potential energy and the penalty function, the optimization of the potential energy itself sometimes deteriorates by additional forces from the penalty function. Because obtained structures match the reference XRD pattern well, we can improve the success rate to find the correct structure. If α is set too large, the atoms hardly move due to the restriction by the penalty function, and the success rate decreases. Since the role of the penalty function is to support the optimization of the potential energy, it is necessary to select an appropriate value of α that maximizes the success rate.

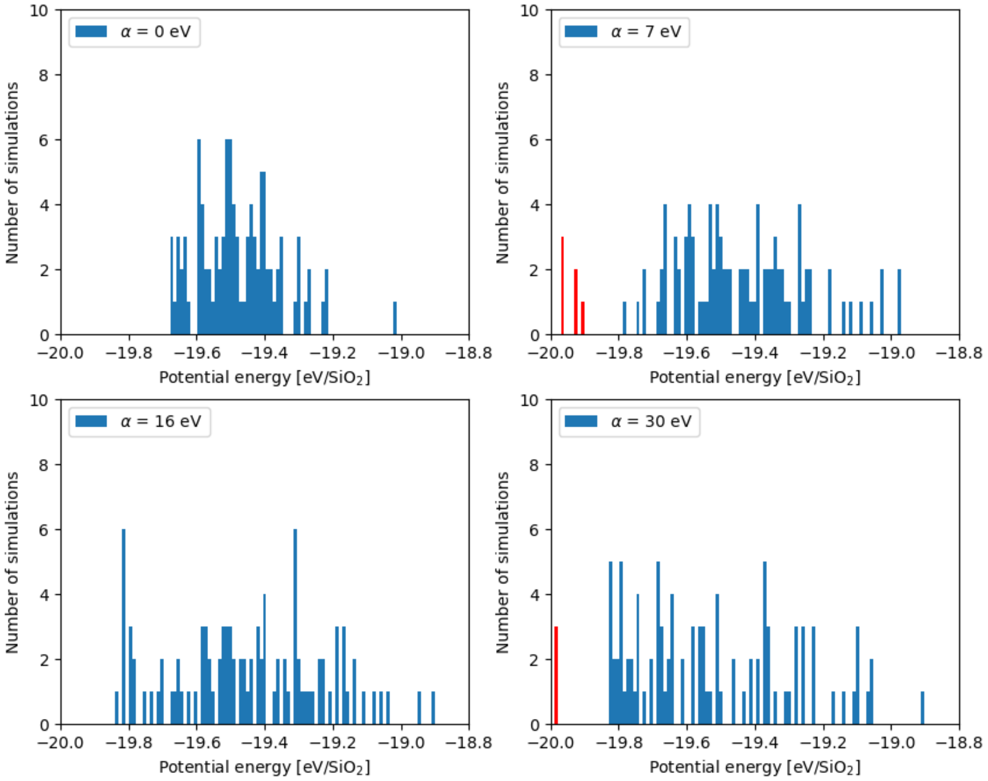

Fig. 3.5. Potential energy histogram of obtained structures in the structure search for coesite using Tersoff potential as the potential energy and the correlation-coefficient-type penalty function of XRD. The number of simulations for each α is 100. The potential energy for the correct structure is about 19.96 eV/SiO$_2$, and the red lines represent structures judged to be correct.

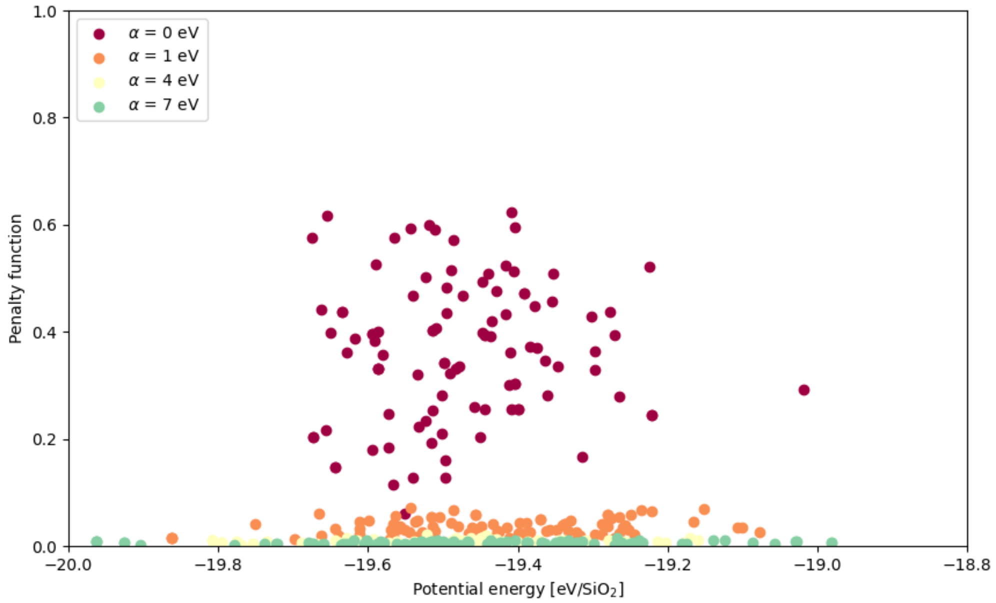

Fig. 3.6. Relationship between the potential energy and the penalty function of the obtained structures in the structure search for coesite using Tersoff potential as the potential energy and the correlation-coefficient-type penalty function of XRD. The vertical axis represents D, which takes a value from 0 to 1 excluding the coefficient αN in Eq. 2.1.

In order to find out the appropriate value of α in advance, we perform isothermal simulations at 10,000 K, and measure the mean squared displacement (MSD) for each atomic species and the penalty function. MSD is expressed by the following formula,

$$\mathrm{MSD} = \frac{1}{N}\sum_{i=1}^{N} |\boldsymbol{r}_i(t) - \boldsymbol{r}_i(t_0)|^2, \tag{3.1}$$

where subscript i is the index of atoms, t is the simulation time, t_0 is the start time, N is the number of atoms, and $\boldsymbol{r}$ is the atomic positions. In the initial stage of the simulation with a random structure as the initial structure, atoms move greatly in specific directions along the forces. In order to investigate the diffusion of atoms, t_0 must be set to a time after the sudden decrease in the energy.

Figure 3.7 shows the average MSD of Si and penalty function. We perform 10 short simulations for each α and average them at each step. As α increases, the penalty function is limited to be small, but at the same time, the diffusion range is narrowed. The diffusion range can be expanded by increasing the temperature, but the number of steps required for Simulated Annealing increases accordingly, and the computational cost also increases. Therefore, it is necessary to select a value of α that balances the limitation of the penalty function and the diffusion range of atoms. These pre-simulations do not give us α to maximize the success rate, but we can find out the appropriate range of α, with which the atoms move sufficiently and the penalty function is kept small enough. The diffusion range needs to be large enough to allow the atomic positions to be exchanged with each other. The average penalty function does not decrease any more for α above a certain value. In this case, we can roughly estimate that the appropriate value of α is around 7 eV. This result is consistent with the success rate in Fig. 3.4.

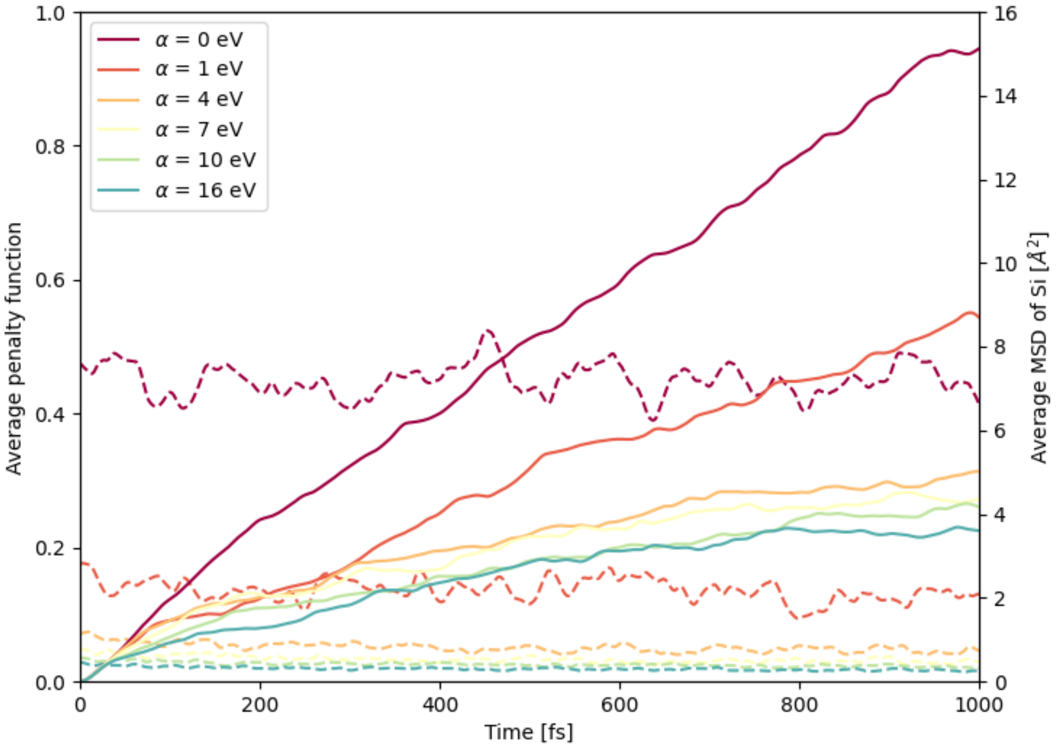

Fig. 3.7. Average MSD of Si and penalty function in isothermal simulations at 10,000 K using Tersoff potential as the potential energy and the correlation-coefficient-type penalty function of XRD. The solid lines represent the average MSD of Si on the right vertical axis for each α. The dotted lines represent the average penalty function on the left vertical axis for each α. Diffusion is measured from 100th step, and the origin on the horizontal axis is also set to 100th step.

3.1.3 XRD pattern with noise

In order to verify the noise-robustness of the structure prediction method based on data assimilation, we implement white Gaussian noise as the artificial noise on the reference XRD pattern. Compared to the crystallinity-type penalty function that uses the peak positions, the noise-robustness is the advantage of the correlation-coefficient-type one that uses the peak intensities as it is. Figure 3.8 shows the XRD pattern for coesite used in these test calculations. The magnitude of the artificial noise is set to bury small peaks other than two large peaks within the reference range of the diffraction angle. The peak width of the calculated XRD pattern I_{calc} is desirable to match that of the reference one I_{ref}, but in this case we set it wider to prevent overfitting. Other computational conditions for the structure optimization are the same as those in Section 3.1.1. We use Tsuneyuki potential as the potential energy.

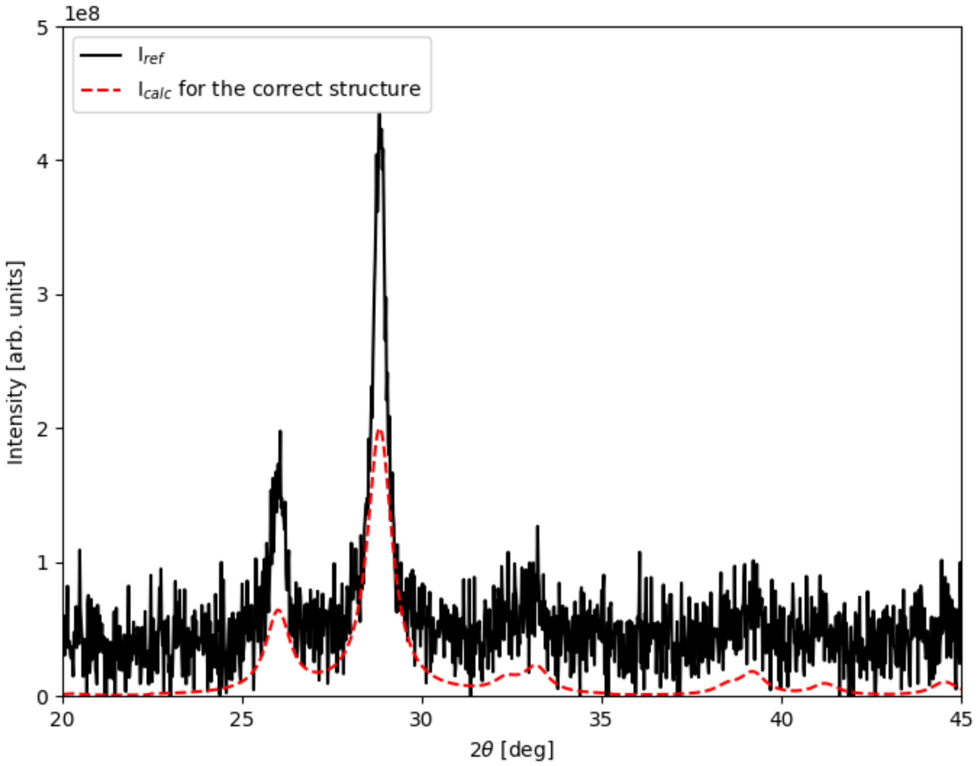

Fig. 3.8. Reference XRD pattern I_{ref} with artificial noise and the calculated one I_{calc} for coesite in the test structure search simulations. The diffraction peaks are smoothed with Lorentzian smearing function, where the scale parameters γ are 0.2 and 0.4 degrees for I_{ref} and I_{calc}, respectively. The white Gaussian noise on I_{ref} is randomly generated with the standard deviation $\sigma = 2 \times 10^7$ for each simulation. The reference range of the diffraction angle $[2\theta_{\mathrm{min}}, 2\theta_{\mathrm{max}}]$ is set to $[20, 45]$ degrees.

Figure 3.9 shows the results of the success rate in our simulations using the reference XRD pattern with artificial noise. We perform 100 simulations for each value of the control parameter α. Compared with tha case with the ideal reference XRD pattern without noise, the success rate is worse overall, but as high as 80 % at the maximum. From this result, even if there is a certain amount of noise on the reference data, we can improve the search efficiency by the structure prediction method based on data assimilation with the correlation-coefficient-type penalty function. If α is set too large, the success rate decrease due to the penalty function with the incomplete reference data. It is necessary to select an appropriate value of α that maximizes the success rate.

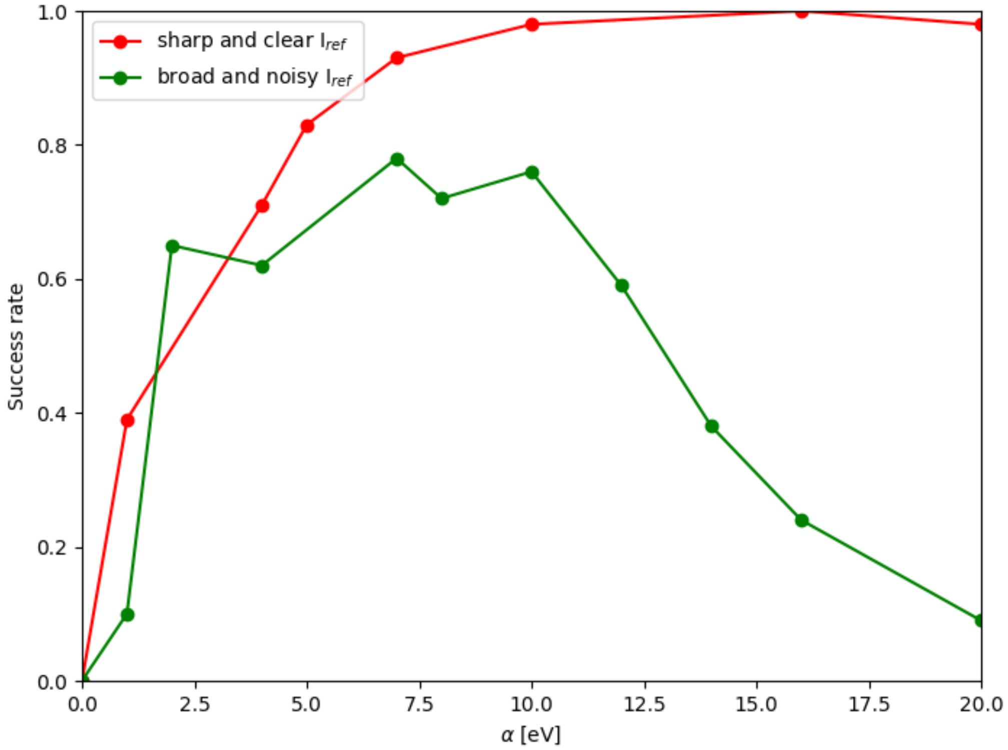

Fig. 3.9. Success rate of the structure search for coesite using Tsuneyuki potential as the potential energy and the correlation-coefficient-type penalty function of XRD with artificial noise. For comparison, the red line is the same as the success rate using the ideal reference XRD pattern without noise in Fig. 3.3.

In order to find out the appropriate value of α in advance, we perform isothermal simulations at 10,000 K. Figure 3.10 shows the average MSD of Si and penalty function for 10 short simulations for each α. Due to the noise on the reference XRD pattern and the different peak width between the reference and the calculated one, the penalty function does not become zero even with the correct structure. The overall trend is similar to Fig. 3.7, and we can roughly estimate that the appropriate value of α is around 7 eV. This result is also consistent with the success rate in Fig. 3.9.

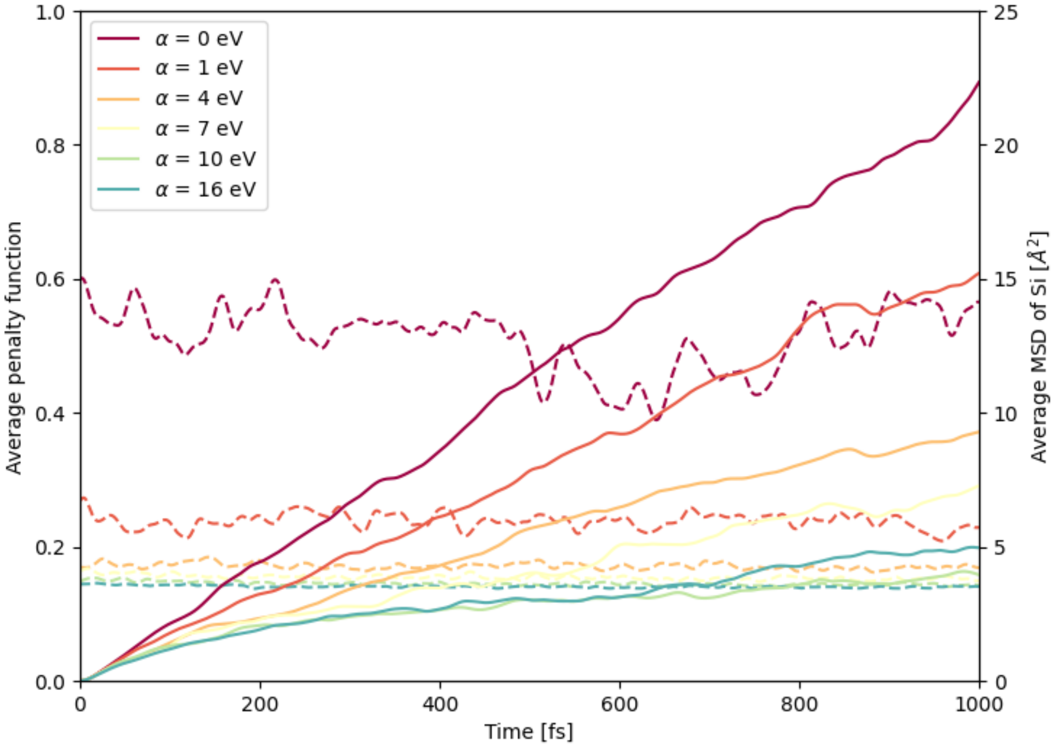

Fig. 3.10. Average MSD of Si and penalty function in isothermal simulations at 10,000 K using Tsuneyuki potential as the potential energy and the correlation-coefficient-type penalty function of XRD with artificial noise. The solid lines represent the average MSD of Si on the right vertical axis for each α. The dotted lines represent the average penalty function on the left vertical axis for each α. Diffusion is measured from 100th step, and the origin on the horizontal axis is also set to 100th step.

3.2 System containing hydrogen

Since hydrogen contributes little to the XRD pattern, the penalty function of XRD hardly restricts hydrogen positions. In this section, we perform the structure prediction based on data assimilation for ϵ-Zn(OH)$_2$ in order to verify its effectiveness in the system containing hydrogen. In these test calculations, we use the reactive force field (ReaxFF) [30] as the potential energy (cf. Section 2.2.2). In order to stabilize the optimization of hydrogen, we set the atomic mass of hydrogen to 12 u.

The heavier the atoms, the larger the contribution to the XRD pattern. There is a problem that the diffusion range of the heavier atoms is narrowed due to the restriction by the penalty function of XRD. In Section 3.2.2, we try to solve this problem by setting the different bath temperatures for each atomic species.

We implement the calculation of the neutron diffraction (ND) which can detect the positions of hydrogen in our code. In Section 3.2.3, we test the hybrid penalty function of XRD and ND in Zn-O-H system.

3.2.1 Test calculations for ϵ-Zn(OH)$_2$

We perform test calculations of the structure prediction based on data assimilation using the correlation-coefficient-type penalty function in Zn-O-H system. We use ReaxFF potential as the potential energy, and set ϵ-Zn(OH)$_2$ in the $2 \times 1 \times 1$ supercell as the target material (see Fig. 3.11). We take the supercell because the atoms directly reach the correct structure in the simulations with the primitive cell due to the small number of atoms, as in the case for coesite of Section 3.1.1.

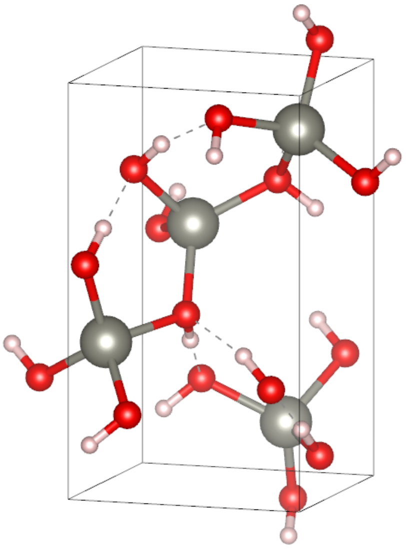

Fig. 3.11. Crystal structure of ϵ-Zn(OH)$_2$. The grey, red and pink spheres represent zinc, oxygen and hydrogen atoms, respectively. The primitive cell contains 20 atoms. The crystallographic data is shown in Appendix B.2.

We perform structure relaxation from the reference structure of ϵ-Zn(OH)$_2$ [39] using LAMMPS with ReaxFF potential, and set this relaxed structure as the correct one in the structure search. In these test calculations, instead of the actual experimental data, we use the calculated XRD pattern for the correct structure as the reference data in the penalty function. Figure 3.12 shows the XRD pattern for ϵ-Zn(OH)$_2$ used in these test calculations. This is the ideal case that the calculated XRD pattern I_{calc} for the correct structure is equal to the reference one I_{ref}.

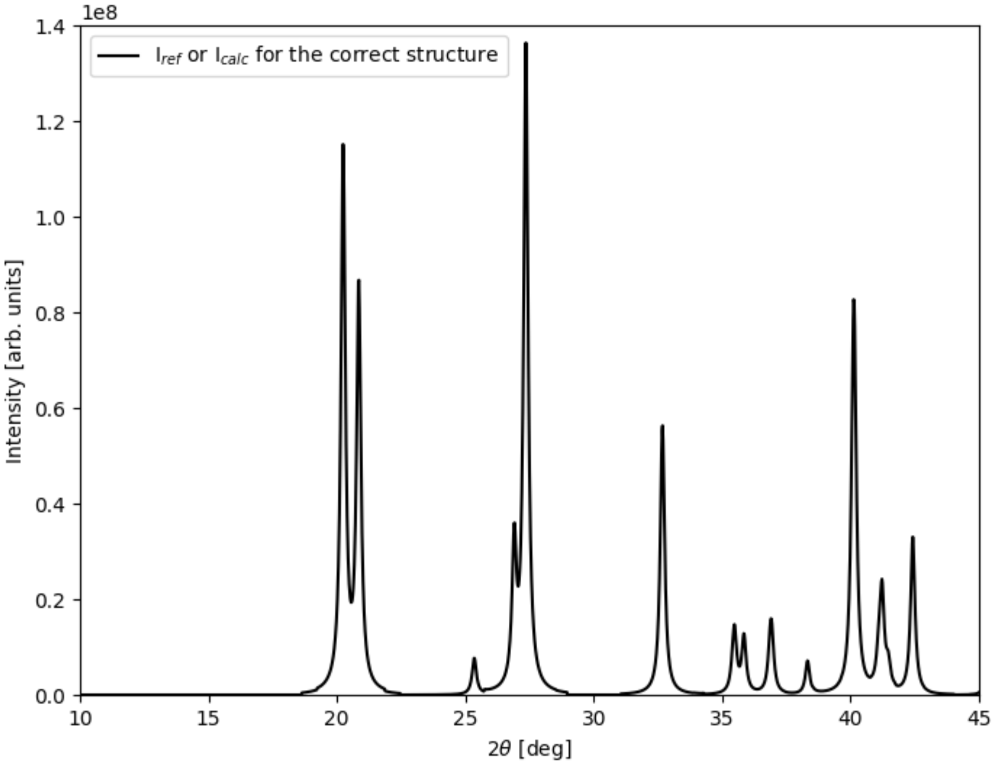

Fig. 3.12. Reference XRD pattern I_{ref} and the calculated one I_{calc} for ϵ-Zn(OH)$_2$ in the test structure search simulations. The wavelength λ is set to 1.54 $\AA$, which corresponds to CuKα-radiation. The diffraction peaks are smoothed with Lorentzian smearing function, where the scale parameter γ is 0.1 degree in Eq. A.9. The reference range of the diffraction angle $[2\theta_{\mathrm{min}}, 2\theta_{\mathrm{max}}]$ in Eq. 2.22 is set to $[10, 45]$ degrees.

Before the structure search, we perform isothermal simulations in order to find out the appropriate value of α, with which the atoms move sufficiently and the penalty function is small enough. Figure 3.13 shows the average MSD of Zn and penalty function for 10 short simulations for each α. The temperature is set to 20,000 K to move Zn atoms sufficiently for an appropriate α. Compared to the diffusion of Si in Figs. 3.7 and 3.10 for coesite, the diffusion of Zn is very large with small α due to the high temperature, but is greatly restricted by increasing α. Since Zn is heavier than Si and contributes more to the XRD pattern, the penalty function more strongly restricts the positions of Zn. From these pre-simulations, similar to the discussion in Fig. 3.7, we can roughly estimate that the appropriate value of α is around 7 eV.

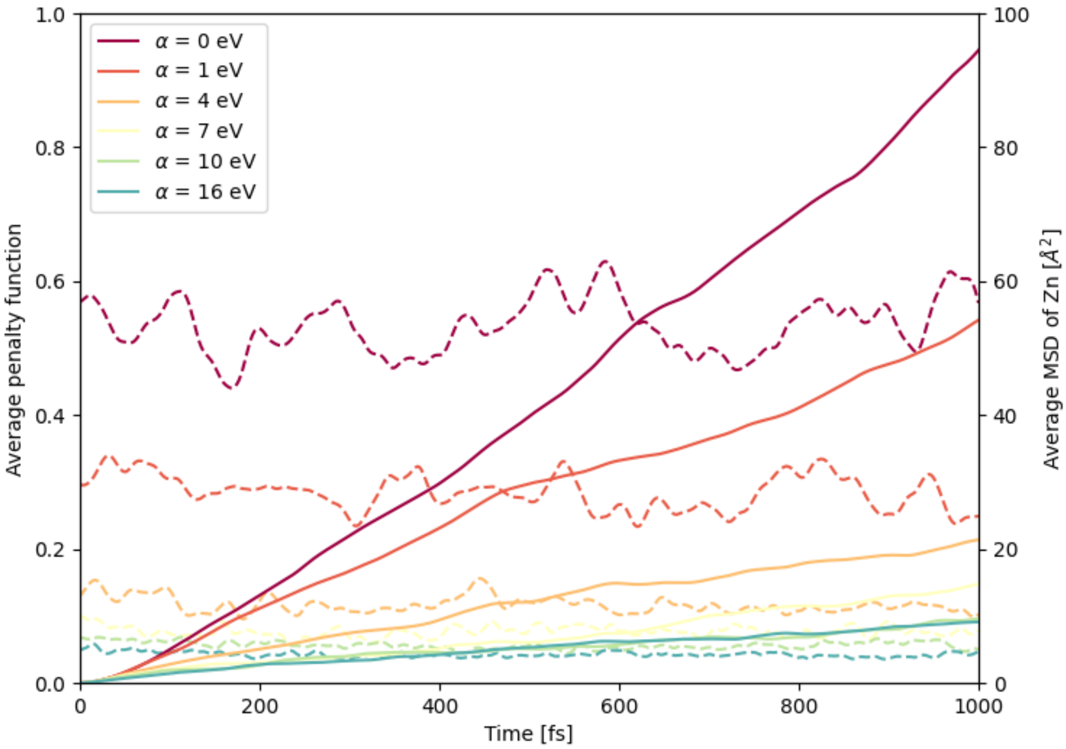

Fig. 3.13. Average MSD of Zn and penalty function in isothermal simulations at 20,000 K using ReaxFF potential as the potential energy and the correlation-coefficient-type penalty function of XRD. The solid lines represent the average MSD of Zn on the right vertical axis for each. The dotted lines represent the average penalty function on the left vertical axis for each. Diffusion is measured from 100th step, and the origin on the horizontal axis is also set to 100th step.

Then, we perform the structure optimization by simulated annealing using molecular dynamics. The initial temperature is set to 20,000 K, the temperature step is −2 K/step, and the time step is 1 fs/step. The cell parameters are fixed to those of the 2 × 1 × 1 supercell, and the initial structure of each simulation is randomly generated under the constraint that the interatomic distances are 0.5 $\AA$ or more.

Figure 3.14 shows the results of the success rate in our simulations. We perform 100 simulations for each value of α. The maximum success rate is about 4 %, and it can be seen that the structure prediction method based on data assimilation using the penalty function of XRD works even in the system containing hydrogen. The appropriate value of α is around 10 eV, which is consistent with the isothermal pre-simulations in Fig. 3.13.

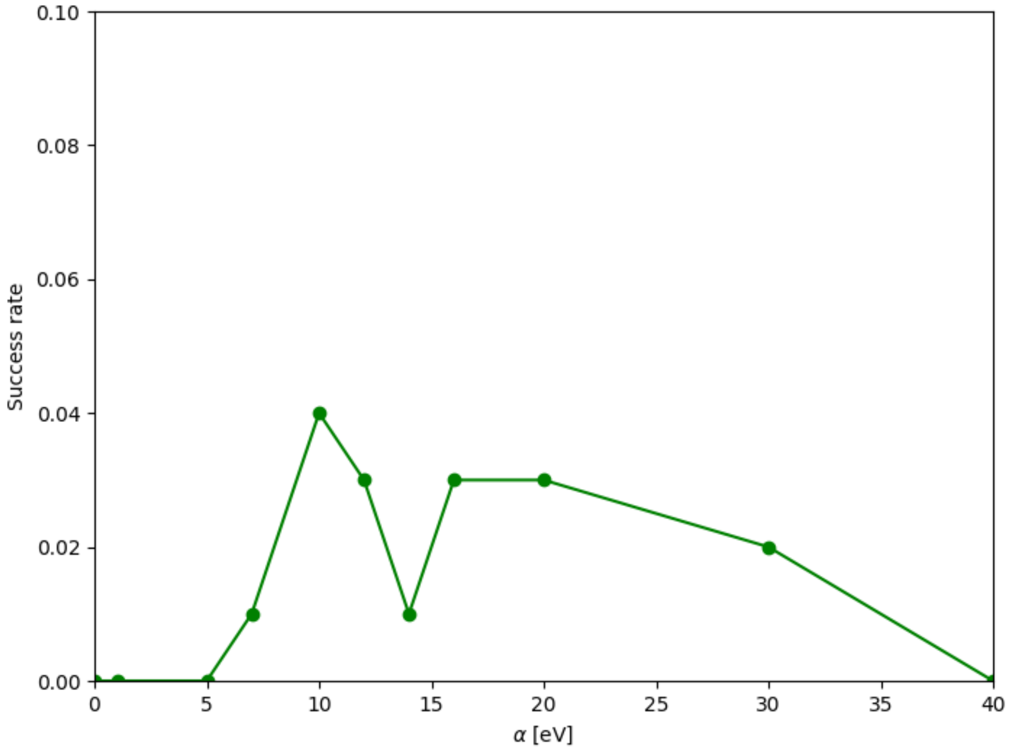

Fig. 3.14. Success rate of the structure search for ϵ-Zn(OH)$_2$ using ReaxFF potential as the potential energy and the correlation-coefficient-type penalty function of XRD.

Figure 3.15 shows the average MSD of O in the same simulations as in Fig. 3.13. It can be seen that the diffusion of O is less restricted by increasing α than that of Zn. Hydrogen atoms are hardly affected by the penalty function, and the diffusion of H also does not change much by increasing α. As a result, there is a large difference in the diffusion range of Zn and O or H. It is expected that by increasing the temperature to move Zn atoms sufficiently, O and H atoms move too much and the optimization efficiency deteriorates. In Section 3.2.2, we try to solve this problem by setting the higher bath temperatures for Zn atoms than that for O and H atoms.

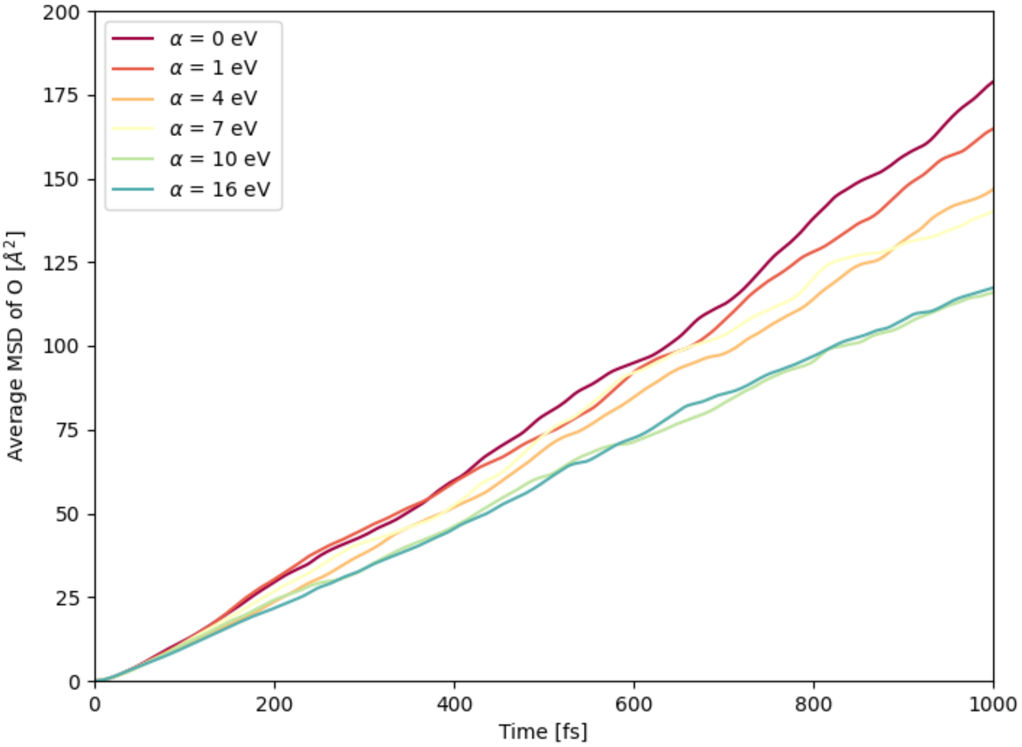

Fig. 3.15. Average MSD of O in isothermal simulations at 20,000 K using ReaxFF potential as the potential energy and the correlation-coefficient-type penalty function of XRD. These simulations are the same as in Fig. 3.13.

3.2.2 Temperature control for each atomic species

The correlation-coefficient-type penalty function of XRD mainly restricts positions of the heavier atoms that contribute more to XRD pattern. As can be seen in Fig. 3.13 and 3.15, if we set the temperature high enough to move heavier atoms, the lighter atoms move too violently. To avoid this problem, we set different bath temperatures for each atomic species in Simulated Annealing using Molecular Dynamics. That is, we control T_{ext} in Eq. 2.26 for the velocity scaling method separately for each atomic species.

Figure 3.16 shows the average MSD of Zn and penalty function of XRD in 10 short isothermal simulations for each α. Figure 3.17 shows the average MSD of O in the same simulations. The temperature for Zn is set to 20,000 K and that for O and H is set to 4,000 K. In an appropriate range of α, for which the penalty function is kept small, each atomic species has a similar diffusion range.

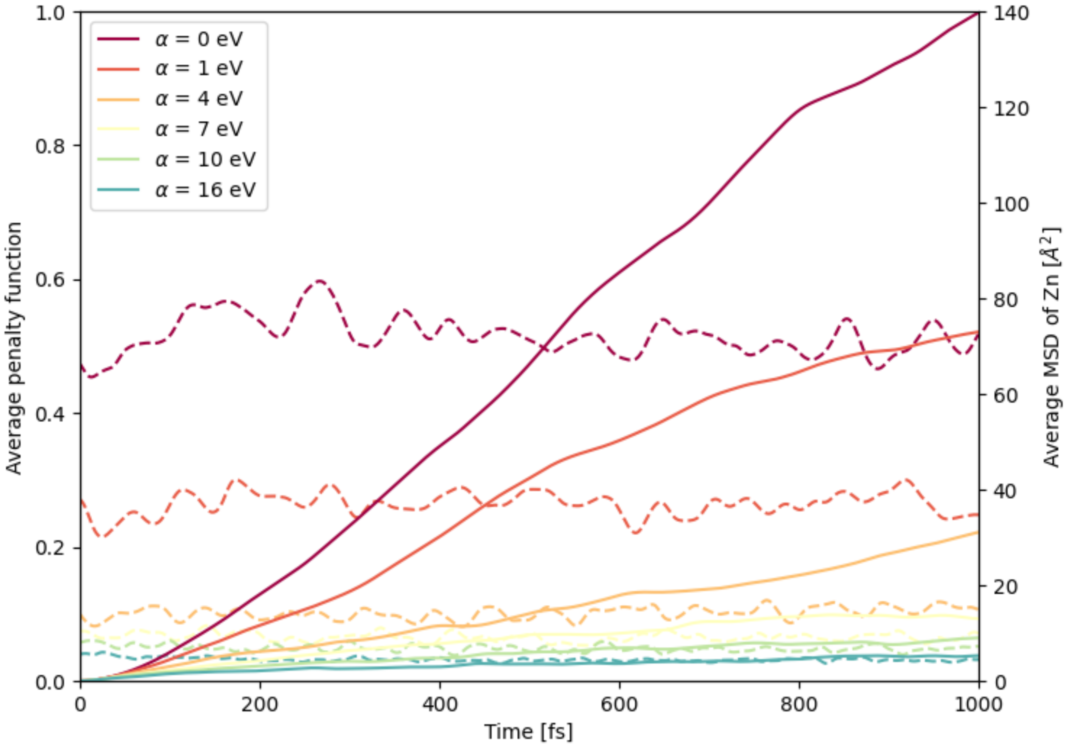

Fig. 3.16. Average MSD of Zn and penalty function in isothermal simulations at 20,000 K for Zn and 4,000 K for O and H using ReaxFF potential as the potential energy and the correlation-coefficient-type penalty function of XRD. The solid lines represent the average MSD of Zn on the right vertical axis for each α. The dotted lines representthe average penalty function on the left vertical axis for each α. Diffusion is measured from 100th step, and the origin on the horizontal axis is also set to 100th step.

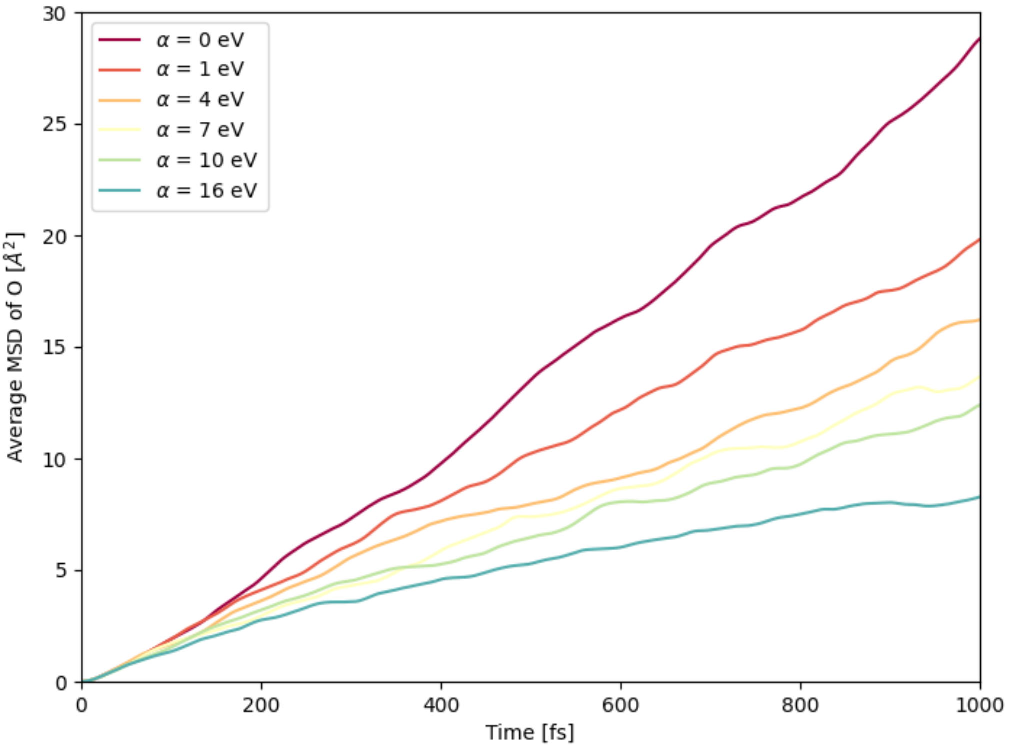

Fig. 3.17. Average MSD of O in isothermal simulations at 20,000 K for Zn and 4,000 K for O and H using ReaxFF potential as the potential energy and the correlation-coefficient-type penalty function of XRD. These simulations are the same as in Fig. 3.16.

Then, we perform the structure optimization by simulated annealing using molecular dynamics with the temperature control for each atomic species. We set the initial temperature for Zn to 20,000 K and that for O and H to 4,000 K, and gradually decrease the temperature for all atoms to zero in 10,000 step. Figure 3.18 shows the results of the success rate in our simulations. We perform 100 simulations for each value of α. By controlling the temperature for each atomic species and adjusting each diffusion range to the same extent, we can improve the search efficiency as a whole. The reason why the success rate drops with α around 14 eV is considered that the number of local minima on the cost function changes depending on α.

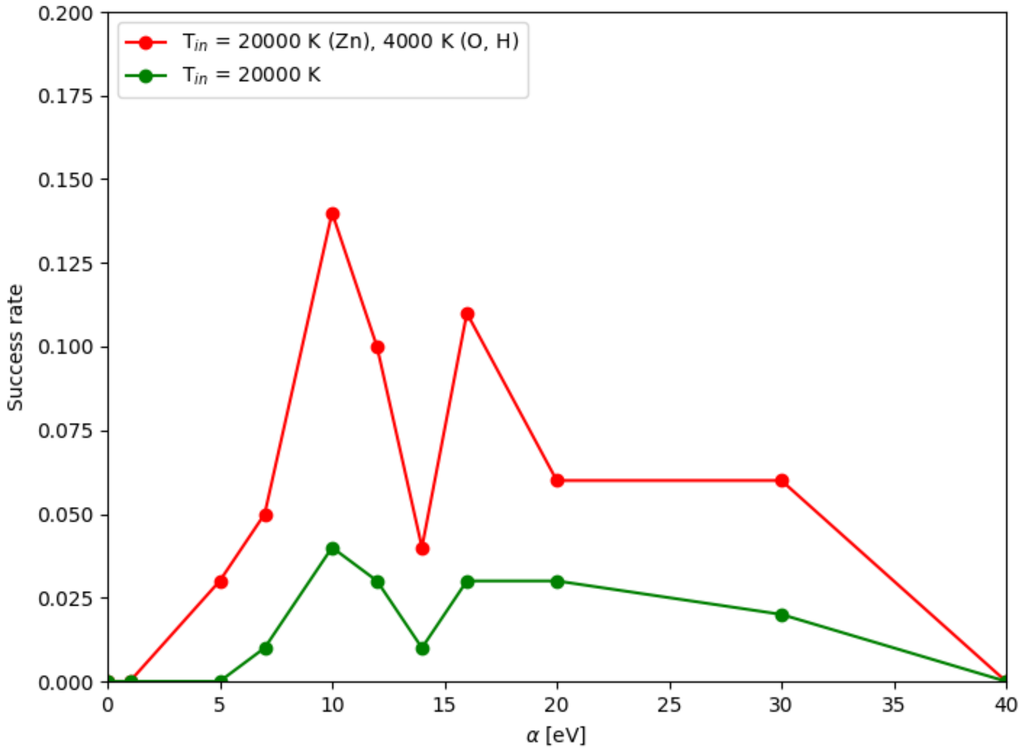

Fig. 3.18. Success rate of the structure search for ϵ-Zn(OH)$_2$ with the temperature control for each atomic species, using ReaxFF potential as the potential energy and the correlation-coefficient-type penalty function of XRD. For comparison, the green line is the same as the success rate with the normal temperature control in Fig. 3.14.

3.2.3 Penalty function of the neutron diffraction

In the X-ray diffraction, the atomic form factor of hydrogen is much smaller than that of a heavy atom. In the neutron diffraction, it is relatively large and hydrogen atoms can be detected. If we have the reference data of ND, we can define the penalty function of ND as well as that of XRD. Furthermore, given both XRD and ND, we can introduce the hybrid penalty function of XRD and ND as follows,

$$D = \frac{1}{2}(D_{\mathrm{XRD}} + D_{\mathrm{ND}}), \tag{3.2}$$

where D_{XRD} is the penalty function of XRD and D_{ND} is that of ND. The ratio of D_{XRD} to D_{ND} is arbitrary, but we set it to 1:1 in our simulations. By minimizing the cost function with this hybrid penalty function, we perform the multi-objective optimization for the potential energy, XRD and ND. Figure 3.19 shows the ND pattern for ϵ-Zn(OH)$_2$ used in these test calculations. The XRD pattern used is the same as that showed in Fig. 3.12, and the reference range of the diffraction angle for ND corresponds to that for XRD.

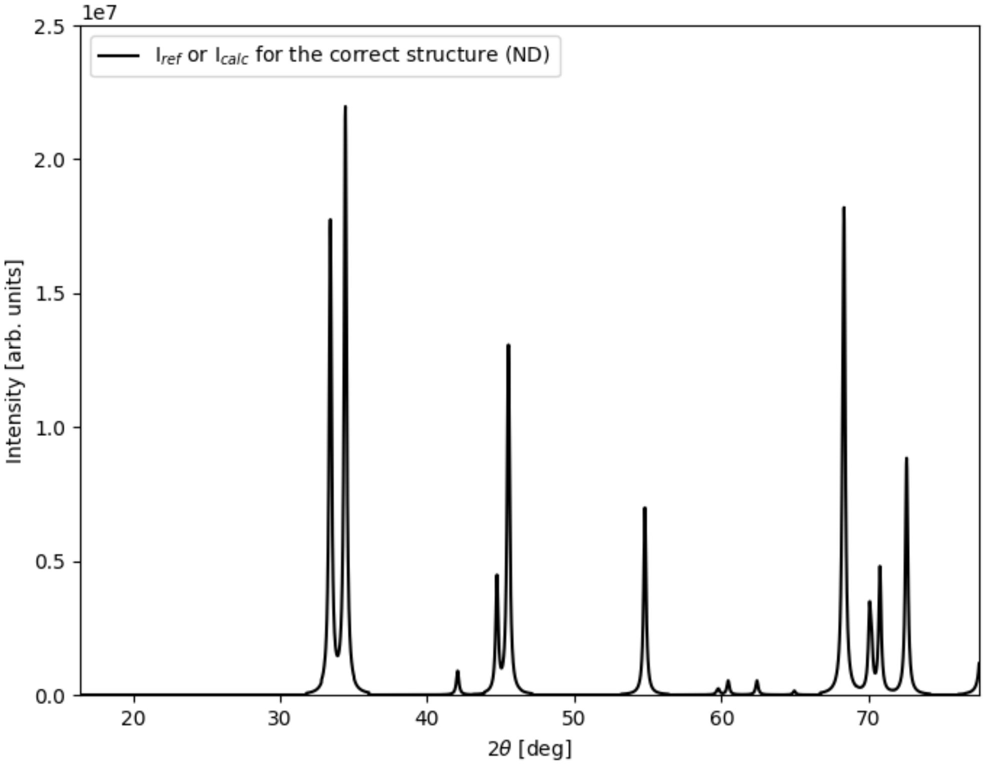

Fig. 3.19. Reference ND pattern I_{ref} and the calculated one I_{calc} for ϵ-Zn(OH)$_2$ in the test structure search simulations. The wavelength λ is set to 2.52 $\mathring{A}$, which corresponds to the neutron radiation. The diffraction peaks are smoothed with Lorentzian smearing function, where the scale parameter γ is 0.1 degree in Eq. A.9. The reference range of the diffraction angle $[2\theta_{\mathrm{min}}, 2\theta_{\mathrm{max}}]$ in Eq. 2.22 is set to $[16.4, 77.5]$ degrees.

Figure 3.20 shows the average MSD of Zn and penalty function of XRD in 10 short isothermal simulations for each α. Figure 3.21 shows the average MSD of O and penalty function of ND in the same simulations. As in Section 3.2.2, the temperature for Zn is set to 20,000 K and that for O and H is set to 4,000 K. Both the penalty functions of XRD and ND can be sufficiently suppressed if α is set to around 8 eV.

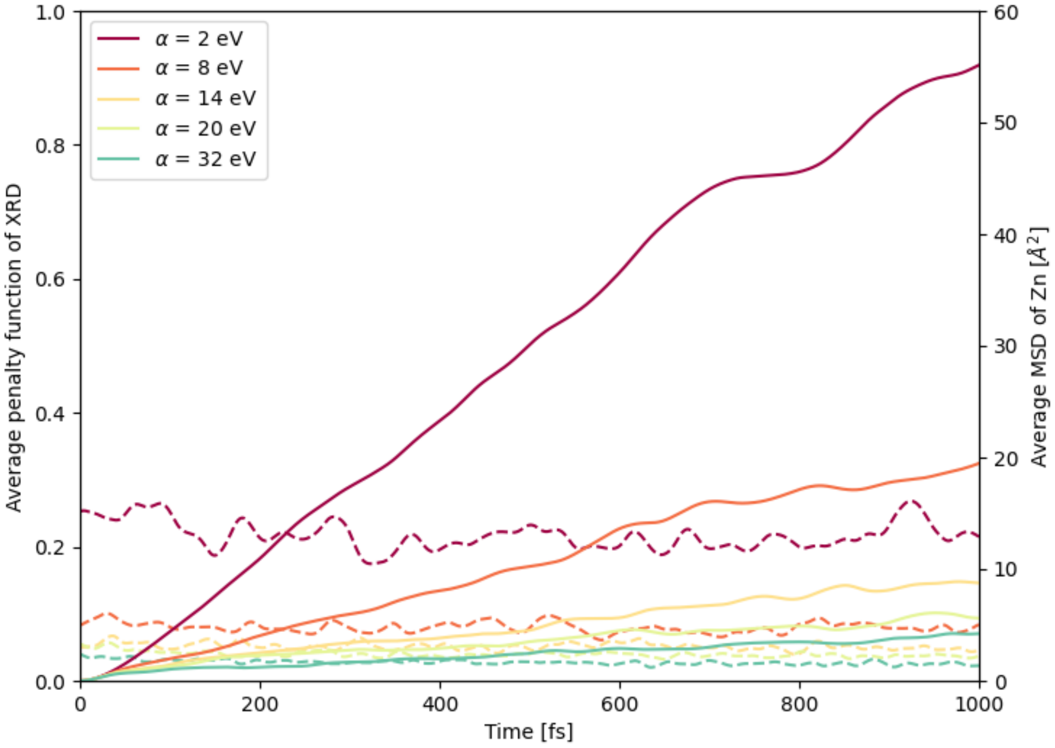

Fig. 3.20. Average MSD of Zn and penalty function of XRD in isothermal simulations at 20,000 K for Zn and 4,000 K for O and H using ReaxFF potential as the potential energy and the correlation-coefficient-type penalty function of XRD. The solid lines represent the average MSD, and the dotted lines representthe average penalty function. Diffusion is measured from 100th step, and the origin on the horizontal axis is also set to 100th step.

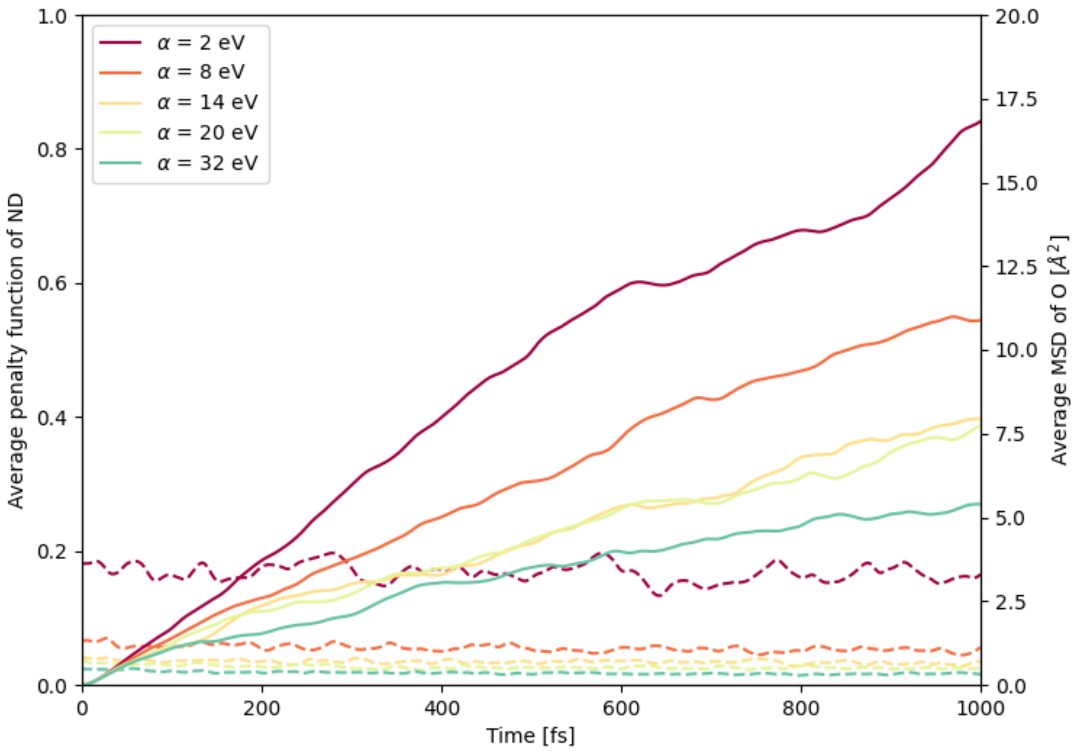

Fig. 3.21. Average MSD of O and penalty function of ND in isothermal simulations at 20,000 K for Zn and 4,000 K for O and H using ReaxFF potential as the potential energy and the correlation-coefficient-type penalty function of XRD. The solid lines represent the average MSD, and the dotted lines representthe the average penalty function. These simulations are the same as in Fig. 3.20.

Then, we perform the structure optimization by simulated annealing using molecular dynamics with the hybrid penalty function. As in Section 3.2.2, we set the initial temperature for Zn to 20,000 K and that for O and H to 4,000 K, and gradually decrease the temperature for all atoms to zero in 10,000 step. Figure 3.22 shows the results of the success rate in our simulations. We perform 100 simulations for each value of α. The hybrid penalty function of XRD and ND improves the search efficiency in the system containing hydrogen. In the case of the hybrid penalty function, the dip in the success rate with α around 14 eV disappears probably because the penalty function of XRD and that of ND have different local minimum points.

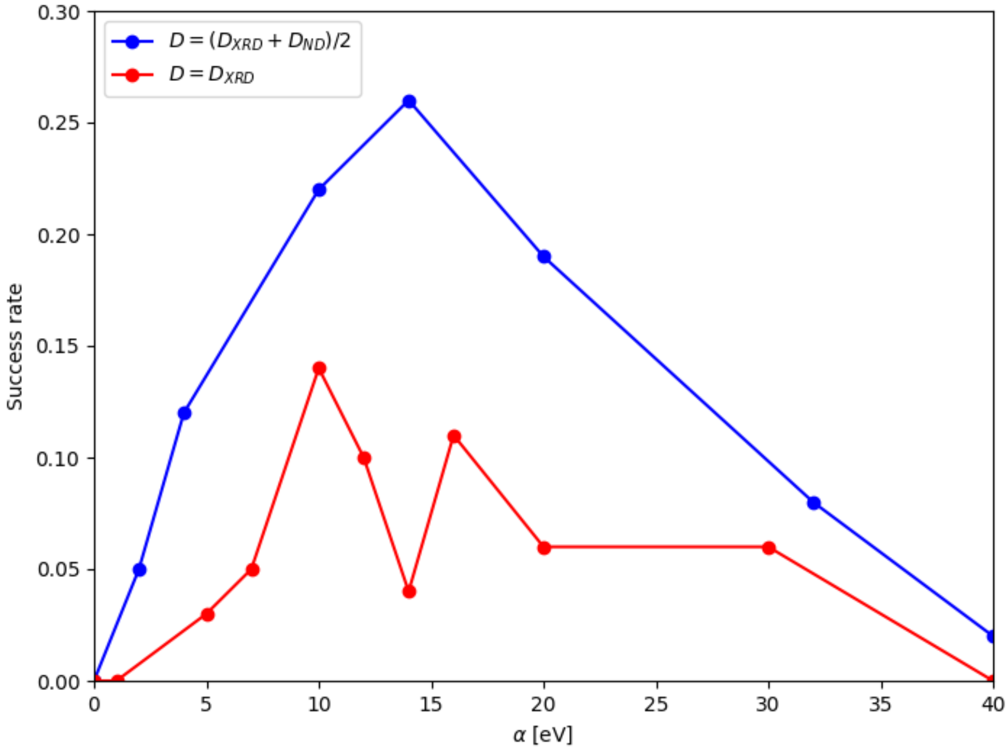

Fig. 3.22. Success rate of the structure search for ϵ-Zn(OH)$_2$ with the temperature control for each atomic species, using ReaxFF potential as the potential energy and the correlation-coefficient-type hybrid penalty function of XRD and ND. For comparison, the green line is the same as the success rate with only the penalty function of XRD in Fig. 3.18.

3.3 New hydride in high-pressure synthesized Al-Ca-H system

As a practical application of the structure prediction based on data assimilation, we perform the structure search in Al-Ca-H system with actual experimental data. This is a collaborative research with Dr. Hiroyuki Saitoh (QST) and Prof. Shin-ichi Orimo (Tohoku Univ., WPI-AIMR/IMR). They discover a new hydride in high-pressure synthesized Al-Ca-H system, and propose us the structure prediction for it with their experimental data.

Section 3.3.1 shows experimental results for the new hydride in Al-Ca-H system, and the cell parameter estimation based on them. Section 3.3.2 describes the discovery of new $Al_{12}Ca_{20}H_{76}$ structure based on data assimilation and first-principles calculations.

3.3.1 Experimental results and cell parameter estimation

The group of Dr. Saitoh and Prof. Orimo discovers a new hydride by hydrogenating Al-Ca alloys under high temperature (670 °C or more) and high pressure (5 GPa or more). This new hydride is a candidate hydrogen storage materials, and the composition ratio has not yet been determined.

They hydrogenate Al-Ca alloys with different mixing ratios, and measure a yield of the new hydride. Since the yield is the highest when AlCa and $AlCa_2$ alloys are hydrogenated, Al-Ca ratio of the new hydride is estimated to be 1:1 to 1:2. In addition, they heat the target materials at a ordinary pressure, and measure weight changes due to hydrogen release. As a result, the composition ratio of the new hydride is estimated to be $AlCaH_6$ to $AlCa_2H_{12}$. This indicates that the new hydride may contain excess hydrogen with respect to Al and Ca. The mass density of the new hydride is measured to be 1.92 ± 0.4 g/cm^3. The error of it is large due to small samples and water absorption.

Figure 3.23 shows the experimental XRD pattern for the target material, which is obtained by hydrogenating $AlCa_2$ alloy. Comparing this with the XRD pattern for known materials in Al-Ca-H system, the target material is considered to be multiphase including pure Al. No other known material has been found so far that corresponds to peaks in this experimental XRD pattern.

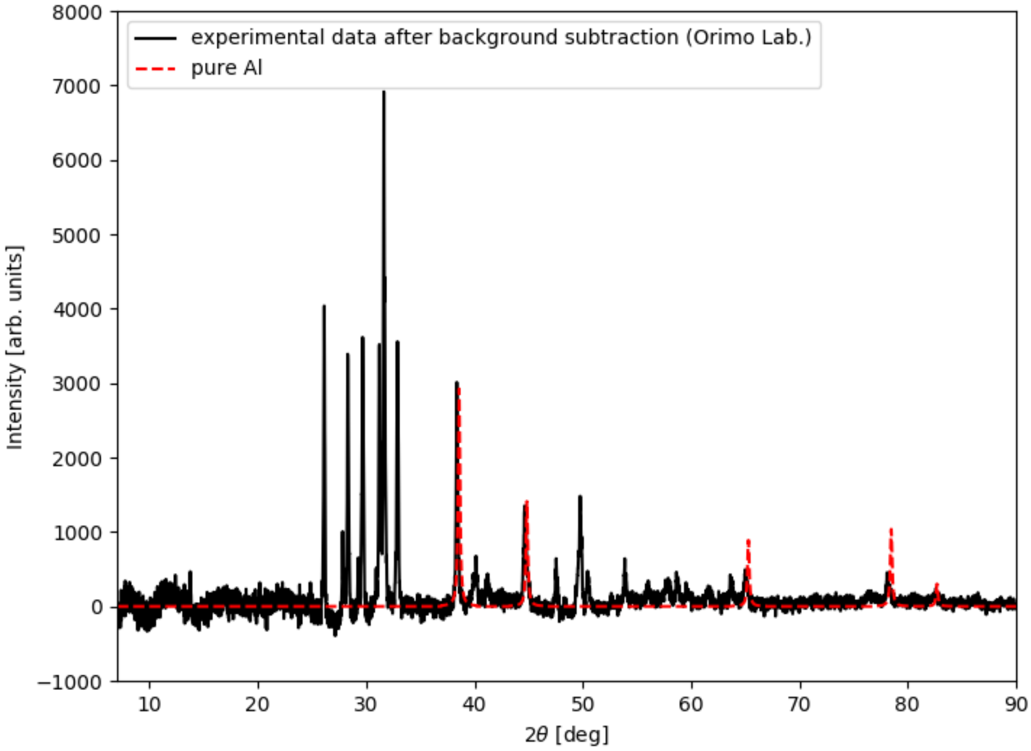

Fig. 3.23. Experimental XRD pattern (CuKα radiation) for the new hydride $AlCa_xH_y$ measured in Orimo Lab. For comparison, the red dotted line represents the calculated XRD pattern for pure Al.

Based on the peak positions in the experiment, we estimate the cell parameters of the new hydride. Assuming an orthorhombic system, the cell parameters are estimated to be $(a, b, c) = (12.81, 6.405, 6.815)\,\mathring{A}$, with which the possible peak positions match the experimental XRD pattern well as shown in Fig. 3.24. We set $a = 2b$ for higher symmetry. In the orthorhombic system, cells smaller than this do not match the experiment.

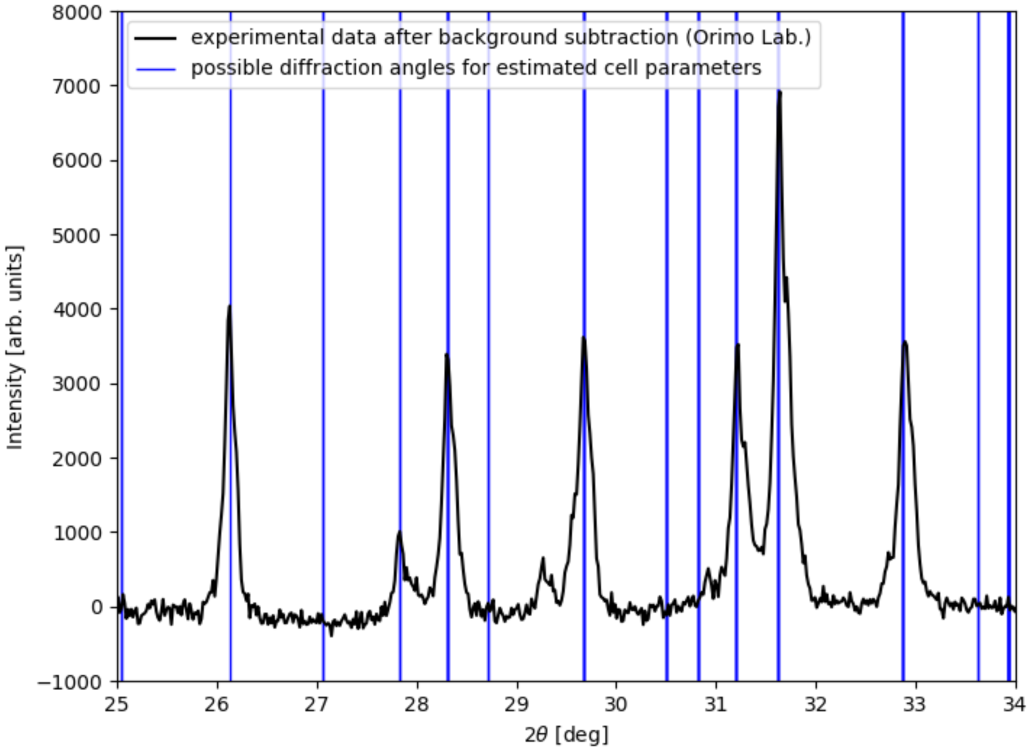

Fig. 3.24. Possible diffraction angles for estimated cell parameters. The experimental XRD pattern for the new hydride $\mathrm{AlCa}_x\mathrm{H}_y$ is the same as that in Fig. 3.23. At larger diffraction angles, the possible peak positions appear at short intervals and not suitable for comparison.

3.3.2 Discovery of a new structure by using data assimilation

We perform the structure prediction based on data assimilation for the new hydride $\mathrm{AlCa}_x\mathrm{H}_y$ in the estimated cell. It is necessary to search with all possible composition ratios to the estimated cell volume and the measured mass density. Since even the primitive cell contains about 50 atoms, the first-principles calculations for the potential energy require enormous costs. As in Chapter 3.1 and 3.2, we first perform Simulated Annealing using Molecular Dynamics to minimize the cost function with the potential energy based on DFT and the correlation-coefficient-type penalty function of XRD. However, due to insufficient number of trials, these structure searches cannot find a candidate structure.

Then, similar to the approach by Putz et al. [15], we search for structures that reproduces the experimental XRD pattern using a model potential as the potential energy. Assuming that hydrogen atoms are ignorable for the XRD pattern, we set the simulation cells with only Al and Ca atoms. We use the EAM potential for Al-Mg system [40] as the potential energy for Al-Ca system. We also use the hybrid penalty function of XRD as follows,

$$\alpha D = \alpha_1 D_{\mathrm{cryst}} + \alpha_2 D_{\mathrm{corr}}, \tag{3.3}$$

where D_{cryst} is the crystallinity-type penalty function, D_{corr} is the correlation-coefficient-type one, and α is the control parameter. The correlation-coefficient-type one emphasizes large peaks, while the crystallinity-type one focuses on disappeared peaks due to the extinction law. We perform SA using MD with possible Al-Ca compositions in the $1 \times 2 \times 2$ supercell, that is, $(a, b, c) = (12.81, 12.81, 13.63)\,\mathring{A}$.

In these simulations, we find $\mathrm{Al}_{24}\mathrm{Ca}_{40}$ structure with relatively high symmetry as shown in Fig. 3.25. Due to the translational symmetry in c-axis direction, the primitive cell of this structure is the $1 \times 2 \times 1$ supercell for the originally estimated cell parameters. Figure 3.26 shows the XRD pattern for this structure. Compared to the experimental XRD pattern, although there are some differences in peak intensities, it satisfies the extinction rule well. The differences may be due to the effect of preferred orientation.

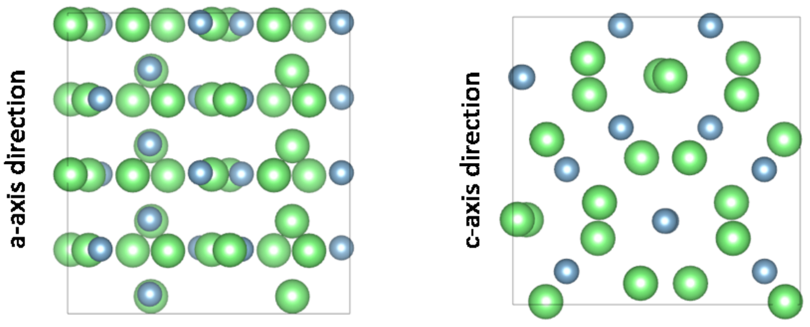

Fig. 3.25. Crystal structure for $\mathrm{Al}_{24}\mathrm{Ca}_{40}$ discovered in our structure search. The left figure is from a-axis direction, and the right figure is from c-axis direction. The blue and green spheres represent aluminum and calcium atoms, respectively. The cell parameters are $(a, b, c) = (12.81, 12.81, 13.63)\,\mathring{A}$, and this structure has translational symmetry in c-axis direction.

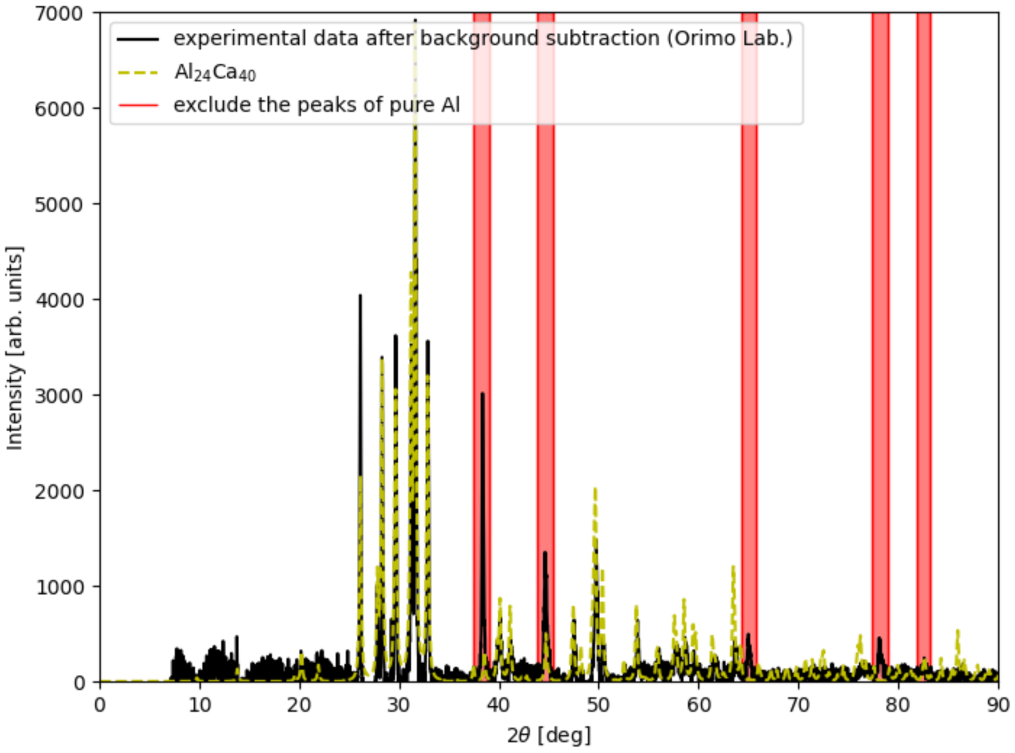

Fig. 3.26. Calculated XRD pattern for the $Al_{24}Ca_{40}$ structure (yellow dotted line). The wavelength λ is set to 1.54 $\AA$, which corresponds to CuKα-radiation. The diffraction peaks are smoothed with Lorentzian smearing function, where the scale parameter γ is 0.1 degree in Eq. A.9. The red bands contain the peaks corresponding to pure Al, and are excluded from the reference range of the diffraction angle in the penalty function.

We arrange 76 hydrogen atoms in this $Al_{12}Ca_{20}$ structure so that the coordination number of Al is 6. The composition is set to $AlH_3 \times 12 + CaH_2 \times 20$ for ease of comparison with known structures, and the amount of hydrogen is less than that estimated in the experiment. We perform the structure relaxation from this structure with fixed cell parameters based on first-principles calculations, and obtain $Al_{12}Ca_{20}H_{76}$ structure as shown in Fig. 3.27. In first-principles calculations based on DFT, pseudopotentials are PAW_GGA, and the k-point mesh is set to Γ-centered 2 × 2 × 4 grid. The positions of Al and Ca do not change much from those in the $Al_{24}Ca_{40}$ structure. Figure 3.28 shows the XRD pattern for this structure. Since hydrogen atoms hardly contributes to the XRD pattern, this is almost the same as the XRD pattern for the $Al_{24}Ca_{40}$ structure in Fig. 3.26.

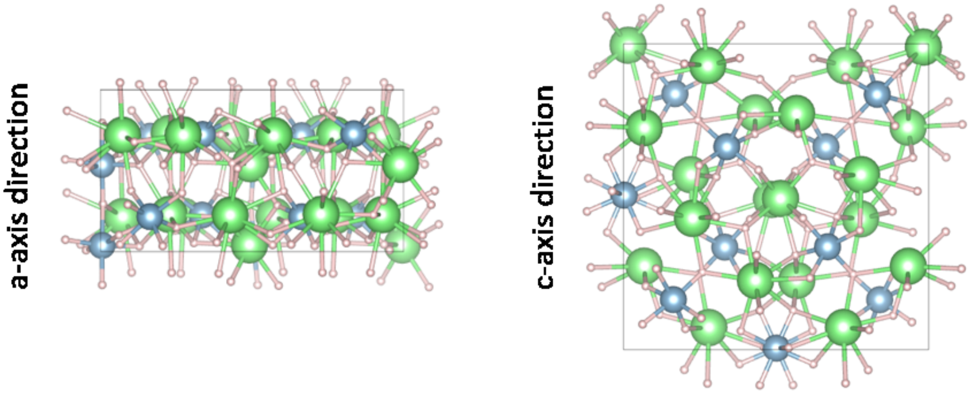

Fig. 3.27. Crystal structure for $Al_{12}Ca_{20}H_{76}$ after the structure relaxation. The left figure is from a-axis direction, and the right figure is from c-axis direction. The blue, green and pink spheres represent aluminum, calcium and hydrogen atoms, respectively. The cell parameters are $(a, b, c) = (12.81, 12.81, 6.815)\,\mathring{A}$. The crystallographic data is shown in Appendix B.3.

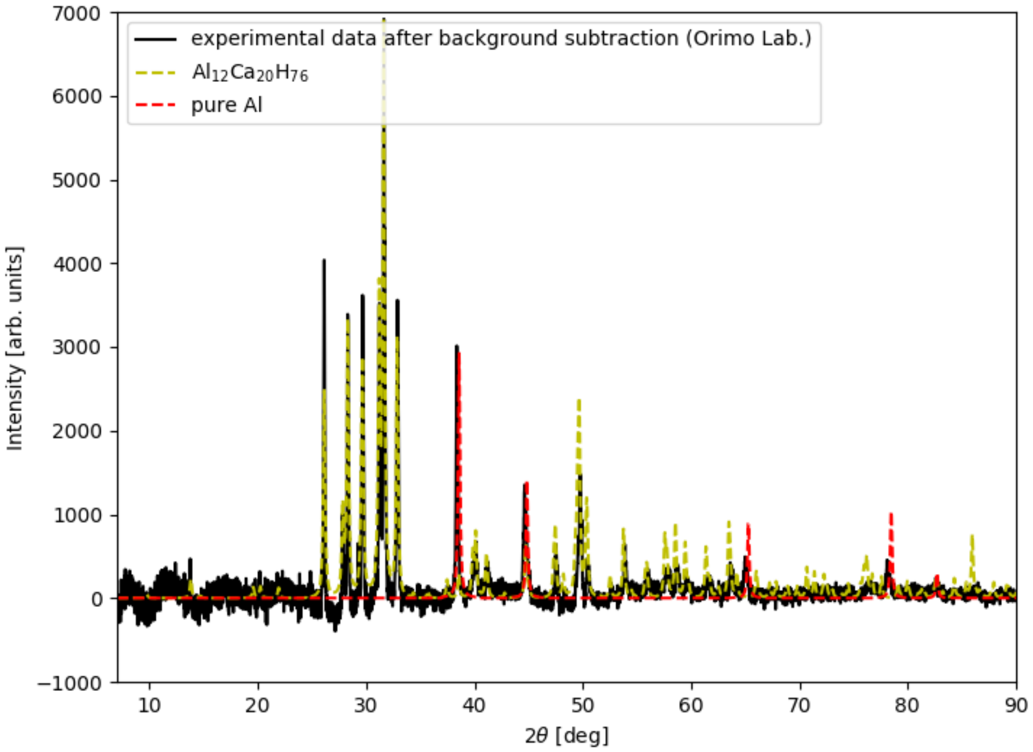

Fig. 3.28. Calculated XRD pattern for the $Al_{24}Ca_{40}H_{76}$ structure (yellow dotted line). The wavelength λ is set to 1.54 $\mathring{A}$, which corresponds to CuKα-radiation. The diffraction peaks are smoothed with Lorentzian smearing function, where the scale parameter γ is 0.1 degree in Eq. A.9. For comparison, the red dotted line represents the calculated XRD pattern for pure Al.

In order to compare the stability with known structures in Al-Ca-H system, we perform the structure relaxation for the $Al_{12}Ca_{20}H_{76}$ structure in a variable cell with higher accuracy. The k-point mesh is set to Γ-centered $4 \times 4 \times 8$ grid. The cell parameters are relaxed to $(a, b, c) = (12.71, 12.73, 6.75)\,\AA$, and the cell volume decreases by about $3\,\%$. This indicates that this structure may contain more hydrogen. In the composition ratio of $Al_{12}Ca_{20}H_{76}$, $AlCaH_5$ (Pnnm)+CaH_2 (Pnma) is the most stable in the combinations of known structures. In our calculations, the new $Al_{12}Ca_{20}H_{76}$ structure has about 2.6 meV/atom higher energy than this composite material. As a result, we discover the candidate $Al_{12}Ca_{20}H_{76}$ structure that is almost as stable as known structures and somewhat consistent with the experiment.

The reason why some peaks in the experiment are not reproduced may be that Al-Ca ratio is wrong or the simulation cell is small. Since the computational costs of DFT potential are too high, we have not yet searched for candidate structures with different compositions in larger cells. It is a future task to expand the search range. One way is to construct the Neural Network (NN) potential in Al-Ca-H system. By replacing DFT potential with NN potential, the computational costs are significantly reduced.

Chapter 4

Conclusion and Outlook

In this study, we introduce the new correlation-coefficient-type penalty function of the diffraction pattern, and work on the improvement and application of the structure prediction method based on data assimilation. The practicality of this method is remarkably improved by the improvement of the success rate and noise-robustness, the pre-determination of the control parameter α, and the temperature control for each atomic species. This chapter describes the summary of outcomes and future improvements of our method.

4.1 Outcomes of the correlation-coefficient-type penalty function

In the test calculations for coesite, the correlation-coefficient-type penalty function significantly improves the search efficiency over the crystallinity-type one used in the previous study [1]. Comparing the cases where Tsuneyuki potential and Tersoff potential are used as the model potentials for Si-O system, the success rate with Tersoff potential is considerably lower than that with Tsuneyuki potential. The complexity of the potential energy is considered to be the cause of this. The search efficiency may be further deteriorated when we use the potential energy obtained by first-principles calculations. By adding artificial noise to the reference diffraction pattern, we confirm the noise-robustness of our method. The correlation-coefficient-type penalty function works well even if we use the reference data with noise as it is.

In the test calculations for ϵ-$\mathrm{Zn(OH)_2}$, although XRD pattern is nearly independent of hydrogen positions, the penalty function of XRD is effective to some extent. The penalty function restrict the positions of other atoms, and the potential energy determines hydrogen positions associated with them. If we have the reference ND pattern detecting hydrogen positions, the hybrid penalty function of XRD and ND is more effective. The diffraction pattern cannot distinguish the atomic species corresponding to the peaks, and we can improve the search efficiency by assimilating the reference data of different probes at the same time.

Generally, in order to improve the search efficiency by our method, it is necessary to set an appropriate value of the control parameter α in Eq. 2.1. The penalty function restricts the search space to match the reference data, but narrows the diffusion range of atoms. By performing isothermal pre-simulations with different α, we can estimate an appropriate range of α such that the penalty function is kept low and the atoms move sufficiently.

The heavier the atoms, the greater the contributions to the XRD pattern and the stronger the forces from the penalty function. For the system containing the atomic species with significantly different weights from each other, if we set an appropriate α for heavier atoms to move sufficiently, the diffusion range of lighter atoms becomes too large. In such a case, we can improve the search efficiency by setting different bath temperatures for each atomic species so that the diffusion ranges are about the same.

We performed the structure prediction for new hydride in high-pressure synthesized Al-Ca-H system based on first-principles calculations and actual experimental XRD pattern. We search for Al-Ca structures that match the experimental data based on data assimilation using a model potential as the potential energy. Then, we add hydrogen atoms in a obtained Al-Ca structure, and perform the structure relaxation based on first-principles calculations. As a result, we discovered new $\mathrm{Al_{12}Ca_{20}H_{76}}$ structure that is almost as stable as known structures and somewhat consistent with the experiment.

4.2 Expandability of the structure prediction method based on data assimilation

In order to succeed in the structure search in large systems, it is necessary to combine our structure prediction method with other approaches. The data assimilation scheme can be incorporated into existing structure prediction methods by replacing the potential energy with the cost function. One way is to use more advanced structure optimization methods than SA using MD in order to improve the efficiency of the global optimization. Another way is to construct the NN potential with the training data of DFT potential in order to reduce the computational costs of the potential energy.

In our method, we search for the common minimum of the potential energy and the penalty function. Unlike the conventional multi-objective optimization, Pareto optimal solutions should be avoided. In the previous study [2], the Combined Optimization Method (COM) is proposed for such multi-objective optimization.

In order to expand the applicable range of our method, it is possible to assimilate other experimental data than XRD and ND. The reference data should reflect the atomic arrangement for the structure prediction. The positron diffraction for surface analysis is one such experimental data.

We also need an implementation to make our method used widely and conveniently. One way is to incorporate the penalty function calculation into existing packages of the potential energy calculation.

Acknowledgments

Foremost, I would like to express my sincere gratitude toward my supervisor Prof. Shinji Tsuneyuki for his insightful suggestions and continuous support. I am deeply grateful to Assistant Prof. Ryosuke Akashi for his encouragement and helpful comments. I also thank all the members of Tsuneyuki research group.

I would like to show my greatest appreciation to my research collaborators. Dr. Naoto Tsujimoto and Dr. Ryuhei Sato in Tsuneyuki Lab., Prof. Synge Todo, and Dr. Daiki Adachi in Todo Lab. have developed the structure prediction method based on data assimilation with me. Dr. Hiroyuki Saitoh (QST) and Prof. Shin-ichi Orimo (Tohoku Univ., WPI-AIMR/IMR) propose us the structure search for the new hydride in Al-Ca-H system with their experimental data.

I would like to offer my special thanks to MERIT program for leading graduate schools, especially my supervisor Prof. Masashi Takigawa. Not to mention the financial support, discussions with researchers in other fields and exchanges with non-researchers through MERIT activities are a great asset to me.

Finally, I would like to express my thanks to my family for their great long-term support. In particular, I would like to express my special gratitude to my mother, who always encourage me despite fighting against her serious disease.

Bibliography

[1] N. Tsujimoto, D. Adachi, R. Akashi, S. Todo, and S. Tsuneyuki. Crystal structure prediction supported by incomplete experimental data. *Physical Review Materials*, 2:053801, 2018.

[2] D. Adachi, N. Tsujimoto, R. Akashi, S. Todo, and S. Tsuneyuki. Search for common minima in joint optimization of multiple cost functions. *Computer Physics Communications*, 241:92, 2019.

[3] H. M. Rietveld. A profile refinement method for nuclear and magnetic structures. *Journal of Applied Crystallography*, 2:65, 1969.

[4] R. L. McGreevy and L. Pusztai. Reverse monte carlo simulation: a new technique for the determination of disordered structures. *Molecular Simulation*, 1:359, 1988.

[5] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671, 1983.

[6] D. J. Wales and J. P. Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101:5111, 1997.

[7] S. Goedecker. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *The Journal of Chemical Physics*, 120:9911, 2004.

[8] A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99:12562, 2002.

[9] C. J. Pickard and R. Needs. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23:053201, 2011.

[10] J. H. Holland. *Adaptation in natural and artificial systems: an introductory 850 analysis with applications to biology, control, and artificial intelligence.* University of Michigan Press, 1975.

[11] R. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, page 39. IEEE, 1995.

[12] A. R. Oganov and M. Valle. How to quantify energy landscapes of solids. *The Journal of Chemical Physics*, 130:104504, 2009.

[13] R. Černỳ and V. Favre-Nicolin. Direct space methods of structure determination from powder diffraction: principles, guidelines and perspectives. *Zeitschrift für Kristallographie-Crystalline Materials*, 222:105, 2007.

[14] P. Gao, Q. Tong, J. Lv, Y. Wang, and Y. Ma. X-ray diffraction data-assisted structure searches. *Computer Physics Communications*, 213:40–45, 2017.

[15] H. Putz, J. C. Schön, and M. Jansen. Combined method for ab initio structure

solution from powder diffraction data. *Journal of Applied Crystallography*, 32:864, 1999.

[16] O. J. Lanning, S. Habershon, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, E. Tedesco, and G. W. Turner. Definition of aguiding function'in global optimization: a hybrid approach combining energy and R-factor in structure solution from powder diffraction data. *Chemical Physics Letters*, 317:296, 2000.

[17] G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston, and B. M. Kariuki. A method for understanding characteristics of multi-dimensional hypersurfaces, illustrated by energy and powder profile R-factor hypersurfaces for molecular crystals. *Zeitschrift für Kristallographie-Crystalline Materials*, 216:187, 2001.

[18] S. M. Santos, J. Rocha, and L. Mafra. Nmr crystallography: toward chemical shift-driven crystal structure determination of the β-lactam antibiotic amoxicillin trihydrate. *Crystal Growth & Design*, 13:2390, 2013.

[19] B. Meredig and C. Wolverton. A hybrid computational–experimental approach for automated crystal structure solution. *Nature Materials*, 12:123, 2013.

[20] L. Ward, K. Michel, and C. Wolverton. Automated crystal structure solution from powder diffraction data: Validation of the first-principles-assisted structure solution method. *Physical Review Materials*, 1:063802, 2017.

[21] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140:A1133, 1965.

[22] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136:B864, 1964.

[23] J. P. Perdew. Accurate density functional for the energy: Real-space cutoff of the gradient expansion for the exchange hole. *Physical Review Letters*, 55:1665, 1988.

[24] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 354:11169, 1996.

[25] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59:1758, 1999.

[26] M. Shiga, M. Tachikawa, and S. Miura. Ab initio molecular orbital calculation considering the quantum mechanical effect of nuclei by path integral molecular dynamics. *Chemical Physics Letters*, 332:396, 2000.

[27] M. Shiga, M. Tachikawa, and S. Miura. A unified scheme for ab initio molecular orbital theory and path integral molecular dynamics. *The Journal of Chemical Physics*, 115:9149, 2001.

[28] S. Tsuneyuki, M. Tsukada, H. Aoki, and Y. Matsui. First-principles interatomic potential of silica applied to molecular dynamics. *Physical Review Letters*, 61:869, 1988.

[29] J. Tersoff. New empirical approach for the structure and energy of covalent systems. *Physical Review B*, 37:6991, 1988.

[30] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard. Reaxff:a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*, 105:9396, 2001.

[31] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren. Neural network models of potential energy surfaces. *The Journal of Chemical Physics*, 103:4129, 1995.

[32] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal*

of Computational Physics, 117:1, 1995.

[33] Lammps molecular dynamics simulator. `https://lammps.sandia.gov/`.

[34] P. J. van Laarhoven and E. H. Aarts. *Simulated annealing: Theory and applications.* Springer, 1987.

[35] L. V. Woodcock. Isothermal molecular dynamics calculations for liquid salts. *Chemical Physics Letters*, 10:257, 1971.

[36] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81:511, 1984.

[37] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31:1695, 1985.

[38] K. L. Geisinger, M. A. Spackman, and G. V. Gibbs. Exploration of structure, electron density distribution, and bonding in coesite with Fourier and pseudoatom refinement methods using single crystal X-ray diffraction data. *The Journal of Chemical Physics*, 91:3237, 1987.

[39] R. Stahl, C. Jung, H. D. Lutz, W. Kockelmann, and H. Jacobs. Kristallstrukturen und wasserstoffbrueckenbindungen bei β-Be(OH)$_2$ und ϵ-Zn(OH)$_2$. *Zeitschrift fuer Anorganische und Allgemeine Chemie*, 624:1130, 1998.

[40] M. I. Mendelev, M. Asta, M. J. Rahman, and J. J. Hoyt. Development of interatomic potentials appropriate for simulation of solid-liquid interface properties in Al-Mg alloys. *Philosophical Magazine*, 89:3269, 2009.

[41] P. J. Brown, A. G. Fox, E. N. Maslen, M. A. O' keefe, and B. T. M. Willis. *International tables for crystallography*, volume C, chapter 6.1, page 554. Wiley, 2006.

[42] F. S. Varley. Neutron scattering lengths and cross sections. *Neutron News*, 3:29, 1992.

Appendix A

Calculation of the powder diffraction pattern

This appendix describes the calculation of the powder diffraction pattern used in the penalty function.

A.1 Formulation of the powder diffraction pattern

We calculate the powder diffraction pattern according to the following formula,

$$J(\theta, R) = |S(\theta, R)|^2 L(\theta) P(\theta), \tag{A.1}$$

where θ is the diffraction angle, R is the crystal structure, S is the structure factor, L is the Lorentz factor, and P is the polarization factor. In order to bring the theoretical value closer to the experimental one, it is necessary to multiply this formula by additional factors, such as the absorption factor and the temperature (Debye-Waller) factor. Since these factors are complicated to calculate during the simulation, we omit them in the calculation of the penalty function. On the other hand, the Lorentz factor and the polarization factor are expressed by the simple following formulas that depend only on the diffraction angle,

$$L(\theta) = \frac{1}{\sin^2\theta\cos\theta}, \tag{A.2}$$

$$P(\theta) = \frac{1 + \cos^2(2\theta)}{2}. \tag{A.3}$$

The structure factor is the main part that depends on the atomic coordinates and is expressed by the following formula,

$$S(\theta, R) = \sum_{j=1}^{N} F_j(\theta) \exp(2\pi i \boldsymbol{h} \cdot \tilde{\boldsymbol{r}}_j), \tag{A.4}$$

where subscript j is the index of atoms, N is the number of atoms, $\tilde{\boldsymbol{r}}$ is the atomic fractional coordinate, $\boldsymbol{h}$ is the vector of the Miller index, and F is the atomic form factor. From Bragg’ s law, the diffraction angle corresponding to the Miller index $\boldsymbol{h}$ is as follows (see Fig. A.1),

$$\theta_{\boldsymbol{h}} = \arcsin\left(\frac{\lambda}{2d_{\boldsymbol{h}}}\right) = \arcsin\left(\frac{\lambda|B\boldsymbol{h}|}{4\pi}\right), \tag{A.5}$$

where λ is the wavelength, d is the inter-planar spacing, and B is the matrix consisting of the reciprocal lattice vectors ($\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$).

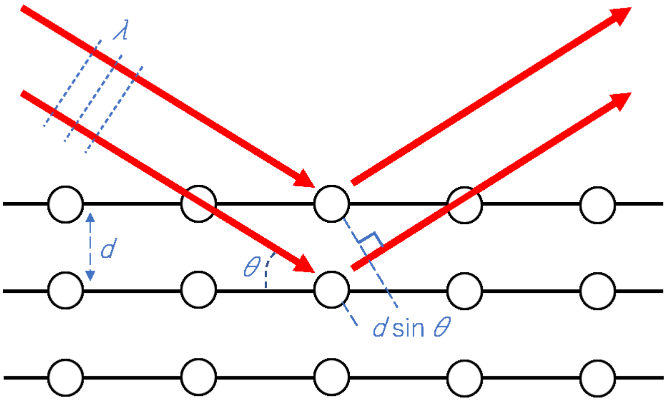

Fig. A.1. Schematic diagram of Bragg's law. The white circles are atoms arranged with the inter-planar spacing d, and red arrows are rays with the wave length λ and the diffraction angle θ. The path difference between the two rays is equal to $2d\sin\theta$.

The atomic form factor depends on atomic species, and in the case of the X-ray diffraction, it is expressed by the following formula,

$$F(\theta) = \sum_{i=1}^{4} a_i \exp\left(-b_i\left(\frac{\sin\theta}{\lambda}\right)^2\right) + c. \tag{A.6}$$

We use the database [41] for the parameters (a, b, c), which depend on atomic species. A similar equation holds in the case of the neutron diffraction, and we also use the database [42]. In the X-ray diffraction, the atomic form factor of hydrogen is much smaller than that of a heavy atom, whereas in the neutron diffraction, it is relatively large and hydrogen atoms can be detected.

A.2 Diffraction peak smearing

In the equation A.1, the diffraction peak is a delta function at a diffraction angle θ, but in the experiment, it has a finite peak width. We implement two smearing functions, Gaussian smearing and Lorentzian smearing. The powder diffraction pattern with the finite peak width is expressed by the following formula,

$$I(\theta) = \sum_{\boldsymbol{h}} J(\theta_{\boldsymbol{h}}) f(|\theta - \theta_{\boldsymbol{h}}|), \tag{A.7}$$

where f is the smearing function. In our calculations, in order to reduce the computational costs, f is regarded as zero at large $|\theta - \theta_{\boldsymbol{h}}|$. Gaussian and Lorentzian smearing function are as follows,

$$f_{\text{gaussian}} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta - \theta_{\boldsymbol{h}})^2}{2\sigma^2}\right) \tag{A.8}$$

$$f_{\text{lorentzian}} = \frac{1}{\pi} \frac{\gamma}{(\theta - \theta_{\boldsymbol{h}})^2 + \gamma^2} \tag{A.9}$$

where σ is the standard deviation, and γ is the scale parameter. σ in Gaussian smearing and γ in Lorentzian smearing correspond to the peak width, respectively.

In the Rietveld analysis [3], a smearing function that combines Gaussian and Lorentzian smearings is used to reproduce the experimental result. In the structure prediction method based on data assimilation, it is not necessary to accurately refine the peak width, but in order to improve the search efficiency, it is better to set the peak width close to that in the experiment.

A.3 Range of the Miller indices

The Miller index is an arbitrary integer, but the range of it to be considered is limited by the range of the diffraction angle. Assuming that the range of the diffraction angle is [θ_{m}, θ_{M}], from Eq. A.5, the range of the Miller index $\boldsymbol{h}$ is as follows,

$$\frac{4\pi \sin\theta_{\text{m}}}{\lambda} \leq |B\boldsymbol{h}| \leq \frac{4\pi \sin\theta_{\text{M}}}{\lambda}. \tag{A.10}$$

Let r be the right side of Eq. A.10, and consider the upper limit of the Miller index. Then, Eq. A.10 can be interpreted that the grid point $\boldsymbol{h}$ is included in the curved surface S' (see Fig. A.2). S' is obtained on a linear transformation of the sphere S with radius r by the inverse matrix $B^{-1} = \frac{1}{2\pi}A^{\mathrm{T}}$. A is the matrix consisting of the lattice vectors ($\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$). It is sufficient to find the maximum value of the (x, y, z)-coordinates on the curved surface S' .

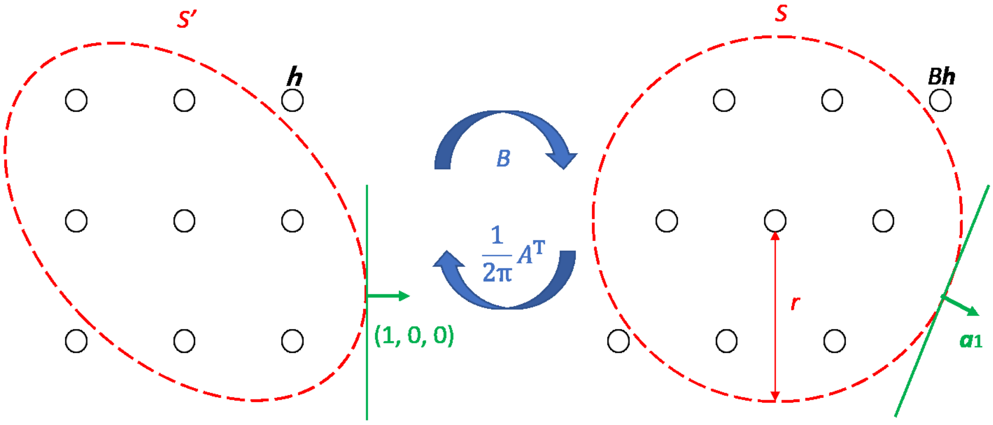

Fig. A.2. Schematic diagram for finding the upper limit of the Miller index. The left and right figures are transferred to each other on linear transformations by the reciprocal lattice vectors B and the lattice vector A, respectively. The white circles are the grid points corresponding to the Miller indices $\boldsymbol{h}$. The curved surface S is the sphere with radius r, and S' is obtained on a linear transformation of the sphere S. The maximum value of the x-coordinate on S' is the x-coordinate of the tangent point with the normal vector (1, 0, 0) on S'. This point corresponds to the tangent point with the normal vector $\boldsymbol{a}_1$ on S.

The tangent plane at the point where the x-coordinate is maximum on S' is orthogonal to x-axis and contains (0, 1, 0) and (0, 0, 1) vectors. The plane obtained on a linear transformation of this tangent plane by the matrix B is the tangent plane of S and contains the reciprocal lattice vectors $\boldsymbol{b}_2$ and $\boldsymbol{b}_3$, that is, orthogonal to the lattice vector $\boldsymbol{a}_1 \parallel \boldsymbol{b}_2 \times \boldsymbol{b}_3$. Therefore, the tangent point on S is $\frac{r}{|\boldsymbol{a}_1|}\boldsymbol{a}_1$, and the tangent point on S' is $\frac{r}{|\boldsymbol{a}_1|}B^{-1}\boldsymbol{a}_1$. Finally, the upper limit of the x-coordinate is expressed as follows,

$$\frac{r}{|\boldsymbol{a}_1|}\left(B^{-1}\boldsymbol{a}_1\right)_x = \frac{r}{2\pi|\boldsymbol{a}_1|}\left(A^{\mathrm{T}}\boldsymbol{a}_1\right)_x = \frac{r}{2\pi}|\boldsymbol{a}_1|. \tag{A.11}$$

Since the same applies to the lower limit of the x-coodinate and the limits of the y and z-coordinates, the upper and lower limits of the Miller index $\boldsymbol{h}$ are as follows.

$$\left(\pm\frac{r}{2\pi}|\boldsymbol{a}_1|, \pm\frac{r}{2\pi}|\boldsymbol{a}_2|, \pm\frac{r}{2\pi}|\boldsymbol{a}_3|\right). \tag{A.12}$$

A.4 Derivative with respect to atomic coordinates

In order to calculate the force applied to each atom from the penalty function, it is necessary to differentiate the diffraction pattern with respect to the atomic coordinates. From Appendix A.1, only the structure factor depends on the atomic coordinates. The derivative of the structural factor with respect to the atomic coordinates can be calculated as follows,

$$\begin{aligned}
\frac{\partial S}{\partial \boldsymbol{r}} &= \frac{\partial}{\partial \boldsymbol{r}} \sum_{j=1}^{N} F_j(\theta) \exp(2\pi i \boldsymbol{h} \cdot \tilde{\boldsymbol{r}}_j) && \text{(A.13)} \\
&= \sum_{j=1}^{N} F_j(\theta) \frac{\partial}{\partial \boldsymbol{r}_j} \exp(2\pi i \boldsymbol{h} \cdot A^{-1} \boldsymbol{r}_j) && \text{(A.14)} \\
&= \sum_{j=1}^{N} F_j(\theta) \exp(2\pi i \boldsymbol{h} \cdot \tilde{\boldsymbol{r}}_j) \times 2\pi i \frac{\partial}{\partial \boldsymbol{r}_j} (\boldsymbol{h} \cdot \frac{1}{2\pi} B^{\mathrm{T}} \boldsymbol{r}_j) && \text{(A.15)} \\
&= \sum_{j=1}^{N} F_j(\theta) \exp(2\pi i \boldsymbol{h} \cdot \tilde{\boldsymbol{r}}_j) \times i B \boldsymbol{h} && \text{(A.16)} \\
&= i S B \boldsymbol{h}. && \text{(A.17)}
\end{aligned}$$

The stress tensor can be calculated by differentiating with respect to the cell parameters. As Eq. A.5 shows, the diffraction angle θ depends on the cell parameters, so it is sufficient to differentiate with respect to θ. However, as seen in Eq. A.7, the smearing function also depend on θ, and its derivative is very large near the peak position. In order to optimize the cell parameters by the structure prediction method based on data assimilation, it is necessary to define a penalty function which is smooth for changes in the peak position.

Appendix B

Crystallographic data in this study

This appendix shows the crystallographic data in this study.

B.1 Coesite

Table. B.1. The correct structure of the target material coesite used in Section 3.1 [38].

Lattice parameters	Atomic coordinates	
($\AA$, degree)	(fractional)	
$a = 7.1367$	Si (0.18197, 0.14169, 0.07227)	O (0.26683, 0.14625, 0.32787)
$b = 12.3695$	Si (0.81803, 0.85831, 0.92773)	O (0.73317, 0.85375, 0.67213)
$c = 7.1190$	Si (0.81803, 0.14169, 0.42773)	O (0.73317, 0.14625, 0.17213)
$\alpha = 90.00$	Si (0.18197, 0.85831, 0.57227)	O (0.26683, 0.85375, 0.82787)
$\beta = 119.57$	Si (0.68197, 0.64169, 0.07227)	O (0.76683, 0.64625, 0.32787)
$\gamma = 90.00$	Si (0.31803, 0.35831, 0.92773)	O (0.23317, 0.35375, 0.67213)
	Si (0.31803, 0.64169, 0.42773)	O (0.23317, 0.64625, 0.17213)
	Si (0.68197, 0.35831, 0.57227)	O (0.76683, 0.35375, 0.82787)
	Si (0.28394, 0.09194, 0.54066)	O (0.28918, 0.03805, 0.02165)
	Si (0.71606, 0.90806, 0.45934)	O (0.71082, 0.96195, 0.97835)
	Si (0.71606, 0.09194, 0.95934)	O (0.71082, 0.03805, 0.47835)
	Si (0.28394, 0.90806, 0.04066)	O (0.28918, 0.96195, 0.52165)
	Si (0.78394, 0.59194, 0.54066)	O (0.78918, 0.53805, 0.02165)
	Si (0.21606, 0.40806, 0.45934)	O (0.21082, 0.46195, 0.97835)
	Si (0.21606, 0.59194, 0.95934)	O (0.21082, 0.53805, 0.47835)
	Si (0.78394, 0.40806, 0.04066)	O (0.78918, 0.46195, 0.52165)
	O (0.07598, 0.12685, 0.55943)	O (0.00000, 0.36631, 0.25000)
	O (0.92402, 0.87315, 0.44057)	O (0.00000, 0.63369, 0.75000)
	O (0.92402, 0.12685, 0.94057)	O (0.50000, 0.86631, 0.25000)
	O (0.07598, 0.87315, 0.05943)	O (0.50000, 0.13369, 0.75000)
	O (0.57598, 0.62685, 0.55943)	O (0.25000, 0.25000, 0.00000)
	O (0.42402, 0.37315, 0.44057)	O (0.75000, 0.75000, 0.00000)
	O (0.42402, 0.62685, 0.94057)	O (0.75000, 0.25000, 0.50000)
	O (0.57598, 0.37315, 0.05943)	O (0.25000, 0.75000, 0.50000)

B.2 ϵ-Zn(OH)$_2$

Table. B.2. The correct structure of the target material ϵ-Zn(OH)$_2$ used in Section 3.2 [39].

Lattice parameters	Atomic coordinates	
($\mathring{A}$, degree)	(fractional)	
$a = 4.87176$	Zn (0.067570, 0.642226, 0.123390)	H (0.757409, 0.324785, 0.841643)
$b = 5.06348$	Zn (0.432430, 0.357774, 0.623390)	H (0.742591, 0.675215, 0.341643)
$c = 8.75226$	Zn (0.932430, 0.142226, 0.376610)	O (0.120869, 0.111773, 0.577116)
$\alpha = 90.00$	Zn (0.567570, 0.857774, 0.876610)	O (0.379131, 0.888227, 0.077116)
$\beta = 90.00$	H (0.030919, 0.641940, 0.847283)	O (0.879131, 0.611773, 0.922884)
$\gamma = 90.00$	H (0.469081, 0.358060, 0.347283)	O (0.620869, 0.388227, 0.422884)
	H (0.969081, 0.141940, 0.652717)	O (0.188286, 0.316113, 0.228961)
	H (0.530919, 0.858060, 0.152717)	O (0.311714, 0.683887, 0.728961)
	H (0.242591, 0.824785, 0.658357)	O (0.811714, 0.816113, 0.271039)
	H (0.257409, 0.175215, 0.158357)	O (0.688286, 0.183887, 0.771039)

B.3 Al-Ca-H

Table. B.3. The new $Al_{12}Ca_{20}H_{76}$ structure proposed in Section 3.3.

Lattice parameters	Atomic coordinates	
($\mathring{A}$, degree)	(fractional)	
$a = 12.809756$	Al (0.500011, 0.005163, 0.537170)	H (0.393267, 0.766778, 0.863840)
$b = 12.809756$	Al (0.003018, 0.505047, 0.537047)	H (0.429963, 0.705195, 0.535184)
$c = 6.814734$	Al (0.502574, 0.005254, 0.037029)	H (0.247695, 0.602437, 0.873155)
$\alpha = 90.00$	Al (0.331902, 0.335357, 0.224586)	H (0.304244, 0.565146, 0.542758)
$\beta = 90.00$	Al (0.168770, 0.836904, 0.222126)	H (0.104170, 0.894427, 0.710286)
$\gamma = 90.00$	Al (0.839311, 0.163375, 0.215090)	H (0.199422, 0.934485, 0.043467)
	Al (0.663408, 0.663518, 0.214548)	H (0.275273, 0.889934, 0.362151)
	Al (0.831443, 0.834761, 0.727048)	H (0.251859, 0.746446, 0.129567)
	Al (0.668759, 0.336545, 0.722520)	H (0.090047, 0.932491, 0.312873)
	Al (0.338640, 0.663467, 0.715482)	H (0.110123, 0.736419, 0.369379)
	Al (0.163486, 0.163750, 0.712352)	H (0.389815, 0.232646, 0.368334)
	Al (1.000437, 0.505635, 0.036693)	H (0.302659, 0.434728, 0.048838)

Ca (0.015420, -0.004752, 0.039557)	H (0.415507, 0.426118, 0.316167)
Ca (0.484803, 0.491280, 0.040281)	H (0.435968, 0.308499, 0.052714)
Ca (0.514707, 0.494307, 0.541170)	H (0.227379, 0.389938, 0.366369)
Ca (0.982685, 0.989151, 0.537713)	H (0.247598, 0.247117, 0.130386)
Ca (0.932544, 0.727437, 0.229506)	H (0.073280, 0.208127, 0.533699)
Ca (0.432343, 0.773718, 0.220318)	H (0.254947, 0.103135, 0.870501)
Ca (0.567838, 0.226603, 0.228757)	H (0.258935, 0.253491, 0.650720)
Ca (0.067889, 0.273081, 0.217527)	H (0.202937, 0.068335, 0.538267)
Ca (0.220985, 0.572058, 0.218955)	H (0.107362, 0.264613, 0.864283)
Ca (0.279747, 0.072355, 0.214524)	H (0.062637, 0.083103, 0.779928)
Ca (0.725710, 0.928484, 0.226622)	H (0.755229, 0.603578, 0.372325)
Ca (0.776591, 0.427751, 0.223853)	H (0.572095, 0.705949, 0.035581)
Ca (0.432544, 0.226399, 0.729145)	H (0.758564, 0.753375, 0.150522)
Ca (0.933172, 0.272773, 0.720397)	H (0.702528, 0.567688, 0.040467)
Ca (0.067933, 0.725746, 0.728927)	H (0.607314, 0.765620, 0.364382)
Ca (0.567350, 0.772103, 0.721009)	H (0.562899, 0.583327, 0.284326)
Ca (0.720780, 0.072674, 0.718421)	H (0.801804, 0.933755, 0.550899)
Ca (0.779762, 0.572221, 0.715348)	H (0.889878, 0.731794, 0.869815)
Ca (0.226267, 0.429268, 0.725852)	H (0.916010, 0.924802, 0.817192)
Ca (0.275707, 0.928949, 0.723158)	H (0.934911, 0.807369, 0.553836)
H (0.962768, 0.518215, 0.793265)	H (0.727701, 0.890272, 0.868660)
H (0.041755, 0.496431, 0.293066)	H (0.746668, 0.746553, 0.634524)
H (0.948064, 0.627964, 0.523266)	H (0.699092, 0.434587, 0.544864)
H (0.055494, 0.385916, 0.584412)	H (0.610534, 0.234624, 0.868264)
H (0.120078, 0.562359, 0.594489)	H (0.774165, 0.390176, 0.863575)
H (0.880794, 0.451961, 0.513059)	H (0.752661, 0.246713, 0.630062)
H (0.885366, 0.560309, 0.108509)	H (0.587929, 0.429989, 0.813679)
H (0.122494, 0.454243, 1.004958)	H (0.564135, 0.309860, 0.552110)
H (0.057008, 0.628784, 0.042721)	H (0.883523, 0.112002, 0.676588)
H (0.947454, 0.383660, 0.055655)	H (0.929624, 0.205285, 0.033478)
H (0.461099, 0.008106, 0.293226)	H (0.893547, 0.268716, 0.361043)
H (0.541870, 0.005978, 0.793358)	H (0.944753, 0.090128, 0.288546)
H (0.384702, 0.061043, 0.604221)	H (0.742902, 0.250292, 0.147886)
H (0.622056, 0.953058, 0.508453)	H (0.750116, 0.101437, 0.374681)
H (0.556149, 0.128408, 0.534006)	H (0.605151, 0.394932, 0.211249)
H (0.446979, 0.884057, 0.568232)	H (0.383282, 0.613126, 0.176826)
H (0.380434, 0.953085, 0.007598)	H (0.618348, 0.613711, 0.671557)
H (0.619000, 0.061626, 0.101415)	H (0.398676, 0.397375, 0.721478)
H (0.555263, 0.884589, 0.072071)	H (0.899486, 0.899628, 0.226555)
H (0.446988, 0.128357, 0.031924)	H (0.119240, 0.115050, 0.167075)
H (0.243024, 0.751148, 0.647358)	H (0.804579, 0.064654, 0.043351)
H (0.442122, 0.587476, 0.788658)	H (0.063647, 0.810769, 0.052614)