A Noise-Robust Data Assimilation Method for Crystal Structure Prediction Using Powder Diffraction Intensity

Seiji Yoshikawa,1 Ryuhei Sato,1, 2 Ryosuke Akashi,1 Synge Todo,1, 3, 4, 5 and Shinji Tsuneyuki1, 4, 5, 6
1) Department of Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
2) Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan
3) Institute for Solid State Physics, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8581, Japan
4) Mathematics and Informatics Center, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
5) Institute for Physics of Intelligence, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
6) UTokyo Research Institute for Photon Science and Laser Technology, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

(*Electronic mail: ryuhei.sato.c1@tohoku.ac.jp)
(Dated: 14 September 2022)

Crystal structure prediction for a given chemical composition has long been a challenge in condensed-matter science. We have recently shown that experimental powder X-ray diffraction (XRD) data are helpful in a crystal structure search using simulated annealing, even when they are insufficient for structure determination by themselves (N. Tsujimoto et al., Phys. Rev. Materials 2, 053801 (2018)). In the method, the XRD data are assimilated into the simulation by adding a penalty function to the physical potential energy, where we used a crystallinity-type penalty function defined by the difference between experimental and simulated diffraction angles. To improve the success rate and noise robustness, we introduce a correlation-coefficient-type penalty function adaptable to XRD data with significant experimental noise. We apply the new penalty function to SiO$_2$ coesite and ε-Zn(OH)$_2$ to determine its effectiveness in the data assimilation method.

I. INTRODUCTION

The prediction of crystal structures at the atomic level is one of the most fundamental challenges in condensed matter science. From the viewpoint of theoretical calculations, one problem is finding the global minimum of the potential-energy hypersurface. Thus, a number of methods have been developed to generate new candidate structures (ex. random sampling, genetic algorithm, and particle swarm optimization (PSO)) and to overcome the potential barrier on the way to the target structure (ex. simulated annealing (SA), basin hopping, minima hopping, and metadynamics). Although these methods successfully yield the new structure represented by superconducting hydrides and other materials, they can only handle systems with up to a few hundred atoms because the number of candidate structures increases exponentially with system size. Therefore, it is necessary to accelerate these structure search methods when applied to the large systems.

One approach to accelerating the theoretical structure search is the utilization of the experimental data with conventional calculations. A typical approach is to calculate physical properties and directly compare them with experimental data to choose the best structure among the candidates. Another is to optimize the crystal structure in accordance with the cost function, which is the combination of the interatomic potential and the kind of “score” (penalty function) obtained from the similarity between the calculated and experimental physical properties during these simulations. In these data-assimilation techniques, we can predict structures from the experimental data, even if we do not know the crystal symmetry or do not have any candidate structures as initial configurations. Also, we can easily exclude the other stable structures, whose physical property does not agree with the experiment, since the structure is optimized in accordance with the cost function (i.e., interatomic potential + penalty function) taking into account the experimental data.

Putz et al. confirmed that crystal structures such as TiO$_2$ (rutile) and SiO$_2$ (quartz) can be obtained from random structures by the Monte-Carlo method based on the linear combination of the empirical two-body potential and penalty function calculated from the similarity between experimental and calculated X-ray diffraction (XRD) patterns. Tsujimoto et al. showed that short-time (5000 step (=5 ps)) simulated annealing based on XRD-data-assimilated molecular dynamic (MD) simulation can even yield a low-symmetry structure such as SiO$_2$ coesite, although the success rate of finding the target structure is low (up to 30% with an ideal experimental XRD pattern without noise). One of the problems in the previous study was the robustness against noise, since we need to manually determine whether each peak in the experimental data is the signal or noise by introducing a cutoff value as a pretreatment of experimental data.

Under such circumstances, we introduce a new correlation-coefficient-type penalty function to the experimental (XRD) data-assimilated approach to improve the success rate and robustness to the experimental noise. We discuss whether this approach is applicable for hydrogen-containing material, in
which hydrogen atoms scarcely contribute to the XRD pattern, to determine the structure.

II. METHOD

A. Experimental data assimilation

The central idea of the data assimilation method is based on Bayes’ theorem. This method can be interpreted as a kind of maximum likelihood estimation. From Bayes’ theorem, the conditional probability of the crystal structure given the experimental data is

\[ \rho(R|I_{\text{ref}}) = \frac{\rho(R) \times \rho(I_{\text{ref}}|R)}{\rho(I_{\text{ref}})} \]  

(1)

where \( R \) represents the atomic positions and lattice parameters and \( I_{\text{ref}} \) is the reference experimental data. In this work, we adopt the diffraction pattern as \( I_{\text{ref}} \), though it can be arbitrary experimentally observable quantities. In the case of optimization by simulated annealing, the probability of finding the crystal structure follows the Boltzmann distribution:

\[ \rho(R) \propto \exp\{-\beta E(R)\} \]  

(2)

where \( E(R) \) is the interatomic potential energy and \( \beta \) is the inverse temperature. Meanwhile, by presuming that \( R \) giving calculated data \( I_{\text{calc}} \) similar to \( I_{\text{ref}} \) should resemble the target structure, \( \rho(I_{\text{ref}}|R) \) is defined in the following manner:

\[ \rho(I_{\text{ref}}|R) \propto \exp\{-\beta \alpha D[I_{\text{ref}}, I_{\text{calc}}(R)]\} \]  

(3)

where \( N \) is the number of atoms and \( \alpha \) is a weight parameter. \( D \) is the penalty function obtained from the similarity between calculated and experimental data: \( D \) takes zero when \( I_{\text{ref}} \) and \( I_{\text{calc}} \) match perfectly, whereas it takes positive value otherwise. As a result, from eqs. (1)-(3), \( \rho(R|I_{\text{ref}}) \) is proportional to the exponential of the cost function \( F \) (eq.(5)) in the following manner:

\[ \rho(R|I_{\text{ref}}) \propto \exp\{-\beta F(R; I_{\text{ref}})\} \]  

(4)

\[ F(R; I_{\text{ref}}) = E(R) + \alpha ND[I_{\text{ref}}, I_{\text{calc}}(R)] \]  

(5)

Therefore, by minimizing \( F \), we can maximize \( \rho(R|I_{\text{ref}}) \) to find the target structure referring to the given experimental data and efficiently obtain the target structure with this data-assimilated approach. The concept of our method is summarized in the schematic diagram shown in Fig. 1. The target structure is the common minimum for both the interatomic potential energy and penalty function. If the target structure is metastable or the experimental data is incomplete, the correct structure may not be the common global minimum of the potential energy and penalty function, yet it should be the best compromise as the global minimum of the cost function, \( F \). In the case of XRD and Neutron diffraction patterns, the penalty function is mainly dependent on the symmetry and long-range order of the structure, while the interatomic potential mainly reflects the local interaction between atoms. Therefore, the positions of their local minima are different from each other and these local minima are expected to have significantly higher cost or become unstable by these two functions. As a result, the correct structure is more emphasized in the cost function than in the interatomic potential energy. Therefore, by optimizing the cost function, we can find the correct structure more efficiently than optimizing the structure with only the interatomic potential energy.

B. Penalty function using the powder diffraction pattern

As the penalty function in the previous study\[^{23}\] we used the crystallinity-type penalty function defined as

\[ D_{\text{cryst}}(R) = 1 - \frac{\int_{\theta_{\text{ref}}}^{\theta_{\text{max}}} I_{\text{calc}} d\theta}{\int_{\theta_{\text{min}}}^{\theta_{\text{max}}} I_{\text{calc}} d\theta} \]  

(6)

where \( I_{\text{calc}} \) is the calculated diffraction intensity, \( \theta \) is the diffraction angle, \([\theta_{\text{min}}, \theta_{\text{max}}]\) is the reference angle range, and \( \theta_{\text{ref}} \) is the peak position observed in the reference experimental diffraction pattern \( I_{\text{ref}} \). This penalty function does not depend on the experimental intensity information and represents the degree of coincidence 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. Because of experimental errors such as the effect of preferred orientation, the peak intensity rate in experiments sometimes deviates from the ideal one. In such a case, this penalty function works...
well even with an unreliable experimental intensity rate. However, 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 owing to noise.

In this study, we propose to use the correlation coefficient between \( I_{\text{calc}} \) and \( I_{\text{ref}} \) as follows:

\[
D_{\text{CC}}(R) = 1 - \frac{\int_{\theta_{\text{min}}}^{\theta_{\text{max}}} I_{\text{calc}}(R) d\theta}{\sqrt{\int_{\theta_{\text{min}}}^{\theta_{\text{max}}} I_{\text{calc}}^2(R) d\theta \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} I_{\text{ref}}^2 d\theta}},
\]

where \( I = I - \bar{I} \) and \( \bar{I} \) is the average intensity over the reference angle range \([\theta_{\text{min}}, \theta_{\text{max}}]\). This penalty function represents how well the shape of the calculated diffraction pattern agrees with that of the experiment. Therefore, it mainly fits the large peaks and their intensity rates in the experiment and includes not only the peak position but also the intensity information.

The other feature of this penalty function is that it rarely requires pretreatment of experimental data. Even when the experimental noise is so large that we cannot distinguish some small diffraction peaks from noise, background subtraction is sufficient to capture the feature of the experimental XRD pattern using the correlation-coefficient-type penalty function.

### C. XRD-assimilated Molecular Dynamic Simulation

In this study, we perform data assimilated MD simulations to predict \( \text{SiO}_2 \) (coesite) and \( \varepsilon\text{-Zn(OH)}_2 \) in the following manner: First, we set the appropriate simulation cell size to reproduce the experimental data in the target system. Although we used the known lattice parameters \( a = 7.14, b = 12.37, c = 7.12 \, \text{Å}, \beta = 119.57^\circ \) for coesite: \( a = 4.87, b = 5.06, c = 8.75 \, \text{Å} \) for \( \varepsilon\text{-Zn(OH)}_2 \) in this report, in practice, we can use excellent programs for Miller indexing and subsequent lattice parameter determination, such as ConoGraph. We employed a \( 2 \times 1 \times 1 \) cell containing 96 (coesite) and 40 atoms (\( \varepsilon\text{-Zn(OH)}_2 \)) for the following data-assimilated MD simulation, since the primitive cell is considered to be too simple to test the applicability and effectiveness of the data-assimilated MD simulation. After choosing the lattice parameter, we generate random structures as initial atomic coordinates for each simulation. A structure with too small an interatomic distance is inappropriate as the initial arrangement, because the interatomic potential energy calculation sometimes diverges. Hence, to avoid the divergence of potential energy calculation during these simulations, atomic positions are chosen randomly, but the minimum interatomic distance is larger than 1.7 Å for coesite and 1.5 Å for \( \varepsilon\text{-Zn(OH)}_2 \) when we generate the initial atomic coordinates. Then, we minimize the cost function by SA using MD simulation. Note that coesite is chosen for its low symmetry and \( \varepsilon\text{-Zn(OH)}_2 \) as a test sample of a hydrogen-containing structure. Also note that other \( \text{SiO}_2 \) structures such as alpha quartz and cristobalite can be determined by the same method, as shown in the previous study.

For MD simulations, LAMMPS packages were employed. During a 10,000-step (10ps) NVT simulation, the temperature is decreased from 10,000 (5,000 for \( \varepsilon\text{-Zn(OH)}_2 \)) to 0 K linearly by the velocity scaling method. For calculating the total energy and forces, we adopted the Tsuneyuki potential for \( \text{SiO}_2 \), whereas we executed the DFT calculation (VASP) for \( \varepsilon\text{-Zn(OH)}_2 \). The total energy is examined to determine whether the correct structure has been obtained. In DFT calculation, the Perdew–Burke–Ernzerhof generalized gradient approximation (PBE-GGA) and projector augmented wave (PAW) method were employed for the exchange-correlation function and pseudopotential, respectively. The plane-wave energy cutoff was set to 400 eV and only \( \Gamma \) point was sampled for the Brillouin zone integration. The reference diffraction data in Figs. 2(b) and (d) are calculated from the experimentally known structure in Figs. 2(a) and (c). Here, as in the previous work, we intentionally limit the range of diffraction angles to 20–45 degrees for coesite and 10–45 degrees for \( \varepsilon\text{-Zn(OH)}_2 \). One reason is that the peak intensity decreases with increasing diffraction angle due to the XRD atomic scattering factor and the Debye-Waller factor, making it difficult to separate the peak intensity from the noise. Another reason is that we cannot measure large-
angle diffraction in ultrahigh-pressure experiments where the data-assimilation method is highly required. Gaussian white noise is also introduced to the reference diffraction data (Fig. 2(b)) to examine the noise robustness of the data assimilation method.

III. RESULTS AND DISCUSSION

A. Correlation-coefficient-type penalty function

The success rate of finding the target structure is significantly improved by using the new correlation-coefficient-type penalty function. Figure 3(a) shows the success rate of finding the coesite structure using XRD-assimilated MD simulations as a function of the weight parameter $\alpha$. Note that the obtained structure is regarded as that for coesite when its interatomic potential is about $-1716.78$ eV. We performed 50 simulations for each value of $\alpha$ and with two types of penalty function. The weight parameter $\alpha$ represents how strongly the search space is restricted by the penalty function so that the calculated XRD pattern matches that of the experiment. If $\alpha$ is set to zero, it means that MD simulations have been performed using only the interatomic potential energy without the experimental data. As shown in the figure, with appropriate $\alpha$, the success rate of finding the coesite structure almost exceeds 80% using the correlation coefficient ($D_{CC}$). On the other hand, the success rate obtained using the crystallinity ($D_{Crys}$) is at most 50%. It is also found that the success rate for the simulations without the penalty function ($\alpha=0$) is about 0%, showing that the efficiency of crystal structure search is enhanced by the penalty function.

The above conclusion is further confirmed by the interatomic potential energy of the optimization trajectories as a function of $D_{CC}$ (Fig.3(b)). This figure shows the values of interatomic potential energy and $D_{CC}$ from 0 to 10,000 steps for all 50 data-assimilated MD simulations. According to the definition of the correlation-coefficient-type penalty function, the calculated XRD pattern is identical to the reference pattern when $D_{CC}=0$, whereas the calculated XRD pattern is totally different from the reference when $D_{CC}=1$. We evaluated $D_{CC}$ for the three optimization cases, “No XRD”, “Crystallinity” and “Correlation coefficient”, in order to compare the distribution of the structural similarity. Here we note that, for the former two cases, $D_{CC}$ was not referred to for the optimization. Without the XRD penalty function, most of the structures obtained during MD simulations have small energies as low as $-1700$ eV, but $D_{CC}$ is widely distributed from 0.06 to 0.8, meaning that this material has many metastable structures and tends to become amorphous. On the other hand, by introducing the correlation-coefficient-type penalty function, most of these configurations become unstable and are rarely obtained during XRD-assimilated MD simulation. Therefore, the structures with an energy of about $-1700$ eV during the XRD-assimilated MD simulation are distributed only around $D_{CC}=0.0$. Compared with the correlation coefficient, the crystallinity weakly restricts the search space, as shown in the distribution around $-1700$ eV, which has wide ranging $D_{CC}$ from 0.0 to 0.6. Thus, we conclude that the correlation-coefficient-type penalty significantly accelerates the structure search compared with the previous crystallinity-type penalty.

B. Diffraction pattern with noise

The use of correlation coefficient improves not only the success rate but also the noise robustness. To verify the noise robustness of the data-assimilated structure prediction method, we add Gaussian white noise to the reference XRD pattern, as shown by the red line in Fig. 2(b). In this figure, the standard deviation of the noise, $\sigma$, is set to 0.08 times the height of the maximum diffraction peak. The intensity with noise at each $\theta$ is calculated using the reference intensity ($\sigma=0$) in the

![FIG. 3. (a) Success rates of finding coesite structure using correlation-coefficient-type and crystallinity-type penalty functions. The horizontal axis $\alpha$ is the weight parameter in eq. (5). (b) Interatomic potential energies and penalty function, $D_{CC}$, values for the trajectories of 50 data-assimilated MD simulations targeting SiO$_2$ coesite. Data obtained by the data-assimilated MD simulations with crystallinity-type (blue) and correlation-function-type (red) penalty functions and without (black) penalty function.](image-url)
A Noise-Robust Data Assimilation Method for Crystal Structure Prediction Using Powder Diffraction Intensity

FIG. 4. Success rate of finding coesite structure using XRD-assimilated MD simulations with noisy reference XRD data as a function of the weight parameter $\alpha$. $\sigma$ represents the standard deviation for the introduced Gaussian white noise.

Following manner:

$$I(\theta) = I(\theta)_{\sigma=0} + \xi I_{\text{MAX}, \sigma=0}.$$  

(8)

Here, $I_{\text{MAX}, \sigma=0}$ and $\xi$ represent the intensity of the largest diffraction peak and a random number, respectively. The random number is chosen so that its probability distribution equals to Gaussian distribution function with $\sigma$:

$$P(\xi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\xi^2}{2\sigma^2}\right).$$  

(9)

As shown in the figure, because of the large noise, it is almost impossible to distinguish the signal from the noise except for the largest and second largest peaks at 25.9 and 28.9°, respectively. Figure 4 shows the success rate of finding the coesite structure using XRD-assimilated MD simulations with the reference XRD pattern including Gaussian white noise. We perform 50 simulations for each value of $\alpha$ and amplitude of noise $\sigma$. According to eq.(8) and (9), the amplitude of noise is roughly proportional to the standard deviation ($\sigma$) for the Gaussian white noise. Compared with the case for the ideal reference XRD pattern without noise ($\sigma=0$), the success rate decreases as the amplitude of noise increases. However, even with the largest $\sigma$ in these test simulations ($\sigma=0.12$), the success rate is about 33% with appropriate $\alpha$, showing the noise robustness of the correlation-coefficient-type penalty function.

C. System containing hydrogen

The applicability of data-assimilated MD simulations to hydrogen containing materials is discussed in this section. Figure 5 shows the interatomic potential energy distribution obtained by XRD-assimilated MD simulations as a function of $D_{CC}$. Note that, for $\varepsilon$-Zn(OH)$_2$, various structures hosting different hydrogen bonding or H$_2$O molecules also have comparably small $D_{CC}$. To highlight the subtle difference in $D_{CC}$, we plotted $\log D_{CC}$ in the horizontal axis. As shown in the figure, without the XRD penalty function, there is a local minimum at around log($D_{CC}$) = -0.9-0.0, suggesting that conventional SA using DFT-MD simulation rarely leads to the $\varepsilon$-Zn(OH)$_2$ structure. On the other hand, XRD-assimilated DFT-MD simulation can yield the $\varepsilon$-Zn(OH)$_2$ structure, as shown by the structures obtained at around log $D_{CC} = -5$ (purple dots in the figure). However, there are still local minima at around log $D_{CC} = -2.35$ and -4, which reduce the success rate. These local minima are probably related to H$_2$O molecule formation, since almost all of these structures at log $D_{CC} = -2.35$ and -4 (purple dots in Fig. 5) include H$_2$O molecules. Note that 97% of the structures obtained after DFT-MD simulation without the XRD penalty function include H$_2$O molecules, suggesting that the XRD penalty function prevents H$_2$O molecule formation to some extent, although 70% of the structures obtained by XRD-assimilated DFT-MD simulations still include H$_2$O molecules. Therefore, introducing some constraint to reduce H$_2$O molecule formation during data-assimilated MD simulations could further increase the success rate. Nevertheless, the success rate of finding the $\varepsilon$-Zn(OH)$_2$ structure is 30% (the rate of finding structures whose interatomic potential energy is about -190.3 eV) with the XRD penalty function with weight parameter $\alpha = 15$, showing the effectiveness of XRD-assimilated simulation in case of the hydrogen-containing system.
IV. CONCLUSION

In this study, we introduce a new correlation-coefficient-type penalty function for XRD data-assimilated simulations to improve the success rate of finding the target structure and noise robustness. Using the correlation-coefficient-type penalty function, the success rate of finding the coesite structure with noise-free reference data almost exceeds 80%, which is almost twice as much as that obtained with the crystallinity-type penalty function. In addition, the use of this penalty function enable us to predict the coesite structure even with noisy reference data in which the minor peaks are almost indistinguishable from the noise. It is also found that this XRD data-assimilated MD simulation can be applied to the hydrogen-containing materials even though the hydrogen atoms negligibly contribute to the XRD data. The success rate of finding e-Zn(OH)$_2$ structure is about 30% using the reference XRD pattern and DFT potential, suggesting that this method is useful for the structure determination of high-pressure metal hydrides, where the angle range of XRD patterns is too small to determine the structure through an experimental approach.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research on Innovative Areas "Hydrogenomics", No. JP18H05519, and Elements Strategy Initiative to Form Core Research Center in Japan. S.Y. was supported by the Japan Society for the Promotion of Science through the Program for Leading Graduate Schools (MERIT).

AUTHOR DECLARATIONS

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCE

1 S. M. Woodley and R. Catlow, “Crystal structure prediction from first principles,” Nat. Mater. 7, 937 (2008).
2 J. Maddox, “Crystals from 1st principles,” Nature 335, 201 (1988).
3 C. J. Pickard and R. J. Needs, “Ab initio / random structure searching,” J. Phys. Condens. Matter 23, 053201 (2011).
4 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).
5 A. R. Oganov and C. W. Glass, “Crystal structure prediction using ab initio evolutionary techniques: Principles and applications,” J. Chem. Phys. 124, 244704 (2006).
6 A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, “New developments in evolutionary structure prediction algorithm USPEX,” Comput. Phys. Comm. 184, 1172 (2013).
7 A. R. Oganov, A. O. Lyakhov, and M. Valle, “How evolutionary crystal structure prediction works-and why,” Acc. Chem. Res. 44, 227 (2011).
8 A. Kruglov, D. Semenok, V. H. Song, R. Szczesniak, I. A. Wrona, R. Ai, D. M. D. Esfahani, D. Duan, T. Cui, A. G. Kvasnishin, and A. R. Oganov, “Superconductivity of LaH$_{10}$ and LaH$_{16}$ polyhydrides,” Phys. Rev. B 101, 024508 (2020).
9 D. V. Semenok, I. A. Troyan, A. V. Sadakov, D. Zhou, M. Galasso, A. G. Kvasnishin, I. A. Kruglov, A. A. Bykov, K. Y. Terent’ev, A. V. Cherepahin, O. A. Sobolevskiy, K. S. Pervakov, A. Y. Seregin, T. Helm, T. Forster, A. D. Grockowiak, S. W. Tozer, Y. Nakamoto, K. Shimizu, V. M. Pudalov, I. S. Lyubutin, and A. R. Oganov, “Effect of magnetic impurities on superconductivity in LaH$_{10}$,” Adv. Mater. 33, 2204038 (2021).
10 K. R. Eberhart, ed., A new optimizer using particle swarm theory, in MHS’95. (Proceedings of the Sixth International Symposium on Micro Machine and Human Science, IEEE, 1995).
11 Y. Wang, J. Lv, L. Zhu, and Y. Ma, “Crystal structure prediction via particle-swarm optimization,” Phys. Rev. B 82, 094116 (2010).
12 Y. Wang, J. Lv, L. Zhu, and Y. Ma, “CALYPSO: A method for crystal structure prediction,” Comput. Phys. Comm. 183, 2063 (2012).
13 P. Gao, Q. Tong, J. Lv, Y. Wang, and Y. Ma, “X-ray diffraction data-assisted structure searches,” Comput. Phys. Comm. 213, 40 (2016).
14 S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” Science 220, 671 (1983).
15 D. Wales and J. Doye, “Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms,” J. Phys. Chem. A 101, 5111 (1997).
16 S. Goedecker, “Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems,” J. Chem. Phys. 120, 9911 (2004).
17 T. Wang, M. Hirayama, T. Nonoto, T. Koretsune, R. Arita, and J. A. Flores-Livas, “Absence of conventional room-temperature superconductivity at high pressure in carbon-doped H$_2$S,” Phys. Rev. B 104 (2021).
18 A. Laio and M. Parrinello, “Escaping free-energy minima,” Proc. Nat. Acad. Sci. 99, 12562 (2002).
19 Y. Li, J. Hao, H. Liu, Y. Li, and Y. Ma, “The metallization and superconductivity of dense hydrogen sulfide,” J. Chem. Phys. 140, 174712 (2014).
20 D. Duan, Y. Liu, F. Tian, D. Li, X. Huang, Z. Zhao, H. Yu, B. Liu, W. Tian, and T. Cui, “Pressure-induced metallization of dense (H$_2$S)$_2$H$_2$ with high-T$_c$ superconductivity,” Sci. Rep. 4, 6968 (2014).
21 Y.-L. Hai, N. Lu, H.-L. Tian, M.-J. Jiang, W. Yang, W.-J. Li, W.-Y. Yan, C. Zhang, X.-J. Chen, and G.-H. Zhong, “Cage structure and near room-temperature superconductivity in ThB$_{12}$ (n=1-12),” J. Phys. Chem. C 125, 3640 (2021).
22 W. Zhang, A. R. Oganov, A. F. Goncharov, Q. Zhu, S. E. Boulfelfel, A. O. Lyakhov, E. Stavrou, M. Somayazulu, V. B. Prakapenka, and Z. Konopkova, “Unexpected stable stoichiometries of sodium chlorides,” Science 342, 1502 (2013).
23 M.-s. Miao, “Caesium in high oxidation states and as a p-block element,” Nature Chem. 5, 846 (2013).
24 Q. Li, D. Zhou, W. Zheng, Y. Ma, and C. Chen, “Global structural optimization of tungsten borides,” Phys. Rev. Lett. 110, 136403 (2013).
25 N. Tsujimoto, D. Adachi, R. Akashi, S. Todo, and S. Tsuneyuki, “Crystal structure prediction supported by incomplete experimental data,” Phys. Rev. Mater. 2, 053801 (2018).
26 B. Meredith and C. Wolverton, “A hybrid computational-experimental approach for automated crystal structure solution,” Nature Mater. 12, 123 (2013).
27 S. M. Santos, J. Rocha, and L. Maia, “NMR crystallography: Toward chemical shift-driven crystal structure determination of the beta-lactam antibiotic amoxicillin trihydrate,” Cryst. Growth Design 13, 2390 (2013).
28 H. Putz, J. C. Schön, and M. Jansen, “Combined method for ab initio structure solution from powder diffraction data,” J. Appl. Cryst. 32, 864 (1999).
29 D. Adachi, N. Tsujimoto, R. Akashi, S. Todo, and S. Tsuneyuki, “Search for common minima in joint optimization of multiple cost functions,” Comput. Phys. Comm. 241, 92 (2019).
A Noise-Robust Data Assimilation Method for Crystal Structure Prediction Using Powder Diffraction Intensity

30A. Esmaeili, T. Kamiyama, and R. Oishi-Tomiyasu, “New functions and graphical user interface attached to powder indexing software CONO-GRAPH,” J. Appl. Cryst. 50, 651 (2017).

31S. Plimpton, “Fast parallel algorithms for short-range molecular dynamics,” J. Comput. Phys. 117, 1 (1995).

32S. Tsuneyuki, M. Tsukada, H. Aoki, and Y. Matsui, “First-principles interatomic potential of silica applied to molecular dynamics,” Phys. Rev. Lett. 61, 869 (1988).

33G. Kresse and J. Hafner, “ab initio molecular dynamics for liquid metals,” Phys. Rev. B 47, 558 (1993).

34G. Kresse and J. Furthmüller, “Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set,” Phys. Rev. B 54, 11169 (1996).

35J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized gradient approximation made simple,” Phys. Rev. Lett. 77, 3865 (1996).

36P. E. Blöchl, “Projector augmented-wave method,” Phys. Rev. B 50, 17953 (1994).

37G. Kresse and D. Joubert, “From ultrasoft pseudopotentials to the projector augmented-wave method,” Phys. Rev. B 59, 1758 (1999).

38The potential energy for the obtained ε-Zn(OH)$_2$ structures (≈–190.3 eV) is slightly larger than that of reference ε-Zn(OH)$_2$ (–190.4 eV). This suggests that the hydrogen bond orientation in the obtained structure does not perfectly agrees with that of reference. However, this energy difference is considered to be within the range of the error, because this energy difference (2.5 meV/atom = 0.1 eV for 40 atoms) is much smaller than thermal energy at 300K (≈25 meV/atom = 1 eV for 40 atoms). In addition, in terms of XRD pattern, we can hardly distinguish the obtained structure from reference one, since $D_{CC}$ is about $10^{-5}$. Therefore, here, we regard the obtained structure with total energy at around −190.3 eV as ε-Zn(OH)$_2$. 