PHYSICAL REVIEW B **94**, 184109 (2016)

Molecular dynamics study on β-phase vanadium monohydride with machine learning potential

Kazutoshi Miwa* and Hiroshi Ohno
Toyota Central Research & Development Laboratories, Inc., Nagakute 480-1192, Japan

(Received 30 August 2016; revised manuscript received 31 October 2016; published 23 November 2016)

The formalism to construct the machine learning potentials (MLPs) is presented. We introduce the spilling factor for the simultaneous error estimation and the recursive bisection method for the reduction of the computational cost. The formalism is applied for the β-phase vanadium monohydride. The first-principles calculations based on density functional theory (DFT) are used to prepare the sample data set from which the MLP for the vanadium monohydride (VH) system is constructed. In the molecular dynamics simulation with the MLP, the time-averaged structure of β-VH is predicted correctly to be the body-centered tetragonal structure with the octahedral (O) site occupation of H. The average lattice constants are in good agreement with the experimental data which are not able to be reproduced by the static DFT calculation. The O-site occupation of H observed in the average structure is, however, a saddle point on the potential-energy surface, and the actual hydrogen occupation is found to be the $4T$ configuration.

DOI: 10.1103/PhysRevB.94.184109

I. INTRODUCTION

Molecular dynamics (MD) simulations have successfully been applied in a wide variety of research fields [1], which provide the microscopic dynamical behavior of a system by solving the time evolution of motions for constituent atoms. In addition to the dynamical properties, MD can also be utilized as a sampling scheme for statistics to determine equilibrium properties. The crucial issue for the MD simulations is how to represent the interatomic potentials. First-principles calculations based on density functional theory (DFT) [2,3] allow us to describe the atomic interactions with an accurate quantum-mechanical treatment, which is known as an *ab initio* MD simulation. On the other hand, because of high computational demands of DFT calculations, the empirical interatomic potentials [4–9] have also been used to perform large-scale simulations. The accuracy of these methods strongly depends on the potential functions and their parameters which have to be determined carefully. Large effort and time are often required to construct the potentials.

In this context, machine learning approaches have recently been adopted to construct the interatomic potentials [10–13]. In the machine learning potentials (MLPs), the total energy of a system is expressed by simple and flexible functions which are chosen from a mathematical viewpoint rather than physical motivation. Different types of bonding can be treated on the same footing. Well-established training procedures are also available. In this paper, we construct the MLP for the vanadium hydrogen (VH) system and apply it for the β-phase vanadium monohydride to investigate its structural property.

It is known that the β-phase vanadium monohydride β-VH$_x$ has a body-centered tetragonal (bct) structure where H atoms occupy the octahedral (O) sites of the V sublattice [14,15]. The axial ratio c/a for the tetragonal distortion has been reported to be 1.11, 1.12, and 1.13 for VH$_{0.68}$ [14], VH$_{0.82}$ [15], and VH$_{0.90}$ [16], respectively: Although the β-phase hydride has some composition range, the axial ratio is nearly constant regardless of the composition. A static DFT calculation cannot reproduce this axial ratio which is overestimated to be $c/a \simeq 1.3$. In order to take into account the effect of the thermal fluctuation, we carry out the MD simulation with the MLP.

The rest of this paper is organized as follows: In Sec. II, our formalism for the machine learning potential is presented. Section III gives the results in which we validate the formalism and discuss the structural property of vanadium monohydride. Section IV summarizes the main results of this paper and gives brief comments for future work.

II. METHODOLOGY

A. Descriptor

In the MLP, atomic configurations are represented by the descriptors [17] which should be invariant with respect to atomic permutations, uniform translations, and rotations because the energy of a system is also invariant under these operations. Let us consider an atomic configuration around the Ith atom in a system located at $\boldsymbol{R}_I$. Fourier transform of the atomic distribution function around $\boldsymbol{R}_I$, $\rho_I(\boldsymbol{r}) = {\sum}'_i \delta(|\boldsymbol{r} - \boldsymbol{R}_i^I|)$ in the spherical coordinates is given by

$$\int \rho_I(\boldsymbol{r}) e^{-\imath \boldsymbol{q}\cdot\boldsymbol{r}}\, d\boldsymbol{r} = 4\pi \sum_l (-\imath)^l \sum_{m=-l}^{l} \left[{\sum_i}' j_l(qR_i^I)\, Y_{lm}(\hat{\boldsymbol{R}}_i^I) \right] Y_{lm}(\hat{\boldsymbol{q}}), \qquad (1)$$

where $\boldsymbol{R}_i^I = \boldsymbol{R}_i - \boldsymbol{R}_I$, $R_i^I = |\boldsymbol{R}_i^I|$, j_l are the spherical Bessel functions, and Y_{lm} are the real spherical harmonics. The prime in the summation excludes the $i = I$ term. The quantity in the square brackets in Eq. (1) is closely related to angular-dependent radial distribution functions. Therefore, we use it with the finite number of sampling wave-vectors q to characterize the local atomic configuration,

$$f_{tnlm}^I = {\sum_{i\in t}}' f_{\rm cut}(R_i^I)\, j_l(q_n R_i^I)\, Y_{lm}(\hat{\boldsymbol{R}}_i^I), \qquad (2)$$

where indices t for atom types are introduced to distinguish the distribution of different types of atoms and $f_{\rm cut}$ is a smooth radial cutoff function with compact support. The sampling

---

*miwa@cmp.tytlabs.co.jp

2469-9950/2016/94(18)/184109(6)

wave-vector lengths are set to be $q_n = 2\pi n/R_{\mathrm{cut}}$, where R_{cut} is a cutoff radius of f_{cut}. The function f^I_{tnlm} is invariant with respect to atomic permutations and translations but not under rotations since Y_{lm} depends on the choice of an azimuthal axis. A uniform rotation $\hat{S}$ transforms Y_{lm} into $\sum_{m'} D^l_{mm'}(\hat{S})Y_{lm'}$ where $D^l_{mm'}$ is a unitary matrix (the Wigner matrix [18] in the case of complex Y_{lm}), $(D^l)^T = (D^l)^{-1}$. Therefore, the following quantity becomes rotationally invariant:

$$\boldsymbol{g}^I = g^I_{tt'nl} = \sum_{m=-l}^{l} f^I_{tnlm} f^I_{t'nlm}, \tag{3}$$

which is the descriptor used in this paper. Our descriptor is categorized as the power spectrum [17]. The maximum number $n_{\max}$ for n as well as $l_{\max}$ for l control the resolution of the descriptor. Because Eq. (1) provides a complete representation for local atomic configurations, it is expected that the descriptor can systematically be improved by increasing $n_{\max}$ and $l_{\max}$.

Using the descriptors given in Eq. (3), the similarity measure between two local atomic environments is defined as follows: When the Ith and Jth atoms belong to the same atom type,

$$d_l^2(I,J) = \sum_n \sum_t \sum_{t' \leqslant t} \left| g^I_{tt'nl} - g^J_{tt'nl} \right|^2, \tag{4}$$

otherwise $d_l^2(I,J) = \infty$. For later convenience, l dependency is retained for the similarity measure.

B. Machine learning potential

In the MLP, the total energy of a system is evaluated by interpolating prepared DFT data sets. Using the similarity measure of Eq. (4), the total energy is expressed by a sum of weighted Gaussians,

$$\begin{aligned} E_{\mathrm{tot}} &= \sum_I \sum_{J \in \mathrm{ref}} \alpha_J \exp\left[-\sum_l \frac{d_l^2(I,J)}{2\theta_{tl}^2} \right] \\ &= \sum_{J \in \mathrm{ref}} \alpha_J \sum_I \mathcal{Q}(\boldsymbol{g}^I, \boldsymbol{g}^J), \end{aligned} \tag{5}$$

where α are the regression coefficients, θ are the scale parameters, and J runs over the reference configurations (see Sec. II C). In this paper, we assume that θ's are anisotropic and depend on the Ith atom type t and the angular momentum l. The first derivative of the total energy with respect to atomic displacements and uniform strain provides the forces acting on atoms $\boldsymbol{F}$ and the macroscopic stress tensor $\boldsymbol{\sigma}$, respectively, which are given by

$$\begin{aligned} \boldsymbol{F}_i &= -\sum_I \sum_{J \in \mathrm{ref}} \alpha_J \frac{\partial \mathcal{Q}(\boldsymbol{g}^I, \boldsymbol{g}^J)}{\partial \boldsymbol{g}^I} \frac{\partial \boldsymbol{g}^I}{\partial \boldsymbol{R}_i^I} \\ &= \sum_{J \in \mathrm{ref}} \alpha_J \sum_I \boldsymbol{\Phi}_i(\boldsymbol{g}^I, \boldsymbol{g}^J), \end{aligned} \tag{6}$$

$$\begin{aligned} \boldsymbol{\sigma} &= \frac{1}{\Omega} \sum_I \sum_{J \in \mathrm{ref}} \alpha_J \sum_i{}' \frac{\partial \mathcal{Q}(\boldsymbol{g}^I, \boldsymbol{g}^J)}{\partial \boldsymbol{g}^I} \frac{\partial \boldsymbol{g}^I}{\partial \boldsymbol{R}_i^I} (\boldsymbol{R}_i^I)^T \\ &= \sum_{J \in \mathrm{ref}} \alpha_J \sum_I \boldsymbol{\Upsilon}(\boldsymbol{g}^I, \boldsymbol{g}^J), \end{aligned} \tag{7}$$

where Ω is the unit-cell volume. Although $\boldsymbol{\sigma}$ in Eq. (7) is a symmetric second-order tensor, it is, hereafter, treated as a vector with the Voigt notation.

In the training process, the potential parameters are determined so as to reproduce the DFT results calculated for sample structures. To enhance the quality of the MLP, in addition to the total energy, the forces acting atoms and the macroscopic stress tensor also are utilized as the target quantities for training. Let $\boldsymbol{y}$ is a collective vector for DFT results,

$$\boldsymbol{y}^T = \left(E_{\mathrm{tot}}^{(1)}, \left\{ \boldsymbol{F}_i^{(1)} \right\}^T, \{\boldsymbol{\sigma}^{(1)}\}^T, \ldots, E_{\mathrm{tot}}^{(N)}, \left\{ \boldsymbol{F}_i^{(N)} \right\}^T, \{\boldsymbol{\sigma}^{(N)}\}^T \right), \tag{8}$$

where N denotes the number of the sample structures. In the MLP, since E_{tot}, $\boldsymbol{F}_i$, and $\boldsymbol{\sigma}$ are the linear functions of the same regression coefficients as shown in Eqs. (5)–(7), the corresponding vector can be expressed by $\boldsymbol{y}_{\mathrm{MLP}} = \mathcal{K}\boldsymbol{\alpha}$ with $\mathcal{K} = (\boldsymbol{k}_1, \ldots, \boldsymbol{k}_M)$ and

$$\begin{aligned} \boldsymbol{k}_J^T = \Bigg(&\sum_I \mathcal{Q}(\boldsymbol{g}^{I(1)}, \boldsymbol{g}^J), \left\{ \sum_I \boldsymbol{\Phi}_i(\boldsymbol{g}^{I(1)}, \boldsymbol{g}^J) \right\}^T, \\ &\left\{ \sum_I \boldsymbol{\Upsilon}(\boldsymbol{g}^{I(1)}, \boldsymbol{g}^J) \right\}^T, \ldots \Bigg), \end{aligned} \tag{9}$$

where M is the number of the reference configurations. Since $\boldsymbol{y}$ and $\boldsymbol{y}_{\mathrm{MLP}}$ are composed of the quantities with different physical units, it is necessary to make them dimensionless. To this end, we introduce the diagonal matrix $\mathcal{D}$, whose elements are chosen to be the inverses of the target criteria for the total energy, the force, and the macroscopic stress. With the standard L_2 regularization, the squared sum of residuals is given by

$$Z = (\boldsymbol{y} - \mathcal{K}\boldsymbol{\alpha})^T \mathcal{D}^T \mathcal{D} (\boldsymbol{y} - \mathcal{K}\boldsymbol{\alpha}) + \lambda \boldsymbol{\alpha}^T \boldsymbol{\alpha}, \tag{10}$$

where λ is the regularization parameter to prevent overfitting. When the scale parameters θ and the regularization parameter λ are given, the regression coefficients $\boldsymbol{\alpha}$ can be obtained by solving the following normal equation:

$$(\mathcal{K}^T \mathcal{D}^T \mathcal{D} \mathcal{K} + \lambda \mathcal{I})\boldsymbol{\alpha} - \mathcal{K}^T \mathcal{D}^T \mathcal{D} \boldsymbol{y} = 0, \tag{11}$$

where $\mathcal{I}$ denotes the unit matrix. The minimization of Z with respect to θ and λ is carried out using the simplex method [19]. When the last term in Eq. (10) for the L_2 regularization, $\lambda \boldsymbol{\alpha}^T \boldsymbol{\alpha}$, is replaced by $\lambda \boldsymbol{\alpha}^T \mathcal{Q} \boldsymbol{\alpha}$, the corresponding normal equation coincides with the expression for the predicted value in the Gaussian process with sparesification [11].

C. Reference configurations

The selection of all local configurations contained in the sample structures as the reference configurations is a natural choice. A large number of the references is, however, computationally unfavorable because the training cost is proportional to the cube of this size. When some configurations in the reference set are highly correlated, one will be able to reduce the number of the reference configurations without significant loss of accuracy. In this paper, we adopt a recursive bisection approach with the k-means clustering [20]. For simplicity, we consider a monoatomic system in this section: In the case of a multiatomic species system, the following procedure should be repeated for each atom type.

First, the local configurations in the sample structures are divided into two clusters by the k-means method where the gravity centers of clusters are defined by arithmetic means and a distance between two configurations is evaluated by $r = |\boldsymbol{g}^I - \boldsymbol{g}^J|$. Then, checking the cluster size (defined by the maximum distance from the center); if the size is smaller than the predetermined criterion δr, the center is selected as a new reference configuration, and the cluster is removed from the iteration; otherwise the configurations in the cluster are classified into two clusters, and the sizes of these clusters are checked again. This procedure is iterated until all clusters are removed.

This approach is computationally efficient, and the number of the reference configurations can be controlled by a single parameter δr.

D. Error estimation

The predictive power of the MLP strongly depends on the distribution of the reference configurations in the configurational space. A poor accuracy is expected if the references are rarely spanned around a prediction point and vice versa. Because $\mathcal{Q}(\boldsymbol{g}^I,\boldsymbol{g}^J)$ in Eq. (5) is thought to be an approximated overlap between two configurations, an error of the MLP at the prediction point $\boldsymbol{g}$ can be estimated by

$$s(\boldsymbol{g}) = 1 - \sum_{I\in\mathrm{ref}}\sum_{J\in\mathrm{ref}} \mathcal{Q}(\boldsymbol{g},\boldsymbol{g}^I)\mathcal{Q}^{-1}(\boldsymbol{g}^I,\boldsymbol{g}^J)\mathcal{Q}(\boldsymbol{g}^J,\boldsymbol{g}). \quad (12)$$

Hereafter, we refer this function as the spilling factor, which is zero when the configuration $\boldsymbol{g}$ is fully projected by the reference configurations, whereas $s = 1$ when $\boldsymbol{g}$ has no overlaps with them.

Equation (12) resembles the expression for the variance of the joint probability in the Gaussian process [11,21]. Note that our formalism is not based on the Gaussian process. Unlike the covariance matrix in the Gaussian process, $\mathcal{Q}$ is non-negative definite, not positive definite. If the reference set includes two or more identical configurations, $\mathcal{Q}^{-1}$ becomes singular. The recursive bisection method described in the previous section is useful not only for reducing the computational cost, but also for avoiding this problem.

III. RESULTS AND DISCUSSION

A. DFT sample data set

A DFT sample data set has been generated using a variable cell-shape *ab initio* MD simulation with the isothermal-isobaric ensemble. A supercell containing 32 formula units is used, and the hydrogen content is assumed to be $x = 1$ for simplicity. Temperature and pressure are controlled to be $T = 300$ K and $P = 0$ GPa with the Nosé-Hoover thermostat [22,23] and the Wentzcovitch barostat [24], respectively. A calculation is carried out using the ultrasoft pseudopotential method [25,26] with the generalized gradient approximation for the exchange-correlation energy [27,28]. The cutoff energies are set to be 15 and 120 hartrees for the pseudowave functions and the charge density, respectively. The Brillouin-zone integration is performed with a $2 \times 2 \times 2$ k-point mesh where partial occupation numbers for the states near the Fermi level are determined from the Fermi-Dirac distribution function with $k_BT = 3 \times 10^{-3}$ hartree.

In order to recover a wide range of the configurational space, a hypothetical zinc-blende structure is chosen as the initial configuration in which V atoms form a face-centered cubic (fcc) sublattice and H atoms occupy half of the tetrahedral (T) sites between them. Starting from this configuration, we run 20 000 MD steps, and then 100 snapshots are collected as the sample data.

In the *ab initio* MD simulation, the phase transition from fcc to bct occurs during the first 5000 MD steps. The time average over the last 2000 steps gives the O-site occupation of H in the bct structure correctly. The average axial ratio is $c/a = 1.22$: Although the improvement from the static calculation can be found, the amount of the correction is still insufficient. This is probably due to poor statistics (the small cell size and the short time period for sampling). As shown in Sec. III C, the MD simulation with the MLP generated from these sample data reproduces the experimental lattice constants quite well.

B. MLP construction

The parameters for the descriptor are set to be $n_{\max} = 12$, $l_{\max} = 6$, and $R_{\mathrm{cut}} = 10$ bohr. The prepared DFT sample data set contains 6400 local configurations. From them, we construct two sets of the reference configurations with the parameters for the recursive bisection method $\delta r^2 = 0.07$ and 0.1. The obtained reference sets, the MLP1 and the MLP2, are composed of 3289 and 1521 configurations, respectively.

We use fivefold cross validation for training [13,29]. The target criteria for the diagonal matrix $\mathcal{D}$ are chosen as 1×10^{-4} hartree/atom, 1×10^{-3} hartree/bohr, and 1 GPa for the energy, force, and stress, respectively. Table I summarizes the results of training. The quality of fitting is about 10 meV/atom for both the MLP1 and the MLP2. Because of smaller Z for the MLP1, the overall quality of the MLP1 is expected to be somewhat higher than that of the MLP2.

Both potentials are applied for four model structures, zinc blende (fcc T-site occupation), rock salt (fcc O-site), and T-site and O-site occupations in the bct structure. Except for zinc blende, these structures are not included explicitly in the DFT sample set. Table II shows the results of the MLP calculations together with the DFT results. The agreement with the DFT results is fairly good. The lattice constants agree within 2%. The cohesive energies are reproduced in the correct order, although the MLPs underestimate them about 15 meV/atom for the bct phases. Comparing the results between the MLP1 and the MLP2, no remarkable reduction in accuracy is found

TABLE I. Root-mean-square errors for the energy ΔE (hartree/atom), force ΔF (hartree/bohr), and stress $\Delta\sigma$ (GPa). Squared sum of residuals Z and the number of the reference configurations N_{ref}.

	N_{ref}	ΔE	ΔF	$\Delta\sigma$	Z
MLP1	3289	3.54×10^{-4}	3.24×10^{-3}	1.32	8.04
MLP2	1521	3.38×10^{-4}	3.51×10^{-3}	1.37	8.26

TABLE II. Comparison of the results given by the MLP and DFT calculations for VH. Lattice constants a (Å), axial ratios c/a, and cohesive energies E_{coh} (eV/atom).

	MLP1			MLP2			DFT		
	a	c/a	E_{coh}	a	c/a	E_{coh}	a	c/a	E_{coh}
fcc T site	4.075	1	4.543	4.075	1	4.542	4.065	1	4.546
fcc O site	3.940	1	4.636	3.934	1	4.635	3.932	1	4.641
bct T site	3.150	1.007	4.653	3.130	1.034	4.652	3.172	0.998	4.666
bct O site	2.903	1.267	4.638	2.878	1.288	4.640	2.870	1.293	4.654

for the MLP2. The recursive bisection method provides an efficient way to reduce the size of the reference set.

C. β-phase vanadium monohydride

Using the MLP1 model, a variable cell-shape MD simulation has been performed to investigate the structural property of β-VH. Temperature and pressure are controlled to be $T = 300$ K and $P = 0$ GPa by the same way as the *ab initio* MD. A supercell containing 256 formula units is used. The bct structure with the O-site occupation of H obtained by the static DFT calculation is chosen as the starting configuration. The time step is set to be 0.48 fs. After 10 000 MD steps for equilibration, the average is taken over 20 000 MD steps.

Figure 1 shows the profiles of the MD run. The conservation of the constant of motion (the extended Hamiltonian for the isothermal-isobaric ensemble) is excellent. Temperature and pressure correctly fluctuate around the target value. During the first 1000 MD steps, the c axis is shrunk, and the square base is expanded. The average lattice constants are $a = 3.06$ Å and $c/a = 1.12$, which are in good agreement with experimental data for $VH_{0.90}$ [16], $a = 3.03$ Å, and $c/a = 1.13$.

The time-averaged crystal structure is shown in Fig. 2(a). The O-site occupation of H in the bct structure is cleanly seen (more exactly, the O_z site where H atoms occupy the middle of the nearest-neighboring V atoms along the c axis). However, the atom trajectories given in Fig. 2(b) indicate that H atoms notably fluctuate around the O sites along the basal plane. From the trajectories, we calculate atom density functions: All trajectories are folded into the bct primitive cell, and atom positions are broadened by a Gaussian with a width of 0.1 bohr. The result is shown in Fig. 3. The thermal vibration of V is found to be almost isotropic, whereas the distribution of H shows four peaks which are placed near the T sites on the basal plane. The displacement of the peaks from the O site is $\delta/a = 0.16$. This indicates that H atoms have the 4T configuration which has been proposed as one of the candidates for the self-trapped state of hydrogen isotopes in bcc metals [30]. The symmetry of the 4T configuration is the same as the O-site occupation.

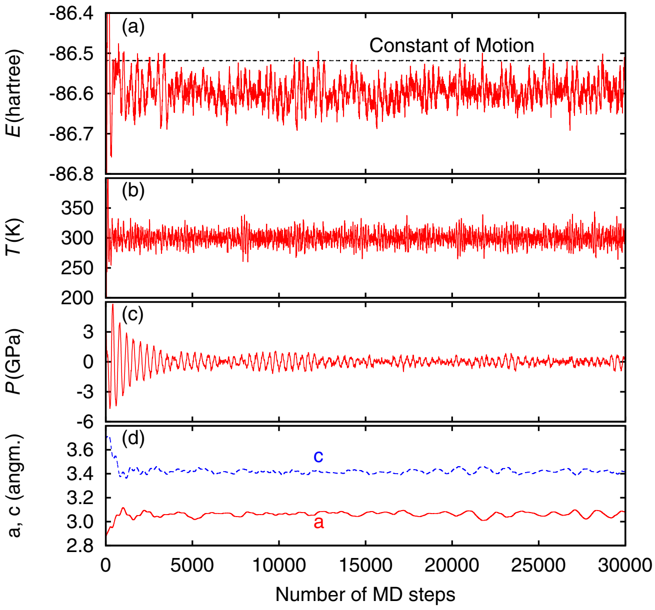

FIG. 1. Profiles of the MLP MD simulation. (a) Energy, (b) temperature, (c) pressure, and (d) lattice constants.

Since the hydrogen density at the O site is relatively low, it is expected to be a saddle point on the potential-energy surface rather than a local minimum. Figure 4(a) depicts the total energy of bct-VH as a function of the rigid displacement of the H sublattice δ along the [100] direction where the lattice constants are kept fixed at the average values of the MD run. We can confirm that the O site is a saddle point. The energy gradually decreases with increasing δ and shows the minimum at $\delta/a = 0.18$. There is a small deviation between this value and the peak position of the atom density because this calculation is imposed on the periodicity of the bct primitive cell.

Another important point is that, although the energies obtained by the MLP1 agree very well with the DFT results for $\delta/a \lesssim 0.35$, they deviate from the DFT results for δ/a larger than it. The structure at $\delta/a = 0.5$ corresponds to the O_y-site occupation of H, which has a relatively high energy due to the short V-V distance along the a axis. Because the high-energy configurations are rarely sampled during the *ab initio* MD run, the accuracy of the MLP is expected to become low for them. The spilling factor given in Fig. 4(b) supports this consideration, which correlates well with the energy deviation between the MLP and the DFT results. The results of the MLP2 are also given in Figs. 4(a) and 4(b). As expected, the accuracy of the MLP2 is comparable with that of the MLP1, particularly, in the range of $\delta/a \lesssim 0.35$.

Figure 5 shows the profile of the spilling factor during the MD run with the MLP1. Overall, the maximum value of the spilling factor at each MD step is smaller than 0.02. The corresponding error of the energy is estimated as $\Delta E \simeq 20$ meV/atom from Fig. 4(b). Although several spikes exceed this tolerance, they are attenuated quickly and so have little

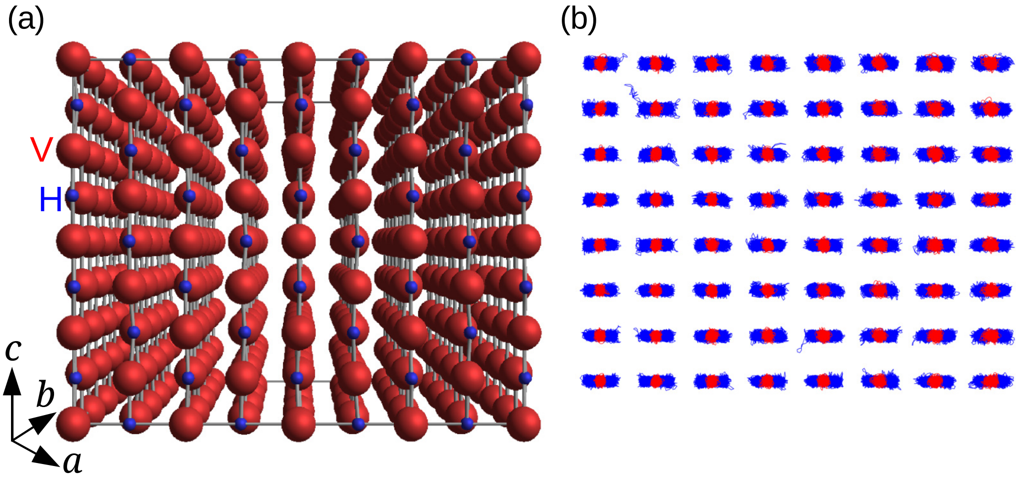

FIG. 2. Results of the MLP MD simulation. (a) Time-averaged structure and (b) atom trajectories projected onto the (110) plane.

effect on the statistics. The average spilling factors are always one order of magnitude smaller than the tolerance. Therefore, good accuracy for the statistics with $\Delta E \ll 20$ meV/atom is expected for the present MD simulation.

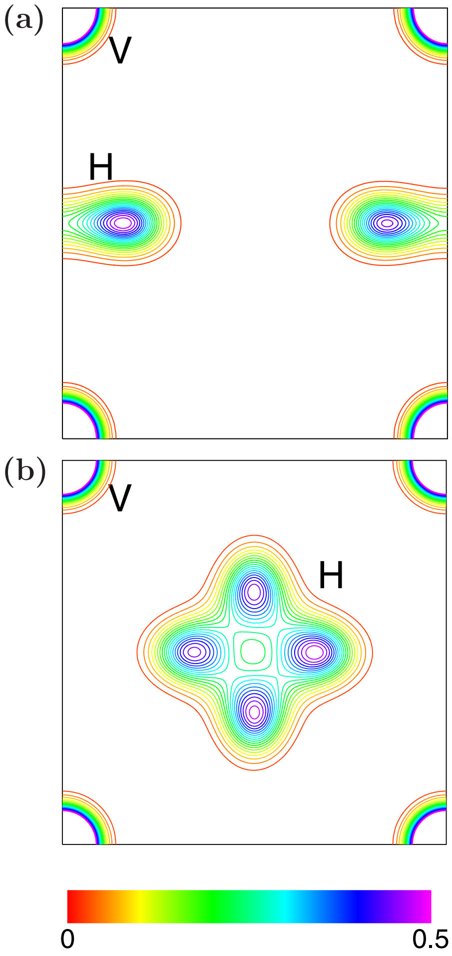

FIG. 3. Contour plots for atom density functions on (a) the (100) and (b) the (001) planes. The contour spacing is 0.025 bohr^{-3}, and the lines are omitted for densities higher than 0.5 bohr^{-3}.

IV. SUMMARY

We have developed the formalism to construct the machine learning potentials, which includes the spilling factor for the simultaneous error estimation and the recursive bisection method for the reduction in the computational cost. The formalism is applied for β-phase vanadium monohydride. In the MD simulation, the time-averaged structure of β-VH is predicted to be the O-site occupation of H in the bct structure as stated in the literature. The average lattice constants are in good agreement with the experimental data which are not able to be reproduced by the static DFT calculation. The O-site occupation of H observed in the average structure is, however, a saddle point on the potential-energy surface, and the actual hydrogen occupation is found to be the $4T$ configuration. Note

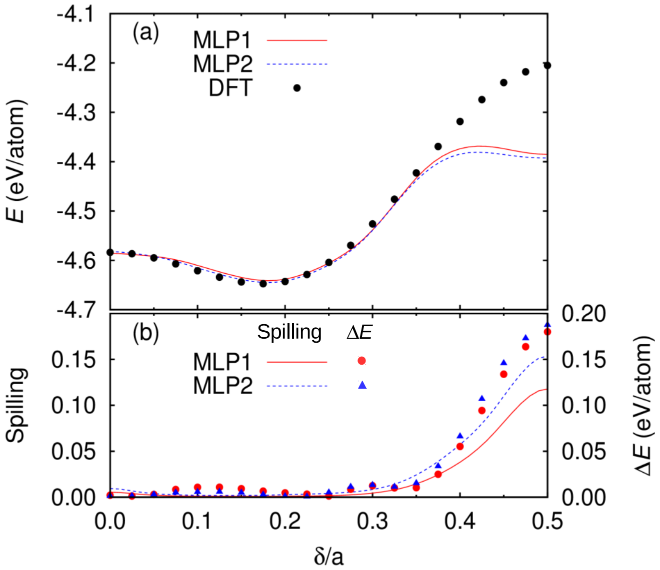

FIG. 4. (a) Energy of bct-VH as a function of the rigid displacement of the H sublattice δ and (b) the spilling factor for the MLP calculations and the energy deviation $\Delta E = |E_{\mathrm{DFT}} - E_{\mathrm{MLP}}|$. For the spilling factor, the values of V are only plotted because those of H are always smaller than them in this calculation.

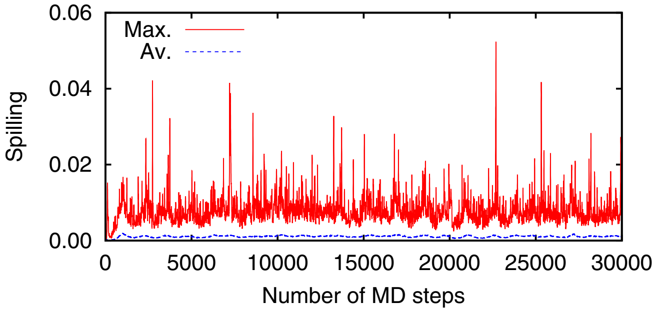

FIG. 5. Profiles of the spilling factor during the MLP MD run. The red (blue) line indicates the maximum (average) value at each MD step.

that recent neutron-diffraction experiments for the V-based alloy also suggest a significant thermal fluctuation of H in the β-phase hydride [31,32].

The remaining issues for future work are as follows: In this paper, the DFT sample data are prepared from the *ab initio* MD run, which might be computationally prohibitive for complex systems and/or phenomena. An alternative way is to execute the MD run with the MLP itself. The spilling factor gives us useful information, which snapshots should be included in the sample data set. The DFT calculations are carried out only for the collected sample structures. The quality of the MLP can be improved by repeating this training process.

In an application viewpoint, the present calculation treats H atoms as classical particles. This treatment sometimes becomes inadequate due to the light mass of hydrogen. Although the path-integral molecular dynamics (PIMD) [33] can be used to incorporate the nuclear quantum effect, a large-scale simulation with the DFT-based PIMD is still a challenging task. A PIMD simulation combined with the MLP will provide an efficient way to treat the quantum-mechanical effect for hydrogen with the accuracy comparable to DFT.

ACKNOWLEDGMENTS

We thank Dr. R. Jinnouchi for fruitful discussions about the MLP construction. We also thank Professor E. Akiba, Dr. K. Asano, and Dr. K. Sakaki for valuable discussions and for informing us about the experimental details.

---

[1] M. E. Tuckerman and G. J. Martyna, J. Phys. Chem. B **104**, 159 (2000).
[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[3] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[4] J. E. Lennard-Jones, Proc. Phys. Soc. **43**, 461 (1931).
[5] F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).
[6] J. Tersoff, Phys. Rev. B **39**, 5566 (1989).
[7] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, J. Phys.: Condens. Matter **14**, 783 (2002).
[8] M. S. Daw and M. I. Baskes, Phys. Rev. B **29**, 6443 (1984).
[9] M. I. Baskes, Phys. Rev. B **46**, 2727 (1992).
[10] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[11] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
[12] A. Seko, A. Takahashi, and I. Tanaka, Phys. Rev. B **90**, 024101 (2014).
[13] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, J. Chem. Theory Comput. **9**, 3404 (2013).
[14] K. Asano, S. Hayashi, Y. Nakamura, and E. Akiba, J. Alloys Compd. **507**, 399 (2010).
[15] B. Bandyopadhyay and S. Hayashi, Phys. Rev. B **60**, 10302 (1999).
[16] A. J. Maeland, J. Phys. Chem. **68**, 2197 (1964).
[17] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).
[18] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1965), Vol. II.
[19] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).
[20] A. K. Jain, M. N. Murty, and P. J. Flynn, ACM Comput. Surv. **31**, 264 (1999).
[21] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (MIT Press, Cambridge, MA, 2005).
[22] S. Nosé, Mol. Phys. **52**, 255 (1984).
[23] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[24] R. M. Wentzcovitch, Phys. Rev. B **44**, 2358 (1991).
[25] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
[26] K. Miwa, Phys. Rev. B **84**, 094304 (2011).
[27] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[28] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396(E) (1997).
[29] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
[30] H. Sugimoto and Y. Fukai, Phys. Rev. B **22**, 670 (1980).
[31] K. Kamazawa, M. Aoki, T. Noritake, K. Miwa, J. Sugiyama, S. Towata, M. Ishikiriyama, S. K. Callear, M. O. Jones, and W. I. F. David, Adv. Energy Mater. **3**, 39 (2013).
[32] S. Itano and E. Akiba (private communication).
[33] D. Marx and M. Parrinello, J. Chem. Phys. **104**, 4077 (1996).