Rapid Communications

Self-learning hybrid Monte Carlo: A first-principles approach

Yuki Nagai,[1,2,*] Masahiko Okumura,[1] Keita Kobayashi,[3] and Motoyuki Shiga[1,†]
[1]*CCSE, Japan Atomic Energy Agency, 178-4-4, Wakashiba, Kashiwa, Chiba, 277-0871, Japan*
[2]*Mathematical Science Team, RIKEN Center for Advanced Intelligence Project (AIP), 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan*
[3]*Research Organization for Information Science and Technology (RIST), 2-4, Shirakata, Tokai-mura, Ibaraki 319-1106, Japan*

(Received 6 September 2019; revised 26 March 2020; accepted 13 July 2020; published 29 July 2020)

We propose an approach called self-learning hybrid Monte Carlo (SLHMC), which is a general method to make use of machine learning potentials to accelerate the statistical sampling of first-principles density-functional-theory (DFT) simulations. The trajectories are generated on an approximate machine learning (ML) potential energy surface. The trajectories are then accepted or rejected by the Metropolis algorithm based on DFT energies. In this way, the statistical ensemble is sampled exactly at the DFT level for a given thermodynamic condition. Meanwhile, the ML potential is improved on the fly by training to enhance the sampling, whereby the training data set, which is sampled from the exact ensemble, is created automatically. Using the examples of α-quartz crystal SiO_2 and phonon-mediated unconventional superconductor YNi_2B_2C systems, we show that SLHMC with artificial neural networks (ANN) is capable of very efficient sampling, while at the same time enabling the optimization of the ANN potential to within meV/atom accuracy. The ANN potential thus obtained is transferable to ANN molecular dynamics simulations to explore dynamics as well as thermodynamics. This makes the SLHMC approach widely applicable for studies on materials in physics and chemistry.

DOI: 10.1103/PhysRevB.102.041124

I. INTRODUCTION

First-principles molecular dynamics based on density functional theory (DFT-MD) is a powerful tool to simulate a variety of materials in physics and chemistry [1]. However, reducing the computational effort required for DFT-MD remains a key issue for its broader application to phenomena on large length- and timescales. The use of artificial neural networks (ANNs), which imitate DFT energies by machine learning, is seen as a promising solution to this issue [2–5]. The branch of research about machine learning molecular simulations has grown rapidly in the last decade [6–18] after an influential paper by Behler and Parrinello [5] laid down a general framework of setting up and training ANN potentials from DFT data sets and running ANN-MD simulations for condensed matter systems.

The training of machine learning (ML) potentials must be based on sufficient amounts of DFT-derived results to cover all the configuration space, which corresponds to a statistical ensemble in the case of systems in thermal equilibrium. Usually, the training sets for ML potentials are chosen before starting ML-MD simulations. Many useful methods, such as generic algorithms [16] and CUR decompositions [17], etc., were suggested for properly choosing training sets. If sufficient care is not taken, it is possible that ML-MD simulations may break down suddenly when the trajectory finds its way into an uncovered part of the phase space, see Fig. 1. Therefore it would be beneficial to establish a way to somehow cover the ensemble space in an automatic manner.

Here we propose an approach for training ML potentials based on the exact statistical ensemble at a given thermodynamic condition, e.g., isothermal, isobaric ensembles, using the hybrid Monte Carlo technique [19–21]. This allows one to access exact results out of approximate ML potentials, and, at the same time, create an objective and unique data set for ML training. Herein the exactness means that the ensemble created by this method is exactly equivalent to that created by the DFT-MD. Importantly, we want to circumvent the use of costly DFT-MD computations as much as possible, which can be achieved by training the ML potential on systems with small sizes and short timescales. Our idea is to use the dual-level HMC method [22–25]. The ensemble is sampled by generating trial moves of the trajectory from an approximate ML potential energy surface, which are then either accepted or rejected by performing the DFT energy calculations at long time-step intervals. Note that the ensemble created is theoretically exact irrespective of the quality of the approximate ML potentials. Its efficiency is strongly dependent, however, on the quality of the ML potential, as this affects acceptance ratio. It will be shown herein that the acceptance ratio tends to improve as the ML potential is trained and updated iteratively, which is done automatically on the fly during the computations. Note that the ML potential is used here in an auxiliary manner to produce exact ensembles, which distinguishes our concept from previous works where the ML potential was used as an accurate approximation of the DFT potentials.

We call our method the self-learning hybrid Monte Carlo (SLHMC) method, since it is akin to the self-learning Monte Carlo (SLMC) method [26]. The SLMC method was recently introduced in the field of many-electron systems to speed up MC simulations by using efficient global updates informed by machine learning techniques [26–32]. SLHMC extends

*nagai.yuki@jaea.go.jp
†shiga.motoyuki@jaea.go.jp

 ©2020 American Physical Society

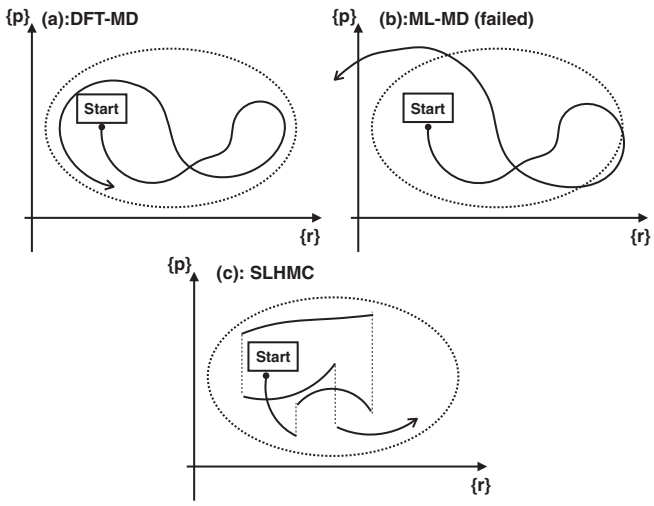

FIG. 1. Schematic figure of trajectories in phase space. (a) Trajectories of DFT-MD, (b) ML-MD, and (c) SLHMC. If the ML potential is not well-trained, the ML-MD trajectory might fall outside the range of the DFT ensemble (depicted by a dotted oval), while the accepted SLHMC trajectory always stays inside.

the idea of SLMC to equation-of-motion-based moves to enable efficient global updates for atomistic and molecular simulations.

This Rapid Communication is organized as follows: We first explain the basic theory and computational procedure of SLHMC. We use example calculations on SiO_2 (α quartz) to demonstrate the accuracy and efficiency of the method. Calculations for the phonon-mediated superconductor YNi_2B_2C are then used to demonstrate the ability to construct accurate ML potentials such that the ML-MD simulations are stable at long times.

II. SELF-LEARNING HYBRID MONTE CARLO METHOD

A trial move uses Hamilton's equations of motion derived from the ML potential energy surface, $V_{\rm ML}$,

$$\dot{\boldsymbol{p}}_i = -\frac{\partial V_{\rm ML}(\{\boldsymbol{r}\},t)}{\partial \boldsymbol{r}_i}, \quad \dot{\boldsymbol{r}}_i = \frac{\boldsymbol{p}_i}{m_i}, \tag{1}$$

where $\boldsymbol{p}_i$, $\boldsymbol{r}_i$, and m_i are the momentum, coordinates, and mass of the ith atom, respectively. Starting with a random initial momentum generated from the Maxwell-Boltzmann distribution, the equations of motion are solved for a discrete time step, $\Delta t_{\rm ML}$, using a time-reversible and area-preserving algorithm (in the present study, the velocity-Verlet algorithm). The ML potential surface $V_{\rm ML}(\{\boldsymbol{r}\},t)$ depends on time t as it is trained on the fly, but is kept constant within the time interval $\Delta t_{\rm T}$, i.e.,

$$V_{\rm ML}(\{\boldsymbol{r}\},t) = V^n_{\rm ML}(\{\boldsymbol{r}\}), \quad n\Delta t_{\rm T} < t < (n+1)\Delta t_{\rm T}, \tag{2}$$

where $V^n_{\rm ML}$ is the ML potential for the nth update. $V_{\rm ML}$ is trained every $\Delta t_{\rm T} = n_{\rm DFT}\Delta t_{\rm DFT}$, where $\Delta t_{\rm DFT} \equiv n_{\rm ML}\Delta t_{\rm ML}$ is the interval of acceptance/rejection in the Metropolis algorithm. Thus, $n_{\rm DFT}$ is the number of times the DFT energy is computed for training and $n_{\rm ML}$ is the number of steps in a trial move. The acceptance probability for a trial move within the phase space from $\{\boldsymbol{p},\boldsymbol{r}\}$ to $\{\boldsymbol{p}',\boldsymbol{r}'\}$ is given by (see Supplemental Material for the technical details [33])

$$P_{\rm acc}(\{\boldsymbol{p},\boldsymbol{r}\} \to \{\boldsymbol{p}',\boldsymbol{r}'\}) = \min(1, e^{-\beta(H_{\rm DFT}(\{\boldsymbol{p}',\boldsymbol{r}'\})-H_{\rm DFT}(\{\boldsymbol{p},\boldsymbol{r}\}))}), \tag{3}$$

where $\beta = 1/T$ is the inverse of temperature and

$$H_{\rm DFT} = \sum_i^N \frac{|\boldsymbol{p}_i|^2}{2m_i} + V_{\rm DFT}(\{\boldsymbol{r}\}) \tag{4}$$

is the Hamiltonian based on the DFT potential energy, $V_{\rm DFT}(\{\boldsymbol{r}\})$. We note that the detailed balance condition is preserved exactly on the basis of the DFT (not ML) potential as long as the ML potential $V_{\rm ML}(\{\boldsymbol{r}\},t)$ does not change during a trial move.

According to Eqs. (1) and (3), ML force calculations are required $n_{\rm ML}$ times while DFT energy calculations are required once. SLHMC is computationally efficient when the former is less expensive than the latter, which is usually the case unless we set a huge value for $n_{\rm ML}$. In such a case, $\Delta t_{\rm ML}$ could be chosen to be sufficiently small so as to conserve the ML energy within a trial move. (This situation is different from conventional HMC where the step size should be large to break energy conservation.) When the ML energy is conserved within a trial move, the acceptance probability Eq. (3) becomes

$$P_{\rm acc}(\{\boldsymbol{p},\boldsymbol{r}\} \to \{\boldsymbol{p}',\boldsymbol{r}'\}) \sim \min(1, e^{-\beta\Delta\Delta V}), \tag{5}$$

where $\Delta\Delta V \equiv \Delta V(\{\boldsymbol{r}'\}) - \Delta V(\{\boldsymbol{r}\})$ and $\Delta V(\{\boldsymbol{r}\}) \equiv V_{\rm DFT}(\{\boldsymbol{r}\}) - V_{\rm ML}(\{\boldsymbol{r}\},t)$. Therefore, the accuracy of the ML potential influences the acceptance ratio, and thus the efficiency of the SLHMC method. In practice, SLHMC is efficient by setting $\Delta t_{\rm DFT}$ such that the acceptance ratio is more than 20%.

In this Rapid Communication, we use the Behler-Parrinello ANN potentials as ML potentials. We follow the standard ANN training protocol to minimize the mean-square error $(\Delta V(\{\boldsymbol{r}\})^2)$ [5]. The ANN variables are restarted from the last update $(n-1)$ and optimized using all the DFT energy data up to the current update (n). Training is possible using either only the accepted structures or both the accepted and rejected structures (the latter is employed here). We recommend the latter since the ML potential is improved more rapidly in this case. By including the rejected samples, the sampled space naturally expands in the course of the simulation.

Since the size of the training data set needed is small ($\sim$2000 here), the computational cost of training the ANN potential is usually not dominant in SLHMC. As the number of training data increase with increasing n, one can skip the training steps after the ANN variables are optimized sufficiently by judging from the average acceptance ratio calculated in SLHMC.

III. DEMONSTRATION I: THERMODYNAMICS OF SiO_2

The SLHMC method was implemented in the PIMD code [34] which has access to ANN potentials and DFT calculations via the ATOMIC ENERGY NETWORK (ænet) [35] and VIENNA AB INITIO SIMULATION PACKAGE [36]. The first

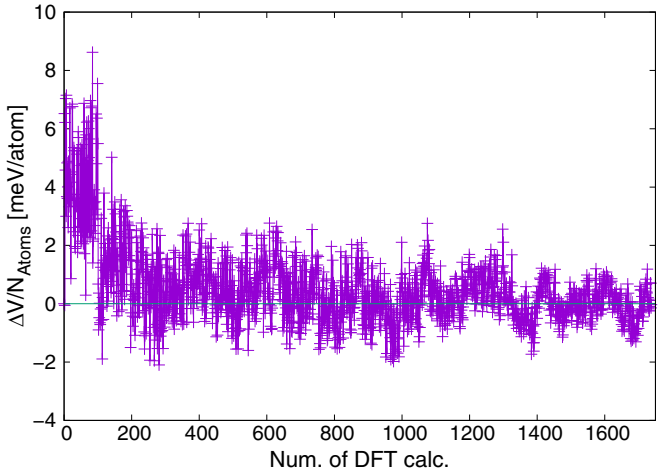

FIG. 2. Difference between the DFT and ANN energies of SiO_2 in SLHMC. The step size of ML, DFT, and training were chosen to be $\Delta t_{\mathrm{ML}} = 0.25$ fs, $\Delta t_{\mathrm{DFT}} = 5$ fs, and $\Delta t_{\mathrm{T}} = 500$ fs, respectively.

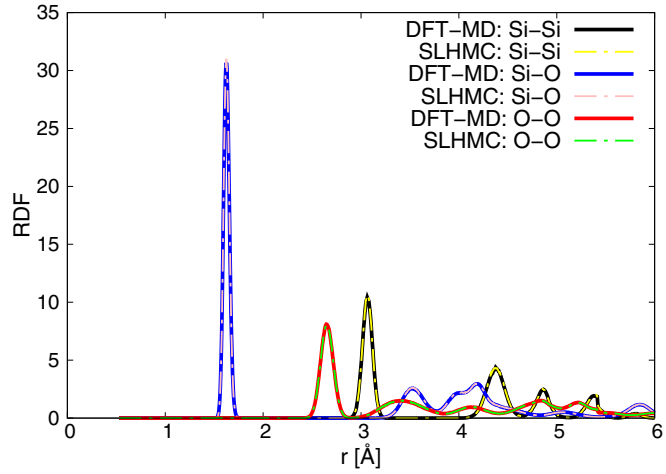

FIG. 3. Radial distribution functions of SiO_2 compared between DFT-MD and SLHMC. The discrete time step in the DFT-MD is $\Delta t = 1$ fs.

test was on the thermodynamics of the α-quartz phase of SiO_2 crystal [37]. The SLHMC simulations were at 300 K in the canonical ensemble for a periodic system of 24 Si atoms and 48 O atoms. The DFT calculations used the Perdew-Burke-Ernzerhof functional [38]. The projected augmented-wave method [39] was employed, while the cutoff energy was 500 eV and the sampling points were gamma point only. The ANN potentials were trained using the limited-memory Broyden-Fletcher-Goldfarb-Shanno method [40,41]. We adopt the Chebyshev basis set as a descriptor for atomic environments [11] and the corresponding parameters are given in Ref. [42]. The step size of ML, DFT, and training were chosen to be $\Delta t_{\mathrm{ML}} = 0.25$ fs, $\Delta t_{\mathrm{DFT}} = 5$ fs, and $\Delta t_{\mathrm{T}} = 500$ fs, respectively. The initial guess of the ANN potential was prepared and trained using a short DFT-MD trajectory of 300 steps starting from the crystal structure. This could be done in other ways if the SLHMC run gets stuck at the initial step.

Figure 2 shows that the difference between the ANN and DFT potentials quickly diminishes to about 1 meV/atom on average as the SLHMC simulation proceeds and the ANN potential is trained on the fly. This accuracy can be ascribed to the fact that the ANN potential has been trained in a confined configuration space corresponding to the exact ensemble at the DFT level. Figure 3 shows that the radial distribution functions (RDFs) obtained from the SLHMC and DFT-MD simulations are identical, as they should be. Note that this is the case even when the ANN potentials are changed during the SLHMC simulation. This demonstrates that the SLHMC method is able to gain statistics on thermodynamic properties while training the ANN potential at the same time. The RDFs in the SLHMC converge faster than those in the DFT-MD.

Figure 4 shows the results of the mean squared displacement (MSD) as a function of the number of DFT calculations, obtained with SLHMC and DFT-MD simulations. The MSD is defined as $(1/N_{\mathrm{Si}}) \sum_{k=1}^{N_{\mathrm{Si}}} (\mathbf{r}'_k(t) - \mathbf{r}'_k(0))^2$, where N_{Si} is the number of Si atoms and $\mathbf{r}'_k(t)$ is the position of the kth atom relative to the center-of-mass of the system at time t. These MSD curves, which should converge to a constant value for large t in solid states, indicate the length scale of the autocorrelation of atomic displacement. As expected, the MSD converges faster as Δt_{DFT} increases, since the ANN-MD trajectory is longer. Assuming that the DFT calculations are the bottleneck of the SLHMC simulations, which was mostly the case in our computations, SLHMC simulations become more efficient as the MSD converges. This is not only because the sampling of statistics becomes more efficient, but also because the training data becomes less correlated. When Δt_{DFT} is long enough to be uncorrelated in a single MC step, the efficiency of SLHMC should become proportional to the acceptance ratio, which is eventually reflected by the difference between the ANN and DFT potentials and thus the quality of the ANN. In the present simulation, the acceptance ratio with the well-trained ANNs with $\Delta t_{\mathrm{DFT}} = 50$ fs was around 40%, and the difference between the ANN and DFT potentials tended to 0.23 meV/atom on average.

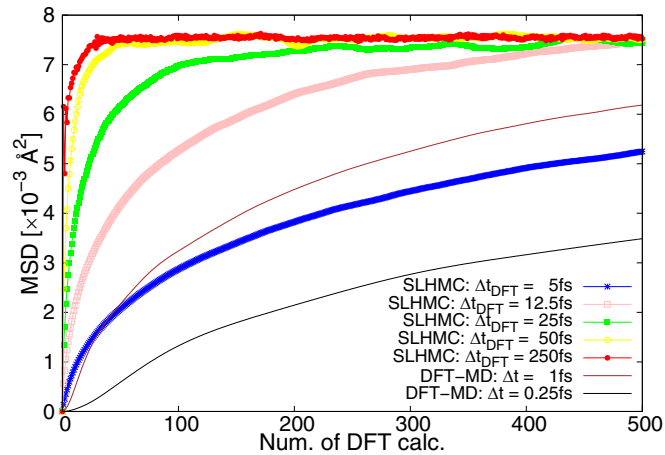

FIG. 4. Mean-squared displacement (MSD) of Si atoms for SiO_2 calculated by the SLHMC at 300 K. The horizontal axis is the number of the DFT calculations. We set $\Delta t_{\mathrm{ML}} = 0.25$ fs in SLHMC.

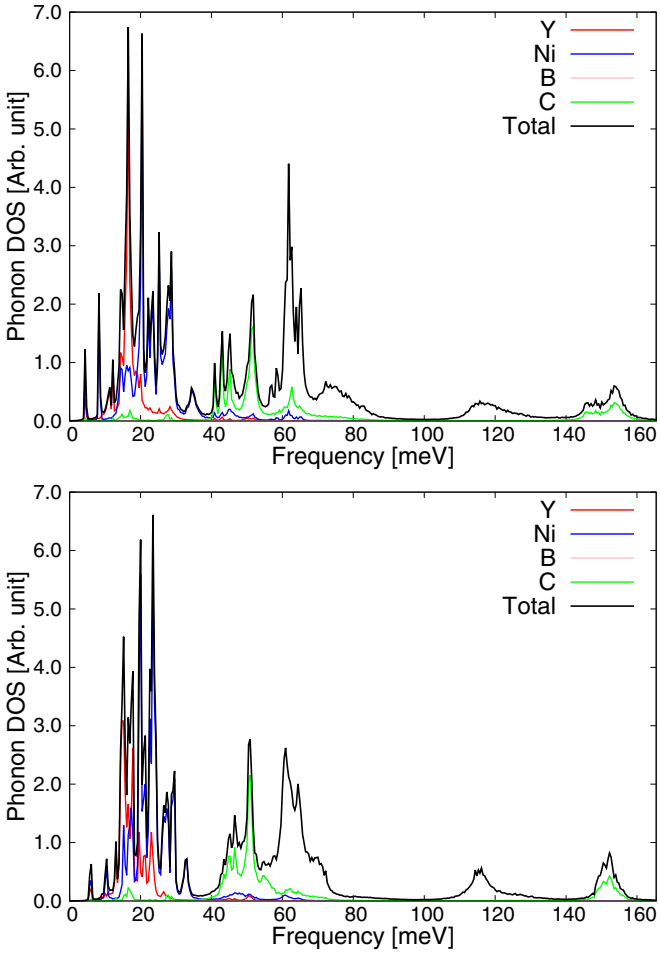

FIG. 5. Phonon density of states in YNi_2B_2C calculated by ANN-MD with the ANN trained by the SLHMC. The temperatures are 60 K (upper panel) and 300 K (lower panel).

IV. DEMONSTRATION II: DYNAMICS OF YNi_2B_2C

The second test was on the dynamics of an unconventional phonon-mediated superconductor YNi_2B_2C ($T_c \sim 15$ K) [43–46]. Neutron-scattering experiments have shown a strong temperature dependence of the phonon density of states (DOS) in this superconducting compound [47]. The temperature dependence arises from anharmonic vibrations that are not taken into account in static DFT calculations based on (quasi) harmonic analysis at zero temperature [47,48]. We show that a combination of the SLHMC and ANN-MD methods could be useful in this case. Since the SLHMC method optimizes the ANN potential in the configuration space of a given ensemble, the accuracy of ANN potential is guaranteed in ensuing ANN-MD simulations as the trajectories stay confined within this configuration space. Therefore, once the ANN potential is optimized in SLHMC, the ANN-MD runs should be stable for long times, which may not necessarily be the case for complex systems with several elements (in the present case, four elements) for the methods proposed earlier.

The SLHMC and ANN-MD simulations were carried out for a supercell containing 16 Y atoms, 32 Ni atoms, 32 B atoms, and 16 C atoms. The ANN and DFT simulations were set up in the same way as those in the previous section. The SLHMC simulations were at 1000 K with the step sizes $\Delta t_{\rm ML} = 0.25$ fs, $\Delta t_{\rm DFT} = 2.5$ fs, and $\Delta t_{\rm T} = 250$ fs. The difference of the trained ANN and DFT potentials was found to be about 0.4 meV/atom on average. The DOS was computed via the Fourier transform of velocity autocorrelations [49] from the ANN-MD simulations. For statistical convergence, ten Newtonian trajectories were run independently for 100 ps with the step size $\Delta t_{\rm MD} = 1$ fs. These trajectories were restarted from the equilibrated configurations of NVT ensemble at 60 K and 300 K to reflect the temperature dependence [50] of the DOS. As expected, all the ANN-MD trajectories were found to be stable for 100 ps.

As shown in Fig. 5, the phonon DOS depends on temperature, which is consistent with the neutron-scattering experiments [47]. This result confirms that anharmonic effects of phonons are important in this material. The crystal structure of YNi_2B_2C is similar to that of high-T_c cuprates and strongly anisotropic superconducting pairing has been suggested [44–46]. Thus, anharmonic effects of phonons appear to play a key role in this unconventional superconductor.

V. CONCLUSIONS

We proposed a method called SLHMC to compute thermodynamic properties exactly based on DFT using an approximate ML potential that is trained on the fly to accelerate sampling. The ML potential is optimized automatically by using a training data set of a given ensemble that is generated exactly. The ML potential thus obtained can be used safely in ML-MD simulations to compute dynamic properties in thermal equilibrium, since the ML-MD simulations are stable for long times. Proof-of-concept calculations were presented for the thermodynamics of SiO_2 and the dynamics of YNi_2B_2C, which demonstrated the usefulness of SLHMC.

As can be expected from the acceptance probability in Eq. (5), the efficiency of SLHMC depends strongly on the balance between the system size and the quality of the ML potential. Thus, it is recommended to keep the system reasonably small for SLHMC in practice. However, like other ANN methods in the framework of the Behler-Parrinello approach, the ANN potentials are transferable to larger systems once they are well-trained by SLHMC. Unlike the original HMC method, the small step size $\Delta t_{\rm ML}$ prevents deterioration of the efficiency of SLHMC. The SLHMC approach established herein is not limited to solids but could be applied generally to many kinds of systems such as molecular clusters and liquids. In principle, this idea could be extended to other statistical ensembles, such as the isobaric ensemble and quantum ensembles via imaginary-time path integral theory [51].

ACKNOWLEDGMENTS

The calculations were performed on the supercomputing system SGI ICE X at the Japan Atomic Energy Agency. This work was supported by JSPS KAKENHI Grants No. 18K03552 (Y.N.), No. 18K05208 (M.S. and M.O.), No. 18H01693, No. 18H05519, and No. 16K05675 (M.S.), and MEXT Fugaku Battery & Fuel Cell Project (M.S.). We thank Alex Malins for proofreading the manuscript.

[1] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, Cambridge, 2009).

[2] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, Neural network models of potential energy surfaces, J. Chem. Phys. **103**, 4129 (1995).

[3] D. F. R. Brown, Combining *ab initio* computations, neural networks, and diffusion Monte Carlo: An efficient method to treat weakly bound molecules, J. Chem. Phys. **105**, 7597 (1996).

[4] S. Lorenz, A. Groß, and M. Scheffler, Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks, Chem. Phys. Lett. **395**, 210 (2004).

[5] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[6] J. Behler, Constructing high-dimensional neural network potentials: A tutorial review, Int. J. Quantum Chem. **115**, 1032 (2015).

[7] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, Without the Electrons, Phys. Rev. Lett. **104**, 136403 (2010).

[8] N. Artrith, T. Morawietz, and J. Behler, High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide, Phys. Rev. B **83**, 153101 (2011).

[9] N. Artrith and A. M. Kolpak, Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: A combination of DFT and accurate neural network potentials, Nano Lett. **14**, 2670 (2014).

[10] Z. Li, J. R. Kermode, and A. D. Vita, Molecular Dynamics with On-The-Fly Machine Learning of Quantum-Mechanical Forces, Phys. Rev. Lett. **114**, 096405 (2015).

[11] N. Artrith, A. Urban, and G. Ceder, Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species, Phys. Rev. B **96**, 014112 (2017).

[12] J. S. Smith, O. Isayev, and A. E. Roitberg, ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost, Chem. Sci. **8**, 3192 (2017).

[13] W. Li, Y. Ando, E. Minamitani, and S. Watanabe, Study of Li atom diffusion in amorphous Li_3PO_4 with neural network potential, J. Chem. Phys. **147**, 214106 (2017).

[14] W. Li and Y. Ando, Construction of accurate machine learning force fields for copper and silicon dioxide, Phys. Chem. Chem. Phys. **20**, 30006 (2018).

[15] T. L. Jacobsen, M. S. Jørgensen, and B. Hammer, On-The-Fly Machine Learning of Atomic Potential in Density Functional Theory Structure Optimization, Phys. Rev. Lett. **120**, 026102 (2018).

[16] N. Artrith, A. Urban, and G. Cederb, Constructing first-principles phase diagrams of amorphous LixSi using machine-learning-assisted sampling with an evolutionary algorithm, J. Chem. Phys. **148**, 241711 (2018).

[17] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, and M. Ceriotti, Automatic selection of atomic fingerprints and reference configurations for machine- learning potentials, J. Chem. Phys. **148**, 241730 (2018).

[18] R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam, Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference, Phys. Rev. Lett. **122**, 225701 (2019).

[19] S. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, Hybrid-molecular-dynamics algorithms for the numerical simulation of quantum chromodynamics, Phys. Rev. D **35**, 2531 (1987).

[20] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo, Phys. Lett. B **195**, 216 (1987).

[21] B. Mehlig, D. W. Heermann, and B. M. Forrest, Hybrid Monte Carlo method for condensed-matter systems, Phys. Rev. B **45**, 679 (1992).

[22] R. Iftimie, D. Salahub, D. Q. Wei, and J. Schofield, Using a classical potential as an efficient importance function for sampling from an ab initio potential, J. Chem. Phys. **113**, 4852 (2000).

[23] L. D. Gelb, Monte Carlo simulations using sampling from an approximate potential, J. Chem. Phys. **118**, 7747 (2003).

[24] A. Nakayama, T. Taketsugu, and M. Shiga, Speed-up of ab initio hybrid Monte Carlo and ab initio path integral hybrid Monte Carlo simulations by using an auxiliary potential energy surface, Chem. Lett. **38**, 976 (2009).

[25] K. Suzuki, M. Tachikawa, and M. Shiga, Efficient *ab initio* path integral hybrid Monte Carlo based on the fourth-order Trotter expansion: Application to fluoride ion-water cluster, J. Chem. Phys. **132**, 144108 (2010).

[26] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, Self-learning Monte Carlo method, Phys. Rev. B **95**, 041101(R) (2017).

[27] J. Liu, H. Shen, Y. Qi, Z. Y. Meng, and L. Fu, Self-learning Monte Carlo method and cumulative update in fermion systems, Phys. Rev. B **95**, 241104(R) (2017).

[28] X. Y. Xu, Y. Qi, J. Liu, L. Fu, and Z. Y. Meng, Self-learning quantum Monte Carlo method in interacting fermion systems, Phys. Rev. B **96**, 041119(R) (2017).

[29] C. Chen, X. Y. Xu, J. Liu, G. Batrouni, R. Scalettar, and Z. Y. Meng, Symmetry-enforced self-learning Monte Carlo method applied to the Holstein model, Phys. Rev. B **98**, 041102(R) (2018).

[30] Y. Nagai, H. Shen, Y. Qi, J. Liu, and L. Fu, Self-learning Monte Carlo method: Continuous-time algorithm, Phys. Rev. B **96**, 161102(R) (2017).

[31] Y. Nagai, M. Okumura, and A. Tanaka, Self-learning Monte Carlo method with Behler–Parrinello neural networks, Phys. Rev. B **101**, 115111 (2020).

[32] H. Shen, J. Liu, and L. Fu, Self-learning Monte Carlo with deep neural networks, Phys. Rev. B **97**, 205140 (2018).

[33] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.102.041124 for the technical details of the SLHMC.

[34] M. Shiga, PIMD: An open-source software for parallel molecular simulations, version 2.2.1 (2019), `https://ccse.jaea.go.jp/software/PIMD/index.en.html`.

[35] N. Artrith and A. Urban, An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO_2, Comput. Mater. Sci. **114**, 135-150 (2016).

[36] G. Kresse and J. Hafner, *Ab initio* molecular dynamics for liquid metals, Phys. Rev. B **47**, 558 (1993); *Ab initio* molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium, **49**, 14251 (1994); G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. **6**, 15 (1996); Efficient iterative

schemes for *ab initio* total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[37] H. Kimizuka, H. Kaburaki, and Y. Kogure, Molecular-dynamics study of the high-temperature elasticity of quartz above the α-β phase transition, Phys. Rev. B **67**, 024105 (2003).

[38] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[39] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994); G. Kresse and D. Joubert, From ultra-soft pseudopotentials to the projector augmented-wave method, *ibid.* **59**, 1758 (1999).

[40] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. **16**, 1190 (1995).

[41] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Trans. Math. Software **23**, 550 (1997).

[42] The ANN fingerprint parameters were set as radial_Rc = 8.0, radial_N = 16, angular_Rc = 6.5, and angular_N = 4 for all the atomic elements. We used two hidden layers with 15 units and the tanh-type activation function.

[43] R. J. Cava, H. Takagi, H. W. Zandbergen, J. J. Krajewski, W. F. Peck Jr., T. Siegrist, B. Batlogg, R. B. van Dover, R. J. Felder, K. Mizuhashi, J. O. Lee, H. Eisaki, and S. Uchida, Superconductivity in the quaternary intermetallic compounds $LnNi_2B_2C$, Nature **367**, 252 (1994).

[44] Y. Nagai, Y. Kato, N. Hayashi, K. Yamauchi, and H. Harima, Calculated positions of point nodes in the gap structure of the borocarbide superconductor YNi_2B_2C, Phys. Rev. B **76**, 214514 (2007).

[45] T. Park, M. B. Salamon, E. M. Choi, H. J. Kim, and S. I. Lee, Evidence for Nodal Quasiparticles in the Nonmagnetic Superconductor YNi_2B_2C via Field-Angle-Dependent Heat Capacity, Phys. Rev. Lett. **90**, 177001 (2003).

[46] K. Izawa, K. Kamata, Y. Nakajima, Y. Matsuda, T. Watanabe, M. Nohara, H. Takagi, P. Thalmeier, and K. Maki, Gap Function with Point Nodes in Borocarbide Superconductor YNi_2B_2C, Phys. Rev. Lett. **89**, 137006 (2002).

[47] F. Weber, L. Pintschovius, W. Reichardt, R. Heid, K.-P. Bohnen, A. Kreyssig, D. Reznik, and K. Hradil, Phonons and electron-phonon coupling in YNi_2B_2C, Phys. Rev. B **89**, 104503 (2014).

[48] L. Liu, X. Wu, W. Xu, R. Wang, C. Shu, and D. Liu, Phonon and thermodynamic properties in YNi_2B_2C and $LuNi_2B_2C$ from first-principles study, J. Supercond. Novel Magn. **31**, 1925 (2018).

[49] S.-T. Lin, M. Blanco, and W. A. Goddard III, The two-phase model for calculating thermodynamic properties of liquids from molecular dynamics: Validation for the phase diagram of Lennard-Jones fluids, J. Chem. Phys. **119**, 11792 (2003).

[50] A. Pérez, M. E. Tuckerman, and M. H. Müser, A comparative study of the centroid and ring-polymer molecular dynamics methods for approximating quantum time correlation functions from path integrals, J. Chem. Phys. **130**, 184105 (2009).

[51] M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford Univ Press, Oxford, 2010).