ILL@UMKC

The document below may be protected by U.S. Copyright Law.

Please report problems with document quality, missing pages, etc. immediately to the Interlibrary Loan office by replying to the email notification you received.

Descriptors representing two- and three-body atomic distributions and their effects on the accuracy of machine-learned inter-atomic potentials ⊘

Ryosuke Jinnouchi; Ferenc Karsai; Carla Verdi; ... et. al

Check for updates

🌐 View Online

Export Citation

CrossMark

Articles You May Be Interested In

Atomistic structure search using local surrogate model

J. Chem. Phys. (November 2022)

Design and performance of the SLD vertex detector, a 120 Mpixel tracking system

AIP Conference Proceedings (February 1992)

Neural network potential from bispectrum components: A case study on crystalline silicon

J. Chem. Phys. (August 2020)

The Journal of Chemical Physics

Special Topic: Adhesion and Friction

Submit Today!

AIP Publishing

AIP Publishing

Descriptors representing two- and three-body atomic distributions and their effects on the accuracy of machine-learned inter-atomic potentials

View Online   Export Citation   CrossMark

Ryosuke Jinnouchi,[1,a] Ferenc Karsai,[2] Carla Verdi,[1] Ryoji Asahi,[3] and Georg Kresse[1,b]

AFFILIATIONS

[1] Computational Materials Physics, Faculty of Physics, University of Vienna, Sensengasse 8/16, 1090 Vienna, Austria
[2] VASP Software GmbH, Sensengasse 8, 1090 Vienna, Austria
[3] Toyota Central Research and Developments Laboratories, Inc., Aichi 480-1192, Japan

Note: This paper is part of the JCP Special Topic on Machine Learning Meets Chemical Physics.
[a] **Also at:** Toyota Central Research and Developments Laboratories, Inc., Aichi, Japan.
Author to whom correspondence should be addressed: e1262@mosk.tytlabs.co.jp
[b] **Also at:** VASP Software GmbH, Vienna, Austria.

ABSTRACT

When determining machine-learning models for inter-atomic potentials, the potential energy surface is often described as a non-linear function of descriptors representing two- and three-body atomic distribution functions. It is not obvious how the choice of the descriptors affects the efficiency of the training and the accuracy of the final machine-learned model. In this work, we formulate an efficient method to calculate descriptors that can separately represent two- and three-body atomic distribution functions, and we examine the effects of including only two- or three-body descriptors, as well as including both, in the regression model. Our study indicates that non-linear mixing of two- and three-body descriptors is essential for an efficient training and a high accuracy of the final machine-learned model. The efficiency can be further improved by weighting the two-body descriptors more strongly. We furthermore examine a sparsification of the three-body descriptors. The three-body descriptors usually provide redundant representations of the atomistic structure, and the number of descriptors can be significantly reduced without loss of accuracy by applying an automatic sparsification using a principal component analysis. Visualization of the reduced descriptors using three-body distribution functions in real-space indicates that the sparsification automatically removes the components that are less significant for describing the distribution function.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0009491

I. INTRODUCTION

Machine-learning force fields (MLFFs) are an emerging powerful method to accurately and efficiently predict inter-atomic potentials.[1–15] In the MLFFs, the potential energy is described as a function of descriptors representing the atomic structure of a material, and the parameters in the function are optimized to reproduce first principles (FP) training data comprised of total energies, forces, and stress tensor components. The constructed MLFFs can be used in molecular dynamics (MD) and Monte Carlo simulations, and the simulations can efficiently provide quantitative finite-temperature properties of materials with almost FP accuracy. A particular strength of this method is that significantly less human intervention is required in the force field (FF) construction compared to conventional classical FFs. Thanks to all these features, the method has been successfully applied to predict static and dynamical properties of gaseous molecules, liquids, solids, surfaces, and interfaces.[16–33]

For an efficient and accurate MLFF generation, a wide variety of descriptors and potential energy functions have been proposed.[1,7,13,14,34–40] Although the formulations differ significantly, many of them describe the potential energy surface as a non-linear

function of descriptors representing two- and three-body atomic distribution functions, i.e., radial and angular distribution functions, respectively. The restriction to two and three-body terms is partly justified by a study of Glielmo et al.[13] Glielmo et al. showed that a non-linear function of two- and three-body descriptors can generate many-body descriptors necessary for describing higher order atomic interactions. In practical applications, efficiency and accuracy depend on the type of functions used. Generally, non-linear functions of higher order can more accurately describe inter-atomic potential energy surfaces but require more training data.[13,35] Hence, one usually needs to compromise on either accuracy or efficiency. Various types of functions, such as activation functions in neural networks[3] and Gaussian or polynomial kernels[13,14,37,41] in kernel ridge regressions, have been suggested. However, little is known on general guidelines for choosing the functional form. A similar trade-off exists for the type of descriptors and the number of descriptors. The descriptors should have the property to compactly represent structural features that strongly influence the inter-atomic potentials. To this end, Willatt and co-workers[37] have proposed radially scaled kernels, where atomic distributions near the center atom are more strongly weighted so that the descriptors put more emphasis on local features in close proximity to the central atom. Furthermore, the descriptors need to be efficiently computable to allow for fast training and applications of the machine learned potentials. For better computational performance, Caro[39] suggested an efficient method to calculate the analytic form of descriptors where the radial and angular channels in the atomic distributions are approximately separated. Given a certain activation function or kernel, adding more descriptors can lead to higher accuracy, but again at the expense of requiring more training data. A solution to this problem in the form of sparsification has been recently proposed by Imbalzano et al.[42] Sparsification is promising for an automatic selection of a small number of descriptors from a large pool of candidates relying on correlations that are intrinsic to the training data. However, the correlations between the selected (or deleted) descriptors and the training data are often difficult to understand, although such intuitive information is useful to construct better descriptors and optimize the number of descriptors from the outset.

In this study, we construct descriptors that are clearly separable into two- and three-body distribution functions. We examine how two- and three-body descriptors alone can reproduce the energies and forces and give guidelines on how to include both. We also investigate the effects of the sparsification of the descriptors. In particular, by visualizing the sparsified descriptors as real-space distributions, we clarify the role of sparsification. In Sec. II, the potential energy surface is formulated as a function of the two- and three-body distribution functions. Details of the sparsification and training are also included in this section. Results and discussions are presented in Sec. III. The paper is finally concluded in Sec. IV.

II. METHOD

A. Two- and three-body distribution functions

Similar to the smooth overlap of atomic positions (SOAP) formulation proposed by Bartók et al.,[34] our descriptors set out from a smoothed three-dimensional atomic distribution function ρ_i around

atom i defined as

$$\tilde{\rho}_{ij}(\mathbf{r}) = f_{\text{cut}}(|\mathbf{r}_j - \mathbf{r}_i|)g(\mathbf{r} - (\mathbf{r}_j - \mathbf{r}_i)),$$

$$\rho_i(\mathbf{r}) = \sum_{j \neq i}^{N_a} \tilde{\rho}_{ij}(\mathbf{r}). \tag{1}$$

Here, $\tilde{\rho}_{ij}(\mathbf{r})$ describes the likelihood to find atom j at position $\mathbf{r}$ relative to atom i, N_a is the number of atoms in a certain cutoff sphere R_{cut}, $\mathbf{r}_i$ denotes the position of atom i, g is a smoothed δ-function, and f_{cut} is a cutoff function that smoothly eliminates the contribution from atoms outside a given cutoff radius R_{cut}. We note on passing that this distribution function obviously depends explicitly on the cutoff radius R_{cut} and thus implicitly on the number of atoms N_a in the sphere. The two-body distribution function $\rho_i^{(2)}$ is defined as the probability to find an atom $j(j \neq i)$ at a distance r from atom i, that is,

$$\rho_i^{(2)}(r) = \frac{1}{4\pi} \int \rho_i(r\hat{\mathbf{r}})d\hat{\mathbf{r}}. \tag{2}$$

The three-body distribution function $\rho_i^{(3)}$ is defined as the probability to find an atom $j(j \neq i)$ at a distance r from atom i and another atom $k(k \neq i,j)$ at a distance s from atom i spanning the angle $\angle kij = \theta$ between them. It can be calculated as

$$\rho_i^{(3)}(r,s,\theta) = \iint d\hat{\mathbf{r}}d\hat{\mathbf{s}}\, \delta(\hat{\mathbf{r}} \cdot \hat{\mathbf{s}} - \cos\theta) \sum_{j\neq i}^{N_a}\sum_{k\neq i,j}^{N_a} \tilde{\rho}_{ij}(r\hat{\mathbf{r}})\tilde{\rho}_{ik}(s\hat{\mathbf{s}}), \tag{3}$$

$$= \iint d\hat{\mathbf{r}}d\hat{\mathbf{s}}\, \delta(\hat{\mathbf{r}} \cdot \hat{\mathbf{s}} - \cos\theta)$$
$$\times \left[\rho_i(r\hat{\mathbf{r}})\rho_i^*(s\hat{\mathbf{s}}) - \sum_{j\neq i}^{N_a} \tilde{\rho}_{ij}(r\hat{\mathbf{r}})\tilde{\rho}_{ij}(s\hat{\mathbf{s}}) \right]. \tag{4}$$

In the second line, we have subtracted the self-term $j = k$, which in practice greatly reduces the computational cost, since this formulation avoids a double summation over the indices j and k. The important difference from the power spectrum method of Bartók and Csányi et al.[14,34] is obviously in the second term in the square bracket in Eq. (4). If we replace the smoothing function g in Eq. (1) by the δ-function and disregard the cutoff function, one can easily show that this term is equivalent to

$$\sum_{j\neq i}^{N_a} \tilde{\rho}_{ij}(r\hat{\mathbf{r}})\tilde{\rho}_{ij}(s\hat{\mathbf{s}}) = \rho_i(r\hat{\mathbf{r}})\delta(r\hat{\mathbf{r}} - s\hat{\mathbf{s}}). \tag{5}$$

This means that in real space, the "diagonal," $\theta = 0$ and $r = s$, of the three-body term in the power spectrum corresponds to the two-body correlation function. By eliminating this term, the three-body distribution function becomes free of two-body components, and this enables us to separately control the importance of the two- and three-body descriptors.

In practical calculations, we do not use the identity derived in Eq. (5) since it does not hold when a cutoff function or a smoothening is introduced. Instead, our descriptors are formulated as expansion coefficients c_n^i and p_{nvl}^i of the two- and three-body distribution functions with respect to orthonormal radial and angular basis sets

χ_{nl} and P_l,

$$\rho_i^{(2)}(r) = \frac{1}{\sqrt{4\pi}} \sum_{n=1}^{N_R^0} \check{c}_n^i \chi_{n0}(r), \qquad (6)$$

$$\rho_i^{(3)}(r, s, \theta) = \sum_{l=0}^{L_{max}} \sum_{n=1}^{N_R^l} \sum_{v=1}^{N_R^l} \sqrt{\frac{2l+1}{2}} \, p_{nvl}^i \chi_{nl}(r) \chi_{vl}(s) P_l(\cos\theta). \qquad (7)$$

We note that the radial basis functions χ_{nl} depend parametrically on the angular quantum number l here as well as in Eq. (10). For χ_{nl} and P_l, normalized spherical Bessel functions $\chi_{nl} = j_l(q_n r)$ and Legendre polynomials of order l are used in this work, respectively (for details, we refer to Ref. 14). The expansion coefficients can be converted to an efficiently computable form as

$$\check{c}_n^i = c_{n00}^i, \qquad (8)$$

$$p_{nvl}^i = \sqrt{\frac{8\pi^2}{2l+1}} \sum_{m=-l}^{l} \left[c_{nlm}^i c_{vlm}^{i*} - \sum_{j\neq i}^{N_a} \tilde{c}_{nlm}^{ij} \tilde{c}_{vlm}^{ij*} \right], \qquad (9)$$

where $\tilde{c}_{nlm}^{ij}$ denotes the expansion coefficient of the distribution $\tilde{\rho}_{ij}(\mathbf{r})$ with respect to the radial basis sets χ_{nl} and spherical harmonics Y_{lm},

$$\tilde{\rho}_{ij}(\mathbf{r}) = \sum_{l=0}^{L_{max}} \sum_{m=-l}^{l} \sum_{n=1}^{N_R^l} \tilde{c}_{nlm}^{ij} \chi_{nl}(r) Y_{lm}(\hat{\mathbf{r}}) \qquad (10)$$

and

$$c_{nlm}^i = \sum_{j\neq i}^{N_a} \tilde{c}_{nlm}^{ij}. \qquad (11)$$

Similar to Eq. (4), the main difference to the power spectrum is in the second term in the bracket of Eq. (9). The present three-body descriptor p_{nvl}^i cannot be expressed by linear combinations of the power spectrum components. We note, however, that the term adds extra compute cost of the order of 10%, since in the power spectrum, the coefficients $\tilde{c}_{nlm}^{ij}$ are not stored or manipulated but immediately contracted to obtain c_{nlm}^i. The advantage of the present approach is that we can take direct control of the two-body correlation functions and tune their weight to obtain faster convergence with respect to the number of configurations used for training, as we will show in Sec. III A. The separation between two and three-body descriptors has a long history in ML and goes back to the first work on ML potentials published by Behler and Parrinello.[1]

Figure 1 shows the three-body distribution function of liquid Si at 2000 K reconstructed from the three-body descriptors with (a) and without (b) the self-overlapping terms. Figure 1(a) illustrates that the power spectrum can be visualized as a more intuitively understandable real-space distribution function through Eq. (7). Figure 1(b) indicates that the self-overlapping peak in Fig. 1(a) shown as a white-colored region is correctly removed by eliminating the self-overlapping terms in Eq. (4).
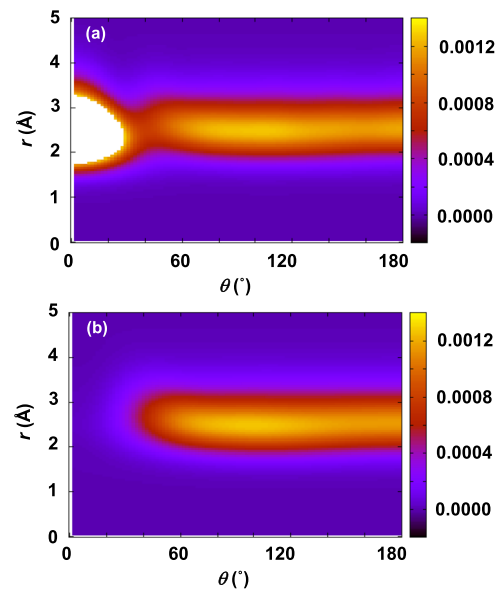
FIG. 1. Three-body distribution function of liquid Si at 2000 K reconstructed from the three-body descriptors p_{nvl}^i with (a) and without (b) the self-overlapping term. The distributions were averaged over all local reference configurations. For this plot, s was set to $s = r$.

B. Potential energy functional

Similar to previous machine-learning approaches,[1,2] the potential energy U of a structure with N_a atoms is approximated as a summation of local energies U_i written as

$$U = \sum_{i=1}^{N_a} U_i. \qquad (12)$$

The local energy U_i of atom i is described as a functional of the two- and three-body distribution functions,

$$U_i = F\left[\rho_i^{(2)}, \rho_i^{(3)}\right]. \qquad (13)$$

In practice, the distribution functions $\rho_i^{(2)}$ and $\rho_i^{(3)}$ are discretized as explained in Sec. II A and represented by a descriptor vector $\mathbf{x}_i$, which collects all two- and three-body coefficients,

$$\mathbf{x}_i^{(2)T} = (c_1^i, c_2^i, \ldots), \qquad (14)$$

$$\mathbf{x}_i^{(3)T} = (p_{110}^i, p_{111}^i, \ldots, p_{120}^i, p_{121}^i, \ldots), \qquad (15)$$

$$\mathbf{x}_i^T = \left(\mathbf{x}_i^{(2)T}, \mathbf{x}_i^{(3)T} \right), \qquad (16)$$

into a single vector $\mathbf{x}_i$ (the superscript T means the transpose of the vector). Thus,

$$U_i = F\left[\rho_i^{(2)}, \rho_i^{(3)}\right] \rightarrow F(\mathbf{x}_i).$$

Following the Gaussian approximation potential pioneered by Bartók and co-workers,[2] in this work, the function F is approximated as a weighted sum of functions $K(\mathbf{x}_i, \mathbf{x}_{i_B})$ centered at reference

points $\{\mathbf{x}_{i_B}|i_B = 1,\ldots,N_B\}$,

$$F(\mathbf{x}_i) = \sum_{i_B=1}^{N_B} w_{i_B} K(\mathbf{x}_i, \mathbf{x}_{i_B}). \tag{17}$$

The reference configurations ρ_{i_B} ($\mathbf{x}_{i_B}$ after discretization) describe the local distribution of atoms around central atoms i_B. They are chosen from a set of reference structures picked from trajectories generated by FPMD simulations. The coefficients $\{w_{i_B}|i_B = 1,\ldots,N_B\}$ are optimized to best reproduce the FP energies, forces, and stress tensor components as obtained by the FPMD simulations. The function K is supposed to measure the similarity between a local configuration $\mathbf{x}_i$ of interest and the reference configuration $\mathbf{x}_{i_B}$. It usually approaches unity if two configurations are similar, and it decays toward a small value if the two configurations are different. A sharply peaked functional form for K allows for more accurate predictions, while it requires more reference points i_B to allow for a smooth interpolation among the reference configurations $\mathbf{x}_{i_B}$.

In this study, we adopt two functional forms for K. In the first case, we define the kernel to depend non-linearly on the two- and three-body contributions,

$$\mathbf{x}_i^T \rightarrow \left(\sqrt{\beta^{(2)}}\mathbf{x}_i^{(2)T}, \sqrt{\beta^{(3)}}\mathbf{x}_i^{(3)T}\right), \tag{18}$$

$$K(\mathbf{x}_i, \mathbf{x}_{i_B}) = (\hat{\mathbf{x}}_i \cdot \hat{\mathbf{x}}_{i_B})^\zeta. \tag{19}$$

The hat symbol $\hat{\mathbf{x}}_i$ denotes a normalized vector of $\mathbf{x}_i$. Without loss of generality, one can set $\beta^{(2)} = 1 - \beta^{(3)}$. Equation (18) implies that the inner product $\mathbf{x}_i \cdot \mathbf{x}_{i_B}$ is proportional to a sum of the two weighted inner products,

$$\mathbf{x}_i \cdot \mathbf{x}_{i_B} = \beta^{(2)}\mathbf{x}_i^{(2)} \cdot \mathbf{x}_{i_B}^{(2)} + \beta^{(3)}\mathbf{x}_i^{(3)} \cdot \mathbf{x}_{i_B}^{(3)}, \tag{20}$$

and hence, $\beta^{(2)}$ and $\beta^{(3)}$ can be viewed as a metric in the super-space. The normalization and exponentiation in Eq. (19), however, introduce non-linear terms that mix two- and three-body contributions. It is fairly straightforward to show that in the limit of an infinite basis set, each inner product corresponds to an overlap integral of the original distribution functions as follows:

$$\mathbf{x}_i^{(2)} \cdot \mathbf{x}_{i_B}^{(2)} = 4\pi \int_0^\infty \rho_i^{(2)}(r)\rho_{i_B}^{(2)}(r)r^2 dr, \tag{21}$$

$$\mathbf{x}_i^{(3)} \cdot \mathbf{x}_{i_B}^{(3)} = (4\pi)^2 \int_0^\infty \int_0^\infty \int_0^\pi \rho_i^{(3)}(r,s,\theta)\rho_{i_B}^{(3)}(r,s,\theta)$$
$$\times r^2 s^2 \sin\theta \, drdsd\theta. \tag{22}$$

When the power spectrum instead of the supervector $\mathbf{x}_i$ is used as the descriptor, Eq. (19) turns to the functional form commonly termed smooth overlap of atomic positions (SOAP) proposed by Bartók et al.[34] Equations (18) and (19) differ from the original SOAP by two points. First, SOAP non-linearly mixes the power spectrum that involves two-body descriptors as squared coefficients $\sum_j \tilde{c}_{nlm}^{ij}\tilde{c}_{vlm}^{ij*}$, while Eq. (19) mixes the three-body descriptors with non-squared two-body radial descriptors c_n^i. Second, the parameters $\beta^{(2)}$ and $\beta^{(3)}$ in Eq. (18) allow us to control the weights of the two- and three-body terms unlike in the original SOAP.

The second functional form K considered here is defined as follows:

$$K(\mathbf{x}_i, \mathbf{x}_{i_B}) = \beta^{(2)}\left(\hat{\mathbf{x}}_i^{(2)} \cdot \hat{\mathbf{x}}_{i_B}^{(2)}\right)^{\zeta^{(2)}} + \beta^{(3)}\left(\hat{\mathbf{x}}_i^{(3)} \cdot \hat{\mathbf{x}}_{i_B}^{(3)}\right)^{\zeta^{(3)}}. \tag{23}$$

This function comprises the same overlap integrals. An essential difference from Eq. (19) and SOAP is that this function does not allow for non-linear terms involving the two- and three-body descriptors. The comparison between the two functions in Eqs. (19) and (23) enables us to examine how the cross terms between two- and three-body descriptors affect the learning efficiency and accuracy.

C. Sparsification of three-body descriptors

The number of three-body descriptors defined in Eq. (9) is, in practice, often much larger than the number of degrees of freedom of the atoms inside the sphere with radius R_{cut}. This means that the three-body descriptors are likely to generate redundant linearly dependent descriptors. To reduce the number of three-body descriptors, sparsification is applied. In this work, the CUR algorithm[43] is adopted as described below.

The method is similar to the algorithm that is used to reduce the number of local reference configurations N_B in our on-the-fly force field generation scheme.[14] The sparsification of the descriptors starts from the construction of an $N_D \times N_D$ square matrix $\mathbf{A}$ defined as

$$\mathbf{A} = \mathbf{X}\mathbf{X}^T, \tag{24}$$

where N_D denotes the number of three-body descriptors given by $N_D = \sum_{l=0}^{L_{\max}} N_R^l(N_R^l + 1)(L_{\max} + 1)/2$. In this number, the symmetry of the descriptors, $p_{nvl} = p_{vnl}$, has been taken into account. $\mathbf{X}$ is an $N_D \times N_B$ matrix containing the vectors $(\mathbf{x}_1^{(3)},\ldots,\mathbf{x}_{N_B}^{(3)})$. The matrix $\mathbf{A}$ is diagonalized as

$$\mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{L} = \text{diag}(l_1,\ldots,l_{N_D}), \tag{25}$$

where $\mathbf{U}$ is the matrix of eigenvectors defined as

$$\mathbf{U} = (\mathbf{u}_1,\ldots,\mathbf{u}_{N_D}), \tag{26}$$

$$\mathbf{u}_j^T = (u_{1j},\ldots,u_{N_Dj}). \tag{27}$$

By using the notations in Eqs. (26) and (27), Eq. (25) can be rewritten as

$$\mathbf{a}_j = \sum_{\xi=1}^{N_D} (u_{j\xi}l_\xi)\mathbf{u}_\xi, \tag{28}$$

where $\mathbf{a}_j$ denotes the jth column vector of the matrix $\mathbf{A}$. As in the original CUR algorithm,[43] the columns of the matrix $\mathbf{A}$ are kept when they are strongly correlated with the k eigenvectors $\mathbf{u}_\xi$ with the largest eigenvalues l_ξ. The correlation is measured by the statistical leverage scoring determined for each column of $\mathbf{A}$ as

$$\omega_j = \frac{1}{k}\sum_{\xi=1}^k u_{j\xi}^2. \tag{29}$$

TABLE I. Parameters for the descriptors and the basis functions. The parameters for the two-body descriptors are the same as those for the three-body descriptors. The parameter σ_{atom} is the width of the Gaussian function used to smoothen the δ-function (see details of the definition in Ref. [14]). Units of σ_{atom} and R_{cut} are Å.

σ_{atom}	0.5	Z	4	N_R^3 and N_R^4	7	N_R^{10}	6
R_{cut}	5	N_R^0	9	N_R^5 and N_R^6	6		
L_{max}	6 or 10	N_R^1 and N_R^2	8	N_R^7, N_R^8 and N_R^9	5		

In this study, k is set to 5. After the calculation of the leverage scoring, descriptors with the largest leverage scoring are selected so that the ratio of the selected descriptors to the total number of descriptors reaches a predetermined value x_{spars}.

D. Details of the training data and validation of MLFFs

MLFFs were generated for solid Si at 1600 K, liquid Si at 2000 K, solid MgO at 3000 K, and liquid MgO at 3000 K. Local reference configurations with a density distribution ρ_{i_B} were randomly selected from a trajectory of a 10 ps FPMD simulation on each material. The selected ρ_{i_B} were mapped to the descriptors $\mathbf{x}_{i_B}$. The number of local reference configurations N_B was varied from 5 to 1000 for Si and 5 to 4000 for MgO. The regression coefficients w_{i_B} were optimized to reproduce all FP energies, forces, and stress tensor components taken from the same FPMD simulations using Bayesian linear regression.[14] The accuracy of the generated MLFFs is examined by comparing their energies and forces to FP energies and forces taken from another 10 ps FPMD trajectory. For all materials, unit cells with 64 atoms were used. The FPMD calculations were executed in the isothermal and isobaric ensemble,[44,45] with the time step set to 3 fs for all materials. The Perdew–Burke–Ernzerhof functional[46] was used to describe the exchange–correlation interactions among the electrons, and a $2 \times 2 \times 2$ **k**-point mesh was used in all simulations. Plane wave basis sets and the projector augmented wave (PAW) method[47,48] were adopted in all FP calculations. The PAW atomic reference configuration was $3s^2 3p^2$ for Si, $3s^2 3p^0$ for Mg, and $2s^2 2p^2$ for O. The plane wave cutoff energy was set to 325 eV and 520 eV for Si and MgO, respectively. The parameters used for the MLFFs are tabulated in Table I. All FP calculations and MLFF generations were performed using the Vienna *Ab initio* Simulation Package (VASP).[49,50]

III. RESULTS AND DISCUSSION

A. Effects of two- and three-body descriptors

Figure 2 shows the root mean square differences (RMSDs shown as σ_E and σ_F) of the energies and forces between the MLFF and FP calculations as a function of the number of local reference configurations N_B. Results for the power spectrum (PS) of Bartók and Csányi[34] and results obtained using Eqs. (18) and (19) with $\beta^{(3)} = 0.0$ (two-body term only), 0.1, 0.5, 0.8, and 1.0 (three-body term only) are shown. For a concise comparison of the newly proposed descriptors to the conventional PS, we show differences in RMSDs as $\Delta\sigma_E$ and $\Delta\sigma_F$ in the same figure. Here, the negative values mean smaller RMDSs than those by PS.
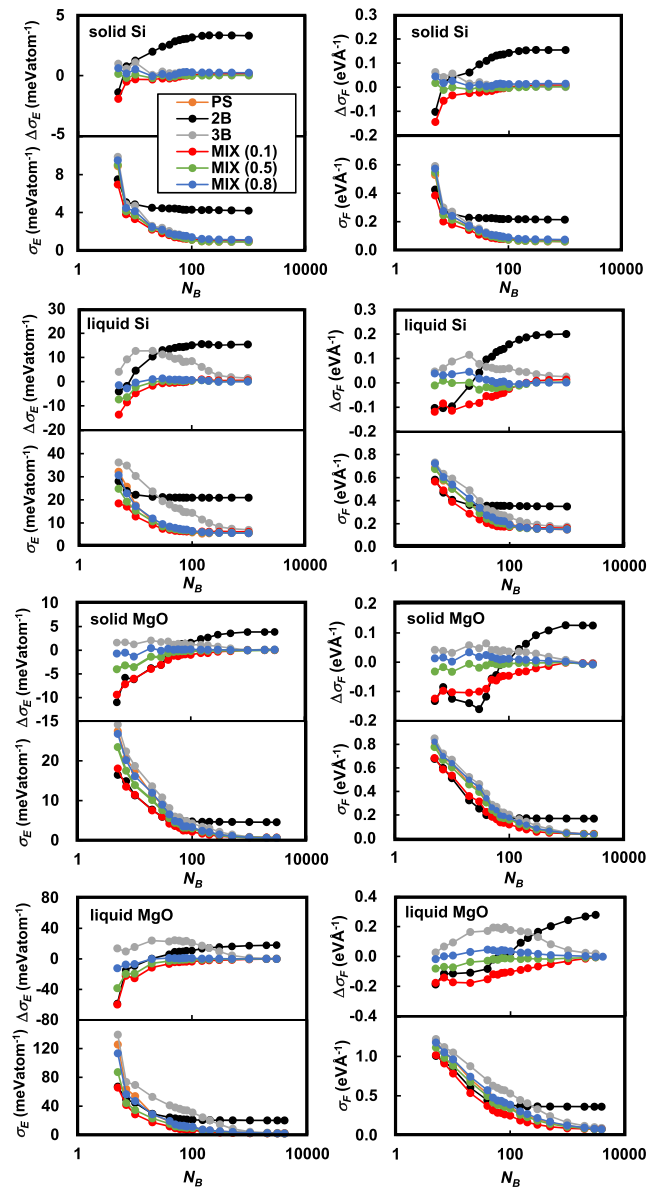
FIG. 2. Root mean square differences of the energies σ_E and forces σ_F between MLFF and FP calculations for solid and liquid Si and MgO as function of the number of local reference configurations N_B. Results for the power spectrum (PS) of Bartók et al.[34] and two-body (2B) and three-body (3B) descriptors are shown. For "MIX," Eq. (19) is used with $\beta^{(3)}$ = 0.1, 0.5, or 0.8 mixing the two- and three-body descriptors (smaller values of MIX correspond to a smaller weight of the three body descriptors). $\Delta\sigma_E$ and $\Delta\sigma_F$ indicate relative differences of the RMSDs from those obtained by PS. Negative values indicate smaller RMSDs than those for PS. In all calculations shown in this figure, L_{max} is set to 6.

Sparsification is not carried out in any of the calculations shown in this subsection. In both Si and MgO, the RMSDs for solids are smaller than those for liquids. The results indicate that, for solids, the MLFF can more easily reproduce the FP potential energy surfaces.

Although clear differences between solids and liquids are present, several common features are visible in both phases. When N_B is small, the two-body descriptors exhibit smaller RMSDs than the three-body descriptors, while the three-body descriptors generally lead to much smaller RMSDs at larger N_B than the two body descriptors. This trend is due to the higher dimensionality and flexibility of the three-body descriptors compared to the two-body ones. The number of variables for the three-body descriptors (281 for Si and 3372 for MgO at $L_{max} = 6$) is much larger than that for the two-body descriptors (9 and 27 for Si and MgO, respectively). Hence, using the three-body descriptors, one needs more local reference configurations for an accurate interpolation.

In both solids and liquids, the power spectrum (PS) and the non-linearly mixed descriptors (MIX = 0.1, 0.5 or 0.8) yield smaller RMSDs with less N_B than the three-body descriptors alone, although the number of descriptors is the same or larger than that of three-body descriptors. In addition, for the mixed descriptors, the initial RMSDs decrease when a smaller weight is used for the three-body descriptors $\beta^{(3)}$, although a too small weight of $\beta^{(3)} = 0.1$ causes a slight increase in the RMSDs at very large N_B. We explain this observation in the following way: by explicitly using two-body descriptors and weighting them strongly, we allow the machine learning algorithm to quickly model two-body interactions in the inter-atomic potentials. This conjecture is supported by the observation that the red curve initially follows closely the black curve for two-body descriptors only. As more configurations are added, the two-body descriptors are not sufficiently flexible and their accuracy quickly levels off. However, if three-body descriptors are also available, the MLFF is able to recover the accuracy of the power spectrum and, in fact, outperforms the PS except for very large number of configurations N_B, where the PS has a very slight advantage compared to the mixed descriptors. We have numerical evidence that this is related to the conditioning number of the problem being worse, if the three-body descriptors are weighted less strongly.

Finally, we examine the effects of the cross terms between two- and three-body descriptors generated by the non-linear mixing, Eqs. (18) and (19), by comparing it with the linear mixing given in Eq. (23). Figure 3 summarizes the results. The non-linear mixing is superior to the linear mixing. Particularly for liquids, the effect of the non-linear mixing is significant. The results indicate that the cross terms are essential for efficiently describing the potential energy surfaces.

The same observation can be made in the radial distribution functions (RDFs) of solid and liquid Si summarized in Fig. 4. For solid Si, MLFFs including those without the cross terms accurately reproduce the FP result even with $N_B = 100$ except for 2B. In contrast, for liquid Si, MLFFs without the cross-terms yield fatal errors in the RDFs if N_B is set to 100. These descriptors need more local reference configurations to reproduce the FP result as illustrated by the result for $N_B = 1000$.

B. Sparsification of descriptors and its visualization

The number of three-body descriptors (281 for Si and 3372 for MgO at $L_{max} = 6$) is much larger than the degrees of freedom of atoms inside the sphere within the cutoff radius (roughly 90 and 120 for Si and MgO, respectively). This means that the
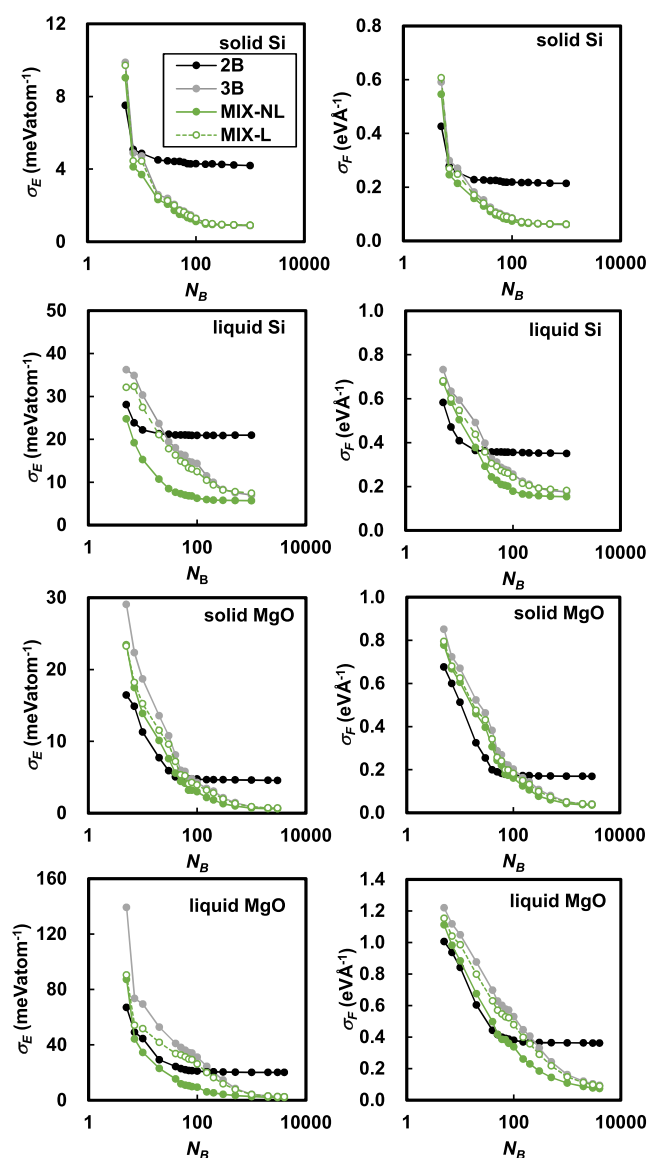
FIG. 3. Root mean square differences between MLFF and FP calculations using non-linear mixing of the two- and three-body descriptors [MIX-NL defined as Eqs. (18) and (19)] or linear mixing [MIX-L defined as Eq. (23)]. For comparison, errors for the two-body (2B) and three-body descriptors (3B) are also reproduced. In all calculations shown in this figure, L_{max} is set to 6. $\beta^{(3)}$ is set to 0.5 for both linearly and non-linearly mixed descriptors.

descriptors are likely redundant. Hence, we have tried to reduce the three-body descriptors using the CUR decomposition method introduced in Sec. II C. The effects of the sparsification are examined in this section. Figure 5 shows the RMSDs for energies and forces as functions of the degree of sparsification x_{spars}, which is defined as the ratio of the number of selected descriptors to the total number before the sparsification. For comparison, results obtained by manually reducing the descriptors by decreasing the maximum
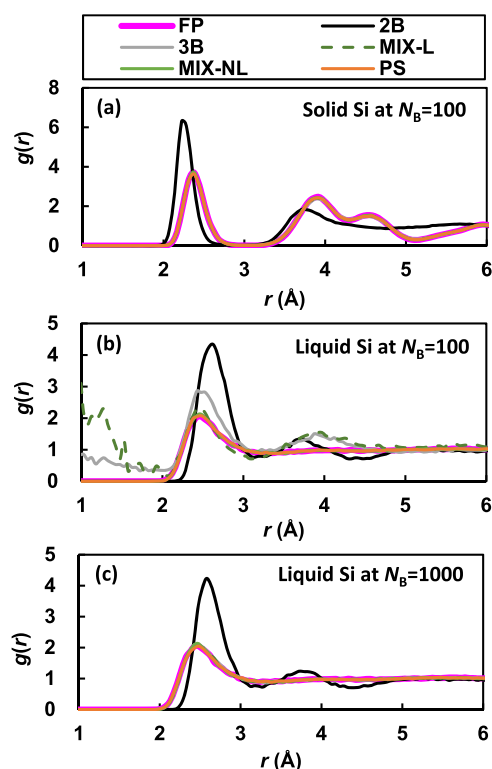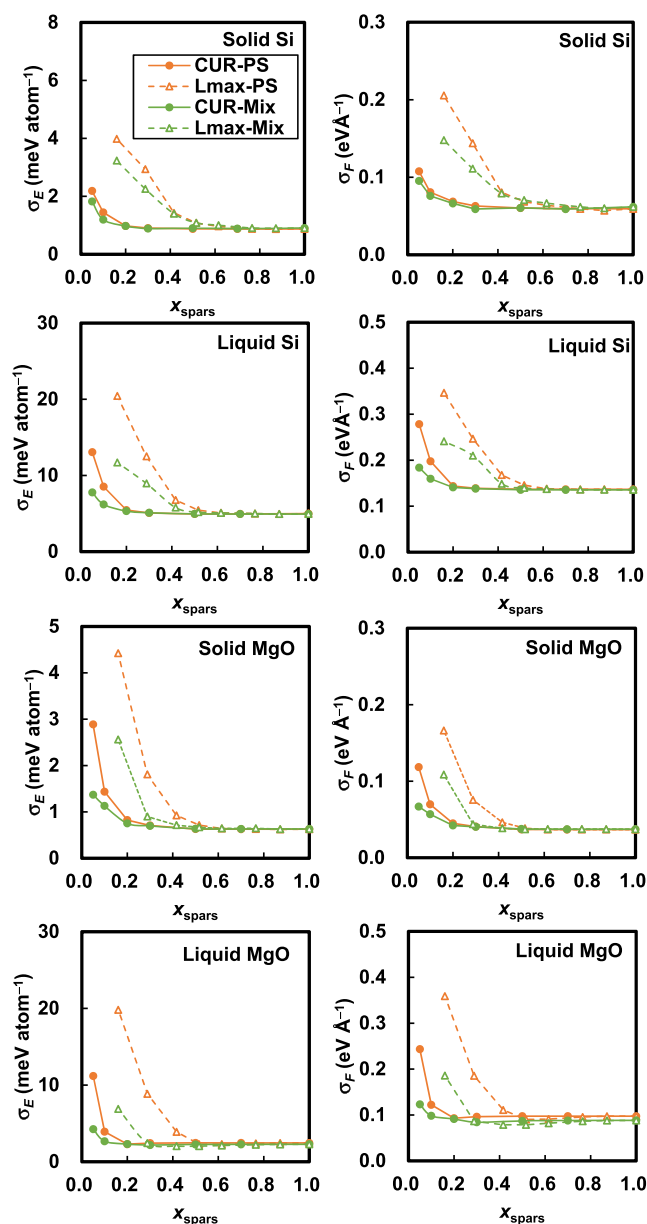
FIG. 4. Radial distribution functions (RDFs) $g(r)$ of solid Si at 1500 K and liquid Si at 2000 K. The machine learned models in (a) and (b) are generated using 100 local reference configurations, and the models in (c) are generated using 1000 local reference configurations. In (a) and (c), the RDFs generated by the machine-learned models are almost identical to the FP result (except for 2B). In (b), the RDFs are indistinguishable from the FP result only for MIX-NL and PS. $\beta^{(3)}$ is set to 0.5 for both linearly and non-linearly mixed descriptors. All RDFs are calculated by isobaric MD simulations for 20 ps.


FIG. 5. Root mean square differences of energies σ_E and forces σ_F for liquid and solid Si and MgO as a function of the ratio of descriptors retained (x_{spars}). Solid lines with filled circles and dashed lines with open triangles show the results obtained via CUR decomposition and manual reduction of L_{max}, respectively. Orange and green lines refer to the results obtained by using the power spectrum (PS) and the mixed descriptors (MIX), respectively. The CUR algorithm sparsifies the descriptors with L_{max} = 10. In the manual reduction, L_{max} is reduced from 10 to 0.

angular momentum number L_{max} are also shown in the same figure. The CUR decomposition method allows us to reduce the number of descriptors to 30% without significantly lowering the accuracy, while manually reducing L_{max} shows adverse effects already when the number of descriptors falls to 40%–60% of the original number. It should be also noted that mixed descriptors can be more efficiently sparsified than the power spectrum of Bartók and Csányi. This behavior is likely related to the fact that the energetically dominant two-body descriptors are fully retained in the case of the mixed descriptors, as sparsification is only performed on the three-body descriptors.

To gain a better understanding of the role of the CUR decomposition method, the three-body distribution functions for liquid Si are reconstructed from the sparsified three-body descriptors and are visualized in Fig. 6. The CUR decomposition algorithm more accurately retains the structure of the three-body distribution function even when it reduces the number of descriptors to 30%. For x_{spars} < 0.3, the reconstructed distribution function largely overshoots or undershoots because of deficiencies of the basis sets as indicated by the white-colored region in x_{spars} = 0.1, where the distribution

significantly undershoots yielding unphysical negative values. The results indicate that the CUR algorithm automatically removes less significant components in the representation of the real-space distribution functions.
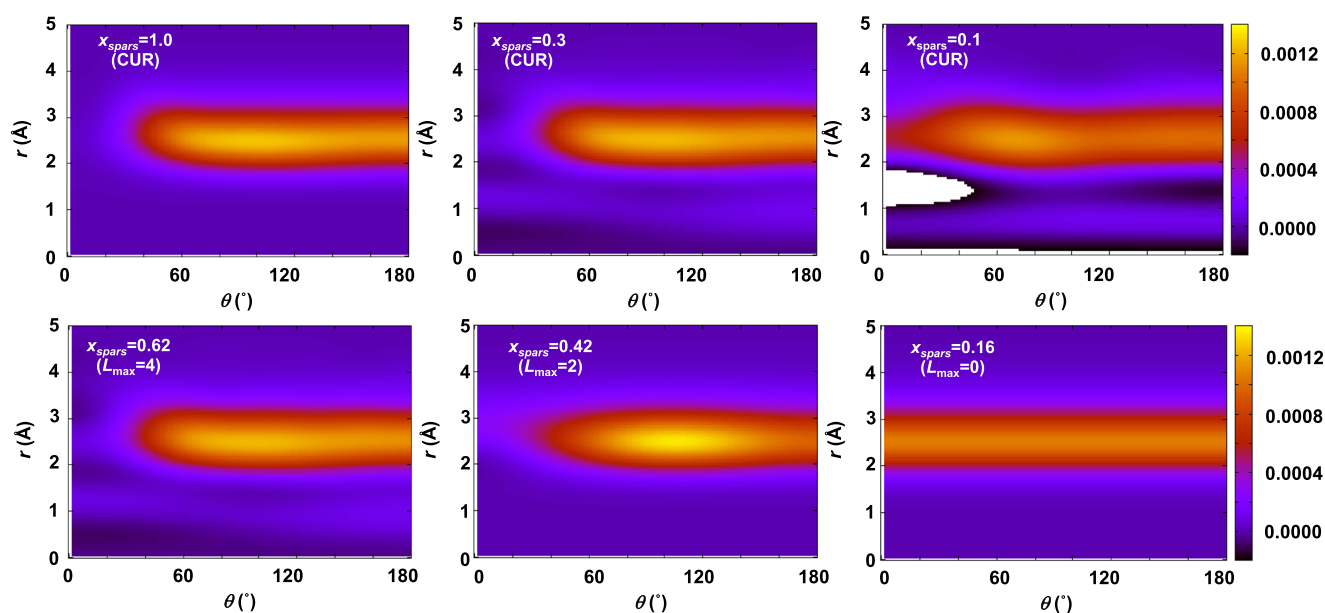
FIG. 6. Three-body distribution functions of liquid Si at 2000 K reconstructed from the sparsified three-body descriptors. Upper and lower figures show the results by the CUR algorithm and manual reduction of L_{max}, respectively.

IV. CONCLUSION

We have formulated descriptors representing the two- and three-body atomic distribution functions in a separable form and examined the effects of each descriptor on the accuracy and efficiency in learning of FP potential energy surfaces. Our results clearly indicate that non-linear mixing of these two descriptors is essential for an accurate and efficient description of the potential energy surfaces and that the learning efficiency can be improved by increasing the weight of the two-body descriptors. We have also examined the effects of reducing the number of three-body descriptors. A sparsification based on the CUR decomposition method automatically selects significant components for the representation of the three-body distribution functions and dramatically reduces the number of descriptors without losing accuracy.

All in all, the present method possesses several advantages compared to the power spectrum method of Bartók and Csányi. The learning efficiency is improved at the beginning, since one can weigh the two-body terms more strongly than the three-body terms. The improved learning efficiency is maintained all the way up to very many training data, although admittedly for very large training sets, the advantages compared to the power spectrum are diminishingly small. Finally, sparsification using the CUR algorithm is also somewhat more efficient if performed exclusively on the three-body descriptors.

ACKNOWLEDGMENTS

DATA AVAILABILITY

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

REFERENCES

[1] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Phys. Rev. Lett. **98**, 146401 (2007).

[2] 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).

[3] J. Behler, "Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations," Phys. Chem. Chem. Phys. **13**, 17930–17955 (2011).

[4] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," Phys. Rev. Lett. **108**, 058301 (2012).

[5] Z. Li, J. R. Kermode, and A. De Vita, "Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces," Phys. Rev. Lett. **114**, 096405 (2015).

[6] K. Miwa and H. Ohno, "Molecular dynamics study on β-phase vanadium monohydride with machine learning potential," Phys. Rev. B **94**, 184109 (2016).

[7] A. V. Shapeev, "Moment tensor potentials: A class of systematically improvable interatomic potentials," Multiscale Model. Simul. **14**, 1153–1173 (2016).

[8] J. Behler, "First principles neural network potentials for reactive simulations of large molecular and condensed systems," Angew. Chem., Int. Ed. **56**, 12828–12840 (2017).

[9] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields," Sci. Adv. **3**, e1603015 (2017).

[10] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, "Machine learning unifies the modeling of materials and molecules," Sci. Adv. **3**, e1701816 (2017).

[11]V. Botu, R. Batra, J. Chapman, and R. Ramprasad, "Machine learning force fields: Construction, validation, and outlook," J. Phys. Chem. C **121**, 511–522 (2017).

[12]T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, and R. Ramprasad, "A universal strategy for the creation of machine learning-based atomistic force fields," npj Comput. Mater. **3**, 37 (2017).

[13]A. Glielmo, C. Zeni, and A. De Vita, "Efficient nonparametric n-body force fields from machine learning," Phys. Rev. B **97**, 184307 (2018).

[14]R. Jinnouchi, F. Karsai, and G. Kresse, "On-the-fly machine learning force field generation: Application to melting points," Phys. Rev. B **100**, 014105 (2019).

[15]V. L. Deringer, M. A. Caro, and G. Csányi, "Machine learning interatomic potentials as emerging tools for materials science," Adv. Mater. **31**, 1902765 (2019).

[16]J. Behler, S. Lorenz, and K. Reuter, "Representing molecule-surface interactions with symmetry-adapted neural networks," J. Chem. Phys. **127**, 014705 (2007).

[17]J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, "Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential," Phys. Rev. Lett. **100**, 185501 (2008).

[18]R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler, and M. Parrinello, "Nucleation mechanism for the direct graphite-to-diamond phase transition," Nat. Mater. **10**, 693–697 (2011).

[19]N. Artrith and J. Behler, "High-dimensional neural network potentials for metal surfaces: A prototype study for copper," Phys. Rev. B **85**, 045439 (2012).

[20]G. C. Sosso, G. Miceli, S. Caravati, F. Giberti, J. Behler, and M. Bernasconi, "Fast crystallization of the phase change compound GeTe by large-scale molecular dynamics simulations," J. Phys. Chem. Lett. **4**, 4241–4246 (2013).

[21]N. Artrith, B. Hiller, and J. Behler, "Neural network potentials for metals and oxides-first applications to copper clusters at zinc oxide," Phys. Status Solidi B **250**, 1191–1203 (2013).

[22]W. J. Szlachta, A. P. Bartók, and G. Csányi, "Accuracy and transferability of Gaussian approximation potential models for tungsten," Phys. Rev. B **90**, 104108 (2014).

[23]S. K. Natarajan and J. Behler, "Neural network molecular dynamics simulations of solid-liquid interfaces: Water at low-index copper surfaces," Phys. Chem. Chem. Phys. **18**, 28704–28725 (2016).

[24]T. Morawietz, A. Singraber, C. Dellago, and J. Behler, "How van der Waals interactions determine the unique properties of water," Proc. Natl. Acad. Sci. U. S. A. **113**, 8368–8373 (2016).

[25]V. L. Deringer and G. Csányi, "Machine learning based interatomic potential for amorphous carbon," Phys. Rev. B **95**, 094203 (2017).

[26]K. Miwa and H. Ohno, "Interatomic potential construction with self-learning and adaptive database," Phys. Rev. Mater. **1**, 053801 (2017).

[27]R. Jinnouchi and R. Asahi, "Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning algorithm," J. Phys. Chem. Lett. **8**, 4279–4283 (2017).

[28]A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, "Machine learning a general-purpose interatomic potential for silicon," Phys. Rev. X **8**, 041048 (2018).

[29]K. Miwa and R. Asahi, "Molecular dynamics simulations with machine learning potential for Nb-doped lithium garnet-type oxide $Li_{7-x}La_3(Zr_{2-x}Nb_x)O_{12}$," Phys. Rev. Mater. **2**, 105404 (2018).

[30]F. C. Mocanu, K. Konstantinou, T. H. Lee, N. Bernstein, V. L. Deringer, G. Csányi, and S. R. Elliott, "Modeling the phase-change memory material, $Ge_2Sb_2Te_5$, with a machine-learned interatomic potential," J. Phys. Chem. B **122**, 8998–9006 (2018).

[31]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).

[32]M. Huang, X. Zhou, Y. Zhang, L. Zhou, M. Alducin, B. Jiang, and H. Guo, "Adiabatic and nonadiabatic energy dissipation during scattering of vibrationally excited CO from Au(111)," Phys. Rev. B **100**, 201407 (2019).

[33]K. Gubaev, E. V. Podryabinkin, G. L. W. Hart, and A. V. Shapeev, "Accelerating high-throughput searches for new alloys with active learning of interatomic potentials," Comput. Mater. Sci. **156**, 148–156 (2019).

[34]A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," Phys. Rev. B **87**, 184115 (2013).

[35]B. Huang and O. A. von Lilienfeld, "Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity," J. Chem. Phys. **145**, 161102 (2016).

[36]F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, "Alchemical and structural distribution based representation for universal quantum machine learning," J. Chem. Phys. **148**, 241717 (2018).

[37]M. J. Willatt, F. Musil, and M. Ceriotti, "Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements," Phys. Chem. Chem. Phys. **20**, 29661–29668 (2018).

[38]M. J. Willatt, F. Musil, and M. Ceriotti, "Atom-density representations for machine learning," J. Chem. Phys. **150**, 154110 (2019).

[39]M. A. Caro, "Optimizing many-body atomic descriptors for enhanced computational performance of machine learning based interatomic potentials," Phys. Rev. B **100**, 024112 (2019).

[40]S. N. Pozdnyakov, M. J. Willatt, A. P. Bartok, C. Ortner, G. Csanyi, and M. Ceriotti, "On the completeness of atomic structure representations," arXiv:2001.11696 [physics.chem-ph] (2020).

[41]A. P. Bartók and G. Csányi, "Gaussian approximation potentials: A brief tutorial introduction," Int. J. Quantum Chem. **115**, 1051–1057 (2015).

[42]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).

[43]M. W. Mahoney and P. Drineas, "CUR matrix decompositions for improved data analysis," Proc. Natl. Acad. Sci. U. S. A. **106**, 697–702 (2009).

[44]M. Parrinello and A. Rahman, "Crystal structure and pair potentials: A molecular-dynamics study," Phys. Rev. Lett. **45**, 1196–1199 (1980).

[45]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**, 7182 (1981).

[46]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865–3868 (1996).

[47]P. E. Blöchl, "Projector augmented-wave method," Phys. Rev. B **50**, 17953 (1994).

[48]G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," Phys. Rev. B **59**, 1758–1775 (1999).

[49]G. 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).

[50]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).