

Pair-distribution-function guided optimization of fingerprints for atom-centered neural network potentials

Cite as: J. Chem. Phys. 152, 224102 (2020); doi: 10.1063/5.0007391

Submitted: 12 March 2020 • Accepted: 18 May 2020 •

Published Online: 8 June 2020



View Online



Export Citation



CrossMark

Lei Li,¹ Hao Li,¹ Ieuan D. Seymour,¹ Lucas Koziol,² and Graeme Henkelman^{1,a)}

AFFILIATIONS

¹Department of Chemistry and the Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712-0231, USA

²Corporate Strategic Research, ExxonMobil Research and Engineering Company, 1545 US Route 22 East, Annandale, New Jersey 08801, USA

Note: This paper is part of the JCP Special Topic on Machine Learning Meets Chemical Physics.

a) Author to whom correspondence should be addressed: henkelman@utexas.edu

ABSTRACT

Atom-centered neural network (ANN) potentials have shown promise in computational simulations and are recognized as both efficient and sufficiently accurate to describe systems involving bond formation and breaking. A key step in the development of ANN potentials is to represent atomic coordinates as suitable inputs for a neural network, commonly described as fingerprints. The accuracy and efficiency of the ANN potentials depend strongly on the selection of these fingerprints. Here, we propose an optimization strategy of atomic fingerprints to improve the performance of ANN potentials. Specifically, a set of fingerprints is optimized to fit a set of pre-selected template functions in the f^*g space, where f and g are the fingerprint and the pair distribution function for each type of interatomic interaction (e.g., a pair or 3-body). With such an optimization strategy, we have developed an ANN potential for the Pd_{13}H_2 nanoparticle system that exhibits a significant improvement to the one based upon standard template functions. We further demonstrate that the ANN potential can be used with the adaptive kinetic Monte Carlo method, which has strict requirements for the smoothness of the potential. The algorithm proposed here facilitates the development of better ANN potentials, which can broaden their application in computational simulations.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0007391>

I. INTRODUCTION

Atomic simulation approaches based on quantum mechanics, empirical interatomic potentials, or classical force fields have demonstrated a remarkable power in gaining chemical and physical insights and providing guidance for the design of novel materials. Quantum mechanical methods, primarily based on density functional theory (DFT), offer sufficient accuracy for the simulation of chemical process and prediction of material properties. However, the computational cost of such methods remains high, which limits their application in large systems containing hundreds of atoms and where extensive sampling is required. Relatively inexpensive methods based on classical force fields are capable of modeling systems with thousands or even millions of atoms. However, classical

force fields are carefully parameterized to a specific chemical environment and suffer from a lack of transferability. In recent decades, machine learning (ML) algorithms have been introduced to describe interatomic interactions by fitting the potential energy surface with reference data obtained from quantum mechanical calculations.¹⁻⁷ ML potentials have been demonstrated to have DFT-level accuracy and orders of magnitude lower computational cost and show applicability to chemical and material systems involving bond formation and breaking.⁸⁻¹⁰

Until now, various ML techniques have been introduced for atomic simulation, including artificial neural networks (NNs),¹¹⁻¹⁶ kernel-ridge regression,^{1,17-19} and Gaussian process regression.²⁰⁻²² A key step in these techniques is to represent the atomic geometries with descriptors that are invariant to the operation of rotation

and translation so as to preserve the intrinsic symmetries of the system. Such descriptors, commonly described as atomic fingerprints, are designed to reflect the unique features of each geometric structure. Several different fingerprint algorithms have been developed, i.e., atom-centered symmetry functions (ACSFs),¹¹ the smooth overlap of atomic positions (SOAP),²³ bispectrum fingerprints,²⁰ and Zernike fingerprints.²⁴ The parameters in these fingerprints are boundless and able to form infinite dimensional vectors. Theoretically, the more complete the fingerprints, the higher accuracy can be reached, but at the expense of a higher computational cost. Therefore, an optimal selection of fingerprint parameters is essential to the performance of ML techniques.

In this work, we present a simple but effective algorithm for selecting fingerprint functions without relying on pre-training of the ML potential. The fingerprints we use here are based upon the ACSFs proposed by Behler and Parrinello (denoted as BP fingerprints).¹¹ The BP fingerprints have been widely used for the construction of atom-centered neural network (ANN) potentials that have a broad applicability across the periodic table and are able to provide accuracy comparable with the methods used to construct the reference data. Here, we first evaluate the performance of the ANN technique by fitting the potential energy surface of a 7-atom Lennard-Jones (LJ₇) particle and determine how the accuracy of the model varies with the set of BP fingerprints. Inspired by the LJ₇ particle case, we developed an algorithm to optimize the BP fingerprints to improve the performance of the ANN model.

To test our algorithm, we have fit forces and energies from DFT calculations of a Pd₁₃H₂ nanoparticle (NP) with the ANN technique. The PdH system is of interest in various areas, including hydrogen storage,^{25,26} hydrogen separation membranes,²⁷ hydrocarbon reforming,^{28–31} etc. Many scientific issues, i.e., the reactivity and dynamics of PdH nanoparticles, are important for these relevant applications. Here, we choose to model the dynamics of a PdH NP over long time scales using the adaptive kinetic Monte Carlo (AKMC) method.^{32,33} In our tests, we have not been able to find a force field with even qualitative agreement with forces and energies from DFT for these PdH NPs. Thus, here, we developed an ANN model for the Pd₁₃H₂ NP system and used AKMC to understand the time scales of the dynamics of the PdH NPs.

II. COMPUTATIONAL DETAILS

A. Atom-centered neural-network model

We used the TensorFlow version of the ANN algorithm implemented in the Atomistic Machine-Learning Package (AMP) developed by Khorshidi and Peterson for ML potential fitting.³⁴ As shown in Scheme 1, the ANN algorithm first interprets the structural information into unique fingerprints by representing the atomic configurations as single-atom chemical environments with generalized numerical representations. Here, we applied modified BP fingerprints of radial (f_i^I) and angular (f_i^II) forms, with centers at R_s and θ_s , respectively,

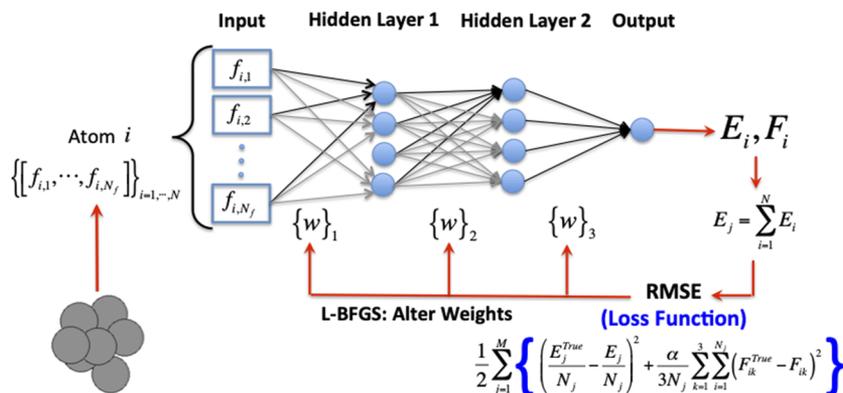
$$f_i^I = \sum_{j \neq i}^{\text{all}} e^{-\eta \frac{(R_{ij}-R_s)^2}{R_c^2}} f_c(R_{ij})$$

$$f_i^{II} = 2^{1-\zeta} \sum_{\substack{j,k \neq i \\ j \neq k}}^{\text{all}} (1 + \lambda \cos(\theta_{ijk} - \theta_s))^\zeta \\ \times e^{-\eta \frac{(R_{ij}^2 + R_{jk}^2 + R_{ik}^2)}{R_c^2}} f_c(R_{ij}) f_c(R_{jk}) f_c(R_{ik}),$$

where

$$f_c(R_{ij}) = \begin{cases} 0.5 \cdot \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1 \right] & \text{for } R_{ij} \leq R_c \\ 0 & \text{for } R_{ij} > R_c. \end{cases}$$

We set the cutoff radius (R_c) as 3.0 Å and 6.0 Å for the LJ₇ and Pd₁₃H₂ particle system, respectively. The parameters η and ζ are used to control the radial and angular fingerprint widths, respectively. For example, a larger η (ζ) value yields a narrower distribution of the radial (angular) symmetry function. A set of symmetry functions with different η and ζ values is required to describe the chemical environment of each atomic center. The numerical representations generated were weighted through an (n , m) NN model. Here, the length of the tuple (n , m) represents the number of hidden layers in the NN model with n and m corresponding to the number of neurons in each hidden layer. In this study, a (5) and a (50) NN model was used for the LJ₇ and Pd₁₃H₂



SCHEME 1. Flow chart of the atom-centered neural-network model. $f_{i,j}$ represents the j th fingerprint centered at the i th atom. w , N_f , N , and M represent the weight of the neural-network model, the number of fingerprints, atoms, and structures, respectively. E_i and F_i are the energy and force of the i th atom.

particle system, respectively. This corresponds to a NN model with one single hidden layer consisting of 5 or 50 neurons. Each neuron uses an activation function (the sigmoid function is selected here) to transform its input values to the output value. The output atomic energies and forces are evaluated with a loss function as defined in Scheme 1. The weights connecting each layer were initialized from a standard normal distribution and then optimized to minimize the loss function with the Limited-Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm. Finally, the resulting ANN model was used as a LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) force field through the OpenKIM module (refer to Part II of the [supplementary material](#) for more details).³⁵

B. Data collection

For the Pd₁₃H₂ particle, high-temperature molecular dynamics (high-T MD) simulations followed by geometry optimization (GEO-OPT) and saddle-point searching based on the dimer method were performed to obtain reference data.^{36–38} In the high-T MD/GEO-OPT simulations, we first ran three independent MD simulations with an NVT ensemble and the temperature fixed at 1000 K to produce high-energy structures. Each MD simulation ran for 20 ps with a time step of 1 fs. In this step, we used Gaussian and plane-wave basis functions with the PBE functional as implemented in the CP2K package.³⁹ A small single- ζ Gaussian basis set and the Goedecker–Teter–Hutter norm-conserved pseudopotentials were used to treat valence and the core electrons, respectively.^{40,41} The energy cutoffs for the finest grid level and Gaussian wavefunctions were set as 250 Ry and 30 Ry, respectively. Next, the fast inertial relaxation engine (FIRE) optimizer, as coded in the atomic simulation environment (ASE), was used to optimize the high-energy structures; 40 structures were collected along each optimization.⁴² Additional structures were also included to describe the chemical environments that were not found in either the dimer saddle search or high-T MD/GEO-OPT simulations. Specifically, geometries for the dissociation of the H₂ molecule and H atoms embedding inside the Pd₁₃ NP were also included. The data collected from each method was randomly split into training and testing sets in the ratio of 4:1 (Table I).

The reference energies and forces for the Pd₁₃H₂ NPs were evaluated with DFT as implemented in the Vienna *ab initio* simulation package (VASP 5.4.4). Projected augmented wave based pseudopotentials and a plane-wave basis set with an energy cutoff of 300 eV were employed to treat the core and the valence electrons, respectively.^{43–45} The generalized gradient approximation with the PBE functional was selected to calculate the exchange–correlation interaction.⁴⁶ A 15 × 15 × 15 Å³ cell was used to avoid interaction from neighboring particles. Note that the force field developed here was fit to the hydride system and is not expected to be transferrable even

TABLE I. Number of structures used for training and testing the ANN model in each category.

Categories	MD/GEO-OPT	Dimer	H ₂ dissociation	Embedded H atom
Training	5414	6312	1213	6863
Testing	1354	1578	304	1716

to similar systems such as Pd₁₃ with molecular H₂. In fact, tests show that the ANN is qualitatively different from DFT for recombinative H₂ desorption from Pd₁₃ (gray areas shown in Fig. S1). This discrepancy is due to the lack of training data in those regions. The objective here, however, is not to develop a general PdH potential; rather, it is to develop a strategy to optimize fingerprints and demonstrate the feasibility of combining an ANN model with the AKMC algorithm.

III. RESULTS AND DISCUSSION

A. PES fitting of the LJ₇ particle

First, we selected a simple system, the 7-atom LJ particle, to test the sensitivity of the ANN model with respect to variations of the fingerprint functions. We performed a basin-hopping calculation to locate all local minima of the LJ₇ particle and then constructed a disconnectivity graph by finding transition states that connect the local minima using the Pele package.⁴⁷ Reduced units are used for the Lennard-Jones potential, which is equivalent to setting the pair equilibrium well depth (ϵ) and separation (σ) to 1. The reason to focus on a disconnectivity graph as a measure of accuracy of the PES is that we are interested in modeling dynamics where the energy of minima and saddles are of particular importance. A total of 2609 structures were selected from the basin-hopping simulation, with 133 being selected from those with a minimum distance between atoms $R_{min} < 1.0$, 165 from those with $R_{min} > 2.5$, and 2307 from the rest plus the four local minima. This set of structures was used to fit the LJ₇ PES. The performance of the ANN model was evaluated by comparing the calculated ANN disconnectivity graph with that obtained from the reference LJ potential.

Figure 1(a) shows the disconnectivity graph of the LJ₇ particle obtained from the LJ potential. The LJ₇ particle has four local minima (black dots) that are connected by three unique transition states (the vertices that connect each pair of local minima). The red counterpart shows the disconnectivity graph obtained from the ANN model; note that the red and black plots are so similar that they appear superimposed on one another. One hidden layer with five neurons was used in this ANN model. Five radial symmetry functions centered at zero with $\eta = 0.05, 4, 20, 40, 80, \text{ and } 120$ were used as fingerprints. Note that this set of fingerprints was determined after numerous trials. We first followed the recommended fingerprints in Ref. 34 and then varied the interval between η values to find fingerprints that reproduced the disconnectivity graph of the LJ₇ particle. The near perfect match between the black and red lines indicates a near perfect agreement of the disconnectivity graph obtained from the ANN model with that from the LJ potential and that we have used a suitable set of fingerprints.

To understand the importance of each fingerprint function, we retrained the ANN model with one fingerprint removed and calculated the difference of the disconnectivity graph between the LJ potential and the ANN model as the root mean square error

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (E_i^{ANN} - E_i^{LJ})^2}{N}},$$

where E_i^{ANN} and E_i^{LJ} are the energies of the local minima and saddle points in the disconnectivity graph from the ANN model and the LJ potential, respectively, and N is the total number of local minima

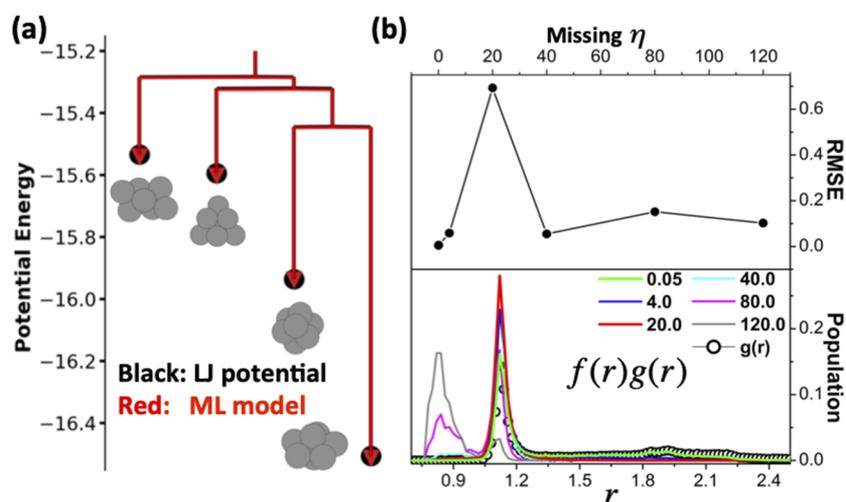


FIG. 1. (a) Disconnectivity graph of the LJ_7 particle obtained from basin-hopping with the LJ potential (black) and the ANN model (red). (b) Plot of $f(r)g(r)$ (bottom panel) with different η values and the RMSE variation of the disconnectivity graph with a missing η value.

and saddles. If the ANN finds more or fewer critical points than the LJ potential, the energies of the additional stationary points found in one model are evaluated in the other without relaxation, to provide a measure of the difference between the two landscapes in the RMSE.

The variation of the RMSE with one fingerprint removed is presented in Fig. 1(b) (top panel). Different sensitivities are observed for different fingerprints. For example, the removal of the $\eta = 20$ fingerprint, corresponding to a mean distance of 1.1, produces the largest error, suggesting that the ANN model is most sensitive to this fingerprint. In contrast, the ANN model with the $\eta = 0.05$ fingerprint removed, corresponding to a mean distance of 1.5, performs as accurately as with all fingerprints applied. The mean distance mentioned here corresponds to the center of the $f(r)g(r)$ plot in the bottom panel of Fig. 1(b). This analysis indicates that the $\eta = 0.05$ fingerprint is redundant for the ANN model. The removal of other fingerprints had a similar influence on the accuracy of the ANN model, indicating that only the $\eta = 20$ fingerprint is important for accuracy, which is consistent with modeling a simple pair potential in which the local energy of an atom is based upon the number of neighbors.

To understand the different sensitivities of the fingerprints, we computed the product of each fingerprint, $f(r)$, with the pair distribution function, $g(r)$, of the training data [denoted as $f(r)g(r)$]. To see this more clearly, an enlarged plot for $\eta = 4$ and 20 is shown in Fig. S2. This product describes how much a specific r value is weighted in the training data by this fingerprint. Such a measure helps describe which fingerprint distances are present in the training data. As shown in the bottom panel in Fig. 1(b), the fingerprints selected here emphasize the description of various radial distributions. For example, only one sharp peak centered at the equilibrium distance ($r = 1.1$) is found for the $\eta = 20$ fingerprint. This indicates that this fingerprint describes the two-body interaction around the equilibrium distance of the system and explains why the ANN model is most sensitive to this fingerprint. In contrast to the $\eta = 20$ fingerprint, the plot of the $f(r)g(r)$ with $\eta = 0.05$ has almost the same shape as that of the $g(r)$. Besides the peak at $R_{ij} = 1.1$, contributions to this fingerprint from other regions beyond 1.3 are also considerable due

to a significant population of those distances in the training data. Nevertheless, their contributions to the LJ potential are negligible compared to those from the region near the equilibrium distance. In other words, the $\eta = 0.05$ fingerprint is unable, on its own, to correctly describe the interaction near the equilibrium distance due to the inclusion of contribution from other regions. Therefore, the removal of the $\eta = 20$ fingerprint leads to a large deviation from the real LJ potential whereas a small error is induced upon removal of the $\eta = 0.05$ fingerprint. These results suggest that to reach a reasonable accuracy for the ANN model, each fingerprint should probe a specific region of the atom-centered chemical environment. The product of $f(r)$ and $g(r)$ is potentially a good criterion for the determination of fingerprints that are distinguishable from each other in this regard.

B. Fingerprints optimization

A principle for designing an ANN algorithm is to have fingerprints distinguishable from each other such that each one probes a specific region of the atom-centered chemical environment. Following this rule, Smith *et al* developed extensible neural network potentials for multiple common organic molecules by evenly distributing the value of R_s and θ_s with fixed η and ζ values.¹⁶ An even distribution of the fingerprint center guarantees inclusion of all possible chemical environments to be explored. However, a biased population of a specific type of chemical environments in the training data will overweight or underweight the corresponding fingerprint, leading to a poor performance of the ANN model. Following the sensitivity test of the LJ_7 particle, the product of each fingerprint with the pair distribution function $g(r)$ can be used to determine how a specific fingerprint is biased in training data. In other words, we need to find a set of fingerprints that are distinguishable from each other upon taking the product with the pair distribution function. This means finding distinguishable fingerprints in the $f(r)g(r)$ (simplified as f^*g) space instead of the $f(r)$ space.

Here, we propose an optimization algorithm to find distinguishable fingerprints in the f^*g space and apply the algorithm

to locate fingerprints for the Pd₁₃H₂ nanoparticle system. Figure 3 presents the pair distribution function and radial fingerprints near the equilibrium distance for the Pd–H pair before and after optimization. We first defined a set of template fingerprints $\left\{f_i^{TP}(r)\right\}_{i=1,N_f}$ that describe separate regions in the pair distribution space to be described. Then, an objective function (τ) defined by the following equation is minimized to find fingerprints $\left\{f_i^{Opt}(r)\right\}_{i=1,N_f}$ that reproduce the template in the f^*g space,

$$\tau = \sum_i^{N_f} \left\{ \int_r^{R_c} \alpha (f_i^{Opt}(r)g(r) - f_i^{TP}(r))^2 dr \right\},$$

where α is the weight parameter set to ensure that the error in the tail region is properly weighted. The tail region is defined as that region where the value of the template fingerprint is smaller than 1/100 of its maximum value. The α value can be determined as the minimum value that eliminates the error in the tail region. An example is shown in Fig. 2, where the tail corresponds to the region with $r > 2.0$ Å. A small α leads to mismatch between the optimized and the template fingerprint in this tail region, i.e., the black line (the optimized fingerprint with $\alpha = 1$) as compared to the gray-circled line (the template fingerprint). Increasing the value of α reduces the mismatch in the tail region and leads to a right shift of the fingerprint center (the R_s value) as shown in the right panel of Fig. 2. The R_s value reaches a maximum value with $\alpha = 100$ —that is, the minimum α value to be used to eliminate the mismatch in the tail region. As such, we set α as follows:

$$\alpha = \begin{cases} 1, & f_i^{TP}(r = r_j) \geq 0.01 * \max(f_i^{TP}(r)|_{r=0}^{r=R_c}) \\ 100, & f_i^{TP}(r = r_j) < 0.01 * \max(f_i^{TP}(r)|_{r=0}^{r=R_c}). \end{cases}$$

Three fingerprints with $R_s = 1.6, 1.75,$ and 1.9 Å and $\eta = 1500$, as shown by dashed lines in the middle panel of Fig. 3, are selected as templates to describe the interactions centered at the respective R_s value in the f space. The $g(r)$ of the H–Pd pair distribution is most populated around the first-nearest distance (1.77 Å) as shown

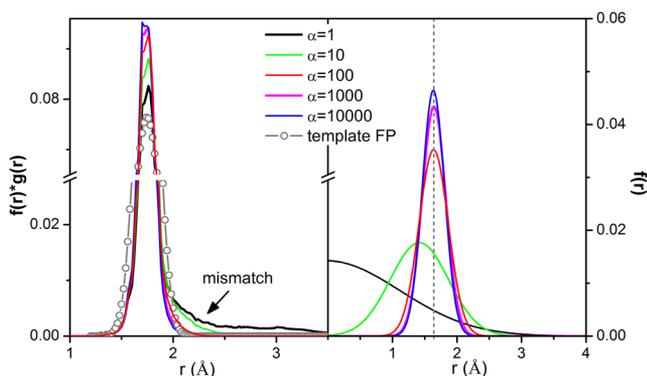


FIG. 2. The right panel shows a plot of fingerprints that are optimized to the template functions (the black line) centered at 1.75 Å with different α values. The left panel shows a plot of the corresponding product with the pair distribution function.

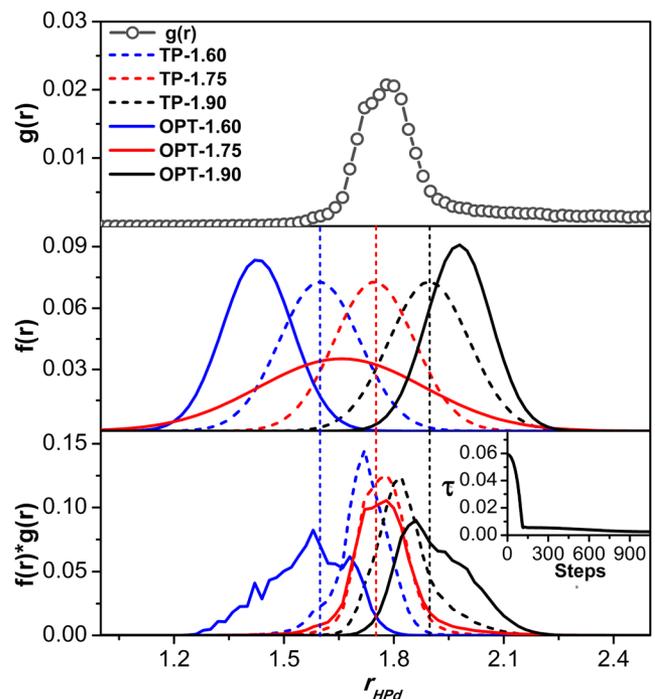


FIG. 3. Plot of the pair distribution function of the H–Pd pair (top panel), the template fingerprints (dashed lines in the middle panel), the optimized fingerprints (solid lines in the middle panel), the $f(r)g(r)$ before (dashed lines in the bottom panel) and after (solid lines in the bottom panel) optimization. The inset image presents the evolution of the objective function (τ) with optimization steps. The dashed vertical lines indicate the radial center of the template fingerprints.

in the top panel of Fig. 3. The strong population bias at the first-neighbor distance leads to a reduction of the fingerprints in the f^*g space toward the peak position of the $g(r)$. The center of these populated fingerprints (dashed lines in the bottom panel of Fig. 3) ranges from 1.72 Å to 1.82 Å, much narrower than our target (1.6–1.9 Å). This indicates that the combination of template fingerprints with our training data tends to emphasize the interaction in a region close to the first-nearest neighbor distance (± 0.05 Å).

By optimizing the objective function (τ), we shifted the centers of the populated fingerprints to our target range. The objective function is continuous and smooth with respect to η and R_s . An example contour map of τ for the template fingerprint centered at 1.6 Å is given in Fig. S3. There exists only one critical point, which is the global minimum. The global minimum of τ can be located in the fast inertial relaxation engine (FIRE) as shown in the inset image of the bottom panel in Fig. 3. The fingerprints obtained after optimization are presented as solid lines in the middle panel of Fig. 3. Compared to the template fingerprints, the two outermost fingerprints exhibit a similar shape but are shifted away from the first-nearest distance in order to lower the weight in that region. The coverage of the middle fingerprint is expanded with a slightly shifted center. As a result, the optimized fingerprints cover radial regions as defined by the template in the f^*g space (solid lines in the bottom panel

of Fig. 3). To fully describe the Pd–H interaction, another six fingerprints are selected and optimized following the same procedure (Fig. S4). These fingerprints are expected to improve the description of interactions centered at the target R_s values for the training data.

C. An improved ANN model with optimized fingerprints for the Pd₁₃H₂ system

Next, we located fingerprints for other interactions following the proposed optimization procedure, including the H–H, and Pd–Pd radial, and the Pd–Pd–Pd angular terms. Only the Pd–Pd–Pd angular term is included given that the angular term with hydrogen has little contribution to the system. We followed the exact same procedure used in the radial term to optimize the angular term. The template and optimized fingerprints are presented in Figs. S5–S7 and the parameters are included in the attached potential files. To evaluate the performance of the optimized fingerprints, we trained ANN models to fit the DFT-based PES of the Pd₁₃H₂ NP with both the template and optimized fingerprints. Ten independent models are trained for each set of fingerprints and the one with the best performance discussed here. In the models from both sets of fingerprints, the training and the test sets show comparable accuracy without obvious overfitting (Figs. S8 and S9).

Figure 4(a) shows the predicted energy (left panel) and forces (right panel) from the template (blue) and the optimized (red) models against the reference DFT values. In both panels, the optimized model exhibits a tighter distribution along the $y = x$ line as compared to the template model. This indicates that the ANN model with optimized fingerprints reaches a higher accuracy compared to those with the template fingerprints. Specifically, the template model shows a lower performance on force prediction as indicated by the sparse distribution of blue dots in the right panel of Fig. 4(a). These results demonstrate that the ANN model is improved for the training and test data upon fingerprint optimization.

To further validate our models, we artificially pulled one Pd atom and one H atom away from the particle and evaluated the energy of structures along this trajectory as indicated in the inset in Fig. 4(b). The structures used in this validation were not included in the training data. The resulting energy profiles from the template (blue) and optimized (red) models are compared to the reference DFT values in Fig. 4(b). The energy profile from the optimized model shows better agreement with the reference DFT results than that from the template model. Specifically, the template model underestimates the energy at $r < 2.7$ Å but overestimates at $r > 2.7$ Å in the Pd–Pd case and predicts a flat basin from 2.5 Å to 2.7 Å, inconsistent with the reference DFT data. The optimized model largely minimizes this discrepancy. In the H–Pd case, the optimized model

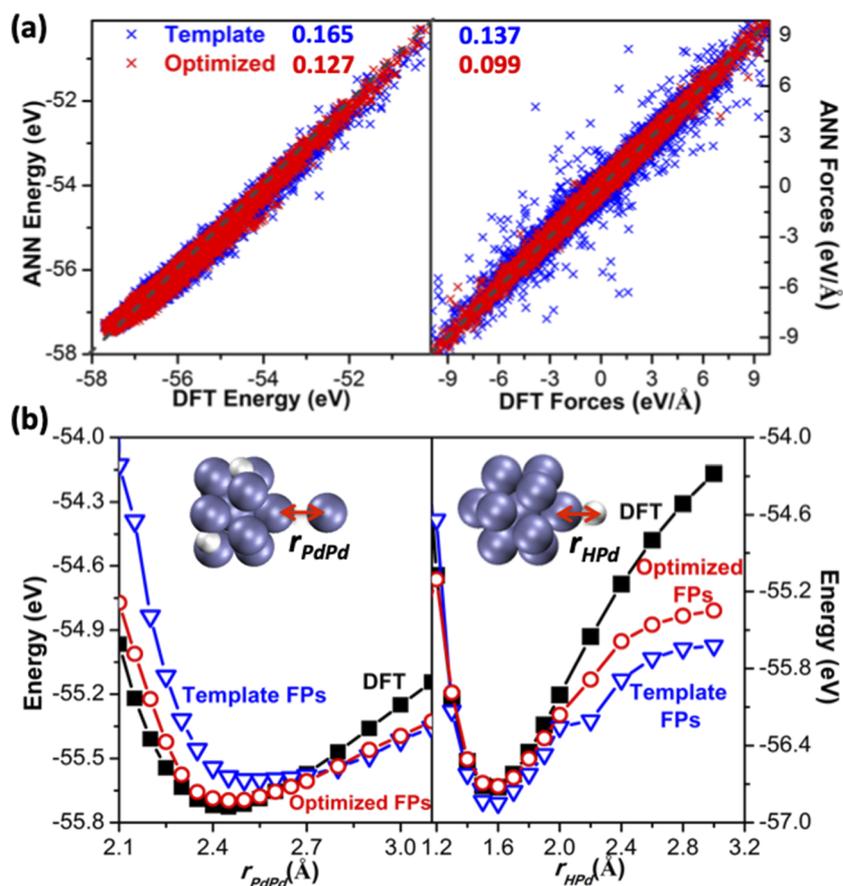


FIG. 4. (a) Comparison of the energy (left panel) and forces (right panel) evaluated by the ANN model with the reference DFT values. The blue and red dots represent values calculated from the ANN model based on the template and the optimized fingerprints, respectively. Corresponding root mean square errors are presented in the figure. (b) Energy variation with the Pd–Pd and the H–Pd distance obtained from the DFT method (black) and the ANN model based on the template (blue) and the optimized (red) fingerprints.

is able to reproduce DFT energies of structures with $r < 2.0 \text{ \AA}$ and increases the prediction accuracy for structures beyond that region. More importantly, the optimized model yields a smooth energy profile along the trajectory, whereas a turning point is observed in the template model. These validation results further indicate that our fingerprint optimization method improves the global performance of the ANN model.

D. AKMC simulations with the ANN model

The AKMC method is a coarse-grain dynamic simulation method with time evolution of the system accumulating on the time scale of the state-to-state transitions.^{32,33} In our previous study, we showed that the AKMC method is able to mitigate the time scale problem in molecular dynamics simulations and extend the time scale to that of experiments.⁴⁸ Nevertheless, a significant challenge for the AKMC method is the availability of reliable and inexpensive force fields that are accurate and smooth enough to facilitate the location of transition states.

Here, we performed AKMC simulations to compute the dynamical evolution of a Pd_{13}H_2 NP at 300 K with energies and forces evaluated by our optimized ANN model. In the AKMC simulation, a minimum distance (R_{min}) between each atomic type was set to avoid the collapse of the ANN model beyond the range of the training data. Any distance below R_{min} was set to R_{min} . The AKMC method as implemented in the EON package was used to perform the simulation.⁴⁹ The dimer method was used to find all possible low-energy transition events, and the kinetic Monte Carlo algorithm was used to determine which event the system proceeds with. At each new state, such dimer/KMC cycles were repeated to simulate the state-to-state evolution of the system. More details about the AKMC method can be found in Ref. 48.

The blue line in Fig. 5(a) shows the time evolution of the total energy of the system from the AKMC simulation based on the ANN model. Roughly, five energy plateaus associated with five types of structures, as grouped in the disconnectivity graph [Fig. 5(b)], are observed in the simulation. Each drop in the energy evolution plot corresponds to a structural transition, except for the transition from 4 to 5, which leads to a slight increase in energy. For comparison, we also calculated the energy of the local-minima structures extracted from the AKMC simulation with DFT [red line in Fig. 5(a)]. The time evolution of the total energy for the AKMC simulation based on the ANN model is consistent with the DFT reference. Using DFT, we further relaxed the most stable structures, corresponding to the green dots in the disconnectivity graph. The structures obtained with the corresponding energy (E_{diff}) and structure (S_{diff}) difference are shown in Fig. 5(c). The gold (light blue) and green (white) colors represent Pd and H atoms in the DFT-relaxed (ANN-relaxed) structures, respectively. The structures obtained from both methods are in agreement with S_{diff} values smaller than 0.05 \AA/atom . The energy differences range from 0.05 eV to 0.20 eV . These results demonstrate the reliability and accuracy of our ANN.

As discussed above, the NP undergoes a structural transition accompanied by diffusion of H atoms on the NP surface. The overall energy barrier for the structural transition from type k to $k + 1$ can be extracted from the disconnectivity graph, taken as the energy difference between the minimum-energy structure (green dots in Fig. 5(b)) of type k and the vertex that connects to a $k + 1$ state.

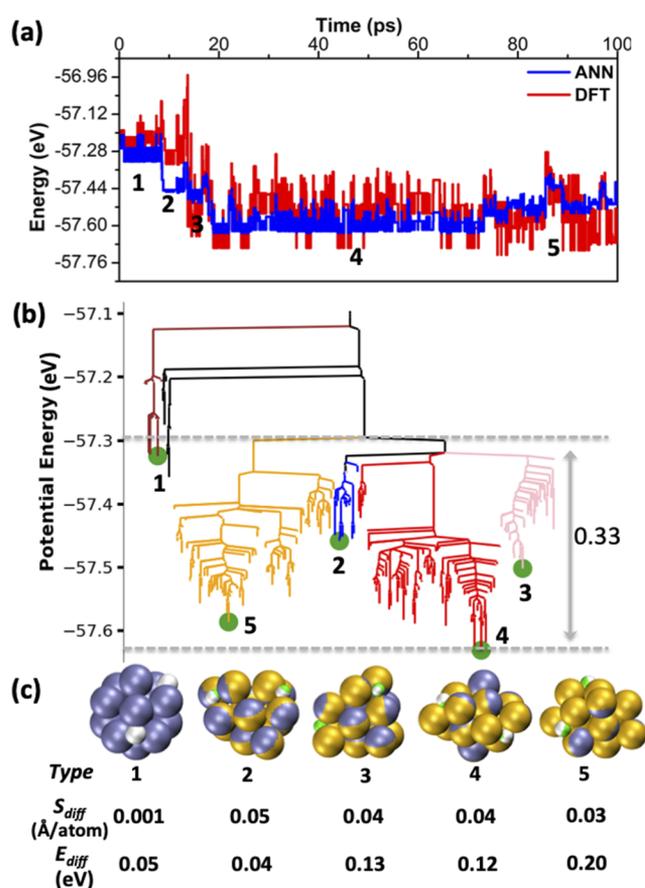


FIG. 5. (a) Time evolution of the total energy and (b) disconnectivity graph obtained from an AKMC simulation of the Pd_{13}H_2 cluster at 300 K. The numbers in (a) and (b) indicate specific structures shown in (c), which are the lowest-energy structures for different types of isomers (highlighted as green dots in the disconnectivity graph). The gold (light blue) and green (white) colors represent H and Pd atoms in the structure obtained from DFT (optimized ANN model). S_{diff} and E_{diff} represent the structural and energy differences between geometries obtained from the DFT method and the optimized ANN model, respectively.

For example, the $4 \rightarrow 5$ transition has an overall barrier $\sim 0.33 \text{ eV}$ as indicated by the gray dotted lines in Fig. 5(b). In this way, we are able to characterize overall energy barriers for the $k \rightarrow k + 1$ transitions ranging from 0.15 eV to 0.33 eV . Each structural type entails a specific Pd skeletal structure with H atoms that can almost freely diffuse on the surface of the Pd structure with very low barriers ($< 0.07 \text{ eV}$). These results suggest that the Pd_{13}H_2 particle stays in a dynamic state with facile movement of H. Throughout the AKMC simulation, we did not observe embedding of H atoms into the nanoparticle. Likely, a higher barrier is required for such embedding events, whereas the diffusion of surface H atoms with low barriers limits the timescale of the AKMC simulation.

IV. CONCLUSIONS

In this study, we first tested the sensitivity of the ANN model with respect to the ACSF by training an ANN model for the LJ_7

particle and found that the product of the fingerprint functions with the corresponding pair distribution function are responsible for the performance variation of the ANN model. Inspired by this sensitivity test, we defined an objective function that can be minimized to find optimal fingerprints, thus improving the performance of the ANN model. In this optimization process, we first define a set of template fingerprints based on the radial and angular regions to be explored and then minimize the objective function to locate fingerprints that can reproduce the template fingerprints in the f^*g space. Accordingly, we successfully developed an ANN model for the Pd₁₃H₂ NP system that is superior to the model trained with the template fingerprints. Furthermore, we applied the developed ANN model to AKMC simulations of dynamic behavior of the Pd₁₃H₂ NP, proving the reliability and accuracy of the model.

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for additional details about the structure of the optimized fingerprints and their efficacy.

ACKNOWLEDGMENTS

Funding for this project was provided by the National Science Foundation (Grant No. CHE-1764230) and Exxon-Mobil (Grant No. UTA17-000317 EM10480.13). Computational resources were provided by the Texas Advanced Computing Center and the XSEDE program (Grant No. CHE190010). G.H. acknowledges the ongoing support from the Welch Foundation (Grant No. F-1841).

DATA AVAILABILITY

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

REFERENCES

- 1 F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K. R. Muller, "Bypassing the Kohn-Sham equations with machine learning," *Nat. Commun.* **8**, 872 (2017).
- 2 J. C. Snyder, M. Rupp, K. Hansen, K. R. Muller, and K. Burke, "Finding density functionals with machine learning," *Phys. Rev. Lett.* **108**(25), 253002 (2012).
- 3 Z. W. Li, J. R. Kermode, and A. De Vita, "Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces," *Phys. Rev. Lett.* **114**(9), 096405 (2015).
- 4 T. L. Jacobsen, M. S. Jorgensen, and B. Hammer, "On-the-fly machine learning of atomic potential in density functional theory structure optimization," *Phys. Rev. Lett.* **120**(2), 026102 (2018).
- 5 S. Chiriki, S. Jindal, and S. S. Bulusu, "Neural network potentials for dynamics and thermodynamics of gold nanoparticles," *J. Chem. Phys.* **146**(8), 084314 (2017).
- 6 S. Jindal, S. Chiriki, and S. S. Bulusu, "Spherical harmonics based descriptor for neural network potentials: Structure and dynamics of Au₁₄₇ nanocluster," *J. Chem. Phys.* **146**(20), 204301 (2017).
- 7 R. Ouyang, Y. Xie, and D.-e. Jiang, "Global minimization of gold clusters by combining neural network potentials and the basin-hopping method," *Nanoscale* **7**(36), 14817–14821 (2015).
- 8 A. A. Peterson, R. Christensen, and A. Khorshidi, "Addressing uncertainty in atomistic machine learning," *Phys. Chem. Chem. Phys.* **19**(18), 10978–10985 (2017).
- 9 F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, "Prediction errors of molecular machine learning models lower than hybrid DFT error," *J. Chem. Theory Comput.* **13**(11), 5255–5264 (2017).
- 10 A. A. Peterson, "Acceleration of saddle-point searches with machine learning," *J. Chem. Phys.* **145**(7), 074106 (2016).
- 11 J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.* **98**(14), 146401 (2007).
- 12 J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *J. Chem. Phys.* **134**(7), 074106 (2011).
- 13 M. Gastegger and P. Marquetand, "High-dimensional neural network potentials for organic reactions and an improved training algorithm," *J. Chem. Theory Comput.* **11**(5), 2187–2198 (2015).
- 14 L. Shen, J. Wu, and W. Yang, "Multiscale quantum mechanics/molecular mechanics simulations with neural networks," *J. Chem. Theory Comput.* **12**(10), 4934–4946 (2016).
- 15 K. T. Schutt, F. Arbabzadah, S. Chmiela, K. R. Muller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nat. Commun.* **8**, 13890 (2017).
- 16 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**(4), 3192–3203 (2017).
- 17 S. Chmiela, H. E. Sauceda, K. R. Muller, and A. Tkatchenko, "Towards exact molecular dynamics simulations with machine-learned force fields," *Nat. Commun.* **9**(1), 3887 (2018).
- 18 O. T. Unke and M. Meuwly, "Toolkit for the construction of reproducing kernel-based representations of data: Application to multidimensional potential energy surfaces," *J. Chem. Inf. Model.* **57**(8), 1923–1931 (2017).
- 19 D. Hu, Y. Xie, X. Li, L. Li, and Z. Lan, "Inclusion of machine learning kernel ridge regression potential energy surfaces in on-the-fly nonadiabatic molecular dynamics simulation," *J. Phys. Chem. Lett.* **9**(11), 2725–2732 (2018).
- 20 A. P. Bartok, M. C. Payne, R. Kondor, and G. Csanyi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," *Phys. Rev. Lett.* **104**(13), 136403 (2010).
- 21 S. T. John and G. Csanyi, "Many-body coarse-grained interactions using Gaussian approximation potentials," *J. Phys. Chem. B* **121**(48), 10934–10949 (2017).
- 22 O.-P. Koistinen, F. B. Dagbjartsdóttir, V. Ásgeirsson, A. Vehtari, and H. Jónsson, "Nudged elastic band calculations accelerated with Gaussian process regression," *J. Chem. Phys.* **147**(15), 152720 (2017).
- 23 A. P. Bartok, R. Kondor, and G. Csanyi, "On representing chemical environments," *Phys. Rev. B* **87**(18), 184115 (2013).
- 24 M. Novotni and R. Klein, "Shape retrieval using 3D Zernike descriptors," *Comput. - Aided Des.* **36**(11), 1047–1062 (2004).
- 25 M. Yamauchi, H. Kobayashi, and H. Kitagawa, "Hydrogen storage mediated by Pd and Pt nanoparticles," *ChemPhysChem* **10**(15), 2566–2576 (2009).
- 26 Y. E. Cheon and M. P. Suh, "Enhanced hydrogen storage by palladium nanoparticles fabricated in a redox-active metal-organic framework," *Angew. Chem., Int. Ed.* **48**(16), 2899–2903 (2009).
- 27 P. Kamakoti, B. D. Morreale, M. V. Ciocco, B. H. Howard, R. P. Killmeyer, A. V. Cugini, and D. S. Sholl, "Prediction of hydrogen flux through sulfur-tolerant binary alloy membranes," *Science* **307**, 569–573 (2005).
- 28 D. Teschner, J. Borsodi, A. Wootsch, Z. Révay, M. Havecker, A. Knop-Gericke, S. D. Jackson, and R. Schlögl, "The roles of subsurface carbon and hydrogen in palladium-catalyzed alkyne hydrogenation," *Science* **320**(5872), 86–89 (2008).
- 29 D. Teschner, Z. Révay, J. Borsodi, M. Hävecker, A. Knop-Gericke, R. Schlögl, D. Milroy, S. D. Jackson, D. Torres, and P. Sautet, "Understanding palladium hydrogenation catalysts: When the nature of the reactive molecule controls the nature of the catalyst active phase," *Angew. Chem., Int. Ed.* **47**(48), 9274–9278 (2008).
- 30 Y. Chin, R. Dagle, J. Hu, A. C. Dohnalkova, and Y. Wang, "Steam reforming of methanol over highly active Pd/ZnO catalyst," *Catal. Today* **77**(1), 79–88 (2002).

- ³¹M. A. Goula, S. K. Kontou, and P. E. Tsiakaras, "Hydrogen production by ethanol steam reforming over a commercial Pd/ γ -Al₂O₃ catalyst," *Appl. Catal., B* **49**(2), 135–144 (2004).
- ³²L. Xu and G. Henkelman, "Adaptive kinetic Monte Carlo for first-principles accelerated dynamics," *J. Chem. Phys.* **129**(11), 114104 (2008).
- ³³G. Henkelman and H. Jónsson, "Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table," *J. Chem. Phys.* **115**(21), 9657 (2001).
- ³⁴A. Khorshidi and A. A. Peterson, "Amp: A modular approach to machine learning in atomistic simulations," *Comput. Phys. Commun.* **207**, 310–324 (2016).
- ³⁵E. B. Tadmor, R. S. Elliott, J. P. Sethna, R. E. Miller, and C. A. Becker, "The potential of atomistic simulations and the knowledgebase of interatomic models," *JOM* **63**(17), 17 (2011).
- ³⁶A. Heyden, A. T. Bell, and F. J. Keil, "Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method," *J. Chem. Phys.* **123**(22), 224101 (2005).
- ³⁷J. Kästner and P. Sherwood, "Superlinearly converging dimer method for transition state search," *J. Chem. Phys.* **128**(1), 014106 (2008).
- ³⁸G. Henkelman and H. Jónsson, "A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives," *J. Chem. Phys.* **111**(15), 7010–7022 (1999).
- ³⁹J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, "cp2k: Atomistic simulations of condensed matter systems," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **4**(1), 15–25 (2014).
- ⁴⁰J. VandeVondele and J. Hutter, "Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases," *J. Chem. Phys.* **127**(11), 114105 (2007).
- ⁴¹S. Goedecker, M. Teter, and J. Hutter, "Separable dual-space Gaussian pseudopotentials," *Phys. Rev. B* **54**(3), 1703–1710 (1996).
- ⁴²A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment—A Python library for working with atoms," *J. Phys.: Condens. Matter* **29**(27), 273002 (2017).
- ⁴³G. 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**(1), 15–50 (1996).
- ⁴⁴G. Kresse and J. Furthmüller, "Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set," *Phys. Rev. B* **54**(16), 11169–11186 (1996).
- ⁴⁵P. E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B* **50**(24), 17953–17979 (1994).
- ⁴⁶J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.* **77**(18), 3865–3868 (1996).
- ⁴⁷See <https://github.com/pele-python/pele> for Pele.
- ⁴⁸L. Li, X. Li, Z. Duan, R. J. Meyer, R. Carr, S. Raman, L. Koziol, and G. Henkelman, "Adaptive kinetic Monte Carlo simulations of surface segregation in PdAu nanoparticles," *Nanoscale* **11**(21), 10524–10535 (2019).
- ⁴⁹S. T. Chill, M. Welborn, R. Terrell, L. Zhang, J.-C. Berthet, A. Pedersen, H. Jónsson, and G. Henkelman, "EON: Software for long time simulations of atomic scale systems," *Modell. Simul. Mater. Sci. Eng.* **22**(5), 055002 (2014).