Computational Materials Science 208 (2022) 111330

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Full length article

Batch active learning for accelerating the development of interatomic potentials

Nathan Wilson [a], Daniel Willhelm [a], Xiaoning Qian [b,c], Raymundo Arróyave [a,d,e,*], Xiaofeng Qian [a,b,*]

[a] Department of Materials Science and Engineering, Texas A&M University, 3003 TAMU, College Station, 77843, TX, USA
[b] Department of Electrical and Computer Engineering Texas A&M University, 3128 TAMU, College Station, 77843, TX, USA
[c] Department of Computer and Science Engineering Texas A&M University, 3112 TAMU, College Station, 77843, TX, USA
[d] Department of Mechanical Engineering Texas A&M University, 3123 TAMU, College Station, 77843, TX, USA
[e] Department of Industrial and Systems Engineering Texas A&M University, 3131 TAMU, College Station, 77843, Tx, USA

ARTICLE INFO

Keywords:
Interatomic potentials
Active learning
Molecular dynamics
Density functional theory

ABSTRACT

Classical molecular dynamics (MD) has been widely used to study atomistic mechanisms and emergent behavior in materials at length and time scales beyond the capabilities of first-principles approaches. The success of classical MD simulations relies on the ability of classical interatomic potentials to accurately map complex many-body interacting systems of electrons and nuclei into effective few-body interacting systems of atoms. In practice, the development of interatomic potentials is a nontrivial process and requires considerable amount of effort. Recently, machine learning has become a promising approach to accelerate interatomic potential development. However, these machine learning approaches are often computation and data intense, as they require a large amount of training data from first-principles calculations, such as total energies, atomic forces, and stress tensors of many atomistic structures. Here we propose an active learning approach combined with first-principles theory calculations to expedite the development of machine learning interatomic potentials. In particular, we develop a batch active learning method which combines both energy uncertainty and structure similarity metrics to efficiently sample the highly uncertain structures that are difficult to predict. This active sampling approach maximizes the utility of the dataset in each batch and generates interatomic potential with highly accurate and robust model coefficients which are difficult to achieve with conventional sampling approaches. To demonstrate this batch active learning method, we develop an active learning potential for monolayer GeSe, a two-dimensional ferroelectric-ferroelastic material, and compare the quality and robustness of the active learning potential with the potential obtained from random sampling. Batch active learning method opens up avenues for accelerating the development of robust and accurate machine learning potential using a small set of atomistic structures which will be valuable for computational materials, physics, and chemistry community.

1. Introduction

The potential energy surface (PES) of many-body systems governs the thermodynamic and kinetic properties of materials under specific boundary conditions and external fields. For a given atomistic system, its exact PES lives in an extremely high-dimensional space where electronic degrees of freedom pose particular challenges due to the inherent quantum-mechanical wave nature of electrons. Classical interatomic potentials, *i.e.* classical force fields, are introduced to map the complex quantum system onto a classical system using effective few-body interatomic interaction potentials. The latter enables efficient calculation of interatomic forces and subsequently the dynamics of atomistic structures under different conditions. Classical interatomic potentials calculate the forces and energy of atomistic structures without the explicit description of electrons, thus facilitate large-scale molecular dynamics (MD) simulations which provide atomistic insight into the behavior of materials, ranging from nanomaterials [1–3] to irradiated metals [4]. However, the development of classical interatomic potentials, the cornerstone of MD simulations, is not only time-consuming, but also requires considerable expertise and extensive benchmarking. Moreover, the transferability and interpretability of classical interatomic potentials remains a significant challenge.

* Corresponding authors at: Department of Materials Science and Engineering, Texas A&M University, 3003 TAMU, College Station, 77843, TX, USA
E-mail addresses: raymundo.arroyave@tamu.edu (R. Arróyave), feng@tamu.edu (X. Qian).

https://doi.org/10.1016/j.commatsci.2022.111330
Received 27 December 2021; Received in revised form 16 February 2022; Accepted 1 March 2022
Available online 17 March 2022

Recent advances in machine learning have enabled the rapid development of machine learning interatomic potentials combining first-principles density functional theory (DFT) calculations with machine learning algorithms [5]. These machine learning potentials are generated by fitting a set of environment-dependent atomistic descriptors to physical properties calculated by DFT, such as total energies, atomic forces, and stresses of atomistic structures. Many different atomistic descriptors and models have been proposed, for example, Spectral Neighbor Analysis Potential (SNAP) [6], Atomic Cluster Expansion (ACE) [7], Gaussian Approximation Potential (GAP) [8,9], Moment Tensor Potentials [10], AGNI interatomic potential [11], and Neural Network Potential (NNP) [12–14]. These atomistic descriptors and physical properties are used to train the models with various machine learning algorithms, such as kernel ridge regression, linear regression, Gaussian processes, support vector machines, and neural networks. Some descriptors adopt specific models, *e.g.* NNP [12,15–17], or even build their own models, *i.e.* Potential Optimization by Evolutionary Techniques (POET) [18].

Unlike conventional interatomic potentials, machine learning potentials often require performing first-principles calculations to sample thousands of structures, making their development computationally expensive. While these potentials are trained on large sets of structures, it is possible that only a subset of these structures are truly necessary to properly train the potential, which implies that a large number of computationally intensive first-principles calculations would be wasted on generating data with marginal utility to the training task. Active learning methods, on the other hand, constitute an effective approach to train machine learning potentials by optimally selecting the training structures which are most likely to improve the performance of the model. Among a pool of potential structures that have yet to be simulated via first-principles methods (unlabeled data), the active learning algorithms will select the ones to simulate (labeled data) based on an *a priori* defined acquisition function. The procedure above is carried out iteratively, enabling the development of accurate and consistent models in an optimal manner. Active learning methods become particularly beneficial when the unlabeled (i.e. yet to be acquired/simulated) data is plentiful and the labeling of data is expensive. For classical MD potentials, generating new structures can be done quickly and cheaply, while labeling them would require DFT calculations that are highly computationally demanding.

Active machine learning potentials have been predominately developed in two categories: the on-the-fly approaches and the structure exploration approaches. The on-the-fly approaches perform an MD simulation using the currently trained machine learning potential and then determines if the generated structures fail to meet some criteria, *i.e.* too large of an uncertainty or over-extrapolation. If the generated structures fail, those structures will be cast into DFT calculations and subsequently included in the training set to retrain the potential [19–23]. These on-the-fly approaches can be used during real simulations to test the validity of predictions. However, they do not intend to actively search for those structures that will improve the potential the most. In addition, these on-the-fly approaches rely on the prior trained potential to guide the exploration of configuration space, so it may miss important structures if the prior trained potential avoids those spaces. The second category is the structure exploration approaches which search for new and unique structures in the configuration space [24,24–28]. In this category, the selected structures are the most dissimilar ones from the other structures and various methods have been proposed to quantify the dissimilarity [27,29–32]. An important benefit of the structure exploration approaches is that they can be used when there is no information available on the material and can run with almost no human input. However, they often ignore the predictive capability of the prior trained model for exploring new structures, hence they may select structures that are already well-predicted, or choose structures that are far away from the desired simulation environment.

In this work, we introduce a batch active learning method that selects structures using both the structure's dissimilarity and the potential's predictive capability. By using both metrics, our method selects multiple structures without updating the potential while avoiding redundant information. Selecting multiple structures in a single batch allows for running multiple DFT calculations to label the sampled structures in parallel, *i.e. batch active sampling*, thereby significantly accelerating DFT-based structure labeling. We will show that our active learning method produces consistent and accurate potentials by performing training on difficult structures, *i.e.* structures with larger errors in the predicted energy/forces. The accuracy of our active learning potential is validated by the atomic forces, phonon dispersions, and transition temperature of multiferroic monolayer GeSe from a series of MD calculations. Our results show that our active learning method offers an efficient approach for developing accurate and consistent potentials with well-converged coefficients for monolayer GeSe using only 3,000 structures from a large database of 13,006 structures [33].

2. Methods

2.1. Batch active learning

A schematic of our active learning method is shown in Fig. 1. Our method starts with a pool of candidate structures, *i.e.* a set of pre-generated structures representing the input structure space. N_I structures (N_I=10 in the present case) are randomly selected from the pool of structures with their energies calculated using DFT to create the initial labeled structures, while the other structures are the unlabeled structures. The labeled structures are used to train N_M models (N_M = 10 in this work), and these trained models are then used to estimate the energy uncertainty $\mathcal{U}$ of all the unlabeled structures. Subsequently, the feature distances $\mathcal{D}$ is the distance in feature space between the unlabeled structures and all the selected structures are computed. Both energy uncertainty and feature distance metrics will be explained in detail in the latter part of this section. Using the energy uncertainty and feature distance metrics, a combined score $\mathcal{S}$ is determined for each unlabeled structure. The unlabeled structure with the highest score is selected, and the distance metrics are recalculated for all the unlabeled structures by including the newly selected structures. The uncertainty metric does not need to be recalculated, as the labeled structures (training data) have not changed. This procedure is repeated until all the structures for a batch, N_{SB} structures per batch, are completed (N_{SB} = 10 structures/batch in this work). As an example, the first structure is selected solely based on the uncertainty. The second structure is selected by the uncertainty and the feature distance from the first structure. The third structure is selected by the uncertainty and the feature distance from the first two structures. Once N_{SB} structures in the batch are selected, all the selected structures will be evaluated using DFT and transferred to the labeled structures. Then a termination criterion will be tested. In this work, the termination criterion is met simply when 3,000 structures were calculated, which can be considered as the given DFT computation budget. If needed, any other user-defined criterion can be applied here. If the termination criterion is not met, the process restarts by retraining the N_M models with the new labeled structures and re-evaluating the uncertainty. This outer loop is repeated until the termination criterion is reached, and a final model is trained on all of the labeled data.

As previously mentioned, the first step in our active learning method is to pre-generate a pool of structures representing the structure space relevant to the physical processes under the anticipated simulation conditions. These structures can be meta-stable states, transition pathways, compressed/expanded structures, etc. along with perturbations to these structures due to thermal fluctuations. The pre-generated structures will be cast into one of three mutually exclusive groups throughout the active learning process: labeled, unlabeled, and selected structures. The labeled structures are those which have been calculated/evaluated

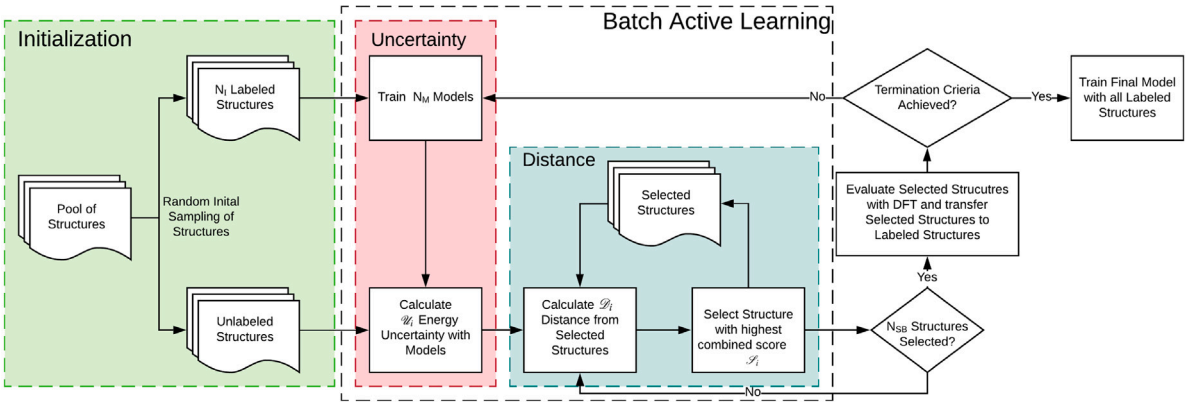

Fig. 1. Schematic of the flow diagram of our batch active learning method.

using DFT, the unlabeled structures are those which have not been calculated, and the selected structures are those unlabeled structures that have been selected by active learning in the current batch and will be labeled by DFT calculation/evaluation. Additional details on our pool of structures and energy calculations can be found in the Methods section.

From all the unlabeled structures, our active learning method will select the next structure based on two criteria: the ith structure's energy uncertainty $\mathcal{U}_i$ and its feature distance $\mathcal{D}_i$ from the current selected structures. The energy uncertainty $\mathcal{U}_i$ is estimated by using a query by bootstrap aggregating (bagging) methodology [34,35]. This uncertainty approach is similar to the methods used in other works to estimate the uncertainty [36–39]. N_M subsamples each having 80% of the labeled structures are sampled using bagging, which are then used to train N_M models. Each model m predicts a potential energy $E_{i,m}$ for every unlabeled structure i. The uncertainty is estimated as the standard deviation between the N_M energy predictions, $\mathcal{U}_i$ as follows,

$$\mathcal{U}_i = \sqrt{\frac{\sum_{m=1}^{N_M} \left(E_{i,m} - \overline{E_i}\right)^2}{N_M - 1}} \tag{1}$$

where the summation is taken over all N_M models. Intuitively, the more uncertain the energy predictions of the structure are, the more difficult the reliable prediction can be achieved. Therefore, obtaining the DFT calculated labels (*i.e.* energy in this work) and adding the corresponding labeled structures have a better chance of improving machine learning predictions overall than the classical machine learning settings with random sampling of training data.

Selecting multiple structures at each batch allows for running DFT calculations in parallel to label the selected structures, significantly reducing the total time to generate the interatomic potential model. However, selecting multiple structures by using exclusively the energy uncertainty as the utility metric/acquisition function may result in redundant information from selecting similar structures which are expected to have similar energy uncertainty. A simple example of this would be if the same structure appears twice in the dataset. Both structures would have identical energy uncertainty and therefore be selected together using only the energy uncertainty. However, these two structures would be completely redundant and provide no new information. To remedy this issue, we use a distance metric $\mathcal{D}$ to penalize the structures that are close to the current selected structures in the feature space $\vec{\chi}$. The distance metric for the ith unlabeled structure, $\mathcal{D}_i$, is the summation of the Euclidean distances between the ith unlabeled structure's feature vector, $\vec{\chi}_i$, and the jth selected structure's feature vector, $\vec{\chi}_j$, summed over all the current selected structures j, *i.e.* ,

$$\mathcal{D}_i = \sum_j \left\| \vec{\chi}_i - \vec{\chi}_j \right\|. \tag{2}$$

This $\mathcal{D}$ metric will be small for the structures that are similar to the current selected structures.

Using the energy uncertainty $\mathcal{U}$ and feature distance $\mathcal{D}$, the algorithm will iteratively select the unlabeled structure with the highest score $\mathcal{S}_i$, which is a weighted sum of $\mathcal{U}_i$ and $\mathcal{D}_i$ normalized by their mean value over all unlabeled structures, $\overline{\mathcal{U}}$ and $\overline{\mathcal{D}}$. That is,

$$\mathcal{S}_i = \alpha \left(\frac{\mathcal{U}_i}{\overline{\mathcal{U}}} \right) + \beta \left(\frac{\mathcal{D}_i}{\overline{\mathcal{D}}} \right). \tag{3}$$

In this work, α was 0.9 and β was 0.1. The weighting factor was chosen to minimize the model uncertainty while generating a stable model. However, in our testing we found that the weighting factors had minimal effect on the accuracy of the model when the α-to-β coefficient ratio varies between 0.9:0.1 and 0.4:0.6. As an example of our active learning method, the first structure is selected purely based on energy uncertainty $\mathcal{U}$. The second structure is selected by the energy uncertainty $\mathcal{U}$ and the feature distance $\mathcal{D}$ from the first structure. The third structure is selected by the uncertainty and the feature distance from the first two structures, and so on. This procedure is repeated for as many structures as desired, and all the selected structures are then labeled by first-principles DFT calculations. Once N_{SB} structures are selected and calculated, a batch is completed and the selection process for the next batch will be restarted by recalculating the energy uncertainty of N_M models. The stopping criteria can be decided by the user, e.g using a convergence threshold of desired physical properties. In this work, the stopping criterion was set to 300 batches (*i.e.* $N_B = 300$), corresponding to 3,000 structures (*i.e.* $N_B \times N_{SB} = 300 \times 10 = 3,000$).

Adopting this batch active learning method, we successfully created a potential for monolayer GeSe using less than 25% (*i.e.* 3,000) of the total 13,006 structures from DFT calculations generated by Yang et al. [33]. By examining the training and test error, we found that our active learning method tends to select structures that are difficult to predict as expected. As will be shown below, the models trained on these difficult structures result in a more consistent and accurate potential with well-converged model coefficients. Subsequently, we validate this approach by applying the active learning potential to predict atomic forces and comparing it with the DFT-calculated results. For comparison, each test was performed using both the above active learning approach and a random sampling approach for N_S times with different random starting conditions (N_S=10 in this work). Finally we successfully predict GeSe's phase transition using this active learning potential in a MD simulation.

2.2. Interatomic potential

The potential form adopted here is based on the AGNI interatomic potential [11]. We consider two types of structural descriptors, including two-body descriptors V_i^{2b} and three-body descriptors V_i^{3b}, with $E = V_i^{2b} + V_i^{3b}$. That is, the total energy is the sum of two-body and three-body energies. V_i^{2b} only depends on the bond length, while V_i^{3b}

Table 1
Parameters for the machine learning interatomic potential.

Parameters	#1	#2	#3	#4	#5	#6	#7	#8
η_2 (Å)	1.0	1.329	1.766	2.347	3.120	4.146	5.510	7.333
κ (Å^{-1})	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
η_3 (Å)	1	1.766	3.120	5.510	–	–	–	–
$f(\cos(\theta_{jik}))$	$\cos^3(\theta_{jik})$	$\cos^2(\theta_{jik}) - 4\cos^4(\theta_{jik})$	$1 - \frac{4}{3}\cos^2(\theta_{jik})$	–	–	–	–	–

depends on both bond lengths and bond angles. The explicit forms of V_i^{2b} and V_i^{3b} are given below,

$$V_i^{2b} = \sum_{j \neq i} e^{-r_{ij}/\eta_2} \cdot \cos(k \cdot r_{ij}) \cdot f_c(r_{ij}). \tag{4}$$

$$V_i^{3b} = \sum_{j,k \neq i} e^{-(r_{ij}^2 + r_{ik}^2)/\eta_3^2} \cdot f(\cos(\theta_{jik})) \cdot f_c(r_{ij}) \cdot f_c(r_{ik}) \tag{5}$$

where

$$f_c(r_{ij}) = \begin{cases} \frac{1}{2}\left[\cos\left(\frac{\pi r_{ij}}{R_c}\right) + 1\right], & \text{for } r_{ij} < R_c \\ 0, & \text{for } r_{ij} \geq R_c. \end{cases} \tag{6}$$

Here, r_{ij} stands for the bond length between atom i and atom j. $f_c(r_{ij})$ is a radial cutoff function, which is zero for $r_{ij} \geq R_c$. In this work, R_c is set to 5.5 Å. Furthermore, θ_{jik} stands for the bond angle between two bonds r_{ij} and r_{ik}. η_2, η_3, and κ are three model parameters in the above two exponential functions and the cosine function, respectively. $f(\cos(\theta_{jik}))$ in the V_i^{3b} descriptor are angle-dependent functions. η_2, κ, η_3, and $f(\cos(\theta_{jik}))$ are provided in Table 1.

For GeSe monolayer, V_i^{2b} consists of 3 two-body families of bond length dependent energy contribution [33], including Ge–Ge, Ge–Se, and Se–Se pairs [8,40–44]. In each two-body family, 8 sets of model parameters (η_2, κ) are used. For the 3 two-body families, this leads to total $3 * 8 = 24$ V_i^{2b} descriptors.

V_i^{3b} consists of 8 three-body families of bond length and bond angle dependent energy contribution, including ∠Ge–Ge–Ge, ∠Ge–Ge–Se, ∠Ge–Se–Ge, ∠Se–Ge–Ge, ∠Ge–Se–Se, ∠Se–Ge–Se, ∠Se–Se–Ge, and ∠Se–Se–Se. For GeSe monolayer, we use 3 $f(\cos(\theta_{jik}))$ functions and 4 η_3 model parameters for each three-body family. For 8 three-body families, this leads to $3 * 4 * 8 = 96$ three-body V_i^{3b} descriptors. In addition, we include a constant as an additional descriptor. In total, we therefore have $24 + 96 + 1 = 121$ descriptors, resulting in a feature vector with length of 121.

For the machine learning model, we use kernel ridge regression (KRR) with a linear kernel using the 121 highly-nonlinear features discussed above. The KRR was implemented through `sklearn` [45] with the regularization coefficient set at 0.0001. This model was fit to the average energy per atom, *i.e.* , the total energy divided by the number of atoms in the supercell.

2.3. First-principles dataset

In this work, we use the dataset from Yang et al. [33] which was generated by using the Vienna ab-initio simulation package (VASP) [46,47] based on first-principles DFT approach [48,49] with Perdew–Burke–Ernzerhof (PBE) [50] exchange–correlation function within the generalized gradient approximation (GGA) [51,52]. According to the paper by Yang et al. [33], the structure pool was created by *ab initio* MD simulations of a supercell with 32 Ge and 32 Se atoms at constant volume and temperature (50 K, 300 K, 500 K and 800 K) for 6,000 time steps with a time step of 3 fs. The training database also contains the relaxed structures at 0 K under strain-free conditions as well as under various uniaxial and biaxial strains. In addition, it includes a transition pathway of a 90° domain switching where a linear interpolation scheme was adopted to generate 100 structures along the domain switching pathway. In total, the structure pool contains 13,006 structures.

2.4. Molecular dynamics simulations

To apply the active learning potential in MD simulations, we constructed a monolayer of GeSe in a 20×20 supercell with each unit cell containing two Ge atoms and two Se atoms. Thus, there are 800 Ge atoms and 800 Se atoms in this model. MD simulations were carried out using `LAMMPS` [53] with the trained active learning potential to simulate the temperature induced phase transition in monolayer GeSe. In the MD simulation, monolayer GeSe was relaxed and equilibrated in an NPT ensemble at 50 K for 40 ps. We then used a step-heating method to heat the system from 50 K to 450 K with a temperature step of 50 K. At each temperature, the structure was relaxed for 40 ps. All the potential energies and lattice constants were obtained by averaging the data in the last 20 ps at each temperature.

3. Results

Our batch active learning process selects the structures with high energy uncertainty and structural dissimilarity, which are relatively difficult to predict among the pool of candidate structures. As shown in Fig. 2(a), the training error is higher than the error on the un-sampled data for the active learning potential, which has a different trend compared to the random sampling procedure as shown in Fig. 2(b). This confirms our expectation that the structures in the training data selected by active learning dependent on the prediction uncertainty are more difficult to predict than the structures in the remaining un-sampled data. It is also clear from Fig. 2 that the error on the un-sampled data during the batch active learning procedure across all the batches is consistently lower than the one by random sampling, demonstrating the effectiveness of our batch active learning procedure.

By directly examining the model coefficients, our potential is shown to be more consistent and is more accurate as expected. As mentioned in Section 2.2, the model used in this paper is a kernel ridge regression with a linear kernel of 121 model coefficients ($\vec{C}$) associated with 121 descriptors, including 24 two-body structure descriptors, 96 three-body structure descriptors, and one constant. Fig. 3(a) and Fig. 3(b) show the Euclidean distance ($\left\|\vec{C} - \vec{C}_{all}\right\|$) between our model coefficients from active learning ($\vec{C}$) and the model coefficients trained on all 13,006 structures ($\vec{C}_{all}$), respectively. For most batches, our active learning potentials have less variance in their coefficients when compared to the potential trained with random sampling. Moreover, with higher batch numbers (e.g. $N_B > 200$), all the active learning potentials converge to similar coefficients, whereas the potentials trained with random sampling still have significant variance among the models. On average, the active sampling potentials are closer to the potential trained on all the data. As these are data-driven potentials, the model trained on all the data is expected to be the best predictor for the data. As our model coefficients converge closer to the model trained on all the data, we expect our model to be more accurate on the represented population. In Fig. 3(c) and Fig. 3(d), we also show that the potential's coefficients converge with active sampling, whereas with the random sampling the potential's coefficients still vary significantly even after 300 batches. All 121 coefficients are shown in Supplementary Fig. S1 and Fig. S2. The large variation from the potentials trained on randomly-sampled structures is particularly worrisome, as it could result in different predictions for the same simulations by using random sampling based interatomic potential. In contrast, all 121 coefficients of the 10 active sampling based interatomic potentials quickly converge with ~2,000

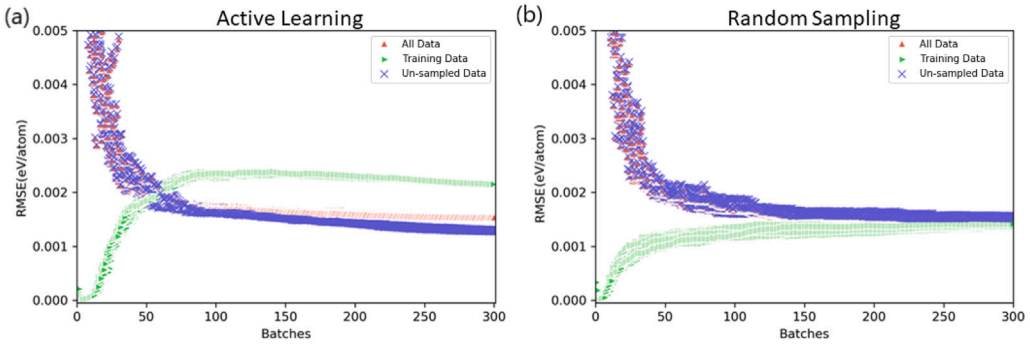

Fig. 2. Root mean square error (RMSE) of the potential energy based on **(a)** active learning potential and **(b)** random sampling potential as a function of the number of batches for N_S different starting conditions ($N_S = 10$). "All Data" includes all 13,006 structures in the pool of candidate structures. "Training Data" are the labeled structures ($N_B{*}N_{SB}$ structures for batch number N_B). "Un-sampled Data" are the unlabeled structures (13,006-$N_B{*}N_{SB}$ structures for batch number N_B). The ratio of training to un-sampled data depends on the batch number.

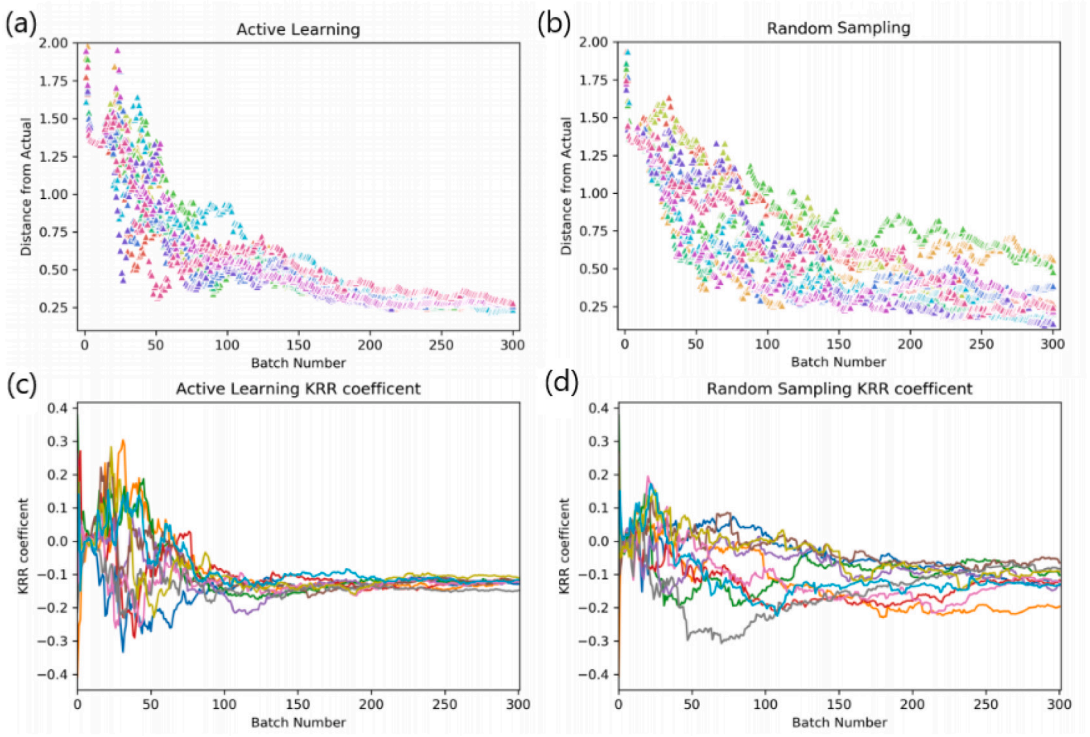

Fig. 3. (a) Distance between the coefficients from the active learning models and the coefficients of the model trained on all the data as a function of the number of batches for N_S different starting conditions ($N_S = 10$). (b) Distance between the coefficients from the random sampling models and the coefficients of the model trained on all the data as a function of the number of batches for N_S different starting conditions. (c) One of the 121 model coefficients from the active learning model as a function of the number of batches for N_S different starting conditions. (d) One of the 121 model coefficients of the random sampling model as a function of the number of batches for N_S different starting conditions. Each color in (a), (b), (c), and (d) represents a different starting condition.

structures (*i.e.* ~200 batches), implying that our active learning method generates interatomic potentials that are more consistent and robust than the ones from the random sampling approach.

Using our active learning method, the forces are more consistent and accurate, and have fewer outliers when compared to random sampling, as can be seen in Fig. 4. In Fig. 4(a), the active learning potentials have a smaller root mean square error (RMSE) and have significantly less variance in the RMSE when predicting the forces than using the potentials from random sampling. Furthermore, Fig. 4(b) and Fig. 4(c) show the comparison between the forces from DFT calculations and the forces predicted by the potentials trained on 1,000 structures from active learning (Fig. 4(b)) and random sampling (Fig. 4(c)). It demonstrates that random sampling has a significant number of outliers in the predictions when trained on 1,000 structures, whereas the active learning has no outliers in the predicted forces across the whole range. Both results indicate that our active learning method consistently produces accurate and robust models. The robustness is especially important as outliers in the force predictions can result in different dynamical processes and atomic trajectories in MD simulations.

Besides the improvement of atomic force predictions, the phonon dispersions are also more consistent when using the active learning method, as seen in Fig. 5. Using the active learning method, the predicted phonon dispersions have a small variance for the potentials trained on 1,000 structures, and the variance becomes negligible for the potentials trained on 3,000 structures as shown in Fig. 5(a) and Fig. 5(b). In the case of random sampling, the predicted phonon dispersions have a considerably larger variance for the potentials trained on 1,000 structures which remains significant even for the potentials trained on 3,000 structures as evidenced in Fig. 5(c) and Fig. 5(d). The large variance in the phonon dispersions indicates different random

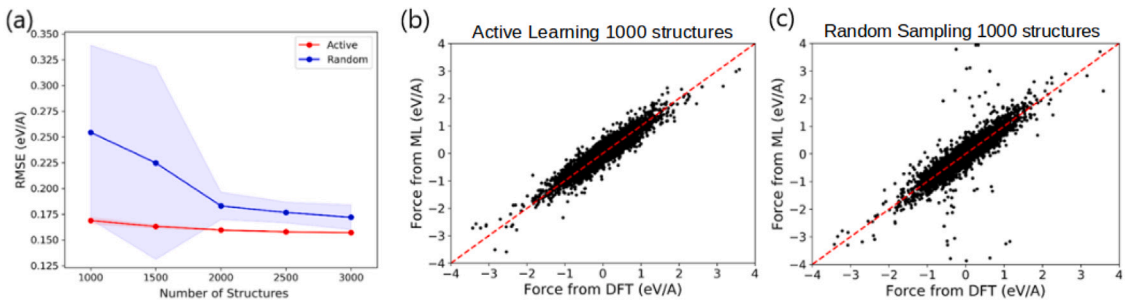

Fig. 4. (a) Root mean square error (RMSE) of the atomic forces predicted by active learning and random sampling as a function of the number of sampled structures. The shaded region indicates the standard deviation based on N_S different starting conditions ($N_S = 10$). (b), (c) Comparison between the forces from DFT calculations and the predicted forces from the potential trained on (b) 1,000 actively-sampled structures and (c) 1,000 randomly-sampled structures.

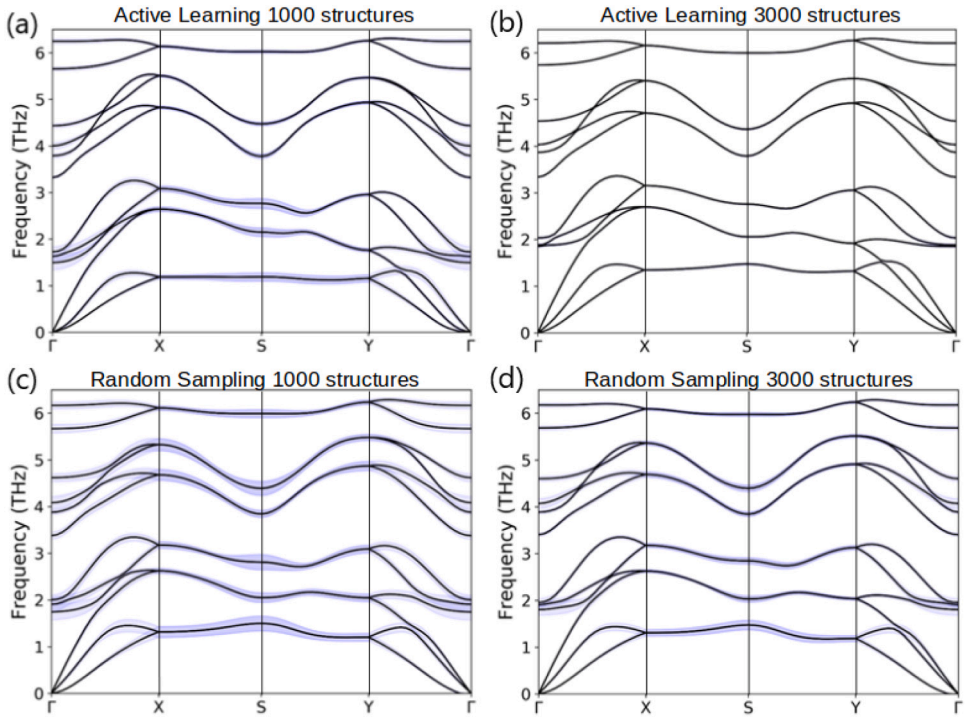

Fig. 5. (a), (b) Phonon dispersion predicted by the potential trained on (a) 1,000 and (b) 3,000 actively sampled structures. (c), (d) Phonon dispersion predicted by the potential trained on (c) 1,000 and (d) 3,000 randomly sampled structures. The shaded blue region indicates the standard deviation from N_S different starting conditions ($N_S = 10$).

sampling may significantly affect the finite temperature simulations and result in very different predictions of thermal and thermodynamic properties.

To further demonstrate our active learning potential, we carry out MD simulations to predict the phase transition temperature of monolayer GeSe, which is the temperature where GeSe changes from an orthorhombic structure to a cubic structure and loses its ferroic order. The phase transition temperature was determined by simulating monolayer GeSe using MD at different temperatures and then calculating the potential energy and lattice constants. The results are shown in Fig. 6(a) where the potential energy becomes discontinuous between 300 K and 350 K. The same discontinuity is observed in the average lattice constants along x and y directions in Fig. 6(b). The lattice constants in the x and y directions converge to the same value, implying a cubic lattice on average corresponding to a transformation to a cubic phase. These discontinuities indicate a first-order phase transition occurring between 300 K and 350 K, which is in agreement with previous results [33,54]. Our active learning approach provides an effective approach to bridge first-principles DFT to large scale atomistic simulations with robust and consistent active learning potentials.

4. Discussion

Our active learning method could be particularly useful when a pre-sampled pool of structures are available in advance. This may occur when in-depth knowledge of the physics and corresponding simulation environment (*e.g.* strain/temperature, phase transitions, or other environmental conditions) is available, and one can therefore generate the relevant structures. One example may be a transition pathway where one can interpolate the initial, final, and other relevant intermediate structures. Meanwhile, more databases are becoming available to the community which sometimes contain a pre-sampled and well-sampled pool of structures. However, the databases may be unintentionally biased towards some specific structure space. In these cases, this active learning strategy could help achieve machine learning potentials with reduced bias.

However, in real-world scenarios a sampled pool of structures very often does not exist. One therefore has to iteratively generate and sample new structures in order to explore PES. In this case, if a classical potential exists, one could use this classical potential in MD or Monte Carlo (MC) simulations to generate a pool of structures. Otherwise, one could first train a machine learning potential on a

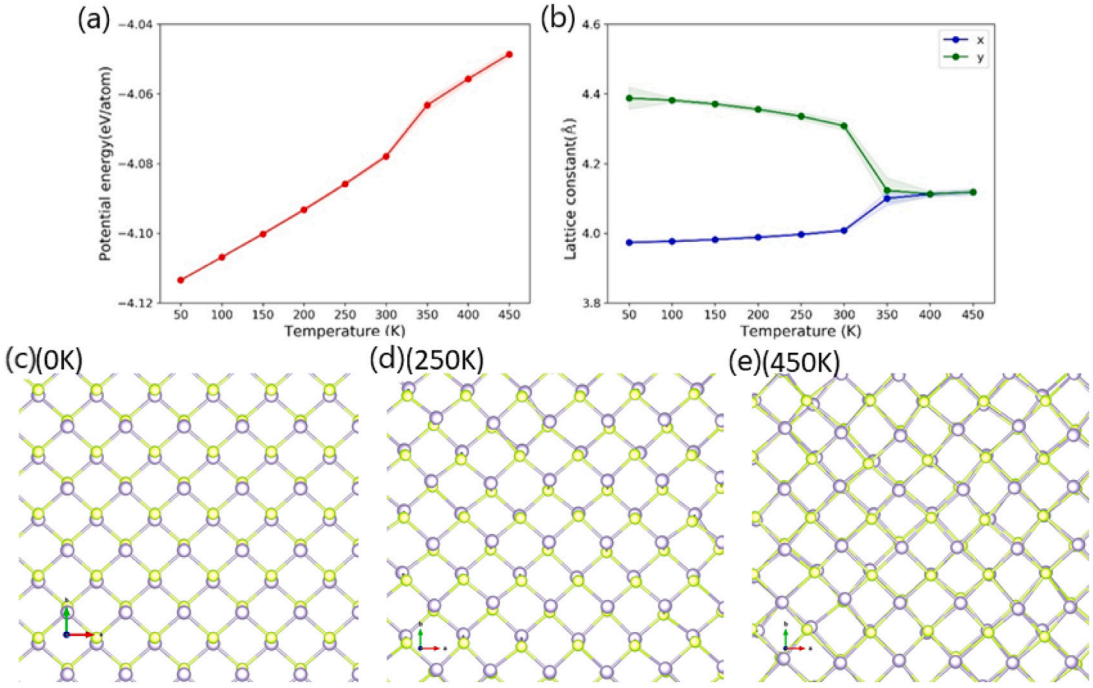

Fig. 6. (a) Temperature dependent potential energy for monolayer GeSe from MD simulation under NPT ensemble using the active learning potential. (b) Temperature dependent lattice constant in the x and y direction for monolayer GeSe from MD simulation under NPT ensemble. The shaded region indicates the standard deviation obtained from MD simulation. (c), (d), (e) Small section of monolayer GeSe at 0 K, 250 K, and 450 K, respectively. Full atomistic structures can be found in Supplementary Material.

small set of initial structures from either random sampling approach or advanced sampling methods, and then perform MD/MC simulations to generate new structures and update the machine learning potential with actively sampled structures iteratively. For those cases where the PES is difficult to explore using conventional MD, one may need to use metadynamics [55] to explore free energy surface and sample the structures. Moreover, if little information is known about a material, USPEX [56] or Nested Sampling Monte Carlo method [57] can be used to find initial stable and metastable configurations.

It is worth emphasizing that only the structures sampled by active learning need to be calculated by DFT. This will greatly reduce the computational cost while generating a potential that can accurately simulate the desired environment. If one is interested in simulating a different environment (such as defects, surfaces, or a new metastable phase), then those structures can be easily added to the pool of structures and the active learning can be restarted from there.

Furthermore, the total computational cost of generating an active learning potential is dominated by DFT calculations on the structures *selected by active sampling*. While both the energy uncertainty, $\mathcal{U}$, and feature distance, $\mathcal{D}$, need to be evaluated for every unlabeled structure, these are computationally very cheap. Specifically, for energy uncertainty $\mathcal{U}_i$, the N_M interatomic potential models, instead of DFT calculations, will be used to quickly evaluate the total energy of all unlabeled structures for each batch, which is computationally very efficient. For computing feature distance $\mathcal{D}_i$, the Euclidean distance between the selected structures (N_{SB} structures per batch) and all the unlabeled structures in the feature space will be calculated for each batch, which is computationally efficient as well. Even if the number of the initial structures is large, the overall computational cost is still dominated by the DFT calculations for labeling the N_{SB} selected structures.

It is worth pointing out that our active learning method can generate a model that is more accurate than a model trained on the entire pool of structures under certain conditions. One example would be a pool of structures with most of them located close to equilibrium. If trained on the entire pool of structures, the model would be biased towards the equilibrium structures, and less accurate for describing the structures away from equilibrium. In contrast, our active learning method would avoid the repetitive selection of the structures once they can be well predicted, and select other less similar structures such as those at high temperatures and transition states. Under these conditions, our active learning model would be more accurate than the one trained on the entire pool of structures. The dataset used in this work was mostly generated from AIMD simulations [33], which usually follow the Boltzmann distribution, making it more likely to sample lower energy structures. As a result, the model trained on all the 13,006 structures is intrinsically biased towards low energy structures.

Our active learning method is also model-agnostic and can be implemented with any machine learning model. Other active learning methods often need the machine learning model to provide prediction uncertainty (e.g. using a Gaussian process), but our active learning method estimates the uncertainty internally by using the predictions from multiple models. Many machine learning potentials can be efficiently trained with the proposed active learning method, such as SNAP and ACE [6,7]. However, it is worth noting that the chosen machine learning models have to be trained many times throughout the batch active learning process, so models that are expensive to train, such as neural networks, would increase the computational cost.

On top of being model-agnostic, our active learning method could also use ensemble learning. With ensemble learning, different machine learning models could be trained using all of the data, instead of training the same machine learning model multiple times with bagging. Then the uncertainty would be the standard deviation of the different model predictions [58]. This ensemble approach would avoid overfitting to a single model and explore the configuration space in a more universal way. However, this approach might not produce the model with the smallest error.

5. Conclusions

In summary, we have introduced a pool-based batch active learning method. It produces an accurate, consistent, and robust MD potential using a small amount of structures via active sampling. This is accomplished by directly and iteratively selecting highly uncertain structures

which are difficult to predict. By training on the actively sampled structures, the model coefficients of interatomic potentials quickly and consistently converge to the same coefficients, in contrast to the large variation in the random-sampling based interatomic potentials. In addition, our active learning method selects multiple (N_{SB}) structures in a single batch from energy uncertainty and feature distance *without DFT calculations*. The N_{SB} actively sampled structures can be efficiently labeled by running N_{SB} DFT simulations in parallel, which reduces the total time to generate interatomic potentials. Our active learning method is particularly useful when a pre-sampled pool of structures are available. For the cases without pre-sampled pool of structures, one can use other advanced sampling approaches such as USPEX and Nested Sampling Monte Carlo to sample initial structures, and iteratively label structures, train machine learning potential, generate new structures using MD/MC with the potential, and sample structures using the active learning method. We envision this active learning method will help generate trustworthy and reliable interatomic potentials at a reduced computational cost that are highly valuable for computational materials, physics, and chemistry community.

CRediT authorship contribution statement

Nathan Wilson: Conceptualization, Methodology, Software, Investigation, Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Daniel Willhelm:** Formal analysis, Writing – review & editing. **Xiaoning Qian:** Methodology, Conceptualization, Formal analysis, Writing – review & editing. **Raymundo Arróyave:** Supervision, Funding acquisition, Conceptualization, Writing – review & editing, Formal analysis. **Xiaofeng Qian:** Supervision, Project administration, Funding acquisition, Conceptualization, Writing – original draft, Writing – review & editing, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The raw/processed data required to reproduce these findings cannot be shared at this time due to technical or time limitations.

Acknowledgments

We acknowledge the support by the US National Science Foundation (NSF) under award number OAC-1835690, the US NSF NRT-DESE: Data-Enabled Discovery and Design of Energy Materials (D3EM) under Grant No. 1545403, and the US Air Force Research Laboratory (AFRL) Minority Leaders Program under sub-contract 165852-19F5830-19-02-C1. We also acknowledge Yang Yang for providing the GeSe dataset.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.commatsci.2022.111330.

References

[1] S. Legoas, V. Coluci, S. Braga, P. Coura, S. Dantas, D.S. Galvao, Molecular-dynamics simulations of carbon nanotubes as gigahertz oscillators, Phys. Rev. Lett. 90 (5) (2003) 055504, http://dx.doi.org/10.1103/PhysRevLett.90.055504.
[2] Q. Zeng, X. Jiang, A. Yu, G.M. Lu, Growth mechanisms of silver nanoparticles: a molecular dynamics study, Nanotechnology 18 (3) (2007) 035708, http://dx.doi.org/10.1088/0957-4484/18/3/035708.
[3] A. Fasolino, J. Los, M.I. Katsnelson, Intrinsic ripples in graphene, Nature Mater. 6 (11) (2007) 858, http://dx.doi.org/10.1038/nmat2011.
[4] D. Bacon, F. Gao, Y.N. Osetsky, The primary damage state in fcc, bcc and hcp metals as seen in molecular dynamics simulations, J. Nuclear Mater. 276 (1–3) (2000) 1–12, http://dx.doi.org/10.1016/S0022-3115(99)00165-8.
[5] J. Behler, Perspective: Machine learning potentials for atomistic simulations, J. Chem. Phys. 145 (17) (2016) 170901, http://dx.doi.org/10.1063/1.4966192.
[6] A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, G.J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, J. Comput. Phys. 285 (2015) 316–330, http://dx.doi.org/10.1016/j.jcp.2014.12.018.
[7] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, Phys. Rev. B 99 (1) (2019) 014104, http://dx.doi.org/10.1103/PhysRevB.99.014104.
[8] A.P. Bartók, M.C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, Phys. Rev. Lett. 104 (13) (2010) 136403, http://dx.doi.org/10.1103/PhysRevLett.104.136403.
[9] A.P. Bartók, G. Csányi, Gaussian approximation potentials: A brief tutorial introduction, Int. J. Quantum Chem. 115 (16) (2015) 1051–1057, http://dx.doi.org/10.1002/qua.24927.
[10] A.V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, Multiscale Model. Simul. 14 (3) (2016) 1153–1173.
[11] T.D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, R. Ramprasad, A universal strategy for the creation of machine learning-based atomistic force fields, npj Comput. Mater. 3 (1) (2017) 37, http://dx.doi.org/10.1038/s41524-017-0042-y.
[12] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. 98 (2007) 146401, http://dx.doi.org/10.1103/PhysRevLett.98.146401, URL https://link.aps.org/doi/10.1103/PhysRevLett.98.146401.
[13] J. Behler, Representing potential energy surfaces by high-dimensional neural network potentials, J. Phys.: Condens. Matter 26 (18) (2014) 183001, http://dx.doi.org/10.1088/0953-8984/26/18/183001.
[14] H. Wang, L. Zhang, J. Han, E. Weinan, DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics, Comput. Phys. Comm. 228 (2018) 178–184, http://dx.doi.org/10.1016/j.cpc.2018.03.016.
[15] H. Gassner, M. Probst, A. Lauenstein, K. Hermansson, Representation of intermolecular potential functions by neural networks, J. Phys. Chem. A 102 (24) (1998) 4596–4605, http://dx.doi.org/10.1021/jp972209d.
[16] S. Lorenz, A. Groß, M. Scheffler, Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks, Chem. Phys. Lett. 395 (4–6) (2004) 210–215, http://dx.doi.org/10.1016/j.cplett.2004.07.076.
[17] G.P. Pun, R. Batra, R. Ramprasad, Y. Mishin, Physically informed artificial neural networks for atomistic modeling of materials, Nature Commun. 10 (1) (2019) 2339, http://dx.doi.org/10.1038/s41467-019-10343-5.
[18] A. Hernandez, A. Balasubramanian, F. Yuan, S.A. Mason, T. Mueller, Fast, accurate, and transferable many-body interatomic potentials by symbolic regression, npj Comput. Mater. 5 (1) (2019) 112, http://dx.doi.org/10.1038/s41524-019-0249-1.
[19] E.V. Podryabinkin, E.V. Tikhonov, A.V. Shapeev, A.R. Oganov, Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning, Phys. Rev. B 99 (6) (2019) 064114, http://dx.doi.org/10.1103/PhysRevB.99.064114.
[20] R. Jinnouchi, F. Karsai, G. Kresse, On-the-fly machine learning force field generation: Application to melting points, Phys. Rev. B 100 (2019) 014105, http://dx.doi.org/10.1103/PhysRevB.100.014105, URL https://link.aps.org/doi/10.1103/PhysRevB.100.014105.
[21] J. Vandermause, S.B. Torrisi, S. Batzner, Y. Xie, L. Sun, A.M. Kolpak, B. Kozinsky, On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events, npj Comput. Mater. 6 (1) (2020) 20, http://dx.doi.org/10.1038/s41524-020-0283-z.
[22] K. Miwa, H. Ohno, Interatomic potential construction with self-learning and adaptive database, Phys. Rev. Mater. 1 (5) (2017) 053801, http://dx.doi.org/10.1103/PhysRevMaterials.1.053801.
[23] T. Jacobsen, M. Jørgensen, B. Hammer, On-the-fly machine learning of atomic potential in density functional theory structure optimization, Phys. Rev. Lett. 120 (2) (2018) 026102, http://dx.doi.org/10.1103/PhysRevLett.120.026102.
[24] N. Bernstein, G. Csányi, V.L. Deringer, De novo exploration and self-guided learning of potential-energy surfaces, npj Comput. Mater. 5 (2019) 99, http://dx.doi.org/10.1038/s41524-019-0236-6.
[25] V.L. Deringer, C.J. Pickard, G. Csányi, Data-driven learning of total and local energies in elemental boron, Phys. Rev. Lett. 120 (15) (2018) 156001, http://dx.doi.org/10.1103/PhysRevLett.120.156001.

[26] G. Sivaraman, A.N. Krishnamoorthy, M. Baur, C. Holm, M. Stan, G. Csányi, C. Benmore, A. Vázquez-Mayagoitia, Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide, npj Comput. Mater. 6 (1) (2020) 104, http://dx.doi.org/10.1038/s41524-020-00367-7.

[27] K. Gubaev, E.V. Podryabinkin, G.L. Hart, A.V. Shapeev, Accelerating high-throughput searches for new alloys with active learning of interatomic potentials, Comput. Mater. Sci. 156 (2019) 148–156, http://dx.doi.org/10.1016/j.commatsci.2018.09.031.

[28] E.V. Podryabinkin, A.V. Shapeev, Active learning of linearly parametrized interatomic potentials, Comput. Mater. Sci. 140 (2017) 171–180, http://dx.doi.org/10.1016/j.commatsci.2017.08.031.

[29] A.P. Bartók, R. Kondor, G. Csányi, On representing chemical environments, Phys. Rev. B 87 (18) (2013) 184115, http://dx.doi.org/10.1103/PhysRevB.87.184115.

[30] J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, J. Chem. Phys. 134 (7) (2011) 074106, http://dx.doi.org/10.1063/1.3553717.

[31] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, M. Ceriotti, Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials, J. Chem. Phys. 148 (24) (2018) 241730, http://dx.doi.org/10.1063/1.5024611.

[32] S. De, A.P. Bartók, G. Csányi, M. Ceriotti, Comparing molecules and solids across structural and alchemical space, Phys. Chem. Chem. Phys. 18 (20) (2016) 13754–13769, http://dx.doi.org/10.1039/C6CP00415F.

[33] Y. Yang, H. Zong, J. Sun, X. Ding, Rippling ferroic phase transition and domain switching in 2D materials, Adv. Mater. 33 (49) (2021) 2103469, http://dx.doi.org/10.1002/adma.202103469.

[34] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140, http://dx.doi.org/10.1007/BF00058655.

[35] B. Settles, Active learning literature survey, University of Wisconsin-Madison Department of Computer Sciences, 2009, URL http://digital.library.wisc.edu/1793/60660.

[36] J.S. Smith, B. Nebgen, N. Lubbers, O. Isayev, A.E. Roitberg, Less is more: Sampling chemical space with active learning, J. Chem. Phys. 148 (24) (2018) 241733, http://dx.doi.org/10.1063/1.5023802.

[37] F. Musil, M.J. Willatt, M.A. Langovoy, M. Ceriotti, Fast and accurate uncertainty estimation in chemical machine learning, J. Chem. Theory Comput. 15 (2) (2019) 906–915, http://dx.doi.org/10.1021/acs.jctc.8b00959.

[38] L. Zhang, D.-Y. Lin, H. Wang, R. Car, E. Weinan, Active learning of uniformly accurate interatomic potentials for materials simulation, Phys. Rev. Mater. 3 (2) (2019) 023804, http://dx.doi.org/10.1103/PhysRevMaterials.3.023804.

[39] E. Uteva, R.S. Graham, R.D. Wilkinson, R.J. Wheatley, Active learning in Gaussian process interpolation of potential energy surfaces, J. Chem. Phys. 149 (17) (2018) 174114, http://dx.doi.org/10.1063/1.5051772.

[40] V. Botu, J. Chapman, R. Ramprasad, A study of adatom ripening on an Al(111) surface with machine learning force fields, Comput. Mater. Sci. 129 (2017) 332–335, http://dx.doi.org/10.1016/j.commatsci.2016.12.007.

[41] M.A. Caro, V.L. Deringer, J. Koskinen, T. Laurila, G. Csányi, Growth mechanism and origin of high sp^3 content in tetrahedral amorphous carbon, Phys. Rev. Lett. 120 (16) (2018) 166101, http://dx.doi.org/10.1103/PhysRevLett.120.166101.

[42] Z. Li, J.R. Kermode, A. De Vita, Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces, Phys. Rev. Lett. 114 (9) (2015) 096405, http://dx.doi.org/10.1103/PhysRevLett.114.096405.

[43] P. Rowe, G. Csányi, D. Alfè, A. Michaelides, Development of a machine learning potential for graphene, Phys. Rev. B 97 (5) (2018) 054303, http://dx.doi.org/10.1103/PhysRevB.97.054303.

[44] A. Seko, A. Takahashi, I. Tanaka, First-principles interatomic potentials for ten elemental metals via compressed sensing, Phys. Rev. B 92 (5) (2015) 054113, http://dx.doi.org/10.1103/PhysRevB.92.054113.

[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[46] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (1996) 11169–11186, http://dx.doi.org/10.1103/PhysRevB.54.11169, URL https://link.aps.org/doi/10.1103/PhysRevB.54.11169.

[47] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. 6 (1) (1996) 15–50, http://dx.doi.org/10.1016/0927-0256(96)00008-0.

[48] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (3B) (1964) B864, http://dx.doi.org/10.1103/PhysRev.136.B864.

[49] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (4A) (1965) A1133, http://dx.doi.org/10.1103/PhysRev.140.A1133.

[50] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (18) (1996) 3865, http://dx.doi.org/10.1103/PhysRevLett.77.3865.

[51] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A 38 (6) (1988) 3098, http://dx.doi.org/10.1103/PhysRevA.38.3098.

[52] D.C. Langreth, M. Mehl, Beyond the local-density approximation in calculations of ground-state electronic properties, Phys. Rev. B 28 (4) (1983) 1809, http://dx.doi.org/10.1103/PhysRevB.28.1809.

[53] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. 117 (1) (1995) 1–19, http://dx.doi.org/10.1006/jcph.1995.1039.

[54] M. Mehboudi, B.M. Fregoso, Y. Yang, W. Zhu, A. van der Zande, J. Ferrer, L. Bellaiche, P. Kumar, S. Barraza-Lopez, Structural phase transition and material properties of few-layer monochalcogenides, Phys. Rev. Lett. 117 (24) (2016) 246802, http://dx.doi.org/10.1103/PhysRevLett.117.246802.

[55] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, et al., PLUMED: A portable plugin for free-energy calculations with molecular dynamics, Comput. Phys. Comm. 180 (10) (2009) 1961–1972.

[56] C.W. Glass, A.R. Oganov, N. Hansen, USPEX—Evolutionary crystal structure prediction, Comput. Phys. Comm. 175 (11–12) (2006) 713–720.

[57] L.B. Pártay, A.P. Bartók, G. Csányi, Efficient sampling of atomic configurational spaces, J. Phys. Chem. B 114 (32) (2010) 10502–10512.

[58] J.Z. Liu, J. Paisley, M.-A. Kioumourtzoglou, B. Coull, Accurate uncertainty estimation and decomposition in ensemble learning, 2019, arXiv preprint arXiv:1911.04061.