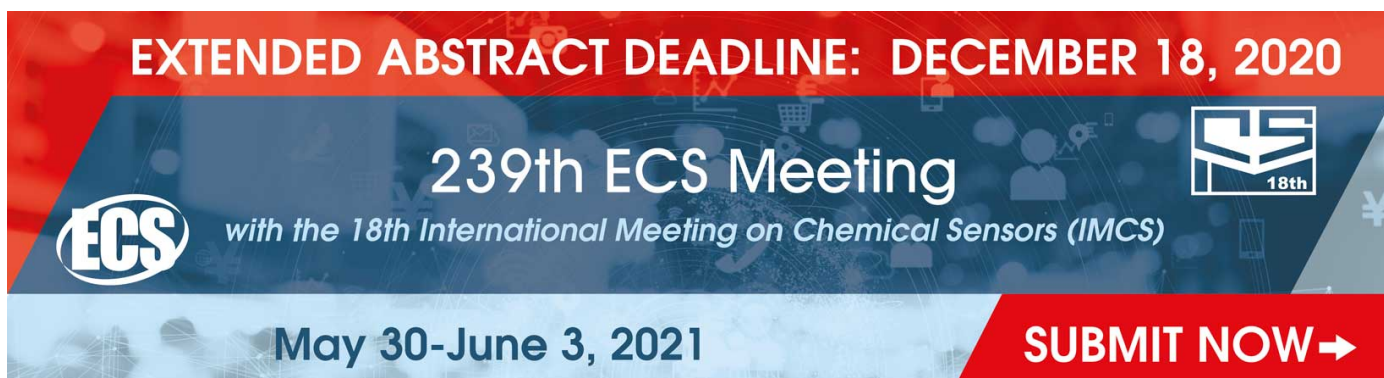


TOPICAL REVIEW • OPEN ACCESS

## High-dimensional neural network atomic potentials for examining energy materials: some recent simulations

To cite this article: Satoshi Watanabe *et al* 2021 *J. Phys. Energy* **3** 012003

View the [article online](#) for updates and enhancements.



**EXTENDED ABSTRACT DEADLINE: DECEMBER 18, 2020**

**239th ECS Meeting**  
with the 18th International Meeting on Chemical Sensors (IMCS)

**May 30-June 3, 2021**

**SUBMIT NOW →**

The banner features a red top section with white text, a dark blue middle section with white and light blue text and icons, and a light blue bottom section with white text. The ECS logo is on the left, and a stylized 'ECS' logo with '18th' is on the right. The background includes faint icons of a shopping cart, a person, and a yen symbol.



## TOPICAL REVIEW

## OPEN ACCESS

RECEIVED  
30 June 2020REVISED  
29 October 2020ACCEPTED FOR PUBLICATION  
5 November 2020PUBLISHED  
15 December 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



# High-dimensional neural network atomic potentials for examining energy materials: some recent simulations

Satoshi Watanabe<sup>1,5</sup> , Wenwen Li<sup>2</sup> , Wonseok Jeong<sup>3</sup> , Dongheon Lee<sup>3</sup> , Koji Shimizu<sup>1</sup> , Emi Mimanitani<sup>4</sup> , Yasunobu Ando<sup>2</sup> and Seungwu Han<sup>3</sup>

<sup>1</sup> Department of Materials Engineering, The University of Tokyo, Tokyo 113-8656, Japan

<sup>2</sup> Research Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology, Ibaraki 305-8568, Japan

<sup>3</sup> Department of Materials Science and Engineering, Seoul National University, Seoul 08826, Republic of Korea

<sup>4</sup> Institute for Molecular Science, Aichi 444-8585, Japan

<sup>5</sup> Author to whom any correspondence should be addressed.

E-mail: [watanabe@cello.t.u-tokyo.ac.jp](mailto:watanabe@cello.t.u-tokyo.ac.jp)

Keywords: neural network, machine learning, ion migration, thermal transport

## Abstract

Owing to their simultaneous accuracy and computational efficiency, interatomic potentials machine-learned using first-principles calculation data are promising for investigating phenomena closely related to atomic motion in various energy materials. We have been working with one type of these potentials, high-dimensional (HD) neural network potentials (NNPs), and their applications, but we realized that our current understanding of HD NNPs, e.g. the meaning of the atomic energy mapping, remained insufficient, and that tuning their prediction performance for different target properties/phenomena often requires much trial and error. In this article, we illustrate the usefulness of NNPs through our studies on ion migration and thermal transport in energy and related materials. We also share our experiences with data sampling and training strategies and discuss the meaning of atomic energy mapping in HD NNPs.

## 1. Introduction

In the research and development of novel energy devices, including information devices with low energy consumption, a deeper understanding of atomic behaviours on the nanoscale has become increasingly important because of their crucial roles in device operation and/or energy consumption. Controlling atomic behaviours through the design of materials and devices is highly desirable to further improve device performance. For example, for resistive switching memory devices, atomic migration is key to the formation and breaking of conductive paths [1, 2] as it determines the switching speed and retention time. In the case of power semiconductor devices for the efficient use of electric power, thermal management is crucial [3], which requires an understanding of the relationship between defects and phonon properties. For secondary batteries, understanding ion motion is obviously important because ion motion causes charging and discharging.

Recently, the development of sophisticated methodologies and high-performance computers have made first-principles calculations a powerful tool for understanding such phenomena. These advances have enabled, for example, the prediction of migration paths and corresponding energy profiles, phonon band structures, and thermal conductivities with satisfactory reliability for ideal systems (e.g. perfect crystals). However, first-principles methods still require heavy computational resources to examine such properties in realistic systems. On the other hand, conventional interatomic potentials, also called force fields, are computationally much more efficient, but their prediction performance is often low.

Considering this situation, interatomic potentials constructed with first-principles calculation data and machine-learning (ML) techniques have attracted considerable attention because they can simultaneously achieve accuracy and computational efficiency. Potentials constructed using various ML techniques [4–10] have already been applied to a wide range of systems and phenomena [11–15].

The high-dimensional (HD) neural network potential (NNP) proposed by Behler and Parrinello [4], one such ML interatomic potential, seems to be promising for complex systems and phenomena because of its flexibility, which arises from its massive degrees of freedom, although some other approaches may be similarly or more promising. In this light, HD NNPs (hereafter referred to simply as NNPs unless this notation could cause confusion) would be suitable for studies on atomic behaviours in novel energy materials because the interatomic potentials in such systems must describe interatomic interactions in a wide variety of local environments. On the other hand, the usefulness of NNPs in such situations should be carefully examined. The flexibility of NNPs does not guarantee a high prediction performance for the target physical properties and phenomena, and it may also involve an increase in the computational load owing to the increased number of parameters and the decreased predictive accuracy because of overfitting.

With these considerations in mind, our goal has been to construct NNPs for various energy materials and explore the possibilities of this technique. In this article, we describe our recent attempts to construct and apply NNPs, and different co-authors of the present article collaborated in each described attempt. After briefly overviewing NNPs in section 2, we present several techniques for reducing the number of parameters/hyperparameters and/or costs of ML and simulations in section 3. In section 4, the results of simulations with NNPs on two topics—ion migration and thermal conductivities in energy materials—are presented. Then, we discuss the meaning and role of atomic energies in NNPs in section 5, and concluding remarks are given in section 6.

## 2. Brief overview of neural network potentials

The NNP methods adopted in the studies presented in this article are essentially the same as the one proposed by Behler and Parrinello [4]. Because several review articles have already described this method [14, 16], we overview it briefly here.

The basic idea of the NNP method is to relate atomic arrangement information to the total energy using one or more neural networks (NNs). In a conventional empirical potential, the input (atomic structural information) and output (energy) are connected using relatively simple analytical functions. In NNPs, such functions are simply replaced with NNs. An NN (more specifically, a multi-layer perceptron) consists of input and output layers and one or more hidden layers in between. Each layer consists of several nodes, and the respective nodes are connected to nodes in the adjacent layers. In addition, we sometimes consider a bias node. Each node has a value calculated from the values of nodes in the previous layer and the bias node via activation functions. The activation functions are introduced to enable fitting arbitrary functions [16] and are often nonlinear functions such as a hyperbolic tangent. In total, taking as an example a case with two hidden layers and a bias node, the output value is expressed as follows.

$$E = f^{\text{out}} \left[ w_{b1}^{\text{out}} + \sum_{n2}^{k=1} w_{k2}^{\text{out}} f^2 \left\{ w_{bk}^2 + \sum_{n1}^{j=1} w_{j1}^2 f^1 \left( w_{bj}^1 + \sum_{n0}^{i=1} w_{ij}^1 G_i \right) \right\} \right], \quad (1)$$

where 0,  $i$ , and out denote the input,  $i$ th hidden, and output layers, respectively;  $b$  is the bias node;  $f^\alpha$  denotes the activation function which produces the values of nodes in the  $\alpha$ th layer; and  $w_{ij}^\alpha$  ( $w_{bj}^\alpha$ ) denotes the weight parameter connecting node  $i$  in layer  $\alpha-1$  (the bias node) to node  $j$  in layer  $\alpha$ , which should be optimised through the training process.

An important feature of the HD NNPs is the representation of the total energy as a sum of the atomic energies, wherein each atomic energy is expressed via an NN. As a result, the same potential can be used for systems with different numbers of atoms as long as the constituent elements are the same.

When we use NNPs, we must choose a descriptor set to represent the information of atomic arrangement. That is, the values of the input nodes are the values of descriptors corresponding to a specific atomic structure. Cartesian coordinates are not suitable as such descriptors, because their values are not invariant when the system is translated or rotated or when the same kinds of atoms are exchanged. The symmetry functions (SFs) introduced by Behler and Parrinello [4] are often used in NNPs for this purpose, as they guarantee the invariance of the NNP in these situations. Note that although SFs seem to be the most popular descriptors for NNPs, other descriptors with the same invariance, such as smooth overlap of atomic positions (SOAPs) [17], bispectrum [6], and Chebyshev [18] and Zernike [19] radial distributions, can be used instead of an SF.

The following is an example of an SF:

$$G^1 = \sum_i e^{-\eta R_i^2} f_c(R_i), \quad (2)$$

where  $R_i$  is the distance between the central atom and atom  $i$ ;  $\eta$  is a hyperparameter which should be determined before training the NN; and  $f_c(r)$  is the cut-off function, which becomes zero for  $r$  larger than the cut-off distance. Note that the introduction of the cut-off distance is justified by the nearsightedness principle [20], although in some cases the explicit consideration of long-range interactions may be necessary.

After setting the structure of the NNs (including the choice of descriptors) and their hyperparameters, the NNP is trained to accurately reproduce the density functional theory (DFT) calculation data ('training data'). More specifically, the loss function is minimised. By using the total energies and forces in the training, the loss function can be expressed as follows:

$$L = \alpha \times \text{MSE}(\{F_i'\}) + (1 - \alpha) \times \text{MSE}(\{E_i\}), \quad (3)$$

where  $\text{MSE}(\{F_i'\})$  and  $\text{MSE}(\{E_i\})$  denote the sums of the mean square errors (MSEs) of the forces and energies, respectively, between the prediction by the NNP and the corresponding first-principles calculation data.

Notably, several open source codes are available such as aenet [21], PiNN [22], and Amp [23] besides those that we developed (SIMPLE-NN [24], hdnnp [25], and DIMP [26]).

### 3. Techniques for the smart construction of potentials

Several groups have demonstrated that it is possible to construct NNPs that agree well with first-principles calculations within the training domain. The next concerns are (a) reducing the complexity to minimise the computational cost of simulations using NNPs and (b) reducing the cost of training while maintaining the prediction accuracy of NNPs. Here, we focus on these issues.

#### 3.1. Hyperparameter optimisation

The most obvious way to reduce the complexity of NNPs is to optimise the NN hyperparameters, that is, the number of hidden layers, the number of nodes in each layer, and the choice of activation function. This has been discussed in, for example [27, 28]. Examples of investigations into NNP accuracy depending on these parameters can also be found in the supporting information of the papers by Li and Ando [26, 29]. Nevertheless, systematically examining this point seems difficult because the quality of the training dataset greatly affects the accuracy of the NNP. For example, we often achieve sufficient accuracy by adopting two hidden layers, and adding more hidden layers does not always drastically improve the accuracy. However, this may be due to an insufficient number of data to train the complex NNs or inadequate parameter optimisation. For the latter, we emphasise that increasing the number of hidden layers often requires sophisticated optimisation techniques to obtain well-optimised results.

There is also room for optimising the number of SFs and the corresponding hyperparameters such as  $\eta$ . We will discuss the number of SFs in the next subsection. We have not yet systematically investigated optimising the hyperparameters of each SF, but we can say that radial distribution functions would be useful in their optimisation: for interatomic distances with high probabilities of neighbouring atoms, a precise description of the local environment is desirable.

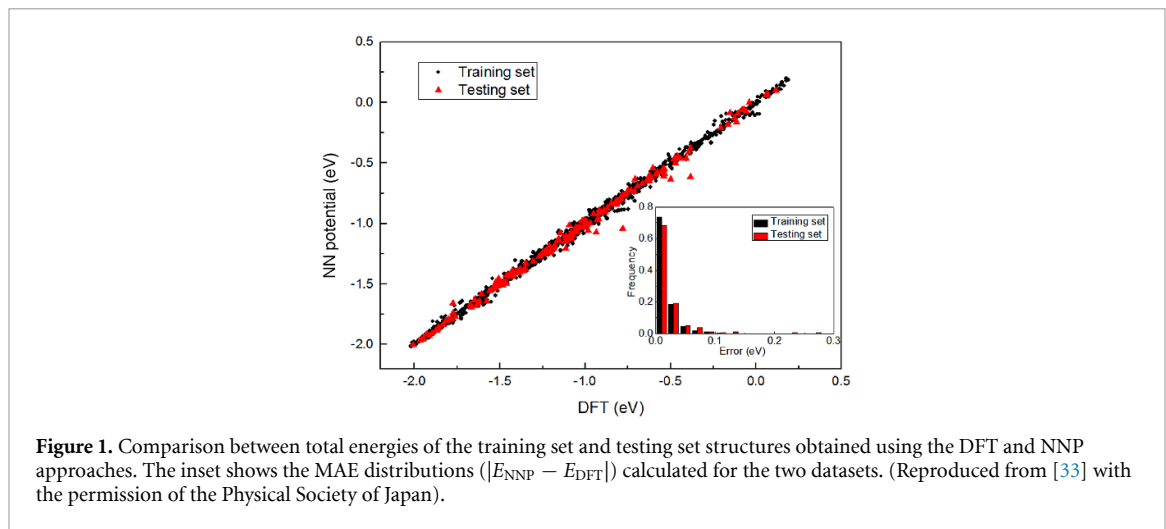
#### 3.2. Smart modelling

Another way to reduce the complexity of NNPs is to simplify the model. Here, we describe two approaches in this category: one using data science techniques and the other based on physical insights.

##### 3.2.1. Simplification by focusing on a target species

Using a data science approach can reduce the number of SFs. For example, principal component analysis (PCA) can illustrate the components which are important for describing the local atomic arrangements in the training data [30]. For instance, we experienced that in reducing the SF components using a PCA, when the number of descriptors was reduced from 41 to 20, the loss of accuracy for the forces and thermal conductivity of crystalline Si was insignificant, whereas further reduction (i.e. to 15 or 10) severely decreased the accuracy.

Other data science approaches can also be used to reduce the number of SFs, e.g. the least absolute shrinkage and selection operator and the genetic algorithm (GA). Li and Ando [31] applied these algorithms to force prediction models as a function of vectorised structure fingerprints based on the SFs for unary (Cu) and binary (SiO<sub>2</sub>) systems. They reduced the number of fingerprints while maintaining the accuracy. Particularly, using the GA enabled decreasing the number of fingerprints for the SiO<sub>2</sub> force prediction model from 2122 to 256 while maintaining a test error of approximately 0.2 eV Å<sup>-1</sup> by linear mixture modelling. Imbalzano *et al* [32] examined CUR decomposition, farthest point sampling, and a Pearson correlation-based method and showed that these approaches are effective in selecting the essential SFs from



large pool of SFs. They also discussed automatic selection of SFs and applied their approach to SOAP fingerprints.

### 3.2.2. Simplification based on physical insights

Another way to reduce the number of SF bases is to consider physical insights. Li *et al* attempted to compose a simplified model in this manner for a diffusing Cu atom to determine all possible diffusion pathways inside an amorphous  $\text{Ta}_2\text{O}_5$  matrix [33]. In this work, they focused on only Cu as a target species instead of constructing a full NNP. They assumed a dilute concentration of diffusing Cu atoms and thus neglected Cu–Cu interactions, and the temperature was low enough to neglect interactions between diffusion and lattice vibrations. Then, the motion of a Cu atom in the  $\text{Ta}_2\text{O}_5$  matrix could be regarded as motion in a ‘medium’. According to this consideration, configuring an NNP for the single species of Cu, which required a small number of parameters, was sufficient. In their study, the numbers of nodes for the input layer, hidden layer 1, hidden layer 2, and output layer were 36, 9, 9, and 1, respectively.

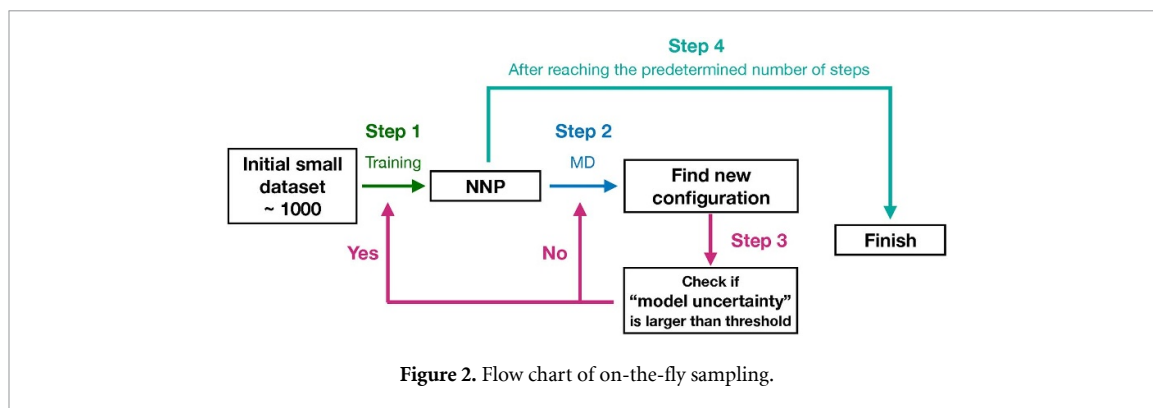
First, the target  $\text{Ta}_2\text{O}_5$  amorphous structure was created, and the energy change  $\Delta E$  for the arrangement of Cu atoms in different locations in the amorphous structure was calculated by DFT. At this time, the Cu atoms were fixed, and the relaxation of the surrounding amorphous structure was considered. As the training data, 2000 points of different Cu arrangements were acquired by random selection from  $50^3$  grid points in the cell, of which 1800 were used for training and 200 were used for model validation. As a result of the training, the root mean square error (RMSE) reached 39 meV/system as shown in figure 1; this is enough to describe the activation barrier of atomic diffusion, which is typically on the order of several hundred millielectronvolts. Using this potential, they optimised the structure  $10^7$  times faster than they did using DFT calculations, making it practically possible to perform a full search for local minima in the amorphous model. The result of the full search is briefly reviewed in section 4.1.1.

Notably, the accuracy of molecular dynamics (MD) simulations using the simplified NNPs must be poor because the simplified NNPs are only a function of descriptors for the guest species and cannot determine the forces acting on atoms in an amorphous host matrix. Fujikake *et al* overcame this limitation by including the difference in the force acting on the host atoms in an ML model to simulate Li diffusion in carbon materials [13].

Even though simplification by focusing on a target species works for MD simulations, dilute conditions must be assumed for the target species to accurately model the applied system. To simulate diffusion processes in dense carrier systems, the NNP should be simplified in a different manner for efficient training. In the following paragraphs, we describe the simulation study conducted by Li *et al* on Li ion diffusion in amorphous  $\text{Li}_3\text{PO}_4$  as an example of reducing the number of input SFs from 27 possible types [34].

To reduce the input dimension, Li *et al* selected the essential SFs by focusing on the structure of  $\text{Li}_3\text{PO}_4$ . For example,  $\text{Li}_3\text{PO}_4$  has typical units of phosphate ( $\text{PO}_4^{3-}$ ) and  $\text{Li}_2\text{O}$ , so O–P–O, O–Li–O, and Li–O–Li would be considered important angular features. In addition, structures in which P atoms are located at short distances from each other would rarely occur considering the features of the above-described units, and thus P–P–P would not contribute much. In this way, they examined whether the NNP could be simplified without reducing the accuracy by considering structural constraints.

Compared with an NNP including all 27 possible types of SFs, they succeeded in constructing an NNP with the same accuracy using 13 types of SFs: nine radial SFs and only four angular SFs. This reduced the



number of model parameters to about half, and thus the authors simplified the training process by selecting essential SFs.

### 3.3. Smart sampling to construct NNPs with high accuracy

The appropriate collection of training data is essential to construct highly accurate NNPs. The training dataset is preferably small while including the essential features necessary to accurately represent the simulated phenomenon. In this subsection, we discuss three smart sampling approaches which we have attempted to achieve preferable data collection.

#### 3.3.1. Heuristic sampling of essential features

Li *et al* [34] demonstrated heuristic sampling based on essential features for the simulation of Li ion diffusion in amorphous  $\text{Li}_3\text{PO}_4$ . Three different classes of initial configurations were constructed using a supercell containing several numbers of atoms up to 64: (1) a snapshot of the structure obtained from the MD trajectory in the temperature range of 300–4000 K, (2) a defect structure extracted from Li or  $\text{Li}_2\text{O}$  from the class (1) configuration, and (3) an atomic configuration of the transition states of Li ion diffusion obtained by a nudged elastic band (NEB) simulation with an NNP trained by data classes (1) and (2). Class (3) configurations are difficult to sample directly by MD simulation because they are ‘rare events’. Therefore, the heuristic sampling of such rare events was efficient. The basic idea of heuristic sampling to improve the transition barrier accuracy of NNPs was reported by Peterson [35].

In total, 38 592 configurations for the three classes and energy results from DFT calculations were generated, 80% of which were used as training data and 20% as testing data. As a result of the training, the RMSE reached 5 meV/atom for the training sets and 5.6 meV/atom for the testing sets. Another verification test was also examined for large-scale structural data using a 500-step MD trajectory of the system with 128 atoms at 4000 K. The RMSE of the energy prediction was 6.2 meV/atom, which was similar to the error in the training and testing sets. This suggests the efficacy of the ‘scaling-up strategy’, wherein training is performed using small configuration datasets with a low computational cost and the model is applied to a large-scale simulation. Because NNP modelling considers the local environment (as a function of a set of SFs) of each atom, training data consisting of small-scale atomic configurations can represent potentials in large-scale atomic configurations.

#### 3.3.2. On-the-fly sampling

Heuristic sampling does not give a sufficient sampling size criterion to achieve the required accuracy for simulating an objective system. Thus, another smart sampling approach, ‘on-the-fly sampling’, has attracted attention to overcome this weak point. It is based on the ML technique called ‘active learning’ and systematically improves the model accuracy by checking the ‘model uncertainty’.

On-the-fly sampling is performed according to the following procedure. (1) First, an NNP model with a coarse potential accuracy is constructed from a small set of training data of about 1000. (2) Next, an MD simulation using the NNP model is executed to acquire new training data candidates as time evolves. (3) The model uncertainty is evaluated in terms of the acquired energy prediction data from the present NNP model. If the uncertainty is higher than a certain threshold, the acquired data are added to the training dataset, and the model is re-optimised. If the uncertainty falls below the threshold, the process returns to step 2 without re-optimising the model. (4) Data collection is terminated according to a criterion such as completing a predetermined number of steps. This process is schematically summarised in figure 2.

The key to on-the-fly sampling is evaluating the model uncertainty. Artrith and Behler defined the model uncertainty as a ‘prediction disagreement’ between a set of NNP models with different network topologies

[16, 36]. A set of NNPs optimised for a common dataset predict energies at each step of the MD simulation, and their difference (or variance) is defined as the model uncertainty. The DFT simulation is run again, and the atomic configuration and DFT total energy are appended to the training dataset if the model uncertainty with regards to the acquired MD configuration is higher than a selected criterion.

Li and Ando applied this method to sample amorphous Si (*a*-Si) reference data along with the melt-quenching method for amorphous modelling [29]. Two different NNPs with 48 nodes in the input layer and three hidden layers were prepared, one with 20 nodes and the other with 25 nodes in the hidden layer. The structure was first determined for an eight-atom system, and then the scale was increased to 32, 64, 96, and 128-atom systems according to the scaling-up strategy. Finally, 49 544 structure-energy data points were collected. The RMSE of the energy prediction was 5.9 meV/atom in the training set (90% of the reference data) and 6.2 meV/atom in the testing set (10% of the reference data). The trained NNP was also applicable to *a*-Si systems on a larger scale than the system in the training dataset: the RMSE of the energy prediction for a 216-atom structure (larger than the reference structures) was 5.3 meV/atom, which is comparable to that of the training and testing sets.

### 3.3.3. Other sampling strategies

Estimating the model uncertainty for the acquired data is not the only way to find candidates for merging into training datasets. From abundant structures obtained by low-cost simulations with a roughly trained NNP, unique structures, i.e. structures distinct from those already in the training dataset, can be identified based on data similarity with PCA, in a way analogous to atomic feature representations [30]. Structures with principal component values far from those of the pre-existing data are selected for addition to the dataset. Shimizu *et al* applied this scheme to Au–Li binary systems [37], and the resulting NNP predicted the mixing energy and dynamical behaviour well.

Data accumulation assisted by classical MD is another approach. Structures for the training dataset are selected from classical MD trajectories, and DFT total energies are calculated only for the selected structures. Minamitani *et al* adopted this strategy with Stillinger–Weber potentials to construct NNPs for crystalline Si and GaN and successfully predicted their thermal properties (see section 4.2 for details) [25].

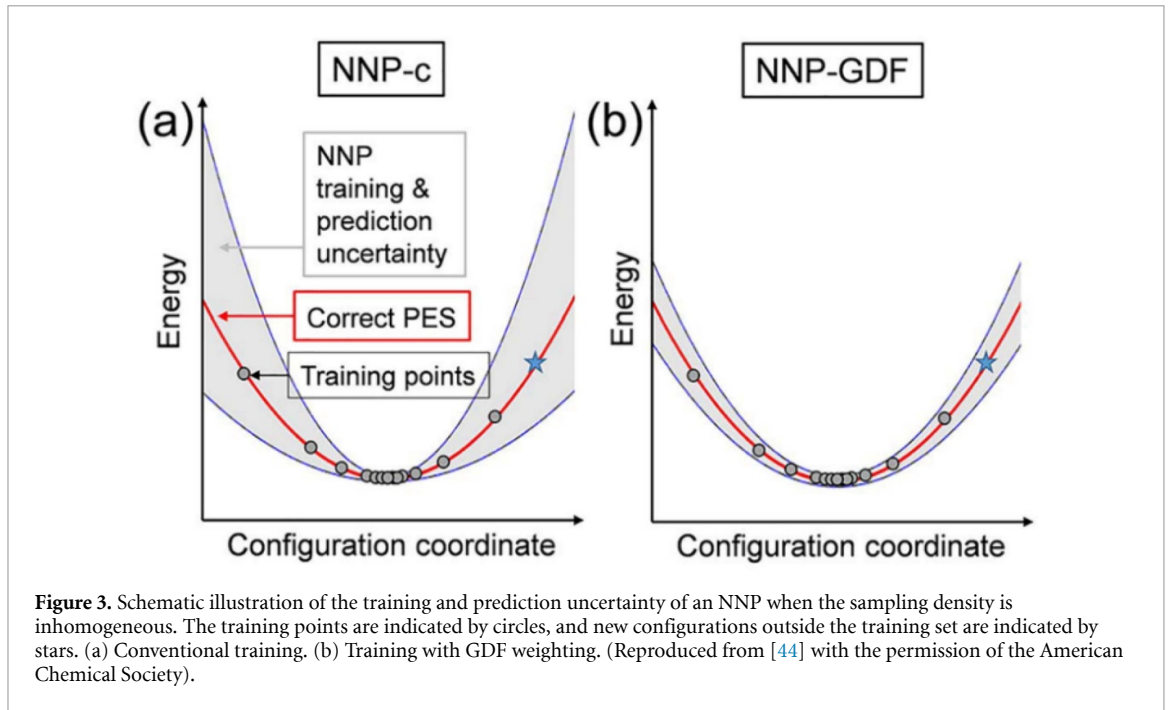
Several other sampling strategies have also been reported. An evolutionary algorithm, which avoids human selection, was employed by Hajinazar *et al* to reproduce various physical quantities of metallic systems up to ternary ones [38]. Farthest-point sampling [32] and k-means clustering methods [39] have been used to classify datasets to maximise their diversity. Several recent works also focused on developing active learning protocols in different ways from the one described in the previous subsection. For instance, Smith *et al* adopted a training dataset selection rule using the query by committee approach for small organic molecules [40], Podryabinkin *et al* reported an evolutionary-algorithm-based structural prediction [41], and Sivaraman *et al* developed a minimum number of training configuration selection scheme using hierarchical density-based spatial clustering of applications with a noise algorithm in conjunction with tuning the hyperparameters using a Bayesian optimisation [42].

## 3.4. Improving training uniformity via weighting

Ideally, a training set should sample the configurational space in an unbiased manner. However, this is not the case if the training set is obtained from MD trajectories. For example, to train defective systems, one constructs a training set from MD snapshots for a large supercell containing one defect. In this case, most of the sample points correspond to the bulk region. The subsequent training procedure tends to focus on reducing errors for over-sampled bulk configurations rather than under-sampled defect configurations, which results in higher training errors for defect configurations than that for bulk atoms. Such sampling bias (or imbalanced learning) is well known in ML, where the typical solution is to omit overrepresented data or add underrepresented data (sometimes synthetic) [43]. However, this solution does not apply to materials modelling because defective atoms must be embedded in crystalline structures and energies are trained as a whole rather than separately for the defect and bulk.

Jeong *et al* addressed this problem in NNP training and suggested a method for overcoming the sampling bias without additional sampling [44]. They first defined the  $\mathbf{G}$  space as the space spanned by SF vectors (note that SF vectors encode the radial and angular distribution of neighbouring atoms within a certain cut-off radius) and then a Gaussian density function (GDF) as a summation of Gaussian functions on each data point in the  $\mathbf{G}$  space.

$$\rho(\mathbf{G}) = \frac{1}{M} \sum_{j=1}^M \exp\left(-\frac{1}{2\sigma^2} \frac{|\mathbf{G} - \mathbf{G}_j|^2}{D}\right), \quad (4)$$



where  $\sigma$  is the Gaussian width, and  $D$  is the dimension of the SF vector. Because the GDF correlates with the number density of training points in the HD space, it can be used as a weighting term to place more emphasis on the under-sampled configurations by modifying the loss function as follows:

$$\Gamma = \frac{1}{N} \sum_{i=1}^N \left( \frac{E_i^{\text{DFT}} - E_i^{\text{NNP}}}{n_i} \right)^2 + \frac{\mu}{3M} \sum_{j=i}^M \Theta \left[ \frac{1}{\rho(\mathbf{G}_j)} \right] \left| \mathbf{F}_j^{\text{DFT}} - \mathbf{F}_j^{\text{NNP}} \right|^2. \quad (5)$$

In equation (5), the force loss function is multiplied by a scaling term because the forces can be separated into individual atoms. The scaling function  $\Theta$  monotonically increases such that it can effectively magnify the weights of under-represented  $\mathbf{G}$  points (low  $\rho$  values). As a result, NNPs trained with the GDF weighting scheme produce more even errors between the bulk and defective atoms, for instance. This uniform training was also found to improve the accuracy in predicting defect properties such as formation and migration energies.

Figure 3 illustrates the advantages of GDF weighting. The red line corresponds to the correct potential energy surface, and the training points are indicated by filled circles. Because the training set is focused on a certain point, the NNP trained by the conventional method (c-NNP) possesses a high accuracy only near the over-represented configurations. However, the prediction uncertainty increases rapidly upon moving farther away from the configuration of focus. On the other hand, GDF weighting can balance the prediction uncertainty (NNP-GDF). One additional benefit of uniform training is that it increases the stability and transferability of the NNP as it can achieve smaller prediction uncertainties for configurations extrapolated outside the training set (stars in figure 3).

## 4. Application examples

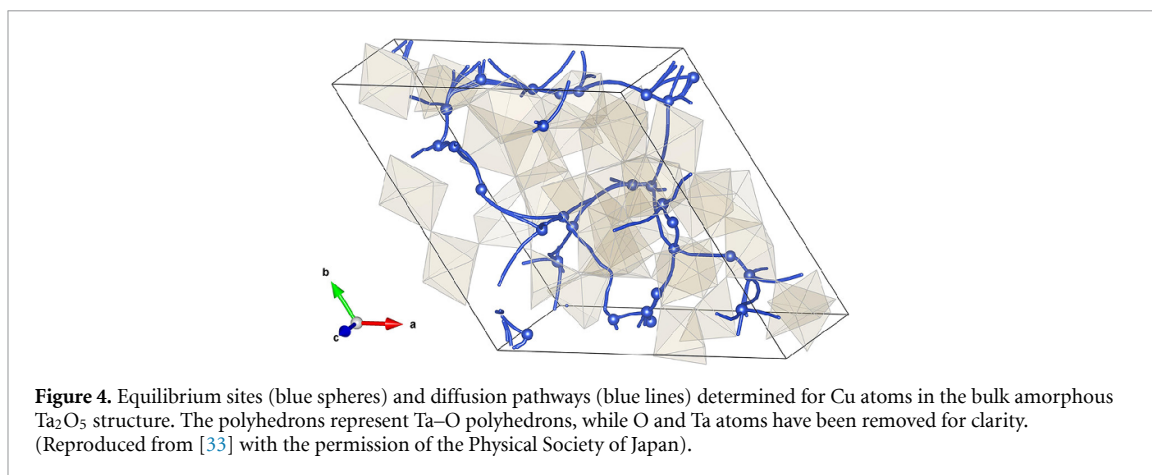
### 4.1. Ion diffusion in amorphous systems

Clarifying the behaviour of atomic diffusion in amorphous materials is important for novel information and energy devices. Simulating amorphous materials requires large-scale systems in order to represent medium-range order, and thus, DFT simulations are difficult to use because of their computational cost. Additionally, empirical potentials are often inadequate in cases where they are optimised for limited crystal structures. Using NNPs is expected to overcome these issues with conventional simulation methods. Therefore, simulation studies on amorphous materials are a promising frontier for the application of NNPs. Here, we review three cases involving the simulation of atomic diffusion in amorphous hosts using NNPs.

#### 4.1.1. Cu diffusion in $\text{Ta}_2\text{O}_5$ by a kinetic Monte Carlo approach

Li *et al* examined Cu atom diffusion in amorphous  $\text{Ta}_2\text{O}_5$  via an NNP to understand the primary process for operating atomic switches [33]. Atomic switches are promising non-volatile memory devices, in which the





diffusion and redox conditions of metallic ions/atoms are controlled to form a conduction path in an amorphous layer between two electrodes [45]. The diffusion rate in the insulating amorphous layer significantly influences the device performance, for example, the turn-on voltage, endurance, and switching speed.

Computational simulations can give insights into the diffusion mechanisms of atomic switches to improve the device performance. The kinetic Monte Carlo (KMC) method is a conventional approach to simulating the diffusion rate on the nanoscale. For the KMC approach, hopping networks of diffusing particles inside amorphous materials must be defined. However, the metastable sites of the diffusing particle, or the nodes of hopping networks, are difficult to determine heuristically because of their low-symmetry structures.

An exhaustive search is one manner of finding all possible metastable sites in amorphous materials. These metastable sites are determined by structural optimisation, wherein atoms are placed at all grid points in the amorphous structural model. Assuming that  $50^3$  grid points are generated and that the exhaustive search is performed using DFT calculations, which cost about 1 h for optimisation, this would require a total of 14 years to finish. Therefore, using NNPs is a promising approach for reducing the cost of exhaustively searching for metastable sites, especially for optimisation.

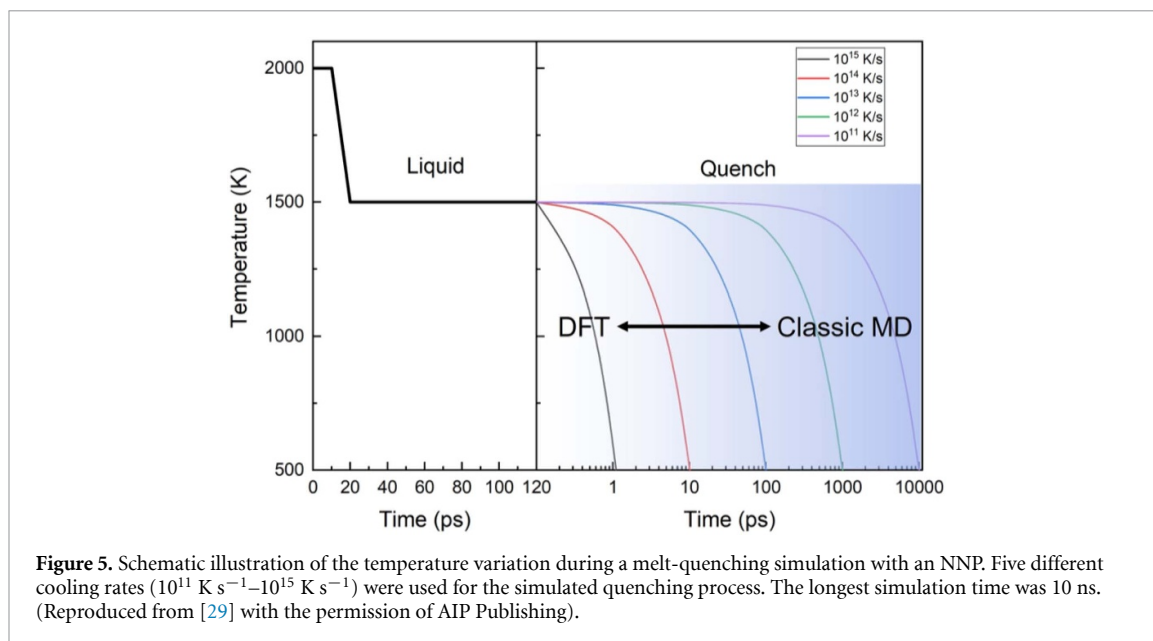
To achieve effective training for the ternary system of Cu/Ta<sub>2</sub>O<sub>5</sub>, a simplified NNP was constructed in the manner explained in section 3.2.2. The structural optimisation using the NNP was  $10^7$  times faster than that using DFT calculations, which means that the total sampling time for 2000 cases as training data was the main concern, whereas almost all of the optimisation processing times could be neglected. Finally, 29 metastable sites were determined in the given amorphous model of Ta<sub>2</sub>O<sub>5</sub>. The NNP also accelerated the NEB calculation to evaluate the hopping barrier between pairs of metastable sites. The determined metastable sites and diffusion pathways in the amorphous structure are shown in figure 4. It should be noted that the metastable sites and hopping pathways determined using the NNP are predicted and should be confirmed by DFT simulations if possible. In this case, the prediction worked well: the estimated maximum error along the diffusion pathway obtained by the NEB method was 0.15 eV. Finally, the diffusion network, including all metastable sites and the diffusion barrier on each edge, was evaluated for the Ta<sub>2</sub>O<sub>5</sub> amorphous structure and was ready for use in a KMC simulation for Cu diffusion.

#### 4.1.2. Amorphous Li<sub>3</sub>PO<sub>4</sub> modelling and Li diffusion simulation by direct MD approach

Solid-state electrolytes for various electrochemical storage systems are another key application of amorphous materials. For a conventional Li<sub>3</sub>PO<sub>4</sub> electrolyte for Li-ion batteries and related devices, a dense Li<sup>+</sup> concentration should be considered to simulate the diffusivity. However, this density makes it difficult to apply a simplified NNP while focusing on a target species, assuming atoms should be dilute, as described in the previous subsection. To overcome difficulties in constructing NNPs for multicomponent systems due to the many SF components, Li *et al* demonstrated the selection of essential SFs by focusing on the structure of Li<sub>3</sub>PO<sub>4</sub> and heuristic sampling of training datasets (for details, see sections 3.2.2 and 3.3.1) [34].

As a result, they successfully simulated the diffusivity of Li<sup>+</sup> with low-cost MD simulations and an NNP for the Li<sub>3</sub>PO<sub>4</sub> system. The computational speed of the NNP was three to four orders of magnitude higher than that of *ab initio* MD. Moreover, the low cost of the MD simulation enabled generating reliable amorphous structures, which is a fundamental obstacle in computational studies on amorphous materials.

A non-stoichiometric large-scale amorphous model of Li<sub>3</sub>PO<sub>4</sub> was also constructed using the NNP, and an amorphous system consistent with experimentally observed features was reproduced. Specifically, a



structural  $\text{P}_2\text{O}_7^{4-}$  unit was formed, which is a dimerisation of  $\text{PO}_4^{3-}$  formed by sharing the O atom at the vertex.

With the NNP, MD simulations were performed on the largest amorphous structure to evaluate the barrier for Li diffusion. The estimated barrier height was 0.55 eV, which was in good agreement with the results obtained by the ion exchange method and impedance measurements. This indicates that the diffusion coefficient can be accurately evaluated even for highly concentrated ion distributions in amorphous structures by using NNPs.

#### 4.1.3. Acceleration of melt-quenching method for modelling amorphous system

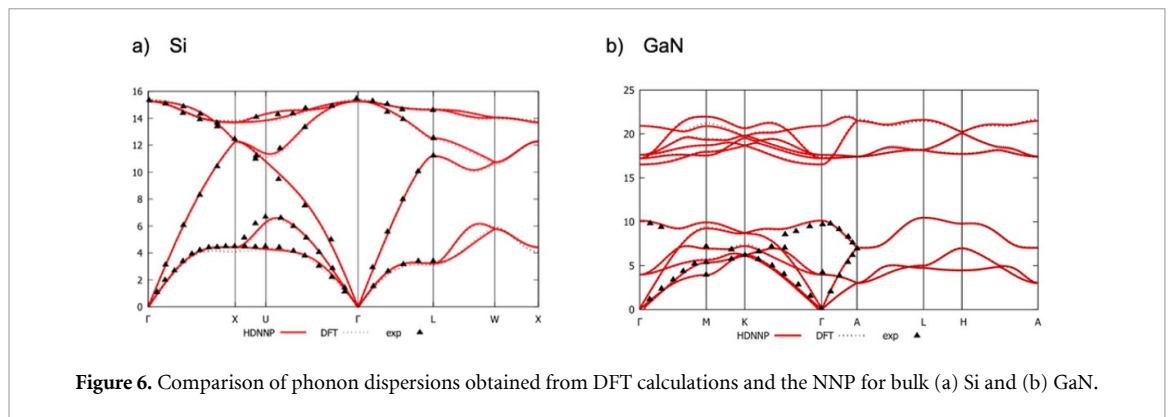
In general, amorphous models are constructed by the melt-quenching method, wherein the atomic configuration is melted at a high temperature and then gradually cooled to room temperature in the MD simulation. The melt-quenching method based on DFT has a high computational cost, which is often reduced by compromising the cooling condition. However, adopting a fast cooling rate (typically  $10^{14} \text{ K s}^{-1}$ ) to accelerate the quenching process can cause artifacts such as high defect concentrations in the generated structure. Because material properties strongly depend on their atomic structures, such artifacts due to the fast cooling rate would affect the simulated properties of amorphous materials. In fact, in experiments, the room temperature Li diffusivity in *a*-Si spans four orders of magnitude between  $10^{-14}$  and  $10^{-10} \text{ cm}^2 \text{ s}^{-1}$ , depending on the fabrication process. Thus, controlling the structural disorder in amorphous materials is in high demand.

As described in previous sections, using NNPs significantly reduces the computational cost while maintaining the accuracy. Therefore, a much slower cooling process than that in compromised DFT-based simulations is possible, and a series of atomic structures of *a*-Si was generated with different degrees of disorder by varying the cooling rates ( $10^{11} \text{ K s}^{-1}$ – $10^{15} \text{ K s}^{-1}$ ) (figure 5) [29]. To collect the training dataset for the NNP, a scaling-up strategy and on-the-fly sampling were adopted (for details, see sections 3.3.1 and 3.3.2). The results showed that the short- and intermediate-range orders increased with the decreasing cooling rate. The corresponding in silico model can be determined for experimental samples prepared with a certain method and thermal history by comparing observed values, although detailed fabrication processes are difficult to simulate directly.

The diffusion pathways of Li ions in a series of *a*-Si models with various degrees of structural disorder were also calculated using an NNP [26]. In this study, diffusion pathways were determined, and KMC simulations were performed. The activation energy of Li ions was estimated to be higher in a more disordered local environment and vice versa, and this energy varied within the range of 1.21 eV–1.46 eV, which agreed well with experimental measurements (1.38 eV–1.46 eV). The simulation results also showed that Li diffusion was enhanced at higher Li concentrations, which was consistent with experimental observations.

## 4.2. Thermal transport

Thermal management is crucial in many information and energy devices. In particular, controlling lattice thermal transport has become an increasingly important and sometimes central issue for further



**Figure 6.** Comparison of phonon dispersions obtained from DFT calculations and the NNP for bulk (a) Si and (b) GaN.

development. Therefore, the simulation of lattice thermal transport has recently become an active area of focus.

The current methods for simulating the lattice thermal conductivity ( $\kappa$ ) can be classified into three categories [46, 47]: (a) equilibrium molecular dynamics (EMD) using the Green–Kubo (GK) formula [48–50], (b) anharmonic lattice dynamics (ALD) combined with phonon transport calculations using the Boltzmann transport equation (BTE) and Fourier’s law [51–55], and (c) the direct evaluation of the heat flux by the nonequilibrium molecular dynamics (NEMD) [56, 57].

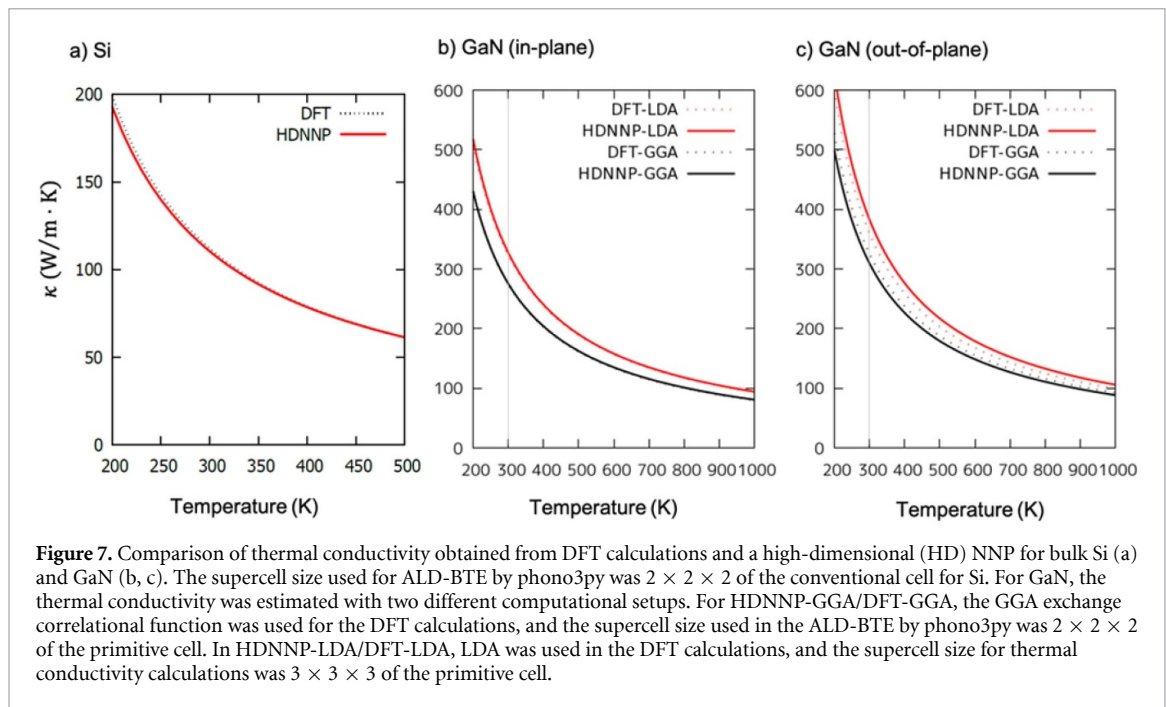
Although first-principles simulations based on the above methods have recently become possible, their computational costs are quite high, and thus their application is limited to simple systems. Therefore, several groups have already applied NNPs to this issue. Sossao *et al* [58] constructed an NNP for amorphous GeTe and investigated  $\kappa$  based on the EMD-GK method. The evaluated  $\kappa$  was close to the experimental values measured in systems with similar compositions. Moreover, based on the Allen–Feldman theory [59], they revealed that the main heat carriers are non-propagating vibrations called ‘diffusons’. Huang *et al* [60] constructed an NNP for *a*-Si using a modified version of the Behler–Parrinello-type NNP called the single atom NNP and effectively reproduced the  $\kappa$  obtained by DFT-EMD data. Wen and Tadmor [61] constructed a hybrid of an NNP and Lennard–Jones-type long-range potential for multilayer graphene. The obtained hybrid potential reproduced the  $\kappa$  of pristine graphene well (2531 W m<sup>−1</sup> K<sup>−1</sup> at 300 K). They also found that  $\kappa$  was dramatically reduced by carbon vacancies in graphene: 415 W m<sup>−1</sup> K<sup>−1</sup> for 0.1% vacancies and 195 W m<sup>−1</sup> K<sup>−1</sup> for 0.2%.

The application of NNPs for evaluating  $\kappa$  is not limited to combination with EMD-GK. Based on ALD-BTE, Minamitani *et al* [25] investigated  $\kappa$  in crystalline Si and GaN by combining a HD NNP and phonopy [62]/phono3py [51] package. The obtained RMSEs of the force predictions were 25.5 meV Å<sup>−1</sup> for Si and 37.8 meV Å<sup>−1</sup> for GaN. These precisions are enough to reproduce the DFT calculation results of phonon dispersions, as seen in the comparison of phonon dispersions (figure 6). The predicted  $\kappa$  also matched the DFT calculation results, as seen in the comparison of thermal conductivities (figures 7(b) and (c)). The dependence of the calculated  $\kappa$  of GaN on several computational conditions is worth noting. The differences between NNP-GGA/DFT-GGA and NNP-LDA/DFT-LDA in figure 7 are the exchange correlational function used in the DFT calculations and the supercell size used in the  $\kappa$  calculations.

In the former case, we constructed a DFT dataset for a 32-atom GaN system using GGA and trained the NNP to predict  $\kappa$  in a 2 × 2 × 2 supercell of the primitive GaN cell (32 atoms). The methods for sampling the atomic configuration, number of data, hyperparameters, and network topology were the same as in [25]. The predictions by the NNP matched the DFT calculation results, but both NNP/DFT calculation results were underestimated compared with previous reports (approximately 400 W m<sup>−1</sup> K<sup>−1</sup> at 300 K [63]).

In the latter case, we constructed a DFT dataset for the same size system by using LDA, and we trained the HD NNP in the same way as in the GGA case but with a slight modification of the hyperparameter  $\eta$  because of the difference in lattice constants for the GGA and LDA cases. Then, we predicted  $\kappa$  using a 3 × 3 × 3 supercell (108 atoms) of the primitive cell. As shown in figure 7, the predicted  $\kappa$  under this condition approaches the previously reported value and agrees with the direct DFT calculations. Notably, even though the system size in the training dataset was smaller than the supercell size in the  $\kappa$  calculation, the prediction accuracy was high. These results clearly show the extendibility and predictive power of NNPs in third-order anharmonic potentials.

The application of NNPs for NEMD has not yet been reported because NEMD requires larger systems to reproduce a reasonable temperature gradient than those used in EMD-GK and ALD-BTE. NEMD is a powerful technique which can directly evaluate heat flux without any assumption of the local equilibrium. In



**Figure 7.** Comparison of thermal conductivity obtained from DFT calculations and a high-dimensional (HD) NNP for bulk Si (a) and GaN (b, c). The supercell size used for ALD-BTE by phono3py was  $2 \times 2 \times 2$  of the conventional cell for Si. For GaN, the thermal conductivity was estimated with two different computational setups. For HDNNP-GGA/DFT-GGA, the GGA exchange correlational function was used for the DFT calculations, and the supercell size used in the ALD-BTE by phono3py was  $2 \times 2 \times 2$  of the primitive cell. In HDNNP-LDA/DFT-LDA, LDA was used in the DFT calculations, and the supercell size for thermal conductivity calculations was  $3 \times 3 \times 3$  of the primitive cell.

this regard, improving the efficiency of NNP-based MD simulations remains an important issue for the application of NNPs to thermal transport problems.

## 5. Atomic energy mapping

Although there have been many methodological advances in the field of NNP development, understanding at the fundamental level remains lacking. Due to their black-box nature, NNPs are often simply considered interpolation models of a given structure and total energy. However, there is a distinct difference between NNPs and typical ML models. That is, unlike other models, the target output of NNPs does not match what they are actually trained upon: the output of an NNP is atomic energies, whereas it is trained over total energies, which are sums of atomic energies. Therefore, the basic concept of NNPs would be justified only when the NNP atomic energy is transferable, i.e. the atomic energy defined in one system can also describe other systems with similar local environments. In addition, as NNPs are trained over DFT total energies, the question of transferable atomic energies must be justified at the DFT level as well. Yoo *et al* [64] examined this issue, which is summarised as follows.

In DFT, the total energy can be expressed as the integral of the local energy density. When the whole space is partitioned into non-overlapping atomic volumes  $V_i$ , an atomic energy can be assigned to each atom, the sum of which is the same as the total energy. The transferability of the atomic energy in DFT can be established by the nearsightedness principle of the electronic structures, which states that the charge density of a given point is determined by the configuration of nearby atoms within a certain cut-off distance. Additionally, in metals and insulators with finite temperatures, the density matrix decays exponentially with distance. As a result, the DFT total energy can be seamlessly decomposed into atomic contributions which depend only on the local environment, thereby securing the transferability [64]. (The Coulomb potential is assumed to be effectively short-ranged or described separately [5, 65]). Therefore, the total energy of the system can be expressed as follows:

$$E_{\text{tot}} = \sum_i E_{\text{at}}(\mathcal{R}_i; R_c), \quad (6)$$

where  $i$  is the atomic index, and  $\mathcal{R}_i$  is the collection of relative position vectors of atoms lying within  $R_c$  from the  $i$ th atom. Thus, the goal of the ML potential is to approximate the underlying atomic energy from the given DFT total energy.

The above discussion also indicates that the ML potential is essentially a variant of the  $O(N)$  method approximating DFT calculations, realised by transferable atomic energies as a function of local configurations. In this respect, the ML potential is conceptually distinct from conventional classical force fields; in classical potentials, chemical bonding is the founding concept, which has no explicit connection to DFT. In contrast, the ML potential is based on local and transferable atomic energies grounded in the DFT

framework, which allows for ML potentials to deal with materials with complicated bonding natures (for instance GeTe [66] and NiSi [67]).

Atomic energy mapping is not unique because of non-unique definition of atomic volumes, which makes it difficult to judge the validity of the mapping chosen by a certain NNP. Therefore, to investigate atomic energy mapping with NNPs, Yoo *et al* utilised invariant  $\mathbf{G}$  space points at which the DFT atomic energy is uniquely defined. One example of invariant points is perfect crystals, in which the atomic energy of each atom is simply the total energy divided by the number of atoms. Through several examples encompassing crystals, surfaces, and nanoclusters, NNPs were demonstrated to correctly reproduce atomic energies at invariant  $\mathbf{G}$  points although they were not explicitly included in the training set. Nevertheless, the atomic energies could deviate significantly from the reference DFT atomic energies, even when the total energy was precisely reproduced. Such ‘ad hoc’ energy mapping can undermine the transferability of NNPs.

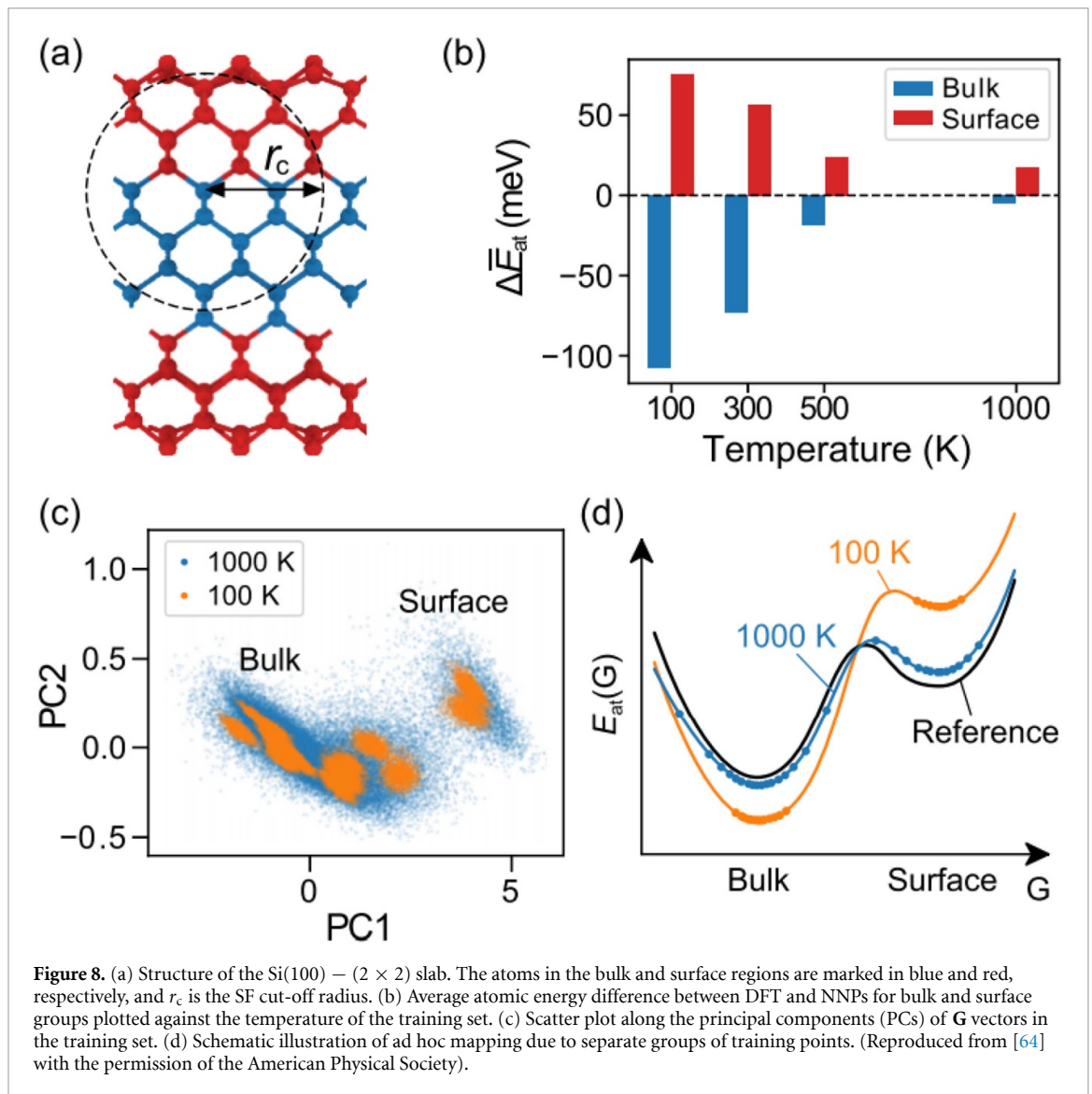
One example of ad hoc energy mapping occurs in the Si slab model (figure 8). When the slab structure is divided into bulk (blue atoms) and surface (red atoms) regions, the bulk atoms should have atomic energies close to the bulk crystalline DFT atomic energy. Figure 8(b) shows the errors in the energy mapping,  $\Delta\bar{E}_{\text{at}}(\text{bulk})$  and  $\Delta\bar{E}_{\text{at}}(\text{surface})$  for NNPs trained over MD trajectories at different temperatures (100–1000 K). At low temperatures, the mapping error is high, with underestimated average atomic energies in the bulk and overestimated ones in the surface regions. Nevertheless, the NNP correctly predicts the total energy because the errors cancel out. At higher temperatures, the error becomes small and comparable to the RMSE in the total energy (3 meV/atom). This observation can be understood based on the distribution of training points in the  $\mathbf{G}$  space (figure 8(c)). At low temperatures, the training points corresponding to the bulk and surface regions are well separated, whereas the training points from high-temperature MD simulations are connected, albeit weakly. Figure 8(d) schematically illustrates this tendency. When the training points in the  $\mathbf{G}$  space are grouped separately, as shown by the orange line, the NNP is vulnerable to ad hoc mapping because arbitrary offsets can still produce the same total energies. On the other hand, when all the training points are connected to some degree, as shown by the blue line, the  $E_{\text{at}}$  values at intermediate configurations help mitigate the spurious energy offset.

Multicomponent systems have an additional degree of freedom, which is the relative offset in  $E_{\text{at}}$  among different chemical species. If the training set consists of only structures with a single stoichiometry, the atomic energy mapping among chemical types can be arbitrary, leading to ad hoc mapping. For example, Lee *et al* [66] reported that when MD simulations were performed on liquid GeTe using an NNP whose training set consists of GeTe with compositions only near 1:1, phase separations were always observed. They found that average atomic energies of Ge and Te in the liquid states differed from those of unary Ge and Te by more than 1 eV, implying ad hoc mapping in the stoichiometric GeTe. On the other hand, when the training set encompassed the entire compositional range, average atomic energies of Ge and Te changed smoothly along the compositional variation, and the issue of the phase separation was resolved.

In addition to the ad hoc energy mapping, we also note that valid energy mapping is still not unique. This means that a certain degree of randomness can be introduced to atomic energies each time an NNP is trained on a certain training set. Such randomness could be critical when a set of NNPs are used as the uncertainty estimator (see section 3.3.2). That is to say, if an ensemble of NNPs is trained with the same DFT total-energy sets, each NNP can produce different atomic energies for the same structures, which may degrade the atomic-scale resolution of the uncertainty estimation. To cope with this issue, Jeong *et al* proposed an NNP uncertainty indicator based on replica NNPs [67], which are a set of NNPs with different sizes and initial weight distributions. These NNPs are trained directly over atomic energies of the reference NNP that runs MD simulations. As the replica ensemble is trained by the atomic energies of reference NNP, they produce the same atomic energies for chemical environments included in the training set while the output diverges for those outside the training set. In [67], the method was applied to the Ni–Si interfacial reaction, and the replica ensemble could identify atomic configurations with high uncertainty at the interface region. This information was then utilized for augmenting the training set, and the refined NNP allowed for MD simulations up to 3.6 ns without configurations with high uncertainties.

## 6. Concluding remarks

In this article, we first discussed the construction and training of HD NNPs based on our experiences. We mentioned that the complexity of the NN and the accuracy of the NNP can be controlled via the optimisation of the hyperparameters, simplification of modelling based on physical insights, or utilisation of data-science techniques. Furthermore, we discussed how the size of the training dataset can be reduced or how the training uniformity can be improved to efficiently attain the desired prediction accuracy via smart sampling techniques such as heuristic sampling of essential features, on-the-fly sampling, and weighting with a GDF. Then, we described a few examples that demonstrate the prediction accuracy of NNPs: ion migration



in Ta<sub>2</sub>O<sub>5</sub>, Li<sub>3</sub>PO<sub>4</sub>, and *a*-Si; construction of *a*-Si models; and thermal transport in crystalline Si and GaN. Finally, we discussed the meaning of atomic energies in NNPs and suggested the possibility of improving their accuracy by considering the behaviour of atomic energies. These examples show that NNPs can achieve good agreement with the first-principles calculations used for training, not only in terms of the total energy and atomic forces but also the target physical properties and phenomena. As mentioned in the original publications of these examples, the results confirm that, consequently, NNPs can exceed the spatial and time scales of first-principles simulations.

As described above, NNPs are highly promising; however, they offer much room for improvement. For sampling and training, although we have described several effective techniques, we still rely on considerable trial and error when constructing NNPs for new systems. A deeper understanding of the effectiveness and limitations of the suggested techniques is highly desired. In relation to this, active-learning approaches should be further explored, including the on-the-fly sampling described in section 3.3.2 and automatic selection of descriptors and atomic configurations for training data [32]. A more serious issue is the construction of NNPs for systems with more than three elements because the complexity of the NNP increases exponentially with the number of elements. Although several NNP trials have been performed on four-element [12, 68, 69] and five-element [70] systems, to the best of our knowledge, their accuracies are inferior to those of the NNPs constructed for systems with one to three elements. Overcoming this difficulty will be an important future issue.

Another issue is the comparison of various ML potentials. Recently, Zuo *et al* [71] comparatively evaluated the performance of NNP, Gaussian approximation potential (GAP) [6], momentum tensor potential (MTP) [9], and linear [7] and quadratic [10] spectral neighbour analysis potentials for several single-element systems, and their comparison demonstrates that all of them are similarly promising.

Although this work is of significance, we would like to note that fairly comparing different potentials appears to be highly difficult because the quality and volume of training data, tuning of hyperparameters, etc, affect the accuracy of the potentials, and evaluating whether the compared potentials approach their performance limits to a similar extent is challenging. There remains a room for further studies on this point, although Zuo *et al* carefully designed the evaluation protocol. In relation to this, it is worth noting that mathematically, the ultimate accuracy of GAPs and NNPs is known to be the same [72]. They also mentioned several limitations of their study, such as the lack of evaluation for multi-element systems. Notably, MTP was recently constructed successfully for a five-element alloy without an essential loss of accuracy [73]. Thus, a different type of ML potential may be the most suitable depending on the purpose. Nevertheless, fairly comparing the quality and performance of various ML potentials remains challenging.

## Acknowledgments

Some of our results presented in this article were obtained in collaboration with S. Okugawa and M. Ogura. SW acknowledges support from JST-CREST program Grant Nos. JPMJCR1523, and JSPS KAKENHI Grant Nos. 17H05330 and 19H04536, Japan. EM acknowledges support from JST-PRESTO program Grant No. JPMJPR1717, Japan.

## ORCID iDs

Satoshi Watanabe  <https://orcid.org/0000-0002-8069-6938>

Wonseok Jeong  <https://orcid.org/0000-0001-8894-1857>

Dongheon Lee  <https://orcid.org/0000-0001-8774-2821>

Koji Shimizu  <https://orcid.org/0000-0001-5622-9582>

Emi Mimanitani  <https://orcid.org/0000-0002-8003-6526>

Yasunobu Ando  <https://orcid.org/0000-0003-3702-034X>

Seungwu Han  <https://orcid.org/0000-0003-3958-0922>

## References

- [1] Waser R and Aono M 2006 Nanoionics-based resistive switching memories *Nat. Mater.* **6** 833–40
- [2] Sawa A 2008 Resistive switching in transition metal oxides *Mater. Today* **11** 28–36
- [3] Ball P 2012 Computer engineering: feeling the heat *Nature* **492** 174–6
- [4] Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401
- [5] Artrith N, Morawietz T and Behler J 2011 High-dimensional neural-network potentials for multicomponent systems: applications to zinc oxide *Phys. Rev. B* **83** 153101
- [6] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons *Phys. Rev. Lett.* **104** 136403
- [7] Thompson A P, Swiler L P, Trott C R, Foiles S M and Tucker G J 2015 Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials *J. Comput. Phys.* **285** 316–30
- [8] Botu V, Batra R, Chapman J and Ramprasad R 2017 Machine learning force fields: construction, validation, and outlook *J. Phys. Chem. C* **121** 511–22
- [9] Shapeev A V 2016 Moment tensor potentials: a class of systematically improvable interatomic potentials *Multiscale Model. Simul.* **14** 1153–73
- [10] Wood M A and Thompson A P 2018 Extending the accuracy of the SNAP interatomic potential form *J. Chem. Phys.* **148** 241721
- [11] Artrith N 2019 Machine learning for the modeling of interfaces in energy storage and conversion materials *J. Phys. Energy* **1** 032002
- [12] Artrith N and Kolpak A M 2014 Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: a combination of DFT and accurate neural network potentials *Nano Lett.* **14** 2670–6
- [13] Fujikake S, Deringer V L, Lee T H, Krynski M, Elliott S R and Csányi G 2018 Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures *J. Chem. Phys.* **148** 241714
- [14] Behler J 2017 First principles neural network potentials for reactive simulations of large molecular and condensed systems *Angew. Chem., Int. Ed.* **56** 12828–40
- [15] Bartók A P, De S, Poelking C, Bernstein N, Kermode J R, Csányi G and Ceriotti M 2017 Machine learning unifies the modeling of materials and molecules *Sci. Adv.* **3** e1701816
- [16] Behler J 2015 Constructing high-dimensional neural network potentials: a tutorial review *Int. J. Quantum Chem.* **115** 1032–50
- [17] Bartók A P, Gillan M J, Manby F R and Csányi G 2013 Machine-learning approach for one- and two-body corrections to density functional theory: applications to molecular and condensed water *Phys. Rev. B* **88** 054104
- [18] Artrith N, Urban A and Ceder G 2017 Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species *Phys. Rev. B* **96** 014112
- [19] Novotni M and Klein R 2004 Shape retrieval using 3D Zernike descriptors *Comput. Aided Des.* **36** 1047–62
- [20] Kohn W 1996 Density functional and density matrix method scaling linearly with the number of atoms *Phys. Rev. Lett.* **76** 3168–71
- [21] Artrith N and Urban A 2016 An implementation of artificial neural-network potentials for atomistic materials simulations: performance for TiO<sub>2</sub> *Comput. Mater. Sci.* **114** 135–50
- [22] Shao Y, Hellström M, Mitev P D, Knijff L and Zhang C 2020 PiNN: a Python library for building atomic neural networks of molecules and materials *J. Chem. Inf. Model.* **60** 1184–93

- [23] Khorshidi A and Peterson A A 2016 Amp: a modular approach to machine learning in atomistic simulations *Comput. Phys. Commun.* **207** 310–24
- [24] Lee K, Yoo D, Jeong W and Han S 2019 SIMPLE-NN: an efficient package for training and executing neural-network interatomic potentials *Comput. Phys. Commun.* **242** 95–103
- [25] Minamitani E, Ogura M and Watanabe S 2019 Simulating lattice thermal conductivity in semiconducting materials using high-dimensional neural network potential *Appl. Phys. Express* **12** 095001
- [26] Li W and Ando Y 2020 Effect of local structural disorder on lithium diffusion behavior in amorphous silicon *Phys. Rev. Mater.* **4** 045602
- [27] Artrith N, Hiller B and Behler J 2013 Neural network potentials for metals and oxides - first applications to copper clusters at zinc oxide *Phys. Status Solidi Basic Res.* **250** 1191–203
- [28] Sun G and Sautet P 2019 Toward fast and reliable potential energy surfaces for metallic Pt clusters by hierarchical delta neural networks *J. Chem. Theory Comput.* **15** 5614–27
- [29] Li W and Ando Y 2019 Dependence of a cooling rate on structural and vibrational properties of amorphous silicon: a neural network potential-based molecular dynamics study *J. Chem. Phys.* **151** 114101
- [30] Cubuk E D, Malone B D, Onat B, Waterland A and Kaxiras E 2017 Representations in neural network based empirical potentials *J. Chem. Phys.* **147** 024104
- [31] Li W and Ando Y 2018 Comparison of different machine learning models for the prediction of forces in copper and silicon dioxide *Phys. Chem. Chem. Phys.* **20** 30006–20
- [32] Imbalzano G, Anelli A, Giofré D, Klees S, Behler J and Ceriotti M 2018 Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials *J. Chem. Phys.* **148** 241730
- [33] Li W, Ando Y and Watanabe S 2017 Cu diffusion in amorphous Ta<sub>2</sub>O<sub>5</sub> studied with a simplified neural network potential *J. Phys. Soc. Japan* **86** 104004
- [34] Li W, Ando Y, Minamitani E and Watanabe S 2017 Study of Li atom diffusion in amorphous Li<sub>3</sub>PO<sub>4</sub> with neural network potential *J. Chem. Phys.* **147** 214106
- [35] Peterson A A 2016 Acceleration of saddle-point searches with machine learning *J. Chem. Phys.* **145** 074106
- [36] Artrith N and Behler J 2012 High-dimensional neural network potentials for metal surfaces: a prototype study for copper *Phys. Rev. B* **85** 045439
- [37] Shimizu K et al 2020 Phase stability of Au-Li binary systems studied using neural network potential (arXiv:2008.05094)
- [38] Hajinazar S, Shao J and Kolmogorov A N 2017 Stratified construction of neural network based interatomic models for multicomponent materials *Phys. Rev. B* **95** 014114
- [39] Miwa K and Ohno H 2016 Molecular dynamics study on  $\beta$ -phase vanadium monohydride with machine learning potential *Phys. Rev. B* **94** 184109
- [40] Smith J S, Nebgen B, Lubbers N, Isayev O and Roitberg A E 2018 Less is more: sampling chemical space with active learning *J. Chem. Phys.* **148** 241733
- [41] Podryabinkin E V, Tikhonov E V, Shapeev A V and Oganov A R 2019 Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning *Phys. Rev. B* **99** 064114
- [42] Sivaraman G, Krishnamoorthy A N, Baur M, Holm C, Stan M, Csányi G, Benmore C and Vázquez-Mayagoitia Á 2020 Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide *Npj Comput. Mater.* **6** 104
- [43] He H and Garcia E A 2009 Learning from imbalanced data *IEEE Trans. Knowl. Data Eng.* **21** 1263–84
- [44] Jeong W, Lee K, Yoo D, Lee D and Han S 2018 Toward reliable and transferable machine learning potentials: uniform training by overcoming sampling bias *J. Phys. Chem. C* **122** 22790–5
- [45] Hasegawa T, Terabe K, Tsuruoka T and Aono M 2012 Atomic switch: atom/ion movement controlled devices for beyond von-Neumann computers *Adv. Mater.* **24** 252–67
- [46] Cahill D G et al 2014 Nanoscale thermal transport. II. 2003–2012 *Appl. Phys. Rev.* **1** 011305
- [47] Schelling P K, Phillpot S R and Keblinski P 2002 Comparison of atomic-level simulation methods for computing thermal conductivity *Phys. Rev. B* **65** 144306
- [48] Kubo R, Yokota M and Nakajima S 1957 Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance *J. Phys. Soc. Japan* **12** 1203–11
- [49] Volz S G and Chen G 2000 Molecular-dynamics simulation of thermal conductivity of silicon crystals *Phys. Rev. B* **61** 2651–6
- [50] Ladd A J C, Moran B and Hoover W G 1986 Lattice thermal conductivity: a comparison of molecular dynamics and anharmonic lattice dynamics *Phys. Rev. B* **34** 5058–64
- [51] Togo A, Chaput L and Tanaka I 2015 Distributions of phonon lifetimes in Brillouin zones *Phys. Rev. B* **91** 094306
- [52] Mizokami K, Togo A and Tanaka I 2018 Lattice thermal conductivities of two SiO<sub>2</sub> polymorphs by first-principles calculations and the phonon Boltzmann transport equation *Phys. Rev. B* **97** 224306
- [53] Zhou J, Liao B and Chen G 2016 First-principles calculations of thermal, electrical, and thermoelectric transport properties of semiconductors *Semicond. Sci. Technol.* **31** 043001
- [54] Lindsay L, Broido D A and Reinecke T L 2013 *Ab initio* thermal transport in compound semiconductors *Phys. Rev. B* **87** 165201
- [55] Broido D A, Malorny M, Birner G, Mingo N and Stewart D A 2007 Intrinsic lattice thermal conductivity of semiconductors from first principles *Appl. Phys. Lett.* **91** 231922
- [56] Müller-Plathe F 1997 A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity *J. Chem. Phys.* **106** 6082–5
- [57] El-Genk M S, Talaat K and Cowen B J 2018 Thermal conductivity of silicon using reverse non-equilibrium molecular dynamics *J. Appl. Phys.* **123** 205104
- [58] Soso G C, Donadio D, Caravati S, Behler J and Bernasconi M 2012 Thermal transport in phase-change materials from atomistic simulations *Phys. Rev. B* **86** 104301
- [59] Allen P B and Feldman J L 1993 Thermal conductivity of disordered harmonic solids *Phys. Rev. B* **48** 12581–8
- [60] Huang Y, Kang J, Goddard W A and Wang L W 2019 Density functional theory based neural network force fields from energy decompositions *Phys. Rev. B* **99** 064103
- [61] Wen M and Tadmor E B 2019 Hybrid neural network potential for multilayer graphene *Phys. Rev. B* **100** 195419
- [62] Togo A and Tanaka I 2015 First principles phonon calculations in materials science *Scr. Mater.* **108** 1–5
- [63] Lindsay L, Broido D A and Reinecke T L 2012 Thermal conductivity and large isotope effect in GaN from first principles *Phys. Rev. Lett.* **109** 095901



- [64] Yoo D, Lee K, Jeong W, Lee D, Watanabe S and Han S 2019 Atomic energy mapping of neural network potential *Phys. Rev. Mater.* **3** 093802
- [65] Morawietz T, Sharma V and Behler J 2012 A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges *J. Chem. Phys.* **136** 064103
- [66] Lee D, Lee K, Yoo D, Jeong W and Han S 2020 Crystallization of amorphous GeTe simulated by neural network potential addressing medium-range order *Comput. Mater. Sci.* **181** 109725
- [67] Jeong W, Yoo D, Lee K, Jung J and Han S 2020 Efficient atomic-resolution uncertainty estimation for neural network potentials using a replica ensemble *J. Phys. Chem. Lett.* **11** 6090–6
- [68] Marcolongo A, Binninger T, Zipoli F and Laino T 2020 Simulating diffusion properties of solid-state electrolytes via a neural network potential: performance and training scheme *ChemSystemsChem* **2** e1900031
- [69] Eckhoff M and Behler J 2019 From molecular fragments to the bulk: development of a neural network potential for MOF-5 *J. Chem. Theory Comput.* **15** 3793–809
- [70] Zhang Y, Hu C and Jiang B 2019 Embedded atom neural network potentials: efficient and accurate machine learning with a physically inspired representation *J. Phys. Chem. Lett.* **10** 4962–7
- [71] Zuo Y *et al* 2020 Performance and cost assessment of machine learning interatomic potentials *J. Phys. Chem. A* **124** 731–45
- [72] Neal R M 1996 *Priors for Infinite Networks* (New York: Springer) pp 29–53
- [73] Grabowski B, Ikeda Y, Srinivasan P, Körmann E, Freysoldt C, Duff A I, Shapeev A and Neugebauer J 2019 Ab initio vibrational free energies including anharmonicity for multicomponent alloys *Npj Comput. Mater.* **5** 80