CJC Chinese Journal of Chemistry 中国化学 · An International Journal

For submission: https://mc.manuscriptcentral.com/cjoc
For published articles: https://onlinelibrary.wiley.com/journal/16147065

Critical Review

Cite this paper: *Chin. J. Chem.* **2021**, *39*, 3144—3154. **DOI: 10.1002/cjoc.202100299**

Structure and Dynamics of Energy Materials from Machine Learning Simulations: A Topical Review†

Shu-Hui Guan,[a,b] Cheng Shang,*[,b] and Zhi-Pan Liu*[,b]

[a] *Shanghai Academy of Agricultural Sciences, Shanghai 201403, China*
[b] *Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200438, China*

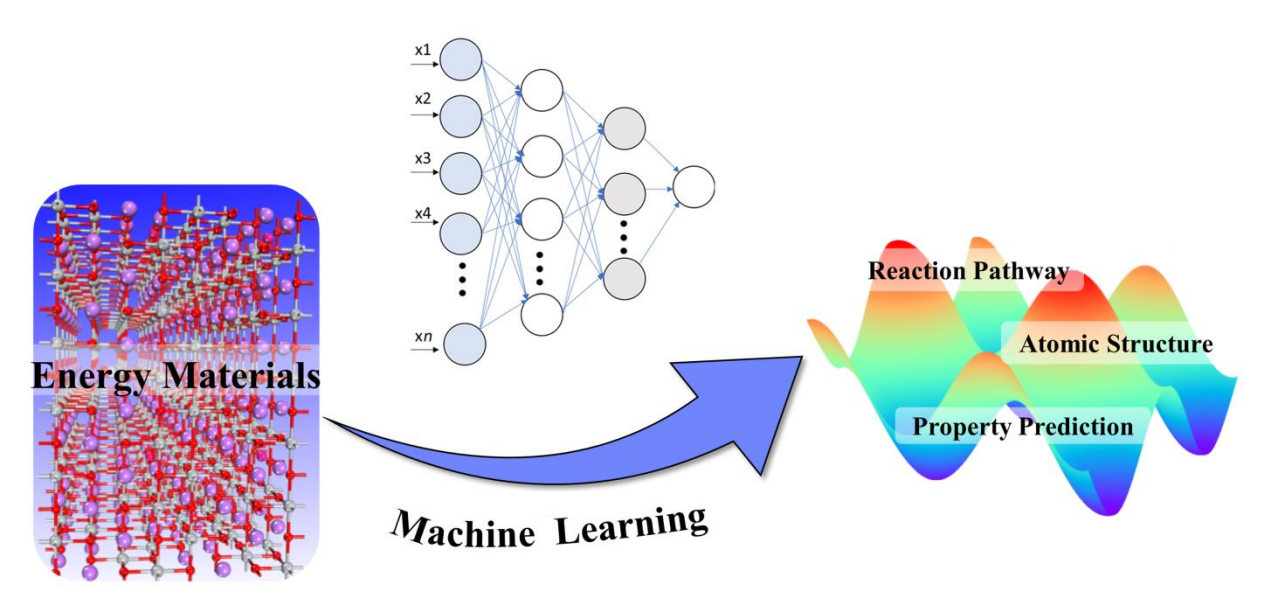
Keywords

Machine learning | Materials science | Atomic simulation | Thermodynamics | Kinetics

Abstract

Energy materials featuring the capability to store and release chemical energy reversibly involve generally complex geometrical structures with multiple elements. It has been a great challenge to establish the quantitative relationship between the structure of materials and their dynamic physicochemical properties. In recent years, machine learning (ML) technique has demonstrated its great power in accelerating the research on energy materials. This topical review introduces the key ingredients and typical applications of ML to energy materials. We mainly focus on the ML based atomic simulation via ML potentials in different architectures/implementations, including high dimensional neural networks (HDNN), Gaussian approximation potential (GAP), moment tensor potentials (MTP) and stochastic surface walking global optimization with global neural network potential (SSW-NN) method. Three cases studies, namely, Si, LiC and LiTiO systems, are presented to demonstrate the ability of ML simulation in assessing the thermodynamics and kinetics of complex material systems. We highlight that the SSW-NN method provides an automated solution for global potential energy surface data collection, ML potential construction and ML simulation, which boosts the current ability for large-scale atomic simulation and thus holds the great promise for fast property evaluation and material discovery.

Comprehensive Graphic Content



*E-mail: cshang@fudan.edu.cn, zpliu@fudan.edu.cn
† Dedicated to the Special Issue of Xplorer Prize in 2020.

View HTML Article

Shu-Hui Guan (left) received her PhD degree from Fudan University (Shanghai, China) in 2016 under the supervision of Prof. Zhi-Pan Liu. She went to Shanghai Academy of Agricultural Sciences as a full-time research worker, joining the persistent organic pollutants group. Her current research focuses on the atomic simulation of solid materials and its applications on environments.

Cheng Shang (middle) graduated in Chemistry at Fudan University where he also received his PhD degree in 2013 under the supervision of Prof. Zhi-Pan Liu. He then moved to the University of Cambridge, where he joined the theory group supervised by Prof. David Wales. In 2015 he joined Fudan University, where he currently holds a position as an associate professor in Physical Chemistry. His current research interests are mainly focused on methodology development for potential energy surface exploration and applications on crystal structure predication.

Zhi-Pan Liu (right) graduated in Chemistry at Shanghai Jiao Tong University (Shanghai, China). He received his PhD degree in 2003 at Queens University of Belfast under the supervision of Prof. Peijun Hu. He then moved to the University of Cambridge, where he joined the surface science group supervised by Prof. Sir David King. In 2005 he joined Fudan University, where he has since stayed, with a position as a Changjiang professor in Physical Chemistry. His current research interests include methodology development for potential energy surface exploration, theory on heterogeneous catalysis, and machine-learning-based atomic simulation.

Contents

1. Introduction

The innovation of energy materials is a key force to shape a better, cleaner and more sustainable future. The properties of energy materials are influenced by many aspects of material, from geometrical to electronic structure. It is thus of fundamental importance to establish the quantitative structure-property relationship for designing better materials. However, due to the structure complexity during the energy storing and releasing, contemporary experimental techniques were frustrated to achieve both high spatial and high temporal resolution in the operational conditions.[1-3] On the other hand, the computational simulation has long been a valuable complement to experiment, which is able to follow closely the material transformation with fine details. To date, a range of computational techniques are available to simulate systems at different scales, from the electronic level based on quantum mechanics (QM),[4-6] to the atomic level via molecular dynamics (MD),[7-8] to the mesoscopic level using Metropolis/kinetics Monte Carlo techniques,[9] and to the macroscopical level using the phase-field method[10-11] and continuum macroscopic approaches. Among them, the atomic simulation is of fundamental importance, which is the key linkage to correlate the electronic structure of material with their dynamic properties.

The main purpose of atomic simulation is to establish the (free) energy landscape of material, based on which the configuration dependent thermodynamics and kinetics properties can be derived based on statistics mechanics. Thus, the total energy evaluation is the foundation of atomic simulation, dictating largely the efficiency and accuracy of energy landscape. To date, QM calculation, often via first principles density functional theory (DFT), is the first-choice work horse in total energy evaluation. While it is certainly more reliable, QM calculation is limited by the poor scaling in treating large size systems. As a result, DFT calculation is most applied to relatively small systems in relatively short time simulation, *e.g.*, within a few hundreds of atoms and less than a few hundreds of picoseconds. This limitation in system size severely hampers the predictive power of QM. For example, to design energy materials, the exhaustive phase space search to examine diverse structure configurations and phases[12-14] and the long-time atomic simulation are often two key ingredients in making property prediction, both of which, however, require a fast and robust total energy calculation tool.

It is no wonder that the empirical force field method is widely used for large-scale material simulation in the past decades. The force field method has a high efficiency in evaluating total energy of system for its simple mathematic form of interatomic potentials. However, the force field method often lacks transferability and high accuracy, especially, for systems with complex chemical bondings. For unknown solid materials with multiple elements typical in energy materials, the force field method is much less popular due to the great difficulty to construct the robust interatomic potential.

In recent years, machine learning (ML) technique, particularly, using the deep-learning methods, demonstrates its great value in solving challenging scientific problems,[3] where the traditional methods fail to do so. In the field of material science, ML is being utilized as an efficient tool to correlate material performances with their chemical compositions, reveal the underlying physics of material, and accelerate the material discovery. Among the large volume of ML applications, ML potential method[15-16] has attracted considerable interests for its great promise in treating large-scale systems. Unlike the empirical force field potentials, ML potential can in principle achieve a high accuracy even for complex material systems involving chemical reactions as long as the training dataset covers the essential structural patterns from the target potential energy surface (PES). Therefore, ML potential method is becoming a convenient and robust choice for large-scale material simulation, where QM simulations are simply too expensive and the force field calculations are not available or too inaccurate. Nevertheless, the construction of ML potential is highly computational demanding, requiring a large and representative PES dataset that is computed *a priori* by QM calculations. Since the quality of the PES dataset is critical, determining largely the predictivity of ML potential, many PES sampling methods have been tested for this purpose, including the simulated annealing[17] PES sampling method, basin-hopping method,[18] genetic algorithm,[19] metadynamics[20] and stochastic surface walking (SSW) methods.[21-22]

This focused review serves to outline the recent progress in ML applications for structure and property predictions of energy materials. We will introduce the general workflow of ML technique in predicting material property, and overview the representative applications in recent years. We will then move on to

outline the representative methodologies on ML potentials, including high dimensional neural networks (HDNN), Gaussian approximation potential (GAP), moment tensor potentials (MTP) and stochastic surface walking global optimization with global neural network potential (SSW-NN) method[23-24] developed in our group, and illustrate their applications on energy materials. Finally, the challenges of ML potentials in atomistic simulations are discussed.

2. ML for Energy Materials

2.1. Workflow of ML for problem solving

A successful application of ML for material prediction relies on (i) the quality of the dataset and (ii) the ML model where specific algorithms are utilized to correlate the dataset with target properties. The dataset can either from experiment or from QM calculations, which should contain the basic information on the materials and their performances associated with the material properties. In general, the workflow of ML applications should contain four essential steps (Figure 1a), which are elaborated in the following:[25-27]

(i) Data collection and curation. This involves the generation and selection of a relevant subset of available data. While sufficient and representative data samples are desirable for building a reliable ML model, the amount of the data depends much on the source and the ease of data generation in reality. For example, in training ML potentials for atomic simulation, the dataset that describes the PES can be very large, often more than a few tens of thousands of structures that are computed from QM calculations.

(ii) Featuring engineering. The features of materials are extracted and quantified by appropriate mathematic representations that correlate features with known physical and chemical quantities of material, including the geometrical and electronic information. The number and the quality of the features are crucial to a ML model and often determine the upper limit of its performance.

(iii) Model testing and optimization. As there are a varieties of ML algorithms, this step involves the testing of different algorithms, the selection of the best, the training of the data and benchmarking predictions. The testing is via the evaluation of the prediction of the validation dataset that is a subset of total data and not included in training set. Parameters that cannot be learned (the so-called hyperparameters) are to be optimized during the testing stage. The common validation method includes K-fold cross-validation, leave-one-outcross-validation (LOOCV), repeating learning test (RLT) cross-validation, and bootstrap cross-validation.[28]

(iv) Application. The ML model is adopted to predict unknown/unexplored material systems and the prediction should be examined to further validate the model. To improve the prediction and promulgate the unknown, the model should be re-trained by adding the data that are poorly predicted.

2.2. ML applications for energy materials

ML technique as a general-purpose tool can be applied to different aspects of energy material research, from energy storage (*e.g.*, fuel cells, supercapacitor, *etc.*) to energy-related catalysis (*e.g.*, noble metal for CO_2 reduction), and to energy conversion (*e.g.*, thermoelectrics),[29-32] as shown in Figure 1b. It is most popular that ML models are utilized to provide a rapid prediction on material properties, which facilitates to find materials with desirable properties.[3,33] Broadly speaking, these applications may be divided into two types according to their data source and the physical laws involved, namely, the experiment-based and first-principles approaches. The former makes prediction via pure mathematics based on existing experimental data. The latter learns properties from QM calculations and often involves extensive atomic simulation to compute properties. It can thus be regarded as a high-speed upgrade to QM calculations, as shown in Figure 1b.

Experiment-based approach. This approach utilizes experimental data to build ML models. The aim of ML is to correlate various experimental conditions with the target properties. The input of ML model, *i.e.*, data descriptors, are typically the controlling parameters from experiment, such as temperature, pressure, material elements, compositions, *etc.*, and the output of ML model is the predicted property, which can be measured by different characterization techniques, such as conductivity, bandgap, peak positions of spectra, power factor, *etc.*

The typical ML application for material discovery is to discover the optimum ratio of element. Hou *et al.*[34] employed a Gaussian process regression (GPR) model to predict power factor (PF) for unknown compositions of AlFeSi system (a potential thermoelectric material). The flowchart for the design of intermetallic $Al_{23.5+x}Fe_{36.5}Si_{40-x}$ (off-stoichiometric samples of the $Al_2Fe_3Si_3$ compound) with a higher PF is summarized in Figure 2, where the

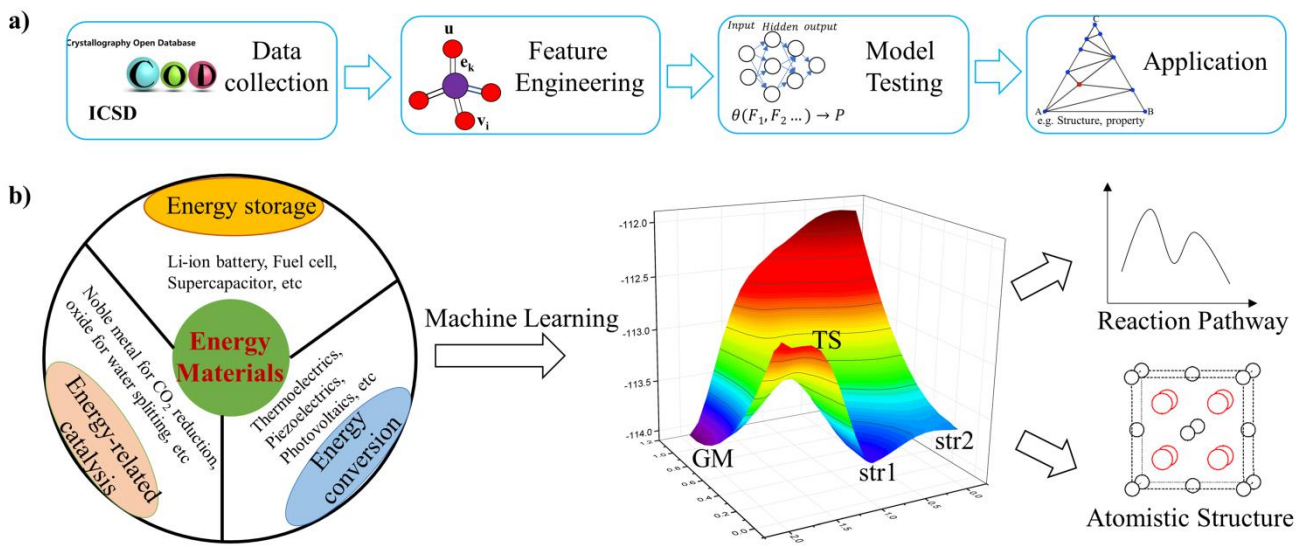


Figure 1 Workflow of machine learning for structure and property prediction (a) and its application in energy materials via PES exploration by atomic simulation (b).

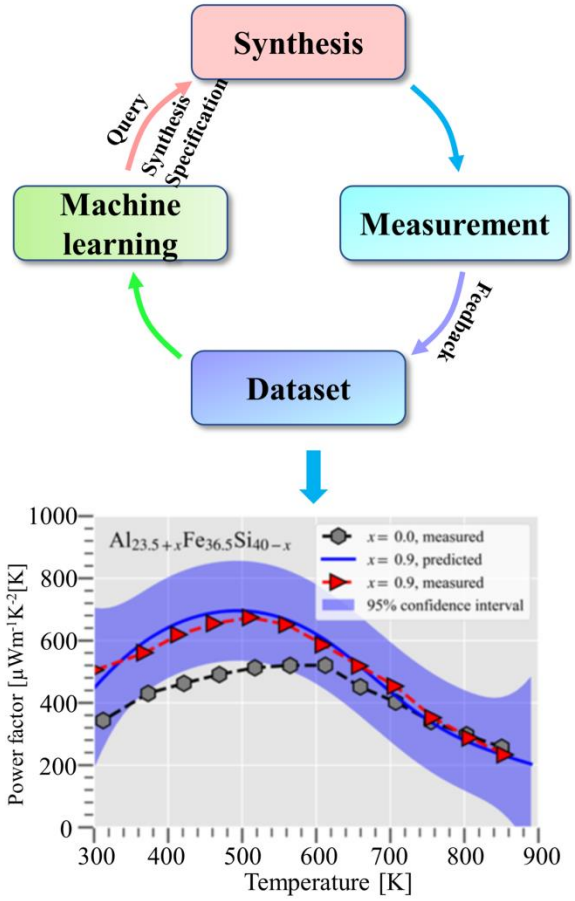


Figure 2 Framework for the design of $Al_{23.5+x}Fe_{36.5}Si_{40-x}$ toward an optimized power factor by learning from experimental data. Reprinted with permission from Ref [33]. (Copyright 2019 American Chemical Society).

input features contain the composition *x* and the experimental temperature. The initial dataset consists of PF measurements from experiments on several compositions at certain temperatures (*T* = 300—800 K), based on which the model is trained. Then this model is used to predict the PFs of the unexplored compositions, and the ones with the highest PFs in the temperature range of 450—650 K are chosen for further synthesis and property measurement. Next, with the new feedback the dataset is expanded, and the ML model can be improved by re-training, which restarts the iteration. After several iterations, the optimal Al/Si ratio for a high PF was found with the optimal ratio of the sample *x* = 0.9 that increases the PF at ~510 K by about 40% with respect to that of the initial sample at *x* = 0.0. Similar ML models were developed for $RE_{12}Co_5Bi$ (*RE*=Gd, Er) thermoelectric materials,[35] superionic conductive materials[36] and ABX_3 halide perovskites materials,[37] by using various ML techniques, such as random forest model,[38] logistic regression model, support vector classification.

Since there are a large volume of published experimental data, ML-assisted literature analysis, *i.e.*, the data mining technology, has been considered for identifying the useful experimental data and eliminating the inconsistency between data. Attempts have been made to extract data and insights from literatures using text mining. For example, in the thermoelectrics field, the relevant literature consists of 3.3 million published papers between 1992 and 2018 in more than 1000 journals. Tshitoyan *et al.*[39] have shown that, by feeding these abstracts to a "machine", one can obtain embeddings of words that incorporate the chemical relationships. Without any explicit insertion of chemical knowledge, these embeddings capture complex materials science concepts such as the underlying structure of the periodic table and structure-property relationships in materials. The trained embeddings can be used as a powerful tool for discovering new candidate.

Another effective approach to expand the dataset involves the unsupervised learning, which does not require labelled data with properties and thus alleviates the data scarcity problem. For finding new solid-state Li-ion conductors, a modified XRD (mXRD) feature was developed by Zhang *et al.*[40] to describe the anion lattice of Li-containing compounds. They showed that the compounds that have similar conductivity tend to be close in mXRD feature space, revealing the unique structure-property relationship between anion lattice and Li^+ conduction over a large materials space. As a result, 16 new compounds with conductivities higher than 10^{-4} $S\cdot cm^{-1}$ were predicted. These discovered candidates have highly different structures and chemical compositions from known fast Li-ion conductors, demonstrating the capability of unsupervised learning for discovering materials over a wide materials space.

First-principles approaches. The alternative way for ML to make prediction is to learn from first principles data, *i.e.*, without any input from experiment. The straightforward way is simply replacing experimental data with first principles data since it is often much more convenient to produce some property data by calculations (*e.g.*, battery voltages, ionic conductivity, bandgaps, *etc.*). Consistently, the data descriptors of ML model need to be changed to the available structural information from calculations. Indeed, recent years have seen a large volume of ML-based material discovery and property prediction using first principles data. For example, to identify the rechargeable Li-ion batteries electrolyte material with fast Li^+ hopping, Jalem *et al.* have investigated 66 olivine-type oxides with an ordered structure $LiMXO_4$.[41] As shown in Figure 3a, they extracted 42 original variables (OVs) including lattice parameters, Born effective charges of cations, Li-M distance at edge sharing, and intra/inter-polyhedron parameters from DFT-relaxed structures and adopted the partial least-square (PLS) model for ML. The fitting of the model (Figure 3b) yields a good agreement between DFT calculations and PLS-predicted results for the Li-ion hopping energy (< 35 meV). They revealed that an increase in the ionic size of M results in a significant distortion of the M octahedron, which, in turn, leads to a slower Li-ion migration caused by the energy-penalizing local lattice distortion. Similarly, the properties of pervoskites, mainly the stability and bandgap, have been investigated using ML methods by Sun and Yin, [42] Im *et al.*,[43] and Li *et al.*[44]

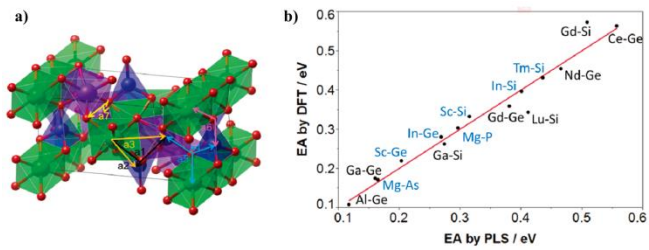


Figure 3 a) Interpolyhedraon parameters within the unit cell that were used in PLS model building, the label corresponding to the angle descriptor and more descriptors could be seen in literature [40]. b) Calculated hopping energies (EA) via NEB method versus predicted hopping energies via PLS model for different M-X pairs of olivine compositions. Blue data points represent M-X pairs with published experimental data in Inorganic Crystal Structure Database (ICSD).[40] (Copyright 2012 American Chemical Society).

A more general but rather sophisticate way of ML-guided material discovery involves the atomic simulation based on ML potentials. These potentials need to be constructed first by learning

the PES data and the subsequent large-scale atomic simulation using the ML potential will result in the material property and thus facilitate the material discovery. To date, the ML potential method has demonstrated its power in treating complex PES problems for many energy materials,[26-27] *e.g.*, $LiMO_2$,[45] Li_xSi,[46] $Ca_3Al_2Si_3O_{12}$,[47] Li_3N,[48] LiC_x,[49] *etc*. The ML potentials can be based on neural network (NN) framework[50] and many other forms, such as Gaussian approximation potential (GAP),[49] spectral neighbour analysis potential (SNAP) model,[48] support Vector Machines (SVMs)[51] and moment tensor potentials (MTP),[52-53] as illustrated in Figure 4. Below we will briefly overview several key ML potential methods and illustrate their applications in energy material systems by providing three case studies.

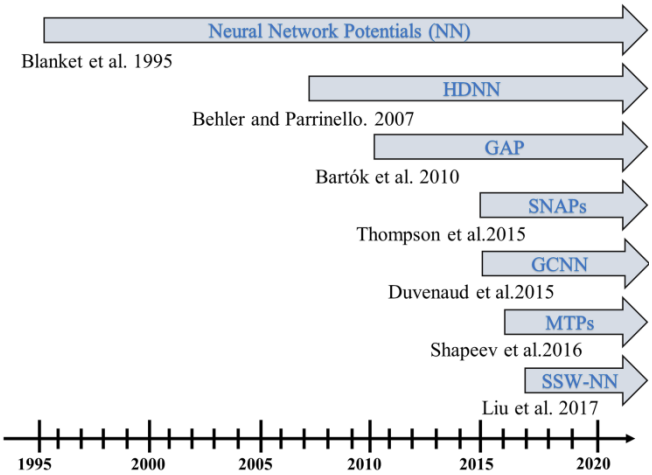


Figure 4 Historic development of machine learning potentials.

3. Overview of ML Potentials

HDNN method. Among ML potentials, the Neural Network (NN)[50] based potential is no question the most sophisticate and more promising one suitable for treating complex system. While this machine-learning technique was inspired by the central nervous system and has been developed since 1990s, the high-dimensional neural network (HDNN) scheme introduced by Behler and Parrinello represents an important step toward the routine use of NN-PESs in computational chemistry.[54-56] In the HDNN scheme, the total energy of system is assumed to be a sum of individual atomic energy contributions as written in Eq. 1.

$$E = \sum_i^N E_i \tag{1}$$

Each atomic energy can then be fitted as a function of the chemical environment of each atom that are obtained by NN machine learning. The NN parameters can be trained via the standard back-propagation technique by learning quantum mechanics PES dataset. Behler and Parrinello proposed a special set of atom centered symmetry functions (ACSFs)[57] as the input of NN, which can describe the environment of atom. These functions are confined in certain cutoff range f_c, *e.g.*, in Eq. 2,[57] that are at least smooth up to the first derivative. In the cutoff function, R_{ij} is the distance between atoms *i* and *j*, and if R_{ij} is larger than the cutoff radius R_c, the cutoff function and its derivative become zero.

$$f_c(R_{ij}) = \begin{cases} 0.5 \times \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1\right] & for\ R_{ij} \le R_c \\ 0 & for\ R_{ij} > R_c \end{cases} \tag{2}$$

The most used two-body G^2 and three-body G^4 are shown in Eqs.3—4,

$$G_i^2 = \sum_{j \ne i} e^{-\eta(r-r_s)^2} \cdot f_c(R_{ij}) \tag{3}$$

$$G_i^4 = 2^{1-\zeta} \sum_{j,k \ne j}^{all} (1 + \lambda \cos\theta_{ijk}{}^{\zeta}) \cdot e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} \cdot f_c(R_{ij}) \cdot f_c(R_{ik}) \cdot f_c(R_{jk}) \tag{4}$$

where θ_{ijk} is the angle centered at the *i* atom with neighbors *j* and *k*. Due to the usage of internal coordinates, pair distance and angles among atoms, these structural descriptors are rotational and permutational invariant, and thus can produce a continuous PES for deriving the first energy derivatives, *i.e.*, the atomic force. The utility of HDNN potentials is demonstrated by a number of applications to solid state systems,[58-62] surface interactions,[63-64] and water clusters.[65-66]

GAP and SNAP method. Gaussian approximation potential (GAP) and its successor, the spectral neighbor analysis potential (SNAP), are another type of ML potential being widely used. GAP framework proposed by Bartók[67-68] in 2010 adopts a Gaussian process regression technique to establish the connection between structure and energy. In this model, the atomic energies E_i in Eq.1 are interpolated in the bispectrum descriptor space using Gaussian process regression, as given by

$$E_i = \sum_{n=1}^{N_{\text{ref}}} \alpha_n e^{-0.5 \cdot \sum_{l=1}^{L} \left[\frac{(b_l - b_{n,l})}{\theta_l}\right]^2} = \sum_{n=1}^{N_{\text{ref}}} \alpha_n e^{G(b, b_n)} \tag{5}$$

where *N* and *l* range over the reference configurations and bispectrum components, respectively, and the θ_l is fitting parameter (hyperparameter). Thus, the atomic energy is a weighted sum over the energies of the atomic environments in the reference training set. The parameter α_n is determined by the inversion of the covariance matrix, which is a rather demanding operation.[69] The GAP potential was successfully used for material systems, such as GaN,[67] metal systems (*e.g.*, W),[70] silicene[71] and molecular systems.[72]

Spectral neighbor analysis potential (SNAP) introduced by Thompson *et al.*[73] in 2015 represents a linear version of GAPs and is based on the same bispectrum components. The atomic energy of atom *i* in SNAP is given as a linear combination of the *K* projected bispectrum components by

$$E_i = \beta_0^{\alpha_i} + \sum_{k=1}^{K} \beta_k^{\alpha_i} \cdot B_k^i \tag{6}$$

where the coefficient β_k^{α} is determined via least squares linear regression of the reference dataset. α_i distinguishes the atom type of atom *i*, B_k^i is the bispectrum component *k* of atom *i*, and $\beta_0^{\alpha_i}$ is a constant element-specific contribution. The SNAP method was utilized to describe the solid phases of tantalum and the liquid structures.[73-74]

GCNN. Convolutional neural network was first introduced by Duvenaud *et al.* in 2015, that operates directly on graphs.[75] Because 3D molecules and materials can be viewed as fully connected graphs with nodes and edges to represent the atoms and bonds, respectively, the features of them can be automatically learned via a feature hierarchy in CNN by applying multistage concatenated convolution operations.[75-77] This automated feature generation is much simpler than the previous methods that rely on manually constructed structure descriptors. GCNN has been proved to be successful for a variety of systems, *e.g.*, small molecules (benzene, toluene),[78] large proteins (deca alanine Ala_{10}),[79] crystalline system ($ThCr_2Si_2$-type materials)[80] and aqueous alkaline electrolyte (NaOH).[81] SchNet is a leading example of GCNN.[77-78] The computational efficiency of GCNN could be the major concern due to the high cost in feature generation and model training of large dataset.

MTPs. The moment tensor potential[52-53] was proposed by Shapeev in 2016, which adopts a linear regression model with polynominal-like functions of atomic coordinates as the basis

functions. The descriptors of MTPs can be interpreted as tensors of inertia of atomic environments. Because of the efficiency of its polynominal basis of interatomic distances and angles, MTP is found to be significantly faster than GAP and has already been shown to be able to reach equivalent accuracy for modelling chemical reactions,[82] single-element systems,[83-84] single-phase binary systems,[85] ground states of multicomponent systems[86] or binary metallic system.[87]

SSW-NN method. The SSW-NN method is a solution for constructing the global neural network potential (G-NN) that aims for the global PES exploration.[23-24] In SSW-NN method, the G-NN potential follows the architecture of HDNN potential, but the structure descriptors are replaced by a set of power-type structure descriptors[88] (PTSDs), which are generally described in the forms of the following equations (Eqs. 7—14), and the training dataset of G-NN potential is obtained from the SSW global optimization in a global-to-global scheme. The first "global" indicates the dataset from SSW global optimization and the second "global" means the ability of NN potential to describe the global PES.

$$f_c(r_{ij}) = \begin{cases} 0.5 \times \tanh^3\left[1 - \frac{r_{ij}}{r_c}\right], & for\ r_{ij} \le r_c, \\ 0 & for\ r_{ij} > r_c \end{cases} \quad (7)$$

$$R^n(r_{ij}) = r_{ij}^n \cdot f_c(r_{ij}), \quad (8)$$

$$S_i^1 = \sum_{j\neq i} R^n(r_{ij}), \quad (9)$$

$$S_i^2 = \left[\sum_{m=-L}^{L}\left|\sum_{j\neq i} R^n(r_{ij}) Y_{Lm}(\boldsymbol{r}_{ij})\right|^2\right]^{\frac{1}{2}} \quad (10)$$

$$S_i^3 = 2^{1-\zeta}\sum_{j,k\neq i}(1+\lambda\cos\theta_{ijk})^{\zeta} \cdot R^n(r_{ij}) \cdot R^m(r_{ik}) \cdot R^p(r_{jk}), \quad (11)$$

$$S_i^4 = 2^{1-\zeta}\sum_{j,k\neq i}(1+\lambda\cos\theta_{ijk})^{\zeta} \cdot R^n(r_{ij}) \cdot R^m(r_{ik}), \quad (12)$$

$$S_i^5 = \left[\sum_{m=-L}^{L}\left|\sum_{j,k\neq i} R^n(r_{ij}) \cdot R^m(r_{ik}) \cdot R^p(r_{jk}) \cdot \left(Y_{Lm}(\boldsymbol{r}_{ij}) + Y_{Lm}(\boldsymbol{r}_{ik})\right)\right|^2\right]^{\frac{1}{2}}, \quad (13)$$

$$S_i^6 = 2^{1-\zeta}\sum_{j,k,l\neq i}(1+\lambda\cos\delta_{ijkl})^{\zeta} \cdot R^n(r_{ij}) R^m(r_{ik}) R^p(r_{il}), \quad (14)$$

where r_{ij} is inter-nuclear distance between atom i and j, θ_{ijk} is the angle centred at i atom with j, k being neighbours (i, j, k are atom indices). The key ingredients in PTSD are the cut-off function f_c that decays to zero beyond the r_c (Eq. 7), power-type radial function, trigonometric angular functions and spherical harmonic function. The replacement of the Gaussian-type structural descriptor which is proposed by Behler and Parrinello by the PTSD has several advantages: (i) the computational cost in numerical calculations is reduced; (ii) the adjustable parameters are reduced to one (n) that simplifies the search for optimal parameters for two-body function; (iii) the power function when combining with the decaying cut-off function can create radial distributions with flexible peak and shape, which fulfils the similar purpose of Gaussian function; (iv) the introduction of different powers (n, m, p) in three-body function can couple conveniently different radial distributions; (v) the introduction of spherical function greatly improves the description of the angular environment of atom.

The SSW method[21-22] is a global optimization algorithm that utilizes the second-order information of PES to identify low energy reaction channels. It[21] contains a series of structure operation steps, including the random bias potential generation, the softening of bias potential, the repeated addition of bias potentials, the local optimization and the Metropolis Monte Carlo structure selection. The whole process of SSW is automated without the need to preconfigure the bias potential, and thus convenient to use for generating representative global PES dataset in unknown systems. In one particular SSW step, labeled as t, a modified PES V_{mod}, as shown in Eq.(15), is utilized for moving from the current minimum, $\boldsymbol{R_t^0}$, to a high energy configuration, $\boldsymbol{R_t^H}$ (the climbing), in which a series of bias Gaussian potential V_n (n is the index bias potential, n–1, 2, ⋯, H) is added one by one consecutively along the direction $\boldsymbol{N_t^n}$,

$$V_{\text{mod}} = V_{\text{real}} + \sum_{n=1}^{NG} v_n = V_{\text{real}} + \sum_{n=1}^{NG} w_n * \exp[-((\boldsymbol{R_t} - \boldsymbol{R_t^n}) \cdot \boldsymbol{N_t^n})^2 / (2 \times ds^2)] \quad (15)$$

where $\boldsymbol{R_t}$ is the coordination vector of the structure and V_{real} represents the unmodified PES; $\boldsymbol{R_t^n}$ is the nth local minima along the movement trajectory on the modified PES that is created after adding n Gaussian functions. The Gaussian function is controlled by its height w and its width ds, and is always added along one particular walking direction as defined by $\boldsymbol{N^n}$. Once the $\boldsymbol{R_t^H}$ is reached, all bias potentials are removed and the local optimization is performed to quench the structure to a new minimum. Since its invention in 2013, the SSW method has been successfully used in both aperiodic systems,[22] such as molecules and clusters, and periodic crystals.[89]

As shown in Figure 5, the SSW-NN method involves three tasks iteratively performed to achieve the self-learning of global PES: (i) SSW global PES sampling, initially based on DFT and later using G-NN potential, is utilized to collect PES data; (ii) DFT calculations with a high accuracy setup are then utilized to compute the selected data from SSW, which expands the dataset; (iii) the G-NN potential is trained based on the dataset. The data from SSW trajectories are selected randomly and screened based on two criteria, *i.e.*, the new atomic environment as judged by structure descriptors, and the poor energy and force by comparing DFT and NN predictions. In our practice, up to 100 iterations are often necessary in constructing the G-NN potential. The final G-NN potential can typically achieve the accuracy of 5—10 meV/atom for root mean square errors (RMSE) of energy and 0.1—0.2 eV/ Å for RMSE of force. In the past several years, our group have mapped out the PESs for a number of systems, *e.g.*, single element crystal (boron),[88] molecular crystal (ice),[90] metal oxide (TiO_2),[91] ternary metal oxides ($ZnCr_2O_4$).[92]

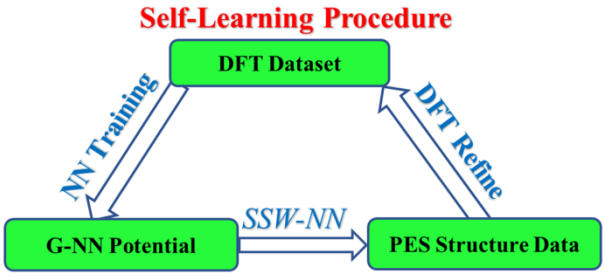


Figure 5 Self-learning procedure of global NN potential. The global dataset is first generated using high accuracy DFT calculations, which is then trained to obtain the global NN potential (G-NN). Then, an additional dataset is generated by SSW sampling based on G-NN potential. This additional dataset is then fed back into global dataset and a new cycle self-learning starts.

The SSW-NN method is now implemented in LASP program (www.lasphub.com), which includes a variety of common atomic simulation tools. As a key feature, LASP provides several TS search and SSW global optimization methods developed in the group, which have been well tested in heterogeneous catalysis and solid phase transition.[92-93] Importantly, it is convenient in LASP plat-

form to exchange data between different modules, particularly, QM calculations, the PES exploration and NN training. We have generated more than 220 G-NN potentials in the past three years, covering most periodic table elements. The LASP code has been used worldwide for materials, catalysis and many other fields.

4. Application of ML Potentials for Energy Materials

Case1: HDNN method to resolve the phase diagram of silicon under high pressures

Silicon is one of the most studied elements in the world for its use in microelectronics, semiconductors, Li-ion battery technologies and its importance in earth science.[94] Under high pressures, the silicon phases can exhibit metallic properties, and even have the superconductivity.[95-96] The study of the Si phase diagram under high pressures is of great challenge for both experiment and theory.

Behler *et al*.[60] performed the metadymanics simulations of the high-pressure phases of silicon by utilizing the HDNN potential. This potential was fitted from DFT calculation results,[54] where 8200 were used for optimizing the NN potential and the remaining 800 were used as an independent test set to investigate the predictive capability of the NN for structures outside the optimization set. The RMSE of the training set is 4—5 meV/atom, and the RMSE of the test set is 5—6 meV. For the NN atomic forces, a RMSE relative to DFT is about 0.2 eV/Å. The NN calculation is about 5 orders of magnitude faster than the DFT and can also provide analytic forces and the stress tensor required in metadynamics simulations.

Silicon at different temperatures and pressures was thus investigated by using the NN-based metadynamics in a 64-atom cell, as shown in Figure 6. Under 300 K and 12 GPa, it is found that the cubic diamond undergoes a very large volume change (~17%) to form the Imma phase, followed by the simple hexagonal (sh) phase until the β-tin phase is reached, which is the stable phase at this pressure (Figure 6). Next, the transitions between these phases at slightly increased pressure of 16 GPa and 300 K have been examined in detail. The enthalpy of the Imma phase is found to be the lowest, in agreement with DFT. In particularly, it is found that the cubic diamond to β-tin transition is strongly first-order and nucleated by the formation of defects. Similarly, the transitions under higher pressure (45 GPa, 50 GPa and 90 GPa) were also investigated, yielding the formation of the *Cmca* phase at 800 K and 45 GPa, the transition to the *hcp* (hexagonal close packed structure) phase at 300 K and 50 GPa, the transition from the *hcp* to the *fcc* phase at 300 K and 90 GPa. The sequence of these phases is consistent with experimental findings[97-98] that β-tin phase exists near 12GPa,[99] which transforms to the *Imma* phase at 16 GPa,[100] and the *hcp* phase has been observed at 42 GPa,[101] which transforms to the *fcc* phase at 80 GPa.[102]

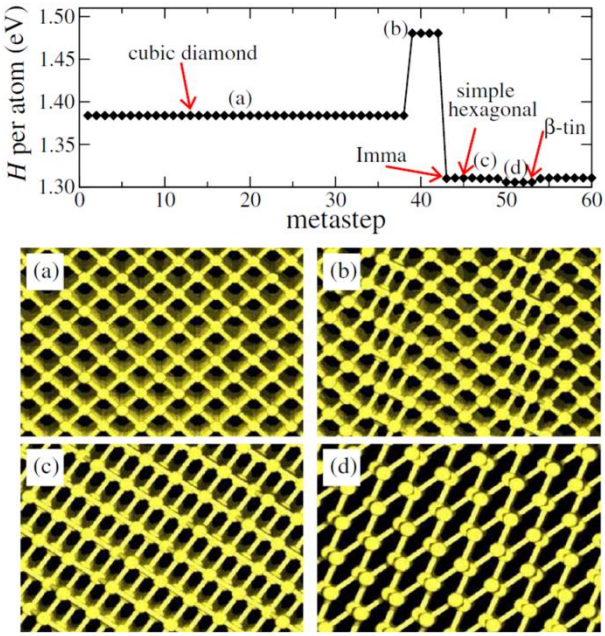


Figure 6 Enthalpies of the optimized crystal structures obtained in a metadynamics simulation at 300 K starting from the cubic diamond structure at an external pressure of 12 GPa using HDNN potential.[59] (Copyright 2008 American Physical Society)

Case 2: GAP method to model Li intercalation in LiC_x system

GAP models have been introduced to study the Li intercalation in carbon nanostructures,[49] which is a key process occurred in the anode of Li-ion battery.[103] The approach employed for producing GAP models was overviewed in Figure 7a, which glues two GAP potentials for C–C interaction model and Li–C interaction

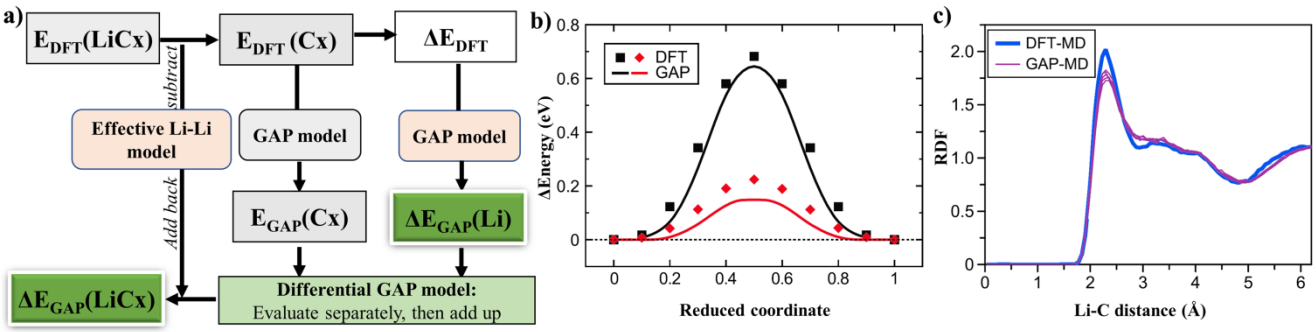


Figure 7 **a)** Overview of the approach employed to generate Li-C GAP potential. Rather than fitting directly to energies (and forces) for Li*Cx* systems, DFT calculations were repeated for the same structures without Li and a GAP model is fitted to the energy and force differences. The latter is combined with a previous GAP model for C–C interactions. A baseline model for long-range Li–Li interactions is subtracted from the reference dataset prior to fitting and added back onto the output. Summing up the terms yields the total GAP energy for the system under study, and similar procedures hold for the forces on atoms. b) the diffusion of an Li atom between two potential-minimum sites in graphite across two high-symmetry pathways using DFT and GAP method. c) Radial distribution function (RDF) analysis for Li–C contacts in this DFT-MD trajectory and for five separate GAP-MD trajectories computed in the same structure.[48] (Copyright 2018 American Institute of Physics)

model with the effective long-range Li–Li interaction model. In order to improve the transferability, the initial training data were produced by randomly placing Li atoms in slightly distorted graphite, graphene and amorphous carbon structures, and the final training set contains 561 graphite, 192 graphene, and 1664 amorphous configurations. Albeit notable remaining numerical energy errors, reaching up to 0.4 eV/atom, the present GAP performs very well in identifying the Li diffusion pathways and reproducing dynamical properties in MD trajectories.

By computed characteristic energy profiles for the most fundamental atomic-scale mechanisms in Li-graphite system, it is found that all three high-symmetry adsorption configurations are correctly captured by the GAP and the diffusion of an Li atom between two potential-minimum sites in graphite is also qualitatively correctly described, with the energy barrier of ~0.13 eV and 0.62 eV for the two low energy pathways (Figure 7b).

Several parallel MD simulations of Li diffusion in a disordered graphite-like structure at 1000 K were further performed and MD trajectories were obtained to analyze the radial distribution function (RDF) of Li environment during the intercalation. As shown in Figure 7c, it shows a maximum at around 2.3 Å, consistent with the DFT data. The vibration densities of states for the host framework and the Li atoms, where the frequency range is commonly associated with the diffusion process, show that in the higher-frequency range (> 15 THz), there is a particularly agreement between DFT and GAP data. While, there is small deviation at lower frequencies, which was considered to be due to the short-time of the calculation and especially the size of the systems from DFT methods as benchmark.

Case 3: SSW-NN method to identify Li_xTiO_2 electrode for Li-ion battery

The rechargeable lithium-ion battery has proved itself as a ground-breaking technology that powers the mobile transportation. Since the graphite anode suffers from low energy density (the engineering limit for the gravimetric capacity C_g is 372 mA·h/g at LiC_6) and low charging speed (the reversible capacity is ~370 mA·h/g at the C/24 rate =15.5 mA/g),[104] significant efforts have been made to search for better Li-ion anode materials, such as Li-M (M = Si, Ge, Sn, *etc.*) alloys[105-106] and lithium transition-metal oxide (*i.e.*, TiO_2).[107-108] A critical quantity for assessing the anode material is the extent of volume expansion at the high capacity, which often occurs due to complex solid reactions and destroys the structural integrity of material.

To search for both high-rate capacity and high stability Li-ion battery, we recently proposed a high-throughput material screening method based on SSW global PES exploration. The PES of lithiated material with the Li/X ratio = 1 for 11 candidates (X = C, Si, TiO_2 and others) were first explored to screen the possible candidates. It is found that TiO_2 and Si are the materials most likely to achieve a balanced performance for capacity and stability,[109] where the delithiated structures for Si are generally less stable than those for TiO_2, suggesting TiO_2 can maintain a better structural integrity during Li intercalation.

Next, a Li-Ti-O three elements G-NN potential was constructed to describe the global PES of LiTiO system. The SSW-NN was utilized to generate a large LiTiO dataset of 50040 structures. The final G-NN potential has five layers NN (324-80-60-60-1), equivalent to 103743 network parameters in total and the RMSE for energy and force are 5.6 meV/atom and 0.16 eV/Å, respectively.[109] The SSW-NN simulation identified the most stable lithiated products of TiO_2 for Li/TiO_2 ratio from 0.5 to 3, including $Li_{0.5}TiO_2$, $LiTiO_2$, Li_2TiO_2, and Li_3TiO_2 for TiO_2. These structures are shown in Figure 8a, which illustrates the general structure evolution trend during the intercalation of Li: the 3-D dense network of TiO_2, originally in anatase form (TiO_2-A), becomes porous, then decays to 2-D layer, and finally decomposes to Ti_2O_3 with O atoms releasing to Li. For Li/TiO_2 ratio at 0.5 to 1, *e.g.*, $Li_{0.5}TiO_2$ and $LiTiO_2$ (denoted as TiO_2-S), the most stable structures are in the spinel form, which has 3-D tunnels to store Li.

The TiO_2-S phase exhibits a facile Li (dis)charging kinetics even at the high Li content, as shown in Figure 8b, where the Li diffusion barrier (0.55 eV) in TiO_2-S is not much different from that in $Li_{0.5}TiO_2$ (0.53 eV). The Li diffusion becomes, however, much more difficult in TiO_2-A at the high Li content ($LiTiO_2$): the diffusion barrier is much higher compared to that at the low Li content ($Li_{0.5}TiO_2$, see Figure 8c). The presence of 3-D tunnels in TiO_2-S is thus critical for fast charging, which is consistent with the fact that nano-porous TiO_2 structures improve the (dis)charging performance of lithium ion battries.[110-111]

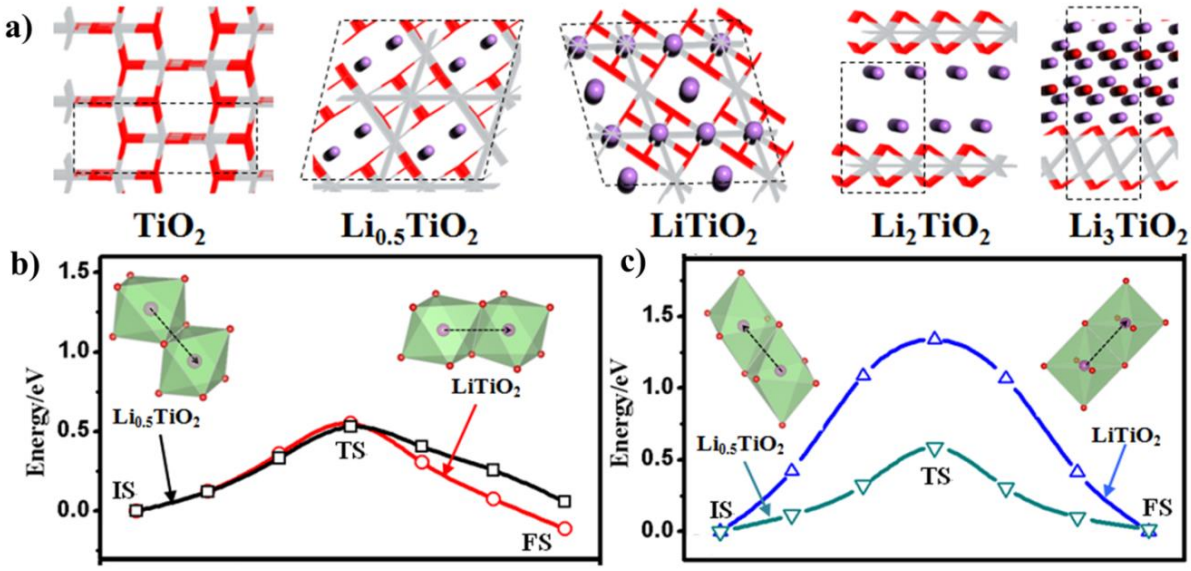


Figure 8 Most stable structures of Li_xTiO_2 obtained from SSW-NN global search (a), Li atoms are represented by magenta balls, while the host TiO_2 is represented by lines. The lowest energy pathway of Li diffusion in (b) TiO_2-S and (c) TiO_2-A at different Li concentrations. Adapted with permission from Ref [101]. (Copyright 2020 American Chemical Society)

5. Conclusions and Perspectives

With the growing impact of ML on material innovation, this review serves to highlight the ML applications for the property prediction and material discovery in the field of energy materials. Among the applications, the atomic simulation based on ML potentials represents one of the most rapid developing field, where ML technique cooperates intimately with quantum mechanics and statistics mechanics for PES description and exploration. Several flavors of ML potential methods, including high dimensional neural networks (HDNN), Gaussian approximation potential (GAP), moment tensor potentials (MTP) and stochastic surface walking global optimization with global neural network potential (SSW-NN) are overviewed and discussed. Atomic simulations based on ML potentials can now be performed to solve problems in chemistry, physics, and materials science that have hitherto been inaccessible due to the lack of suitable potential with first principles accuracy. In this field, the SSW-NN method developed by us is a powerful tool for generating G-NN potential and performing the global PES search. Due to the high cost of generating G-NN potentials, the worldwide colloration between different groups is essential for constructing and benchmarking G-NN potentials with well-defined standards.

In spite of the success, the construction of ML potentials is still very demanding for dataset generation and the potentials need to be tested very carefully due to the lack of physical bases of the numerical functional form. It becomes even more challenging when the structure and composition complexity increases. Consequently, the usage of ML potentails is not as cost effective and as simple as other types of potentials, which hinders the wide usage of these methods. New algorithms to reduce the computational cost for both first principles calculations and PES sampling are called for in the coming years. On the other hand, it is also desirable to benchmark systematically different ML potentials on well-defined datasets to identify new research directions, *e.g.*, combining the advantages of different methods. New approaches are expected to extend the applicability of ML potentails to simulate multiple-element systems and to determine the long-time dynamic properties.

Acknowledgement

This work was supported by Shanghai Sailing Program (19YF1442800), the National Key Research and Development Program of China (2018YFA0208600), and the National Natural Science Foundation of China (22003040, 22033003, 91945301, 91745201 and 21533001).

References

[1] Olson, G. B. Designing a new material world. *Science* **2000**, *288*, 993–998.
[2] Balachandran, P. V.; Kowalski, B.; Sehirlioglu, A.; Lookman, T. Experimental search for high–temperature ferroelectric perovskites guided by two-step machine learning. *Nat. Commun.* **2018**, *9*, 1668.
[3] Edward, S.; Alán, A. G. Use machine learning to find energy materials. *Nature* **2017**, *552*, 23–25.
[4] Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, 864–871.
[5] Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, 1133–1138.
[6] Parr, R. G. Density functional theory of atoms and molecules. In *Horizons of Quantum Chemistry*, Springer, Dordrecht, **1980**, pp. 5–15.
[7] Alder, B. J.; Wainwright, T. E. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* **1959**, *31*, 459–466.
[8] Rahman, A. Correlations in the motion of atoms in liquid argon. *Phys. Rev.* **1964**, *136*, 405–411.
[9] Binder, K. *The Monte Carlo Method in Condensed Matter Physics*, Vol. 71, Springer-Verlag, Berling, **1995**, pp. 285–316.
[10] Chen, L. Q. Phase-field models for microstructure evolution. *Annu. Rev. Mater. Res.* **2002**, *32*, 113–140.
[11] Steinbach, I. Phase–field models in materials science. *Modell. Simul. Mater. Sci. Eng.* **2009**, *17*, 073001.
[12] Khodakov, A. Y.; Chu, W.; Fongarland, P. Advances in the development of novel cobalt Fischer-Tropsch catalysts for synthesis of long–chain hydrocarbons and clean fuels. *Chem. Rev.* **2007**, *107*, 1692–1744.
[13] Seth, S. K.; Saha, I.; Estarellas, C.; Frontera, A.; Kar, T.; Mukhopadhyay, S. Supramolecular self-assembly of M-IDA complexes involving lone-pair...π interactions: crystal structures, hirshfeld surface analysis, and DFT calculations [H2IDA= iminodiacetic acid, M= Cu (II), Ni (II)]. *Cryst. Growth Des.* **2011**, *11*, 3250–3265.
[14] Kotakoski, J.; Krasheninnikov, A.; Kaiser, U.; Meyer, J. From point defects in graphene to two-dimensional amorphous carbon. *Phys. Rev. Lett.* **2011**, *106*, 105505.
[15] Jordan, M. I.; Mitchell, T. M. Machine learning: Trends, perspective, and prospects. *Science* **2015**, *349*, 255–260.
[16] Kohonen, T. An introduction to neural computing. *Neural Networks* **1988**, *1*, 3–16.
[17] Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.
[18] Wales, D. J.; Doye, J. P. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
[19] Sivanandam, S.; Deepa, S. *Introduction to Genetic Algorithms*, Springer, Berlin, **2008**, pp. 165–209.
[20] Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*, 126601.
[21] Shang, C.; Liu, Z. P. Stochastic Surface Walking Method for Structure Prediction and Pathway Searching. *J. Chem. Theory Comput.* **2013**, *9*, 1838–1845.
[22] Zhang, X. J.; Shang, C.; Liu, Z. P. From Atoms to Fullerene: Stochastic Surface Walking Solution for Automated Structure Prediction of Complex Material. *J. Chem. Theory Comput.* **2013**, *9*, 3252–3260.
[23] Huang, S. D.; Shang, C.; Zhang, X. J.; Liu, Z. P. Material discovery by combining stochastic surface walking global optimization with a neural network. *Chem. Sci.* **2017**, *8*, 6327–6337.
[24] Huang, S. D.; Shang, C.; Kang, P. L.; Zhang, X. J.; Liu, Z. P. LASP: Fast global potential energy surface exploration. *WIREs Comput. Mol. Sci.* **2019**, *9*, e1415.
[25] Wei, J.; Chu, X.; Sun, X. Y.; Xu, K.; Deng, H. X.; Chen, J.; Wei, Z.; Lei, M. Machine learning in materials science. *InfoMat* **2019**, *1*, 338–358.
[26] Wang, H.; Ji, Y.; Li, Y. Simulation and design of energy materials accelerated by machine learning. *WIREs Comput. Mol. Sci* **2019**, *10*, e1421.
[27] Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **2020**, *10*, 1903242.
[28] Picard R. R. ; Cook, R. D. Cross-validation of regression models. *J. Am. Stat. Assoc.* **1984**, *79*, 575–583.
[29] Ma, X.; Li, Z.; Achenie, L. E.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO_2 Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 3528–3533.
[30] Liu, M.; Pang, Y.; Zhang, B.; De Luna, P.; Voznyy, O.; Xu, J.; Zheng, X.; Dinh, C. T.; Fan, F.; Cao, C.; de Arquer, F. P.; Safaei, T. S.; Mepham, A.; Klinkova, A.; Kumacheva, E.; Filleter, T.; Sinton, D.; Kelley, S. O.; Sargent, E. H. Enhanced electrocatalytic CO_2 reduction via field-induced reagent concentration. *Nature* **2016**, *537*, 382–386.
[31] Gómez–Bombarelli, R.; Aguilera–Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood–Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; Markopoulos, G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.;

Adams, R. P.; Aspuru-Guzik, A. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **2016**, *15*, 1120–1127.

[32] Li, D.; Li, X.; Zhang, Y.; Sun, L.; Yuan, S. Four Methods to Estimate Minimum Miscibility Pressure of CO_2-Oil Based on Machine Learning. *Chin. J. Chem.* **2019**, *37*, 1271–1278.

[33] Ji, H.; Jung, Y. Artificial neural network for the configuration problem in solids. *J. Chem. Phys.* **2017**, *146*, 064103.

[34] Hou, Z.; Takagiwa, Y.; Shinohara, Y.; Xu, Y.; Tsuda, K. Machine-Learning-Assisted Development and Theoretical Consideration for the $Al_2Fe_3Si_3$ Thermoelectric Material. *ACS Appl. Mater. Interfaces* **2019**, *11*, 11545–11554.

[35] Gaultois, M. W.; Oliynyk, A. O.; Mar, A.; Sparks, T. D.; Mulholland, G. J.; Meredig, B. Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Mater.* **2016**, *4*, 053213

[36] Sendek, A. D.; Yang, Q.; Cubuk, E. D.; Duerloo, K. A. N.; Cui, Y.; Reed, E. J. Holistic computational structure screening of more than 12 000 candidates for solid lithium-ion conductor materials. *Energy Environ. Sci.* **2017**, *10*, 306–320.

[37] Pilania, G.; Balachandran, P. V.; Kim, C.; Lookman, T. Finding New Perovskite Halides via Machine Learning. *Front. Mater.* **2016**, *3*, 19.

[38] Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32.

[39] Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.; Persson, K. A.; Ceder, G.; Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **2019**, *571*, 95–98.

[40] Zhang, Y.; He, X.; Chen, Z.; Bai, Q.; Nolan, A. M.; Roberts, C. A.; Banerjee, D.; Matsunaga, T.; Mo, Y.; Ling, C. Unsupervised discovery of solid-state lithium ion conductors. *Nat. Commun.* **2019**, *10*, 5260.

[41] Jalem, R.; Aoyama, T.; Nakayama, M.; Nogami, M. Multivariate Method-Assisted Ab Initio Study of Olivine-Type $LiMXO_4$ (Main Group M^{2+}-$X^{5+}$and M^{3+}-X^{4+}) Compositions as Potential Solid Electrolytes. *Chem. Mater.* **2012**, *24*, 1357–1364.

[42] Sun, Q.; Yin, W. J. Thermodynamic Stability Trend of Cubic Perovskites. *J. Am. Chem. Soc.* **2017**, *139*, 14905–14908.

[43] Im, J.; Lee, S.; Ko, T. W.; Kim, H. W.; Hyon, Y.; Chang, H. Identifying Pb-free perovskites for solar cells by machine learning. *npj Comput. Mater.* **2019**, *5*, 37.

[44] Li, Z.; Xu, Q.; Sun, Q.; Hou, Z.; Yin, W. J. Thermodynamic Stability Landscape of Halide Double Perovskites via High-Throughput Computing and Machine Learning. *Adv. Funct. Mater.* **2019**, *29*, 1807280.

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

[46] Artrith, N.; Urban, A.; Ceder, G. Constructing first-principles phase diagrams of amorphous LixSi using machine–learning–assisted sampling with an evolutionary algorithm. *J. Chem. Phys.* **2018**, *148*, 241711.

[47] Ye, W.; Chen, C.; Wang, Z.; Chu, I. H.; Ong, S. P. Deep neural networks for accurate predictions of crystal stability. *Nat. Commun.* **2018**, *9*, 3800.

[48] Deng, Z.; Chen, C.; Li, X. G.; Ong, S. P. An electrostatic spectral neighbor analysis potential for lithium nitride. *npj Comput. Mater.* **2019**, *5*, 75.

[49] Fujikake, S.; Deringer, V. L.; Lee, T. H.; Krynski, M.; Elliott, S. R.; Csanyi, G. Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures. *J. Chem. Phys.* **2018**, *148*, 241714.

[50] Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural network models of potential energy surfaces. *J. Chem. Phys.* **1995**, *103*, 4129–4137.

[51] Balabin, R. M.; Lomakina, E. I. Support vector machine regression (LS-SVM)-an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? *Phys. Chem. Phys. Chem.* **2011**, *13*, 11710–11718.

[52] Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Sim.* **2016**, *14*, 1153–1173.

[53] Podryabinkin, E. V.; Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **2017**, *140*, 171–180.

[54] Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

[55] Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *J. Phys. Condens. Matt.* **2014**, *26*, 183001.

[56] Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Int. Ed.* **2017**, *56*, 12828–12840.

[57] Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106.

[58] Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Graphite-diamond phase coexistence study employing a neural-network mapping of the ab initio potential energy surface. *Phys. Rev. B* **2010**, *81*, 100103(R).

[59] Eshet, H.; Khaliullin, R. Z.; Kühne, T. D.; Behler, J.; Parrinello, M. Ab initio quality neural-network potential for sodium. *Phys. Rev. B* **2010**, *81*, 184107.

[60] Behler, J.; Martonak, R.; Donadio, D.; Parrinello, M. Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. *Phys. Rev. Lett.* **2008**, *100*, 185501.

[61] Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **2011**, *83*, 079914.

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

[63] Seema, P.; Behler, J.; Marx, D. Adsorption of Methanethiolate and Atomic Sulfur at the Cu(111) Surface: A Computational Study. *J. Phys. Chem. C* **2013**, *117*, 337–348.

[64] Carbogno, C.; Behler, J.; Grob, A.; Reuter, K. Fingerprints for spin-selection rules in the interaction dynamics of O_2 at Al(111). *Phys. Rev. Lett.* **2008**, *101*, 096104.

[65] Morawietz, T.; Sharma, V.; Behler, J. A neural network potential energy surface for the water dimer based on environment-dependent atomic energies and charges. *J. Chem. Phys.* **2012**, *136*, 064103.

[66] Morawietz, T.; Behler, J. A Full–Dimensional Neural Network Potential Energy Surface for Water Clusters up to the Hexamer. *Z. Phys. Chem.* **2013**, *227*, 1559–1581.

[67] Bartók, A. P.; Payne, M. C.; Kondor, R.; Csanyi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

[68] Bartók, A. P.; Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051–1057.

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

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

[71] Zhang, C.; Sun, Q. Gaussian approximation potential for studying the thermal conductivity of silicene. *J. Appl. Phys.* **2019**, *126*, 105103.

[72] John, S. T.; Csányi, G. b. Many-Body Coarse-Grained Interactions Using Gaussian Approximation Potentials. *J. Phys. Chem. B* **2017**, *121*, 10934–10949.

[73] Thompsona, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum–accurate interatomic potentials. *J. Comput. Phys.* **2015**, *285*, 316–330.

[74] Wood, M. A.; Thompson, A. P. Extending the accuracy of the SNAP interatomic potential form. *J. Chem. Phys.* **2018**, *148*, 241721.

[75] Duvenaud, D.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. In *Advances in Neural Information Processing Systems 28*, Curran Associates, Inc., **2015**, pp. 2224–2232.
[76] Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol .Des.* **2016**, *30*, 595–608.
[77] Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Muller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.
[78] Schütt, K. T.; Sauceda, H. E.; Kindermans, P. J.; Tkatchenko, A.; Muller, K. R. SchNet -A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.
[79] Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.
[80] Park, C. W.; Wolverton, C. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Phys. Rev. Mater.* **2020**, *4*, 063801.
[81] Shao, Y.; Hellstrom, M.; Mitev, P. D.; Knijff, L.; Zhang, C. PiNN: A Python Library for Building Atomic Neural Networks of Molecules and Materials. *J. Chem. inf. Model.* **2020**, *60*, 1184–1193.
[82] Novikov, I. S.; Suleimanov, Y. V.; Shapeev, A. V. Automated calculation of thermal rate coefficients using ring polymer molecular dynamics and machine-learning interatomic potentials with active learning. *Phys. Chem. Chem. Phys.* **2018**, *20*, 29503–29512.
[83] Podryabinkin, E. V.; Tikhonov, E. V.; Shapeev, A. V.; Oganov, A. R. Accelerating crystal structure prediction by machine–learning interatomic potentials with active learning. *Phys. Rev. B* **2019**, *99*, 064114.
[84] Novoselov, I. I.; Yanilkin, A. V.; Shapeev, A. V.; Podryabinkin, E. V. Moment tensor potentials as a promising tool to study diffusion processes. *Comput. Mater. Sci.* **2019**, *164*, 46–56.
[85] Novikov, I. S.; Shapeev, A. V. Improving accuracy of interatomic potentials: more physics or more data? A case study of silica. *Mater. Today Commun.* **2019**, *18*, 74–80.
[86] Gubaev, K.; Podryabinkin, E. V.; Hart, G. L. W.; Shapeev, A. V. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials. *Comput. Mater. Sci.* **2019**, *156*, 148–156.
[87] Rosenbrock, C. W.; Gubaev, K.; Shapeev, A. V.; Pártay, L. B.; Bernstein, N.; Csányi, G.; Hart, G. L. W. Machine-learned interatomic potentials for alloys and alloy phase diagrams. *npj Comput. Mater.* **2021**, *7*, 24.
[88] Huang, S. D.; Shang, C.; Kang, P. L.; Liu, Z. P. Atomic structure of boron resolved using machine learning and global sampling. *Chem. Sci.* **2018**, *9*, 8644–8655.
[89] Shang, C.; Zhang, X. J.; Liu, Z. P. Stochastic surface walking method for crystal structure and phase transition pathway prediction. *Phys. Chem. Chem. Phys.* **2014**, *16*, 17845–17856.
[90] Guan, S. H.; Shang, C.; Huang, S. D.; Liu, Z. P. Two–Stage Solid–Phase Transition of Cubic Ice to Hexagonal Ice: Structural Origin and Kinetics. *J. Phys. Chem. C* **2018**, *122*, 29009–29016.
[91] Ma, S.; Huang, S. D.; Fang, Y. H.; Liu, Z. P. Microporous Titania Crystals with Penta-oxygen Coordination. *ACS Appl. Energy Mater.* **2018**, *1*, 22–26.
[92] Ma, S.; Huang, S. D.; Liu, Z. P. Dynamic coordination of cations and catalytic selectivity on zinc-chromium oxide alloys during syngas conversion. *Nat. Catal.* **2019**, *2*, 671–677.
[93] Ma, S.; Shang, C.; Liu, Z. P. Heterogeneous catalysis from structure to activity via SSW-NN method. *J. Chem. Phys.* **2019**, *151*, 050901
[94] Pandolfi, S.; Renero-Lecuna, C.; Le Godec, Y.; Baptiste, B.; Menguy, N.; Lazzeri, M.; Gervais, C.; Spektor, K.; Crichton, W. A.; Kurakevych, O. O. Nature of Hexagonal Silicon Forming via High-Pressure Synthesis: Nanostructured Hexagonal 4H Polytype. *Nano. Lett.* **2018**, *18*, 5989–5995.
[95] Hanfland, M.; Alouani, M.; Syassen, K.; Christensen, N. E. Optical properties of metallic silicon. *Phys. Rev. B* **1988**, *38*, 12864–12867.
[96] Chang, K. J.; Dacorogna, M. M.; Cohen, M. L.; Mignot, J. M.; Chouteau, G.; Martinez, G. Superconductivity in high-pressure metallic phases of Si. *Phys. Rev. Lett.* **1985**, *54*, 2375–2378.
[97] Mujica, A.; Rubio, A.; Muñoz, A.; Needs, R. J. High-pressure phases of group-IV, III-V, and II-VI compounds. *Rev. Mod. Phys.* **2003**, *75*, 863.
[98] Hanfland, M.; Schwarz, U.; Syassen, K.; Takemura, K. Crystal Structure of the High-Pressure Phase Silicon VI. *Phys. Rev. Lett.* **1999**, *82*, 1197.
[99] Jamieson, J. C. Crystal Structures at High Pressures of Metallic Modifications of Compounds of Indium, Gallium, and Aluminum. *Science* **1963**, *139*, 3557.
[100] McMahon, M. I.; Nelmes, R. J. New high-pressure phase of Si. *Phys. Rev. B* **1993**, *47*, 8337–8340.
[101] Olijnyk, H.; Sikka, S. K.; Holzapfel, W. B. Structural phase transitions in Si and Ge under pressures up to 50 GPa. *Phys. Lett.* **1984**, *103A*, 137.
[102] Duclos, S. J.; Vohra, Y. K.; Ruoff, A. L. hcp to fcc transition in silicon at 78 GPa and studies to 100 GPa. *Phys. Rev. Lett.* **1987**, *58*, 775–777.
[103] Etacheri, V.; Marom, R.; Elazari, R.; Salitra, G.; Aurbach, D. Challenges in the development of advanced Li-ion batteries: a review. *Energy Environ. Sci.* **2011**, *4*, 3243.
[104] Zaghib, K.; Song, X.; Guerfi, A.; Kostecki, R.; Kinoshita, K. Effect of Particle Morphology on Lithium Intercalation Rates in Natural Graphite. *J. Power Sources* **2003**, *124*, 505–512.
[105] Stokes, K.; Fynn, G.; Geaney, H.; Bree, G.; Ryan, K. M. Axial Si-Ge Heterostructure Nanowires as Lithium-Ion Battery Anodes. *Nano Lett.* **2018**, *18*, 5569–5575.
[106] Liu, D. Q.; Liu, Z. J.; Li, X. W.; Xie, W. H.; Wang, Q.; Liu, Q. M.; Fu, Y. J.; He, D. Y. Group IVA Element (Si, Ge, Sn)-Based Alloying/Dealloying Anodes as Negative Electrodes for Full-Cell Lithium-Ion Batteries. *Small* **2017**, *13*, 1702000.
[107] Park, M. S.; Wang, G. X.; Kang, Y. M.; Wexler, D.; Dou, S. X.; Liu, H. K. Preparation and Electrochemical Properties of SnO_2 Nanowires for Application in Lithium-ion Batteries. *Angew. Chem. Int. Ed.* **2007**, *46*, 750–753.
[108] Lai, C.; Li, G. R.; Dou, Y. Y.; Gao, X. P. Mesoporous Polyaniline or Polypyrrole/Anatase TiO_2 Nanocomposite as Anode Materials for Lithium-Ion Batteries. *Electrochim. Acta* **2010**, *55*, 4567–4572.
[109] Fang, Y. H.; Ma, S. C.; Liu, Z. P. 3-D Tunnel TiO_2 Crystal Phase as a Fast Charging Lithium Battery Anode from Stochastic Surface Walking-Based Material Screening. *J. Phys. Chem. C* **2019**, *123*, 19347–19353.
[110] Liu, H. S.; Bi, Z. H.; Sun, X. G.; Unocic, R. R.; Paranthaman, M. P.; Dai, S.; Brown, G. M. Mesoporous TiO_2-B Microspheres with Superior Rate Performance for Lithium Ion Batteries. *Adv. Mater.* **2011**, *23*, 3450–3454.
[111] Kim, S. W.; Han, T. H.; Kim, J.; Gwon, H.; Moon, H. S.; Kang, S. W.; Kim, S. O.; Kang, K. Fabrication and Electrochemical Characterization of TiO_2 Three-Dimensional Nanonetwork Based on Peptide Assembly. *ACS Nano* **2009**, *3*, 1085–1090.

Manuscript received: April 28, 2021
Manuscript revised: June 7, 2021
Manuscript accepted: June 12, 2021
Accepted manuscript online: June 16, 2021
Version of record online: September 7, 2021

