

Construction of High Accuracy Machine Learning Interatomic Potential for Surface/Interface of Nanomaterials—A Review

Kaiwei Wan, Jianxin He, and Xinghua Shi*

The inherent discontinuity and unique dimensional attributes of nanomaterial surfaces and interfaces bestow them with various exceptional properties. These properties, however, also introduce difficulties for both experimental and computational studies. The advent of machine learning interatomic potential (MLIP) addresses some of the limitations associated with empirical force fields, presenting a valuable avenue for accurate simulations of these surfaces/interfaces of nanomaterials. Central to this approach is the idea of capturing the relationship between system configuration and potential energy, leveraging the proficiency of machine learning (ML) to precisely approximate high-dimensional functions. This review offers an in-depth examination of MLIP principles and their execution and elaborates on their applications in the realm of nanomaterial surface and interface systems. The prevailing challenges faced by this potent methodology are also discussed.

1. Introduction

Advancements in synthesis and characterization technologies have led to the development of a wide array of nanomaterials, including nanoparticles, nanotubes, and nanosheets. One standout characteristic of these nanomaterials is their vast specific surface area, endowing them with distinct nanoscale surfaces and interfaces compared to bulk materials. As a result, nanomaterials have found applications in diverse areas such as heterogeneous catalysis,^[1,2] gas sensors,^[3,4] microchip integrated circuits,^[5,6] and biological ion channels.^[7,8] Thus, understanding the surface and interfacial properties of nanomaterials becomes crucial for grasping their microstructure and emerging properties.

Generally, surfaces refer to the boundary separating a condensed phase from the gas phase or vacuum, while interfaces

describe the division between two condensed phases (Figure 1). Given the inherent discontinuity of these boundaries, their study is intricate. Wolfgang Pauli aptly encapsulated this challenge, stating, “God made bulk; the surface was invented by the devil”. He highlighted the complexities of surfaces, noting that their direct exposure makes them especially sensitive to external influences. Fortunately, modern science boasts advanced characterization techniques surpassing those available in Pauli’s era. For instance, atomic force microscopy (AFM) offers insights into nanomaterials’ surface topographies.^[9] Q-plus-based AFM further provides detailed visualizations of hydrogen bonds and complex structures at the water–nanomaterial interface with an atomic resolution.^[10–12] The recent spherical aberration-corrected transmission electron microscope stands out for its sub-atomic resolution capabilities.^[13,14] Surface-enhanced spectroscopy techniques have been pivotal for material composition and structure analysis on nanomaterial surfaces, tracking reaction intermediates at interfaces, and studying water layer structures.^[15–17] Additionally, sum-frequency spectroscopy can selectively probe surface and interfacial molecular vibrations, providing insights into their structures and dynamics without bulk interference.^[18,19]

While experimental techniques have advanced, challenges persist in investigating nanomaterial surfaces and interfaces due to limitations like specific sample requirements, potential sample damage from high-energy illumination, and limited spatial or temporal resolution. In contrast, molecular dynamics (MD) simulations provide a comprehensive *in situ* method to study interfacial phenomena in nanomaterials without these constraints. MD is a computational approach that models molecular motion, using Newton’s equations of motion to calculate forces on atoms. It operates under the Born-Oppenheimer approximation, presuming electrons achieve their ground state swiftly after atomic nuclei movement. The accuracy of the force field, used in calculating atomic forces, directly impacts the alignment between simulated trajectories and real-world physical behaviors. Typically, all-atom MD simulations use empirical potentials as their force fields. While empirical potentials have effectively modeled phenomena including nanomaterial structural changes,^[20–23] nanomaterial–water interface dynamics,^[24–27] and nanomaterial self-assembly,^[28–30] concerns about their precision and

K. Wan, J. He, X. Shi
 Laboratory of Theoretical and Computational Nanoscience
 National Center for Nanoscience and Technology
 Chinese Academy of Sciences
 Beijing 100190, China
 E-mail: qiaoxiaoai@gmail.com; shixh@nanoctr.cn

K. Wan, J. He, X. Shi
 University of Chinese Academy of Sciences
 No. 19A Yuquan Road, Beijing 100049, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adma.202305758>

DOI: 10.1002/adma.202305758

transferability persist. For instance, their fixed parameters can occasionally falter in accuracy, especially at interfaces where the behavior starkly contrasts with bulk phases.^[31–33] The use of fixed charges in these potentials makes them ill-suited to capturing molecular polarization, notably at solid–liquid junctures.^[34,35] They also remain inept at simulating chemical reactions. While the scientific community has witnessed the advent of advanced force fields, their rigid character curtails their adaptability across a diverse array of nanomaterial interfaces.

Ab initio MD methods, anchored by density functional theory (DFT), are exceptional in determining atomic forces. However, their intensive computational demands limit their applicability for extensive nanomaterial interfaces and extended simulations. This constraint complicates the process of deriving thermodynamic properties using statistical physics. Such challenges highlight the pressing need for a simulation approach that seamlessly combines accuracy and efficiency in the modeling of nanomaterial surfaces and interfaces.

A silver lining in this space is the rapid evolution of ML algorithms, particularly the likes of large-scale language models such as GPT-4. Their applications are vast and varied, extending to sectors like material development, 3D printing, and cutting-edge wearable technology.^[36–41] Some even posit that this ML renaissance heralds the onset of the fifth industrial revolution. A case in point from 2020: a team adeptly employed DP MLIP, attaining DFT-grade accuracy in simulating a copper crystal comprising a billion atoms.^[42] Subsequent to this, another research group applied the SNAP MLIP technique to achieve similar precision for ten billion carbon atoms under extreme pressure conditions.^[43] Such achievements accentuate the formidable potential of MLIPs, acting as a bridge between simulated and experimental outcomes. This has fueled a growing enthusiasm for employing MLIPs in simulating surfaces and interfaces of nanomaterials, marking it as an emergent and promising research field.^[44]

This review delves into three primary facets: the underlying principles and execution of MLIPs, their utilization in the surface and interfacial domains of nanomaterials, and the prevailing challenges of this robust technique. We aspire to equip both experimental and computational researchers

in the arena of nanomaterials with deeper insights into this method.

2. Principles and Implementation of MLIPs

The MLIPs essentially represent potential energy surfaces (PES) derived through ML techniques. PESs are multidimensional surfaces that describe the variation of potential energy with respect to atomic coordinates in configuration space. Based on the potential energy surface, the energy and force of a molecular system can be obtained without requiring the solution of the complex Schrödinger equation. Therefore, the reliability of MD simulations depends directly on the accuracy of the potential energy surface. Traditional methodologies for approximating potential energy surfaces typically decompose atomic interactions into various components, each reflective of specific physical properties. Analytical expressions are then tailored for each of these terms. However, these expressions are often grounded in physical intuition rather than rigorous mathematical validation, introducing the potential for error. Although methods like mathematical interpolation can provide more precise PESs, they often demand intricate parameterization and impose substantial computational burdens. As universal function approximators, ML methods—specifically neural networks—are theoretically capable of modeling high-dimensional functions with an arbitrary degree of precision.^[45–47] This offers a compelling solution to the challenges inherent to multidimensional PESs. By harnessing ML technologies, one can generate potential energy surfaces that are not only highly accurate but also computationally economical.

2.1. Basic Concept

In the development of MLIPs, two principal algorithms are predominantly employed: artificial neural networks and Gaussian process regression (GPR). The subsequent section elucidates the key features of these two methodologies.

Artificial neural networks function as computational models inspired by the architecture of biological neural systems. Generally, these models are composed of input, hidden, and output layers; each hidden layer encompasses an array of artificial neurons. A standard mathematical representation of a neuron is given by:

$$H_{\omega,b}(x) = f(\omega^T x + b) = f\left(\sum_{i=1}^n \omega_i x_i + b\right) \quad (1)$$

where x denotes the input from the preceding layer, ω symbolizes the weights, b stands for the bias term, and f signifies the activation function. Common activation functions typically map their inputs to a pre-defined range. Figure 2a illustrates a classical feedforward, fully connected artificial neural network, wherein each input undergoes successive transformations through various neurons in the hidden layers before arriving at the output. Training these networks involves the iterative optimization of weights and biases via the backpropagation algorithm to minimize a specified loss function. Owing to its robust fitting capabilities, this approach is among the most widely utilized in MLIPs.^[48–55]

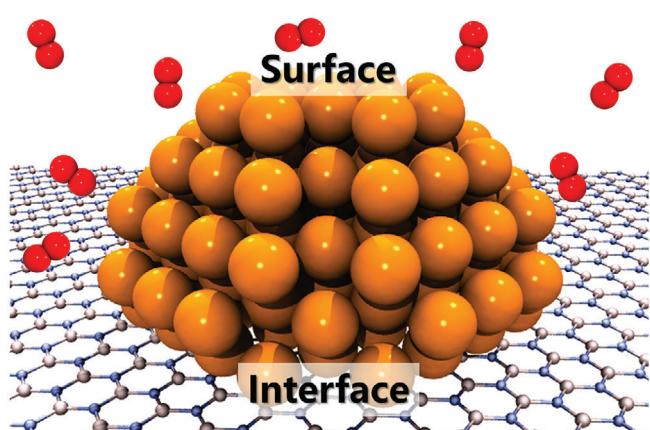


Figure 1. Schematic of the surface/interface of nanomaterials.

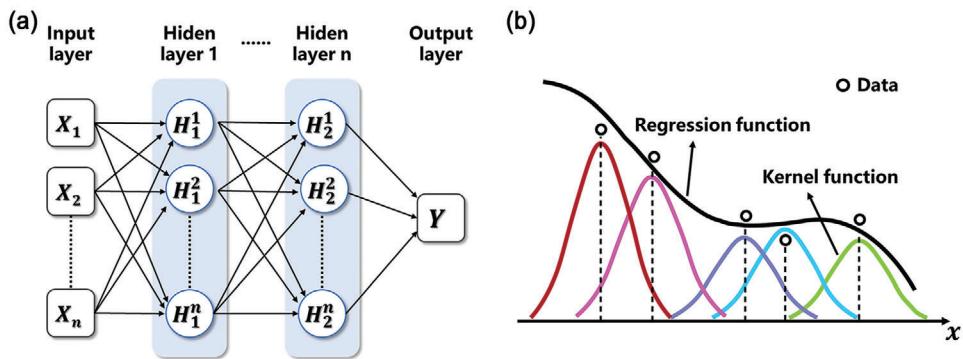


Figure 2. Artificial neural network and kernel regression algorithms. a) Schematic of an artificial neural network model. b) Schematic of a Gaussian process regression model.

GPR, on the other hand, operates within the Bayesian framework as a potent non-parametric regression technique. It leverages a kernel function to craft a nonlinear regression model, as illustrated in Figure 2b. The corresponding formula for this model is:

$$\tilde{y}(\mathbf{x}) = \sum_{m=1}^M c_m k(\mathbf{x}, \mathbf{x}_m) \quad (2)$$

where $\tilde{y}(\mathbf{x})$ is the predicted value at point \mathbf{x} , \mathbf{x}_m represents the set of reference points, $k(\mathbf{x}, \mathbf{x}_m)$ is the kernel function, and $\{c_m\}$ is the set of coefficients with dimension M , determined by minimizing the loss function:

$$\mathcal{L} = \sum_{n=1}^N \frac{[y_n - \tilde{y}(\mathbf{x}_n)]^2}{\sigma_n^2} + R \quad (3)$$

In this equation, R is a regularization term, $\{\mathbf{x}_n\}$ is the training dataset with dimension N , and σ_n is a tunable parameter. For large dataset dimensions and $M = N$, computing the kernel function for every reference point can be computationally prohibitive. Hence, a sparse Gaussian process regression technique, where $M \ll N$, is commonly employed to mitigate computational complexity.

In GPR-based MLIPs, the kernel function acts as a similarity metric between different configurations. Using this similarity, the energy of previously unencountered configurations can be predicted. Relative to neural network-based MLIPs, those leveraging GPR often achieve comparable accuracy with smaller training sets, making them another prevalent choice in this domain.^[56–63]

2.2. The Birth of MLIPs

The initial exploration of using ML methods for constructing potential energy surfaces dates back to 1992,^[65] although these early endeavors markedly differ from contemporary MLIPs. Bobby et al. pioneered the use of an artificial neural network model to predict anharmonic vibrational parameters for a polymer PES, leveraging a dataset of 51 vibrational spectra. While rudimentary by today's standards, this seminal work indisputably laid the groundwork for subsequent advancements in the field of MLIPs.

The advent of the first bona fide MLIP was in 1995 when Blank et al. utilized an artificial neural network to approximate the reaction potential energy surface governing the formation of molecular hydrogen H₂ on a Si(100) surface.^[66] Their model followed the architecture of a standard feedforward neural network, employing the 3D coordinates of four atoms as inputs, which were then processed through hidden layers to yield the final potential energy as output. The model achieved a training accuracy quantified by a mean absolute deviation (MAD) of 1.7 kcal mol⁻¹. Following this landmark study, similar approaches were extended to small molecules and elementary reactions,^[67–71] corroborating the substantial promise of ML techniques for the accurate parameterization of PES. Although there were successful attempts to construct ML-based PESs during this period, their application in MD simulations remained relatively scarce, chiefly hindered by the nascent state of theoretical methods and constrained computational resources of the era.

2.3. Behler-Parrinello Framework-Based MLIPs

The early MLIPs encountered two serious issues,^[72] which greatly limited their widespread application in MD simulations at that time. First, the fixed-value weights in artificial neural networks presented a limitation: the sequence of atoms in the input layer had to remain constant. As a result, interchanging two identical atoms led to disparate potential energy outputs, an outcome that contradicted physical images. Second, the matrix dimensions of the input layer were rigidly defined, necessitating that the number of atoms in both the training dataset and the eventual MD simulation system remain consistent. This posed a significant obstacle, as quantum chemical methods could only accommodate systems far smaller in size than those typically modeled in MD, thereby complicating the acquisition of training data. Fortunately, these two issues were resolved in the well-known work of Behler and Parrinello,^[48] wherein they proposed a new MLIP model as shown in Figure 3a.

In order to address the issues, a locality approximation was introduced, where the energy of each atom can be determined by its local atomic environment within a cutoff radius. Moreover, a permutation-invariant design for the input layer was introduced,

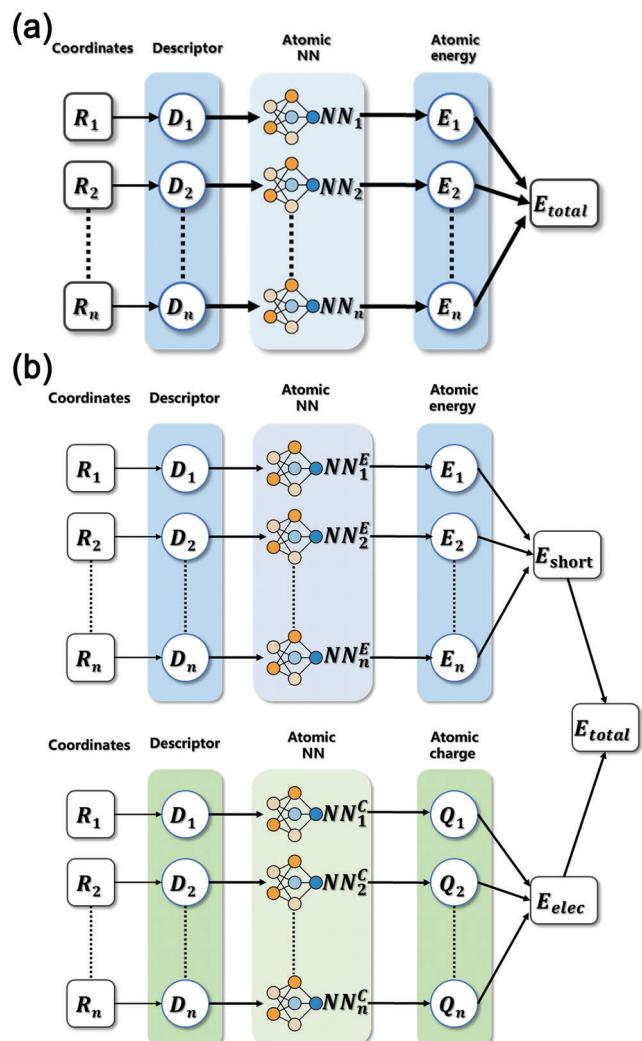


Figure 3. Two types of MLIP models. a) MLIP model proposed by Behler and Parrinello.^[48] b) MLIP with electronic interaction correction.^[64]

which resembles an empirical potential function given by the expression:

$$E = \sum_i E_i \quad (4)$$

The total energy E of the system can be expressed as the sum of the energies of each individual atom E_i , thereby rendering the energy of the system independent of the ordering of the atoms. Consequently, there is no requirement for the training dataset and the eventual MD simulation system to have a consistent number of atoms; one simply needs to adequately sample the local atomic environments within a smaller training set. This flexibility also eliminates the necessity to pre-define the topological structure of the system, enabling the simulation of chemical reactions using this methodology.

Using this novel MLIP model, Behler and Parrinello performed MD simulations of the radial distribution function (RDF) of melted silicon at 3000 K.^[48] The results were highly consistent with those obtained from DFT and outperformed other

empirical potential functions. Subsequent applications of this model across various systems have validated its accuracy and broad applicability.^[73–75] Furthermore, the conceptual framework and principles underlying this groundbreaking work have been widely adopted, serving as the foundation for the development of many subsequent MLIP models.^[51,52,76,77]

2.3.1. Descriptors

The model of Behler-Parrinello has successfully addressed most issues of the early MLIPs. However, the construction of descriptors, which involves mapping the local atomic environment to atomic energies, remains an active area of research for scientists. A well-defined descriptor should satisfy the following properties: 1) it should produce identical results for any given structure, even after undergoing translation, rotation, and permutation operations; 2) both the descriptor and its derivatives should ensure smoothness. Additionally, the computational cost of the descriptor is also an important consideration that needs to be considered.

Atomic Central Symmetric Function: The atomic central symmetric function (ACSF) is a method conceived by Behler and Parrinello within their MLIP model. This approach employs symmetric functions to encapsulate the influence of surrounding atoms within a specified cutoff radius around a given atom. Generally, ACSF incorporates two categories of functions: radial and angular functions. The formulas can be articulated as follows:

$$G_i^{\text{rad}} = \sum_{j \neq i}^{N_{\text{atom}} \in R_c} e^{-\eta(R_j - R_s)^2} f_c(R_{ij}) \quad (5)$$

$$G_i^{\text{ang}} = 2^{1-\zeta} \sum_{i,j,k} (1 + \lambda \cos \theta_{ijk})^\zeta e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}) \quad (6)$$

$$f_c(R_{ij}) = \begin{cases} 0.5 \left[\cos \left(\pi \frac{R_{ij}}{R_c} \right) + 1 \right], & R_{ij} \leq R_c \\ 0, & R_{ij} > R_c \end{cases} \quad (7)$$

where i is the atomic index, while η , R_s , λ , and ζ serve as tunable parameters. By fine-tuning these parameters, it becomes feasible to derive symmetric functions that are attuned to different atomic centers and varying spatial extents, thereby accommodating a range of elemental combinations. The cutoff function f_c is designed to gradually reduce the function value and its higher-order derivatives to zero, mitigating discontinuities in energy and force as atoms move in and out of the cutoff radius.

In addition to the original ACSF, subsequent research has led to refined variants. To address the complication of proliferating symmetric functions as the variety of atom types increases, Gastegger et al. introduced the concept of the weighted atomic central symmetric function (wACSF),^[78] which incorporates atom-type-specific information directly into the ACSF. Further, Liu et al. have enhanced the granularity of the symmetric functions by adding descriptors of the chemical environment, including dihedral angles and contributions from non-bonded atoms that are not directly linked to the central atom.^[49,50]

Deep Potential: In 2019, Zhang et al. introduced the DP descriptor, distinguished by its minimal hyperparameter tuning demands,^[51,52] which has since become one of the most popular MLIP models. The DP descriptor consists of two core components: the environment matrix and the embedding matrix.

For a given atom i , the environment matrix is defined as:

$$(\tilde{\mathcal{R}}_i)_j = s(r_{ij}) \times \left(1, \frac{x_{ij}}{r_{ij}}, \frac{y_{ij}}{r_{ij}}, \frac{z_{ij}}{r_{ij}} \right) \quad (8)$$

where $s(r_{ij})$ is the smoothing function, r_{ij} represents the distance between atoms i and j , and x_{ij}, y_{ij}, z_{ij} denote the distances between these atoms in the x, y , and z direction, respectively.

Two distinct approaches are used for constructing the embedding matrix. In the case of the two-body embedding matrix, each row is articulated as follows:

$$\left(\mathcal{G}_i^{(2)} \right)_j = \left(G_1^{(2)}(s(r_{ij}), Z_j), \dots, G_M^{(2)}(s(r_{ij}), Z_j) \right) \quad (9)$$

Here, Z_j is the atomic type, and the vector $(G_1^{(2)}, \dots, G_M^{(2)})$ is represented by a neural network with the input of $(s(r_{ij}), Z_j)$, also known as the embedding network. For the three-body embedding matrix, each element is defined by:

$$\left(\mathcal{G}_i^{(3)} \right)_{j,k} = \left(G_1^{(3)}((\theta_i)_{jk}, Z_j, Z_k), \dots, G_M^{(3)}((\theta_i)_{jk}, Z_j, Z_k) \right) \quad (10)$$

where $(\theta_i)_{jk}$ contains the angle information of the atoms j, i, k , and Z_j and Z_k denote the atomic types.

Finally, three types of DP descriptors are formulated based on the environment and embedding matrices, given by:

$$\mathcal{D}_i^{(2,r)} = \frac{1}{N_i} \sum_j \left(\mathcal{G}_i^{(2)} \right)_j \quad (11)$$

$$\mathcal{D}_i^{(2,a)} = \frac{1}{N^2(i)} \left(\mathcal{G}_i^{(2,M)} \right)^T \tilde{\mathcal{R}}_i (\tilde{\mathcal{R}}_i)^T \mathcal{G}_i^{(2)} \quad (12)$$

$$\mathcal{D}_i^{(3)} = \frac{1}{N^2(i)} \left(\tilde{\mathcal{R}}_i (\tilde{\mathcal{R}}_i)^T \right) : \mathcal{G}_i^{(3)} \quad (13)$$

where $\mathcal{G}_i^{(2,M)}$ represents the first M columns of $\mathcal{G}_i^{(2)}$. It is noteworthy that $\mathcal{D}_i^{(3)}$ contains more angular information and, therefore, exhibits the highest accuracy among the three descriptors. The order of precision for these three descriptors is as follows: $\mathcal{D}_i^{(3)} > \mathcal{D}_i^{(2,a)} > \mathcal{D}_i^{(2,r)}$, with computational costs inversely related to their precision.

2.3.2. Long-Range Interaction Correction

In the framework of Behler-Parrinello, the introduction of the locality approximation has successfully addressed the biggest drawback of the primary MLIPs. Under this framework, thanks to the powerful fitting ability of ML algorithms, the description of short-range interactions is highly successful. However, for systems with significant long-range interactions, this assumption is clearly not reasonable. Typically, long-range interactions mainly

include electrostatic and van der Waals interactions, and the decay rate of the former is much slower than the latter. Therefore, the correction of long-range interactions mainly focuses on correcting electrostatic interactions.

Within empirical potentials, the incorporation of electrostatic interactions predominantly relies on predetermined atomic charges. Building on this notion, Deng et al. introduced eSNAP,^[79,80] which augments the SNAP MLIP with fixed formal charges. The total energy expression is defined as:

$$E_{\text{tot}} = E_{\text{SR}} + E_{\text{ele}} \quad (14)$$

where E_{SR} is directly given by the SNAP MLIP, and E_{ele} is calculated using formal charges through the Ewald summation method.^[81] While this approach exhibits commendable fitting accuracy in systems like $\alpha\text{-Li}_3\text{N}$, it fails to account for the dynamic variations in atomic charges, particularly in complex systems undergoing phase transitions or chemical reactions.

A more general solution was proposed by Artrith et al., which mainly involves training an additional neural network to predict atomic charges.^[64] The process begins by utilizing electronic structure methods to calculate reference atomic charges, which then serve as training data for the neural network. The electrostatic energy is subsequently subtracted from the total reference energy to yield short-range energy, which is employed to train another neural network. The two networks are eventually integrated to produce an MLIP that accommodates corrections for long-range interactions. The network structure is shown in Figure 3b. It is crucial to note that atomic charges are not directly observable quantities. The calculated values can vary depending on the partitioning methods employed, thus potentially influencing simulation outcomes. However, for systems with larger interatomic distances, different atomic charge calculation methods generally produce similar charge values. In systems with smaller interatomic distances, any significant variations in electrostatic energy are typically smoothed over by the short-range neural network, rendering the impact of different atomic charge partitioning methods on simulation outcomes relatively insignificant.

Compared to the arbitrariness of atomic charge partitioning, Zhang et al. proposed a more rigorously defined method called DPLR.^[82] This method also divides the total energy into short-range and electrostatic energy, and the difference from other methods lies in the calculation of electrostatic energy using spherical Gaussian charge distributions on ions and electron sites, with the latter defined by the molecular Wannier centroid, and its position is fitted by a separate neural network. However, since this method assumes that the Wannier centroid is related to a specific atom, it cannot simulate chemical reactions.

2.3.3. Nonlocal Charge Transfer Correction

Although incorporating a separate neural network to describe electrostatic interactions can significantly improve the performance of MLIPs in describing long-range interactions, the neural network itself still relies on the local approximation, i.e., describing the corresponding electrostatic interactions based on local atomic environments. As a result, this approach is unable to capture nonlocal charge transfer phenomena, which are commonly

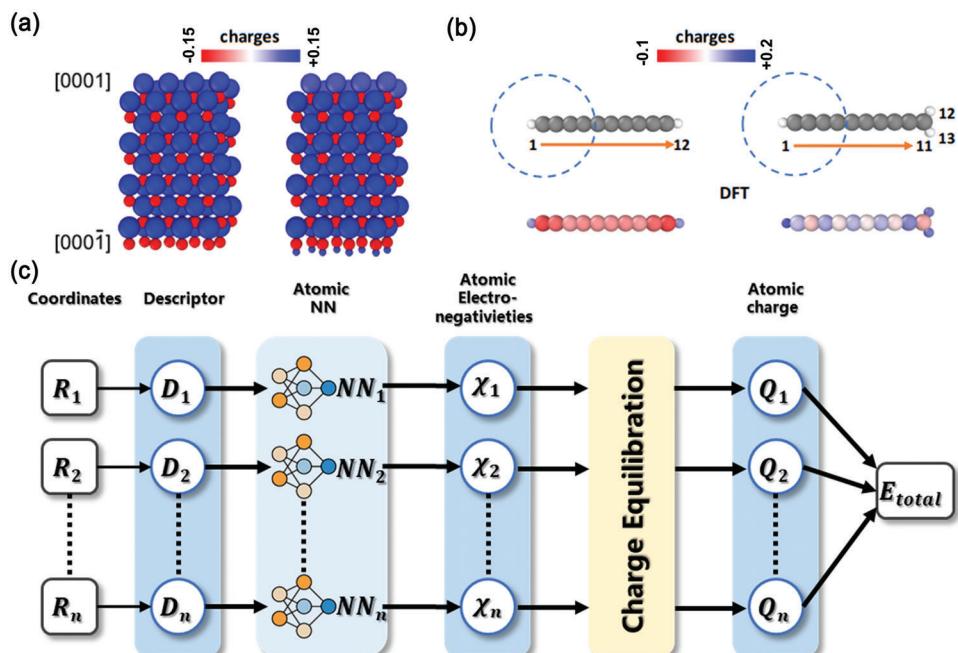


Figure 4. Nonlocal charge transfer phenomena and MLIP with nonlocal charge transfer correction. a) Nonlocal charge transfer in ZnO slab. Reproduced with permission.^[72] Copyright 2021, American Chemical Society. b) Nonlocal charge transfer in C₁₀H₂ molecule. Reproduced under the terms of the CC BY 4.0 license.^[83] Copyright 2021, the Authors. Published by Springer. c) Charge equilibration neural network technique.^[84]

observed in interfacial systems. As demonstrated in Figure 4a, the bottom-layer H atoms in the ZnO slab can affect the charge distribution of the top-layer Zn atoms. When the cut-off radius is not sufficiently large, the top-layer Zn's local environment does not include the bottom-layer H atoms. In this case, relying solely on the local environment description fails to account for the influence of non-local electron transfer. And this phenomenon has an even more significant impact on the charge distribution of molecular systems (Figure 4b).

In fact, this issue also exists in empirical force fields. A well-known solution was proposed by Rappe and Goddard, termed the charge equilibration method,^[85] which can be simply understood as redistributing the system's charges by solving equations related to atomic electronegativity and other physical quantities. This method was later successfully applied in advanced force fields such as ReaxFF.^[86] Ghasemi et al. proposed an MLIP capable of describing non-local electron transfer, called the charge equilibration neural network technique (CNET).^[84] The main idea is to achieve charge equilibration by solving the energy expression concerning atomic electronegativity, which can be written as:

$$E_{\text{tot}}(\{q_i\}, \{\mathbf{r}_i\}) = \sum_{i=1}^N \left[E_i^0 + \chi_i q_i + \frac{1}{2} \left(J_i + \frac{2}{\sqrt{(\alpha_i^2 + \alpha_j^2) \pi}} \right) q_i^2 \right] \\ + \sum_{i>j}^N q_i q_j \frac{\text{erf}\left(\frac{r_{ij}}{\sqrt{\alpha_i^2 + \alpha_j^2}}\right)}{r_{ij}} \quad (15)$$

where E_i^0 represents the energy of a free neutral atom, χ_i is the atomic electronegativity, J_i is the element-specific hardness obtained through fitting, q_i is the atomic charge, r_{ij} is the distance between atoms i and j , and α_i is an adjustable parameter. The specific neural network structure is shown in Figure 4c, predicting electronegativity based on local atomic environments. The equilibrated charge distribution and system energy are then obtained by minimizing Equation (15). The CNET method has achieved satisfactory results in ionic crystal systems, but its accuracy is still an order of magnitude lower than that of general MLIPs.

Building on this, Ko et al. introduced a more precise approach dubbed the fourth-generation high-dimensional neural network potential (4G-HDNNP).^[83] The calculation of the total energy in 4G-HDNNP requires two steps: First, the neural network is used to obtain atomic electronegativity, and a method similar to CNET is employed to obtain the equilibrated charge distribution and electrostatic energy. The difference lies in training with the goal of reproducing reference charges, and the element hardness J_i in Equation (15) can be optimized during training instead of being a fixed parameter. In the second step, the obtained atomic charges and ACSF descriptors are combined as inputs, and the short-range energy is obtained through the atomic neural network mapping. The total energy of the system is then acquired by summing the two parts of energy. This method has demonstrated promising fitting accuracy in various systems.^[83,87] However, whether electronegativity itself, like atomic charges, is affected by the nonlocal environment, as well as issues such as the extensive computational cost of the method, make the development of MLIPs capable of describing nonlocal charge transfer still an open problem.

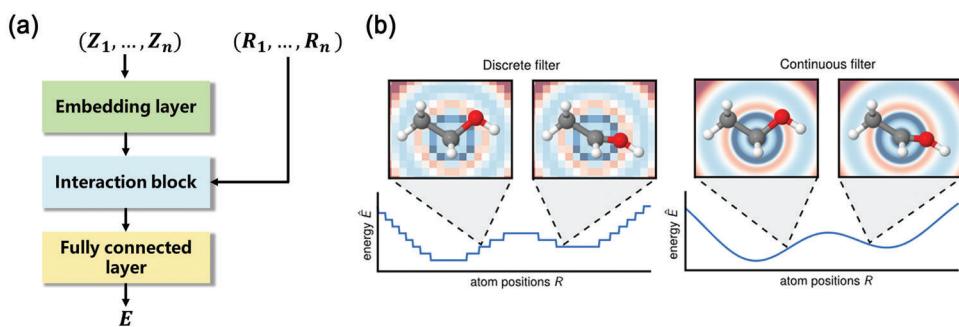


Figure 5. GNN-based MLIPs models. a) The key structure of GNN-based MLIPs models. b) The comparison of the energy output after two different filtering convolutional layers.^[53] (left denotes discrete filter, and right denotes continuous filter). Reproduced with permission.^[53] Copyright 2017, Princeton University.

2.4. Graph Neural Network-Based MLIPs

The Behler-Parrinello framework has shown great potential in MLIP modeling. In addition, other MLIP models based on graph neural networks (GNN) exist that obviate the need for manual construction of descriptors. The GNN approach is particularly well-suited for constructing MLIP models as molecular and material systems can be naturally represented in graph form. However, it should be noted that designing appropriate network structures is an essential aspect of implementing such types of MLIPs. As illustrated in Figure 5a, a typical GNN-based MLIP should comprise three core components: an embedding layer, an interaction block, and a fully connected layer.

- 1) Embedding layer: It represents molecules or material systems as graphs, where atoms are nodes and bonds are edges. Node features can include atomic properties such as nuclear charge number or atomic number, while edge features can represent bond type and distance.
- 2) Interaction block: It allows nodes to send messages to their neighbors based on their current features and the features of connected edges. These messages are then aggregated by receiving nodes and processed by a neural network to update node features. This process is usually repeated for several cycles.
- 3) Fully connected layer: After passing through the interaction blocks, a pooling operation is applied to aggregate information from all nodes into a single representation of the entire system. The pooled graph passes through one or more fully connected layers to produce the final output, usually atomic potential energy or force.

An early implementation of this method is the deep tensor neural network (DTNN),^[88] which uses the nuclear charge number vector V and the atomic distance matrix D as input to construct a graph representation. To achieve rotational and translational invariance, a Gaussian feature transformation is applied to the atomic distance matrix. The formula for this transformation is:

$$\hat{d}_{ij} = \left[\exp\left(-\frac{(D_{ij} - (\mu_{\min} + k\Delta\mu))^2}{2\sigma^2} \right) \right]_{0 \leq k \leq \frac{\mu_{\max}}{\Delta\mu}} \quad (16)$$

where μ_{\min} , μ_{\max} , $\Delta\mu$ and σ are adjustable parameters.

Although DTNN demonstrates commendable precision in forecasting the static properties of molecules, its dynamic simulation capabilities are constrained. This limitation arises because the traditional convolutional layer is designed for discrete signals and fails to output continuously varying potential energy during atomic motion (Figure 5b). To address this limitation, the SchNet MLIP was developed,^[53] which designs a continuous filtering convolutional layer capable of handling unevenly spaced data, especially atoms at arbitrary positions. For a feature representation $X^l = (x_1^l, \dots, x_n^l)$ of n elements located at $R = (r_1, \dots, r_n)$, the expression is as follows:

$$x_i^{l+1} = (X^l * W^l)_i = \sum_j x_j^l \circ W^l(r_i - r_j) \quad (17)$$

where \circ represents the Hadamard product multiplication, and W^l represents the filter generation function modeled by the neural network.

In GNN-based MLIPs, the number of message-passing iterations is inherently limited, resulting in a description of the local atomic environment. Consequently, it also shares a similar challenge as the Behler-Parrinello framework-based MLIPs, where long-range interactions and non-local electron transfer cannot be adequately captured. For the former, the PhysNet MLIP provides a good correction to its long-range interaction behavior by simultaneously outputting atom energy and charge.^[54] The expression for the total potential energy is:

$$E = \sum_{i=1}^N E_i + k_e \sum_{i=1}^N \sum_{j>i}^N \tilde{q}_i \tilde{q}_j \chi(r_{ij}) + E_{D3} \quad (18)$$

Moreover, the SpookyNet MLIP has augmented the model's performance concerning long-range interactions and non-local electron transfer by incorporating the electron count and spin states as additional input features, modeled through self-attention in a transformer architecture.^[55]

Recent advancements in equivariant GNN-based MLIPs have garnered considerable attention. Utilizing this novel approach substantially enhances fitting accuracy while substantially reducing the data requirements for training. For instance, the Neural Equivariant Interatomic Potentials (NequIP) MLIP has demonstrated similar accuracy to the Behler-Parrinello

framework-based MLIP trained on the full dataset, using less than 0.1% of the data for training.^[89]

2.5. Gaussian Process Regression-Based MLIPs

In addition to MLIPs that rely on the Behler-Parrinello framework or GNN-based methodologies, both of which employ deep neural network algorithms, another prevalent class of MLIPs is grounded in Gaussian process regression (GPR). Similar to Behler-Parrinello framework-based MLIPs, GPR-based MLIPs also require the design of descriptors to distinguish the local atomic environment of the system.

2.5.1. Gaussian Approximation Potential

One of the pioneering methods for actualizing this approach is the Gaussian Approximation Potential (GAP), which combines Gaussian Process Regression (GPR) with bispectrum descriptors.^[59] The core concept behind these descriptors is to utilize spherical harmonics and radial basis functions for expanding the local atomic density. Subsequently, the associated coefficients are combined to achieve translational, rotational, and permutational invariance with respect to atomic coordinates. Starting from the local atomic density in 3D space:

$$\rho_i(R) = \sum_{j \in \{i\} \cup \mathcal{N}_i} \delta(R - R_{ij}) f_c(R_{ij}) \quad (19)$$

$$f_c(R_{ij}) = \begin{cases} \frac{1}{2} + \frac{1}{2} \cos\left(\frac{\pi |R_{ij}|}{R_c}\right), & R_{ij} \leq R_c \\ 0, & R_{ij} > R_c \end{cases} \quad (20)$$

Equation (20) is projected onto the surface of a 4D unit sphere using the following coordinate transformation:

$$(\phi, \theta, \theta_0) = \left(\tan^{-1}\left(\frac{y}{x}\right), \cos^{-1}\left(\frac{z}{R}\right), \frac{R}{R_0} \right) \quad (21)$$

where $R_0 > \frac{R_c}{\pi}$. Next, the density function is expanded using 4D spherical harmonics:

$$\rho(\phi, \theta, \theta_0) = \sum_{j=0}^{\infty} \sum_{-j \leq m, m' \leq j} c_{m', m}^j U_{m', m}^j(\phi, \theta, \theta_0) \quad (22)$$

Using the expansion coefficients $c_{m', m}^j$ and $U_{m', m}^j$, the bispectrum descriptor is constructed according to the following expression:

$$B_{j_1 j_2 j} = \sum_{m'_1, m_1 = -j_1}^{j_1} \sum_{m'_2, m_2 = -j_2}^{j_2} \sum_{m', m = -j}^j C_{j_1 m_1 j_2 m_2}^{j m'} C_{j_1 m'_1 j_2 m'_2}^{j m'} c_{m'_1 m_1}^{j_1} c_{m'_2 m_2}^{j_2} \quad (23)$$

where $C_{j_1 m_1 j_2 m_2}^{j m'}$ is the Clebsch-Gordan coefficient and the truncated version with $j, j_1, j_2 \leq J_{\max}$ is used. Ultimately, the bispectrum descriptor is incorporated into a sparse Gaussian process regression framework in order to implement the GAP.

2.5.2. Smooth Overlap of Atomic Positions

To improve the limited accuracy of GAP in certain cases, Bartok et al. proposed an improved version, namely the smooth overlap of atomic positions (SOAP).^[60,90] Unlike GAP, which directly represents local atomic densities, SOAP focuses on quantifying the similarity between two configurations, making it especially well-suited for kernel regression techniques. Initially, the local atomic density is formulated via a Gaussian distribution:

$$\rho_{\mathcal{X}}(\mathbf{R}) = \sum_{i \in \mathcal{X}} \exp\left(-\frac{(\mathbf{R}_i - \mathbf{R})^2}{2\sigma^2}\right) \quad (24)$$

The overlap between two local densities, which serves as a measure of similarity between two configurations:

$$S(\rho_{\mathcal{X}}, \rho_{\mathcal{X}'}) = \int \rho_{\mathcal{X}}(\mathbf{r}) \rho_{\mathcal{X}'}(\mathbf{r}) d\mathbf{r} \quad (25)$$

The SOAP kernel function is defined as the overlap between the local atomic neighbor densities of two configurations, integrated over all 3D rotations $\hat{\mathbf{R}}$,

$$\tilde{k}(\mathcal{X}, \mathcal{X}') = \int \left| S(\rho_{\mathcal{X}}, \rho_{\mathcal{X}'}) \right|^n d\hat{\mathbf{R}} = \int d\hat{\mathbf{R}} \left| \int \rho_{\mathcal{X}}(\mathbf{r}) \rho_{\mathcal{X}'}(\hat{\mathbf{R}}\mathbf{r}) d\mathbf{r} \right|^n \quad (26)$$

when $n = 1$, the integration over symmetry operations and over coordinates can be interchanged, and the kernel function loses the relative position information of the atoms. Therefore, only the case of $n = 2$ is considered. After normalizing the kernel function, we have:

$$k(\mathcal{X}, \mathcal{X}') = \frac{\tilde{k}(\mathcal{X}, \mathcal{X}')}{\sqrt{\tilde{k}(\mathcal{X}, \mathcal{X}) \tilde{k}(\mathcal{X}', \mathcal{X}')}} \quad (27)$$

The next question is how to compute the integral in Equation (25). First, Equation (23) is expanded on a basis consisting of spherical harmonics $Y_{lm}(\hat{\mathbf{R}})$ and a set of orthogonal radial basis functions $g_b(|\mathbf{R}|)$:

$$\rho_{\mathcal{X}}(\mathbf{R}) = \sum_{blm} c_{blm} g_b(|\mathbf{R}|) Y_{lm}(\hat{\mathbf{R}}) \quad (28)$$

where c_{blm} are the expansion coefficients. Next, the definition of each element in the power spectrum vector $\hat{\mathbf{p}}(\mathcal{X})$ is given:

$$p(\mathcal{X})_{b_1 b_2 l} = \pi \sqrt{\frac{8}{2l+1}} \sum_m (c_{b_1 l m})^* c_{b_2 l m} \quad (29)$$

Finally, the expression for the SOAP kernel function is:

$$\tilde{k}(\mathcal{X}, \mathcal{X}') = \hat{\mathbf{p}}(\mathcal{X}) \cdot \hat{\mathbf{p}}(\mathcal{X}') \quad (30)$$

One salient advantage of GPR-based MLIPs is the capability to deploy Bayesian inference for estimating the uncertainties of predicted configurations. This allows for the incorporation of configurations with large prediction errors into the training set after performing first-principles calculations, enabling on-the-fly functionality. This unique advantage has garnered considerable scholarly attention. For instance, Jinnouchi et al. employed straightforward radial and angular distribution descriptors coupled with a

Table 1. Selected common databases.

Database	Year	Authors	Refs.
QM series		Lilienfeld et al.	[96–100]
MD17	2017	Tkatchenko et al.	[61]
ANI-1	2017	Isayev et al.	[101]
rMD17	2020	Lilienfeld et al.	[102]
ANI-1ccx and ANI-1x	2020	Isayev et al.	[103]
MD22	2022	Tkatchenko et al.	[104]

GPR approach akin to GAP.^[91] Their methodology was incorporated into the Vienna Ab initio Simulation Package (VASP).^[92] Similarly, Vandermause et al. proposed a novel kernel function for describing two-body and three-body interactions, which facilitated a fully Bayesian automatic training of this GPR-based MLIP model called FLARE.^[93] Moreover, by circumventing the need for high-dimensional descriptors, they reduced the size of the phase space, thereby achieving comparable accuracy to Behler-Parrinello framework-based MLIP on smaller datasets.

It is worth noting that, due to the introduction of local approximations in the construction of descriptors, MLIP based on GPR also suffers from limitations in describing long-range interactions. To address this issue, Grisafi et al. introduced the long-distance equivariant (LODE) representation.^[94,95] By formulating an atomic structure representation based on local atomic density potential values in a Coulombic form, and then combining this with SOAP descriptors within a GPR framework, they significantly enhanced the accuracy of SOAP-GAP in systems with pronounced long-range effects.

2.6. Construction of MLIP

In summary, the construction of an MLIP typically involves three major steps: dataset preparation, descriptor generation, and model training.

The dataset preparation is typically from an AIMD simulation trajectory, which should be performed at a temperature as high as possible above the final simulation temperature to sample a larger phase space. Alternatively, a series of configurations can be obtained using a sampling algorithm, such as meta-dynamics, which can overcome the limitations of standard MD simulations in sampling chemical reactions. Monte Carlo algorithms can also be used to sample equilibrium systems. Another possible way to get data is from publicly available datasets, which are commonly used in this field. A summary of frequently utilized datasets is provided in **Table 1**. Additionally, combining active learning methods with iterative sampling can result in a more complete dataset.

Descriptor generation is a crucial step, and different descriptors have varying levels of computational cost and accuracy. In some cases, the choice of descriptor may depend on the specific system being studied. The optimal descriptor can be determined through testing or by referencing previous literature. For descriptors with many parameters, setting the parameters can be a tedious and experience-dependent task, and global optimization algorithms such as genetic algorithms can be employed.^[105] Fur-

thermore, cutoff radius testing of the descriptor is necessary,^[106] and a radius of 6 Å is often appropriate but may need to be increased for systems with significant medium-range interactions, but it also depends on the specific MLIP model.

During model training, the loss function should first be defined. Typically, the error from the total system potential energy and the atomic forces is a suitable choice. The number of training steps required depends on the descriptor and system but is generally proportional to the size of the dataset. The completion of training is determined by monitoring the learning curve. After training, thorough testing is necessary to ensure that the potential function can accurately simulate the target system.

It's worth noting that additional details critical to the successful construction of MLIPs have been well-documented in existing literature.^[72,113–115] Furthermore, the user documentation for various MLIP construction software packages (outlined in **Table 2**) can serve as an invaluable resource. These key elements, though not elaborated upon in this manuscript, are instrumental in ensuring the success of an MLIP implementation.

2.7. Performance of MLIP

Following the training of an MLIP, high-precision MD simulations can be conducted under a wide array of conditions. As shown in **Table 3**, the mean absolute error (MAE) of the energy relative to the fitting target typically lies within the 10^{-1} to 10^1 meV range, and the MAE of the force exhibits a similar range of 10^{-1} to 10^1 meV Å⁻¹. As a point of comparison, the energy-fitting MAE of empirical potential generally approximates 10^1 to 10^2 meV.^[116–119] It is essential to note that empirical force fields are constrained by their limited fitting capacity, and often necessitate a top-down fitting method to approach experimental data. This method can only guarantee precise fitting for specific conditions or properties, making it challenging to ensure fitting accuracy across all properties. An illustrative example is the simulation of phonon spectrum, MLIP can achieve a level of consistency with DFT calculations that significantly surpasses that of empirical potential (**Figure 6a**).^[120] Another example is the simulation of the water phase diagram, empirical potential simulations can only partially reproduce experimental results, whereas MLIP generally yields results in line with experimental data (**Figure 6b**).^[121]

However, one must be cognizant of the fact that the numerous parameters in MLIPs contribute to increased computational costs, typically reducing the computational speed by 1 to 2 orders of magnitude compared to empirical potential (**Figure 6c**).^[122] Despite this, MLIPs maintain a speed

Table 2. Tools for construction of MLIPs (Note: The full names of the MLIPs models are: atomic cluster expansion (ACE), accurate neural network engine for molecular energies (ANAKIN-ME, ANI for short), fourth-generation high-dimensional neural network potential (4G-HDNNP), directional message passing neural network (DimeNet), deep potential (DP), deep tensor neural network (DTNN), atomic central symmetric function (ACSF), neural equivariant interatomic potentials (NeQUIP), Faber–Christensen–Huang–Lilienfeld (FCHL), Gaussian approximation potential (GAP), smooth overlap of atomic positions (SOAP), polarizable atom interaction neural network (PaiNN), symmetric gradient domain machine learning (sGMDL).

Tools	Model	Dataset prepare	MD interface
ACE1pack	ACE ^[107]	Disable	ASE, LAMMPS
Allergro	Allergro ^[108]	Disable	ASE, LAMMPS
ASE_ANI	ANI ^[101]	Disable	ASE
Charge_transfer_nnp	4G-HDNNP ^[83]	Disable	Disable
DimeNet	DimeNet ^[109]	Disable	Disable
DeePMD-kit	DP, ^[51,52] DPLR ^[82]	Concurrent learning	ASE, LAMMPS, GROMACS
DTNN	DTNN ^[88]	Disable	Disable
n2p2	ACSF ^[48]	Disable	LAMMPS, CabanaMD
NeQUIP	NeQUIP ^[89]	Disable	ASE, LAMMPS
PhysNet	PhysNet ^[54]	Disable	ASE
QML	FCHL ^[58]	Disable	Disable
QUIP	GAP ^[110] , SOAP-GAP ^[60,90]	on the fly	ASE, LAMMPS, CP2K, CASTEP
SchNetPack	SchNet ^[111] , PaiNN ^[112]	Disable	ASE, LAMMPS, SchNetPack
sGMDL	sGMDL ^[57]	Disable	ASE
SpookyNet	SpookyNet ^[55]	Disable	ASE

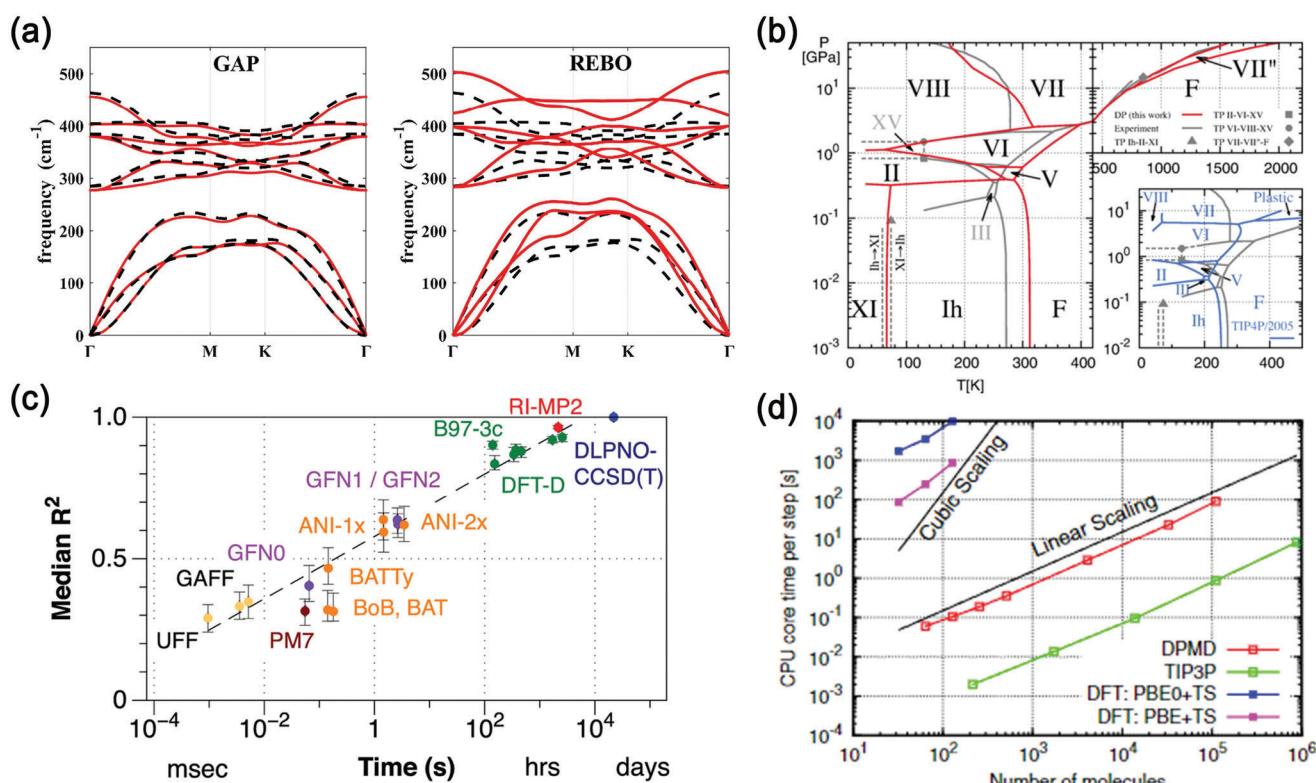


Figure 6. Accuracy and speed performance of MLIP. a) The phonon spectrum of single-layer MoS₂ is calculated by GAP (left) MLIP and REBO (right) force field, where the dashed line denotes the DFT calculation result. Reproduced with permission.^[120] Copyright 2022, Elsevier. b) The water phase diagram is simulated by DP MLIP (red line) and TIP4P/2005 force field (blue line in the inset), where the gray line denotes the experimental result. Reproduced with permission.^[121] Copyright 2021, American Physical Society. c) The correlation between the median R^2 , which signifies the accuracy of various calculation methods compared to DLPNO-CCSD(T) energies, and the corresponding single-core computational time. Reproduced under the terms of the CC BY 4.0 license.^[122] Copyright 2020, the Authors. Published by Wiley-VCH GmbH. d) Time cost versus system atom number with different computational methods. Reproduced with permission.^[52] Copyright 2018, American Physical Society.

advantage that exceeds two orders of magnitude when compared to a variety of first-principles methods. The specific magnitude will vary depending on the method utilized.^[123,124] Moreover, the computational costs of MLIPs and the number of atoms in the simulated system exhibit a linear growth relationship,^[52] while, for DFT methods, the computational cost is generally believed to increase with the cube of the atom number of system (Figure 6d).^[52,125] This renders MLIPs more advantageous for large-scale system calculations. Therefore, MLIPs currently offer the most optimal balance between accuracy and speed.

3. Applications of MLIP for Surfaces/Interfaces of Nanomaterials

While the foundational concepts of MLIP have existed for over 15 years, their prominence in the study of nanomaterials only surged recently due to advancements in computational tools, data processing, and model architecture. In this section, we highlight pivotal computational studies illustrating MLIP applications for nanomaterial surfaces/interfaces. These studies are broadly classified into two categories: inorganic systems and molecular systems. We will also delve into the design of MLIP-centric methodologies, emphasizing simulation tasks and workflows tailored for nanomaterial surface/interface investigations.

3.1. Inorganic Systems

Inorganic systems, characterized by a limited elemental variety and absence of intricate covalent structures, predominantly encompass metals, alloys, inorganic non-metals, and oxides. For the purposes of this review, “inorganic systems” pertain to typical inorganic nanomaterials and nanostructures, which markedly differ from molecular system configurations. The interfaces of such systems predominantly comprise solid surfaces, nanoparticle exteriors, and their interactive boundaries with the ambient environment. Our discussion centers on the utility of MLIPs in

deciphering the intrinsic and environmental interplay of these inorganic system surfaces.

3.1.1. Intrinsic Stability of Surfaces/Interfaces

The surface/interface attributes are pivotal in crafting stable, functional nanomaterials suitable for diverse applications, given their significant influence on nanomaterial stability through factors like surface energy, defects, functionalization, and morphology. Crafting thermodynamically stable and producible nanostructures remains a daunting task in computational material design.^[126] Inorganic systems typically involve metals, alloys, semiconductors, and oxides, either supported on substrates or existing as independent nanoparticles. While periodic slab models are prevalent for elementary surface/interface modeling, real-world surfaces, which often exhibit imperfections and operational modifications, present additional structural intricacies. Accurately simulating such a system mandates the creation of expansive initial unit cells, a requirement that frequently outstrips the capabilities of traditional first-principles methods. Empirical potential-driven MD and Monte Carlo (MC) simulations, though popular for studying nanomaterial surfaces/interfaces, fall short in considering quantum effects, thus undermining their accuracy and reliability. Presently, MLIPs, ensuring an optimal blend of efficiency and precision, are the potentially go-to choice for investigations into nanomaterial surfaces/interfaces, including surface restructuring, nanoparticle structural fine-tuning, and heterogeneous catalyst design.^[127,128]

MLIPs for Surface Reconstruction: The nature of solid surfaces lies at the heart of many device functionalities related to substrate and film attributes. The atomic configurations of these surfaces significantly influence their inherent properties.^[129] However, modeling certain complex surface reconstructions, like the silica surface, is challenging.

Take the Si(111) surface as an example. It displays intricate $(2n+1) \times (2n+1)$ reconstructions, with the 7×7 structure being the energetically most favored, characterized by adatoms, 10-atom rings, and stacking errors.^[130] Conventional empirical

Table 3. Different MLIP's MAE on MD17 database (Note: The molecule I–VII represents aspirin, ethanol, malonaldehyde, naphthalene, salicylic acid, toluene, and uracil, respectively. All data are from previous literature.^[51,54,89,107,108] The data inside and outside the brackets are the MAE of energy and force respectively, and the corresponding units are meV and meV Å⁻¹).

	I	II	III	IV	V	VI	VII
GAP ^[110]	17.7 (44.9)	3.5 (18.1)	4.8 (26.4)	3.8 (16.5)	5.6 (24.7)	4.0 (17.8)	3.0 (17.6)
ANI ^[101]	16.6 (40.6)	2.5 (13.4)	4.6 (24.5)	11.3 (29.2)	9.2 (29.7)	7.7 (24.3)	5.1 (21.4)
DP ^[51,52]	6.7 (12.1)	2.2 (3.1)	3.3 (4.4)	5.2 (5.5)	5.0 (6.6)	4.4 (5.8)	4.7 (2.8)
ACE ^[107]	6.1 (17.9)	1.2 (7.3)	1.7 (11.1)	0.9 (5.1)	1.8 (9.3)	1.1 (6.5)	1.1 (6.6)
sGMDL ^[57]	7.2 (31.8)	2.4 (16.0)	3.1 (18.8)	0.8 (5.4)	2.1 (12.8)	1.0 (6.3)	1.4 (10.4)
FCHL ^[58]	6.2 (20.9)	0.9 (6.2)	1.5 (10.3)	1.2 (6.5)	1.8 (9.5)	1.7 (8.8)	0.6 (4.2)
SchNet ^[111]	16.0 (58.5)	3.5 (16.9)	5.6 (28.6)	6.9 (25.2)	8.7 (36.9)	5.2 (24.7)	6.1 (24.3)
DimeNet ^[109]	8.8 (21.6)	2.8 (10.0)	4.5 (16.6)	5.3 (9.3)	5.8 (16.2)	4.4 (9.4)	5.0 (13.1)
PaiNN ^[112]	6.9 (14.7)	2.7 (9.7)	3.9 (13.8)	5.0 (3.3)	4.9 (8.5)	4.1 (4.1)	4.5 (6.0)
PhysNet ^[54]	5.2 (1.7)	2.2 (0.9)	3.0 (1.3)	5.2 (1.3)	4.7 (1.3)	4.3 (1.3)	4.3 (1.3)
SpookyNet ^[55]	6.5 (11.2)	2.3 (4.1)	3.4 (7.2)	5.0 (3.9)	4.9 (7.8)	4.1 (3.8)	4.6 (5.2)
NequIP ^[89]	5.7 (8.0)	2.2 (3.1)	3.3 (5.6)	4.9 (1.7)	4.6 (3.9)	4.0 (2.0)	4.5 (3.3)
Allegro ^[108]	2.3 (7.3)	0.4 (2.1)	0.6 (3.6)	0.5 (0.9)	0.9 (2.9)	0.4 (1.8)	0.6 (1.8)

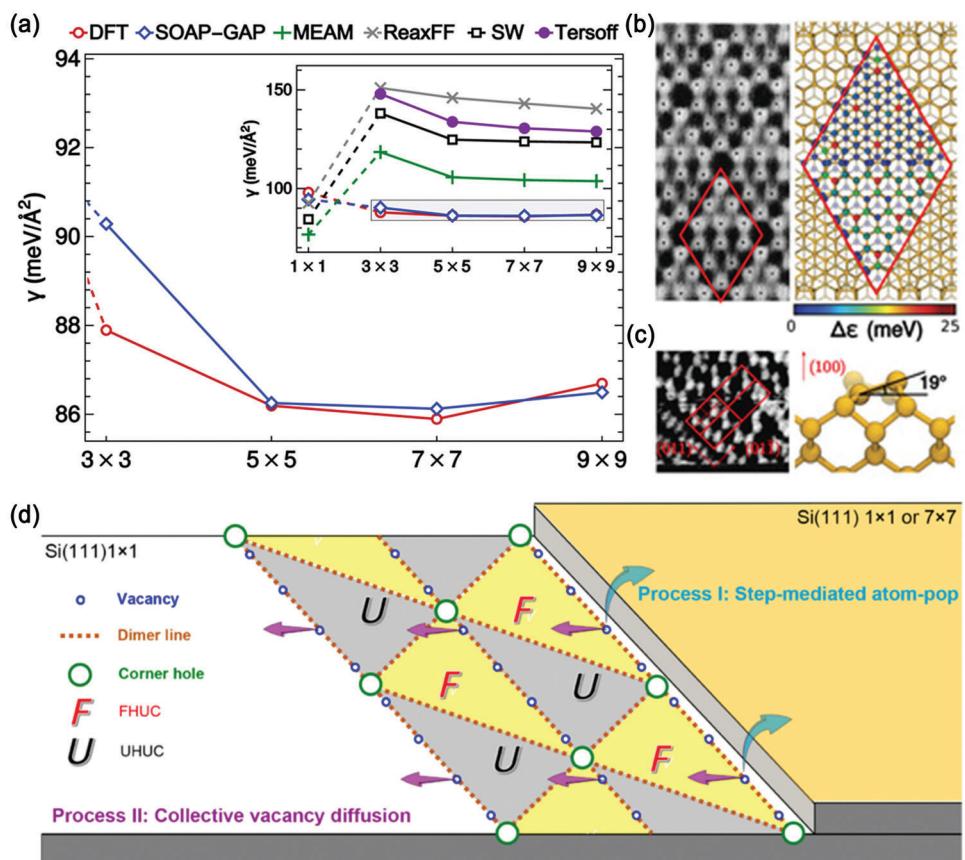


Figure 7. MLIP-based predictions for silicon surface reconstruction. a) SOAP-GAP's energy order prediction versus empirical force fields. b) Si(111)-(7 \times 7) reconstruction: left—STM image; right—SOAP-GAP model output (SOAP-GAP-relaxed structure colored by predicted local energy error when using a training set without adatoms). c) Tilt of dimers on Si(100) due to Jahn-Teller distortion, left—STM image, right—SOAP-GAP output. Empirical models failed to show this tilt. a–c) Reproduced under the terms of the CC BY 4.0 license.^[110] Copyright 2017, The Authors. Published by American Association for the Advancement of Science (AAAS). d) Proposed kinetic processes for Si(111)-(7 \times 7) reconstruction.: Process I involves a step-mediated atom-pop process (two cyan arrows), creating a row of vacancies (blue circles) adjacent to the step. Process II sees collective vacancy diffusion (six purple arrows) leading to the 7 \times 7 reconstruction with faulted ("F") and unfaulted ("U") components. Reproduced with permission.^[133] Copyright 2021, American Physical Society.

potential simulations, while effective for some bulk properties of materials, struggle with accurate surface behavior predictions. As shown in Figure 7a, these methods misidentified the Si(111) surface cell's energy ordering. In contrast, Bartok et al. applied an MLIP approach with notable success to silicon surface reconstructions.^[110] Their model, SOAP-GAP, correctly predicted the Si(111)-(7 \times 7) is an energy minimum (Figure 7a,b) and the 19° tilt of Si(100) dimers (resulting from Jahn-Teller distortion) in agreement with DFT (Figure 7c),^[131] even when trained on limited datasets, showcasing the MLIP's strong extrapolation capabilities. At present, this case has served as a testing ground for many MLIP models.^[113,132]

Further advancements were made by Hu et al., who utilized a high-dimensional artificial neural network (ANN) potential to deeply probe the Si(111)-(7 \times 7) surface reconstruction.^[133] Their approach, which leveraged the precision of the ANN potential and integrated the capabilities of climbing image nudged elastic band (CI-NEB) and MD simulations,^[134] expanded the simulation's scale to encompass approximately 17 000 atoms and extended the time scale to span several nanoseconds.

Through their comprehensive analysis, Hu et al. discovered two primary kinetic processes that pave a feasible path toward the Si(111)-(7 \times 7) reconstruction. The first is a step-mediated atom-pop rate-limiting process that instigates extensive nonconserved atomic rearrangements. The second crucial process involves collective vacancy diffusion, directing a series of selective formations: dimer, corner-hole, stacking-fault, and dimer-line patterns, which culminate in the full reconstruction, as visualized in Figure 7d. Their findings highlight the potential of MLIPs, when combined with other techniques, to decipher even more intricate surface structures and processes in upcoming research.

Carbon nanomaterials, comprising nanostructures like carbon nanotubes, graphene, and carbon nanospheres, are renowned for their unique electrical, thermal, mechanical, and chemical attributes. Among these, amorphous carbon (a-C) stands out due to its high surface area, adsorption capability, and chemical reactivity, making it ideal for catalysis, electrochemistry, and environmental applications.^[135] Nonetheless, its absence of a clear crystalline structure and intricate surface morphology pose

challenges in deciphering and enhancing its surface stability and reactivity.

The computational modeling community has taken a keen interest in a-C, given its diverse bonding versatility, which results in microscopic variability. Initial research gravitated towards empirical force field optimization for a-C, supplemented by ab initio MD simulations. An enduring challenge has been the accurate representation of the high sp³-carbon content in tetrahedral amorphous carbon (ta-C). While empirical force fields have struggled in this endeavor, recent endeavors by Caro and Deringer have harnessed MLIPs for exhaustive studies on a-C.^[136–140]

Caro et al., in 2018, utilized a GAP-based simulation to replicate the observed high sp³ content in ta-C films, offering insights into the microscopic growth mechanisms.^[140] Concurrently, Deringer et al. probed the reconstruction, hydrogenation, and oxidation of amorphous carbon surfaces using GAP MLIP.^[137] They adopted a series of surface slab models, derived from bulk ta-C samples, and systematically analyzed these surfaces, determining a 216-carbon atom model as optimal. Despite the GAP being tailored primarily for bulk carbon, the research merged its results with DFT calculations to traverse a broader chemical domain, especially focusing on hydrogen and oxygen-based functionalizations. Employing methods like grand canonical Monte Carlo (GCMC) simulations and DFT-based ab initio MD simulations, they explored hydrogenation up to 30% and intricate oxygenation processes.

Furthering this research on ta-C surface structures, Caro et al. combined GAP MLIP and SOAP-based clustering to categorize atomic motifs of carbon.^[138] They expanded the model to incorporate electronic-structure fingerprints, melding this with a SOAP element. This dual approach offered superior predictions of hydrogen adsorption energies over a standalone SOAP-GAP model, marking a pivotal step toward predicting chemical reactivity. Building on this, Aarva et al., in 2019, harnessed GAP, DFT, and SOAP-based clustering methods to decipher the X-ray absorption spectroscopy (XAS) and X-ray photoelectron spectroscopy (XPS) signatures of functional groups on carbon surfaces.^[139]

MLIPs, beyond silicon and carbon, have found significant applications in studying other inorganic material surfaces. For instance, Timmermann et al. utilized a GAP MLIP to globally optimize low-index rutile IrO₂ facets' geometry through simulated annealing.^[141] Using the cost-effective GAP for the optimization instead of conventional DFT calculations allowed for comprehensive configuration space sampling. Their analysis of five distinct low-index surfaces revealed notably lower-energy structures, especially for (101) and (111) orientations (Figure 8a,b). This strategy, in tandem with experimental and DFT findings, uncovered many new metal-rich surface compositions.

Similarly, Pun et al. fashioned a physics-based bond-order potential for metallic aluminum (Al), exhibiting precision on par with DFT for predicting various characteristics, from lattice dynamics and crack nucleation to surface tension.^[142] Such advancements underscore MLIPs' potential in delving into the multifaceted physical aspects of surfaces and interfaces.

Metals, traditionally deemed as hard matter, have atomic lattices that display dynamic behavior. Intriguingly, these can reconfigure even at temperatures significantly beneath their melting thresholds.^[143,144] This atomic dynamism profoundly influences the core and surface attributes of metals. Dissecting these deli-

cate structural dynamics is paramount for a host of applications, yet it's a formidable challenge.^[145]

In 2023, Cioni et al. presented a breakthrough using DP MLIP-based MD simulations, illuminating the sophisticated dynamics of diverse copper surfaces at atomic resolution (Figure 8c).^[146] To craft the MLIP, they commenced with an initial dataset—sourced from DFT calculations (containing small Cu FCC bulk and specific surface configurations: (100), (110), (111), (211), and (210); involving approximately 100 atoms)—as a precursor for DP MLIP training. They subsequently employed an iterative learning strategy, optimizing the representation of the copper surface configurations and their underlying atomic transitions. Validating their finalized MLIP against broad copper bulk and surface properties (encompassing around 2400 atoms), they embarked on production MD simulations. Their analysis, harnessing SOAP vectors, revealed that the copper (110) surface was uniquely dynamic, becoming notably unstable at 700 K, as visualized in Figure 8d–g. This research signifies the synergy of MLIPs with high-dimensional descriptor-based methods, enabling deeper insights into complex interfacial dynamics and reconstructions. Furthermore, studies suggest the potential marriage of MLIPs with empirical potentials for nuanced material analysis. As a testament, Wang's research integrated an empirical force field (EAM) with MLIPs, where the latter precisely depicted atomic interactions in distortion-intensive regions, relegating EAM to the remaining zones.^[147]

MLIP for Intrinsic Stability of NPs: Inorganic nanoparticles (NPs) stand out in nanoscience research, primarily due to their unique size and vast specific surface area. The NP surface plays a pivotal role because active surface sites influence catalytic activity, surface defects modify electronic and optical properties, and surface modifications enhance biocompatibility and stability. A deep dive into the interfacial properties of inorganic NPs is thus critical for their application.^[126,148]

The inherent stability of NPs is crucial, given its implications for industrial viability.^[149] Take gold nanoparticles (AuNPs) as an example: they find applications in optics,^[150] nanomedicine,^[151] and catalysis.^[152] The structural stability of AuNPs, influenced by their shape, is paramount for practical uses. Gaining insights into the state transitions of AuNPs, especially under high-temperature conditions, is vital for catalytic applications. Traditional molecular simulations, while invaluable, have limits in capturing the nuances of AuNPs. This is where MLIPs come into play, offering promising results in understanding AuNP stability.

In 2021, Zeni et al. developed a set of MLIPs through an innovative mapped Gaussian process framework, providing insights into the melting dynamics of AuNPs of varying sizes via MD simulations.^[153] They focused on AuNPs spanning 1 to 6 nm in diameter, which corresponds to 147 to 6266 atoms. Their ambitious simulations covered 2.4 ms, a time length far beyond the capabilities of electronic structure calculations, even with cutting-edge computational resources. To pinpoint the nuances between surface and inner melting during the solid–liquid phase transition, they harnessed an unsupervised ML technique. This method drew from a curated database of configurations coupled with a local atomic density representation. By enhancing a 3-body local atomic cluster expansion descriptor, they attributed a 40-D feature set to each atom. Subsequently, a hierarchical k-means clustering discerned atoms into six distinct local

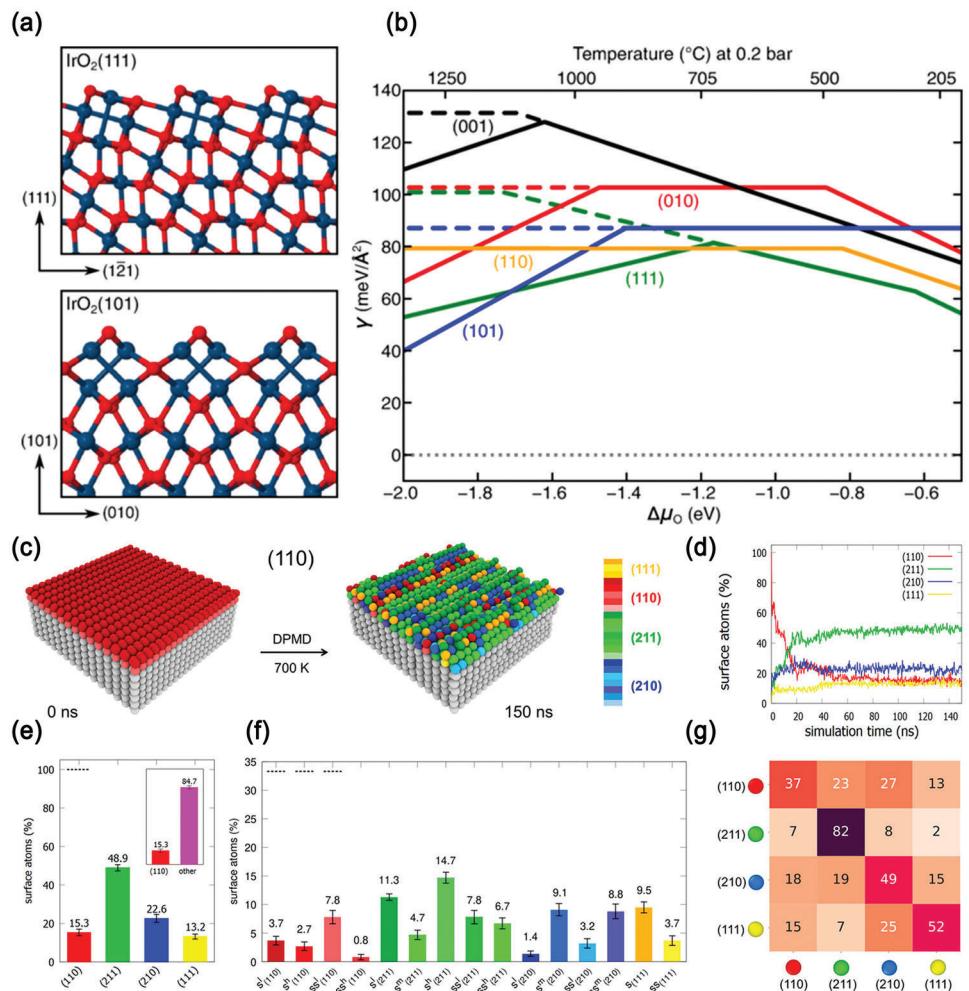


Figure 8. Reconstructions of IrO_2 and $\text{Cu}(110)$ surfaces via MLIPs. a) Side perspectives of IrO_2 complexes with Ir (blue, large) and O (red, small) atoms. b) Calculated surface free energies, γ , for five distinct low-index facets in a pure oxygen atmosphere. On the top x-axis, the dependence on the oxygen chemical potential $\Delta\mu$ is translated to a temperature scale, considering 0.2 bar pressure (oxygen's partial pressure in ambient air). Dashed lines represent surface-free energies in the absence of complexions. a,b) Reproduced with permission.^[144] Copyright 2020, American Physical Society. c) $\text{Cu}(110)$ at 0 ns (left) and post 150 ns DP at 700 K (right). SOAP environments for standard (110) are red, with green, blue, and yellow marking non-native configurations from the (211), (210), and (111) surfaces. d) Variations in (110) surface environments at 700 K over DP time. e) Composition of (110) at 700 K, showing percentages and standard deviations. Inset: red indicates native configurations; pink indicates deviations. f) Composition breakdown of the (110) surface at 700 K; dashed lines represent initial DP composition. g) Transition matrix for atomic shifts on (110) at 700 K within $\Delta t = 300$ ps. c–g) Reproduced with permission.^[146] Copyright 2023, American Institute of Physics.

environmental categories, effectively segregating solid from liquid phases and identifying various coordination surfaces (Figure 9a,b). Crucially, their results unveiled that AuNP melting initiates externally before progressing internally (Figure 9b), a revelation in harmony with experimental data. This groundbreaking study not only paves the way for a deeper understanding of nanoscale phase transitions but also elevates our computational grasp of nanomaterial surfaces and interfaces.

Another relative study on AuNPs was conducted by Chiriki et al. in 2017.^[154] They utilized a neural network potential (NNP) to scrutinize the intrinsic stability of gold clusters. On the one hand, they employed a basin-hopping method coupled with NNPs to search for the global minimum (GM) structures of Au_{17} , Au_{34} , and Au_{58} , where the systematically quenched structures from multiple NNP-based MD trajectories at various tempera-

tures were used. The results showed that the NNP is able to reproduce the GM structures for Au_{17} and Au_{58} from previous DFT calculations, and also predict a new 3-core atom GM structure for Au_{34} (Figure 9c).^[155–157] On the other hand, they further conducted a series of NNP-MD simulations for these AuNPs and demonstrated that the melting process of AuNPs initiates from the core atoms and then spreads out to the surface atoms in the AuNP being studied. Even though the possibility of the GM structure occurring in Au_{34} clusters was low at room temperature, it remained prevalent in Au_{17} and Au_{58} clusters under the same conditions. The same group also constructed NNPs for sodium (Na) clusters ranging from 16 to 40 atoms. They performed NNP-MC simulations to understand the melting process of these Na clusters.^[158] The results of which support the notion of stepwise melting in small Na clusters as previously reported by Aguado

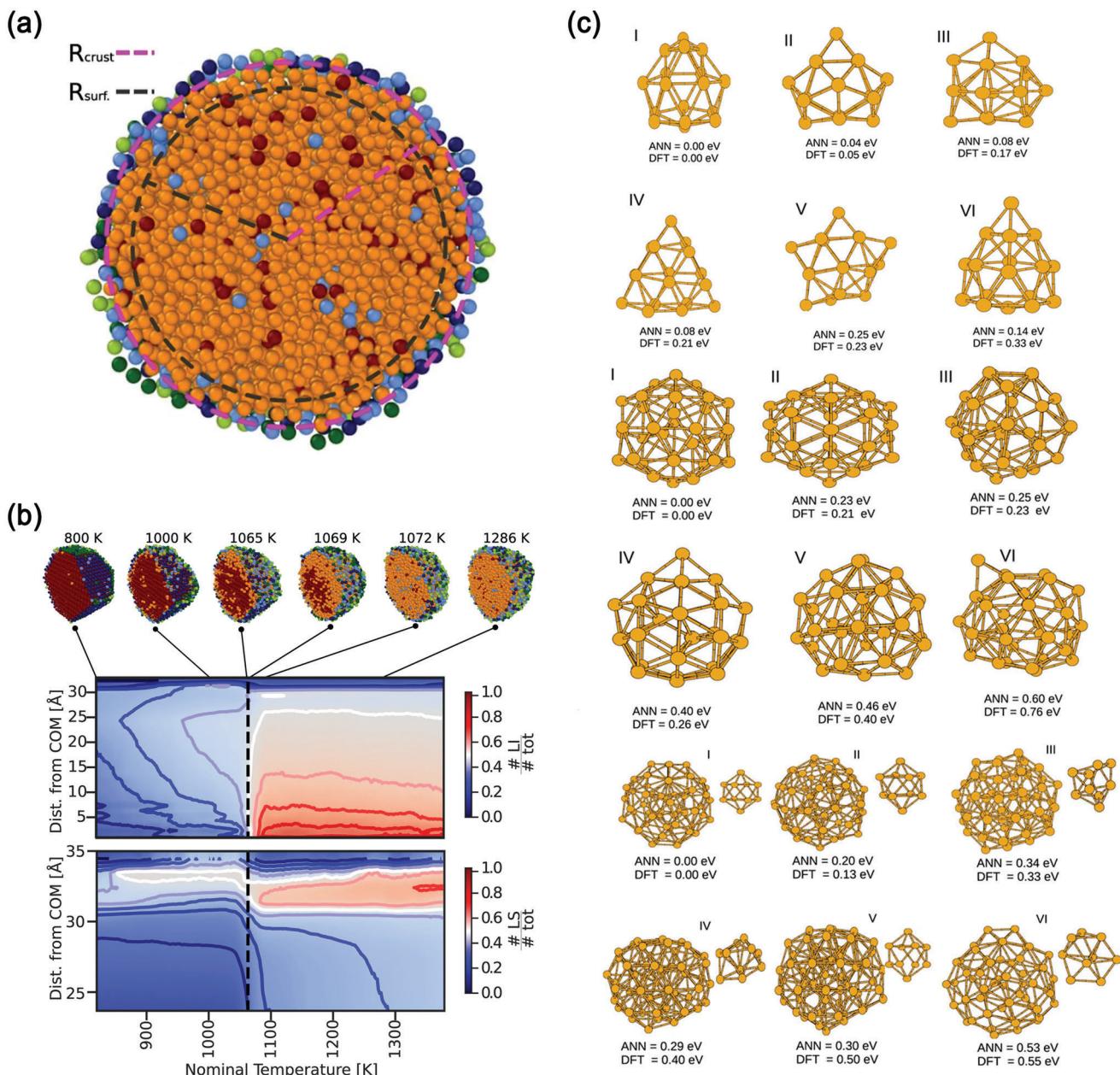


Figure 9. AuNPs melting studied by MLIPs. a) Au₆₂₆₆ NP showcasing R_{crust} values (pink) and R_{surf} values (gray). Color of atoms is differentiated by their class from the clustering algorithm. b) Distribution of liquid environments in an Au₆₂₆₆. Top: Different Au₆₂₆₆ NP snapshots at various MLIP simulation temperatures. Center and bottom: Average fractions of #LI (center) and #LS (bottom) local atomic environments based on radial distance from the COM (y-axis) and system temperature (x-axis). Bold lines in the center and bottom highlight isosurfaces from 0 to 1 at 0.1 intervals, while the black dashed line marks the T_{melt} . a,b) Reproduced under the terms of the CC BY 4.0 license.^[153] Copyright 2021, the Authors. Published by Springer. c) Low energy isomers for Au₁₇ (up sixer isomers), Au₃₄ (center six isomers), and Au₅₈ (bottom six isomers). Reproduced with permission.^[154] Copyright 2017, American Institute of Physics.

and Kostko.^[159] Later on, with the help of MLIP, researchers also studied the intrinsic stability of larger AuNPs and other elemental NPs.^[160–163] Drawing upon the ideas of these studies, utilizing MLIP enables researchers not only to conduct MD simulations but also to incorporate various optimization algorithms for exploring optimal structures along the machine-learned PES. This idea is extremely valuable in identifying local minima on the PES, which in turn aids in exploring the mechanisms

of catalytic reactions through the analysis of these metastable structures.^[44]

Moreover, incorporating structural searching algorithms into the MLIP training dataset construction enhances the sampling of the PES, resulting in a more robust MLIP model. This is particularly crucial for studying nanoalloy (NA) catalysts. As the number of metal elements in these catalysts rises—from binary and ternary to high-entropy alloys with five or more

components—the potential surface structures multiply exponentially. This growing complexity in their configuration space emphasizes the need for efficient methods to identify stable and highly catalytic structures.^[164,165] For instance, many studies on bimetallic NAs combine MLIP with MC or optimization algorithms to delve into their structure–properties relationship, leading to advancements in our understanding of alloy catalyst properties and paving the way for superior heterogeneous catalysts.^[166–169]

A recent study by Han et al. showcases the integration of a symmetry-constrained genetic algorithm (SCGA) with NNP to examine the stability of Pt–Ni bimetallic NAs.^[170] The SCGA, tailored for symmetric NAs, works alongside NNP to explore a vast space of homotops and compositions of Pt–Ni NAs with up to 4033 atoms, shedding light on how shape, size, and composition influence dominant chemical ordering patterns. It is essential to mention that training datasets should encompass configurations of nanoparticles (NPs), surface slabs, and bulk structures to ensure the model's accurate simulation of larger nanoalloys.^[169]

In many scenarios, NPs often interact with substrates, resulting in intricate interface structures. However, MLIPs are adept at handling these complexities. A notable example is Hu et al.'s research, which combined experiments and SSW-NN-based MD simulations to reveal a Sabatier principle of metal–support interaction, setting the stage for ultrastable nanocatalysts.^[171] Another example is Kolsbjerg et al.'s study on MgO-supported Pt nanoparticles.^[172] This study showed that using a NNP-based MD for structural relaxation, combined with an evolutionary algorithm in a DFT framework, dramatically boosts computational speed. This workflow facilitates the accurate identification of global minimum energy structures, a factor critical for understanding the catalytic reactivity of NP shapes. Furthering this line of inquiry, Lim et al. in 2020 provided a comprehensive atomistic examination of the restructuring of Pd on Ag through microscopy, spectroscopy, and GAP-based MD simulation.^[173] With accelerated MLIP MD and an automated analysis method, they unveiled novel surface restructuring mechanisms. This methodology offers deep insights into complex multimetallic systems, presenting a level of atomic resolution previously unreachable.

3.1.2. Interactions Between Surface And Environment

In natural settings, nanomaterials invariably interact with their environment. These interactions, be they with the atmosphere or in solution, profoundly influence the nanomaterials' stability and distinct properties. Take heterogeneous catalysis as an example: environmental gas components can modify active site distributions on catalyst surfaces, altering their catalytic efficiency.^[174] In solution, the impact of solvents and solutes on the surface and interface may be even more complex.^[175] Despite advances, achieving real-time or *in situ* characterization of nanomaterial surfaces and their operando interactions remains a formidable task. And while novel characterization methodologies continue to emerge, we must understand that each has its inherent limitations.

In a 2022 study, for example, Pullanchery et al. employed vibrational sum-frequency scattering spectroscopy (VSFS) to examine the interfacial vibrational spectrum at a hexadecane–water boundary, focusing on the oxygen–deuterium and carbon–hydrogen

stretches.^[176] Pairing this with MD simulations, they found water molecules form atypical interfacial hydrogen bonds with alkyl hydrogens. This bonding phenomenon leads to a charge transfer from water to oil, lending stability to oil droplets. This research sheds light on why oil disperses in water as stable droplets with negative electrophoretic mobility. Although VSFS offers invaluable insights, it cannot solely pinpoint the atomic structures influencing the spectra. Many surface characterization approaches, including SFG, STM, and XPS, rely on molecular simulations for detailed atomic interpretations. In this landscape, the demand for precise and swift computational tools grows. Recently, MLIP rises to this challenge.

Gas Environment: The advent of MLIPs has revolutionized the capability to predict interactions between molecules or atoms and solid surfaces, providing deeper insights into subsequent surface property alterations. This innovation finds promise in diverse areas, such as catalysis, gas storage, and separation.^[128]

Heterogeneous catalytic reactions, especially those at the solid–gas interface, serve as prime examples. The efficiency of a catalyst is determined by the adsorption–desorption patterns of the reactants and products as well as any surface restructuring. In 2018, Sun et al. conducted a study on the catalytic performance of Pt₁₃ NPs.^[177] They first trained an HDNNP potential for Pt_nH_x, then used a genetic algorithm to extensively search for low-energy metastable ensemble (LEME) structures of Pt₁₃ and Pt_nH_x. In their work, Pt₁₃ with varying H coverage (Pt_nH_x) was explored as a catalytic environment under different hydrogen pressures. From the search results, they analyzed the catalytic performance based on the three global minimum structures for hydrogen evolution reaction (HER) and methane activation (Figure 10a–c). Their study highlighted the vital role of metastable structures in catalytic reactivity, specifically for Pt_n NPs under hydrogen pressure. Their results indicate that local rearrangements might generate less stable but more reactive sites, which has significant implications for catalyst design. Rather than focusing solely on the most stable structures, it is also important to consider low-energy metastable isomers with higher reactivities. This idea of metastability-triggered surface reactivity is crucial, not just in catalysis, but also in areas like corrosion science, growth, and other surface-related processes. In 2017, Ulissi et al. introduced a comprehensive method to tackle the challenge of modeling multiple heterogeneous active sites on polycrystalline surfaces.^[178] They established an iterative ML and DFT optimization framework, enabling precise prediction of the adsorption energy of CO on the NiGa bimetallic surface. In this framework, a set of NNPs were employed to relax all potential adsorption configurations, select configurations for DFT study, augment the training set, and refine. The use of NNPs allows for information sharing while drastically cutting computational costs, leading to both accurate and efficient modeling. Furthermore, this framework can be extended to effectively catalog and pinpoint new motifs in a wide variety of intermetallic compounds.

Additionally, a recent study by Xu et al. on MLIP-based GCMC for the oxidation of flat and stepped platinum surfaces is noteworthy.^[179] On one hand, they incorporated a range of exploration algorithms, including MD and GA, combined with model uncertainty estimation and descriptor-based configuration selection, to create an efficient active learning pipeline for MLIP dataset construction. This approach is highly inspiring

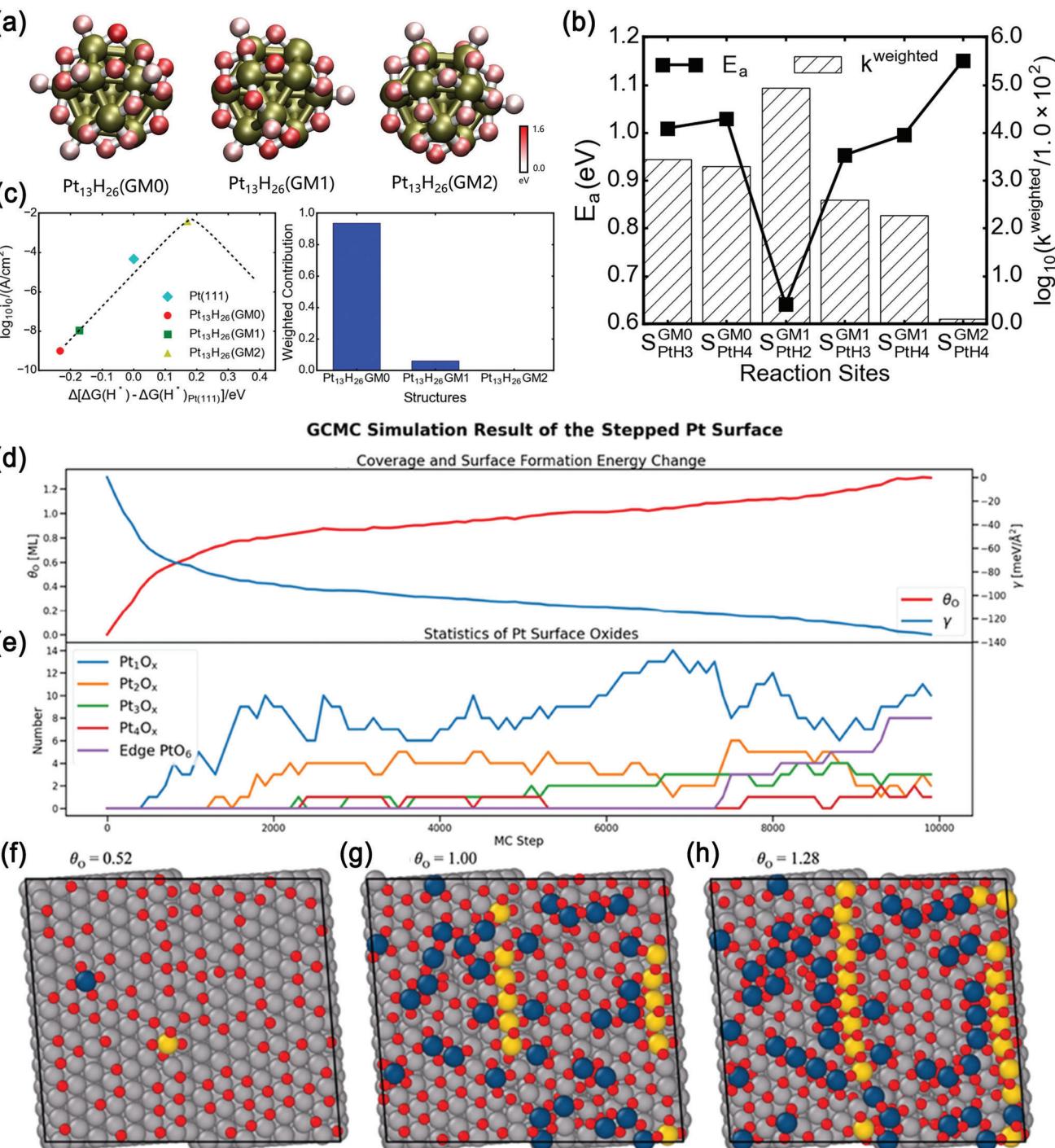


Figure 10. Pt NPs properties under hydrogen environment and Pt surface oxidation under oxygen conditions studied by MLIPs. a) Three of the most stable $\text{Pt}_{13}\text{H}_{26}$ clusters (GM0, GM1, GM2), the teal green color represent Pt atoms, and the red represents H atoms (the darker is the red color, the stronger is the binding of H on Pt). b) Methane activation on the $\text{Pt}_{13}\text{H}_{26}$. c) Activity for HER of the three most stable structures of the $\text{Pt}_{13}\text{H}_{26}$. a–c) Reproduced with permission.^[177] Copyright 2018, American Chemical Society. d–h) Insights from the GCMC simulation on the oxidation of the stepped p(2 × 12) Pt (17 15 15) surface. d) Evolution of coverage and surface energy during the simulation. e) Variation in local surface oxide entities, inclusive of elevated Pt atom clusters ranging from one to four atoms, and edge-based octahedral PtO_6 units. f–h) Selective structural visuals at specific coverages, where blue depicts elevated Pt atoms and yellow illustrates the edge Pt atoms with a minimum of four O neighbors. d–h) Reproduced under the terms of the CC BY 4.0 license.^[179] Copyright 2022, American Chemical Society.

for constructing interfacial MLIP because it allows for the collection of sufficient interface reconstructions, various configurations involved in the interface oxidation process, and rare events, which are difficult to achieve with traditional iterative methods. On the other hand, for production simulations, they utilized a GCMC simulation workflow, enabling the insertion or removal of gas to manage oxygen pressure. This allowed for the simulation of operando conditions, which are challenging to implement in AIMD. Given that numerous interface oxidation simulations for expansive systems lean heavily on reactive force fields—and often lack force field parameters for many elements—the combined MLIP and GCMC approach offers a practical solution for modeling oxidation on metal surfaces and related corrosion phenomena.

Moreover, the process of surface oxidation and surface buckling as depicted by this MLIP-based GCMC study is of notable practical importance. It provides an enriched set of references and avenues for researchers delving into the intricacies of fracture and corrosion mechanisms in metallic materials. Figure 10d–h display GCMC results for a representative trajectory. The surface formation energy declines rapidly with increasing O coverage, slowing considerably once the coverage hits 0.60 ML (as shown in Figure 10d). Figure 10e reveals the first instance of a Pt atom rising on the terrace at approximately 0.40 ML coverage. As more O atoms are adsorbed, additional Pt atoms on the terrace elevate to form local surface oxide units. When the coverage exceeds 1.00 ML, a few octahedral PtO_6 structures appear, each with a Pt atom sharing the same local chemical environment as in $\alpha\text{-PtO}_2$. Figure 10f–h illustrates representative structures at specific coverages. During the initial stages of oxidation, Pt atoms on the terrace rise to form on-surface square planar PtO_4 oxide units, while edge-located Pt atoms also oxidize into a square planar PtO_4 unit but with an additional subsurface O atom. Further oxidation elongates the on-terrace PtO_2 stripes, and the edge Pt atoms are fully oxidized to a sixfold-coordinated state. Such revelations are pivotal for theoretical and mechanical explorations into the corrosion and fracturing of metal surfaces. This is evident in the chemo-mechanical coupling model concerning the oxidation and interlayer fracturing of copper nanowires, as recently put forth by Gong et al.,^[180] where the chemical aspects of metal oxidation assume a vital role. It further underscores the vast potential methods akin to MLIP-GCMC hold for forthcoming applications in this domain.

In another representative study, Liu et al. utilized an MLIP-driven semi-grand canonical Monte Carlo (SGCMC) simulation to explore the surface behavior of Pd on an FCC (111) AgPd surface in the context of acrolein's presence or absence.^[181] They applied MLIP to estimate energies related to the AgPd bulk, the AgPd slab, and the adsorption of acrolein on the AgPd slabs. Subsequent MLIP-guided SGCMC simulations projected the patterns of surface segregation and aggregation across varying Pd bulk concentrations. Their findings revealed that in a vacuum, acrolein induces the aggregation of Pd as dimers on the surface, even at a low 6.5% Pd bulk concentration. In contrast, without acrolein, Pd presence on the surface is negligible for concentrations below 20%. This research not only offers a meticulously devised MLIP-based simulation approach but also illuminates the intricate dynamics between surfaces and their environments. While these investigations underscore MLIP's pivotal role

in understanding surface-environment interactions, the inherent complexity of such interactions remains a crucial consideration. There's optimism that MLIPs might facilitate simulations under even more challenging operando conditions in the future.

Solution Environment: In solutions, interactions between nanomaterials and their surrounding environment differ significantly from gas environments. For instance, when solvents adsorb onto nanomaterial surfaces, properties like stability, catalytic activity, and surface charge may be altered. Concurrently, interfaces or distinct confinement structures between nanomaterials and solvent molecules might display unique properties or influence the system's mass and heat transfer behaviors.^[174,175] Although studying these interactions is crucial, capturing atomic-level information about interfacial behavior within solutions experimentally is challenging. As a remedy, molecular simulations often pair with specific characterization techniques to examine unique interfacial phenomena. Generally, MD simulations are preferred for systems exceeding 1000 atoms. Yet, empirical force fields often do not have specially fitted parameters for some interfacial systems. Even if some force fields have been proposed for interfacial applications, many of them use fixed charge parameterization, which may not necessarily describe many interfacial phenomena in solution.^[182,183] Another solution for modeling these systems is AIMD simulation, it has been employed to explicitly consider solvent effect for such solution interfacial systems in recent years. However, they are limited to hundreds of atoms in tens of picoseconds, which may result in inadequate sampling and problems from finite-size effects. Therefore, more efficient and accurate simulation methods have always been needed. To address this issue, in recent years, MLIPs have been widely applied to the study of solution systems, such as the properties of water molecules and electrochemical interfaces.^[184,185]

For example, in 2014, Artrith and Kolpak used an NNP-based lattice MC method to explore the atomic arrangement of Au/Cu NPs and surface models in water.^[186] Initially, they optimized the structures and compositions of Au/Cu NPs in a vacuum using grand canonical simulated annealing facilitated by the NNP. Their results indicated that Au-coated core-shell NPs exhibit thermodynamic stability in a vacuum, regardless of the Au/Cu chemical potential and NP size. Conversely, in aqueous environments, NPs with mixed Au–Cu surfaces gained preference. Specifically, when a static solvation shell was applied, Cu atom segregation was evident at the interface (Figure 11a,b). Moreover, they identified significant differences in the adsorption energies of CO and O_2 between NPs with equilibrium surface compositions in water compared to those in a vacuum. Such disparities have profound implications on the electrocatalytic properties of the material, emphasizing the critical role of solvent effects in electrocatalyst models.

In 2016, Natarajan and Behler employed an NNP-based MD simulation to scrutinize the structural and dynamic characteristics of interfacial water on different Cu surfaces (Cu (111), Cu (100), and Cu (110)).^[187] They meticulously constructed a high-precision NNP specific to the Cu/water interface and subjected it to rigorous convergence tests. These tests assessed the required system size, focusing on water film diameter and supercell dimensions, given that interfaces are known to modify water structures up to 10 Å from the surface. Building on this foundation, they explored a myriad of properties of the interface water, such

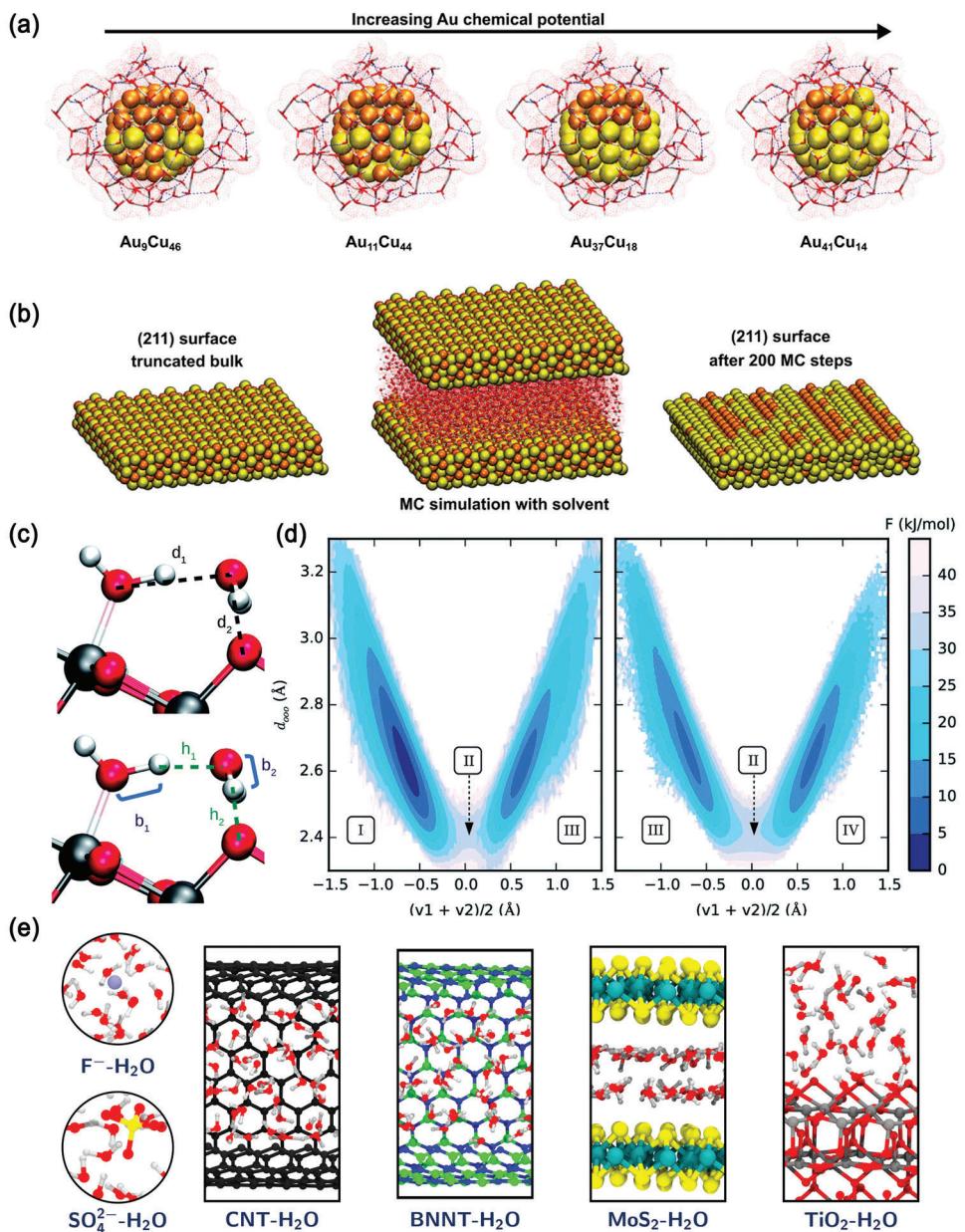


Figure 11. Solid–liquid interface studied by MLIPs. a) Water-solvated Au/Cu clusters (55 atoms) with increasing gold chemical potentials. b) MC simulation snapshots of a (211) surface slab (approximately 6500 atoms) with solvent. a,b) Reproduced with permission.^[186] Copyright 2014, American Chemical Society. c) Interatomic distances for collective variables (CVs) in proton transfer reactions simulation. d) Free energy landscape: water dissociation (left) and proton transport at the anatase (101)–water interface (right). c,d) Reproduced under the terms of the CC BY-NC 3.0 license.^[190] Copyright 2020, Royal Society of Chemistry. e) Complex aqueous systems. Reproduced under the terms of the CC BY 4.0 license.^[193] Copyright 2021, the Authors. Published by National Academy of Sciences.

as oxygen and hydrogen probability densities, angular orientations, and hydrogen bond dynamics. A key insight was that the behavior of water molecules, particularly within the first hydration layer, was intimately tied to the geometric structure of the underlying surfaces. This group further broadened their research to other solid–liquid interfaces, including those involving zinc oxide–water and lithium manganese oxide–water.^[188,189]

Moreover, the combination of MLIP and enhanced sampling techniques offers a promising approach to comprehending the

mechanisms of chemical reactions on solid–liquid interfaces. A notable example from 2020 is the work of Andrade et al., who employed a deep neural network (DNN) potential to map out the free energy surface (FES) of water dissociation and proton transport at the anatase interface in contact with liquid water (Figure 11c,d).^[190] Significantly, they didn't just utilize the enhanced sampling method to reveal the reaction mechanism in production MD; they also integrated it into the iterative DNN potential training process. To date, MLIPs have been

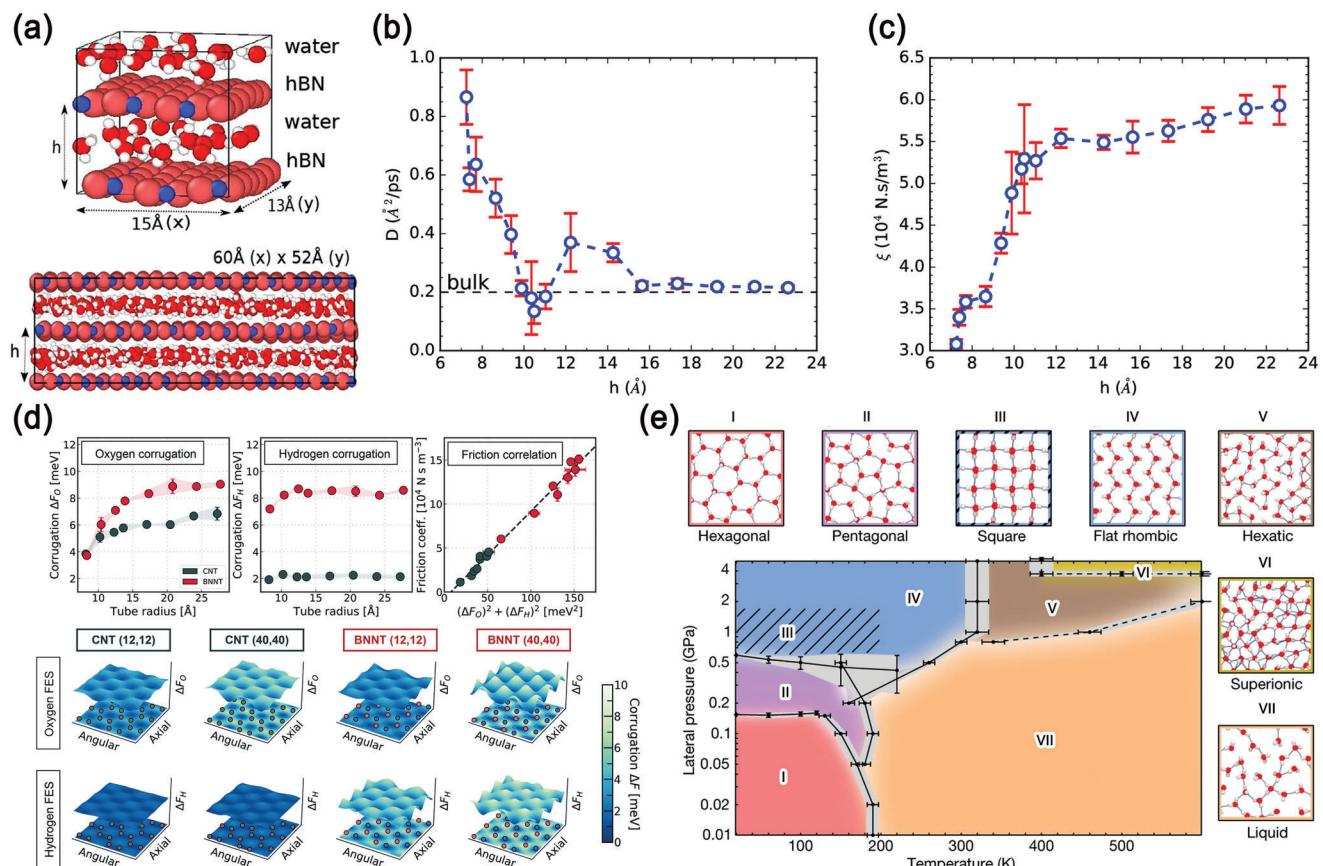


Figure 12. Water is confined in different nanostructures studied by MLIPs. a) Water molecules (red and white) between hBN membranes (blue and pink) showcasing typical DFT (top) and MD supercells (bottom) with a nanometric interlayer distance h . b) Water self-diffusion coefficients versus interlayer hBN distance h . c) Water friction coefficients against hBN interlayer distance h . a–c) Reproduced with permission.^[194] Copyright 2020, American Chemical Society. d) Correlation of friction and the FES in water within CNTs and BNNTs. Reproduced under the terms of the CC BY 4.0 license.^[195] Copyright 2022, the Authors. Published by American Chemical Society. e) Phase diagram of water confined in graphene channel. Reproduced with permission.^[197] Copyright 2022, Springer.

harnessed in various solid–liquid interface studies, encompassing metal/water interfaces,^[191] oxide/water interfaces,^[192] and even intricate aqueous systems (Figure 11e).^[193]

By designing appropriate datasets, MLIPs can be constructed for interfacial systems that empirical potentials struggle to describe. When used in conjunction with MD simulations, these models shed light on nanoscale surface/interface dynamics, especially the behavior of water flow in confined spaces. In 2020, Ghorbanfekr et al. utilized HDNNP-MD to study water permeation in 2D hBN nanocapillaries.^[194] Their findings suggest that at minimal interlayer distances, water permeation is chiefly driven by a monolayer possessing a superior diffusion coefficient and reduced water-wall friction (Figure 12a–c). A subsequent 2022 study by Thiemann et al. probed water transport in both carbon (CNT) and boron nitride nanotubes (BNNT) using c-NNP-MD.^[195] They observed a notable hydrodynamic slippage influenced by the radius of nanotube. Notably, water displayed about 5 times less friction on carbon than on boron nitride. Figure 12d illustrates the variations in the FES of hydrogens and oxygens across materials and curvature. As the radius increases, the oxygen FES becomes more corrugated, yet the favorable positions for oxygens remain consistent with those on flat surfaces. On carbon

surfaces, oxygen atoms predominantly occupy the hollow site of a carbon hexagon, while at the BN interface, they align near the boron atom. In contrast, hydrogen's FES shows minimal radius dependence, with a pronounced corrugation difference between BN and carbon interfaces. The findings underscore a direct correlation between the friction coefficient and the combined corrugation of both oxygen and hydrogen, emphasizing their intertwined contributions to overall frictional properties. Diffusion analyses indicate swift water transport on carbon stems from easy oxygen movement, while increased friction on boron nitride arises from distinct hydrogen–nitrogen interactions. This research not only set a benchmark for quantum accuracy in water transit through single-wall nanotubes but also enriched our understanding of how radius and material type can shape confined water dynamics, paving the way for advanced technological applications

Recently, for the same system, Andrade et al. utilized a different NNP to explore the infrared spectra (IR) of confined water in CNT.^[196] In this work, they employed the DP MLIP framework.^[52] Typically, conventional DP MLIPs are trained solely on a system's energy and force data within a local environment around atoms, lacking long-range correction; thus, they cannot yield information on charges and dipoles. In this

study, the dipole data for water molecules was derived from additional calculations of Wannier centers of atoms during both model training and production simulation. This dipole information not only enhanced the model's capacity to describe long-range interactions,^[82] but also enabled them to determine the dipole data for confined water molecules and subsequently calculate the infrared spectra. This work provides a reference for surface spectroscopy studies. Additionally, in 2022, Kapil et al. also reported an MLIP-based work of nanoconfined water in graphene channels.^[197] They found that water when confined at the nanoscale, exhibits diverse phase behaviors influenced by temperature and van der Waals pressure. Among the molecular phases observed were the hexatic and superionic phases, known for their high conductivity. Such results suggest that nanoconfinement offers a practical avenue to induce superionic behavior under feasible conditions, highlighting the value of MLIP in analyzing these systems (Figure 12e).

3.2. Molecular Systems

Unlike inorganic systems, molecular systems primarily comprise diverse organic molecules. This diversity results in a multitude of elements in the system and a complex configuration space due to molecular structures. Many MLIPs that excel with inorganic systems falter with molecular systems. The interfacial structures in molecular systems, especially those that involve interactions between organic molecules and inorganic surfaces, further complicate MLIP application.^[187,198] Yet, MLIPs still offer significant potential in molecular systems.^[114,199,200] This section emphasizes MLIPs' work related to molecular systems involving surfaces/interfaces.

3.2.1. MLIP for Molecular Porous Nanomaterials

Metal-organic frameworks (MOFs) are a prevalent class of 3D molecular nanoarchitectures, typically composed of inorganic secondary building units (SBUs) interconnected by organic linker molecules through covalent bonds. This modular structure allows for a vast array of combinations between SBUs and linkers, offering a considerable design and modification space tailored for various applications, such as catalysis, drug delivery, and separation engineering.^[201] The porous nature of MOFs provides them with a high specific surface area and the formation of confined channels. These characteristics enable interactions with guest species, leading to unique adsorption and catalytic selectivity. Much of the current theoretical research is centered on pore size regulation, structural stability, and the dynamics of guest-species interactions and their resulting properties. However, the intricate and varied structures of MOFs present challenges for traditional MD or MC methods, often due to the absence of specific force field parameters, which hinder our comprehension and design capabilities for MOFs. The advent of MLIPs is anticipated to overcome these challenges.^[202]

In 2019, Eckhoff and Behler developed an HDNNP model for IRMOF-1 (also named as MOF-5), utilizing molecular fragments centered on the system's geometrically inequivalent atoms for reference data (Figure 13a).^[203] This model could determine the

lattice constants of the MOF-5 structure with an error of just 0.1% when compared to DFT, and it precisely forecasted its negative thermal expansion behavior. Employing a fragment-based dataset construction strategy rooted in chemical environment equivalence, they further undertook MLIP-based investigations on IRMOF-10 and IRMOF-16 systems.^[204] More recently, several MLIP-based studies focusing on MOF–guest interactions have emerged. In 2022, Achar et al. crafted a DP potential for UiO-66, applying it to probe neon and xenon diffusion within the crystal (Figure 13b).^[205] Aiming for optimal efficiency, they introduced a hybrid potential simulation approach. Within this framework, Lennard-Jones (LJ) potentials modeled adsorbate–adsorbate interactions, the DP described adsorbent–adsorbent interactions, and LJ cross-interactions characterized adsorbent–adsorbate interactions. This hybrid potential approach offers valuable insights, especially for more intricate MOF–adsorbate systems. That same year, Yu et al. constructed a GNN potential to investigate the adsorption and migration tendencies of Pt_n clusters in MOF-808 (Figure 13c).^[206] Their findings highlighted that the prime adsorption site's stability is contingent upon the Pt_n cluster's size. Moreover, the potential for migration is significantly swayed by the adsorption energies associated with these Pt_n clusters.

Another class of porous material, similar to MOFs, is covalent organic frameworks (COFs). COFs are porous crystalline materials made up of conjugated organic molecules and linkers that are interconnected either through covalent or noncovalent bonds. The structure and characteristics of COFs can be modified and fine-tuned using chemical synthesis, offering extensive adaptability for their application development. A distinct advantage of COFs over other porous materials is their ability to adjust structures and properties through chemical synthesis, allowing for enhanced customization and refinement for targeted uses.^[207,208] However, both experimental and computational research has identified the presence of various defects and disorders within COF structures. These imperfections can compromise their crystallinity, stability, porosity, and conductivity. In 2023, Huang et al. employed an on-the-fly MLIP-based approach to investigate the structural disorder patterns of certain frameworks at 300 K (Figure 13d).^[209] Their research identified that an initially eclipsed stacking arrangement spontaneously shifts to a zigzag pattern, which subsequently reduces the crystal's free energy. This discovery implies that the zigzag configuration offers greater stability than the original eclipsed stacking pattern.

Furthermore, zeolites, another category of materials similar in nature, deserve mention. Although traditionally categorized as inorganic, zeolites also similarly encompass a diverse array of elements such as silicon, aluminum, sodium, calcium, potassium, magnesium, and more. These constituent elements play a pivotal role in determining the chemical and physical attributes of zeolites, subsequently influencing their adsorption and catalytic performances. Of late, MLIPs have been extensively employed to delve into the multifaceted properties of zeolites. Readers seeking an in-depth exploration are directed to pertinent literature reviews on the subject.^[210]

Lastly, the realm of 2D nanomaterials has seen noteworthy MLIP applications.^[211] In 2021, Mortazavi et al. demonstrated that MLIPs outperform traditional force fields in predicting the mechanical properties of complex interfaces like those in

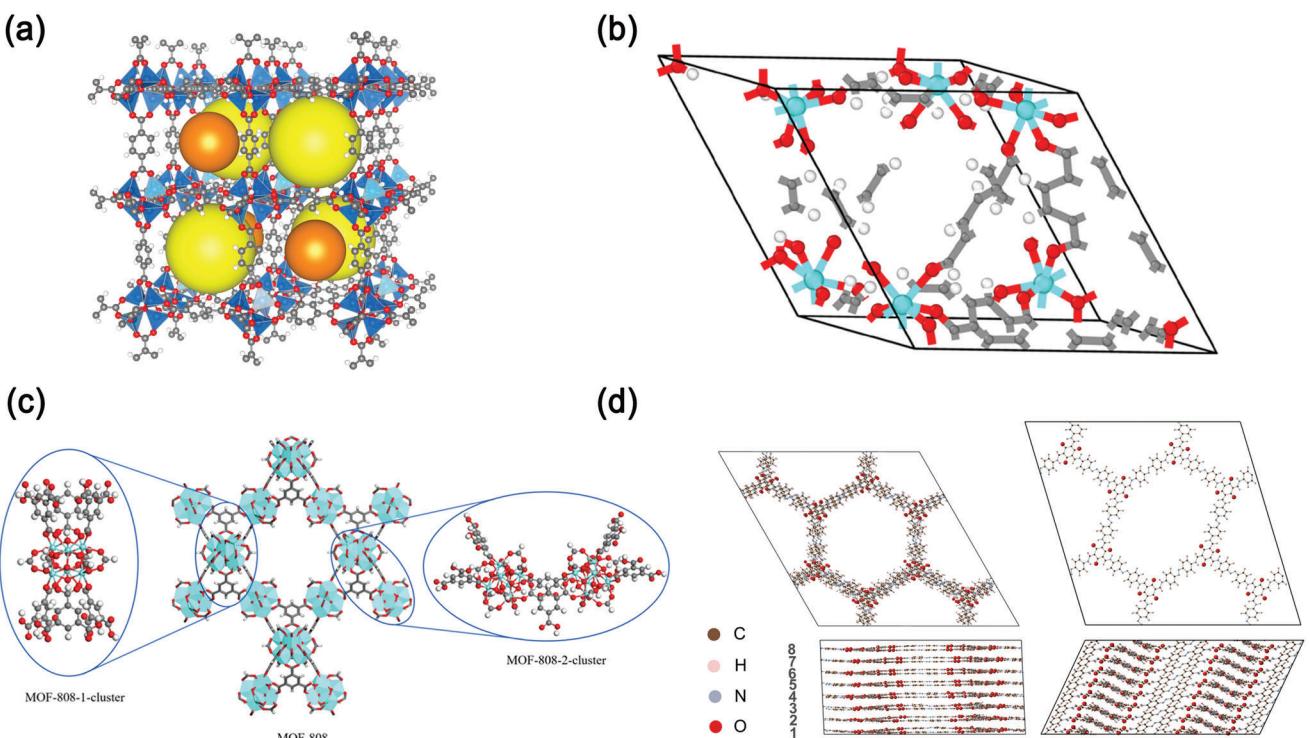


Figure 13. MOFs and COFs are studied by MLIPs. a) MOF-5. Reproduced with permission.^[203] Copyright 2019, American Chemical Society. b) Uio-66. Reproduced with permission.^[205] Copyright 2022, American Chemical Society. c) MOF-808. Reproduced with permission.^[206] Copyright 2022, American Chemical Society. d) Crystal structures eclipsed Tp-Azo (top left), inclined Tp-Azo (top right), eclipsed DAAQ-TFP (down left), and inclined DAAQ-TFP (down right). Reproduced under the terms of the CC BY 3.0 license.^[209] Copyright 2023, the Authors. Published by Royal Society of Chemistry.

graphene–borophene coplanar heterostructures.^[212] Their work underscores MLIPs' potential in uniting first-principles accuracy with multiscale modeling.

3.2.2. MLIP for Solvated Biomolecular Systems

The surface and interfacial properties of biomolecular systems play a pivotal role in understanding their interactions with external environments and other biomolecules. These interactions underpin many vital life processes within biological systems.^[213] For instance, the surfaces of proteins influence their interactions with other proteins, ligands, and cell membranes. These interactions, in turn, shape protein conformation, functionality, and stability. Similarly, the surfaces of cell membranes determine their interactions with the extracellular environment and other cells, affecting crucial processes such as signal transduction, nutrient absorption, and waste removal. Additionally, the interactions of DNA and RNA surfaces with proteins can influence pivotal life processes, including gene expression and protein synthesis. Consequently, in-depth theoretical and experimental research on biomolecular surfaces and interfaces is essential.

Although empirical force field-based all-atom MD is a widely used technique for studying nanoscale biomaterials, it demonstrates a notable force field dependence. For example, the force field selection can alter the results of peptide simulations concerning self-assembly in solution (Figure 14a),^[214,215] while the choice of a water model can impact the conformational integrity

of protein molecules (Figure 14b).^[216] To align simulation results more closely with experimental data, a comprehensive understanding of biomolecule–solvent interactions is necessary. The ongoing development of high-precision MLIPs, which are devoid of empirical parameters, is aiding this domain. While several MLIPs, such as PhysNet, ANI series, and GDML series, are available for simulating specific properties of solvated biomolecules,^[54,76,217] there are comparatively fewer studies focusing on MLIP-based simulations of biomolecular systems using explicit solvent models.

In a recent innovative study, Unke et al. presented a versatile method to develop accurate MLIPs for large-scale molecular simulations (GEMS). They trained their models on “bottom-up” and “top-down” molecular fragments of varying sizes, paving the way to investigate vital physicochemical properties of biomolecules (Figure 14c).^[218] GEMS was utilized to explore the dynamics of poly-alanine peptides and the 46-residue protein crambin in an aqueous environment. Through nanosecond-scale MD simulations, involving an explicit solvent model comprising over 25 000 atoms and nearly ab initio quality, the study found that the force and energy predictions of their model aligned more with density DFT calculations than the Amber force field (AmberFF). Additionally, within an explicit solvent, the structural motifs in protein trajectories derived from GEMS demonstrated greater flexibility than those from AmberFF (Figure 14d). The approach to constructing the MLIP dataset for complex systems in this research has profound implications for the evolution of MLIPs tailored for biomolecular interfaces in solutions. Moreover, the observed

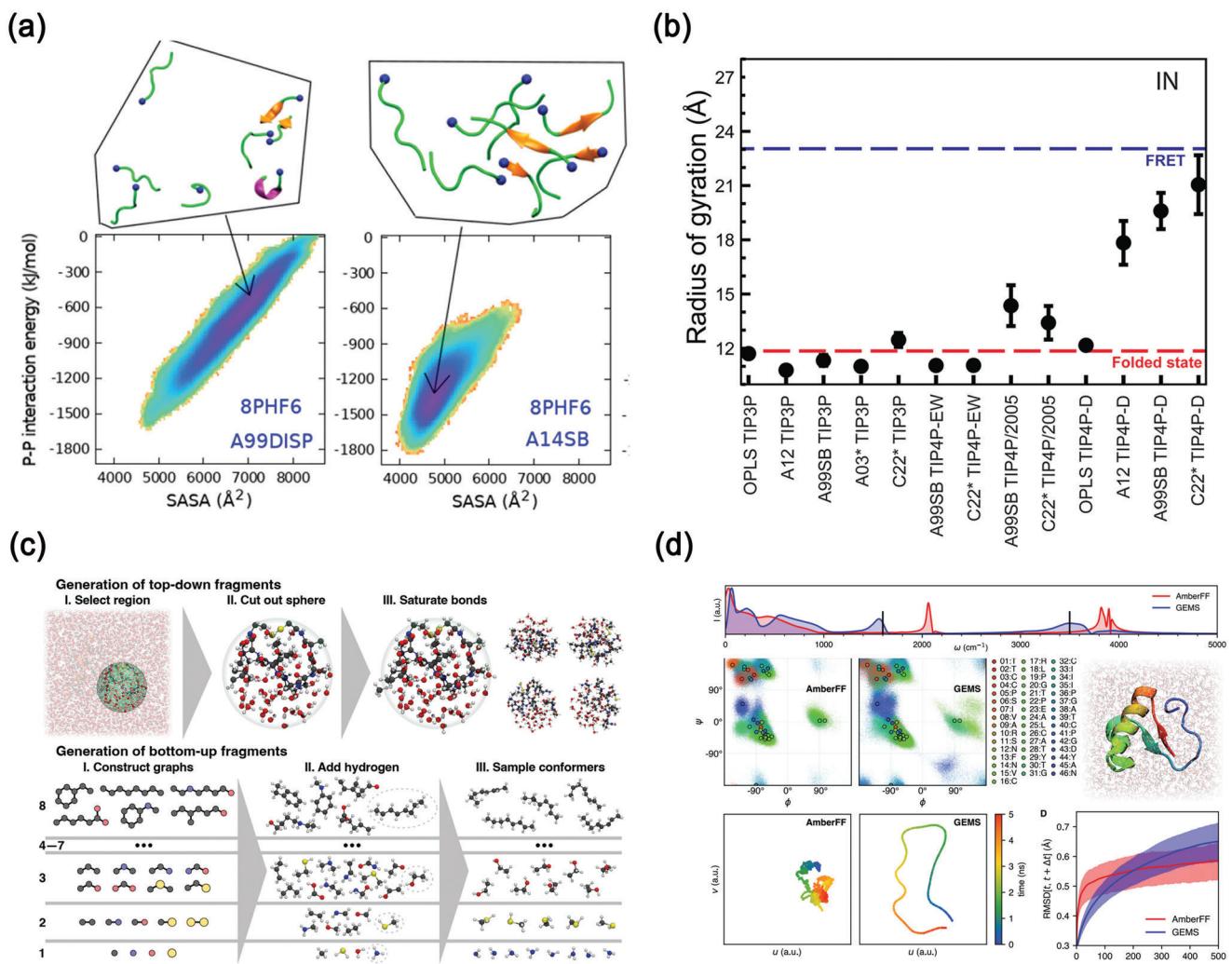


Figure 14. Simulations for biomolecules by empirical potentials and MLIPs. a) Free energy landscapes of 8PHF6 (as a function of surface accessible solvent area (SASA) and peptide–peptide (P–P) interaction energy). Reproduced with permission.^[215] Copyright 2021, American Chemical Society. b) Determination of the gyration radius for the inherently unfolded IN through empirical force field simulations. Reproduced with permission.^[216] Copyright 2015, American Chemical Society. c) Overview of GEMS dataset construction. d) MD simulation results of 46-residue protein crambin with explicit solvent model based on AmberFF and GEMS. c,d) Reproduced with permission.^[218] Copyright 2022, Cornell University.

enhanced flexibility in molecular structural motifs in MLIP-based simulations underscores the need for deeper, more accurate investigations of biomolecular systems.

3.2.3. ML/MM Method for Molecular Systems

The recent rise in methods merging MLIPs with traditional empirical potentials—dubbed the ML/MM methods—marks a significant step forward.^[114] The term “ML/MM” may not have a universally settled definition yet, but its underlying strategy resonates with the QM/MM paradigm, emphasizing system division to achieve computational resource optimization. For example, Lier et al. reported a buffer region neural network (BuRNN), which introduces a buffer region that experiences full electronic polarization by the inner QM region to minimize artifacts.^[219] ML handles the interactions between QM and the buffer zone,

ensuring the hybrid ML/MM system’s blend of computational efficiency and quantum-level precision. Their method’s efficacy was showcased in simulations of the hexa-aqua iron (III) complex within a water solvent. Over a 10 ns MD simulation, the smooth transitions of water molecules in and out of the buffer region became evident. The results revealed Fe^{3+} to oxygen distributions in coordinated water in alignment with QM/MM simulations with electrostatic embedding, closely resembling experimental observations. This exemplifies ML/MM’s potential in efficiently and accurately probing solution-based interfacial water structures and dynamics. Recently, integrating this technique with MLIP has been advantageous in comprehending intricate interfacial phenomena, spanning pore diffusion,^[205] solvent effects,^[220] and protein–ligand interactions,^[221] to name a few.^[222–225]

The prediction of protein–ligand binding free energy is a case in point. In 2020, Rowley et al. reported an ML/MM scheme with

a combination of NNP and MM calculations for this issue.^[226] Their architecture leveraged the ANI-cxx potential for ligand intramolecular interactions, while the remaining interactions were facilitated through a conventional force field. The cumulative system energy is derived from the MM potential surrounding the ligand, the ligand's ANI potential, and the ligand–protein interaction potential. This novel framework was deployed for predicting ligand binding stances and ascertaining the conformational free energies for a series of protein–ligand assemblages. Notably, the methodology could project the PES of some configurations external to the training set, with results more harmonized with quantum computations than those from the CHARMM general force field (CgenFF).^[227] While this case might not encapsulate the entire spectrum of ML/MM techniques across different platforms, its prowess in ligand-binding energy estimation heralds the wider application of ML/MM for upcoming interfacial research endeavors.

3.3. MLIP-Based Methodologies for the Study of Surfaces/Interfaces of Nanomaterials

In earlier sections, we delved into various MLIP applications, categorizing them based on the chemical systems they were applied. It is evident that most of the endeavors involving MLIP are tailored to specific projects or systems, with the choice of MLIP largely contingent on the researcher's familiarity with MLIPs and the peculiarities of the application context. Nonetheless, a handful of generalized methodologies and workflows founded on MLIPs for probing surfaces and interfaces do exist. This section highlights some of these methodological paradigms that cater to the study of nanomaterial surfaces and interfaces.

In the realm of interface and NP stability, MLIPs can be synergized with a plethora of structure search and optimization algorithms. This amalgamation facilitates the identification of both the most stable and metastable system configurations. Such a strategy proves instrumental not just in material stability assessments but also in curating training datasets. Recognizing the pivotal role of adsorption and catalysis at the surface/interface, parallel frameworks can be sculpted to address these specific challenges, thereby bolstering our capabilities in material screening and design.

Undoubtedly, MLIP-integrated MD/MC simulations offer a much larger length and time scale of simulations for nanomaterials. This is evident from studies like the melting simulations of AuNPs and proton transitions at the TiO₂ interface. Such an expanded framework empowers researchers to undertake extensive simulations for a myriad of surface/interface phenomena. By marrying MLIP-MD/MC with advanced sampling techniques, one can derive both dynamic and thermodynamic insights. Even on a reduced scale, these simulations, when coupled with clustering algorithms, prove invaluable in high-throughput materials screening. Furthermore, leveraging MLIPs within the Grand Canonical Monte Carlo (GCMC) framework, as seen in studies of metal surface oxidations and heterogeneous catalysis, has shown potential for scalability. Such blueprints can be adjusted and deployed for a plethora of intricate interfacial challenges, achieving near-quantum accuracy, thereby paving the way for more in-depth studies on surfaces.

In summary, Section 3 offers an overview of MLIP applications in the study of inorganic nanomaterials and molecular systems linked to the surfaces/interfaces of these nanomaterials. Employing highly accurate MLIPs proves advantageous for delving into the intricacies of nanomaterial surfaces and interfaces. To further enrich understanding, Table 4 is appended at the conclusion of this section, serving as a reference for works focused on nanomaterial surfaces and interfaces.

4. Challenges

In the preceding section, MLIPs have showcased their exceptional merit in a diverse range of applications pertaining to nanomaterial interfacial systems. In this section, we shall delve into several persisting challenges associated with this methodology. Should these obstacles be surmounted in future research, MLIPs hold the potential to assume a pivotal role in the simulation and modeling of not only nanomaterial interfacial systems but also other intricate systems.

4.1. More Efficient Data Acquiring Methods

As a data-driven methodology, the quality of data sources for MLIPs is critically important because it dictates the reliability of the resulting models. Well-established databases such as the QM series,^[96–100] and the ANI-1 series,^[101,103] which provide a convenient starting point for researchers constructing MLIPs. However, a considerable volume of data published in academic journals remains underexploited. This underutilization arises partly from the laborious process of literature search required to locate this data, and partly from the absence of comprehensive computational parameters, which complicates assessments of data accuracy. Consequently, there is an immediate need for a standardized data storage platform to serve a broader research community.

When investigating specialized systems, the development of custom datasets may be indispensable. Active learning techniques serve as a powerful strategy for this, as they assess the uncertainty of predicted configurations and selectively augment the training set with high-uncertainty samples. A commonly adopted method for uncertainty representation is the committee model,^[276] which gauges uncertainty through prediction consensus across an ensemble of trained models. However, this approach may sometimes underestimate the level of uncertainty.^[277] To enhance accuracy in uncertainty quantification, recent work has amalgamated Bayesian estimation,^[278] and marginal similarity methods,^[279] albeit at the expense of increased computational overhead. Therefore, finding an efficient yet accurate solution for uncertainty quantification remains an open challenge.

Another pivotal issue is maximizing data efficiency—that is, constructing reliable MLIPs with minimal data inputs. In this context, transfer learning and pre-trained models have shown promise. For instance, Zhang et al. developed a pre-trained DPA-1 model,^[280] based on attention mechanisms, for 56 elements, achieving models with high precision and a 90% reduction in data volume. Additionally, equivariant GNN-based MLIP have demonstrated the capability to drastically reduce data requirements for model training without sacrificing model accuracy.^[89]

Table 4. Overview of applications of MLIPs for surface/interface of nanomaterials.

Surface/interface	System	Refs.
Inorganic material surfaces	Silicon	[110,132,133,228]
	Carbon	[106,137–139,229–231]
	Phosphorus	[232]
	Metal	[146,233–236]
	Alloy	[237–239]
	Oxide	[141,240,241]
	Metal	[153,154,160]
Nanoparticles ^a	Alloy clusters	[238]
	Other inorganic clusters	[242]
	Supported nanoparticles	[243–245]
	Solid–gas	[179,245,246]
	Liquid–gas	[247]
	Solid–liquid	[186–191,193,248–250]
Gas environment	Nanoconfined environment	[194,195,197,251]
Solution environment	Solid–solid interface	[252–254]
Other inorganic systems	MOF	[203–206,255]
Molecular system surfaces/interfaces	COF	[209]
	Zeolite	[256–264]
	2d materials	[212,265–270]
	molecular crystal	[271–274]
	biomolecules	[221,223,275]

^{a)} (Note: Although clusters are typically smaller than nanoparticles and may consist of only a few atoms or molecules, for classification purposes, clusters are categorized as nanoparticles in this context.)

4.2. More Improved and Robust Models

While current MLIPs have been successful across various applications, several critical issues remain unresolved. In the context of the Behler-Parrinello framework-based MLIPs, descriptor construction is a pivotal concern. Despite the array of available descriptors, none have been definitively proven to be complete,^[281] meaning that any two structures that can be superimposed after transformation have the same descriptors and vice versa. Another challenge arises in capturing the intricacies of potential energy surfaces, particularly near certain configurations where the surfaces may exhibit non-smooth behavior.^[115] Many MLIPs, operating on the assumption of smoothness, fall short of accurately modeling these anomalies. Additionally, the methodological concerns around long-range interactions and non-local charge transfer, discussed in the method section, remain without a universally accepted solution strategy.

Complicating matters further is the escalating computational burden associated with the ever-increasing complexity of descriptors and model structures. While this complexity aims to improve the accuracy of MLIPs, it simultaneously inhibits their applicability in modeling intricate systems. The future trajectory of MLIP development could benefit from integrating deeper physical insights into the models, so as to strike a more balanced trade-off between computational efficiency and predictive accuracy.

4.3. More Tools for Developing MLIP of Surface/Interface

While MLIPs have shown promise in various areas, it should also be noted that currently most MLIPs can only address some issues of simple surface/interface systems and are not friendly to multi-element, multi-local environment surface/interface scenarios. In addition to the commonly encountered challenges in surface/interface investigations, the complexities arising from the inherent scale effects of nanomaterials warrant consideration, such as the multi-metastable state, and non-periodicity of the model.^[282,283] Moreover, the application of MLIPs to chemistry and biomolecular systems remains largely unexplored, particularly at complex abiotic–biotic interfaces. This research gap can be attributed to multiple factors. First, the absence of a universally applicable and reliable phase space sampling strategy makes system exploration inefficient. Second, even when comprehensive ab initio quality datasets are used for model training, assessing the model's ultimate reliability and accuracy poses challenges that often necessitate experimental validation. Customizing the selection of relevant observable quantities and evaluation metrics to fit a specific problem is a nuanced task, complicating matters further.

Therefore, fostering collaborations between experimental and theoretical researchers is essential for the development of more transferable, scalable, and reliable MLIP-based tools tailored for studying the surface and interface properties of nanomaterials.

5. Concluding Remarks

In the field of MD simulations, the emergence of MLIPs has been nothing short of revolutionary, comparable to the invention of the steam engine in industrial history. Boasting theoretically limitless fitting capabilities and unprecedented accuracy in representing chemical interactions, MLIPs have solved a long-standing conundrum: the lack of precise force fields for simulating surficial and interfacial phenomena of nanomaterials. Their remarkable performance in predicting physicochemical attributes, replicating experimental spectra, and characterizing surface and interfacial reactions has broken new ground, significantly expanding the scope of computational studies in nanomaterial surficial/interfacial systems.

Though current MLIP models are not without challenges, their impact has ignited an interdisciplinary fervor, bridging the gap between MD and ML communities. The number of researchers committed to the advancement of MLIPs is steadily growing, and interdisciplinary endeavors are increasingly fueling the field's progression. As such, there's a justified anticipation that existing obstacles will be overcome in the near future, paving the way for even more accurate and broadly applicable models.

Furthermore, the rise of automated laboratories is progressively reducing the costs associated with acquiring experimental data. Coupled with the continual improvements in MLIP accuracy, this trend foretells a diminishing margin of error and subsequent validation of MLIP universality. This symbiotic relationship will not only fine-tune theoretical models but also redefine the landscape of scientific investigation. We foresee a future where experimental research is not limited to physical laboratories but extends seamlessly into the realm of *in silico* simulations. In summary, MLIPs are not merely an emerging technology for simulating nanomaterial interfaces; they are a lighthouse guiding us toward an unprecedented epoch of scientific exploration.

Acknowledgements

K.W. and J.H. contributed equally to this work. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB36000000), the National Key R&D Program of China (Grant No. 2022YFA1203200), the Natural Science Foundation of Beijing (2222085, 1202023, 2194092), the National Natural Science Foundation of China (11672079, 12072082, 12125202). The authors thank the Hefei Advanced Computing Center for the computational resources.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

force field, machine learning, molecular dynamics, physical chemistry, surface and interface

Received: June 15, 2023

Revised: August 24, 2023

Published online:

[1] C. Xie, Z. Niu, D. Kim, M. Li, P. Yang, *Chem. Rev.* **2020**, 120, 1184.

- [2] L. Liu, A. Corma, *Chem. Rev.* **2018**, 118, 4981.
- [3] E. Llobet, *Sens. Actuators, B* **2013**, 179, 32.
- [4] S.-J. Choi, I.-D. Kim, *Electron. Mater. Lett.* **2018**, 14, 221.
- [5] J. Hecht, *Nature* **2022**, 608, S2.
- [6] C. Dong, S. Chilstedt, D. Chen, in *2009 Design, Automation & Test in Europe Conf. & Exhibition* (Eds: G. D. Micheli, W. Mueller), IEEE, Piscataway, NJ **2009**, 442.
- [7] C. Maffeo, S. Bhattacharya, J. Yoo, D. Wells, A. Aksimentiev, *Chem. Rev.* **2012**, 112, 6250.
- [8] H. Zhang, X. Li, J. Hou, L. Jiang, H. Wang, *Chem. Soc. Rev.* **2022**, 51, 2224.
- [9] F. J. Giessibl, *Rev. Mod. Phys.* **2003**, 75, 949.
- [10] Y. Tian, J. Hong, D. Cao, S. You, Y. Song, B. Cheng, Z. Wang, D. Guan, X. Liu, Z. Zhao, X. Z. Li, L. M. Xu, J. Guo, J. Chen, E. G. Wang, Y. Jiang, *Science* **2022**, 377, 315.
- [11] R. Ma, D. Cao, C. Zhu, Y. Tian, J. Peng, J. Guo, J. Chen, X. Z. Li, J. S. Francisco, X. C. Zeng, L. M. Xu, E. G. Wang, Y. Jiang, *Nature* **2020**, 577, 60.
- [12] J. Peng, D. Cao, Z. He, J. Guo, P. Hapala, R. Ma, B. Cheng, J. Chen, W. J. Xie, X. Z. Li, P. Jelinek, L. M. Xu, Y. Q. Gao, E. G. Wang, Y. Jiang, *Nature* **2018**, 557, 701.
- [13] A. Kumar, J. N. Baker, P. C. Bowes, M. J. Cabral, S. Zhang, E. C. Dickey, D. L. Irving, J. M. LeBeau, *Nat. Mater.* **2021**, 20, 62.
- [14] C. T. Toh, H. Zhang, J. Lin, A. S. Mayorov, Y. P. Wang, C. M. Orofeo, D. B. Ferry, H. Andersen, N. Kakenov, Z. Guo, I. H. Abidi, H. Sims, K. Suenaga, S. T. Pantelides, B. Ozyilmaz, *Nature* **2020**, 577, 199.
- [15] J. Reguera, J. Langer, D. Jimenez de Aberasturi, L. M. Liz-Marzan, *Chem. Soc. Rev.* **2017**, 46, 3866.
- [16] N. C. Lindquist, C. D. L. de Albuquerque, R. G. Sobral-Filho, I. Paci, A. G. Brolo, *Nat. Nanotechnol.* **2019**, 14, 981.
- [17] Y. H. Wang, S. Zheng, W. M. Yang, R. Y. Zhou, Q. F. He, P. Radjenovic, J. C. Dong, S. Li, J. Zheng, Z. L. Yang, G. Attard, F. Pan, Z. Q. Tian, J. F. Li, *Nature* **2021**, 600, 81.
- [18] C. S. Tian, Y. R. Shen, *Surf. Sci. Rep.* **2014**, 69, 105.
- [19] C. D. Bain, *J. Chem. Soc., Faraday Trans.* **1995**, 91, 1281.
- [20] W. C. D. Cheong, L. C. Zhang, *Nanotechnology* **2000**, 11, 173.
- [21] L. Chen, J. L. Fan, H. R. Gong, *J. Nanopart. Res.* **2017**, 19.
- [22] Z. Rao, S. Wang, F. Peng, *Int. J. Heat Mass Transfer* **2013**, 66, 575.
- [23] Y. Shibuta, T. Suzuki, *J. Chem. Phys.* **2008**, 129, 144102.
- [24] M. Abbaspour, H. Akbarzadeh, S. Salemi, E. Jalalitalab, *J. Mol. Liq.* **2018**, 250, 26.
- [25] T. A. Ho, A. Striolo, *J. Chem. Phys.* **2013**, 138, 054117.
- [26] H. Tang, D. Liu, Y. Zhao, X. Yang, J. Lu, F. Cui, *J. Phys. Chem. C* **2015**, 119, 26712.
- [27] M. Foroutan, S. M. Fatemi, F. Esmaeilian, *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2017**, 40, 19.
- [28] X. Zhao, *J. Phys. Chem. C* **2011**, 115, 6181.
- [29] M. Fuentes-Cabrera, B. H. Rhodes, J. D. Fowlkes, A. Lopez-Benzanilla, H. Terrones, M. L. Simpson, P. D. Rack, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2011**, 83, 041603.
- [30] C. L. Cheng, G. J. Zhao, *Nanoscale* **2012**, 4, 2301.
- [31] A. Berg, C. Peter, K. Johnston, *J. Chem. Theory Comput.* **2017**, 13, 5610.
- [32] A. E. Nel, L. Madler, D. Velegol, T. Xia, E. M. Hoek, P. Somasundaran, F. Klaessig, V. Castranova, M. Thompson, *Nat. Mater.* **2009**, 8, 543.
- [33] J. Liu, J. Zhang, *Chem. Rev.* **2020**, 120, 2123.
- [34] T. Rajh, Z. Saponjic, J. Liu, N. M. Dimitrijevic, N. F. Scherer, M. Vega-Arroyo, P. Zapol, L. A. Curtiss, M. C. Thurnauer, *Nano Lett.* **2004**, 4, 1017.
- [35] N. S. Lewis, *Annu. Rev. Phys. Chem.* **1991**, 42, 543.
- [36] G. D. Goh, J. M. Lee, G. L. Goh, X. Huang, S. Lee, W. Y. Yeong, *Tissue Eng., Part A* **2023**, 29, 20.
- [37] U. Delli, S. Chang, *Procedia Manuf.* **2018**, 26, 865.
- [38] P. Bannigan, M. Aldeghi, Z. Bao, F. Hase, A. Aspuru-Guzik, C. Allen, *Adv. Drug Delivery Rev.* **2021**, 175, 113806.

- [39] A. K. Sinha, G. L. Goh, W. Y. Yeong, Y. Cai, *Adv. Mater. Interfaces* **2022**, *9*, 2200621.
- [40] M. Farhan Khan, A. Alarm, M. Ateeb Siddiqui, M. Saad Alam, Y. Rafat, N. Salik, I. Al-Saidan, *Mater. Today Proc.* **2021**, *42*, 521.
- [41] K. P. Seng, L.-M. Ang, E. Peter, A. Mmonyi, *Electronics* **2023**, *12*, 1509.
- [42] K. Nguyen-Cong, J. T. Willman, S. G. Moore, A. B. Belonoshko, R. Gayatri, E. Weinberg, M. A. Wood, A. P. Thompson, I. I. Oleynik, in *Proc. of the Int. Conf. for High Performance Computing, Networking, Storage and Analysis* (Eds: B. R. de Supinski, M. Hall, T. Gamblin), ACM, Washington, D.C **2021**, 1.
- [43] W. Jia, H. Wang, M. Chen, D. Lu, L. Lin, R. Car, E. Weinan, L. Zhang, in *SC20: Int. Conf. for High Performance Computing, Networking, Storage and Analysis* (Eds: C. Cucci, I. Qualters, W. Kramer), IEEE, Piscataway, NJ **2020**, 1.
- [44] H. Li, Y. Jiao, K. Davey, S. Z. Qiao, *Angew. Chem., Int. Ed. Engl.* **2023**, *62*, 202216383.
- [45] G. Cybenko, *Math. Control Signals Syst.* **1989**, *2*, 303.
- [46] K. Hornik, M. Stinchcombe, H. White, *Neural Networks* **1989**, *2*, 359.
- [47] K. Hornik, *Neural Networks* **1991**, *4*, 251.
- [48] J. Behler, M. Parrinello, *Phys. Rev. Lett.* **2007**, *98*, 146401.
- [49] S. D. Huang, C. Shang, P. L. Kang, Z. P. Liu, *Chem. Sci.* **2018**, *9*, 8644.
- [50] P.-l. Kang, C. Shang, Z.-p. Liu, *Chin. J. Chem. Phys.* **2021**, *34*, 583.
- [51] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 4436.
- [52] L. Zhang, J. Han, H. Wang, R. Car, E. Weinan, *Phys. Rev. Lett.* **2018**, *120*, 143001.
- [53] K. Schütt, P.-J. Kindermans, H. E. Saucedo Felix, S. Chmiela, A. Tkatchenko, K.-R. Müller, *Adv. Neural Inf. Process Syst.* **2017**, *30*, 992.
- [54] O. T. Unke, M. Meuwly, *J. Chem. Theory Comput.* **2019**, *15*, 3678.
- [55] O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schutt, H. E. Saucedo, K. R. Muller, *Nat. Commun.* **2021**, *12*, 7273.
- [56] T. Bereau, R. A. DiStasio Jr, A. Tkatchenko, O. A. von Lilienfeld, *J. Chem. Phys.* **2018**, *148*, 241706.
- [57] S. Chmiela, H. E. Saucedo, I. Poltavsky, K.-R. Müller, A. Tkatchenko, *Comput. Phys. Commun.* **2019**, *240*, 38.
- [58] A. S. Christensen, L. A. Bratholm, F. A. Faber, O. Anatole von Lilienfeld, *J. Chem. Phys.* **2020**, *152*, 044107.
- [59] A. P. Bartok, M. C. Payne, R. Kondor, G. Csanyi, *Phys. Rev. Lett.* **2010**, *104*, 136403.
- [60] A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, *87*, 184115.
- [61] S. Chmiela, A. Tkatchenko, H. E. Saucedo, I. Poltavsky, K. T. Schutt, K. R. Muller, *Sci. Adv.* **2017**, *3*, 1603015.
- [62] B. Huang, O. A. von Lilienfeld, *Nat. Chem.* **2020**, *12*, 945.
- [63] P. O. Dral, *J. Comput. Chem.* **2019**, *40*, 2339.
- [64] N. Artrith, T. Morawietz, J. Behler, *Phys. Rev. B* **2011**, *83*, 153101.
- [65] B. G. Sumpter, D. W. Noid, *Chem. Phys. Lett.* **1992**, *192*, 455.
- [66] T. B. Blank, S. D. Brown, A. W. Calhoun, D. J. Doren, *J. Chem. Phys.* **1995**, *103*, 4129.
- [67] D. F. R. Brown, M. N. Gibbs, D. C. Clary, *J. Chem. Phys.* **1996**, *105*, 7597.
- [68] F. V. Prudente, P. H. Acioli, J. J. S. Neto, *J. Chem. Phys.* **1998**, *109*, 8801.
- [69] F. V. Prudente, J. J. Soares Neto, *Chem. Phys. Lett.* **1998**, *287*, 585.
- [70] S. Lorenz, A. Groß, M. Scheffler, *Chem. Phys. Lett.* **2004**, *395*, 210.
- [71] R. A. Olsen, H. F. Busnengo, A. Salin, M. F. Somers, G. J. Kroes, E. J. Baerends, *J. Chem. Phys.* **2002**, *116*, 3841.
- [72] J. Behler, *Chem. Rev.* **2021**, *121*, 10037.
- [73] N. Artrith, T. Morawietz, J. Behler, *Phys. Rev. B* **2012**, *86*, 153101.
- [74] J. Behler, R. Martonak, D. Donadio, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*, 185501.
- [75] G. C. Sosso, G. Miceli, S. Caravati, J. Behler, M. Bernasconi, *Phys. Rev. B* **2012**, *85*, 174103.
- [76] J. S. Smith, O. Isayev, A. E. Roitberg, *Chem. Sci.* **2017**, *8*, 3192.
- [77] S. D. Huang, C. Shang, P. L. Kang, X. J. Zhang, Z. P. Liu, *WIREs Comput. Mol. Sci.* **2019**, *9*, e1415.
- [78] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, P. Marquetand, *J. Chem. Phys.* **2018**, *148*, 241709.
- [79] Z. Deng, C. Chen, X.-G. Li, S. P. Ong, *npj Comput. Mater.* **2019**, *5*, 75.
- [80] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, G. J. Tucker, *J. Comput. Phys.* **2015**, *285*, 316.
- [81] P. P. Ewald, *Ann. Phys.* **1921**, *369*, 253.
- [82] L. Zhang, H. Wang, M. C. Muniz, A. Z. Panagiotopoulos, R. Car, E. Weinan, *J. Chem. Phys.* **2022**, *156*, 124107.
- [83] T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, *Nat. Commun.* **2021**, *12*, 398.
- [84] S. A. Ghasemi, A. Hofstetter, S. Saha, S. Goedecker, *Phys. Rev. B* **2015**, *92*, 045131.
- [85] A. K. Rappe, W. A. Goddard III, *J. Phys. Chem.* **1991**, *95*, 3358.
- [86] A. C. T. van Duin, S. Dasgupta, F. Lorant, W. A. Goddard, *J. Phys. Chem. A* **2001**, *105*, 9396.
- [87] T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, *Acc. Chem. Res.* **2021**, *54*, 808.
- [88] J. Hermann, D. Alfe, A. Tkatchenko, *Nat. Commun.* **2017**, *8*, 14052.
- [89] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, B. Kozinsky, *Nat. Commun.* **2022**, *13*, 2453.
- [90] S. De, A. P. Bartok, G. Csanyi, M. Ceriotti, *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754.
- [91] R. Jinnouchi, F. Karsai, G. Kresse, *Phys. Rev. B* **2019**, *100*, 014105.
- [92] G. Kresse, J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558.
- [93] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, B. Kozinsky, *npj Comput. Mater.* **2020**, *6*, 20.
- [94] A. Grisafi, M. Ceriotti, *J. Chem. Phys.* **2019**, *151*, 204105.
- [95] A. Grisafi, J. Nigam, M. Ceriotti, *Chem. Sci.* **2021**, *12*, 2078.
- [96] M. Rupp, A. Tkatchenko, K. R. Muller, O. A. von Lilienfeld, *Phys. Rev. Lett.* **2012**, *108*, 058301.
- [97] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, O. Anatole von Lilienfeld, *New J. Phys.* **2013**, *15*, 095003.
- [98] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *Sci. Data* **2014**, *1*, 140022.
- [99] R. Ramakrishnan, M. Hartmann, E. Tapavicza, O. A. von Lilienfeld, *J. Chem. Phys.* **2015**, *143*, 084111.
- [100] J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio Jr, A. Tkatchenko, *Sci. Data* **2021**, *8*, 43.
- [101] J. S. Smith, O. Isayev, A. E. Roitberg, *Sci. Data* **2017**, *4*, 170193.
- [102] A. S. Christensen, O. A. von Lilienfeld, *Mach. Learn.: Sci. Technol.* **2020**, *1*, 045018.
- [103] J. S. Smith, R. Zubatyuk, B. Nebgen, N. Lubbers, K. Barros, A. E. Roitberg, O. Isayev, S. Tretiak, *Sci. Data* **2020**, *7*, 134.
- [104] S. Chmiela, V. Vassilev-Galindo, O. T. Unke, A. Kabylda, H. E. Saucedo, A. Tkatchenko, K. R. Muller, *Sci. Adv.* **2023**, *9*, eadff0873.
- [105] T. Barnard, S. Tseng, J. P. Darby, A. P. Bartók, A. Broo, G. C. Sosso, *Mol. Syst. Des. Eng.* **2023**, *8*, 300.
- [106] V. L. Deringer, G. Csányi, *Phys. Rev. B* **2017**, *95*, 094203.
- [107] D. P. Kovacs, C. V. Oord, J. Kucera, A. E. A. Allen, D. J. Cole, C. Ortner, G. Csanyi, *J. Chem. Theory Comput.* **2021**, *17*, 7696.
- [108] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, B. Kozinsky, *Nat. Commun.* **2023**, *14*, 579.
- [109] J. Gasteiger, C. Yeshwanth, J. Groß, S. Günnemann, in *Advances in Neural Information Processing Systems* (Eds: M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, J. W. Vaughan), Curran Associates, virtual, **2021**, 15421.
- [110] A. P. Bartok, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csanyi, M. Ceriotti, *Sci. Adv.* **2017**, *3*, 1701816.

- [111] K. T. Schutt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko, K. R. Muller, *J. Chem. Phys.* **2018**, *148*, 241722.
- [112] K. Schütt, O. Unke, M. Gastegger, in *proc. of the 38th Int. Conf. on Machine Learning* (Eds: D. Hsu, Y. Liang), ICML, virtual, **2021**, 9377.
- [113] J. D. Morrow, J. L. A. Gardner, V. L. Deringer, *J. Chem. Phys.* **2023**, *158*, 121501.
- [114] S. Kaser, L. I. Vazquez-Salazar, M. Meuwly, K. Topfer, *Digital Discovery* **2023**, *2*, 28.
- [115] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schutt, A. Tkatchenko, K. R. Muller, *Chem. Rev.* **2021**, *121*, 10142.
- [116] C. Chen, S. P. Ong, *Nat. Comput. Sci.* **2022**, *2*, 718.
- [117] X.-G. Li, C. Hu, C. Chen, Z. Deng, J. Luo, S. P. Ong, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2018**, *98*, 094104.
- [118] C. Lu, C. Wu, D. Ghoreishi, W. Chen, L. Wang, W. Damm, G. A. Ross, M. K. Dahlgren, E. Russell, C. D. Von Bargen, R. Abel, R. A. Friesner, E. D. Harder, *J. Chem. Theory Comput.* **2021**, *17*, 4291.
- [119] F. Bianchini, J. R. Kermode, A. De Vita, *Modell. Simul. Mater. Sci. Eng.* **2016**, *24*, 045012.
- [120] Z. Wei, C. Zhang, Y. Kan, Y. Zhang, Y. Chen, *Comput. Mater. Sci.* **2022**, *202*, 111012.
- [121] L. Zhang, H. Wang, R. Car, E. Weinan, *Phys. Rev. Lett.* **2021**, *126*, 236001.
- [122] D. Folmsbee, G. Hutchison, *Int. J. Quantum Chem.* **2020**, *121*, 26381.
- [123] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csanyi, A. V. Shapeev, A. P. Thompson, M. A. Wood, S. P. Ong, *J. Phys. Chem. A* **2020**, *124*, 731.
- [124] X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli, T. Jaakkola, *arXiv preprint arXiv:2210.07237* **2022**.
- [125] L. Fiedler, N. A. Modine, S. Schmerler, D. J. Vogel, G. A. Popoola, A. P. Thompson, S. Rajamanickam, A. Cangi, *npj Comput. Mater.* **2023**, *9*, 115.
- [126] S. Mitchell, R. Qin, N. Zheng, J. Perez-Ramirez, *Nat. Nanotechnol.* **2021**, *16*, 129.
- [127] Y. Guan, D. Chaffart, G. Liu, Z. Tan, D. Zhang, Y. Wang, J. Li, L. Ricardez-Sandoval, *Chem. Eng. Sci.* **2022**, *248*, 117224.
- [128] P. Friederich, F. Hase, J. Proppe, A. Aspuru-Guzik, *Nat. Mater.* **2021**, *20*, 750.
- [129] H. Chen, W. Zhang, M. Li, G. He, X. Guo, *Chem. Rev.* **2020**, *120*, 2879.
- [130] G. Binnig, H. Rohrer, C. Gerber, E. Weibel, *Phys. Rev. Lett.* **1983**, *50*, 120.
- [131] K. D. Brommer, M. Needels, B. Larson, J. D. Joannopoulos, *Phys. Rev. Lett.* **1992**, *68*, 1355.
- [132] A. P. Bartók, J. Kermode, N. Bernstein, G. Csányi, *Phys. Rev.: X* **2018**, *8*, 041048.
- [133] L. Hu, B. Huang, F. Liu, *Phys. Rev. Lett.* **2021**, *126*, 176101.
- [134] G. Henkelman, B. P. Uberuaga, H. Jónsson, *J. Chem. Phys.* **2000**, *113*, 9901.
- [135] I. S. Kim, C. E. Shim, S. W. Kim, C. S. Lee, J. Kwon, K. E. Byun, U. Jeong, *Adv. Mater.* **2022**, 2204912.
- [136] V. L. Deringer, G. Csányi, *Phys. Rev. B* **2017**, *95*, 094203.
- [137] V. L. Deringer, M. A. Caro, R. Jana, A. Aarva, S. R. Elliott, T. Laurila, G. Csányi, L. Pastewka, *Chem. Mater.* **2018**, *30*, 7438.
- [138] M. A. Caro, A. Aarva, V. L. Deringer, G. Csányi, T. Laurila, *Chem. Mater.* **2018**, *30*, 7446.
- [139] A. Aarva, V. L. Deringer, S. Sainio, T. Laurila, M. A. Caro, *Chem. Mater.* **2019**, *31*, 9243.
- [140] M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, G. Csányi, *Phys. Rev. Lett.* **2018**, *120*, 166101.
- [141] J. Timmermann, F. Kraushofer, N. Resch, P. Li, Y. Wang, Z. Mao, M. Riva, Y. Lee, C. Staacke, M. Schmid, C. Scheurer, G. S. Parkinson, U. Diebold, K. Reuter, *Phys. Rev. Lett.* **2020**, *125*, 206101.
- [142] G. P. P. Pun, V. Yamakov, J. Hickman, E. H. Glaessgen, Y. Mishin, *Phys. Rev. Mater.* **2020**, *4*, 113807.
- [143] M.-C. Desjonquieres, D. Spanjaard, *Concepts in Surface Physics*, 2nd ed., Vol. 30, Springer Science & Business Media, Berlin **1996**.
- [144] M. Bernasconi, E. Tosatti, *Surf. Sci. Rep.* **1993**, *17*, 363.
- [145] M. S. Spencer, *Nature* **1986**, *323*, 685.
- [146] M. Cioni, D. Polino, D. Rapetti, L. Pesce, M. Delle Piane, G. M. Pavan, *J. Chem. Phys.* **2023**, *158*, 124701.
- [147] P. Wang, Y. Shao, H. Wang, W. Yang, *Extreme Mech. Lett.* **2018**, *24*, 1.
- [148] G. V. Huerta, K. Hisama, Y. Nanba, M. Koyama, *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier, Amsterdam **2023**.
- [149] L. Xu, H. W. Liang, Y. Yang, S. H. Yu, *Chem. Rev.* **2018**, *118*, 3209.
- [150] L. Jauffred, A. Samadi, H. Klingberg, P. M. Bendix, L. B. Oddershede, *Chem. Rev.* **2019**, *119*, 8087.
- [151] E. C. Dreaden, A. M. Alkilany, X. Huang, C. J. Murphy, M. A. El-Sayed, *Chem. Soc. Rev.* **2012**, *41*, 2740.
- [152] Y. Zhang, J. Zhang, B. Zhang, R. Si, B. Han, F. Hong, Y. Niu, L. Sun, L. Li, B. Qiao, K. Sun, J. Huang, M. Haruta, *Nat. Commun.* **2020**, *11*, 558.
- [153] C. Zeni, K. Rossi, T. Pavloudis, J. Kiouseoglou, S. de Gironcoli, R. E. Palmer, F. Baletto, *Nat. Commun.* **2021**, *12*, 6056.
- [154] S. Chiriki, S. Jindal, S. S. Bulusu, *J. Chem. Phys.* **2017**, *146*, 084314.
- [155] S. Bulusu, X. C. Zeng, *J. Chem. Phys.* **2006**, *125*, 154303.
- [156] C. D. Dong, X. G. Gong, *J. Chem. Phys.* **2010**, *132*, 104301.
- [157] N. Shao, W. Huang, Y. Gao, L. M. Wang, X. Li, L. S. Wang, X. C. Zeng, *J. Am. Chem. Soc.* **2010**, *132*, 6596.
- [158] S. Chiriki, S. S. Bulusu, *Chem. Phys. Lett.* **2016**, *652*, 130.
- [159] A. Aguado, O. Kostko, *J. Chem. Phys.* **2011**, *134*, 164304.
- [160] S. Jindal, S. S. Bulusu, *J. Chem. Phys.* **2020**, *152*, 154302.
- [161] C. Zeni, K. Rossi, A. Glielmo, A. Fekete, N. Gaston, F. Baletto, A. De Vita, *J. Chem. Phys.* **2018**, *148*, 241739.
- [162] J. Kloppenburg, L. B. Partay, H. Jonsson, M. A. Caro, *J. Chem. Phys.* **2023**, *158*, 134704.
- [163] S. Manna, T. D. Loeffler, R. Batra, S. Banik, H. Chan, B. Varughese, K. Sasikumar, M. Sternberg, T. Peterka, M. J. Cherukara, S. K. Gray, B. G. Sumpter, S. Sankaranarayanan, *Nat. Commun.* **2022**, *13*, 368.
- [164] J. D. Lee, J. B. Miller, A. V. Shneidman, L. Sun, J. F. Weaver, J. Aizenberg, J. Biener, J. A. Boscoboinik, A. C. Fouquer, A. I. Frenkel, J. E. S. van der Hoeven, B. Kozinsky, N. Marcella, M. M. Montemore, H. T. Ngan, C. R. O'Connor, C. J. Owen, D. J. Stacchiola, E. A. Stach, R. J. Madix, P. Sautet, C. M. Friend, *Chem. Rev.* **2022**, *122*, 8758.
- [165] E. P. George, D. Raabe, R. O. Ritchie, *Nat. Rev. Mater.* **2019**, *4*, 515.
- [166] J. Weinreich, A. Römer, M. L. Paleico, J. Behler, *J. Phys. Chem. C* **2020**, *124*, 12682.
- [167] J. Weinreich, M. L. Paleico, J. Behler, *J. Phys. Chem. C* **2021**, *125*, 14897.
- [168] S. Zha, D. I. Sharapa, S. Liu, Z. J. Zhao, F. Studt, *J. Phys. Chem. A* **2022**, *126*, 9440.
- [169] Z. Ma, P. Huang, J. Li, P. Zhang, J. Zheng, W. Xiong, F. Wang, X. Zhang, *npj Comput. Mater.* **2022**, *8*, 51.
- [170] S. Han, G. Barcaro, A. Fortunelli, S. Lysgaard, T. Vegge, H. A. Hansen, *npj Comput. Mater.* **2022**, *8*, 121.
- [171] S. Hu, W. X. Li, *Science* **2021**, *374*, 1360.
- [172] E. L. Kolsbjerg, A. A. Peterson, B. Hammer, *Phys. Rev. B* **2018**, *97*, 195424.
- [173] J. S. Lim, J. Vandermause, M. A. van Spronsen, A. Musaelian, Y. Xie, L. Sun, C. R. O'Connor, T. Egle, N. Molinari, J. Florian, K. Duanmu, R. J. Madix, P. Sautet, C. M. Friend, B. Kozinsky, *J. Am. Chem. Soc.* **2020**, *142*, 15907.
- [174] C. Vogt, B. M. Weckhuysen, *Nat. Rev. Chem.* **2022**, *6*, 89.
- [175] R. Sundararaman, D. Vigil-Fowler, K. Schwarz, *Chem. Rev.* **2022**, *122*, 10651.

- [176] S. Pullanchery, S. Kulik, B. Rehl, A. Hassanali, S. Roke, *Science* **2021**, *374*, 1366.
- [177] G. Sun, P. Sautet, *J. Am. Chem. Soc.* **2018**, *140*, 2812.
- [178] Z. W. Ulissi, M. T. Tang, J. Xiao, X. Liu, D. A. Torelli, M. Karamad, K. Cummins, C. Hahn, N. S. Lewis, T. F. Jaramillo, K. Chan, J. K. Nørskov, *ACS Catal.* **2017**, *7*, 6600.
- [179] J. Xu, W. Xie, Y. Han, P. Hu, *ACS Catal.* **2022**, *12*, 14812.
- [180] Y. Gong, X. Yan, J. Wen, Q. Meng, A. Li, X. Shi, *J. Mech. Phys. Solids* **2023**, *174*, 105259.
- [181] M. Liu, Y. Yang, J. R. Kitchin, *J. Chem. Phys.* **2021**, *154*, 134701.
- [182] L. Scalfi, M. Salanne, B. Rotenberg, *Annu. Rev. Phys. Chem.* **2021**, *72*, 189.
- [183] H. Heinz, T. J. Lin, R. K. Mishra, F. S. Emami, *Langmuir* **2013**, *29*, 1754.
- [184] N. Artrith, *J. Phys.: Condens. Matter* **2019**, *1*, 032002.
- [185] Y. Zhou, Y. Ouyang, Y. Zhang, Q. Li, J. Wang, *J. Phys. Chem. Lett.* **2023**, *14*, 2308.
- [186] N. Artrith, A. M. Kolpak, *Nano Lett.* **2014**, *14*, 2670.
- [187] S. K. Natarajan, J. Behler, *Phys. Chem. Chem. Phys.* **2016**, *18*, 28704.
- [188] M. Hellstrom, V. Quaranta, J. Behler, *Chem. Sci.* **2019**, *10*, 1232.
- [189] M. Eckhoff, J. Behler, *J. Chem. Phys.* **2021**, *155*, 244703.
- [190] M. F. Calegari Andrade, H. Y. Ko, L. Zhang, R. Car, A. Selloni, *Chem. Sci.* **2020**, *11*, 2335.
- [191] X. Yang, A. Bhowmik, T. Vegge, H. A. Hansen, *Chem. Sci.* **2023**, *14*, 3913.
- [192] P. Schienbein, J. Blumberger, *Phys. Chem. Chem. Phys.* **2022**, *24*, 15365.
- [193] C. Schran, F. L. Thiemann, P. Rowe, E. A. Muller, O. Marsalek, A. Michaelides, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, 2110077118.
- [194] H. Ghorbanfekr, J. Behler, F. M. Peeters, *J. Phys. Chem. Lett.* **2020**, *11*, 7363.
- [195] F. L. Thiemann, C. Schran, P. Rowe, E. A. Muller, A. Michaelides, *ACS Nano* **2022**, *16*, 10775.
- [196] M. F. Calegari Andrade, T. A. Pham, *J. Phys. Chem. Lett.* **2023**, *14*, 5560.
- [197] V. Kapil, C. Schran, A. Zen, J. Chen, C. J. Pickard, A. Michaelides, *Nature* **2022**, *609*, 512.
- [198] F. Musil, A. Grisafi, A. P. Bartok, C. Ortner, G. Csanyi, M. Ceriotti, *Chem. Rev.* **2021**, *121*, 9759.
- [199] A. Tkatchenko, *Nat. Commun.* **2020**, *11*, 4125.
- [200] M. Pinheiro Jr, F. Ge, N. Ferre, P. O. Dral, M. Barbatti, *Chem. Sci.* **2021**, *12*, 14396.
- [201] H. C. Zhou, S. Kitagawa, *Chem. Soc. Rev.* **2014**, *43*, 5415.
- [202] H. Demir, H. Daglar, H. C. Gulbalkan, G. O. Aksu, S. Keskin, *Coord. Chem. Rev.* **2023**, *484*, 215112.
- [203] M. Eckhoff, J. Behler, *J. Chem. Theory Comput.* **2019**, *15*, 3793.
- [204] M. Herbold, J. Behler, *Phys. Chem. Chem. Phys.* **2023**, *25*, 12979.
- [205] S. K. Achar, J. J. Wardzala, L. Bernasconi, L. Zhang, J. K. Johnson, *J. Chem. Theory Comput.* **2022**, *18*, 3593.
- [206] Y. Y. Yu, W. W. Zhang, D. H. Mei, *J. Phys. Chem. C* **2022**, *126*, 1204.
- [207] S. Y. Ding, W. Wang, *Chem. Soc. Rev.* **2013**, *42*, 548.
- [208] X. Zhao, P. Pachfule, A. Thomas, *Chem. Soc. Rev.* **2021**, *50*, 6871.
- [209] J. Huang, S. J. Shin, K. Tolborg, A. M. Ganose, G. Krenzer, A. Walsh, *Mater. Horiz.* **2023**, *10*, 2883.
- [210] S. Ma, Z. P. Liu, *Chem. Sci.* **2022**, *13*, 5055.
- [211] B. Mortazavi, X. Zhuang, T. Rabczuk, A. V. Shapeev, *Mater. Horiz.* **2023**, *10*, 1956.
- [212] B. Mortazavi, M. Silani, E. V. Podryabinkin, T. Rabczuk, X. Zhuang, A. V. Shapeev, *Adv. Mater.* **2021**, *33*, 2102807.
- [213] B. Bagchi, *Water in Biological and Chemical Processes*, Cambridge University Press, Cambridge **2013**.
- [214] M. Carballo-Pacheco, A. E. Ismail, B. Strodel, *J. Chem. Theory Comput.* **2018**, *14*, 6063.
- [215] V. H. Man, X. He, J. Gao, J. Wang, *J. Chem. Theory Comput.* **2021**, *17*, 6458.
- [216] S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, *J. Phys. Chem. B* **2015**, *119*, 5113.
- [217] C. Devereux, J. S. Smith, K. K. Huddleston, K. Barros, R. Zubatyuk, O. Isayev, A. E. Roitberg, *J. Chem. Theory Comput.* **2020**, *16*, 4192.
- [218] O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. Medrano Sandonas, A. Tkatchenko, K.-R. Müller, arXiv preprint arXiv.2205.08306, **2022**.
- [219] B. Lier, P. Poliak, P. Marquetand, J. Westermayr, C. Oostenbrink, *J. Phys. Chem. Lett.* **2022**, *13*, 3812.
- [220] M. Gastegger, K. T. Schutt, K. R. Muller, *Chem. Sci.* **2021**, *12*, 11473.
- [221] X. Pan, J. Yang, R. Van, E. Epifanovsky, J. Ho, J. Huang, J. Pu, Y. Mei, K. Nam, Y. Shao, *J. Chem. Theory Comput.* **2021**, *17*, 5745.
- [222] T. J. Giese, J. Zeng, S. Ekesan, D. M. York, *J. Chem. Theory Comput.* **2022**, *18*, 4304.
- [223] J. Zeng, T. J. Giese, S. Ekesan, D. M. York, *J. Chem. Theory Comput.* **2021**, *17*, 6993.
- [224] K. Zinovjev, *J. Chem. Theory Comput.* **2023**, *19*, 1888.
- [225] L. Boselt, M. Thurlemann, S. Riniker, *J. Chem. Theory Comput.* **2021**, *17*, 2641.
- [226] S. J. Lahey, C. N. Rowley, *Chem. Sci.* **2020**, *11*, 2362.
- [227] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell Jr, *J. Comput. Chem.* **2010**, *31*, 671.
- [228] V. L. Deringer, N. Bernstein, G. Csanyi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, S. R. Elliott, *Nature* **2021**, *589*, 59.
- [229] P. Rowe, V. L. Deringer, P. Gasparotto, G. Csanyi, A. Michaelides, *J. Chem. Phys.* **2020**, *153*, 034702.
- [230] V. L. Deringer, C. Merlet, Y. Hu, T. H. Lee, J. A. Kattirtzi, O. Pecher, G. Csanyi, S. R. Elliott, C. P. Grey, *Chem. Commun.* **2018**, *54*, 5988.
- [231] J. Wang, H. Shen, R. Yang, K. Xie, C. Zhang, L. Chen, K.-M. Ho, C.-Z. Wang, S. Wang, *Carbon* **2022**, *186*, 1.
- [232] V. L. Deringer, M. A. Caro, G. Csanyi, *Nat. Commun.* **2020**, *11*, 5461.
- [233] D. J. Liu, J. W. Evans, *J. Chem. Phys.* **2022**, *156*, 204106.
- [234] Y.-H. Fang, D.-d. Song, H.-x. Li, Z.-P. Liu, *J. Phys. Chem. C* **2021**, *125*, 10955.
- [235] D. Marchand, A. Jain, A. Glensk, W. A. Curtin, *Phys. Rev. Mater.* **2020**, *4*, 103601.
- [236] M. Stricker, B. Yin, E. Mak, W. A. Curtin, *Phys. Rev. Mater.* **2020**, *4*, 103602.
- [237] C. M. Andolina, J. G. Wright, N. Das, W. A. Saidi, *Phys. Rev. Mater.* **2021**, *5*, 083804.
- [238] C. M. Andolina, M. Bon, D. Passerone, W. A. Saidi, *J. Phys. Chem. C* **2021**, *125*, 17438.
- [239] S. Gao, H. Zhen, B. Wen, J. Ma, X. Zhang, *Nanoscale* **2022**, *14*, 2660.
- [240] R. Wanzenböck, M. Arrigoni, S. Bichelmaier, F. Buchner, J. Carrete, G. K. H. Madsen, *Digital Discovery* **2022**, *1*, 703.
- [241] J. Timmermann, Y. Lee, C. G. Staacke, J. T. Margraf, C. Scheurer, K. Reuter, *J. Chem. Phys.* **2021**, *155*, 244107.
- [242] Z. Tang, S. T. Bromley, B. Hammer, *J. Chem. Phys.* **2023**, *158*, 224108.
- [243] W. Chu, W. A. Saidi, O. V. Prezhdo, *ACS Nano* **2020**, *14*, 10608.
- [244] J. Westermayr, S. Chaudhuri, A. Jeindl, O. T. Hofmann, R. J. Maurer, *Digital Discovery* **2022**, *1*, 463.
- [245] X. T. Li, L. Chen, C. Shang, Z. P. Liu, *J. Am. Chem. Soc.* **2021**, *143*, 6281.
- [246] N. Gerrits, K. Shakouri, J. Behler, G. J. Kroes, *J. Phys. Chem. Lett.* **2019**, *10*, 1763.
- [247] I. Sanchez-Burgos, M. C. Muniz, J. R. Espinosa, A. Z. Panagiotopoulos, *J. Chem. Phys.* **2023**, *158*, 184504.
- [248] X. Li, W. Paier, J. Paier, *Front. Chem.* **2020**, *8*, 601029.
- [249] V. Quaranta, M. Hellstrom, J. Behler, J. Kullgren, P. D. Mitev, K. Hermansson, *J. Chem. Phys.* **2018**, *148*, 241720.

- [250] V. Quaranta, J. Behler, M. Hellström, *J. Phys. Chem. C* **2018**, *123*, 1293.
- [251] W. Zhao, H. Qiu, W. Guo, *J. Phys. Chem. C* **2022**, *126*, 10546.
- [252] S. K. Achar, J. Schneider, D. A. Stewart, *ACS Appl. Mater. Interfaces* **2022**, *14*, 56963.
- [253] E. V. Podryabinkin, A. G. Kvashnin, M. Asgarpour, Maslenikov II, D. A. Ovsyannikov, P. B. Sorokin, M. Y. Popov, A. V. Shapeev, *J. Chem. Theory Comput.* **2022**, *18*, 1109.
- [254] J. Lu, J. Wang, K. Wan, Y. Chen, H. Wang, X. Shi, *J. Chem. Phys.* **2023**, *158*, 204702.
- [255] S. Vandenhaute, M. Cools-Ceuppens, S. DeKeyser, T. Verstraelen, V. Van Speybroeck, *npj Comput. Mater.* **2023**, *9*, 19.
- [256] S. Ma, C. Shang, C. M. Wang, Z. P. Liu, *Chem. Sci.* **2020**, *11*, 10113.
- [257] F. Daeyaert, F. Ye, M. W. Deem, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 3413.
- [258] D. Schwalbe-Koda, S. Kwon, C. Paris, E. Bello-Jurado, Z. Jensen, E. Olivetti, T. Willhammar, A. Corma, Y. Roman-Leshkov, M. Moliner, R. Gomez-Bombarelli, *Science* **2021**, *374*, 308.
- [259] P. L. Kang, C. Shang, Z. P. Liu, *J. Am. Chem. Soc.* **2019**, *141*, 20525.
- [260] T. Bucko, M. Gesvandtnerova, D. Rocca, *J. Chem. Theory Comput.* **2020**, *16*, 6049.
- [261] M. Gešvandtnerová, D. Rocca, T. Bučko, *J. Catal.* **2021**, *396*, 166.
- [262] S. Ma, Z. P. Liu, *Nat. Commun.* **2022**, *13*, 2716.
- [263] A. Erlebach, P. Nachtigall, L. Grajciar, *npj Comput. Mater.* **2022**, *8*, 174.
- [264] M. Bocus, R. Goeminne, A. Lamine, M. Cools-Ceuppens, T. Verstraelen, V. Van Speybroeck, *Nat. Commun.* **2023**, *14*, 1008.
- [265] B. Mortazavi, A. Rajabpour, X. Zhuang, T. Rabczuk, A. V. Shapeev, *Carbon* **2022**, *186*, 501.
- [266] B. Mortazavi, M. Shahrokhi, F. Shojaei, T. Rabczuk, X. Zhuang, A. V. Shapeev, *Nanoscale* **2022**, *14*, 4324.
- [267] B. Mortazavi, *Carbon* **2021**, *182*, 373.
- [268] B. Mortazavi, I. S. Novikov, A. V. Shapeev, *Carbon* **2022**, *188*, 431.
- [269] S. Arabha, A. Rajabpour, *Int. J. Heat Mass Transfer.* **2021**, *178*, 121589.
- [270] S. K. Achar, L. Zhang, J. K. Johnson, *J. Phys. Chem. C* **2021**, *125*, 14874.
- [271] S. Wengert, G. Csanyi, K. Reuter, J. T. Margraf, *J. Chem. Theory Comput.* **2022**, *18*, 4586.
- [272] R. K. Cersonsky, M. Pakhnova, E. A. Engel, M. Ceriotti, *Chem. Sci.* **2023**, *14*, 1272.
- [273] V. Kapil, E. A. Engel, *Proc. Natl. Acad. Sci. USA* **2022**, *119*, 2111769119.
- [274] S. Wengert, G. Csanyi, K. Reuter, J. T. Margraf, *Chem. Sci.* **2021**, *12*, 4536.
- [275] D. Rosenberger, J. S. Smith, A. E. Garcia, *J. Phys. Chem. B* **2021**, *125*, 3598.
- [276] G. Imbalzano, Y. Zhuang, V. Kapil, K. Rossi, E. A. Engel, F. Grasselli, M. Ceriotti, *J. Chem. Phys.* **2021**, *154*, 074102.
- [277] L. Kahle, F. Zipoli, *Phys. Rev. E* **2022**, *105*, 015311.
- [278] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, *Artif. Intell. Rev.*, <https://doi.org/10.1007/s10462-023-10562-9>.
- [279] A. P. Bartók, J. Kermode, *arXiv preprint arXiv:2206.08744* **2022**.
- [280] D. Zhang, H. Bi, F.-Z. Dai, W. Jiang, L. Zhang, H. Wang, *arXiv preprint arXiv:2208.08236* **2022**.
- [281] S. N. Pozdnyakov, M. J. Willatt, A. P. Bartok, C. Ortner, G. Csanyi, M. Ceriotti, *Phys. Rev. Lett.* **2020**, *125*, 166001.
- [282] T. R. Walsh, M. R. Knecht, *Chem. Rev.* **2017**, *117*, 12641.
- [283] T. R. Walsh, *Acc. Chem. Res.* **2017**, *50*, 1617.



Kaiwei Wan received his bachelor's degree from the University of Science and Technology Beijing in 2017, followed by a Ph.D. in physical chemistry, under the supervision of Prof. Xinghua Shi, from the National Center for Nanoscience and Technology, Chinese Academy of Sciences in 2023. His research interests include the interfacial properties of nanomaterials, computational surface catalysis, and the development of machine learning force field methods.



Jianxin He received his Danish master's degree in nanoscience and technology from the University of Copenhagen in January 2023, at the Sino-Danish Center. Currently, he is pursuing his Chinese master's degree at the National Center for Nanoscience and Technology, under the supervision of Prof. Xinghua Shi. His interests lie in organic chemistry and computational physics.



Xinghua Shi received his bachelor's degree from Peking University, Master's degree from the Institute of Mechanics, Chinese Academy of Sciences, and Ph.D. from Brown University. From 2011 to 2016 back in the Institute of Mechanics, CAS as an associate professor, he studied nanoparticle–cell interaction with multiscale modeling. At the beginning of 2016, he moved to the National Center for Nanoscience and Technology as a full professor and principal investigator. There he focused on the simulation of catalysis, self-assembly, and drug delivery, and the development of machine learning force field methods.