ANNUAL REVIEWS

Annual Review of Physical Chemistry

Neural Network Potentials: A Concise Overview of Methods

Emir Kocer,* Tsz Wai Ko,* and Jörg Behler

Institut für Physikalische Chemie, Theoretische Chemie, Universität Göttingen, Göttingen, Germany; email: ekocer@uni-goettingen.de, tko@uni-goettingen.de, joerg.behler@uni-goettingen.de

Annu. Rev. Phys. Chem. 2022. 73:163–86

First published as a Review in Advance on January 4, 2022

The *Annual Review of Physical Chemistry* is online at physchem.annualreviews.org

https://doi.org/10.1146/annurev-physchem-082720-034254

*These authors contributed equally to this article

Keywords

neural network potentials, machine learning, atomistic simulations, molecular dynamics, potential energy surfaces

Abstract

In the past two decades, machine learning potentials (MLPs) have reached a level of maturity that now enables applications to large-scale atomistic simulations of a wide range of systems in chemistry, physics, and materials science. Different machine learning algorithms have been used with great success in the construction of these MLPs. In this review, we discuss an important group of MLPs relying on artificial neural networks to establish a mapping from the atomic structure to the potential energy. In spite of this common feature, there are important conceptual differences among MLPs, which concern the dimensionality of the systems, the inclusion of long-range electrostatic interactions, global phenomena like nonlocal charge transfer, and the type of descriptor used to represent the atomic structure, which can be either predefined or learnable. A concise overview is given along with a discussion of the open challenges in the field.

1. INTRODUCTION AND SCOPE OF THIS REVIEW

Machine learning plays an increasingly important role in many aspects of life (1, 2). The ability to provide, after a training process, accurate predictions makes modern machine learning algorithms a very attractive tool for applications in all fields of science in which the analysis, classification, and interpretation of data are important (3–8). Consequently, data science and machine learning are now called the fourth paradigm of science (9), along with the three established paradigms of empirical experimental studies, theoretical models, and computer simulations.

A particularly fruitful combination of machine learning and computer simulations that has emerged in the past two decades is the representation of the multidimensional potential energy surface (PES) by machine learning potentials (MLPs) (10–20). The PES is of central importance for reaching an atomic-level understanding of any type of system, from small molecules to bulk materials, as it contains all the information about the stable and metastable structures, the atomic forces driving the dynamics at finite temperatures, the transition states and barriers governing reactions and structural transitions, and also the atomic vibrations. For a long time, computer simulations in chemistry, molecular biology, and materials science had to rely either on computationally demanding electronic structure calculations like density-functional theory (DFT) or on efficient but significantly less accurate empirical potentials or force fields based on physical intuition and approximations. Modern MLPs bridge this gap by learning the shape of the PES from reference data obtained from high-level electronic structure calculations. The resulting analytic MLPs represent the atomic interactions, commonly as a function of the atomic positions and nuclear charges, and can then be used without a significant loss in accuracy in large-scale simulations, like molecular dynamics, that are many orders of magnitude faster than the underlying electronic structure calculations. MLPs thus represent an important tool to meet the continuously increasing need for computer simulations of more and more complex systems.

Since the advent of the first MLP in 1995 (21), many types of methods have been proposed, such as neural network potentials (NNPs) (22–25), Gaussian approximation potentials (GAPs) (26, 27), kernel-based approaches like gradient domain machine learning (GDML) (28), spectral neighbor analysis potentials (SNAPs) (29, 30), moment tensor potentials (MTPs) (31), atomic cluster expansion (ACE) (32), atomic permutationally invariant polynomials (aPIPs) (33), and support vector machines (SVMs) (34). MLPs offer many advantages, such as a very flexible functional form, which allows one to represent the available reference data with a very high accuracy (35); generality, which makes them suitable for all types of bonding and atomic interactions, from covalent bonds via metallic bonding to dispersion interactions; and a well-defined analytic form that enables the consistent computation of energies and gradient-based properties like forces and the stress tensor. The main disadvantages of MLPs are the lack of a physical functional form and the resulting limited transferability beyond the structural diversity of the underlying training data, which make a careful validation of the obtained potentials an essential step. For the same reason, the construction of MLPs is computationally very demanding, because large training sets are required to ensure that all required information about the topology of the PES can be learned from the reference data.

The applicability of current MLPs is very diverse in many respects, concerning for example the dimensionality of the systems, the range of the atomic interactions that can be described, and the possible inclusion of physical laws and concepts. Recently, a classification scheme dividing MLPs into four generations has been proposed (36–38), which we adopt in this review. First-generation MLPs can provide accurate PESs for low-dimensional systems like small molecules. Second-generation MLPs make use of the locality of a major part of the atomic interactions and employ the concept of atomic energies, making these potentials applicable to high-dimensional

systems containing thousands of atoms. Third-generation MLPs overcome this locality approximation by explicitly also including long-range interactions, like electrostatics based on Coulomb's law, without truncation. Still, they rely on atomic charges depending on the local chemical environment. Finally, fourth-generation potentials are able to capture phenomena like nonlocal charge transfer by taking into account the global structure and charge distribution of the system.

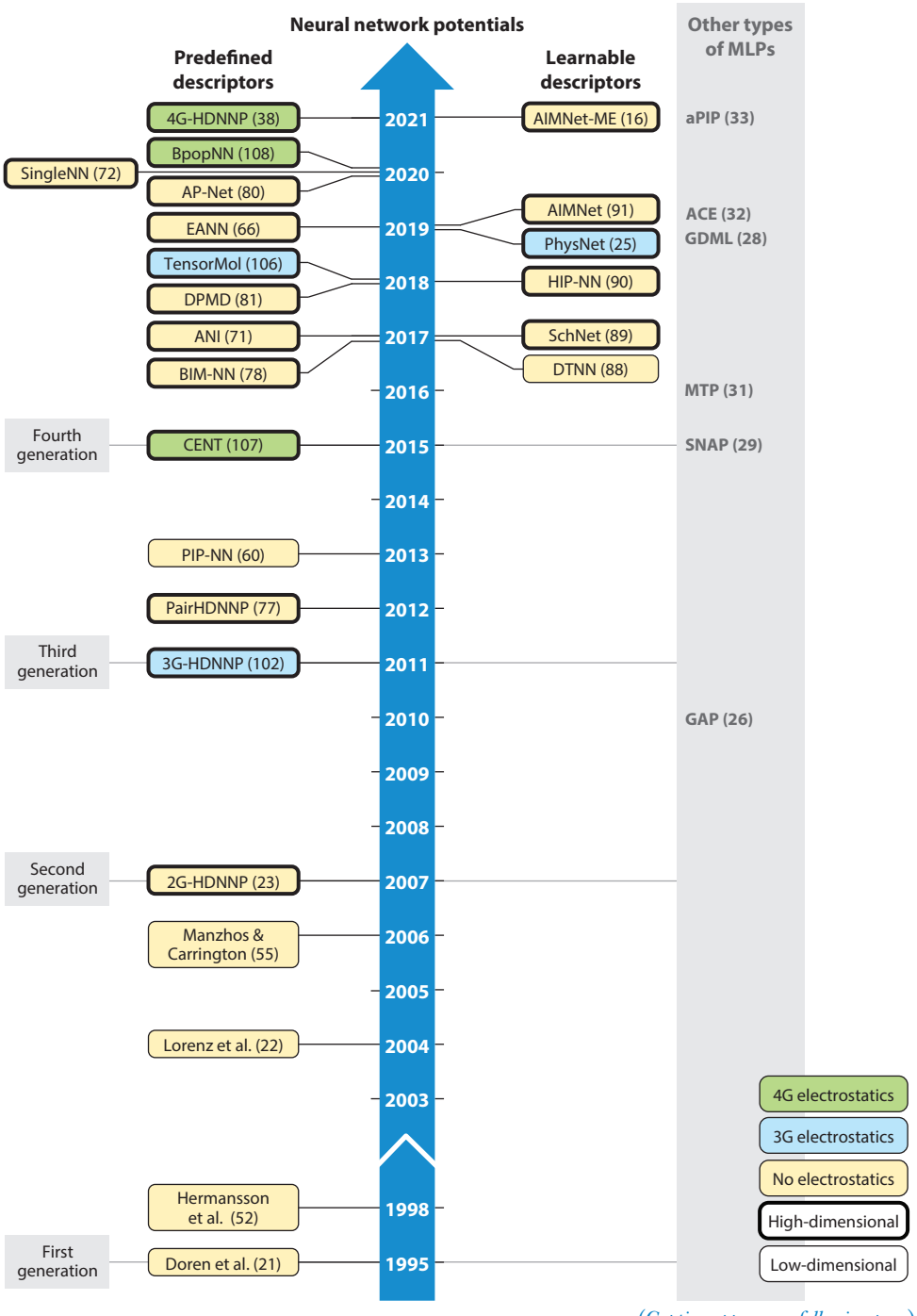
In this review, we focus on discussing MLPs based on neural networks. NNPs not only have the longest history but also have by far the largest diversity in terms of methods and concepts of all classes of MLPs. We restrict our discussion to methods providing continuous PESs for arbitrary atomic positions suitable for molecular dynamics simulations, while we note that many other interesting neural networks–based approaches have been proposed and successfully tested to predict, for example, the atomization energies of large sets of organic molecules in their global minimum geometries (39) and a variety of properties of materials (40–42). Further, we do not include methods used for predicting atomic partial charges or multipoles unless they are used to explicitly compute the electrostatic energy contribution to the PES. Our aim is to present the main methods and concepts, while applications are only briefly mentioned. The interested reader is referred to many reviews providing comprehensive lists of such applications (10, 11, 15, 43).

Even with this restricted scope, giving a comprehensive overview of all available methods in this rapidly expanding field is an impossible endeavor. **Figure 1** shows a schematic timeline of the historic evolution of NNPs using representative examples. The presented methods are classified based on the types of descriptors of the atomic configurations, which can be predefined or learnable; the dimensionality of the systems that can be studied; and the inclusion or absence of long-range electrostatic interactions. These criteria will guide our discussion of the individual methods in the following sections. In addition, the evolution of NNPs is put in the context of other important MLP methods proposed to date, which are mentioned in the gray column in **Figure 1** but are not discussed here in more detail.

2. THE FIRST GENERATION: LOW-DIMENSIONAL SYSTEMS

In 1995, Doren and coworkers published the first MLP trained to electronic structure data that enabled simulations at strongly reduced costs (21). They used a feed-forward neural network to construct a mapping from the atomic positions to the potential energy for the adsorption of H_2 on a cluster model of the Si(100) surface. In this pioneering work, many technical problems were investigated and solved that nowadays are considered textbook knowledge about MLPs, and the advantages and limitations of machine learning for representing PESs have been thoroughly discussed. In the following decade, the applicability of neural networks has further been explored by several groups for numerous molecular systems (44–47). Another branch of applications addressed adsorption and gas-surface dynamics (22, 48–50). The construction of NNPs for condensed systems like liquid and bulk materials was out of reach at that time, but several attempts have been made to augment empirical potentials suitable for larger systems by neural networks (51–53).

In addition to these applications, very promising methodological advances have been made, resulting for example in very accurate potentials based on a sum of products, optimized coordinates, and a high-dimensional model representation proposed by Manzhos & Carrington, which reached spectroscopic accuracy (54–56). In 2009, Malshe and coworkers proposed to use a set of neural networks for the different terms in a many-body expansion (57). Finally, in 2013, Jiang and coworkers used permutationally invariant polynomials, originally developed by Braams and Bowman for the construction of PESs with full permutation invariance (58, 59), as input coordinates for constructing molecular permutation invariant polynomial neural networks (PIP-NNs) (60, 61).



(*Caption appears on following page*)

Figure 1 (*Figure appears on preceding page*)

Timeline for the evolution of NNPs. NNPs using predefined descriptors are shown on the left, and NNPs with learnable descriptors (message passing neural networks) are shown on the right. The frame width and background color label the dimensionality and the absence or inclusion of long-range electrostatic interactions in third-generation (3G) and fourth-generation (4G) potentials. Other important types of MLPs not discussed in this review are shown in the gray column on the right. Abbreviations: ACE, atomic cluster expansion; AP-Net, atomic-pairwise neural network; aPIP, atomic permutationally invariant polynomial; BIM-NN, bonds-in-molecule neural network; BpopNN, Becke population neural network; CENT, charge equilibration via neural network technique; DPMD, deep potential molecular dynamics; DTNN, deep tensor neural network; EANN, embedded atom neural network; GAP, Gaussian approximation potential; GDML, gradient domain machine learning; HDNNP, high-dimensional neural network potential; HIP-NN, hierarchically interacting particle neural network; MLP, machine learning potential; MTP, moment tensor potential; NNP, neural network potential; PIP-NN, permutation invariant polynomial neural network; SNAP, spectral neighbor analysis potential.

In spite of the diversity of methods and systems that have been studied in the initial decade, the common feature of all these first-generation MLPs, which exclusively relied on neural networks, is the small number of degrees of freedom that could be taken into account explicitly, which limited applications to small molecules in a vacuum, to molecules interacting with frozen surfaces, or to low-dimensional additive terms, often in combination with conventional empirical potentials or force fields. A summary of these applications can be found in two early reviews (10, 11).

3. THE SECOND GENERATION: LOCAL METHODS FOR HIGH-DIMENSIONAL SYSTEMS

3.1. Overview

Many important advances have been made in first-generation NNPs, but the use of first-generation NNPs remained limited to rather low-dimensional systems including the degrees of freedom of a few atoms only. The main conceptual problem preventing the application of MLPs to high-dimensional systems containing thousands of atoms was the lack of suitable structural descriptors including the mandatory invariances of the total energy with respect to translation, rotation, and permutation (i.e., the order) of chemically equivalent atoms in the system. Further, the use of a single neural network to express the global energy of the system, which is a central component of many early NNPs, did not allow an application to systems containing variable numbers of atoms, as the dimensionality of the system is related to the number of input neurons, which cannot be changed after the training of the neural network.

Many empirical potentials in the field of materials science construct the total energy as a sum of atomic energies, that is,

$$E_{\text{total}} = \sum_{i=1}^{N_{\text{atoms}}} E_i, \qquad 1.$$

and most classical force fields used, for example, in biomolecular chemistry can also be cast into this form. For these types of potentials, incorporating all invariances into the potential is straightforward, because they consist of very simple low-dimensional terms that are additive and rely on translationally and rotationally invariant internal coordinates like interatomic distances, angles, and dihedral angles. This is different in the case of MLPs, which have a more complex multidimensional functional form that allows them to reach a very high numerical accuracy and requires an ordered vector of input coordinates lacking intrinsic permutation invariance. Consequently, in the first decade of MLP development it has not been possible to make use of Equation 1, and the

search for suitable descriptors taking all invariances into account has been a frustrating challenge, with only a few special-purpose solutions for low-dimensional systems (52, 62).

This problem has now been solved, and from today's perspective it is hard to imagine the difficult situation of early MLP developers, because a huge variety of different descriptors for high-dimensional systems is now readily available (63, 64). The resulting, very powerful second generation of MLPs based on Equation 1 is applicable in principle to systems of arbitrary size, and many combinations of the two main ingredients of modern MLPs—the descriptors of the atomic environments and the machine learning algorithm connecting the descriptor values to the energies—have been explored.

Two major classes of second-generation NNPs can be distinguished, depending on the type of descriptor. The first high-dimensional NNPs that became available made use of descriptors with predefined functional forms, which contained a few parameters defining the spatial shape. In addition, in recent years a second class of second-generation NNPs has received increasing attention, which is based on the automatic learning of descriptors from structural information using message passing neural networks (MPNNs). This is not to be confused with the optimization of the parameters of the predefined descriptors, since no ad-hoc assumptions about the functional form are made. Both approaches are presented in the subsequent sections, and the most important examples are briefly discussed.

3.2. Predefined Descriptors

Predefined descriptors share some similarities with basis functions used in electronic structure calculations. Sometimes these parameters are even optimized during the fitting process (65, 66). Due to their simplicity and efficiency, such predefined descriptors are still dominantly used in applications of current MLPs.

3.2.1. High-dimensional neural network potentials. Using Equation 1 to construct the energy of the system implies the partitioning of the total energy into atomic contributions, and these contributions can often, to a good approximation, be expressed as a function of the local atomic environment, if the spatial cutoff for describing the environment is chosen large enough to include all energetically relevant atomic interactions. The first MLPs making use of this locality approximation are the high-dimensional neural network potentials (HDNNPs) proposed by Behler & Parrinello in 2007 (23). In this approach, which for the first time enabled one to perform large-scale simulations driven by a MLP, each atomic energy is calculated as the output of an individual atomic neural network (see **Figure 2*a*, *b***). To ensure permutation invariance, the architecture and weight parameters of all atomic neural networks for atoms of the same element are constrained to be the same. Alternatively, the method can be viewed as using one separate atomic neural network per element, which is evaluated as many times as atoms of the respective element are present in the system. The training of these atom-based second-generation HDNNPs (2G-HDNNPs) is done using energies and forces. The partitioning into atomic energies is carried out automatically in the fitting process, allowing the direct use of total energies. Once trained, the HDNNP can then be applied to simulations of systems of variable size by adapting the number of atomic neural networks accordingly, resulting in $O(N)$ scaling of the computational costs of the method. The key step for the introduction of HDNNPs has been the development of a novel type of descriptor, atom-centered symmetry functions (ACSFs) (23, 67, 68), which for the first time allowed one to construct structural fingerprints of the local environments with exact translational, rotational, and permutational invariances. Because all structurally equivalent atomic environments give rise

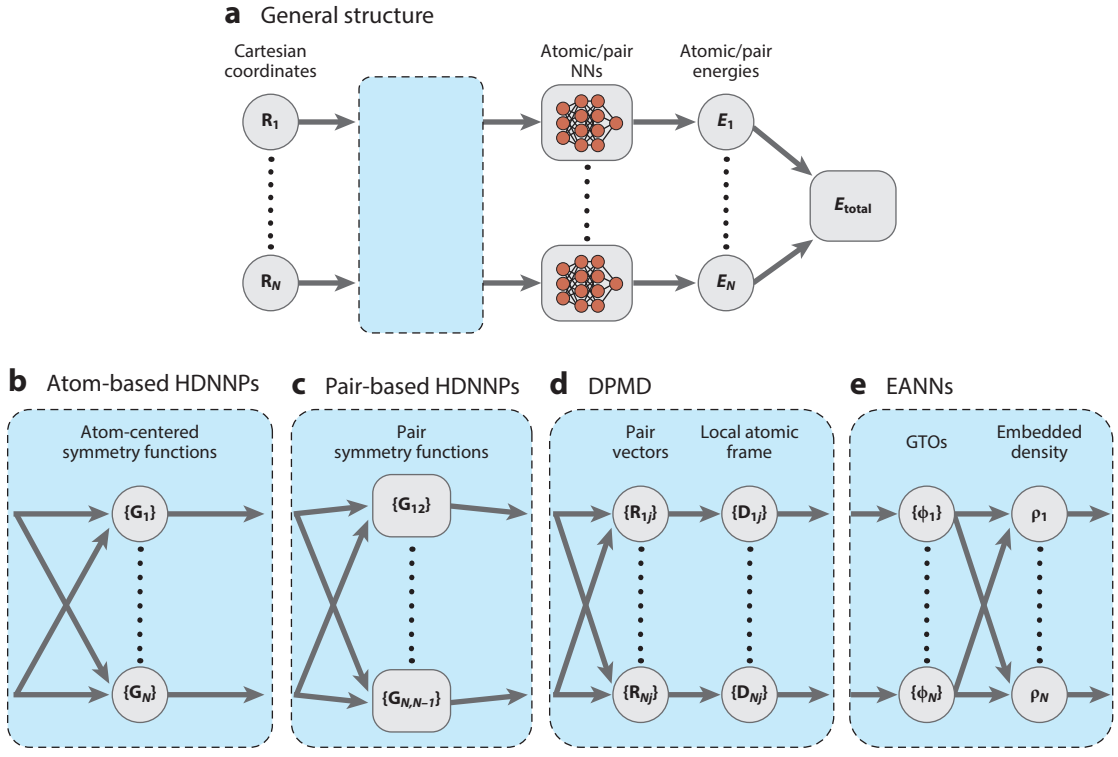


Figure 2

(*a*) General structure of second-generation neural network potentials with predefined descriptors. The total (short-range) energy is the sum of atomic or pair energies provided as outputs of individual atomic or pair NNs. Starting from the Cartesian coordinates of the atoms $\{\mathbf{R}_i\}$, different types of transformations to input descriptors characterizing the local atomic environments are performed for (*b*) atom-based or (*c*) pair-based HDNNPs, (*d*) DPMD, and (*e*) EANNs. Details of these methods are discussed in Sections 3.2.1–3.2.4. Abbreviations: DPMD, deep potential molecular dynamics; EANN, embedded atom neural network; GTO, Gaussian-type orbital; HDNNP, high-dimensional neural network potential; NN, neural network.

to the same symmetry function vector, these invariances are also preserved in the atomic energy outputs of the neural networks.

The local atomic environments are defined by a cutoff function,

$$f_{\mathrm{cut}}(R_{i,j}) = \begin{cases} 0.5\left[\cos\left(\frac{\pi R_{ij}}{R_c}\right)+1\right], & \text{if } R_{ij} \le R_{\mathrm{c}} \\ 0, & \text{if } R_{ij} \ge R_{\mathrm{c}} \end{cases}, \qquad 2.$$

which decays smoothly to zero in value and slope at the cutoff radius R_{c}, and many alternative cutoff functions are available (69, 70). R_{ij} is the distance between the reference atom i and any neighboring atom j. The cutoff function is a central ingredient of ACSFs and many other similar descriptors that describe the positions of all neighboring atoms inside the resulting cutoff spheres. Many types of ACSFs have been proposed in the literature to characterize the radial and angular distributions of neighbors (67). The most commonly used radial ACSF has the form

$$G_{\mathrm{radial},i} = \sum_{j \in R_{\mathrm{c}}} \mathrm{e}^{-\eta(R_{ij}-R_s)^2} \times f_c(R_{ij}), \qquad 3.$$

with the hyperparameters η and R_s defining the shape of a Gaussian sphere around the central atom. Due to the summation over all neighboring atoms, the value of $G_{\text{radial},i}$ can be interpreted as a continuous coordination number, and a combination of different parameter sets can be used to obtain information about the radial distribution of neighbors. In a similar way, the angular distribution of neighbors can be probed by the angular function

$$G_{\text{angular},i} = 2^{1-\zeta} \sum_{j \in R_c} \sum_{k \neq j \in R_c} (1 + \lambda \cos\theta_{ijk})^{\zeta} \times e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} \times f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}), \quad 4.$$

with θ_{ijk} being the angle enclosed by ij and ik, and with ζ, λ, and η being the hyperparameters. For multi-element systems, ACSFs for each pair of elements (radial) or each triple of elements (angular) are constructed, resulting in a combinatorial increase of the number of symmetry functions if many elements are present. A detailed discussion of the properties of ACSFs is beyond the scope of this review and can be found in Reference 67. The typical error of energies and forces obtained from HDNNPs is around 1.0 meV/atom and 0.1 eV/Å, respectively, and in the past decade, HDNNPs have been reported for a variety of systems (43), from bulk materials via aqueous solutions to interfaces, and very generally applicable parameterizations of HDNNPs for a broad range of organic molecules, like ANI (71), which employs a modified angular symmetry function, have also been published.

Some attempts have been made in recent years to overcome the combinatorial increase in the number of ACSFs for multi-element systems. In 2018, Gastegger and coworkers introduced weighted atom-centered symmetry functions (wACSFs) (65). In wACSFs, the ACSFs sharing the same parameters are merged for all element combinations using element-dependent weighting prefactors that make the input dimensionality of the element-specific atomic neural networks independent of the number of elements in the system. In a recent study, Liu & Kitchin used wACSFs and proposed to also combine the atomic neural networks into a single, universal atomic neural network with multiple outputs for each chemical species in a system (72). All elements thus share the same connecting weights in the single atomic neural network, while the weights connecting the output neurons remain element specific, which is based on the assumption of a roughly linear relation between a universal feature vector generated by the atomic neural network and the element-specific atomic energy, which has been investigated for bulk elements. A similar method has also been proposed by Profitt & Pearson for molecular systems (73). Finally, the extraction of the relevant information from a large pool of ACSFs can also be done automatically—for example, by CUR (74, 75) or using convolutional neural networks (76).

3.2.2. Pair-based high-dimensional neural network potentials. The sum over atomic energies in Equation 1 is the most frequently used energy expression in high-dimensional MLPs. However, it can be equivalently written as a sum of atom-pair energies directly reflecting the atomic interactions as

$$E = \sum_{i=1}^{N_{\text{atoms}}} E_i = \frac{1}{2} \sum_{i=1}^{N_{\text{atoms}}} \sum_{j \neq i}^{N_{\text{atoms}}} E_{ij}. \quad 5.$$

In 2012, Behler and coworkers (77) suggested to construct the environment-dependent pair energies using pair symmetry functions, which are a new class of descriptors characterizing the combined environments of both atoms in each pair up to a cutoff radius with full translational, rotational, and permutational symmetry. The resulting structure of the high-dimensional pair NNP is very similar to conventional atom-based HDNNPs, with atom pairs up to a maximum interatomic distance replacing the atoms as central structural entities. For each element combination

there is one distinct neural network to be trained, which is evaluated as many times as atom pairs of the respective type are present in the system (see **Figure 2*c***). The method has been shown to be as accurate as the atom-based approach, but because of the larger number of atom pairs compared to atoms, the computational costs are increased as there is a larger number of neural networks to be evaluated.

An alternative method to express the energy as a sum of bond energies was proposed by Parkhill and coworkers in 2017 (78). In this bonds-in-molecule neural network (BIM-NN), a modified version of the bags of bonds descriptor (79) has been employed using an element-specific distance cutoff to take only covalent bonds into account. The descriptor consists of the bond length as well as information about the directly connected atoms. The resulting pair energies have been interpreted as chemical bond energies. So far the method has only been applied to isolated molecules in vacuum.

A further related approach is the atomic-pairwise neural network (AP-Net), which has recently been introduced by Sherrill and coworkers (80). AP-Net has been developed focusing on noncovalent interactions by representing physically meaningful interaction energies derived from symmetry-adapted perturbation theory, which allows researchers to represent all element combinations by a single neural network for a given type of physical interaction.

3.2.3. Deep potential molecular dynamics. Deep potential molecular dynamics (DPMD), developed by E and coworkers (81), is a force-extended version of the deep potential method (82) that includes energy training only. Like in HDNNPs, the total energy is written as a sum of environment-dependent atomic energies (Equation 1). For describing the positions of the neighboring atoms, a local atomic frame based on the two closest neighboring atoms is defined. All neighboring atoms are then sorted by element and inverse distance, and for each neighbor j a descriptor vector

$$\mathbf{D}_{ij} = \{D^0_{ij}, D^1_{ij}, D^2_{ij}, D^3_{ij}\} = \left\{ \frac{1}{R_{ij}}, \frac{x_{ij}}{R^2_{ij}}, \frac{y_{ij}}{R^2_{ij}}, \frac{z_{ij}}{R^2_{ij}} \right\}, \qquad 6.$$

with x_{ij}, y_{ij}, and z_{ij} being the components of the connecting vector $\mathbf{R}_{ij}$, is defined. Beyond a predefined number of closest neighbors inside the cutoff radius, just the first component $1/R_{ij}$ is used, corresponding to radial information only. The descriptor vectors of all neighboring atoms are then provided jointly as $\{\mathbf{D}_{ij}\}$ into atomic neural networks to yield the atomic energies E_i as a function of the geometry of all atom pairs relative to the reference frame (see **Figure 2*d***). An advantage of DPMD is the absence of any tunable hyperparameters in the descriptors that depend only on the atomic positions, reducing the complexity of the potential. Drawbacks are the simple functional forms of the descriptors without explicit many-body information, which therefore needs to be established by the atomic neural networks, and the presence of small discontinuities in the forces at the cutoff radius, as no smooth cutoff function is applied. The latter limitation has been overcome recently by introducing scalar weight functions in the Deep Potential–Smooth Edition (DeepPot-SE) method (83), in which the atomic neural networks are split into encoding networks that determine self-adapted descriptors by mapping the structure to multiple outputs and fitting networks, both of which are optimized in the training process.

3.2.4. Embedded atom neural network potentials. In 2019 Zhang and coworkers proposed the embedded atom neural network (EANN) approach (66), which was inspired by the embedded atom method (EAM) (84), an empirical potential frequently used in materials science in which each atomic energy is given as an embedding energy arising from the interaction with the density of the

surrounding atoms. In the EANN approach, the scalar density value at the position of each atom used in EAM is replaced by a vector of Gaussian-type orbital (GTO)-based densities, which are used as input for atomic neural networks yielding the atomic energy contributions (see **Figure 2*e***). The expansion coefficients of the GTOs forming angular momentum-specific elements of the density vector are optimized during the training process like in a linear basis set expansion. The density vector of each atom is then used as input vector for the respective atomic neural network yielding the atomic energy. Only neighboring atoms up to a cutoff contribute to the density, and a sufficient orbital overlap must be ensured to obtain accurate potentials. A transformation to an angular basis makes the method computationally very efficient, as three-body descriptors do not need to be computed explicitly.

3.3. Learnable Descriptors

All NNPs discussed so far have in common that the descriptor functions for the atomic structure are predefined and often contain some hyperparameters, which may be manually selected or adapted during the training process. The purpose of these descriptors is to provide local structural fingerprints of the atomic environments as input for atomic neural networks. Thus, the only necessary information is the atomic structure, and a vector of function values needs to be calculated to characterize each environment as uniquely as possible with as few functions as possible.

In the context of machine learning applications in molecular modeling, the idea of replacing predefined static descriptors with learned dynamic ones using structural information was first introduced by Duvenaud and coworkers (85). Inspired by extended-connectivity fingerprints (ECFPs) (86), they treated molecules as graph networks and employed a convolutional layer to process these graphical networks to predict their final feature vectors, that is, the descriptors. Although they did not exploit these methods to construct PESs, their work paved the way for the definition of a new class of MLPs, for which Gilmer and coworkers in 2017 coined the name "message passing neural networks" (MPNNs) (87). All MPNN methods have in common that predefined descriptors are replaced by automatically determined descriptors learned from the geometric structure.

In MPNNs, a molecule is considered as a three-dimensional graph consisting of nodes (vertices) and connections (edges), which are associated, respectively, with the atoms defined by their nuclear charges $\{Z_i\}$ and bonds and with the interatomic distances more generally $\{R_{ij}\}$. Each atom is described by a feature vector, which is iteratively updated in an interaction block in a message passing phase using message functions that process information about neighboring atoms, like distances and their respective current feature vectors. The various flavors of MPNNs differ in the types of message functions that iteratively embed atoms into their chemical environments, the atomic update functions (vertex update functions) describing these interactions, and the distance of the neighbors included. Once the message passing phase has been completed, after a predefined number of steps, the resulting atomic feature vectors, which now contain the information about the atom in its environment, are used in a readout phase to predict the target property. In the context of NNPs, the atomic feature vectors are typically passed to atomic neural networks in this step to yield atomic energy contributions summing up to the total energy.

3.3.1. Deep tensor neural networks. The first MLP employing MPNNs has been the deep tensor neural network (DTNN) proposed by Schütt and coworkers in 2017 (88), which has been designed to describe the properties of small organic molecules. Starting from the nuclear charges and a complete distance matrix $\mathbf{D}$, the pairwise interatomic distances D_{ij} are first expanded in a Gaussian basis, yielding a coefficient vector $\mathbf{d}_{ij}$ for each atomic pair (see **Figure 3*a***). Further, each atom i is represented by a multidimensional feature vector $\mathbf{c}_i$, starting with an initial random

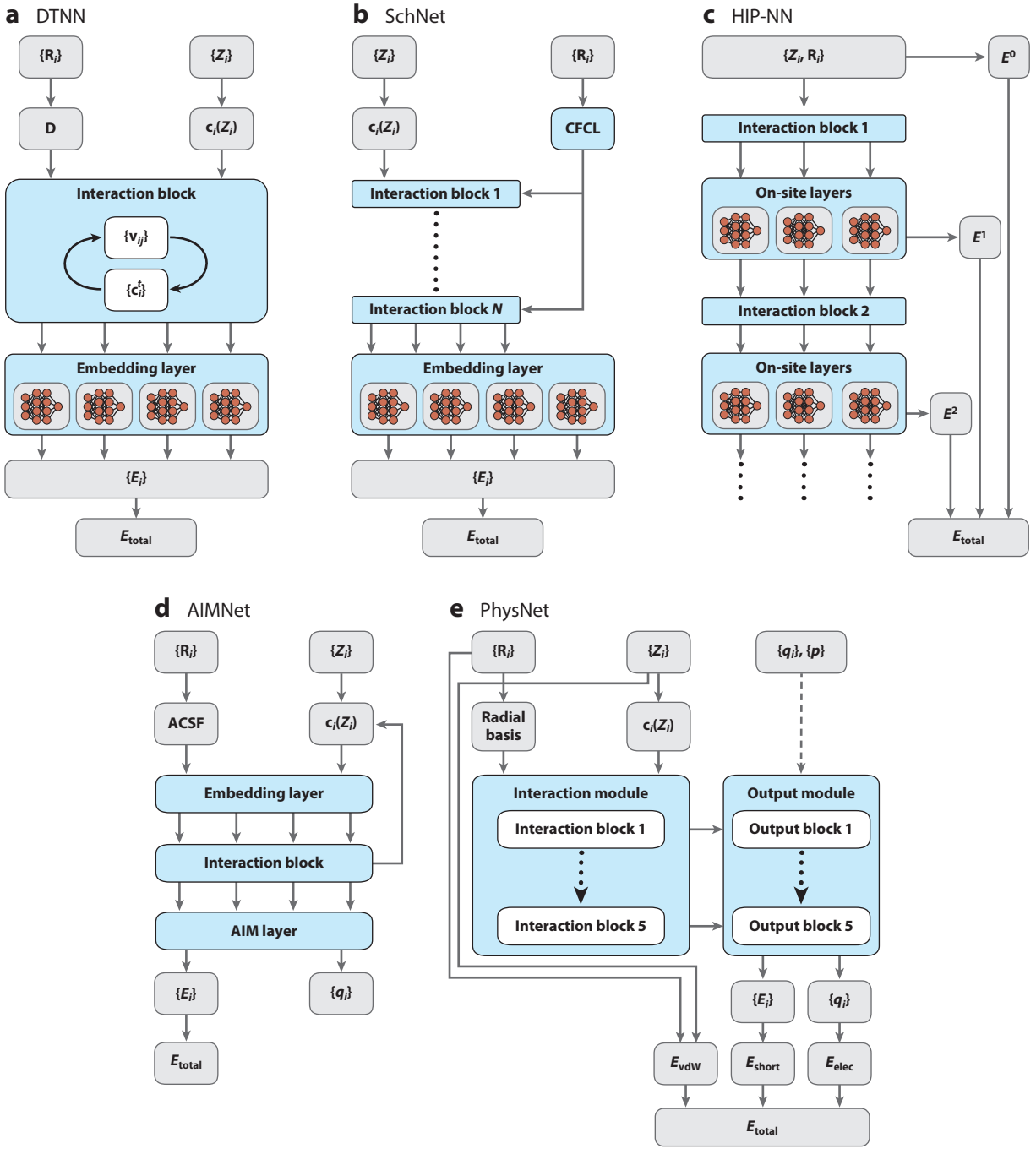


Figure 3

Schematic structure of the message passing neural networks (*a*) DTNN (88), (*b*) SchNet (89), (*c*) HIP-NN (90), (*d*) AIMNet (91), and (*e*) PhysNet (25), discussed in this review in Sections 3.3.1–3.3.4 and 4. Abbreviations: ACSF, atom-centered symmetry function; AIM, atom-in-molecule; CFCL, continuous-filter convolutional layer; DTNN, deep tensor neural network; HIP-NN, hierarchically interacting particle neural network.

nuclear charge-specific vector $\mathbf{c}_i(Z_i)$. The atomic feature vectors are then iteratively refined in an interaction block using message vectors $\mathbf{v}_{ij}$, reflecting the interactions of atom i with all other atoms j by

$$\mathbf{c}_i^{t+1} = \mathbf{c}_i^t + \sum_{j \neq i}^{N_{\text{atoms}}} \mathbf{v}_{ij}. \qquad 7.$$

In total, three of these global refinement steps are carried out in DTNN (88), after which the accuracy has been found to saturate for the investigated systems. The $\mathbf{v}_{ij}$ contain the nonlinear coupling of the neighboring atomic feature vectors and the corresponding interatomic distance vectors, and they are computed in a tensor layer,

$$\mathbf{v_{ij}} = \tanh\left\{\mathbf{W}^{\text{fc}}\left[\left(\mathbf{W}^{\text{cf}}\mathbf{c}_j + \mathbf{b}^{\text{f}_1}\right) \circ \left(\mathbf{W}^{\text{df}}\mathbf{d}_{ij} + \mathbf{b}^{\text{f}_2}\right)\right]\right\}, \qquad 8.$$

where $\circ$ is the element-wise multiplication and $\mathbf{W}^{\text{fc}}$, $\mathbf{W}^{\text{cf}}$, $\mathbf{W}^{\text{df}}$, $\mathbf{b}^{\text{f}_1}$, and $\mathbf{b}^{\text{f}_2}$ are weight and bias matrices to be trained. Consequently, the atomic feature vectors start from the state of free atoms and are then iteratively refined using information about the distances and feature vectors of all other atoms. Once this refinement step through convolutional mapping is complete, the feature vectors are provided to an embedding layer containing atomic neural networks yielding atomic energies as output entering Equation 1.

The accuracy of DTNNs has been demonstrated in compositional space for the atomization energies of optimized molecules, in configurational space for the isomers of composition $C_7O_2H_{10}$, and in conformational space for molecular dynamics trajectories of small molecules in vacuum (88). Due to the use of a complete set of interatomic distances, DTNNs represent a global method applicable to small molecules, but in principle a spatial cutoff for the message vectors could be introduced to extend the method to a second-generation potential.

3.3.2. SchNet. Briefly after the introduction of DTNNs, Schütt and coworkers proposed an improved MPNN called SchNet (24, 89) (see **Figure 3*b***), which shares many properties with the original DTNN approach. In SchNet, so-called continuous-filter convolutional layers (CFCLs) are used to represent the atomic environments, which are continuous generalizations of the discrete convolutional networks typically used in image analysis applications, because in contrast with pixel grids in images, atomic positions are continuous quantities. In this way, the coefficient vector of the Gaussian expansion of the interatomic distances can be replaced by the atomic positions. After the atomic feature vectors have been randomly initialized based on their species, they are distributed to a set of N interaction blocks in which CFCLs act on each local atomic neighborhood through a message passing phase in which they iteratively refine structural representations by processing atomic feature vectors. For the interatomic distances, a cutoff is applied making the method applicable to systems with large number of atoms, including periodic systems like crystals. Like in DTNNs, the parameters of the atomic neural network layers are shared for all atoms, and different elements are distinguished based on their features.

After being refined in the interaction blocks, the atomic feature vectors are used as input of atomic neural networks yielding the atomic energy contributions, which are added to obtain the total energy of the system.

3.3.3. Hierarchically interacting particle neural network. An NNP combining the concepts of MPNNs and conventional many-body expansions is the hierarchically interacting particle neural network (HIP-NN) proposed by Lubbers and coworkers in 2018 (90). Going beyond

Equation 1, the atomic energies E_i are further decomposed into contributions of order n,

$$E_i = \sum_{n=0}^{N_{\text{body}}} E_i^n, \qquad 9.$$

which are learned simultaneously and consistently by a single hierarchical neural network consisting of many layers and intermediate interaction blocks (see **Figure 3*c***). The bases of the structural description are the nuclear charges and those pairwise interatomic distances that are below a predefined cutoff distance. The term $E^0 = \sum E_i^0$ is computed to minimize the least squares error of the training data using nuclear charge information only. The first interaction block corresponds to the first message passing step, in that information between atoms within a cutoff radius is exchanged. The atomic feature vectors are then learned through a sequence of independent atomic on-site layers, the last of which corresponds to a linear regression on the learned atomic descriptor vectors and yields $E^1 = \sum E_i^1$. Then, the second interaction block represents another message passing step, followed by atomic on-site layers to yield features to compute E^2, and so forth. In HIP-NN, this sequence is terminated after the $n = 2$ term. The total energy is then given according to Equation 9 as the sum of all computed terms. As it is assumed that the energy contributions decay with increasing order, the magnitude of the different terms can be used to estimate the uncertainty of the HIP-NN prediction.

3.3.4. AIMNet. Another NNP with learnable descriptors called AIMNet was proposed by Isayev and coworkers in 2019 (91) (see **Figure 3*d***). Starting from the atomic positions and nuclear charges, the atomic feature vectors are constructed by combining two components, which are (*a*) the element-independent geometric description by atomic environment vectors, consisting of modified ACSFs of the ANI HDNNP (71) employing a cutoff, and (*b*) the initial feature vectors, containing the information about the nuclear charges. Therefore, like in DTNN and SchNet, the dimensionality of the resulting atomic feature vectors does not depend on the number of chemical elements in the system. In addition to the atomic feature vectors, atom-pair feature vectors are also used to describe the bonds in the system.

In contrast to what happens in other MPNNs, in AIMNet an embedding layer can be used directly after the first feature vectors have been constructed, because the latter already contain right from the start information about the atomic environments via the ACSFs and do not rely on a preparing message passing step. Like, for example, in SchNet, the effective cutoff of atomic interactions can be increased by multiple repeated updated steps of the feature vectors in interaction blocks. Finally, an atom-in-molecule (AIM) layer allows one to compute the atomic properties.

Apart from the environment-dependent atomic energies yielding the total energy, AIMNet can provide other properties such as atomic partial charges, which can be redistributed within the interaction range covered by the sequence of message passing steps. Typically, in AIMNet the AIM layer is calculated three times, corresponding to the initial evaluation and two subsequent message passing updates of the feature vectors. Long-range interactions like electrostatics are not explicitly included beyond the distance covered by these message passing steps. A recent extension, AIMNet-ME (16), can also be trained to systems in different global charge states by including additional information about the spins and charges in the system.

4. THE THIRD GENERATION: LONG-RANGE INTERACTIONS

Second-generation NNPs, as well as second-generation MLPs in general, have been applied very successfully to a wide range of systems, and they represent today's workhorse methods in machine

learning–based atomistic simulations. Still, the underlying locality approximation and the neglect of interactions beyond the cutoff radius can be expected to result in notable errors for systems in which long-range interactions are important. The most prominent example for such long-range interactions is electrostatics, but also comparably weak dispersion interactions can amount to decisive energy contributions for large systems (92).

A basic prerequisite for considering long-range electrostatic interactions is the availability of atomic partial charges. Starting with the pioneering work of Popelier and coworkers (93, 94), who explored the capabilities of neural networks and other machine learning methods to represent electrostatic mono- and multipoles for the improvement of electrostatics in classical force fields, many groups have proposed to represent atomic properties like charges and multipoles by machine learning (95–101). However, to date only a few NNPs contain long-range electrostatic interactions by explicitly computing the Coulomb interaction without truncation. These potentials define the third generation of NNPs.

The first MLP of the third generation has been the 3G-HDNNP introduced in 2011 (102, 103). In this method, which is an extension of the second-generation HDNNPs proposed by Behler & Parrinello (23) (see Section 3.2.1), in addition to the atomic neural networks yielding the short-range atomic energies, a second set of atomic neural networks is introduced to predict environment-dependent atomic charges, which are then employed in an Ewald summation (104) to compute the electrostatic energy E_{elec}. Often the same ACSF vector $\mathbf{G}_i$ used for describing the atomic environment is employed. The total energy expression of this method is given by

$$E_{\mathrm{total}} = E_{\mathrm{short}} + E_{\mathrm{elec}} = \sum_{i=1}^{N_{\mathrm{atoms}}} E_i\left(\{\mathbf{G}_i\}\right) + \sum_{i>j}^{N_{\mathrm{atoms}}} \frac{q_i(\{\mathbf{G}_i\})q_j(\{\mathbf{G}_j\})}{R_{ij}}. \qquad 10.$$

Since the short-range atomic energies can describe in principle all types of atomic interactions, including electrostatics up to the cutoff radius, double counting of energy terms has to be avoided by a sequential training process. First, the atomic neural networks representing the charges are trained using reference charges from electronic structure calculations. Then, the electrostatic energies and forces are computed and removed from the reference energies and charges to obtain the target properties needed for training the short-range energies. This two-step procedure is also required because the computation of electrostatic forces resulting from environment-dependent charges requires the derivatives of the atomic charges with respect to the atomic positions, which can be provided by the neural networks representing the charges but are not accessible in electronic structure calculations. Another third-generation HDNNP containing long-range electrostatics and also dispersion interactions using Grimme's D2 method (105) is the TensorMol approach introduced in 2018 by Yao and coworkers (106). In this method, the charges are trained by neural networks to reproduce reference dipole moments obtained in electronic structure calculations.

In addition to HDNNPs relying on predefined descriptors, also MPNNs (see Section 3.3) including long-range interactions have been reported. The most important example is PhysNet, introduced by Unke & Meuwly in 2019 (25). This method is closely related to SchNet, but with two important modifications: First, the information flow in the interaction block is enhanced by preactivation residual layers that enable a higher level of expressiveness, and distance-based attention masks are used as messenger functions to increase the efficiency of the learning cycle. Second, long-range interactions and electrostatics are explicitly treated. In PhysNet, atomic coefficient vectors pass through one single refinement box that contains many interaction blocks and are then fed into atomic neural networks to predict the atomic energies and charges, which are in

turn used to compute the long-range electrostatic energy. An important feature of this approach is the simultaneous prediction of both atomic energies and charges by the same neural networks, which increases the computational efficiency of the method (see **Figure 3**).

In spite of these advances, third-generation NNPs are rarely used, because in many condensed systems long-range electrostatic interactions are efficiently screened and contribute very little to the atomic interactions beyond the typical cutoff radii of 6–10 Å. In addition, the explicit computation of the Ewald sum, as well as the introduction of additional neural networks in several methods, substantially increases the computational costs with very small improvements in accuracy.

5. THE FOURTH GENERATION: NONLOCAL INTERACTIONS

5.1. Overview: Nonlocal Dependencies

Even though long-range electrostatic interactions are included in third-generation NNPs, the underlying charges are assumed to be local in that they depend only on the positions of the neighboring atoms up to a cutoff radius. While this is a reasonable assumption for many systems, the locality approximation breaks down for systems in which the atomic partial charges depend on structural features outside the atomic cutoff spheres. These can be distant functional groups in molecules or doping and defects in solid materials. Another typical situation involving nonlocal modifications of the electronic structure is a change in the global charge state of the system—for example, by ionization, protonation, or deprotonation. In these situations, the first three generations of NNPs, which implicitly assume a single fixed total charge, are unable to represent the PES reliably, since it is impossible to represent the atomic partial charges as a function of the local environment only.

Potentials that are able to capture these nonlocal or even global dependencies for high-dimensional systems define fourth-generation NNPs. We note that the terminology in the literature is not consistent, because often there is no clear distinction between long-range interactions and nonlocal interactions. In this review, long-range interactions refer to electrostatics and also to van der Waals interactions, which are not truncated but depend on local properties, like charges, and hence can be described by third-generation NNPs. Nonlocal interactions arise from long-range or even global dependencies in the electronic structure and require fourth-generation NNPs for a qualitatively correct description. In recent years, a few NNPs of the fourth generation have been proposed (see **Figure 4**), which are discussed in the following sections.

5.2. Methods

MLPs of the fourth generation are not commonly used yet, but in recent years several methods have been proposed.

5.2.1. Charge equilibration via neural network technique. The first MLP of the fourth generation has been the charge equilibration via neural network technique (CENT) proposed by Ghasemi and coworkers in 2015 (107). The central idea of CENT is to employ a charge equilibration step (109), which is also used in several advanced force fields and allows the redistribution of the electrons over the whole system to minimize the electrostatic energy. The total energy of CENT is based on a second-order Taylor expansion in the atomic partial charges q_i, and the classical electrostatic energy of the charge density is given by

$$E_{\mathrm{CENT}} = \sum_{i=1}^{N_{\mathrm{atoms}}} \left(E_{i,0} + \chi_i q_i + \frac{1}{2} \left(\eta_i + \frac{2\gamma_{ii}}{\sqrt{\pi}} \right) q_i^2 \right) + \sum_{i>j}^{N_{\mathrm{atoms}}} q_i q_j \frac{\mathrm{erf}(\gamma_{ij} R_{ij})}{R_{ij}}, \qquad 11.$$

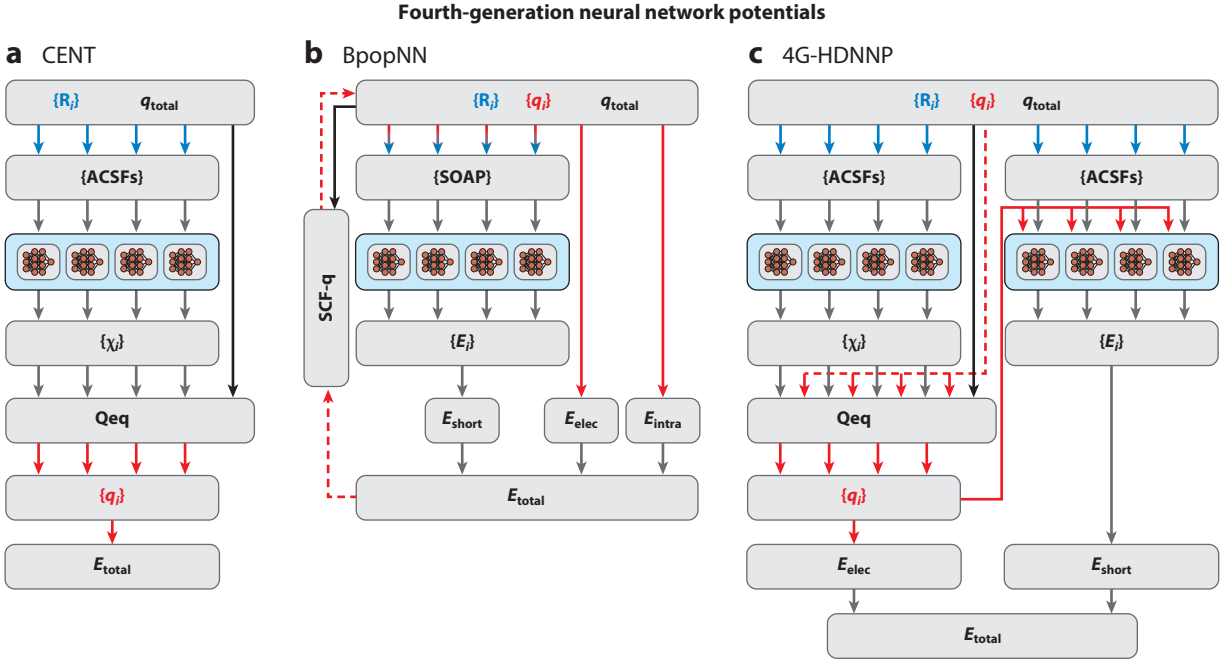


Figure 4

Schematic structure of fourth-generation neural network potentials. (*a*) In the charge equilibration via neural network technique (CENT) (107), the atomic electronegativities χ_i are expressed by atomic neural networks as a function of the local atomic environments described by atom-centered symmetry functions (ACSFs). They are then used in a charge equilibration (Qeq) step to provide globally dependent charges used to minimize the energy in Equation 11. (*b*) In the Becke population neural network (BpopNN) (108), a charge-dependent SOAP descriptor is used to determine the atomic energies. Atomic charges are also used to compute electrostatic and intra-atomic energy terms. When used in predictions, the atomic charges are determined self-consistently (SCF-q) by minimizing the total energy, as indicated by the dotted line. (*c*) In fourth-generation high-dimensional neural network potentials (4G-HDNNP) (38), the atomic charges are determined in a charge equilibration step like in CENT, but the training is done using reference charges, as indicated by the dashed line. Apart from the calculation of the electrostatic energy, the charges are also used as additional nonlocal input information for the atomic energies E_i provided by a second set of atomic neural networks to yield the short-range energy. The total energy is the sum of the short-range and electrostatic energies.

where $E_{i,0}$ are the energies of the free atoms, the coefficients of the linear terms are the atomic electronegativities χ_i, and η_i represent the element-dependent hardness values. The atomic electronegativities are expressed by individual atomic neural networks as a function of the local atomic environments described by ACSFs (see **Figure 4*a***). The parameters γ_{ij} depend on the width of the atomic Gaussian charge densities.

The minimization of this energy expression results in a set of linear equations that in addition contain a constraint for the total charge of the system. This needs to be solved to yield the equilibrated atomic partial charges in the system. The weights of the electronegativity neural networks are optimized during the training process so that the error of E_{CENT} with respect to DFT reference energies is minimized. Like all fourth-generation MLPs, the CENT method allows the construction of potentials that are simultaneously applicable to several global charge states, and due to the charge-based energy expression in Equation 11, the method works best for systems with predominantly ionic bonding.

5.2.2. Becke population neural networks. Another fourth-generation NNP is the Becke population neural network (BpopNN) introduced in 2020 by Xie and coworkers (108). In this method,

for the chosen global charge of the system, the atomic populations (i.e., the partial charges) are adapted self-consistently in a process called SCF-q (see **Figure 4*b***). Starting from an initial guess of the charge distribution, the total energy of the system is minimized with respect to the atomic populations and calculated as

$$E_{\mathrm{BpopNN}} = \sum_{i=1}^{N_{\mathrm{atoms}}} E_i(\mathbf{D}_i) + E_{\mathrm{intra}} + E_{\mathrm{elec}}, \qquad 12.$$

where E_i represent the energy contributions of atoms i determined by atomic neural networks as a function of both the local atomic environments and the atomic populations, which are provided as input in the form of modified SOAP (110) descriptors $\mathbf{D}_i$. The intra-atomic energy E_{intra} is an element-specific quadratic function of the atomic partial charge. The electrostatic energy E_{elec} contains a damping function at very short interatomic distance, facilitating the representation of the atomic short-range energies.

In the training stage, the reference atomic populations, total energies, and atomic forces are obtained from constrained DFT calculations (111) for various charge distributions and states. Then, the atomic neural network weights and element-pair dependent parameters $\kappa(z_i, z_j)$ are adjusted so that the errors of the total energy, the atomic forces, and the population gradients are minimized with respect to the DFT data.

5.2.3. Fourth-generation high-dimensional neural network potentials. Very recently, a fourth generation of HDNNPs (4G-HDNNP) has been proposed (36, 38) that combines the advantages of the CENT method (see Section 5.2.1)—that is, the ability to describe nonlocal charge transfer, global dependencies in the electronic structure, and also the resulting electrostatic interactions—and of second-generation HDNNPs (see Section 3.2.1), which provide a very accurate description of local bonding. The schematic structure of a fourth-generation HDNNP is shown in **Figure 4*c***. Like in CENT, the electrostatic energy is computed from charges obtained in a charge equilibration process based on environment-dependent electronegativities, which are represented by atomic neural networks. In contrast to CENT, in the training process the electronegativities are not adjusted to yield charges minimizing the total energy, but rather the neural network weights are determined to reproduce a set of reference Hirshfeld charges obtained in DFT calculations (see dashed red line in **Figure 4*c***). The long-range energy is then calculated using either Coulomb's law or, in case of periodic systems, an Ewald sum (104).

The second energy contribution, the short-range energy, is given as a sum of atomic energies like in second-generation HDNNPs, but in addition the atomic partial charges of the respective atoms are used as additional input neurons to provide information about the local electronic structure. For the same reasons as in third-generation HDNNPs, the short-range atomic neural networks are trained in a second step after the electrostatic neural networks, and this step is also required because the atomic charges used as additional input for the short-range atomic neural networks need to be available.

The resulting fourth-generation HDNNPs are generally applicable to many types of systems, from organic molecules to ionic solids (38). Like all fourth-generation potentials, they can be constructed to simultaneously describe the different global charge states of a given system.

6. DISCUSSION AND OUTLOOK

Starting with the introduction of second-generation potentials applicable to high-dimensional systems, MLPs for atomistic simulations have attracted significant attention, resulting in a rapid development of this field, which is now very advanced but has not yet reached full maturity. Whereas

initially applications in materials science have been the main driving force for these developments (23), MLPs are now explored and applied to many different fields, and a lot of effort is spent, for example, in the construction of broadly applicable potentials for a wide range of organic molecules (25, 71, 88, 112).

The development of MLPs is full of challenges, some of which have been solved while others require further research. The development of descriptors for high-dimensional systems exactly considering all symmetries and invariances, which in the initial years has been a formidable challenge, can now be considered as accomplished, and there are two alternative approaches based on predefined or learnable descriptors. Both are equally suitable for high-quality potentials, but because the first message passing networks have been introduced only a few years ago, most applications published to date rely on predefined descriptors like ACSFs (67), SOAP (110), and many others (63). Still, a yet unsolved problem is the combinatorial growth in the number of predefined descriptors with the increasing number of elements in the system. First promising steps, like combined descriptors (65) or even combined atomic neural networks (72, 73), have been proposed, and many MPNNs also make use of combined feature vectors (24, 88, 91), but the generality of these methods remains to be explored.

Another challenge, which is addressed by many research teams, is the efficient construction of the reference data sets, which is of crucial importance for obtaining reliable potentials. Because the computationally most expensive part of constructing NNPs is the generation of the reference data by accurate electronic structure calculations, the data sets should be as small as possible while at the same time covering a broad range of structures to obtain transferable potentials. Nowadays, the concept of active learning, which has been known in the machine learning field for a long time (113) and was first used by Artrith & Behler for the construction of an HDNNP for copper (114), is now broadly applied to establish an almost automatic framework for constructing reference data. The central idea of active learning is to identify important missing structures without making an explicit comparison to costly electronic structure calculations but rather using an estimate of the uncertainty of a prediction to select new data points. This can be done, for example, by comparing the energies or forces obtained from different NNPs trained to the same data set, and many different variants for the iterative construction of the reference data by active learning have been proposed (115–122).

Another approach to improve the quality of MLPs with only a moderate amount of high-level data is Δ-learning (123–125). Here, assuming that the difference between two PESs obtained with different electronic structure methods is smooth and can thus be represented by machine learning in a relatively straightforward manner, a baseline potential is obtained using an approximate method, and the difference of this potential from the one obtained with the high-level method is learned by the NNP.

Apart from the optimized selection of the underlying data (74), another main strategy to generate more accurate and transferable potentials is to include the underlying physical laws. An example is the inclusion of long-range electrostatic interactions in third- and fourth-generation NNPs. Whereas most fourth-generation NNPs employ predefined descriptors, novel types of message passing methods are just emerging that aim at describing nonlocal effects (97, 98, 126). Apart from electrostatics, dispersion interactions, which are weak but can be important in large systems, have also been included beyond the local atomic environments in NNPs (25, 106). More recently, information beyond charges, such as atomic spin moments (126, 127) for magnetic systems and external electric fields (128) to study molecular spectra, has also been considered.

A new direction of the field is to combine NNPs with information from electronic structure methods of different levels of complexity. There are several examples, like the combination of neural networks with density-functional tight binding (129) or Hückel theory (130) and the

representation of symmetry-adapted atomic-orbital features (131) by message passing techniques. Furthermore, highly accurate molecular electronic wave functions can be learned by neural networks (132, 133). This trend of including or generating more information at the electronic structure or quantum mechanical level can be expected to continue in the years to come.

In summary, NNPs have become an indispensable tool for atomistic simulations in many fields of chemistry, molecular biology, and materials science. They have been proven to be particularly useful for large-scale molecular dynamics simulations in extensive sampling problems, which require long simulations of extended systems. Because they are able to describe all types of bonding with the same level of accuracy, they are even applicable to systems that are difficult to describe by conventional types of potentials. Still, a main limitation of NNPs, and of MLPs in general, is the restricted transferability beyond the underlying training set. However, it can be anticipated that future generations of potentials will allow this limitation to be overcome by including physical knowledge and laws while maintaining the general applicability that is a major advantage of the family of MLPs.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) (BE3264/13-1, project 411538199).

LITERATURE CITED

1. Charles River Ed. 2018. *Machine Learning: An Overview of Artificial Intelligence*. Scotts Valley, CA: CreateSpace
2. Alpaydin E. 2020. *Introduction to Machine Learning*. Cambridge, MA: MIT Press
3. Mater AC, Coote ML. 2019. Deep learning in chemistry. *J. Chem. Inf. Model.* 59:2545–59
4. Larranaga P, Calvo B, Santana R, Bielza C, Galdiano J, et al. 2005. Machine learning in bioinformatics. *Brief. Bioinform.* 7:86–112
5. von Lilienfeld OA, Burke K. 2020. Retrospective on a decade of machine learning for chemical discovery. *Nat. Commun.* 11:4895
6. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. 2016. Applications of deep learning in biomedicine. *Mol. Pharm.* 13:1445–54
7. Libbrecht MW, Noble WS. 2015. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16(6):321–32
8. Selvaratnam B, Koodali RT. 2021. Machine learning in experimental materials chemistry. *Catal. Today* 371:77–84
9. Agrawal A, Choudhary A. 2016. Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science. *APL Mater.* 4:053208
10. Handley CM, Popelier PLA. 2010. Potential energy surfaces fitted by artificial neural networks. *J. Phys. Chem. A* 114:3371–83
11. Behler J. 2011. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* 13:17930–55
12. Behler J. 2016. Perspective: machine learning potentials for atomistic simulations. *J. Chem. Phys.* 145:170901
13. Dral PO. 2020. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* 11:2336–47

14. Deringer VL, Caro MA, Csányi G. 2019. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* 31:1902765
15. Noé F, Tkatchenko A, Müller KR, Clementi C. 2020. Machine learning for molecular simulation. *Ann. Rev. Phys. Chem.* 71:361–90
16. Zubatiuk T, Isayev O. 2021. Development of multimodal machine learning potentials: toward a physics-aware artificial intelligence. *Acc. Chem. Res.* 54:1575–85
17. Zhang J, Lei YK, Zhang Z, Chang J, Li M, et al. 2020. A perspective on deep learning for molecular modeling and simulations. *J. Phys. Chem. A* 124:6745–63
18. Unke OT, Chmiela S, Sauceda HE, Gastegger M, Poltavsky I, et al. 2021. Machine learning force fields. *Chem. Rev.* 121:10142–86
19. Friederich P, Häse F, Proppe J, Aspuru-Guzik A. 2021. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* 20(6):750–61
20. Grisafi A, Nigam J, Ceriotti M. 2021. Multi-scale approach for the prediction of atomic scale properties. *Chem. Sci.* 12:2078–90
21. Blank TB, Brown SD, Calhoun AW, Doren DJ. 1995. Neural network models of potential energy surfaces. *J. Chem. Phys.* 103(10):4129–37
22. Lorenz S, Groß A, Scheffler M. 2004. Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks. *Chem. Phys. Lett.* 395:210–15
23. Behler J, Parrinello M. 2007. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* 98:146401
24. Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR. 2018. SchNet—a deep learning architecture for molecules and materials. *J. Chem. Phys.* 148:241722
25. Unke OT, Meuwly M. 2019. PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* 15:3678–93
26. Bartók AP, Payne MC, Kondor R, Csányi G. 2010. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* 104:136403
27. Bartók AP, Csányi G. 2015. Gaussian approximation potentials: a brief tutorial introduction. *Int. J. Quant. Chem.* 115:1051–57
28. Chmiela S, Sauceda HE, Poltavsky I, Müller KR, Tkatchenko A. 2019. sGDML: constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* 240:38–45
29. Thompson AP, Swiler LP, Trott CR, Foiles SM, Tucker GJ. 2015. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* 285:316–30
30. Wood MA, Thompson AP. 2018. Extending the accuracy of the SNAP interatomic potential form. *J. Chem. Phys.* 148:241721
31. Shapeev AV. 2016. Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* 14:1153–73
32. Drautz R. 2019. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* 99:014104
33. Allen A, Dusson G, Ortner C, Csanyi G. 2021. Atomic permutationally invariant polynomials for fitting molecular force fields. *Mach. Learn. Sci. Techn.* 2:025017
34. Balabin RM, Lomakina EI. 2011. Support vector machine regression (LS-SVM)—an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? *Phys. Chem. Chem. Phys.* 13:11710
35. Cybenko G. 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* 2:303–14
36. Ko TW, Finkler JA, Goedecker S, Behler J. 2021. General-purpose machine learning potentials capturing nonlocal charge transfer. *Acc. Chem. Res.* 54:808–17
37. Behler J. 2021. Four generations of high-dimensional neural network potentials. *Chem. Rev.* 121:10037–72
38. Ko TW, Finkler JA, Goedecker S, Behler J. 2021. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* 12:398
39. Rupp M, Tkatchenko A, Müller KR, von Lilienfeld OA. 2012. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* 108:058301

40. Zeng M, Kumar JN, Zeng Z, Savitha R, Chandrasekhar VR, Hippalgaonkar K. 2018. Graph convolutional neural networks for polymers property prediction. arXiv:1811.06231 [cond-mat.mtrl-sci]
41. Omprakash P, Manikandan B, Sandeep A, Shrivastava R, Viswesh P, Panemangalore DB. 2021. Graph representational learning for bandgap prediction in varied perovskite crystals. *Comput. Mater. Sci.* 196:110530
42. Miccio LA, Schwartz GA. 2020. From chemical structure to quantitative polymer properties prediction through convolutional neural networks. *Polymer* 193:122341
43. Behler J. 2017. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Ed.* 56:12828
44. Brown DFR, Gibbs MN, Clary DC. 1996. Combining *ab initio* computations, neural networks, and diffusion Monte Carlo: an efficient method to treat weakly bound molecules. *J. Chem. Phys.* 105:7597
45. No KT, Chang BH, Kim SY, Jhon MS, Scheraga HA. 1997. Description of the potential energy surface of the water dimer with an artificial neural network. *Chem. Phys. Lett.* 271:152–56
46. Bittencourt ACP, Prudente FV, Vianna JDM. 2004. The fitting of potential energy and transition moment functions using neural networks: transition probabilities in OH ($A^2\sum^+$ to $X^2\Pi$). *Chem. Phys.* 297:153–61
47. Lee HM, Raff LM. 2008. Cis → trans, trans → cis isomerizations and N–O bond dissociation of nitrous acid (HONO) on an ab initio potential surface obtained by novelty sampling and feed-forward neural network fitting. *J. Chem. Phys.* 128:194310
48. Behler J, Reuter K, Scheffler M. 2008. Nonadiabatic effects in the dissociation of oxygen molecules at the Al(111) surface. *Phys. Rev. B* 77:115421
49. Goikoetxea I, Beltrán J, Meyer J, Juaristi JI, Alducin M, Reuter K. 2012. Non-adiabatic effects during the dissociative adsorption of O_2 at Ag(111)? A first-principles divide and conquer study. *New J. Phys.* 14:013050
50. Manzhos S, Yamashita K. 2010. A model for the dissociative adsorption of N_2O on Cu(100) using a continuous potential energy surface. *Surf. Sci.* 604:554–60
51. Cho KW, No KT, Scheraga HA. 2002. A polarizable force field for water using an artificial neural network. *J. Mol. Struct.* 641:77–91
52. Gassner H, Probst M, Lauenstein A, Hermansson K. 1998. Representation of intermolecular potential functions by neural networks. *J. Phys. Chem. A* 102:4596–605
53. Hobday S, Smith R, Belbruno J. 1999. Applications of neural networks to fitting interatomic potential functions. *Model. Simul. Mater. Sci. Eng.* 7:397–412
54. Manzhos S, Carrington T Jr. 2006. Using neural networks to represent potential surfaces as sums of products. *J. Chem. Phys.* 125:194105
55. Manzhos S, Carrington T Jr. 2007. Using redundant coordinates to represent potential energy surfaces with lower-dimensional functions. *J. Chem. Phys.* 127:014103
56. Manzhos S, Carrington T Jr. 2008. Using neural networks, optimized coordinates, and high-dimensional model representations to obtain a vinyl bromide potential surface. *J. Chem. Phys.* 129:224104
57. Malshe M, Narulkar R, Raff LM, Hagan M, Bukkapatnam S, et al. 2009. Development of generalized potential-energy surfaces using many-body expansions, neural networks, and moiety energy approximations. *J. Chem. Phys.* 130:184102
58. Brown A, Braams BJ, Christoffel K, Jin Z, Bowman JM. 2003. Classical and quasiclassical spectral analysis of CH_5^+ using an *ab initio* potential energy surface. *J. Chem. Phys.* 119:8790
59. Braams BJ, Bowman JM. 2009. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* 28:577–606
60. Jiang B, Guo H. 2013. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. *J. Chem. Phys.* 139:054112
61. Li J, Jiang B, Guo H. 2013. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems. *J. Chem. Phys.* 139:204103
62. Behler J, Lorenz S, Reuter K. 2007. Representing molecule-surface interactions with symmetry-adapted neural networks. *J. Chem. Phys.* 127:014705
63. Langer MF, Goessmann A, Rupp M. 2020. Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning. arXiv:2003.12081 [physics.comp-ph]

64. Himanen L, Jägera MOJ, Morooka EV, Canova FF, Ranawat YS, et al. 2020. DScribe: library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* 247:106949
65. Gastegger M, Schwiedrzik L, Bittermann M, Berzsenyi F, Marquetand P. 2018. WACSF—weighted atom-centered symmetry functions as descriptors in machine learning potentials. *J. Chem. Phys.* 148:241709
66. Zhang Y, Hu C, Jiang B. 2019. Embedded atom neural network potentials: efficient and accurate machine learning with a physically inspired representation. *J. Phys. Chem. Lett.* 10:4962–67
67. Behler J. 2011. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* 134:074106
68. Behler J. 2015. Constructing high-dimensional neural network potentials: a tutorial review. *Int. J. Quantum Chem.* 115:1032–50
69. Singraber A, Morawietz T, Behler J, Dellago C. 2019. Parallel multi-stream training of high-dimensional neural network potentials. *J. Chem. Theory Comput.* 15:3075–92
70. Khorshidi A, Peterson AA. 2016. Amp: a modular approach to machine learning in atomistic simulations. *Comput. Phys. Commun.* 207:310–24
71. Smith JS, Isayev O, Roitberg AE. 2017. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* 8:3192–203
72. Liu M, Kitchin JR. 2020. SingleNN: a modified Behler-Parrinello neural network with shared weights for atomistic simulations with transferability. *J. Phys. Chem. C* 124:17811–18
73. Profitt TA, Pearson JK. 2019. A shared-weight neural network architecture for predicting molecular properties. *Phys. Chem. Chem. Phys.* 21:26175
74. Imbalzano G, Anelli A, Giofre D, Klees S, Behler J, Ceriotti M. 2018. Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. *J. Chem. Phys.* 148:241730
75. Mahoney MW, Drineas P. 2009. CUR matrix decompositions for improved data analysis. *PNAS* 106:697–702
76. Selvaratnam B, Koodali RT, Miro P. 2020. Application of symmetry functions to large chemical spaces using a convolutional neural network. *J. Chem. Inf. Model.* 160:1928–35
77. Jovan Jose KV, Artrith N, Behler J. 2012. Construction of high-dimensional neural network potentials using environment-dependent atom pairs. *J. Chem. Phys.* 136:194111
78. Yao K, Herr JE, Brown SN, Parkhill J. 2017. Intrinsic bond energies from a bonds-in-molecules neural network. *J. Phys. Chem. Lett.* 8:2689–94
79. Hansen K, Biegler F, Ramakrishnan R, Pronobis W, von Lilienfeld OA, et al. 2015. Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.* 6:2326–31
80. Glick ZL, Metcalf DP, Koutsoukas A, Spronk SA, Cheney DL, Sherrill CD. 2020. AP-Net: an atomic-pairwise neural network for smooth and transferable interaction potentials. *J. Chem. Phys.* 153:044112
81. Zhang L, Han J, Wang H, Car R, E W. 2018. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* 120:143001
82. Han J, Zhang L, Car R, E W. 2018. Deep potential: a general representation of a many-body potential energy surface. *Commun. Comput. Phys.* 23:629–39
83. Zhang L, Han J, Wang H, Saidi WA, Car R, E W. 2018. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ed. S Bengio, HM Wallach, H Larochelle, K Grauman, N Cesa-Bianchi, pp. 4441–51. New York: ACM
84. Daw M, Foiles S, Baskes M. 1993. The embedded-atom method: a review of theory and applications. *Mater. Sci. Rep.* 9:251
85. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, Gomez-Bombarelli R, Hirzel T, et al. 2015. Convolutional networks on graphs for learning molecular fingerprints. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, ed. C Cortes, DD Lee, M Sugiyama, R Garnett, pp. 2224–32. New York: ACM
86. Rogers D, Hahn M. 2010. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* 50:742–54

87. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. 2017. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 1263–72. Sydney, Aust.: JMLR
88. Schütt KT, Arbabzadah F, Chmiela S, Müller KR, Tkatchenko A. 2017. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* 8:13890
89. Schütt K, Kindermans PJ, Felix HES, Chmiela S, Tkatchenko A, Müller KR. 2017. SchNet: a continuous-filter convolutional neural network for modeling quantum interactions. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, ed. U von Luxburg, I Guyon, S Bengio, H Wallach, R Fergus, pp. 992–1002. New York: ACM
90. Lubbers N, Smith JS, Barros K. 2018. Hierarchical modeling of molecular energies using a deep neural network. *J. Chem. Phys.* 148:241715
91. Zubatyuk R, Smith JS, Leszczynski J, Isayev O. 2019. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* 5:eaav6490
92. Kronik L, Tkatchenko A. 2014. Understanding molecular crystals with dispersion-inclusive density functional theory: pairwise corrections and beyond. *Acc. Chem. Res.* 47:3208–16
93. Houlding S, Liem SY, Popelier PLA. 2007. A polarizable high-rank quantum topological electrostatic potential developed using neural networks: molecular dynamics simulations on the hydrogen fluoride dimer. *Int. J. Quantum Chem.* 107:2817–27
94. Handley CM, Popelier PLA. 2009. Dynamically polarizable water potential based on multipole moments trained by machine learning. *J. Chem. Theory Comput.* 5:1474–89
95. Nebgen B, Lubbers N, Smith JS, Sifain A, Lokhov A, et al. 2018. Transferable dynamic molecular charge assignment using deep neural networks. *J. Chem. Theory Comput.* 14:4687–98
96. Gastegger M, Behler J, Marquetand P. 2017. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* 8:6924
97. Zubatyuk R, Smith JS, Nebgen BT, Tretiak S, Isayev O. 2021. Teaching a neural network to attach and detach electrons from molecules. *Nat. Commun.* 12:4870
98. Metcalf DP, Jiang A, Spronk SA, Cheney DL, Sherrill CD. 2021. Electron-passing neural networks for atomic charge prediction in systems with arbitrary molecular charge. *J. Chem. Inf. Model.* 61:115–22
99. Cuevas-Zuvira B, Pacios LF. 2021. Machine learning of analytical electron density in large molecules through message-passing. *J. Chem. Inf. Model.* 61:2658–66
100. Wang Y, Fass J, Stern CD, Luo K, Chodera J. 2019. Graph nets for partial charge prediction. arXiv:1909.07903 [physics.comp-ph]
101. Wang J, Cao D, Tang C, Xu L, He Q, et al. 2021. DeepAtomicCharge: a new graph convolutional network-based architecture for accurate prediction of atomic charges. *Brief. Bioinform.* 22(3):bbaa183
102. Artrith N, Morawietz T, Behler J. 2011. High-dimensional neural-network potentials for multicomponent systems: applications to zinc oxide. *Phys. Rev. B* 83:153101
103. Morawietz T, Sharma V, Behler J. 2012. A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges. *J. Chem. Phys.* 136:064103
104. Ewald PP. 1921. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* 64:253–87
105. Grimme S. 2006. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* 27(15):1787–99
106. Yao K, Herr JE, Toth DW, Mckintyre R, Parkhill J. 2018. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* 9:2261–69
107. Ghasemi SA, Hofstetter A, Saha S, Goedecker S. 2015. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Phys. Rev. B* 92:045131
108. Xie X, Persson KA, Small DW. 2020. Incorporating electronic information into machine learning potential energy surfaces via approaching the ground-state electronic energy as a function of atom-based electronic populations. *J. Chem. Theory Comput.* 16:4256–70
109. Rappe AK, Goddard WA III. 1991. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* 95:3358
110. Bartók AP, Kondor R, Csányi G. 2013. On representing chemical environments. *Phys. Rev. B* 87:184115

111. Kaduk B, Kowalczyk T, Voorhis TV. 2011. Constrained density functional theory. *Chem. Rev.* 112:321–70
112. Devereux C, Smith JS, Huddleston KK, Barros K, Zubatyuk R, et al. 2020. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.* 16:4192–202
113. Seung HS, Opper M, Sompolinsky H. 1992. Query by committee. In *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pp. 287–94. New York: ACM
114. Artrith N, Behler J. 2012. High-dimensional neural network potentials for metal surfaces: a prototype study for copper. *Phys. Rev. B* 85:045439
115. Podryabinkin EV, Shapeev AV. 2017. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* 140:171–80
116. Loeffler TD, Manna S, Patra TK, Chan H, Narayanan B, Sankaranarayanan S. 2020. Active learning a neural network model for gold clusters & bulk from sparse first principles training data. *ChemCatChem* 12:4796–806
117. Zhang L, Lin DY, Wang H, Car R, E W. 2019. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Mater.* 3:023804
118. Sivaraman G, Krishnamoorthy AN, Baur M, Holm C, Stan M, et al. 2020. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *NPJ Comput. Mater.* 6:104
119. Smith JS, Nebgen B, Lubbers N, Isayev O, Roitberg AE. 2018. Less is more: sampling chemical space with active learning. *J. Chem. Phys.* 148:241733
120. Schran C, Behler J, Marx D. 2020. Automated fitting of neural network potentials at coupled cluster accuracy: protonated water clusters as testing ground. *J. Chem. Theory Comput.* 16:88–99
121. Lin Q, Zhang L, Zhang Y, Jiang B. 2021. Searching configurations in uncertainty space: active learning of high-dimensional neural network reactive potentials. *J. Chem. Theory Comput.* 17:2691–701
122. Bernstein N, Csányi G, Deringer VL. 2019. De novo exploration and self-guided learning of potential-energy surfaces. *NPJ Comput. Mater.* 5:99
123. Sun G, Sautet P. 2019. Toward fast and reliable potential energy surfaces for metallic Pt clusters by hierarchical delta neural networks. *J. Chem. Theory Comput.* 15(10):5614–27
124. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA. 2015. Big data meets quantum chemistry approximations: the delta-machine learning approach. *J. Chem. Theory Comput.* 11:2087–96
125. Dral PO, Owens A, Dral A, Csányi G. 2020. Hierarchical machine learning of potential energy surfaces. *J. Chem. Phys.* 152:204110
126. Unke OT, Chmiela S, Gastegger M, Schütt KT, Sauceda HE, Müller KR. 2021. SpookyNet: learning force fields with electronic degrees of freedom and nonlocal effects. arXiv:2105.00304 [physics.chem-ph]
127. Eckhoff M, Behler J. 2021. High-dimensional neural network potentials for magnetic systems using spin-dependent atom-centered symmetry functions. arXiv:2104.14439 [physics.comp-ph]
128. Gastegger M, Schütt KT, Müller KR. 2020. Machine learning of solvent effects on molecular spectra and reactions. arxiv:2010.14942 [physics.chem-ph]
129. Li H, Collins C, Tanha M, Gordon GJ, Yaron DJ. 2018. A density functional tight binding layer for deep learning of chemical Hamiltonians. *J. Chem. Theory Comput.* 14:5764–76
130. Zubatyuk T, Nebgen B, Lubbers N, Smith JS, Zubatyuk R, et al. 2019. Machine learned Hückel theory: interfacing physics and deep neural networks. arXiv:1909.12963v1 [cond-mat.dis-nn]
131. Qiao Z, Welborn M, Anandkumar A, Manby FR, Miller TF III. 2020. OrbNet: deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* 153:124111
132. Pfau D, Spencer JS, Matthews AGDG, Foulkes WMC. 2020. *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks. *Phys. Rev. Res.* 2:033429
133. Hermann J, Schätzle Z, Noé F. 2020. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* 12:891–97

Annual Review of Physical Chemistry
Volume 73, 2022

Contents

Errata

An online log of corrections to *Annual Review of Physical Chemistry* articles may be found at http://www.annualreviews.org/errata/physchem