Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Digital Discovery*, 2023, **2**, 28

Neural network potentials for chemistry: concepts, applications and prospects

Silvan Käser, Luis Itza Vazquez-Salazar, Markus Meuwly * and Kai Töpfer *

Artificial Neural Networks (NN) are already heavily involved in methods and applications for frequent tasks in the field of computational chemistry such as representation of potential energy surfaces (PES) and spectroscopic predictions. This perspective provides an overview of the foundations of neural network-based full-dimensional potential energy surfaces, their architectures, underlying concepts, their representation and applications to chemical systems. Methods for data generation and training procedures for PES construction are discussed and means for error assessment and refinement through transfer learning are presented. A selection of recent results illustrates the latest improvements regarding accuracy of PES representations and system size limitations in dynamics simulations, but also NN application enabling direct prediction of physical results without dynamics simulations. The aim is to provide an overview for the current state-of-the-art NN approaches in computational chemistry and also to point out the current challenges in enhancing reliability and applicability of NN methods on a larger scale.

Received 23rd September 2022
Accepted 20th December 2022

DOI: 10.1039/d2dd00102k

rsc.li/digitaldiscovery

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1 Introduction

The *in silico* modeling of chemical and biological processes at a molecular level is of central importance in today's research and will be crucial for future challenges of mankind.[1] The modeling often requires a trade-off between accuracy and computational cost: quantum chemical calculations (*e.g. ab initio* molecular dynamics), at a high level of theory, can be very accurate but also come at a high computational cost rendering the approach impractical except for rather small molecules. Empirical force fields, on the other hand, provide a computationally advantageous approach that scales well with system size but the possibility to carry out quantitative studies is limited due to the assumptions underlying their formulation. Thus, computationally efficient and accurate modelling techniques are required for quantitative molecular simulations.[2]

In this regard, Machine Learning (ML) techniques have emerged as a powerful tool to satisfy such demands for force field models which are limited, in principle, by the accuracy of *ab initio* methods and allow an efficiency approaching that of empirical force fields.[3] Motivated by the advances in computational chemistry techniques and the continuous growth of the performance of computer hardware (Moore's law[4]), ML is becoming a daily tool for modeling molecules and materials. By definition, ML methods are data-driven algorithms based on statistical learning theory with the aim of generating numerical methods that generalize to new data, not used in the learning process.[5,6] This capability renders ML methods highly appealing for modelling molecular systems. It even reaches levels where some authors believe that the use of ML techniques will constitute the "fourth paradigm of science",[7] bridging the gap from atomic-scale molecular properties towards macroscopic properties of materials[8,9] and one of the drivers for a revolution of the simulation techniques of matter.[10] The enthusiasm is reflected in the appearance of an extensive number of ML models and their application in computational chemistry.

Some of the most important publications have focused on the study of potential energy surfaces (PESs), which contain all the information about the many-body interactions of a molecular system including stable and metastable structures.[11] At the same time, it is possible to extract a considerable amount of information from PESs including the atomic forces driving the dynamics of molecular systems, reactions and structural transitions, and atomic vibrations.[12] Additionally, it has been proposed that the chemical information contained in a chemical bond, therefore in the PES, can help in the exploration of chemical space.[13] In a recent work,[14] it was found that the exploration of chemical space can be improved by adding adequate information from the configurational space represented by the PES.

Over the past several decades several ML-based methods have been used to represent continuous PESs.[3,15–17] While a number of those are briefly mentioned below, the focus of the present work is on NN-based approaches. Kernel-based methods provide an efficient solution to highly non-linear optimization problems[17] by finding a representation of the problem which encodes the distribution of the data in

Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland. E-mail: m.meuwly@unibas.ch; kai.toepfer@unibas.ch

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

a complete, unique and efficient way.[18] There is a large number of possible representations of chemical space that can be used in kernel methods. Examples include Coulomb Matrices,[19] Bag of Bonds (BoB),[20] Histograms of Distance, Angles and Dihedrals (HDAD),[21] Spectrum of London and Axilrod–Teller–Muto (SLATM),[22] Faber–Christensen–Huang–von Lilienfeld (FCHL)[23] and Smooth Overlap of Atomic Positions (SOAP).[24] A comprehensive review of representations for kernel and non-kernel methods can be found in ref. 25. It should be noted that variations of kernel methods, such as for Gaussian processes[26] which assume a Bayesian/probabilistic point of view for the solution of the problem or the reproducing kernel Hilbert space (RKHS) method[27,28] which uses polynomials as support functions have been extensively discussed in the literature. While the remainder of the perspective is mainly dedicated to NN-based approaches, many alternative interpolation and representation methods for PES construction exist. These include, *e.g.* modified Shepard interpolation,[29] (interpolative) moving least-squares,[30–32] permutationally invariant polynomial (PIP) PESs by least-squares fitting,[33] or least absolute shrinkage and selection operator (LASSO) constrained least-squares.[34] Several of these approaches have been recently described, reviewed and compared.[3,35,36]

NNs are inspired by the biological model of the intricate networks formed by the brain and how information is passed.[37] The ideas underlying NNs date back to 1960 when "the perceptron" was presented by Rosenblatt.[38] However, computational and theoretical limitations inhibited the development of NNs.[39,40] It was not until 1970 with the development of the automatic differentiation and the introduction of back-propagation[41] that NN models continued to develop. Still, large scale applications were rare until the beginning of the 21st century when considerably more powerful computer hardware became available. In chemistry, the application of NN models dates back to 1990s with first applications in analytical and medicinal chemistry.[42,43] Regarding PES representation, the first application of NNs can be tracked back to the same decade.[44,45] Nowadays, NNs are the most common ones from the field of ML models for the use in chemistry-related applications that are focused on the generation and study of PESs. Some examples of popular NN-based schemes for PES fitting include the High Dimensional Neural Network (HDNN) method,[46,47] Deep Tensor Neural Network (DTNN),[48] SchNet,[49] ANI,[50] or PhysNet,[51] among others.

The purpose of the present perspective is to provide a birds-eye view and an outlook into the conception, generation and use of NN based PESs for the exploration of chemical systems. Additionally, we will present some of the current challenges in the development and application of NN models for the study of PESs. The remainder of the present work is structured as follows. A brief introduction to the theoretical background of PESs and NNs is provided in Section 2. Section 3 discusses existing NN architectures with emphasis on structural information and current developments in the field. Section 4 describes the construction of a PES from the initial sampling to the validation and refinement of the generated models and Section 5 discusses knowledge transfer that allows obtaining PESs at high levels of theory with less data. Selected applications for chemical systems showcasing the concepts introduced and including NN models in established atomistic dynamics models are described in Section 6. Applications of NN models that skip dynamics simulation to predict physical observables are shown in Section 7. Section 8 describes some of the current challenges that we consider critical for the development and enhancement of the current models and the field in general, followed by a short conclusion.

2 Theoretical background

This section introduces the concept of PESs, the principles underlying NNs, their building blocks, such as dense layers and activation functions. A more in-depth overview of descriptors for chemical structures and representative examples of frequently used neural network potentials (NNPs) is given in the next section. In terms of nomenclature, italic symbols denote scalars or functions and bold symbols are n – dimensional tensors ($n \geq 1$) with the special case of a one-dimensional spatial vector (*e.g.* position or distance) denoted as italic symbol with vector arrow.

2.1 Potential energy surfaces

The energetics of a molecular system can be described by solving the electronic Schrödinger Equation (SE). Unfortunately, the SE can only be solved exactly for simple, single-electron atomic systems. In order to obtain solutions for many-electron systems, it is necessary to introduce approximations. The Born–Oppenheimer approximation (BOA),[52] also called the most important approximation in quantum chemistry,[53] assumes that the coupling between the nuclear and electronic motion can be neglected because the mass of the nuclei is several orders of magnitude larger than the mass of the electrons. Under this assumption, it is possible to rewrite the total wavefunction Ψ, which is a solution of the SE, as the product of a nuclear wavefunction $\chi(\mathbf{R})$ with nuclear positions $\mathbf{R}$ and the electronic wavefunction $\psi(\mathbf{r};\mathbf{R})$ with electron coordinates $\mathbf{r}$ for a fixed configuration of nuclear positions

$$\Psi(\mathbf{r},\mathbf{R}) = \psi(\mathbf{r};\mathbf{R})\cdot\chi(\mathbf{R}). \quad (1)$$

As a consequence, the electronic wavefunction can be obtained by solving the electronic time-independent SE:

$$\hat{H}_e\psi_\lambda(\mathbf{r};\mathbf{R}) = \left[\hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee}\right]\psi_\lambda(\mathbf{r};\mathbf{R}) = \epsilon_\lambda(\mathbf{R})\psi_\lambda(\mathbf{r};\mathbf{R}) \quad (2)$$

Here, $\hat{H}_e$ is the electronic (spin-free) Hamiltonian describing the kinetic energy of the electrons $\hat{T}_e$, the Coulomb interaction between the nuclear and electron charges $\hat{V}_{ne}$ and the electron–electron interaction $\hat{V}_{ee}$. The solution is the electronic wavefunction ψ_λ and electronic energy ϵ_λ for the electronic state λ. The so-called adiabatic PES of an atomic system $E_\lambda^{BO}(\mathbf{R})$ in electronic state λ constitutes an effective potential for the nuclear dynamics. It is obtained by the sum of the Coulomb repulsion V_{nn} between the nuclei with nuclear charge Z_i for the

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

total number of atoms N, and the respective electronic energy at the associated nuclear positions.[54]

$$E_\lambda^{\mathrm{BO}}(\mathbf{R}) = V_{\mathrm{nn}}(\mathbf{R}) + \epsilon_\lambda(\mathbf{R}) \tag{3}$$

Eqn (3) defines a PES as a $(3N - 6)$ – dimensional function that can be approximated as an analytical function which is, however, a challenging task. Often, one can only report low-dimensional cuts of such high-dimensional hypersurfaces and one example is shown in Fig. 1. Alternatively, eqn (3) suggests that there should be a mapping between the total electronic energy of a molecular system and the combination of position of the nuclei and the set of nuclear charges $\{Z_i\}_{i=1}^N$. This is the starting point for a ML-based approach described in the following.

PESs lie at the heart of computational chemistry.[55] From the relationship between structure and potential energy E, it is possible to derive many molecular properties by taking derivatives with respect to a perturbation such as atomic positions $\mathbf{R}$, an external electric $\vec{\mathscr{E}}$ or magnetic field $\vec{\mathscr{B}}$, which require additional coupling terms in the Hamiltonian and an analytical representation of the PES.[54] Following this, a general response property takes the form

$$\text{Property} \propto \frac{\partial^{n+m+l} E}{\partial \mathbf{R}^n \partial \vec{\mathscr{E}}^m \partial \vec{\mathscr{B}}^l} \tag{4}$$

where n, m, l indicate the order of the derivative with respect to the perturbation. Derivatives of eqn (4) provide, *e.g.*, the forces $\mathbf{F} = -\partial E/\partial \mathbf{R}$ that constitute the foundation of MD simulations and structure optimization schemes. The second derivatives $\partial^2 E/\partial \mathbf{R}^2$ gives access to the Hessian matrix from which the harmonic frequencies of molecular vibrations can be obtained. Other properties such as the dipole moment ($\vec{\mu} = -\partial E/\partial \vec{\mathscr{E}}$) or the molecular polarizability ($\vec{\alpha} = -\partial^2 E/\partial \vec{\mathscr{E}}^2$) are directly related to experimental observables such as the Infrared (IR) or Raman spectra.[56] Mixed derivatives also provide IR absorption intensities ($\partial^2 E/\partial \vec{\mathscr{E}} \partial \mathbf{R}$) or the optical rotation in circular dichroism ($\partial^2 E/\partial \vec{\mathscr{E}} \partial \vec{\mathscr{B}}$).

Given the versatility and usefulness of PESs, a wealth of approaches to construct PESs have been designed over the years and new ML schemes are proposed with high frequency. Especially NNs have been shown to be general function approximators[57,58] by the universal approximation theorem[59] and hence seem particularly useful to learn intricate relationships such as the PES or even external perturbations.

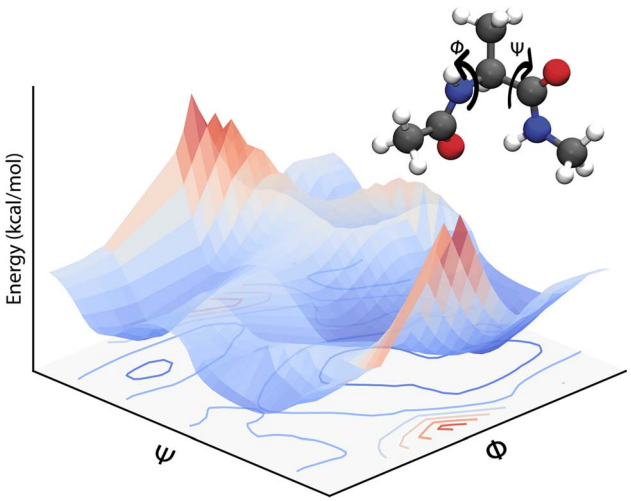

Fig. 1 A two-dimensional PES for the dialanine molecule calculated at the MP2 level with the 6-31G** basis set along dihedral angles Φ and Ψ. A representation of the molecule (ball and stick) indicating the dihedral angles (Φ, Ψ) calculated is given as well. The bottom gives the projection of the 2D PES.

2.2 Artificial neural networks

Artificial NNs (NNs, henceforth) represent a family of computer algorithms and form a subgroup of ML. Nowadays, NNs are applied in diverse areas including, among others, health care,[60] medical imaging,[61] self-driving cars,[62] high-energy physics,[63] particle physics and cosmology,[64] genetics,[65] chemical discovery,[66] reaction planning.[67,68]

Typically, a NN consists of an input layer, a predefined number of hidden layers and an output layer (see Fig. 2A). Deep NNs comprise a larger number of hidden layers while a NN with only one or two hidden layers is a shallow NN. Each layer contains a defined number of nodes (or neurons) that connect to the nodes of the following layer and each connection is associated with weights and biases.

The elementary units of NNs are so-called dense layers, which linearly transform an input vector $\mathbf{x}$ to an output vector $\mathbf{y}$ according to

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}. \tag{5}$$

Here, $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{N,M}$ and $\mathbf{b} = \{b_i\}_{i=1}^N$ are the weights (a matrix) and biases (a vector),[3] M is the dimension of the input and N the number of nodes. The combination of a dense layer with a nonlinear activation function (Fig. 2B) transforms the input $\mathbf{x}$ to an output $\mathbf{y}$ that serves as "input" to the following (hidden) layer.

$$\mathbf{h}_i = \sigma(\mathbf{W}_i\mathbf{x} + \mathbf{b}_i) \tag{6}$$

Modelling non-linear relationships requires the combination of at least two dense layers with an activation function σ according to

$$\mathbf{y} = \mathbf{W}_{i+1}\sigma(\mathbf{W}_i\mathbf{x} + \mathbf{b}_i) + \mathbf{b}_{i+1} = \mathbf{W}_{i+1}\mathbf{h}_i + \mathbf{b}_{i+1} \tag{7}$$

While such shallow architectures are in principle capable of modelling any functional relationship, deeper variants thereof are usually preferred due to improved performance and parameter-efficiency.[69–72] The functional form of the NN is characterized by the number of layers L and number of nodes N in a given layer. With increasing L and N the functional form becomes more flexible, however, overfitting requires careful attention since the obtained form has no underlying physical meaning.[73] A fully connected deep NN is given by the following relation

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

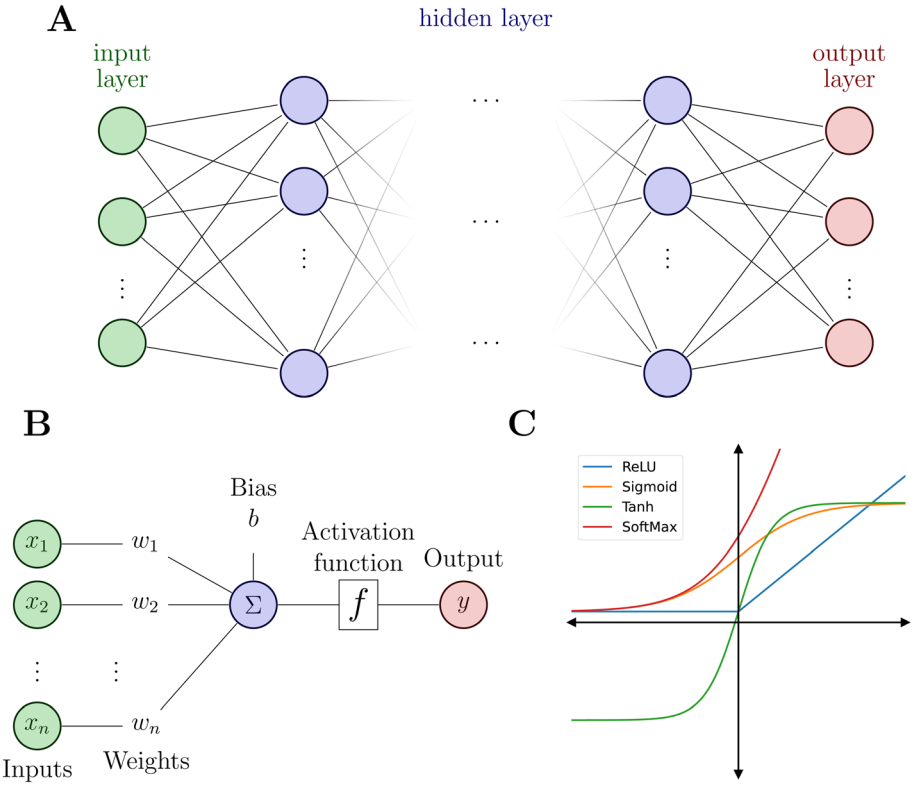

Fig. 2 Neural network and its building blocks. (A) Schematic of a NN model with an input layer (green), *N* hidden layers (blue) and an output layer (red). (B) Illustration of a node inside the hidden layers. Bottom right (C): examples of common activation functions.

$$\mathbf{y} = \mathbf{W}^{L}_{i_L}\sigma\left(\mathbf{W}^{L-1}_{i_L i_{L-1}}\sigma\left(\cdots\sigma\left(\mathbf{W}^{1}_{i_2 i_1}\left(\sigma\mathbf{W}^{0}_{i_1 i_0}\mathbf{x}+\mathbf{b}^{0}_{i_0}\right)+\mathbf{b}^{1}_{i_1}\right)\cdots\right)+\mathbf{b}^{L-1}_{i_{L-1}}\right) + \mathbf{b}^{L}_{i_L} \tag{8}$$

which is usually followed by a linear transformation in the final output layer to yield the prediction $\mathbf{y}_{L+1}$. If the NN is used to construct a PES, a chemical descriptor $\mathbf{x}$ is mapped onto one or multiple scalar values $\mathbf{y} = \{V\}$, which are the energies of one or several electronic states for an atomic configuration.

As mentioned above, the flexibility and power of a NN is related to the number of layers and nodes but the ability to obtain highly non-linear relationships between inputs and outputs is a consequence of the use of appropriate activation functions (Fig. 2C). Activation functions usually satisfy particular mathematical properties, including differentiability (crucial for computing forces or vibrational frequencies)[74] and smoothness, that simplifies the optimization of the model and increasing the quality of the prediction of energy and forces.[75]

Besides the architecture of a NN, the actual training (or "learning") step is important, too. Training comprises the parameter fitting process of the weights and biases to match the prediction $\mathbf{y}(\mathbf{x})$ to the reference results $\mathbf{t}$ for a set of N_{data} data points. The accuracy of the fit is measured by monitoring a loss function $\mathscr{L}$ which has the general form:[75]

$$\mathscr{L} = \frac{1}{N_{\text{data}}}\sum_{n=1}^{N_{\text{data}}}[\mathbf{y}(\mathbf{x}_n)-\mathbf{t}_n]^m + \omega \tag{9}$$

The value of m in eqn (9) mostly takes the value $m = 1$ or 2 (L_1 or L_2 norm) and ω can be a regularization term that helps to improve the generalizability of the model and to prevent overfitting (*i.e.* the model is fitted perfectly against training data losing generalizability). Different loss functions for fitting NNs can be used as well.[76] In general, the loss function is highly nonlinear and is minimized iteratively by a gradient descent algorithm which, preferably, can find the best solution despite potentially many local minima.[56] For PES fitting, convergence behaviour and accuracy can be improved by including additional information such as atomic forces or dipole moments (or other properties of the system) in the loss function.

3 Neural networks for potential energy surfaces

The use of NNs to represent PESs of molecular systems started in the 1990s. However, initially it was only possible to include a few degrees of freedom.[42,77–80] Applicability and transferability of NNs to larger systems and with different system compositions were improved by the approach proposed by Behler and Parrinello who decomposed the total energy of a system into atomic contributions[46]

$$E = \sum_{i=1}^{N} E_i. \tag{10}$$

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Here, N is the total number of atoms and E_i is the energy of atom i that can be predicted by one or multiple NNs (*e.g.* one for each atomic element). The inputs are local, atom-centered descriptors that encode the local chemical environment around atom i. Rooted in eqn (10), the so-called high-dimensional NNP (HDNNP),[46,81] was introduced and followed by further models.[47,48,50,51,82–96] It is important to note that most of the commonly used models are based on the decomposition of the energy in atomic contributions, although models that represent the energy as the sum of bond energies have also been proposed.[97–99] In the following, we will focus on NNs that decompose the potential energy into atomic contributions.

3.1 Descriptors

All NNs are based on a local representation of the chemical environment to correctly predict the reference data.[24,100–103] Such representations require descriptors that, most importantly, are (i) *invariant* with respect to transformations including translation, rotation and permutation of same elements, (ii) *unique* by showing changes when transformation that modify the predicted property are applied and (iii) *continuous* and *differentiable* with respect to the atomic coordinates to determine forces for molecular simulations.[101,103] Based on the type of local representation that incorporates all the conditions above, NNPs can be classified into two major categories: those with predefined and those with learnable descriptors.[15]

3.1.1 Predefined descriptors. Encoding the atomic environment by descriptors that fulfill the previously described characteristics has been a challenge since the early beginnings of the development of NN models and it is still an area of active development. Some of the requirements for a "good" descriptor can be matched with simple transformations of the Cartesian atom positions. For example, rotational and translational invariance can be obtained by using internal coordinates.[11,104] However, permutational invariance is more difficult to incorporate. A solution to this problem is the use of PIPs[33] as input for a NNP, which are still extensively used for small molecule PESs.[105–110] Other solutions are based on using symmetrized input coordinates or symmetry incorporated in the NN.[104]

A better solution to the problems described above was found with predefined descriptors introduced by Behler and Parrinello in 2007 with the development of the HDNNP.[46,47,50,81,84,90,92,93] These descriptors, termed atom-centered symmetry functions (ACSF)[81,111] or variations[50,91] thereof are the prevalent predefined descriptors for NNPs in the literature.

Originally, the local chemical environment of atom i is encoded by sets of radial- and angular-type symmetry functions G_i^{rad} and G_i^{ang} for each element or element combination of atoms j and k individually. A modified version of Gastegger and coworkers, on the other hand, combines them linearly with a weighting factor depending on the respective atoms' element number Z_j and Z_k.[112]

$$G_i^{\mathrm{rad}}(\eta, R_s) = \sum_{j\neq i}^{N} g(Z_j)\cdot \mathrm{e}^{-\eta\left(R_{ij}-R_s\right)^2}\cdot f_c(R_{ij}) \tag{11}$$

$$\begin{aligned} G_i^{\mathrm{ang}}(\zeta,\lambda,\eta,R_s) = 2^{1-\zeta}\sum_{j\neq i}^{N}\sum_{k\neq i,j}^{N} h(Z_j,Z_k)\cdot\left(1+\lambda\cos\theta_{ijk}\right)^{\zeta} \\ \cdot \mathrm{e}^{-\eta\left(R_{ij}-R_s\right)^2}\cdot \mathrm{e}^{-\eta\left(R_{ik}-R_s\right)^2}\cdot \mathrm{e}^{-\eta\left(R_{jk}-R_s\right)^2} \\ \cdot f_c(R_{ij})\cdot f_c(R_{ik})\cdot f_c(R_{jk}) \end{aligned} \tag{12}$$

In this version of weighted ACSF (wACSF) R_{ij}, R_{ik}, R_{jk} are pair distances and the angle θ_{ijk} is defined between the vectors $\vec{R}_{ij}$ and $\vec{R}_{ik}$. The contributions to the symmetry function are limited by the cutoff function $f_c(R)$ which monotonically decrease from 1 to 0 at the cutoff separation R_c. The parameter $\lambda \in \{-1,1\}$ determines the maxima of the cosine term at $\theta_{ijk} = 0°$ or 180°. The resolution and size of the descriptor are determined by the choice and number of combinations of hyperparameters η and R_s for the radial symmetry functions G_i^{rad} as well as ζ and η for the angular symmetry functions G_i^{ang}. The functions $g(Z_j)$ and $h(Z_j,Z_k)$ are the element-dependent weighting functions for which even simple expressions such as $g(Z_j) = Z_j$ and $h(Z_j,Z_k) = Z_jZ_k$ yielded satisfactory results.[112]

Regarding the ACSF representation, each descriptor is a vector for which the length depends on combinations of the sizes of respective hyperparameters η, R_s and ζ with size N_{par} but also the number of different chemical elements N_{el} in the atomic system. These are $N_{\mathrm{par}}\cdot N_{\mathrm{el}}$ for radial-type and $N_{\mathrm{par}}\cdot N_{\mathrm{el}}(N_{\mathrm{el}}+1)/2$ for angular-type symmetry functions. The size of the radial- and angular-type wACSF simply scales by the respective combination of the hyperparameters. HDNNPs with descriptor sizes of 32 wACSFs, 220 ACSFs and 35 ACSFs were trained using the energies of the molecules in the QM9 database with up to five elements. The mean absolute error of the validation and test set is reported even lower for the model with wACSFs (1.84 and 1.83 kcal mol^{-1}, respectively) than the 220 ACSFs (2.49 and 2.39 kcal mol^{-1}) and 35 ACSFs (7.57 and 7.40 kcal mol^{-1}).[112]

ACSFs commonly apply expensive trigonometric cutoff functions but computationally much cheaper polynomial cutoff functions can be designed for the same functionality.[113] Further improvement in the performance is achieved by replacing the exponential function and cosine in radial- and angular-type symmetry function with dedicated polynomials with essentially no loss in accuracy.[114] The speedup is shown by MD simulations of 360 water molecules using a HDNNP that performs about 1.8 times faster with polynomial symmetry and cutoff functions than with the original ACSFs.[114]

Another type of fixed descriptors was introduced by E and coworkers in their Deep Potential (DP) model.[82,115] These are based on the construction of a local coordinate frame which assures the required invariances. Once the positions of the atoms are transformed by a translation and rotational matrix, the local coordinates can be used to construct the descriptor based on radial and/or angular information. However, this descriptor cannot ensure smoothness because of the uncertainty in the choice of the local frame that can lead to discontinuities.[116] E and coworkers proposed the Deep Potential-Smooth Edition (DP-SE) model[117] to solve the mentioned issue by enforcing continuity of the descriptor by multiplying the

 © 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

local coordinate system with a continuous and differentiable function and modifying the embedding matrix to recover two-body and three-body terms of the descriptor.[116]

In addition to the ACSF functions and the DP descriptor, there are other descriptors that utilize the concept of neighbourhood density functions.[118,119] For this type of descriptors the information about the local environment of atom i up to a cutoff radius is represented by a density function $\rho(\mathbf{R}_i)$ depending on the nuclear charge Z_j and position $\mathbf{R}_j$ of neighbouring atoms j.

$$\rho(\mathbf{R}) = \sum_{j,\|\mathbf{R}_j\| \leq R_c} Z_j \delta\left(\|\mathbf{R} - \mathbf{R}_j\|\right) \tag{13}$$

Here, δ is the Dirac delta function. In order to use this function in a NNP, it is necessary to expand $\rho(\mathbf{R}_i)$ in a basis set of fixed dimension. For Gaussian-type basis functions, the ACSF functions are obtained.[118] Other interesting expansions include the use of Zernike basis sets in which radial basis functions and spherical harmonics polynomials are used.[119]

A major problem of using predefined descriptors is that it requires a certain degree of knowledge to define the hyperparameters appropriately.[47,50,81,84,90,92,93] Even though some of the hyperparameters can be optimized during the training as well,[83,91] a poor choice of hyperparameters can lead to limited resolution of certain atomic displacements with quasi-constant descriptors and degenerate values of the predicted energy for different geometrical structures.[120,121] The disadvantages of fixed descriptors motivated the emergence of NNPs which directly learn a suitable representation of atomic positions and element types.[3,74]

3.1.2 Learnable descriptors. The concept of learnable descriptors originates from graph neural networks.[122] In general, atoms are regarded as nodes (not to be confused with nodes of NN layers), each associated with a feature vector, which are connected to their neighbouring atoms within a cutoff sphere by so-called edges. Information between the nodes is passed along the edges over multiple iterations to encode the necessary chemical interaction.

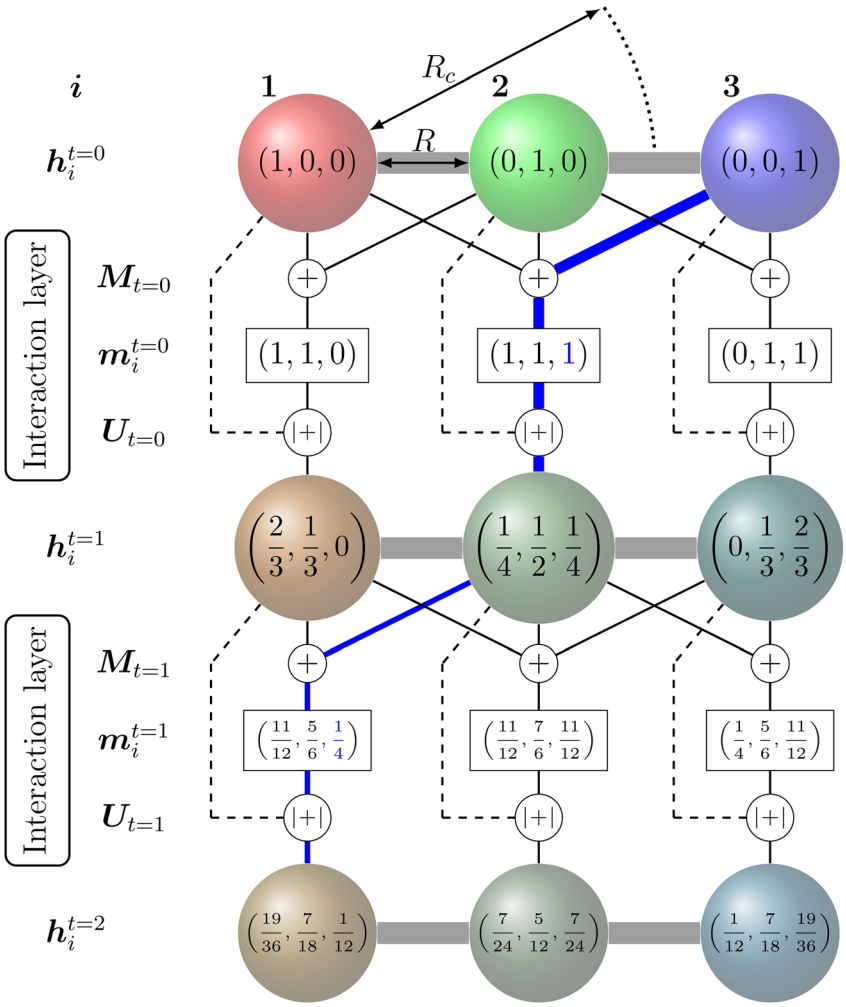

Fig. 3 Message-passing principle visualized on a chain of three nodes with initial feature vectors $h_i^{t=0}$ representing the colour fraction red, green, blue on the mixed colour of node i. The message operation M_t corresponds to the addition of the feature vectors within in cutoff range and the update operation U_t corresponds to an addition of $h_i^{t+1} = h_i^t + m_i^t$ and scaling that sum$\{h_i^{t+1}\} = 1$. Although it is outside the cutoff radius R_c, after two iterations the feature vector of node 1 (h_1^t=2) contains a fraction (information) of the initial feature vector from node 3 (visualized by the blue coloured path).

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

The feature vectors of each node with length N_f are randomly initialized as a function of the atoms' nuclear charge, that is iteratively updated by a message vector encrypting structural information and feature vectors of the atoms within a cutoff sphere by passing through interaction layers which ensure the required invariances. Fig. 3 visualizes the message passing principle on a linear chain of nodes (atoms) with distance R, where the feature vector ${h_i}^t$ at each iteration step t corresponds to the ratio of the colours red, green and blue to the mixed colour. In each interaction layer, the feature vectors of node i and connected nodes within cutoff range R_c are combined by a message function M_t (addition) to the message vector ${m_i}^t$. Note that this message function does not encode distances R. The message vector ${m_i}^t$ is combined with the feature vector ${h_i}^t$ by an update function U_t (addition and scaling to linear sum of 1) to form a refined feature vector ${h_i}^{t+1}$ that contains information of the surrounding nodes. Message and update functions usually include the transformation of feature with update vectors by a NN. For an iteration step $t > 1$, this approach allows that information from nodes that are outside of the cutoff range can still be incorporated in a feature vector of a given node i indirectly. This means that for the case illustrated in Fig. 3, the feature vector ${h_1}^{t=2}$ of node 1 contains a fraction of blue colour after two iterations $(1/12)$ that is passed from node 3 *via* node 2.

Many of the more recently developed NNPs[48,51,85–89,94–96] apply such atom-wise feature vector approaches and are called message-passing NNs (MPNNs).[123,124] Depending on the MPNN model, the atomic feature vectors of either the final iteration or each iteration are passed to a specific NN and transformed into the desired quantity (*e.g.* energy).

Feature vectors with higher number of elements N_f and more complex message and update functions including bond distance and direction dependencies allow higher resolution of the structural encoding. In common NNPs, the number of elements in the feature vectors N_f range from about 64 to 128 per element. A larger number might increase the risk of overfitting.[86] Similarly, a larger number of message passing iterations improves the representation of the structural features but the potential energy accuracy usually shows sufficient saturation after three iteration $(t = 3)$.[48,85–87,94]

3.2 Architectures

Given that the field of NNPs is very active, it is impossible to describe all the available NN architectures. Hence this section is not a comprehensive review of all possible architectures but rather a more history-guided view of architectures and what functionalities were included in subsequent development steps.

Initial models use NNs as a method for the fitting of PES only (no forces).[125] These models were limited to small molecules in gas phase and were fitted to energies of *ab initio* calculations *via* a many-body expansion[126] or a high-dimensional model representation.[127] Therefore, these models take energies and positions to predict coefficients for a defined functional form. These models already achieved spectroscopic accuracy for small molecules.[12]

The introduction of the HDNNP with the concept of decomposing the molecular energy into atomic contributions (eqn (10)) changes the paradigm of NNPs. A new challenge was encoding the local environment information sufficiently well for an accurate energy prediction that lead to the two main approaches of predefined or learnable descriptors. The main development of NN architectures with predefined descriptors goes towards more sophisticated descriptors to encode atom-centered properties which are then provided to standard fully-connected feed-forward NNs.[128] NN architectures with learnable descriptors and the MPNN approach differ in their message and update functions within an interactions layer.

The first MPNN proposed was the deep tensor neural network (DTNN)[48] by Schütt and coworkers that had been further improved into the, to this day, popular SchNet model.[85] An interaction layer in SchNet includes so called continuous-filter convolutional layers that have already been used in image or sound processing.[85] A combination of the popular predefined ACSF descriptors and learnable ones was proposed by Isayev and coworkers and their atoms-in-molecule NN model (AIMNet).[87] Modified ACSF descriptors from the ANI architecture were used for initialization of atomic structure feature vectors, combined with atomic information feature vectors and passed through the interaction layer.

Although these models already achieve good accuracy, long range interactions between chemical compounds can only contribute to the total energy if the information is included in or passed to the descriptor by a sufficiently long cutoff range R_c. Systems with strong electrostatic interactions, especially with highly polar or ionic chemical species, requires larger cutoffs but at the cost of higher computational demand.[51] One solution is to add a Coulomb term to the atomic energy contributions which includes electrostatic interactions between atomic charges q predicted by the NN model.

$$E = \sum_{i=1}^{N}\left[E_i + \frac{1}{2}\sum_{j>1}^{N}\frac{q_i q_j}{R_{ij}}\right] \tag{14}$$

The earliest NN model using eqn (14) was introduced by Artrith and Behler in 2011 that trains a separate NN with reference charges from a Hirshfeld population analysis.[90] Another approach is applied by the TensorMol model that predict atom charges by fitting the *ab initio* and physically determinable molecular dipole moment to the predicted one computed by the atom charges.[91]

Additional physically motivated interactions, such as dispersion interactions, were also included in the TensorMol model but have been employed in PhysNet, too. PhysNet is based on the MPNN architecture and was developed by Unke and Meuwly.[51] It does not only add an energy contribution from the DFT-D3 dispersion correction scheme[129] but also modifies eqn (14) by applying a damping function that smoothly damps Coulomb interactions for small atom distances to avoid singularities

$$E = \sum_{i=1}^{N}\left[E_i + \frac{1}{2}\sum_{j>i}^{N} q_i q_j \cdot \chi\left(R_{ij}\right)\right] + E_{D3}. \tag{15}$$

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

E_{D3} is the DFT-D3 dispersion correction and the damping function $\chi(\mathbf{R}_{ij})$ is defined as:

$$\chi(R_{ij}) = \phi(2R_{ij})\frac{1}{\sqrt{{R_{ij}}^2+1}} + (1-\phi(2R_{ij}))\frac{1}{R_{ij}}. \quad (16)$$

A continuous behaviour is ensured by the cutoff function $\phi(R_{ij})$.

Although adding a Coulomb term to NNPs improves the description of long range interactions while the atomic charges still depend on the local chemical environment.[104] However, chemical systems are inherently non-local. Therefore, the approximation breaks down for systems with changes in the total charge state (*i.e.* ionization, protonation or deprotonation), electronic delocalization or spin density rearrangements.[89] These effects are difficult to capture with NN architectures which model changes in the atom charges by local perturbations.

The most recent generation of NNPs addresses the problem of non-local charge transfer by using different strategies. The first work dedicated to the issue of charge equilibration was the "charge equilibration *via* NN technique" (CENT) developed by Ghasemi and coworkers.[92] The CENT algorithm equilibrates the charge density to minimize the electrostatic energy which depends on environment-dependent atomic electronegativity and hardness besides the charge–charge interaction. Inspired by CENT, Behler and coworkers introduced their fourth generation HDNNP (4G-HDNNP) model where NNs are trained to predict environment-dependent atomic electronegativities (constant element-specific hardness) and the charge equilibration yields the reference atomic charges.[47] In a second training step, NNs provided with ACSFs and the atomic charge information are trained to predict the short-range atomic energy contributions which sum up with the electrostatics to the correct reference energy and forces.

SpookyNet is a MPNN model and introduced by Unke and coworkers that treats the problem of non-locality by creating an embedding for charges and spin.[89] It is capable to predict molecular systems with different spins and charged states as provided in the reference data set within one single model. The general idea of predicting PESs of chemical systems for different electronic states and their coupling strength within one model is an area of active research.[130] One model in this direction that can be mentioned is SchNarc[131] that combines the SchNet model with the surface hopping including arbitrary couplings (SHARC)[132] code.

So far, we have been reporting the effort to improve the models accuracy by introducing more physically motivated interactions. However, current developments for MPNNs focus on passing spatial directions between atoms to the NN that allow the prediction of atom-centered tensorial properties such as atomic polarizability.[94,133,134] Providing solely distance information inherently ensures translational and rotational invariance for atom-centered scalar properties (predictions do not change with respect to, *e.g.*, rotation of the molecule). The challenge with directional information is rotational equivariance which means that predicted atom-centered directional properties $\vec{f}(\mathbf{R})$ keep its amplitude but change in direction equivalent to a rotation $\mathcal{A}$ of the molecular coordinates $\mathbf{R}$.

$$\mathcal{A}\cdot\vec{f}(\mathbf{R}) = \vec{f}(\mathcal{A}\cdot\mathbf{R}) \quad (17)$$

MPNNs that encode directional information (directional message passing) and fulfill eqn (17) are called equivariant NNs (ENNs).[135,136]

ENNs have been proven to be data-efficient and capable of providing better predictions of tensorial quantities (*i.e.* dipole, quadrupole moments) than invariant models. ENN models with different modifications were suggested to include directional information and assure equivariance. Some of them are PaiNN,[94] NeuqIP,[96] and NewtonNet.[137] Still one of the best performing ENNs on the QM9 data set is DimeNet, where rotational equivariance is achieved by representing the local chemical environment of an atom by spherical 2D Fourier–Bessel basis with radial basis functions to represent bond distances and spherical basis functions to represent angles between bonds towards neighbouring atoms.[133]

Many NN potentials are often additionally designed for application on periodic systems including solids and crystals,[49] or were updated to support periodicity.[138] Others are specifically designed to train on reference data to predict formation energy, lattice parameters of the unit cells and other material properties directly from the structural fingerprint.[139,140] The application of ML (including NNs) to materials has been discussed in detail in recent reviews[141–143] and is not further considered in the present work.

The field of NNs in computational chemistry has been and will continue to be steadily developed to improve the capability and accuracy in predicting reference data. In consequence, the selection of a model should be done based on the problem at hand, the availability of the code, its user friendliness, and the computational resources available. It might not be necessary to use the most sophisticated model if the task does not require that level of description. Most of the previously described architectures are based on open source NN frameworks like Tensorflow[144] or PyTorch[145] which open the possibility to modifications and enhancements of the described models.

4 Construction of PESs

The collection of reference structures is an essential step in constructing a molecular PES, especially since the underlying functional form of the potential is not based on physical laws and is inferred purely from reference data.[104] Besides the unfavourable scaling of the configurational space with system size, the computational expense associated with a reference point is usually high and depends on the level of quantum chemical theory used. Thus, the number of expensive and non-trivial *ab initio* calculations needs to be restricted to a minimum and optimally covers the configurational space most important/representative (this is an open question in itself) to the problem at hand.[3,146] Ultimately, the configurational space that is covered by the reference data set defines the boundaries of application

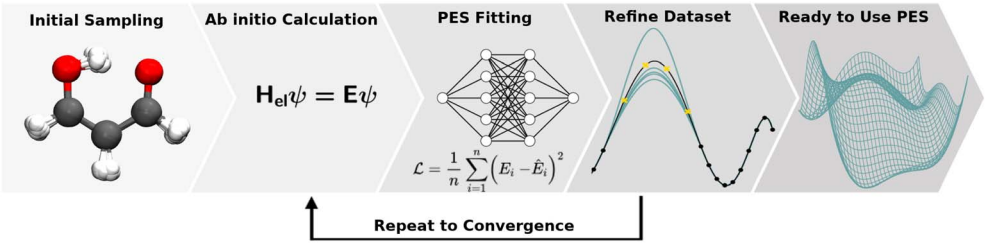

Fig. 4 The process of PES generation: the configurational space of a chemical system (here malonaldehyde) is sampled to obtain an initial set of geometries. A quantum chemical *ab initio* calculation is carried out for each geometry to obtain reference data (including energies). After a NNP is fitted to the initial reference data set the resulting PES is validated thoroughly to find holes. New *ab initio* calculations are run for scarcely sampled regions and a new NNP is fitted. These steps are repeated until the PES has the required quality before the PES can be used to study the chemical system.

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

of the NNP. Therefore, knowing the application(s) for which the PES will be used is essential when generating the data.

Reference data sets can be generated using a multitude of strategies which often requires the generation of an initial data set and refining it iteratively. This iterative process is illustrated in Fig. 4. Commonly employed strategies for structure sampling, which are often combined, will be described in the following. In addition to methods reviewed here, other possibilities include virtual reality sampling,[147–150] Boltzmann machines[151] or sampling based on the AMONS approach.[22]

4.1 Initial sampling

4.1.1 *Ab Initio* MD. *Ab initio* MD (AIMD) constitutes an established means for generating reference data that samples a part of the configuration space of a chemical system.[104] The temperature T (or the velocities that are drawn from a Maxwell–Boltzmann distribution corresponding to T) at which the simulation is run determines which part of a PES is sampled, how strongly the molecular geometries are distorted and whether or not reaction barriers are crossed. If the chemical system under investigation has multiple isomers, AIMD simulations can be run for all of them (partly) avoiding the need of running a long simulation that samples all isomers. Ideally, the sampling temperature T is chosen to be higher than the temperature at which the NNP is used. In other words, if the reference data set that was used to train a NNP was generated at $T = 300$ K the NNP should not be used to run simulations at $T >$ 300 K because (most likely) configurations outside of the reference data set are visited leading to a breakdown of the NNP. Thus, running AIMD at a sufficiently high sampling temperature is needed to guarantee that the production runs do not enter the extrapolation regime, while the lower energy configurations are still sampled.[3]

The obvious disadvantage of running AIMD at the (final) level of theory at which the reference data set is generated is the high computational cost. This either limits the level of quantum chemical rigor or it limits the extent to which the configurational space can be sampled.[152] Alternatively, configurations can be generated using sampling by proxy.[3] This approach involves running AIMD at a lower level of theory to sample the PES and then perform single point *ab initio* calculations for a representative set of geometries at a higher level. This ideally requires that the topologies of the lower and the higher level of theory are similar to guarantee that the "correct" configurations are sampled. If the two PESs differ too much it is possible that the regions explored on the lower level PES do not correspond to relevant regions on the high level PES (which might happen if a force field is used to guide the sampling).[3,104] As a consequence, the NNP could reach an extrapolation regime and exhibit a nonphysical behaviour.

Reactive chemical systems are usually associated with rare events. When NNPs are used to study reactive systems it is, thus, not sufficient to sample the reactant and product states since the reaction path (which is rarely visited in a simulation) needs to be part of the reference data set as well. TS regions can be sampled using AIMD by employing a scheme similar to umbrella sampling,[153] in which geometries around the TS are sampled by harmonically biasing the molecule towards the TS.

A simulation technique that is related to MD simulations and can be used to generate configurations for the construction or refinement of a reference data set is metadynamics.[154] Converse to ordinary MD, metadynamics uses history dependent biasing potentials to artificially increase the potential of visited regions on the PES and enhance the sampling of higher energy regions.

4.1.2 Normal mode sampling. Normal mode sampling (NMS) was proposed to enable accelerated yet chemically/physically relevant sampling of a PES.[50] As the name suggests, NMS uses the normal modes of vibration of a molecule to generate molecular geometries that cover configurational space at which single point calculations can be carried out at a desired level of theory. NMS is carried out as follows:[50] (i) the molecule of interest is optimized at a desired level of theory (ii) normal mode coordinates $Q = \{q_i\}$ (*i.e.* eigenvectors of the mass-weighted Hessian) and corresponding force constants $K = \{k_i\}$ are determined (with $i \in [1, N_f = 3N - 5]$ or $i \in [1, N_f = 3N - 6]$, for linear and non-linear molecules, respectively) (iii) N_f uniformly distributed random numbers c_i with $\sum_i c_i \in [0, 1]$ are generated (iv) displacements for each normal mode are determined as $R_i = \pm\sqrt{\frac{3c_i N_A k_b T}{K_i}}$ with N_A and k_b being the Avogadro number and the Boltzmann constant, respectively. This

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

displacement is obtained by scaling an energy with $c_i\left[E_k = \frac{3}{2}c_i N_A k_b T\right]$ and setting it equal to a harmonic potential $\left[U = \frac{1}{2}kr^2\right]$. (v) Determine the sign of the displacement R_i randomly using a Bernoulli distribution to sample the attractive and repulsive parts of the potential (vi) the normalized normal mode coordinates q_i are scaled using R_i giving a new set of coordinates.

Unlike the consecutive snapshots of an AIMD, NMS yields uncorrelated molecular configurations in a very efficient manner. Nonetheless, the sampling is based on a harmonic approximation of the potential well and usually only geometries close to the respective equilibrium structures are obtained. For larger displacements and large amplitude motions, the harmonic approximation breaks down. Thus, NMS is often used in conjunction with alternative sampling strategies or followed by adaptive sampling.[3]

4.1.3 Diffusion Monte Carlo. Diffusion Monte Carlo (DMC) can be used to determine the zero-point energy (ZPE) and wavefunction of a molecule by appropriately, yet randomly, sampling the configurational space.[155] The foundation of DMC is the similarity of the imaginary time SE

$$\hbar\frac{\partial\Psi(x,\tau)}{\partial\tau} = \frac{\hbar^2}{2m}\nabla^2\Psi(x,\tau) - \left[V(x) - E_0\right]\Psi(x,\tau) \quad (18)$$

with the diffusion equation with a sink term allowing random-walk simulations to estimate the ZPE and wavefunction.[156] Given a molecule, a set of walkers is initialized (usually at some energy minimum), propagated randomly at each time step τ and used to represent the nuclear wavefunction. In one dimension, the displacement assigned to each of the walkers is given by[156]

$$x_{\tau+\Delta\tau} = x_\tau + \sqrt{\frac{\hbar\Delta\tau}{m}}r \quad (19)$$

where x_τ corresponds to coordinates at time step τ, $\Delta\tau$ is the time step of the random-walk simulation, m corresponds to an atomic mass and r is a random number drawn from a Gaussian distribution, $\mathcal{N}(0,1)$. Once the walkers are randomly displaced following eqn (19), their potential energy E_i is determined. Based on E_i with respect to a reference energy E_r, a walker might stay alive, give birth to a new walker or can be killed following the probabilities below:

$$P_{\text{death}} = 1 - e^{-(E_i - E_r)\Delta\tau} \quad (E_i > E_r) \quad (20)$$

$$P_{\text{birth}} = e^{-(E_i - E_r)\Delta\tau} - 1 \quad (E_i < E_r) \quad (21)$$

Once the probabilities have been determined, the dead walkers have been eliminated and new walkers are initialized, E_r is adjusted following

$$E_r(\tau) = \langle V(\tau)\rangle - \alpha\frac{N(\tau) - N(0)}{N(0)} \quad (22)$$

The averaged potential energy of the alive walkers is given by $\langle V(\tau)\rangle$, α governs the fluctuation in the number of walkers and is a parameter, and $N(\tau)$ and $N(0)$ are the number of alive walkers at time step τ and 0, respectively. The ZPE is then approximated as the average of E_r over all imaginary time.[155,156]

The geometries sampled using the DMC scheme are physically meaningful (the ensemble of walkers represents the nuclear ground state wavefunction) and efficiently obtained by only using energies. In comparison to AIMD, the DMC scheme has the advantage that it samples configurations up to the ZPE, which becomes larger for bigger molecules. The (quantum) exploration of a PES using DMC is typically done after a first PES has been fitted and is used to refine the reference data set.[157] DMC has been proposed as a tool to detect holes (regions on a PES that have large negative energies with respect to the global minimum) in ML based PESs.[157] These holes are caused by insufficient data in specific regions in configuration space, for which a NNP without any underlying physical knowledge leads to artifacts. As an adaptation, DMC with artificially reduced masses has been proposed to locate holes more efficiently due to the larger random displacements (which are proportional to $1/\sqrt{m}$, see eqn (19)).

4.2 Validation and refinement of the data set

These holes were found to exhibit energies with large negative values.[158] After an initial PES is fitted, a thorough evaluation of the PES to discover any holes is needed. For this reason, the family of active learning schemes which comprise algorithms to systematically generate reference data sets have gained considerable attention.[159] The necessity for more elaborate sampling schemes is related to the impracticality of an exhaustive sampling of a PES and the high computational cost of extensive *ab initio* calculations. Typically, a first PES is trained on reference data based on representative configurations. This is followed by suitably extending the data set in an iterative fashion in which similar configurations are avoided and configurations from underrepresented regions of the PES are found and included into the data set.[159] This approach is usually termed adaptive sampling (or on-the-fly ML).[160,161] Therefore, a requirement for ML models to autonomously select new reference data is the availability of an uncertainty estimation. If a defined uncertainty threshold is exceeded for a particular configuration electronic structure calculations are performed and used to extend the reference data.

4.2.1 Uncertainty estimation. Given the breadth of NN methods (or ML methods in general), various approaches for uncertainty estimation exist. One of the most popular methods is query-by-committee.[159] This approach involves training/fitting a number of individual NNPs (*e.g.* starting from different parameter initialization or on different splits of the reference data set) and using the ensemble for predictions. In regions of the configuration space where sufficient data is available the predictions of the different models agree well. Conversely, the predictions for configurations for scarcely sampled regions will diverge rapidly, and can be used to autonomously select new configurations. A possible uncertainty metric for NNPs is[152]

$$\sigma_E = \sqrt{\frac{1}{\mathcal{N}-1}\sum_{i=1}^{\mathcal{N}}\left(E_i - \overline{E}\right)^2} \quad (23)$$

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

with $\mathcal{N}$ being the number of individual models, E_i an individual energy prediction and the average of all energy predictions, $\bar{E}$. Similar metrics can certainly also be adapted to other properties including the forces acting on the atoms α:[152]

$$\sigma_{F^\alpha} = \sqrt{\frac{1}{\mathcal{N}-1}\sum_{i=1}^{\mathcal{N}} \left\|F_i^\alpha - \overline{F}^\alpha\right\|^2} \tag{24}$$

The use of query-by-committee requires the training of several independent models which incurs a high computational cost to obtain the uncertainty. In addition to this, it has been found that the uncertainty estimated by NNP ensembles are often overconfident.[162] As a solution to this bottleneck, methods that obtain the uncertainty in a single evaluation have been proposed. Some us[76] recently introduced a modification of the PhysNet architecture that allows the calculation of the uncertainty on the prediction through a method called deep evidential regression.[163] Using this method, the energy distribution of the system is represented with a Gaussian and its uncertainty as a gamma distribution. With this approach, it is possible to obtain the prediction and the uncertainty of the prediction in one single calculation. Other possibilities for the prediction of uncertainties include the use of Bayesian NNs, however, they imply a larger computational cost than the previously described methods.

4.2.2 Elaborate sampling techniques. With the availability of an uncertainty measure and an initial PES, geometries from underrepresented regions on the PESs can easily be identified: the initial PES is used to guide the sampling of new structures (by MD, DMC, metadynamics, …) and if the uncertainty measure (*e.g.* σ_E) exceeds a threshold, *ab initio* calculations are performed for the geometry and the data set is suitably extended. These more systematic approaches of generating reference data sets offer a number of advantages over random methods. Since including similar configurations is avoided and new data is only added for scarcely sampled regions, the approaches are clearly more data efficient requiring less expensive quantum chemical computations. Additionally, since the NNP that is used to guide the sampling of new geometries is topologically very similar to the *ab initio* PES it is assured that configurations, that are similar to the configurations visited in AIMDs, are sampled. The quality of the uncertainty estimate is crucial for all adaptive sampling schemes. While an overconfident estimate leads to an inaccurate PES (in the worst case holes are overlooked) an under-confident estimate leads to the inclusion of redundant configuration and unnecessary, computationally expensive *ab initio* calculations. Zipoli and coworkers report that adding new configurations based on uncertainty estimation from an ensemble of NNPs does not show significant differences from random sampling.[162] Contrary to that, Pernot[164] and Zheng *et al.*[165] find that querying the uncertainties from ensembles are well suited for outlier detection and adaptive sampling. This clearly indicates the necessity for future studies exploring more elaborate sampling techniques.

5 Knowledge transfer

Most ML algorithms (foremost deep learning) heavily rely on abundant training data to extract the underlying patterns in very complex data. This severe data dependence is one of the major drawbacks to deep learning.[166] The collection of big data sets is a cumbersome and expensive task impeding the generation of large, high-quality data sets. While this time-consuming endeavor might be possible for some areas of application (*e.g.* manually labeling images for an image recognition task) insufficient training data/data scarcity is an inevitable problem in other domains (*e.g.* drug discovery).[166,167] Thus, transfer learning (TL)[166,168] and related approaches including Δ-ML,[169,170] dual-level Shepard interpolation,[171] multifidelity learning[172] or the multilevel grid combination technique[173] have been proposed to circumvent the severe data dependence/scarcity or expensive labeling efforts by knowledge transfer. Thereby, exploiting the knowledge acquired by solving one task (a source task) to solve a new, related task (a target task) forms its common ground.[168]

Besides addressing the data scarcity dilemma, knowledge transfer also helps reducing training times, computer resources (which both are significant for large data sets/models[174]) and their energy consumption. Recently, the CO_2 emission for training common natural language processing (NLP) models has been studied, which, depending on their size, can exceed a car's lifetime CO_2 emission.[175]

Traditional ML problems usually proceed in a domain $\mathcal{D}$ and try to solve a specific task $\mathcal{T}$. In the context of molecular PESs, the domain $\mathcal{D}$ is a set of molecular configurations (defined by $\{\mathbf{R},\mathbf{Z}\}$) with their associated descriptors (see Section 3.1) and the task involves the prediction of the corresponding energies $E_\lambda^{\mathrm{BO}}(\mathbf{R})$ (eqn (3)). Considering two domains (a source $\mathcal{D}_s$ and a target domain $\mathcal{D}_t$) and two learning tasks ($\mathcal{T}_s$ and $\mathcal{T}_t$) from the perspective of traditional ML, two separate machines are trained to solve the two tasks (see Fig. 5). In contrast, TL circumvents learning to solve both tasks from scratch by facilitating the learning of $\mathcal{T}_t$ with knowledge from $\mathcal{T}_s$ (see Fig. 5). Here, the domains and/or tasks can differ for TL giving rise to three distinct cases.[167,168] (i) The domains are the same, $\mathcal{D}_s = \mathcal{D}_t$, while the tasks differ, $\mathcal{T}_s \neq \mathcal{T}_t$. This situation can, *e.g.*, be found for TL between molecular properties (inductive learning) (ii) the domains differ, $\mathcal{D}_s \neq \mathcal{D}_t$, while the tasks remain the same $\mathcal{T}_s = \mathcal{T}_t$. This corresponds to transductive learning and can be found for TL between different molecular data sets. (iii) Both, the domains and the tasks differ, $\mathcal{D}_s \neq \mathcal{D}_t$ and $\mathcal{T}_s \neq \mathcal{T}_t$. All three subsettings have in common that they try to learn/improve the target predictive function $f_t(\cdot)$ of $\mathcal{T}_t$ in $\mathcal{D}_t$ using the knowledge in $\mathcal{D}_s$ and $\mathcal{T}_s$ which is the definition of TL.[168]

The training of NNPs typically requires thousands to tens of thousands of *ab initio* calculations even for moderately sized molecules, which often limits the quantum chemical calculations to the level of density functional theory (DFT). If highly accurate molecular properties are needed, researchers usually resort to the coupled cluster with perturbative triples (CCSD(T)) level of theory. This "gold standard" – CCSD(T) – scales as N^7

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

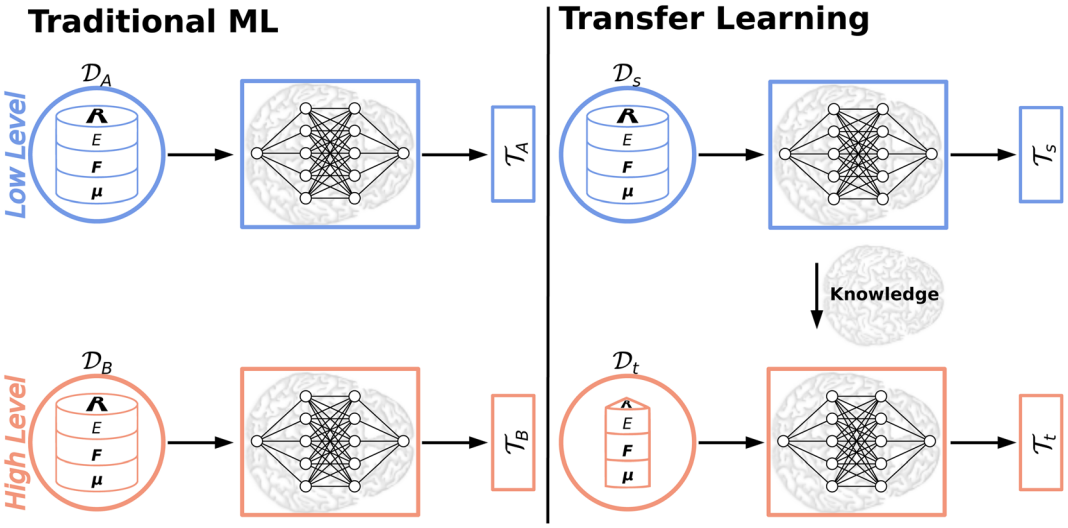

Fig. 5 Illustration of the difference between traditional ML and TL approaches. In traditional ML, two different models are trained for two different tasks ($\mathcal{T}_A$ and $\mathcal{T}_B$), although the two tasks might be related (*e.g.* predicting the MP2 and the CCSD(T) energy of a given configuration). In TL, however, the knowledge gained from solving a source task ($\mathcal{T}_s$) in the source domain ($\mathcal{D}_s$) is used to solve a target task ($\mathcal{T}_t$) (*e.g.* by fine-tuning the weights and biases). In the context of PES generation, typically a (global) PES is developed at a low level of theory and then transfer leaned with less data calculated at a considerable higher level of theory (*e.g.* CCSD(T)).

(with N being the number of basis functions),[176] which makes calculating energies and forces for large data sets and larger molecules impractical. Thus, TL[50,177–180] and related Δ-learning approaches[170,181–183] gained a lot of attention in recent years and were shown to be data and cost effective alternatives to the "brute force" approach in quantum chemistry: a low level PES based on a large data set of cheap reference data (*e.g.* DFT) is generated first, which then is used to obtain a high level PES based on few, well chosen high level of theory (*e.g.* CCSD(T)) data points.

5.1 Deep transfer learning

Deep TL[167] combines deep NN architectures with TL among which fine-tuning is the most commonly used technique. Fine-tuning, which is a parameter-based TL technique, assumes that the weights and biases of a deep NN that was trained on a source task $\mathcal{T}_s$ contain useful information to solve a (related) target task $\mathcal{T}_t$. In the context of molecular PESs, a lower level (LL) PES is obtained by training a deep NN on a large data set of energies/gradients determined at a low level of theory. Then, the parameters (weights and biases) of the LL PES are migrated to the target model for which they serve as the initialization (a good initial guess). The target model (*i.e.* the transfer learned model) is then fine-tuned (retrained) on a small data set of high-level of theory energies/gradients. The fine-tuning technique that migrates the parameters of a LL PES to a high level (HL) PES is shown in Fig. 5.

There are certain subtleties when applying TL in practice. TL can be performed without any further restriction to the fine-tuning for which all weights and biases are allowed to adapt to the new HL data. Conversely, it is possible to fix the weights and biases of particular layers. Usually, the first hidden layers are fixed and only the last layer(s) are allowed to adjust (alternatively a new, final layer can be added keeping the LL model as is). Fixing a portion of the NN parameters limits its flexibility but might help in reducing overfitting for small data sets. Recently, TL in combination with NNs was used for structure-based virtual screenings of proteins.[184] The authors found that fine-tuning a full NN worked best for kinases, proteases and nuclear proteins, however, fine-tuning only the final layer yielded better results for G-protein-coupled receptors (GPCRs). They speculate that this is caused by the limited and less diverse data for GPCR targets. Besides the need to avoid overfitting, it is imaginable that for NNs that employ learnable descriptors of the atomic/molecular configuration it might be beneficial to freeze the parameters that are used to learn the descriptor for the fine-tuning step. Instead of freezing a portion of the layers, fine-tuning with differential learning rates[185] (*i.e.* having different learning rates for different parts of the NN) could allow minimal changes to early layers (*e.g.* where the descriptors are learned) and larger adjustments to the later layers. Although empirical rules are followed in the community, accepted criteria for choosing TL methods are essentially nonexistent.[167]

5.2 Δ-Machine learning

The Δ-machine learning approach was developed in the context of kernel-based methods and is motivated by the fact that the heaviest burden in quantum chemical calculations is the determination of a tiny energy contribution to a (approximate) total energy.[170] The approximate energy often is able to describe the general chemistry/physics of a given system, while the determination of the "Δ" comes at a tremendous computational cost due to adverse scaling with system size of correlated electronic structure methods. For a molecular property, the Δ-ML prediction is modeled as a LL value plus a correction towards a HL value following

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

$$P_{\mathrm{HL}}(R_{\mathrm{HL}}) \approx \Delta_{\mathrm{LL}}^{\mathrm{HL}}(R_{\mathrm{LL}}) = P'_{\mathrm{LL}}(R_{\mathrm{LL}}) + \sum_{i=1}^{N} \alpha_i k(R_{\mathrm{LL}}, R_{\mathrm{i}}). \quad (25)$$

The high level property P_{HL} (*e.g.* enthalpy H_{HL}) at a relaxed molecular geometry (R_{HL}) is approximated as a related property $P'_{\mathrm{LL}}(R_{\mathrm{LL}})$ (*e.g.* energy E_{LL}) obtained at the LL plus a correction term[170] that is obtained from ML (reference 170 employed Slater type basis functions k and kernel ridge regression (KRR) to obtain the regression coefficients α_i). The Δ-ML approach as defined in eqn (25) allows modeling changes in level of theory (*e.g.* DFT → CCSD(T)), molecular property (*e.g.* energy → enthalpy) and molecular geometry. Although the Δ-ML approach is often used in conjunction with kernel-based methods, a correction PES Δ (*i.e.* $V_{\mathrm{HL}} = V_{\mathrm{LL}} + \Delta$) can also be learned using NNs.[186] The resulting HL PES V_{HL} can either be used directly (requiring the evaluation of two models) or can be used as a proxy to generate a larger data set for a final training containing many, though approximate, HL points.[186] As is common for the ML field, different flavours of Δ-ML exist.[146,170,172,173,181,182,186–189]

Recent work proposed "Δ-DFT" that uses Kohn–Sham (KS) electron densities ρ^{KS} to correct the DFT energy towards, *e.g.*, a coupled cluster energy following

$$E^{\mathrm{cc}} = E^{\mathrm{DFT}}[\rho^{\mathrm{KS}}] + \Delta E[\rho^{\mathrm{KS}}] \quad (26)$$

using KRR.[146] While the formalism of DFT and wavefunction based approaches (such as CCSD(T)) differ radically (also note that the CCSD(T) density is not routinely calculated and not needed to obtain the CCSD(T) energy), the "learnability" of DFT and CCSD(T) energies from KS densities was studied alongside the Δ − DFT approach. The authors find starting from ρ^{KS} learning DFT and CCSD(T) energies directly is associated with approximately the same effort. However, learning $\Delta E[\rho^{\mathrm{KS}}]$ was more efficient and yielded lower out-of-sample errors at smaller training set sizes.[146]

6 Exemplary applications of NNPs in molecular simulations

The high flexibility of NNs allows the representation of PESs for a wide range of chemical systems and reactions as long as a sufficiently large reference data set is available from *ab initio* computations at a sufficient level of theory to correctly describe the physics in the system. This section presents several typical applications of NNPs in molecular simulations.

6.1 Gas phase spectroscopy

In a recent review, Manzhos and Carrington report advances of NNPs and applications in classical and quantum dynamics of small and reactive systems.[125] They point out that for small systems modern NNPs are still outperformed by permutationally invariant polynomial (PIP[33,36]) methods in terms of PES fitting error which, however, does not translate to significant deviations in computed observables such as vibrational frequencies.[190] As an example, the RMSE of a Gaussian process regression (GPR) model potential (0.017 kcal mol^{-1}, 5.98 cm^{-1}) is half of that of a NNP (0.034 kcal mol^{-1}, 12.03 cm^{-1}) with regard to 120 000 reference points for formaldehyde. However, the RMSE of the first 50 (100) predicted vibrational frequency levels with respect to their reference is 0.43 cm^{-1} (0.82 cm^{-1}) for the NN and 0.46 cm^{-1} (0.82 cm^{-1}) for the GPR potential. When the potential models are fitted to a subset of reference points with high significance for the vibrational frequency prediction, the RMSE of the first 50 (100) predicted vibrational frequency levels differs substantially with 0.21 cm^{-1} (0.30 cm^{-1}) for the NN and only 0.04 cm^{-1} (0.06 cm^{-1}) for the GPR model.[125,191]

The application of NNPs to determine anharmonic vibrational frequencies in combination with TL has been studied in ref. 179. For that purpose, a NN of the PhysNet type is trained on *ab initio* energies, forces and dipole moments and employed in second order vibrational perturbation theory (VPT2) calculations that are directly compared to their experimental counterpart. A total of eight molecules are studied from which the results for formaldehyde are shown in Fig. 6A as it allows a good comparison of a TL scheme with a model that is trained "from scratch" due to its small size. A PhysNet model that is trained on MP2 data ($\mathrm{NN_{MP2}}$) yields errors up to 40 cm^{-1} with respect to the experimental values, while the CCSD(T)-F12 model ($\mathrm{NN_{CCSD(T)\text{-}F12}}$) has a maximum deviation of ∼20 cm^{-1}. Both $\mathrm{NN_{MP2}}$ and $\mathrm{NN_{CCSD(T)\text{-}F12}}$ were trained on roughly 3400 *ab initio* energies, forces and dipole moments, for which the computation at the CCSD(T)-F12 level of theory requires high computational effort. In contrast, 6% of the CCSD(T)-F12 reference points are sufficient to transfer learn a $\mathrm{NN_{MP2}}$ model and achieve an accuracy that is within ∼7 cm^{-1} of $\mathrm{NN_{CCSD(T)\text{-}F12}}$ trained on the full reference set from scratch.

6.2 Condensed phase simulations

Even though NNPs scale more favourably with the number of atoms, the construction of a reference data set for molecular compounds still requires several thousand *ab initio* calculations. As NNPs are mathematical representations of the input data and are uninformed about the underlying physics governing intermolecular interactions, their extrapolation capabilities are rather limited. This also concerns the transferability of NNPs optimized on smaller molecular clusters towards larger clusters or even periodic systems. This issue has been addressed recently, for instance, by Kästner and coworkers on liquid water and Marx and coworkers on protonated water clusters using NNPs.[194,195]

Kästner and coworkers train a Gaussian moment NN (GM-NN) model on DFT rev-PBE-D3 reference data of water cluster configurations produced by *ab initio* MD simulation at 150, 300 and 800 K, and study its transferability to a periodic bulk water system with 64 molecules from *ab initio* MD simulation at 400 K.[194,196] The GM-NN model trained on clusters containing 30 to 126 water molecules can reproduce the total energy of the periodic bulk water system well, although with a slightly broader error distribution as for the model trained on the periodic system. The potential energy predicted by the cluster

 © 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

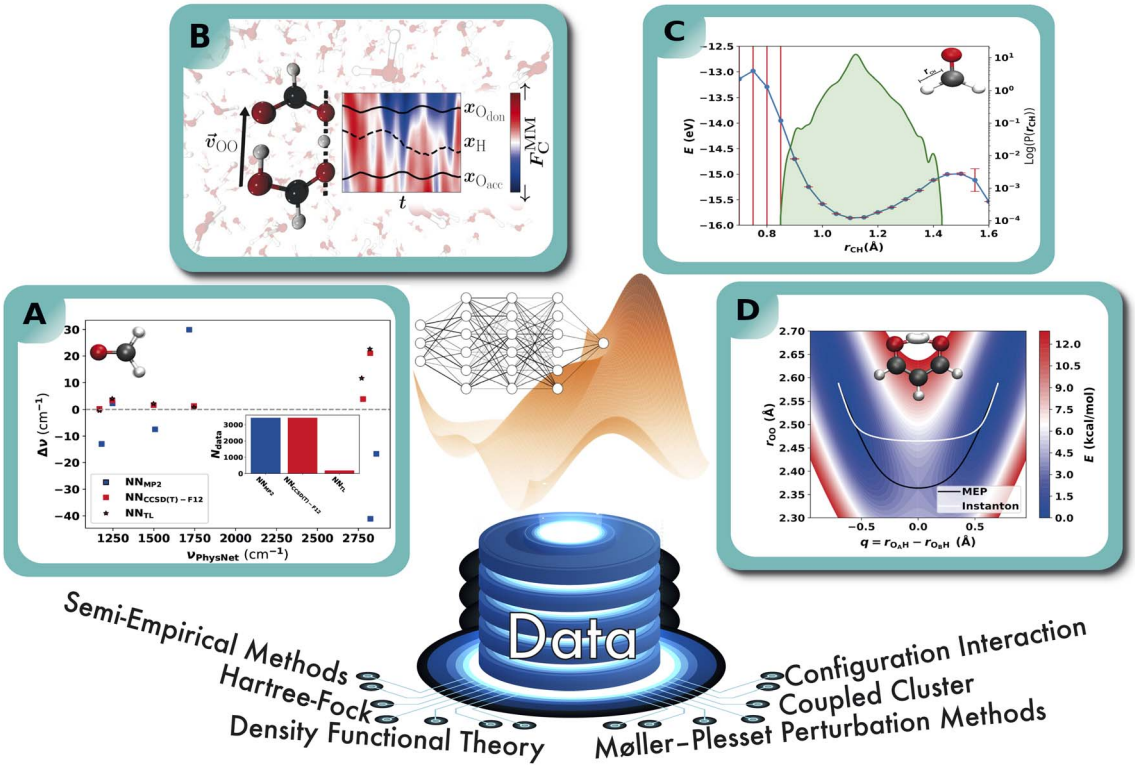

Fig. 6 Schematic representation of the exemplary applications of NNPs. A: performance of a NNPs based on MP2/aVTZ and CCSD(T)-F12/aVTZ-F12 with respect to experiment. NNPs trained from scratch are compared to the more-data efficient TL approach and the anharmonic frequencies are obtained from VPT2 calculations.[179] B: double proton transfer in formic acid dimer from mixed ML/MM/MD simulations.[192] The time series next to the molecular structure shows the variation in the background solvent field depending on time across one proton transfer event. C: 1D cut of the PES of the C–H bond in formaldehyde (upper right) calculated with the PhysNet evidential model (blue curve). Red bars indicate the predicted variance by the model. The green distribution shows the logarithm of the probability distribution of the distances covered by the training set. D: the two-dimensional projection of a NN-trained PES of CCSD(T) quality for proton transfer in malonaldehyde. The white and black traces are the instanton and minimum energy paths, and the PES is used to calculate tunneling splittings.[193]

model for the periodic systems are also arbitrarily shifted mainly due to the differences in the non-periodic and periodic computational system setup. MD simulation of a periodic water box at 300 K with the model potentials trained on clusters (cluster model) and periodic reference data (bulk model) produce radial distribution function that agree well and X-ray diffraction spectra are close to experimental ones. The computed water molecule self-diffusion coefficients and equilibrium density from simulations with the cluster model are about 18% larger ($2.15 \cdot 10^{-9}$ m^2 s^{-1} and 1.02 g cm^{-3}) than with the bulk model ($1.82 \cdot 10^{-9}$ m^2 s^{-1} and 0.86 g cm^{-3}) but closer to the respective experimental values ($2.41 \cdot 10^{-9}$ m^2 s^{-1} and 1.00 g cm^{-3}). Detached from the evaluation of the rev-PBE-D3 method and MD setup to accurately reproduce experimental water properties, the case study shows transferability of the cluster model to reproduce bulk properties. However, the authors mention that further studies are necessary to get insights into the deviation in the computed properties of both models as both water cluster and periodic water system are based on the same physical–mathematical description. Only water molecules closer to the cluster surface experience different strain energy than bulk water due to the lack of bonding partners.

Great transferability is also shown by Marx and coworkers using a HDNNP model trained on protonated water cluster $H^+(H_2O)_n$ (n = 1–4) with up to four water molecules to representing the PES of a protonated water hexamer $H^+(H_2O)_6$.[73,195] The reference data for the protonated water clusters n = 1–4 were produced by an automatic fitting procedure that performs DFT based *ab initio* MD and path integral MD (PIMD) simulation at 1.67, 100 and 300 K to sample relevant configurations. Within a repeated fitting procedure, holes in the reference data set are detected by estimating the uncertainty as described in section 4.2.1 or configurations were included where the local descriptors (ACSFs) of configurations in the MD simulation leave the range of the reference data set.[197] A final data set is created from reference data of the configurations computed at CCSD(T*)-F12a/aug-cc-pVTZ level of theory. Extrapolation of the NN model trained on the smaller cluster n = 1–4 to configuration of the protonated water hexamer yields a mean absolute energy error about three times higher than for the original training data set that is 0.026, 0.031, 0.038 kcal mol^{-1} (0.11, 0.13, 0.16 kJ mol^{-1}) per atom against 0.007, 0.010, 0.012 kcal mol^{-1} (0.03. 0.04, 0.05 kJ mol^{-1}) per atom from the sampling procedure at 1.67, 100 and 300 K, respectively.[195] Again, an arbitrary shift is added to the predicted energies of the

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

hexamer to minimize the error between the predicted and the reference energies. The ability to extrapolate is illustrated by comparing the potential energy sequence for 25 fs between an *ab initio* MD and the MD simulation using the NNP. It is further noticeable, that the extrapolation towards the hexamer potential failed in PIMD simulations for which unphysical configurations are reached if the NNP is trained only on tetramer configurations ($n = 4$). The authors conclude that the transferability towards larger cluster sizes improves if smaller clusters are included within the training data set.

6.3 Reaction rates

The reaction of methane with molecular oxygen is one of the most fundamental but highly complex combustion processes involving more than one hundred different reaction steps as shown by experiments.[198] Zhu and Zhang report MD results of the combustion reaction including 100 methane and 200 oxygen molecules at 3000 K simulated for 1 ns.[199] They used the DeepMD model potential that was fitted to reproduce 578731 reference DFT energies at the MN15 level of theory.[115,200] In their simulation they detected 505 molecular species and 798 different reactions where 130 reaction steps are also reported from experiments.[198] A selection of computed reaction rates deviates from experiment by up to two orders of magnitude, but combustion reactions usually involve the formation of radical species, that might require a non-adiabatic molecular dynamics approach which are highly non-trivial.

Marquetand and coworkers applied the SchNarc approach to investigate the photodissociation reaction of tyrosine that shows a dissociation channel of a hydrogen radical with a chemically non-intuitive path which is called roaming.[201] Roaming was originally explored experimentally and computationally in formaldehyde by Bowman and coworkers in 2004 but real-time experimental observation were not achieved until 2020.[202,203] The NNP is learned to reproduce 29 energy values and force values for electronic singlet and triplet states and 812 spin–orbit couplings. They simulated over 1000 trajectories of at least one picosecond which, in comparison, would take over eight years for *ab initio* MD simulation on a high-performance computer. About 17% of the trajectories show the roaming of the hydrogen atom in photoexcited tyrosine that lead to a higher ratio of subsequent further fragmentation than in non-roaming trajectories. This application marks a major step forward towards atomistic simulations of photoexcitation reactions in larger molecules like proteins that lead to further insight in, *e.g.*, photosynthesis, harmful photodegradation or drug designing for phototherapy.

6.4 Hybrid ML/MM simulations of solvated systems

The use of NNPs as force fields promotes the performance of MD simulations in comparison to the *ab initio* MD counterpart. But even if the computational cost of NNPs scales by a similar factor of $\sim O(N^{1-2})$ as empirical force fields do, due to their more compact and explicit functional form empirical force fields are considerably more efficient in general. Thus, a significant speed-up in MD simulations can be achieved by decomposing the force field into a contribution from a NNP (ML part) for, *e.g.*, a solute of interests or a reactive center in a protein, an empirical force field (MM part) for solvent molecules or protein backbone structures, and a coupling (or embedding) between the ML and MM parts. This approach is well known and applied in QM/MM MD simulations.[204]

One straightforward approach was pursued to investigate the double proton transfer reaction in cyclic formic acid dimers and the electrostatic impact of a water solvent on the reaction rate as shown in Fig. 6B.[192] Here, a PhysNet model was trained with a reference data set including formic acid dimers and monomers in the gas phase at MP2/aug-cc-pVTZ level of theory. The model accurately reproduces the energies, forces and molecular dipole by assigning atom centered charges.[51] The interaction potential between formic acid and the TIP3P water solvent consists of Lennard-Jones terms with parameters from the CGenFF[205] force field and electrostatic interactions between the atom charges from the TIP3P[206] water atoms and the configurational dependent PhysNet charges of the formic acid atoms. The advantage is the lower computational cost to produce trajectories with lengths of multiple nanoseconds to statistically sample the raw double proton transfer events with a rate of just 1 ns^{-1} at 350 K. Furthermore, the NNP fit inherently includes the coupling of the reactive potential path of the proton transfer with other structural dependencies such as the C–O bond order of the acceptor and donor oxygen and the dimer dissociation reaction into formic acid monomers. On the other hand, such an approach does not include the mutual polarization of the formic acid charges and the water solvent which, in the present case, is however expected to be small. This is akin to a mechanical embedding known from QM/MM schemes.[207]

Applications of electrostatic embedding in ML/MM simulation are reported by Riniker and coworkers as well as Gastegger and coworkers.[208,209] Here, the ML-MM interaction potential includes the polarization of the ML system by the electric field originating from the MM compounds. Riniker and coworkers modified the HDNNP by providing two sets of local descriptors for just ML solute atoms and surrounding MM solvent atoms, separately. The model is trained to reproduce either the ML atom potential and the electrostatic component of the ML–MM atom interaction itself (pure ML/MM) or in accordance of the Δ-learning approach an energy correction of both components to improve from computational cheap tight-binding DFT result towards more accurate reference data ((QM)ML/MM).[111,208] This approach demands larger reference data sets from QM calculations to sample solute configurations with different solvent distribution where the solvent is represented as their respective MM point charges. However, the Δ-learning (QM)ML/MM approach applied to tight-binding DFT computations have been shown to achieve higher accuracy even with fewer reference samples than the pure ML/MM model.

The accuracy is illustrated by running *NPT* simulations of *S*-adenosylmethionate and retinoic acid in explicit water solvent at 298 K and 1 bar using the pure ML/MM and the (QM)ML/MM model for 5000 and 2000 integration steps of 0.5 fs, respectively, and comparing it to reference QM/MM results.[208] The mean absolute error for the (QM)ML/MM model is up to one

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

magnitude lower with 1.4 kcal mol^{-1} (5.8 kJ mol^{-1}) and 12.6 kcal mol^{-1} (52.8 kJ mol^{-1}) than the pure ML/MM model with 4.3 kcal mol^{-1} (18.1 kJ mol^{-1}) and 17.9 kcal mol^{-1} (74.9 kJ mol^{-1}). One integration step with the (QM)ML/MM model takes less than a second on 1 CPU while the reference QM/MM model at DFT BP86/def2-TZVP level is about 3 magnitudes slower with about 60 to 80 minutes on 4 CPUs. A potential disadvantage of the (QM)ML/MM model is that certain solute configurations at the tight-binding DFT level may fail to converge or converge only slowly, *e.g.*, during a reaction.

Gastegger and coworkers presented the FieldSchNet model, a modification of the SchNet model that includes energy contributions from interactions between predicted atomic charges and dipoles, but also with an external field such as the electric field originating from a set of point charges.[85,209] The advantage of such elaborated models is the sensitivity of the potential energy to changes in atomic positions, electric and magnetic fields that enable the computation of response properties such as forces, molecular dipole moments, polarizabilities, and atomic shielding tensors that are crucial for the direct prediction of, *e.g.*, IR, Raman and NMR spectra. As the atomic charges and dipoles of the ML treated system respond to the external field caused by MM atoms point charges, this model is considered to be electrostatic embedding. Consequently, it has the same requirement for additional sampling of ML system configurations in different arrangements of MM atomic point charges as the model of Riniker and coworkers described above.[208]

For ethanol in vacuum, PIMD simulations with FieldSchNet yield excellent agreement in terms of frequency shifts and widths between predicted IR/Raman spectra and experimentally measured ones. For liquid ethanol, IR spectra were predicted from MD trajectories with an explicit ML/MM solvent model of one ML treated ethanol molecule in a MM treated ethanol solvent. The explicit ML/MM approach shows great agreement with experimental IR spectra in the low frequency region and a blue shift for the C–H and O–H stretch vibrations bands in the high frequency range due to missing anharmonicity effects by the MD approach. MD simulations with an implicit PCM solvent model do not yield an IR spectra with significant differences from gas phase spectra as it fails to capture hydrogen bridging between ethanol molecules.[210] However, the applied ML/MM model still predicts the intermolecular ML–MM potential between ML ethanol and the MM solvent by the CGenFF[205] force field with fixed atomic charges. The implementation of the electrostatic interaction between predicted atomic charges and dipoles by FieldSchNet and the MM point charges is a highly non-trivial task and would further increase the computational costs. It limits the application range to systems where the ML–MM interaction potential is sufficiently well described by the MM force field that may not work for dynamics with complex configurational changes or chemical reactions.

Electrostatic embedding in the QM/MM approach (and the ML/MM approach)[208] includes the QM-MM electrostatic interaction and the polarization of the QM system by the electric field of the MM atoms but not *vice versa*. The highly expensive task to approximate the polarization of the MM system by the electric field of the QM system is part of polarizable embedding schemes.[211] An analogue for the hybrid ML/MM model is developed Westermayr, Oostenbrink and coworkers with their buffer region NN approach (BuRNN).[212] Here, a buffer region around the ML atoms is defined by a cutoff sphere to select MM atoms within the sphere. The ML and selected MM atoms are the input to a modified SchNet model to predict the potential energy between the ML atoms, the ML-MM interaction energy and a polarization correction energy to the classical MM potential of the MM atoms within the buffer sphere to match reference potential data. The modified SchNet model also predicts atomic point charges for the ML atoms and MM atoms within the buffer region, which are used to compute the electrostatic interaction to the remaining MM atoms in the system outside the buffer region. The potential energy of the atoms in the inner region are predicted by a modified SchNet model. As for electrostatic embedding, potential energy and charge distribution of the ML system are impacted by the MM atoms within a buffer region and, additionally, interaction energy and atomic charges of the respective MM atoms are impacted by the ML system. A major disadvantage is the high computation cost for the reference data set, that requires two quantum electronic calculation for configuration samples of (1) the ML system and MM atoms in the buffer region and (2) the MM atoms in the buffer region alone to predict the polarization correction term.

The BuRNN approach was applied to a hexa-aqua iron(III) complex simulated by a ML treated Fe^{3+} ion in a water solvent described by the SPC model. A buffer region was defined by a cutoff radius of 5 Å around the Fe^{3+}. MD simulation of 10 ns shows smooth diffusion of water molecules entering and leaving the buffer region and reveal power spectra that match the low frequency bands around 180, 310 and 500 cm^{-1} observed in experiments very well. Radial and improper and distributions between Fe^{3+} and the oxygens of the coordinated water match with distributions from QM/MM simulation with electrostatic embedding and are within experimental estimations.

All the presented applications show an active field of developments in hybrid ML/MM approaches towards accurate MD simulation of solutes or reactive species in the presence of a solvent. A major gain in computational efficiency and much longer simulation times at comparable accuracy are achieved by replacing QM methods with a NNP. However, the effort to generate a reference data set that sufficiently samples the relevant configurational space of the ML system in combination with different solvent configuration depends significantly on the embedding scheme. The simplest mechanical embedding scheme only requires a converged NNP that predicts the total energy, forces and the charges of the ML system in the gas phase but it neglects polarization of the MM atoms.[192] In comparison, NNPs based on electrostatic embedding require additional sampling with MM atom configurations included as point charges. MD simulation using ML/MM approaches with electrostatic embedding show great agreement with MD simulation of respective QM/MM simulation at the same level of theory as the reference data set.[208] The increase in the quality to describe

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

the impact of the MM solvent on the properties of the ML system is also demonstrated by accurate computational reproduction of experimental IR and Raman spectra.[209] The most complex polarization embedding scheme allows the most complete description of the ML system with the MM environment, but requires more costly reference computations.[212] Even a QM/MM model using polarization embedding is significantly more challenging in terms of computational effort and implementation than the electrostatic embedding schemes.[211]

7 Applications based on but beyond PESs

Up to this point PESs were used in explicit simulations to determine experimental observables from dynamics or Monte Carlo simulations. However, quantum nuclear dynamics or a statistically significant number of (quasi) classical MD simulations and their analysis is often a computationally demanding endeavor in itself. It would be desirable to determine, predict or estimate observables from only a limited amount of such explicit simulations and devise rapidly-to-evaluate models that predict with confidence outcomes for arbitrary input. To set the stage, the full characterization of all state-to-state cross sections for reactive triatomic systems A + BC → AB + C is considered. This problem involves $\sim 10^8$ transitions. Using QCT simulations, convergence of each of the cross sections requires $\sim 10^5$ independent trajectories to be run. Hence, for one collision energy $\sim 10^{13}$ QCT simulations would be required for a full characterization of a reactive triatomic system. This is neither desirable nor meaningful to do. Hence, despite the availability of a full-dimensional NN-based or otherwise represented PES it would be advantageous to reduce the computational burden of explicitly sampling the PES in this case and the task is to extract as much information as possible from only a limited number of simulations.

The two problems considered further below concern the prediction of final states or final state distributions for atom + diatom reactions and predicting thermal rates for bimolecular reactions. Both problems can, in principle, be solved accurately for carefully chosen systems which provides the necessary benchmark to extend the range of applicability of the approaches described below to larger systems.

7.1 Final state distributions for atom + diatom reactions

Exhaustive enumeration and characterization of final state distributions from bimolecular reactions is particularly relevant in combustion and atmospheric re-entry (hypersonics). The particular interest is rooted in need to devise more coarse-grained models for the macroscopic (in space and time) modeling of the chemistry and physics of reactive flows but based on accurate microscopic information.[213,214] For atom + diatom reactions (A + BC → AB + C) this involves complete enumeration of all state-to-state reaction probabilities. As mentioned above, this problem can - in principle – be addressed by brute-force sampling. But this is neither practical nor desirable.

For this reason, ML-based models were devised that allow to either predict final states or final state distributions from discrete initial states. From quasiclassical trajectory (QCT) simulations for the $N(^4S) + NO(^2\Pi) \rightarrow O(^3P) + N_2(X\cdot {}^1\Sigma_g^+)$ reaction the state-to-state cross sections $\sigma_{v,j\rightarrow v'j'}(E_t)$ as a function of the translational energy E_t were explicitly determined for 1232 initial ro-vibrational states (v,j) which amounted to $\sim 10^8$ QCT trajectories in total. This compares with an estimated 10^{15} QCT trajectories required for brute-force sampling of the problem. This information was used as input to train a NN together with features such as the internal energy, the vibrational and rotation energy of the diatoms, or the turning points of the diatoms.[215] The resulting state-to-state (STS) model is capable of predicting the cross section for a final state given an initial collision energy, the vibrational state v of the diatom and its rotational quantum number j. More recently, the approach was extended to predict entire final state distributions from discrete initial conditions, which led to the state-to-distribution (STD) model.[216] Finally, it is also possible to devise distribution-to-distribution (DTD) models.[217]

The prediction quality of STS, DTD, and STD models is universally high and reaches a correlation coefficient $R^2 \sim 0.98$ or better between predicted and QCT-calculated reference data. From these models it is also possible to determine thermal rates as done for the $N(^4S) + O_2(X\cdot {}^3\Sigma_g^-)$ reaction shown in Fig. 7 and further examples are given below. Comparison with rates directly determined from QCT simulations - which themselves are in good agreement with experiments[221–223] – shows that the trained NNs reach accuracies better than 99% over a wide temperature range $(1000 \leq T \leq 20\,000)$ K. Thus, ML-based models based on limited input data from direct simulations on high-quality, full-dimensional PESs are a computationally efficient and accurate substitute for explicit, brute-force evaluations of the relevant properties.

7.2 Predicting thermal rates

Determining thermal rates is one of the major goals of computational chemistry. Carrying out such a calculation in full dimensionality, based on an accurate PES and including nuclear quantum effects is a serious computational undertaking. An accurate rate requires treating the electronic structure, representing the underlying PES, and running the (quantum) dynamics simulations at the highest possible levels and has only been done for a few selected systems. Hence, it is of great interest to develop models that can predict thermal rates based on alternative approaches.

One such effort was based on a library of ~40 bimolecular reactions for which T-dependent rates from transition state theory (TST), the Eckart correction to TST, and a set of tabulated "accurate rates" from two-dimensional calculations at 8 temperatures were available.[224] These calculations required a represented PES for carrying out the necessary dynamics simulations. The data collected was used to learn a correction to the product of the TST-rate and the Eckart correction by using Gaussian process regression. Reactions considered included the Cl + HCl H-atom exchange reaction (in 1d and 3d), the H_2+OH → H + H_2O and for O + CH_4 → OH + CH_3 which was

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

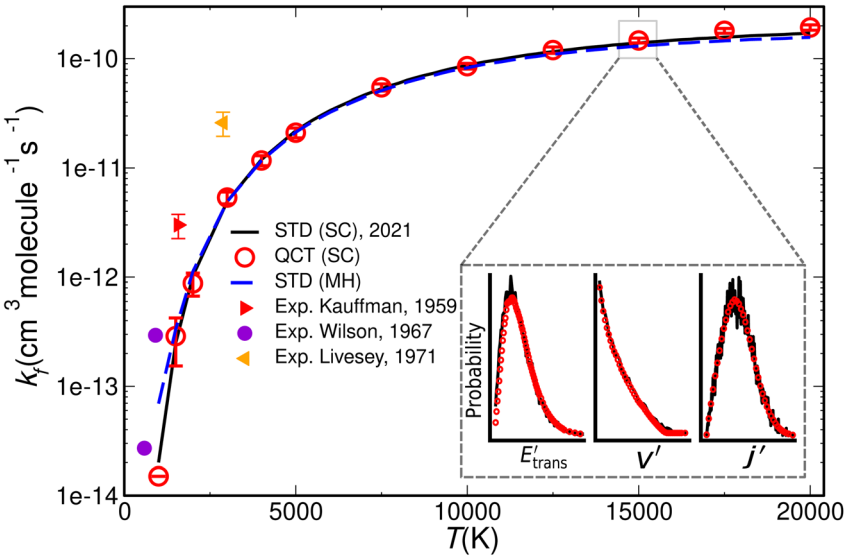

Fig. 7 The thermal forward rate k_f calculated from QCT (open red circle) and STD model (solid black line) for the $^4A'$ state of the N(^4S) + $O_2(X\cdot{}^3\Sigma_g^-)$ → NO($X^2\Pi$) + O(^3P) reaction between 1000 and 20 000 K. Experimental total forward reaction rate k_f (including contributions from the doublet and the quartet states) are also shown for comparison: (red triangle),[218] (orange triangle)[219] and (magenta circle).[220] A comparison is made between QCT and STD model based on model Hamiltonian E_{trans} (dash blue line) for the predicted distributions in the bottom right corner (inset). The evaluation is made at $T = 15\,000$ K with QCT and STD evaluations marked as black and red solid lines respectively. Figure courtesy J. C. San Vicente Veliz.

investigated more in-depth in a separate study.[225] The results for reactions not used in the learning procedure indicate that it is possible to obtain thermal rates close to those from explicit quantum simulations or trajectory-based quantum calculations (ring polymer MD).[226]

7.3 Other applications

In one recent application a mapping between local water cluster arrangement and the frequency of an embedded solute as the spectroscopic probe was used to predict water anharmonic stretch vibrations.[227] Although this application is not dependent on and does not require a full-dimensional NN-based PES it illustrates the potential uses of a mapping between structure and spectroscopy that can be exploited in the future. Another area which links intermolecular interactions, structural dynamics and spectroscopy are ionic and eutectic liquids (ILs and ELs). A strong case for combining rigorous MD simulations with accurate, ML-based FFs for property prediction has been made for ionic liquids.[228] ILs and ELs are characterized by strong interactions that probe the short-range part of electrostatics due to the chemical composition of the systems which consists of a high density of positively and negatively charged building blocks. For ELs a recent combination of MD simulations, two-dimensional infrared and terahertz spectroscopy was able to elucidate the microscopic structure of the liquid depending on the degree of hydration without, however, using a ML-based FF.[229] Further improved agreement between simulations and experiments than that reported can be expected from refined intermolecular interactions.

8 Challenges

This section discusses several challenges the field of NN-based PESs faces. Some of the points discussed may also apply to other ML-based techniques more broadly in other branches of chemistry. As a very general opening point it is noted that one of the challenges in statistical approaches is to extract as much consolidated information, potentially including an error estimate on the prediction, from a statistical model from as little information possible. This point concerns very broadly the aspect of "data efficiency".

8.1 Data management and availability

Given the tremendous computational cost and effort needed for generating robust and high quality reference data sets for PES fitting, data management and availability is a fundamental focus. Yet, the raw *ab initio* data (nuclear geometries, energies (and gradients)) of a published PES is often not publicly available, incomplete or lacks key information such as a precisely specified level of theory or the employed quantum chemical software. This could be avoided by publishing exemplary input files alongside the *ab initio* results. Some of the most popular data sets used for benchmarking NN potentials contain only equilibrium geometries and corresponding energies from different levels of theory and are used to benchmark ML methods. These include the QM7,[19] QM7b,[230] QM9,[231] and ANI-1ccx[232] databases. Databases that contain energy and gradients for equilibrium and distorted structures for different molecules include ANI-1,[233] the refined ANI-1x[232] and QM7-X.[234] A popular data set that provides energies and gradients for configurations

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

visited in MD simulations is the MD-17 dataset[235,236] which is generated from *ab initio* MD.

On a cautionary note regarding publicly available datasets, it was reported that PESs resulting from the MD-17 data are likely to feature holes in high-energy regions which are visited for example in DMC simulations.[237,238] Databases such as ANI-1 (ref. 233) which uses normal mode sampling for multiple species also can generate problems. Recently, it was found that redundancies in databases can compromise the prediction quality of NN models exploring chemical space.[14] For training NNPs, the influence of the distribution of the reference points on the quality of the PES is an open question. Recent efforts in providing data sets for rigorous and global PES gave rise to the VIB5 (ref. 239) and QM-22 (ref. 238) databases that include energies (and gradients) for different molecules calculated at various levels of theory.

An often overlooked step in generating databases is the prepossessing step. It is advisable that the generated data contains as little redundancies as possible by removing correlated states to reduce the number of *ab initio* calculations and training time. Therefore, the generated database can, *e.g.*, be analyzed beforehand by unsupervised machine learning methods which have been successfully applied to evaluate MD trajectories.[240,241] It is also important to consider that the generation of data must be application driven because the properties of interest will determine the amount of data required and should guide the selection of the sampling method. Data generation for NN-based PESs should be considered an iterative process in which it is best to start from a representative and "clean" data set that will be enriched based on the problem at hand as was recently done for tunnelling splittings in malonaldehyde.[193]

Finally, ML models are starting to face some of the same difficulties that the molecular simulation community has been dealing with.[242] This includes the lack of standard file formats, shortage of tools for file sharing, absence of methods to ensure the quality of the generated databases, *etc.* Hence, it is worth mentioning that the young ML community has the unique opportunity to propose solutions to these obstacles before they become unbearable. In this regard, the FAIR principle[243] (findable, accessible, interoperable and reproducible/reusable) must be taken into account. In this regard, some authors have proposed general rules for the application of ML in chemistry[244] and in particular to PESs for small molecules.[245] Specifically, Li and Liu[245] proposed a checklist for reporting PESs of high-quality. As a complement to this, we propose some suggestions for providing data sets underlying molecular PESs. Data sets should:

- Provide sample input and output files for the quantum chemical software.
- Have an easy and understandable format.
- Have a consolidated structure.
- Contain raw data (at least nuclear geometries, energies (and forces)) with clearly defined units, level of theory, employed quantum chemical software.
- Have a clear description of HOW the geometries were generated.
- If possible, provide information whether the PES was developed for a particular purpose/application and whether there a known limitations.
- Be extensible.

8.2 Interpretability

An important ingredient for extending NN methods is the degree and confidence with which a human can understand the relationship between cause (starting database and model) and effect (result or observation, applying the model to a new task).[246,247] This process has also been called "interpretability", and it can be used to assess the relationships learned by the model or contained in the data used for training.[248,249] However, for complex models like NNs the relationship between input and output is not clear as a consequence of the non-linearity and parametric complexity of the models.[56] Therefore, it is not evident if the model is deriving the correct physics of the system from the provided data or whether it is only learning artefacts of the data which limits it's application to narrow settings in what is known as the "clever Hans" predictor.[250] Only a few efforts have been made to derive techniques that can relate the contribution of different structural components (atom, bond type) to the predicted quantity (energy or dipole moment).[251]

Despite its importance and need, interpretability is still not a main topic in developing NNPs. A reason for this might be that the use of conventional techniques is not possible because of the continuous nature of the properties studied in chemistry.[252] However, general guidelines have been proposed.[253] By definition, interpretability is the missing link between the data used for training and the prediction obtained by the NNP. A better understanding of the inner processes of NNs will help to better understand the amount of data required to obtain reliable predictions, understand the completeness of the descriptor, and maybe even some new physical interactions. In contrast, the largest risk that the lack of interpretability presents is that users employ models as "black box" therefore without knowing the limitations of the model and possibly obtaining good results for the wrong reason(s).

8.3 Generation of robust initial models

A NNP is only as good as the data it is trained on. As a consequence, if low-quality data is used the resulting model will under-perform. This is the principle of "Garbage in-Garbage Out" which can be traced back to Charles Babbage.[254] The NNP fitting is usually an iterative process starting from an initial reference data set. This data set ideally covers the full configurational space of the chemical system at hand with as few points as possible (note that, in principle, the number of points on a PES as well as in chemical space is ∞). While an exhaustive sampling of a PES might be possible for systems with up to 3 atoms (*e.g.* by choosing configurations on a regular grid), this becomes impossible for larger systems. Consequently, the initial sampling relies on (partly random) methods including MD or normal mode sampling (see Section 4) that all suffer from distinct weaknesses/disadvantages such as correlated structures or insufficient coverage. These weaknesses lead to

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

additional training time, evaluations, *ab initio* calculations and ultimately to a slower and more expensive convergence of the iterative NNP fitting procedure.

Thus, the generation of data for PESs requires improved methods of (initial) sampling that can warrant sufficient coverage of the PES for a desired application with as few points as possible. An interesting prospect for the generation of PES reference data concerns spreading the data according to the "correct" distribution for different degrees of freedom resulting from methods like Boltzmann[255,256] or Monte Carlo inversion[257] and opens the possibility of deriving interactions from experiments.[258] Other solutions might come from the application of information theory to ensure a number of samples with the maximum amount of information. Alternatively, the use of similarity measures between the initial structures before the actual running of *ab initio* calculations can be a tool to obtain representative structures of the PES. However, the problem of how to best choose initial structures for NNP generation is still open.

On the other hand, the processing of information by the model can be enhanced to facilitate the convergence of the model, make it more data efficient and reduce the dependency on the initial points. This has been explored for equivariant NNs which complement the description of the interactions in the message step of MPNNs (see Section 3.2). Equivariant NNs have been proven to be very data efficient by obtaining an accuracy comparable to the best NNPs using only a fraction of the data that other methods require.[96] As a complement to this strategy, it is possible to obtain data efficient models by including more physics-based information which has been proved to perform better than regular approaches for kernel methods.[259]

8.4 Reliable active learning and uncertainty quantification

A complete exploration of a PES is a challenging task that most likely can not be done in a single step and depends heavily on the application. Therefore, the improvement of PESs is an active topic of research. Algorithms for systematically improving a training dataset are known as "active learning" techniques. Active learning is closely related to uncertainty quantification of the predictions, which by itself is an active area of research. For NNPs, the most common technique for obtaining the uncertainty is the training of ensembles of NNs which are then averaged for the prediction of identical points. This procedure has a high computationally price because it requires the training and evaluation of several NNPs. As mentioned before, ensemble methods present a clear drawback because their estimated uncertainty can only quantitatively relate to the observed error.[162]

Other methods of uncertainty quantification like Bayesian NNs, which impose a prior distribution to each of the parameters of a NNP are computationally too expensive for practical use.[260] However, Gaussian processes are a limiting case of Bayesian NNs,[261] which have been extensively used and applied for the refinement of PESs by means of UQ.[262,263] Therefore, a combination of NNPs and Gaussian process regression is a promising avenue for UQ in NNPs. Another approach for UQ is single network deterministic methods[163,264] which make assumptions about the distribution of the data. These methods appear to be a promising alternative to the mentioned problems by obtaining the uncertainty by training and evaluating a single model (see Fig. 6C). However, it should be noted that single network models are strongly influenced by the initial assumptions and it is necessary to calibrate the model beforehand. The need for adjustments is not an exclusive problem of single network models. All the previously described methods require a step of calibration in order to assure that the predicted uncertainties can be related to the observed error. Finally, it should be mentioned that active learning techniques without uncertainty quantification have not been tested.[265]

8.5 Extrapolation outside the training set covered

One of the major drawbacks of NNs is their limited capability to extrapolate in general beyond the training data.[266] For the case of NNPs this means that evaluating energies and forces for structures not covered in the training/validation are likely to lead to a severe breakdown of the model. This weakness stems from the fact that the functional form lacks a physical basis and is a pure mathematical fitting procedure.[267] This is different for methods such as reproducing kernels (RKHS) and PIPs. RKHS allows to choose kernel functions to follow the physics of the long-range part of the intermolecular interactions.[27,28,268–270] PIPs make use of Morse variables (*i.e.* internuclear distances are usually transformed to Morse variables) which decay to zero for large distances giving the PES fit a qualitatively correct asymptotic behaviour.[271] However, to obtain the correct long-range behaviour, PIP PESs often employ switching functions.[33,272] The inability of extrapolation for NNPs is often revealed at early stages of the NNP generation and can, *e.g.*, be expressed by unphysical short interatomic distances or by a partial or entire fragmentation of the system.[111] Thus, a possible route for improvement is to include explicit physical knowledge, *e.g.*, on the long-range electrostatic interactions,[51,90,91] dispersion corrections,[51] or on nuclear repulsion.[89] Such extensions are likely to allow extrapolation beyond the training data. Besides the extrapolation in configurational space, the extrapolation and transferability across chemical space is of concern.

8.6 Enhancing PESs to higher levels of theory

Transfer learning and Δ-ML is a comparatively new concept for theoretical chemistry and solid evaluations are needed. One of the questions that arises is how to validate the quality of a TL-PES if single point calculations become increasingly expensive. In other words, if the effort to carry out one single point *ab initio* calculation for the HL model required for TL becomes too large, it is preferable to keep this data in the training set for TL instead of using it for testing. This certainly gives rise to the question as to how to probe and validate the NNP for regions that lie outside of the TL data set. One possible strategy to test the improvement of the HL PES with respect to the LL PES is to calculate an observable, compare it to experiment and check for

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

a convergence towards the experiment, as was done by some of us for the determination of tunneling splittings.[193]

Another open question is what the lowest possible level of theory for the LL-model is which still allows reliable TL to a HL-model. The answer to this question will depend on the system and application considered. Ideally, Hartree–Fock calculations would be a suitable surrogate model for TL to CCSD(T) levels of theory, but this needs to be explored for specific systems.[170]

Finally, since the computational cost of the quantum chemical calculations can be appreciable, again the judicious selection of molecular structures for which HL calculations are carried out for TL is crucial. While no simple answer to this question exists as of now, the structures are usually carefully chosen with human intervention. Alternatively, it is conceivable that an approach similar to on-the-fly ML[160,161] ("on-the-fly TL") could be used to select data points to include in the TL data set.

8.7 Other challenges

Finally, a number of other challenges are briefly summarized. With the ever increasing quality of NNPs a better understanding of the relationship between the accuracy of a NNPs based on reference data for a given quality of the electronic structure and the observables determined from simulations using this PES is required. Ultimately, this requires a direct comparison with experiment. This raises the question whether it is possible to determine the underlying PES from inverting the relationship between observables and interaction potential, *e.g.* by using invertible NNs.[273–275] Such an inversion has been done successfully for low-dimensional systems. The Rydberg–Klein–Rees (RKR)[276–278] and rotational RKR (RRKR)[279] procedures are examples for this. However, for high-dimensional systems, this is a formidable task and will require a large number of high-quality data. With respect to the quality of the trained models, more informative statistical measures should be developed because those used at present often hide poor performance in individual structures.

Another challenge ahead is the seamless integration of NNPs – or ML models in general – into standard MD simulation packages while not compromising their computational efficiency. Further improvements of NN-based interaction potentials can be expected from using physics-informed NNs.[280,281] Another possibility is to explore the combination of NN-based representations at short range with physics-based long range models based on multipolar and/or polarizable models.

Technically, the question arises how complete descriptors need to be for a comprehensive and accurate representation of the intermolecular interactions *i.e.* what is a meaningful balance between the size of the descriptor(s) and the accuracy of the final model? Additionally, recent advancement in quantum computing technologies provides opportunities to further reduce the computational cost for generating, training and applying NNPs.[282] Still, whether and how these developments will impact how NNPs evolve and are being used is an open question.

On the more societal side, it is noted that constructing a full-dimensional PES for one given molecule is often a computational investment that requires appreciable resources. Hence, the environmental impact of this should be considered as well.[283–285] Generally, all ML-based PESs require the *ab initio* computation of information (energies, forces, or both) for thousands of nuclear geometries followed by the training of a model which incurs appreciable environmental cost.†.

9 Conclusion

The field of NNPs has reached a considerable degree of maturity in conceiving PESs that can be used in concrete applications, be it within the exploration of individual structures or in dynamics-based studies. Also due to the tremendous progress in efficiency of electronic structure calculations, it is now possible to determine full-dimensional – not necessarily "global" – potential energy surfaces for medium-sized molecules at levels of theory that allow direct comparison and in some cases even prediction of experimental observables. This, combined with techniques such as transfer learning holds promise to design yet improved PESs.

On the other hand, a rather unexplored facet of NNPs concerns questions about the interpretation of the underlying NN from a chemical perspective, aspects relating to the optimal distribution of reference points including minimizing the number of such calculations, or transferring PESs from one chemical system to a related species without recomputing all reference information afresh. Solutions to these questions will considerably increase the efficiency for conceiving and evaluating NNPs, and improve the prospects for generalizing trained models to broader chemistries and applications.

The present contribution aims at consolidating the available technical approaches, their use in constructing PESs and their application in concrete molecular simulations. It is hoped that this will provide a basis for further development because the prospects of NNPs are bright and the future for them is open.

Data availability

As this is a perspective article, no primary research results, data, software or code have been included.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the University of Basel, the Swiss National Science Foundation through grants 200021-117810, 200020-188724, the NCCR MUST, the Air Force Office for Scientific Research (AFOSR), and the European Union's Horizon 2020 research and innovation program under the Marie

† An interesting tool to check the CO_2 production of your algorithms can be found at: http://www.green-algorithms.org/

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Skłodowska-Curie grant agreement No 801459 - FP-RESOMUS. The authors acknowledge the help of Juan Carlos San Vicente Veliz for the preparation of Fig. 7.

References

1 T. Zhou, Z. Song and K. Sundmacher, Big Data Creates New Opportunities for Materials Research: A Review on Methods and Applications of Machine Learning for Materials Design, *Engineering*, 2019, **5**, 1017–1026.
2 P. Friederich, F. Häse, J. Proppe and A. Aspuru-Guzik, Machine-learned potentials for next-generation matter simulations, *Nat. Mater.*, 2021, **20**, 750–761.
3 O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, Machine learning force fields, *Chem. Rev.*, 2021, **121**, 10142–10186.
4 G. E. Moore, Cramming more components onto integrated circuits, *Proc. IEEE*, 1998, **86**, 82–85.
5 V. Vapnik, *The nature of statistical learning theory*, Springer science & business media, 1999.
6 M. Meuwly, Machine learning for chemical reactions, *Chem. Rev.*, 2021, **121**, 10218–10239.
7 A. Agrawal and A. Choudhary, Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science, *APL Mater.*, 2016, **4**, 053208.
8 A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi and M. Ceriotti, Machine learning unifies the modeling of materials and molecules, *Sci. Adv.*, 2017, **3**, e1701816.
9 P. Sinz, M. W. Swift, X. Brumwell, J. Liu, K. J. Kim, Y. Qi and M. Hirn, Wavelet scattering networks for atomistic systems with extrapolation of material properties, *J. Chem. Phys.*, 2020, **153**, 084109.
10 A. Aspuru-Guzik, R. Lindh and M. Reiher, The matter simulation (r)evolution, *ACS Cent. Sci.*, 2018, **4**, 144–152.
11 F. Noé, A. Tkatchenko, K.-R. Müller and C. Clementi, Machine learning for molecular simulation, *Annu. Rev. Phys. Chem.*, 2020, **71**, 361–390.
12 J. Behler and G. Csányi, Machine learning potentials for extended systems: a perspective, *Eur. Phys. J. B*, 2021, **94**, 142.
13 S. Shaik, H. S. Rzepa and R. Hoffmann, One molecule, two atoms, three views, four bonds?, *Angew. Chem., Int. Ed.*, 2013, **52**, 3020–3033.
14 L. I. Vazquez-Salazar, E. D. Boittier, O. T. Unke and M. Meuwly, Impact of the Characteristics of Quantum Chemical Databases on Machine Learning Prediction of Tautomerization Energies, *J. Chem. Theory Comput.*, 2021, **17**, 4769–4785.
15 E. Kocer, T. W. Ko and J. Behler, Neural Network Potentials: A Concise Overview of Methods, *Annu. Rev. Phys. Chem.*, 2022, **73**, 163–186.
16 C. Qu, Q. Yu and J. M. Bowman, Permutationally Invariant Potential Energy Surfaces, *Annu. Rev. Phys. Chem.*, 2018, **69**, 151–175.
17 W. Pronobis and K.-R. Müller, *Machine Learning Meets Quantum Physics*, Springer, 2020, pp. 25–36.
18 B. Huang and O. A. von Lilienfeld, Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity, *J. Chem. Phys.*, 2016, **145**, 161102.
19 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. Von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, *Phys. Rev. Lett.*, 2012, **108**, 058301.
20 K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. Von Lilienfeld, K.-R. Müller and A. Tkatchenko, Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space, *J. Phys. Chem. Lett.*, 2015, **6**, 2326–2331.
21 F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. Von Lilienfeld, Prediction errors of molecular machine learning models lower than hybrid DFT error, *J. Chem. Theory Comput.*, 2017, **13**, 5255–5264.
22 B. Huang and O. A. von Lilienfeld, Quantum machine learning using atom-in-molecule-based fragments selected on the fly, *Nat. Chem.*, 2020, **12**, 945–951.
23 A. S. Christensen, L. A. Bratholm, F. A. Faber and O. Anatole von Lilienfeld, FCHL revisited: Faster and more accurate quantum machine learning, *J. Chem. Phys.*, 2020, **152**, 044107.
24 A. P. Bartók, R. Kondor and G. Csányi, On representing chemical environments, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
25 F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi and M. Ceriotti, Physics-inspired structural representations for molecules and materials, *Chem. Rev.*, 2021, **121**, 9759–9815.
26 V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, Gaussian process regression for materials and molecules, *Chem. Rev.*, 2021, **121**, 10073–10141.
27 T.-S. Ho and H. Rabitz, A general method for constructing multidimensional molecular potential energy surfaces from *ab initio* calculations, *J. Chem. Phys.*, 1996, **104**, 2584–2597.
28 O. T. Unke and M. Meuwly, Toolkit for the construction of reproducing kernel-based representations of data: Application to multidimensional potential energy surfaces, *J. Chem. Inf. Model.*, 2017, **57**, 1923–1931.
29 M. A. Collins, Molecular potential-energy surfaces for chemical reaction dynamics, *Theor. Chem. Acc.*, 2002, **108**, 313–324.
30 P. Lancaster and K. Salkauskas, Surfaces generated by moving least squares methods, *Math. Comput.*, 1981, **37**, 141–158.
31 R. Farwig, Multivariate interpolation of arbitrarily spaced data by moving least squares methods, *J. Comput. Appl. Math.*, 1986, **16**, 79–93.
32 J. D. Bender, S. Doraiswamy, D. G. Truhlar and G. V. Candler, Potential energy surface fitting by a statistically localized, permutationally invariant, local

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

interpolating moving least squares method for the many-body potential: Method and application to N_4, *J. Chem. Phys.*, 2014, **140**, 054302.
33 B. J. Braams and J. M. Bowman, Permutationally invariant potential energy surfaces in high dimensionality, *Int. Rev. Phys. Chem.*, 2009, **28**, 577–606.
34 W. Mizukami, S. Habershon and D. P. Tew, A compact and accurate semi-global potential energy surface for malonaldehyde from constrained least squares regression, *J. Chem. Phys.*, 2014, **141**, 144310.
35 R. Dawes and E. Quintas-Sánchez, The construction of *ab initio*-based potential energy surfaces, *Rev. Comput. Chem.*, 2018, **31**, 199–263.
36 P. L. Houston, C. Qu, A. Nandi, R. Conte, Q. Yu and J. M. Bowman, Permutationally invariant polynomial regression for energies and gradients, using reverse differentiation, achieves orders of magnitude speed-up with high precision compared to other machine learning methods, *J. Chem. Phys.*, 2022, **156**, 044120.
37 W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.*, 1943, **5**, 115–133.
38 F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, *Psychol. Rev.*, 1958, **65**, 386.
39 M. Minsky and S. Papert, *Perceptron: an introduction to computational geometry*, MIT Press, 1969.
40 J. O'Connor, Undercover Algorithm: A Secret Chapter in the Early History of Artificial Intelligence and Satellite Imagery, *Int. J. Intell.*, 2022, 1–15.
41 D. E. Rumelhart, G. E. Hinton and R. J. Williams, Learning representations by back-propagating errors, *Nature*, 1986, **323**, 533–536.
42 J. Gasteiger and J. Zupan, Neural Networks in Chemistry, *Angew. Chem., Int. Ed.*, 1993, **32**, 503–527.
43 J. Zupan and J. Gasteiger, *Neural networks in chemistry and drug design*, John Wiley & Sons, Inc., 1999.
44 B. G. Sumpter and D. W. Noid, Potential energy surfaces for macromolecules. a neural network technique, *Chem. Phys. Lett.*, 1992, **192**, 455–462.
45 T. B. Blank, S. D. Brown, A. W. Calhoun and D. J. Doren, Neural network models of potential energy surfaces, *J. Chem. Phys.*, 1995, **103**, 4129–4137.
46 J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, *Phys. Rev. Lett.*, 2007, **98**, 146401.
47 T. W. Ko, J. A. Finkler, S. Goedecker and J. Behler, A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer, *Nat. Commun.*, 2021, **12**, 1–11.
48 K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller and A. Tkatchenko, Quantum-chemical insights from deep tensor neural networks, *Nat. Commun.*, 2017, **8**, 13890.
49 K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, SchNet–A deep learning architecture for molecules and materials, *J. Chem. Phys.*, 2018, **148**, 241722.
50 J. S. Smith, O. Isayev and A. E. Roitberg, ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost, *Chem. Sci.*, 2017, **8**, 3192–3203.
51 O. T. Unke and M. Meuwly, PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges, *J. Chem. Theory Comput.*, 2019, **15**, 3678–3693.
52 M. Born and R. Oppenheimer, Zur quantentheorie der molekeln, *Ann. Phys.*, 1927, **389**, 457–484.
53 D. J. Tannor, *Introduction to quantum mechanics: a time-dependent perspective*, University Science Books, 2007.
54 F. Jensen, *Introduction to computational chemistry*, John wiley & sons, 2017.
55 E. Lewars, *Computational chemistry: Introduction to the theory and applications of molecular and quantum mechanics*, Springer, 2nd edn, 2011.
56 J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller and A. Tkatchenko, Combining machine learning and computational chemistry for predictive insights into chemical systems, *Chem. Rev.*, 2021, **121**, 9816–9872.
57 G. Cybenko, Approximation by superposition of sigmoidal functions, *Math. Control Signals Syst.*, 1989, **2**, 303–314.
58 K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks*, 1991, **4**, 251–257.
59 K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, 1989, **2**, 359–366.
60 K. Shailaja, B. Seetharamulu and M. Jabbar, Machine learning in healthcare: A review, *2018 Second international conference on electronics, communication and aerospace technology*, ICECA, 2018, pp. 910–914.
61 G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken and C. I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.*, 2017, **42**, 60–88.
62 S. Grigorescu, B. Trasnea, T. Cocias and G. Macesanu, A survey of deep learning techniques for autonomous driving, *J. Field Robot.*, 2020, **37**, 362–386.
63 D. Guest, K. Cranmer and D. Whiteson, Deep Learning and Its Application to LHC Physics, *Annu. Rev. Nucl. Part. Sci.*, 2018, **68**, 161–181.
64 G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.*, 2019, **91**, 045002.
65 C. Angermueller, T. Pärnamaa, L. Parts and O. Stegle, Deep learning for computational biology, *Mol. Syst. Biol.*, 2016, **12**, 878.
66 O. A. Von Lilienfeld and K. Burke, Retrospective on a decade of machine learning for chemical discovery, *Nat. Commun.*, 2020, **11**, 1–4.
67 J. N. Wei, D. Duvenaud and A. Aspuru-Guzik, Neural networks for the prediction of organic chemistry reactions, *ACS Cent. Sci.*, 2016, **2**, 725–732.
68 M. H. Segler, M. Preuss and M. P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI, *Nature*, 2018, **555**, 604–610.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

69 R. Eldan and O. Shamir, The power of depth for feedforward neural networks, *Conference on learning theory*, 2016, pp. 907–940.
70 N. Cohen, O. Sharir and A. Shashua, On the expressive power of deep learning: A tensor analysis, *Conference on learning theory*, 2016, pp. 698–728.
71 M. Telgarsky, Benefits of depth in neural networks, *Conference on learning theory*, 2016, pp. 1517–1539.
72 Z. Lu, H. Pu, F. Wang, Z. Hu and L. Wang, The Expressive Power of Neural Networks: A View from the Width, *Adv. Neural Inf. Process Syst.*, 2017, 6231–6239.
73 J. Behler, First principles neural network potentials for reactive simulations of large molecular and condensed systems, *Angew. Chem., Int. Ed.*, 2017, **56**, 12828–12840.
74 K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Machine Learning Meets Quantum Physics*, Springer, 2020, pp. 215–230.
75 G. Montavon, *Machine Learning Meets Quantum Physics*, Springer, 2020, pp. 37–62.
76 L. I. Vazquez-Salazar, E. D. Boittier and M. Meuwly, Uncertainty quantification for predictions of atomistic neural networks, *Chem. Sci.*, 2022, **13**, 13068–13084.
77 T. B. Blank, S. D. Brown, A. W. Calhoun and D. J. Doren, Neural network models of potential energy surfaces, *J. Chem. Phys.*, 1995, **103**, 4129–4137.
78 D. F. R. Brown, M. N. Gibbs and D. C. Clary, Combining *ab initio* computations, neural networks, and diffusion Monte Carlo: An efficient method to treat weakly bound molecules, *J. Chem. Phys.*, 1996, **105**, 7597–7604.
79 S. Lorenz, A. Groß and M. Scheffler, Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks, *Chem. Phys. Lett.*, 2004, **395**, 210–215.
80 J. B. Witkoskie and D. J. Doren, Neural Network Models of Potential Energy Surfaces: Prototypical Examples, *J. Chem. Theory Comput.*, 2005, **1**, 14–23.
81 J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, *J. Chem. Phys.*, 2011, **134**, 074106.
82 L. Zhang, J. Han, H. Wang, R. Car and W. E, Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics, *Phys. Rev. Lett.*, 2018, **120**, 143001.
83 Y. Zhang, C. Hu and B. Jiang, Embedded Atom Neural Network Potentials: Efficient and Accurate Machine Learning with a Physically Inspired Representation, *J. Phys. Chem. Lett.*, 2019, **10**, 4962–4967.
84 M. Liu and J. R. Kitchin, SingleNN: Modified Behler–Parrinello Neural Network with Shared Weights for Atomistic Simulations with Transferability, *J. Phys. Chem. C*, 2020, **124**, 17811–17818.
85 K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko and K.-R. Müller, SchNet: A continuous-filter convolutional neural network for modeling quantum interactions, *Adv. Neural Inf. Process. Syst.*, 2017, 992–1002.
86 N. Lubbers, J. S. Smith and K. Barros, Hierarchical modeling of molecular energies using a deep neural network, *J. Chem. Phys.*, 2018, **148**, 241715.
87 R. Zubatyuk, J. S. Smith, J. Leszczynski and O. Isayev, Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network, *Sci. Adv.*, 2019, **5**, eaav6490.
88 T. Zubatiuk and O. Isayev, Development of Multimodal Machine Learning Potentials: Toward a Physics-Aware Artificial Intelligence, *Acc. Chem. Res.*, 2021, **54**, 1575–1585.
89 O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda and K.-R. Müller, SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects, *Nat. Commun.*, 2021, **12**, 1–14.
90 N. Artrith, T. Morawietz and J. Behler, High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 153101.
91 K. Yao, J. E. Herr, D. W. Toth, R. Mckintyre and J. Parkhill, The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics, *Chem. Sci.*, 2018, **9**, 2261–2269.
92 S. A. Ghasemi, A. Hofstetter, S. Saha and S. Goedecker, Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **92**, 045131.
93 X. Xie, K. A. Persson and D. W. Small, 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.*, 2020, **16**, 4256–4270.
94 K. Schütt, O. Unke and M. Gastegger, Equivariant message passing for the prediction of tensorial properties and molecular spectra, *International Conference on Machine Learning*, 2021, pp. 9377–9388.
95 V. G. Satorras, E. Hoogeboom and M. Welling, E(n) equivariant graph neural networks, *International conference on machine learning*, 2021, pp. 9323–9332.
96 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. E. Kozinsky, (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, *Nat. Commun.*, 2022, **13**, 1–11.
97 K. V. J. Jose, N. Artrith and J. Behler, Construction of high-dimensional neural network potentials using environment-dependent atom pairs, *J. Chem. Phys.*, 2012, **136**, 194111.
98 K. Yao, J. E. Herr, S. N. Brown and J. Parkhill, Intrinsic Bond Energies from a Bonds-in-Molecules Neural Network, *J. Phys. Chem. Lett.*, 2017, **8**, 2689–2694.
99 Z. L. Glick, D. P. Metcalf, A. Koutsoukas, S. A. Spronk, D. L. Cheney and C. D. Sherrill, AP-Net: An atomic-pairwise neural network for smooth and transferable interaction potentials, *J. Chem. Phys.*, 2020, **153**, 044112.
100 O. A. von Lilienfeld, R. Ramakrishnan, M. Rupp and A. Knoll, Fourier series of atomic radial distribution functions: A molecular fingerprint for machine learning models of quantum chemical properties, *Int. J. Quantum Chem.*, 2015, **115**, 1084–1093.

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

101 H. Huo and M. Rupp, Unified representation of molecules and crystals for machine learning, *Mach. Learn.: Sci. Technol.*, 2022, **3**, 045017.
102 M. Uhrin, Through the eyes of a descriptor: Constructing complete, invertible descriptions of atomic environments, *Phys. Rev. B*, 2021, **104**, 144110.
103 M. F. Langer, A. Goeßmann and M. Rupp, Representations of molecules and materials for interpolation of quantum-mechanical simulations *via* machine learning, *npj Comput. Mater.*, 2022, **8**, 1–14.
104 J. Behler, Four generations of high-dimensional neural network potentials, *Chem. Rev.*, 2021, **121**, 10037–10072.
105 B. Jiang and H. Guo, Permutation invariant polynomial neural network approach to fitting potential energy surfaces, *J. Chem. Phys.*, 2013, **139**, 054112.
106 J. Li, B. Jiang and H. Guo, Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems, *J. Chem. Phys.*, 2013, **139**, 204103.
107 B. Jiang, J. Li and H. Guo, Potential energy surfaces from high fidelity fitting of *ab initio* points: the permutation invariant polynomial-neural network approach, *Int. Rev. Phys. Chem.*, 2016, **35**, 479–506.
108 K. Shao, J. Chen, Z. Zhao and D. H. Zhang, Communication: Fitting potential energy surfaces with fundamental invariant neural network, *J. Chem. Phys.*, 2016, **145**, 071101.
109 B. Fu and D. H. Zhang, *Ab initio* potential energy surfaces and quantum dynamics for polyatomic bimolecular reactions, *J. Chem. Theory Comput.*, 2018, **14**, 2289–2303.
110 R. Chen, K. Shao, B. Fu and D. H. Zhang, Fitting potential energy surfaces with fundamental invariant neural network. II. Generating fundamental invariants for molecular systems with up to ten atoms, *J. Chem. Phys.*, 2020, **152**, 204307.
111 J. Behler, Constructing high-dimensional neural network potentials: A tutorial review, *Int. J. Quantum Chem.*, 2015, **115**, 1032–1050.
112 M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi and P. Marquetand, wACSF—Weighted atom-centered symmetry functions as descriptors in machine learning potentials, *J. Chem. Phys.*, 2018, **148**, 241709.
113 A. Singraber, J. Behler and C. Dellago, Library-Based LAMMPS Implementation of High-Dimensional Neural Network Potentials, *J. Chem. Theory Comput.*, 2019, **15**, 1827–1840.
114 M. P. Bircher, A. Singraber and C. Dellago, Improved description of atomic environments using low-cost polynomial functions with compact support, *J. Mach. Learn. Res.*, 2021, **2**, 035026.
115 H. Wang, L. Zhang, J. Han and E. Weinan, DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics, *Comput. Phys. Commun.*, 2018, **228**, 178–184.
116 T. Wen, L. Zhang, H. Wang, W. E and D. J. Srolovitz, Deep potentials for materials science, *Mater. Future*, 2022, **1**, 022601.
117 L. Zhang, J. Han, H. Wang, W. Saidi, R. Car and W. E, End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems, *Adv. Neural Inf. Process. Syst.*, 2018, 4441–4451.
118 A. Khorshidi and A. A. Peterson, Amp: A modular approach to machine learning in atomistic simulations, *Comput. Phys. Commun.*, 2016, **207**, 310–324.
119 O. T. Unke and M. Meuwly, A reactive, scalable, and transferable model for molecular energies from a neural network approach based on local information, *J. Chem. Phys.*, 2018, **148**, 241708.
120 S. N. Pozdnyakov, M. J. Willatt, A. P. Bartók, C. Ortner, G. Csányi and M. Ceriotti, Incompleteness of Atomic Structure Representations, *Phys. Rev. Lett.*, 2020, **125**, 166001.
121 B. Parsaeifard and S. Goedecker, Manifolds of quasi-constant SOAP and ACSF fingerprints and the resulting failure to machine learn four-body interactions, *J. Chem. Phys.*, 2022, **156**, 034302.
122 F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner and G. Monfardini, The Graph Neural Network Model, *IEEE Trans. Neural Netw. Learn. Syst.*, 2009, **20**, 61–80.
123 D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarelli, T. Hirzel, A. Aspuru-Guzik and R. P. Adams, Convolutional Networks on Graphs for Learning Molecular Fingerprints, *Adv. Neural Inf. Process. Syst.*, 2015, 2224–2232.
124 J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, Neural Message Passing for Quantum Chemistry, *Proceedings of the 34th International Conference on Machine Learning*, 2017, pp. 1263–1272.
125 S. Manzhos and T. Carrington, Neural Network Potential Energy Surfaces for Small Molecules and Reactions, *Chem. Rev.*, 2021, **121**, 10187–10217.
126 M. Malshe, R. Narulkar, L. Raff, M. Hagan, S. Bukkapatnam, P. Agrawal and R. Komanduri, Development of generalized potential-energy surfaces using many-body expansions, neural networks, and moiety energy approximations, *J. Chem. Phys.*, 2009, **130**, 184102.
127 S. Manzhos and T. Carrington Jr, A random-sampling high dimensional model representation neural network for building potential energy surfaces, *J. Chem. Phys.*, 2006, **125**, 084109.
128 M. Hellström and J. Behler, *Machine Learning Meets Quantum Physics*, Springer, 2020, pp. 253–275.
129 S. Grimme, S. Ehrlich and L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, *J. Chem. Theory Comput.*, 2011, **32**, 1456–1465.
130 J. Westermayr and P. Marquetand, Machine learning for electronically excited states of molecules, *Chem. Rev.*, 2020, **121**, 9873–9926.
131 J. Westermayr, M. Gastegger and P. Marquetand, Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics, *J. Phys. Chem. Lett.*, 2020, **11**, 3828–3834.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

132 S. Mai, P. Marquetand and L. Gonzalez, Nonadiabatic Dynamics: The SHARC Approach, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1370.
133 J. Gasteiger, J. Groß and S. Günnemann, Directional Message Passing for Molecular Graphs, *International Conference on Learning Representations*, 2020.
134 B. K. Miller, M. Geiger, T. E. Smidt and F. Noé, Relevance of Rotationally Equivariant Convolutions for Predicting Molecular Properties, *arXiv*, 2020, e-prints arXiv:2008.08461.
135 N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff and P. Riley, Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, *arXiv*, 2018, e-prints, arXiv:1802.08219.
136 T. E. Smidt, Euclidean symmetry and equivariance in machine learning, *Trends Chem.*, 2021, **3**, 82–85.
137 M. Haghighatlari, J. Li, X. Guan, O. Zhang, A. Das, C. J. Stein, F. Heidar-Zadeh, M. Liu, M. Head-Gordon, L. Bertels, *et al.*, Newtonnet: A newtonian message passing network for deep learning of interatomic potentials and forces, *Digit. Discov.*, 2022, **1**, 333–343.
138 V. Zaverkin, D. Holzmüller, I. Steinwart and J. Kästner, Fast and Sample-Efficient Interatomic Neural Network Potentials for Molecules and Materials Based on Gaussian Moments, *J. Chem. Theory Comput.*, 2021, **17**, 6658–6670.
139 X. Zheng, P. Zheng and R.-Z. Zhang, Machine learning material properties from the periodic table using convolutional neural networks, *Chem. Sci.*, 2018, **9**, 8426–8432.
140 T. Xie and J. C. Grossman, Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, *Phys. Rev. Lett.*, 2018, **120**, 145301.
141 R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi and C. Kim, Machine learning in materials informatics: recent applications and prospects, *npj Comput. Mater.*, 2017, **3**, 1–13.
142 A. Y.-T. Wang, R. J. Murdock, S. K. Kauwe, A. O. Oliynyk, A. Gurlo, J. Brgoch, K. A. Persson and T. D. Sparks, Machine learning for materials scientists: an introductory guide toward best practices, *Chem. Mater.*, 2020, **32**, 4954–4965.
143 C. Sutton, M. Boley, L. M. Ghiringhelli, M. Rupp, J. Vreeken and M. Scheffler, Identifying domains of applicability of machine learning models for materials science, *Nat. Commun.*, 2020, **11**, 1–9.
144 M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, Software available from, **https://www.tensorflow.org/**.
145 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, *Adv. Neural Inf. Process. Syst.*, 2019, **32**, 8024–8035.
146 M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller and K. Burke, Quantum chemical accuracy from density functional approximations *via* machine learning, *Nat. Commun.*, 2020, **11**, 1–11.
147 M. O'Connor, H. M. Deeks, E. Dawn, O. Metatla, A. Roudaut, M. Sutton, L. M. Thomas, B. R. Glowacki, R. Sage, P. Tew, *et al.*, Sampling molecular conformations and dynamics in a multiuser virtual reality framework, *Sci. Adv.*, 2018, **4**, eaat2731.
148 S. Amabilino, L. A. Bratholm, S. J. Bennie, A. C. Vaucher, M. Reiher and D. R. Glowacki, Training neural nets to learn reactive potential energy surfaces using interactive quantum chemistry in virtual reality, *J. Phys. Chem. A*, 2019, **123**, 4486–4499.
149 S. Amabilino, L. A. Bratholm, S. J. Bennie, M. B. O'Connor and D. R. Glowacki, Training atomic neural networks using fragment-based data generated in virtual reality, *J. Chem. Phys.*, 2020, **153**, 154105.
150 Q. Chu, K. H. Luo and D. Chen, Exploring complex reaction networks using neural network-based molecular dynamics simulation, *J. Phys. Chem. Lett.*, 2022, **13**, 4052–4057.
151 J. V. Diez, S. R. Atance, O. Engkvist, R. Mercado and S. Olsson, A transferable Boltzmann generator for small-molecules conformers, *ELLIS Machine Learning for Molecule Discovery Workshop (ML4Molecules)*, 2021.
152 M. Gastegger and P. Marquetand, *Machine learning meets quantum physics*, Springer, 2020, pp. 233–252.
153 G. M. Torrie and J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *J. Comput. Phys.*, 1977, **23**, 187–199.
154 A. Barducci, M. Bonomi and M. Parrinello, Metadynamics, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 826–843.
155 I. Kosztin, B. Faber and K. Schulten, Introduction to the diffusion Monte Carlo method, *Am. J. Physiol.*, 1996, **64**, 633–644.
156 J. Li, C. Qu and J. M. Bowman, Diffusion Monte Carlo with fictitious masses finds holes in potential energy surfaces, *Mol. Phys.*, 2021, **119**, e1976426.
157 R. Conte, P. L. Houston, C. Qu, J. Li and J. M. Bowman, Full-dimensional, *ab initio* potential energy surface for glycine with characterization of stationary points and zero-point energy calculations by means of diffusion Monte Carlo and semiclassical dynamics, *J. Chem. Phys.*, 2020, **153**, 244301.
158 A. Nandi, C. Qu and J. M. Bowman, Using gradients in permutationally invariant polynomial potential fitting: A demonstration for CH4 using as few as 100 configurations, *J. Chem. Theory Comput.*, 2019, **15**, 2826–2835.
159 A. Shapeev, K. Gubaev, E. Tsymbalov and E. Podryabinkin, *Machine Learning Meets Quantum Physics*; Springer, 2020, pp. 309–329.
160 G. Csányi, T. Albaret, M. Payne and A. De Vita, Learn on the fly: A hybrid classical and quantum-mechanical molecular dynamics simulation, *Phys. Rev. Lett.*, 2004, **93**, 175503.
161 M. Gastegger, J. Behler and P. Marquetand, Machine learning molecular dynamics for the simulation of infrared spectra, *Chem. Sci.*, 2017, **8**, 6924–6935.

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

162 L. Kahle and F. Zipoli, Quality of uncertainty estimates from neural network potential ensembles, *Phys. Rev. E*, 2022, **105**, 015311.
163 A. Amini, W. Schwarting, A. Soleimany and D. Rus, Deep Evidential Regression, *Adv. Neural Inf. Process. Syst.*, 2020, 14927–14937.
164 P. Pernot, Prediction uncertainty validation for computational chemists, *J. Chem. Phys.*, 2022, **157**, 144103.
165 P. Zheng, W. Yang, W. Wu, O. Isayev and P. O. Dral, Toward Chemical Accuracy in Predicting Enthalpies of Formation with General-Purpose Data-Driven Methods, *J. Phys. Chem. Lett.*, 2022, **13**, 3479–3491.
166 C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang and C. Liu, A survey on deep transfer learning, *International conference on artificial neural networks*, 2018, pp. 270–279.
167 C. Cai, S. Wang, Y. Xu, W. Zhang, K. Tang, Q. Ouyang, L. Lai and J. Pei, Transfer learning for drug discovery, *J. Med. Chem.*, 2020, **63**, 8683–8694.
168 S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, 2009, **22**, 1345–1359.
169 B. Fu, X. Xu and D. H. Zhang, A hierarchical construction scheme for accurate potential energy surface generation: An application to the F+ H_2 reaction, *J. Chem. Phys.*, 2008, **129**, 011103.
170 R. Ramakrishnan, P. Dral, M. Rupp and O. A. von Lilienfeld, Big Data meets quantum chemistry approximations: The Δ-machine learning approach, *J. Chem. Theory Comput.*, 2015, **11**, 2087–2096.
171 K. A. Nguyen, I. Rossi and D. G. Truhlar, A dual-level Shepard interpolation method for generating potential energy surfaces for dynamics calculations, *J. Chem. Phys.*, 1995, **103**, 5522–5530.
172 R. Batra, G. Pilania, B. P. Uberuaga and R. Ramprasad, Multifidelity information fusion with machine learning: A case study of dopant formation energies in hafnia, *ACS Appl. Mater. Interfaces*, 2019, **11**, 24906–24918.
173 P. Zaspel, B. Huang, H. Harbrecht and O. A. von Lilienfeld, Boosting quantum machine learning models with a multilevel combination technique: Pople diagrams revisited, *J. Chem. Theory Comput.*, 2018, **15**, 1546–1559.
174 G. Hinton, O. Vinyals and J. Dean, Distilling the Knowledge in a Neural Network, *arXiv*, 2015, e-prints, arXiv:1503.02531.
175 E. Strubell, A. Ganesh and A. McCallum, Energy and Policy Considerations for Deep Learning in NLP, *arXiv*, 2019, e-prints, arXiv:1906.02243.
176 R. A. Friesner, *Ab initio* quantum chemistry: Methodology and applications, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6648–6653.
177 P. Mo, M. Shi, W. Yao and J. Liu, Transfer Learning of Potential Energy Surfaces for Efficient Atomistic Modeling of Doping and Alloy, *IEEE Electron Device Lett.*, 2020, **41**, 633–636.
178 S. Käser, O. T. Unke and M. Meuwly, Reactive dynamics and spectroscopy of hydrogen transfer from neural network-based reactive potential energy surfaces, *New J. Phys.*, 2020, **22**, 055002.
179 S. Käser, E. D. Boittier, M. Upadhyay and M. Meuwly, Transfer learning to CCSD (T): Accurate anharmonic frequencies from machine learning models, *J. Chem. Theory Comput.*, 2021, **17**, 3687–3699.
180 S. Käser and M. Meuwly, Transfer learned potential energy surfaces: accurate anharmonic vibrational dynamics and dissociation energies for the formic acid monomer and dimer, *Phys. Chem. Chem. Phys.*, 2022, **24**, 5269–5281.
181 A. Nandi, C. Qu, P. L. Houston, R. Conte and J. M. Bowman, Δ-machine learning for potential energy surfaces: A PIP approach to bring a DFT-based PES to CCSD(T) level of theory, *J. Chem. Phys.*, 2021, **154**, 051102.
182 C. Qu, P. L. Houston, R. Conte, A. Nandi and J. M. Bowman, Breaking the coupled cluster barrier for machine-learned potentials of large molecules: The case of 15-atom acetylacetone, *J. Phys. Chem. Lett.*, 2021, **12**, 4902–4909.
183 C. Qu, Q. Yu, R. Conte, P. L. Houston, A. Nandi and J. M. Bowman, A Δ-Machine Learning Approach for Force Fields, Illustrated by a CCSD(T) 4-body Correction to the MB-pol Water Potential, *arXiv*, 2022, e-prints, arXiv:2206.04254.
184 F. Imrie, A. R. Bradley, M. van der Schaar and C. M. Deane, Protein family-specific models using deep neural networks and transfer learning improve virtual screening and highlight the need for more data, *J. Chem. Inf. Model.*, 2018, **58**, 2319–2330.
185 S. Mishra, T. Yamasaki and H. Imaizumi, Improving image classifiers for small datasets by learning rate adaptations, *2019 16th International Conference on Machine Vision Applications (MVA)*, 2019, pp. 1–6.
186 Y. Liu and J. Li, Permutation-Invariant-Polynomial Neural-Network-Based Δ-Machine Learning Approach: A Case for the HO_2 Self-Reaction and Its Dynamics Study, *J. Phys. Chem. Lett.*, 2022, **13**, 4729–4738.
187 J. Zhu, V. Q. Vuong, B. G. Sumpter and S. Irle, Artificial neural network correction for density-functional tight-binding molecular dynamics simulations, *MRS Commun.*, 2019, **9**, 867–873.
188 P. O. Dral, A. Owens, A. Dral and G. Csányi, Hierarchical machine learning of potential energy surfaces, *J. Chem. Phys.*, 2020, **152**, 204110.
189 M. Ruth, D. Gerbig and P. R. Schreiner, Machine Learning of Coupled Cluster (T)-Energy Corrections *via* Delta (Δ)-Learning, *J. Chem. Theory Comput.*, 2022, **18**, 4846–4855.
190 M. Majumder, S. E. Hegger, R. Dawes, S. Manzhos, X.-G. Wang, C. Tucker Jr, J. Li and H. Guo, Explicitly correlated MRCI-F12 potential energy surfaces for methane fit with several permutation invariant schemes and full-dimensional vibrational calculations, *Mol. Phys.*, 2015, **113**, 1823–1833.
191 A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. Carrington and S. Manzhos, Neural networks *vs.* Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy, *J. Chem. Phys.*, 2018, **148**, 241702.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

192 K. Töpfer, S. Käser and M. Meuwly, Double Proton Transfer in Hydrated Formic Acid Dimer: Interplay of Spatial Symmetry and Solvent-Generated Force on Reactivity, *Phys. Chem. Chem. Phys.*, 2022, **24**, 13869–13882.

193 S. Käser, J. O. Richardson and M. Meuwly, Transfer learning for affordable and high quality tunneling splittings from instanton calculations, *J. Chem. Theory Comput.*, 2022, **18**, 6840–6850.

194 V. Zaverkin, D. Holzmüller, R. Schuldt and J. Kästner, Predicting properties of periodic systems from cluster data: A case study of liquid water, *J. Chem. Phys.*, 2022, **156**, 114103.

195 C. Schran, F. Brieuc and D. Marx, Transferability of machine learning potentials: Protonated water neural network potential applied to the protonated water hexamer, *J. Chem. Phys.*, 2021, **154**, 051101.

196 V. Zaverkin and J. Kästner, Gaussian Moments as Physically Inspired Molecular Descriptors for Accurate and Scalable Machine Learning Potentials, *J. Chem. Theory Comput.*, 2020, **16**, 5410–5421.

197 C. Schran, J. Behler and D. Marx, Automated Fitting of Neural Network Potentials at Coupled Cluster Accuracy: Protonated Water Clusters as Testing Ground, *J. Chem. Theory Comput.*, 2020, **16**, 88–99.

198 G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg, C. T. Bowman, R. K. Hanson, S. Song, W. C. Gardiner, *et al.*, *GRI_Mech 30*, 1999, **http://www.me.berkeley.edu/gri_mech/**.

199 J. Zeng, L. Cao, M. Xu, T. Zhu and J. Z. Zhang, Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation, *Nat. Commun.*, 2020, **11**, 1–9.

200 Z. He, X.-B. Li, L.-M. Liu and W. Zhu, The intrinsic mechanism of methane oxidation under explosion condition: A combined ReaxFF and DFT study, *Fuel*, 2014, **124**, 85–90.

201 J. Westermayr, M. Gastegger, D. Vörös, L. Panzenboeck, F. Joerg, L. González and P. Marquetand, Deep learning study of tyrosine reveals that roaming can lead to photodamage, *Nat. Chem.*, 2022, **14**, 914–919.

202 D. Townsend, S. A. Lahankar, S. K. Lee, S. D. Chambreau, A. G. Suits, X. Zhang, J. Rheinecker, L. B. Harding and J. M. Bowman, The Roaming Atom: Straying from the Reaction Path in Formaldehyde Decomposition, *Science*, 2004, **306**, 1158–1161.

203 T. Endo, S. P. Neville, V. Wanie, S. Beaulieu, C. Qu, J. Deschamps, P. Lassonde, B. E. Schmidt, H. Fujise, M. Fushitani, *et al.*, Capturing roaming molecular fragments in real time, *Science*, 2020, **370**, 1072–1077.

204 H. Hu and W. Yang, Free Energies of Chemical Reactions in Solution and in Enzymes with *Ab Initio* Quantum Mechanics/Molecular Mechanics Methods, *Annu. Rev. Phys. Chem.*, 2008, **59**, 573–601.

205 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, *et al.*, Charmm General Force Field: A Force Field for Drug-Like Molecules Compatible with the Charmm All-Atom Additive Biological Force Fields, *J. Comput. Chem.*, 2010, **31**, 671–690.

206 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys.*, 1983, **79**, 926–935.

207 J. Ho, Y. Shao and J. Kato, Do better quality embedding potentials accelerate the convergence of QM/MM models? The case of solvated acid clusters, *Molecules*, 2018, **23**, 2466.

208 L. Böselt, M. Thürlemann and S. Riniker, Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems, *J. Chem. Theory Comput.*, 2021, **17**, 2641–2658.

209 M. Gastegger, K. T. Schütt and K.-R. Müller, Machine learning of solvent effects on molecular spectra and reactions, *Chem. Sci.*, 2021, **12**, 11473–11483.

210 B. Mennucci, Polarizable continuum model, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 386–404.

211 M. Bondanza, M. Nottoli, L. Cupellini, F. Lipparini and B. Mennucci, Polarizable embedding QM/MM: the future gold standard for complex (bio)systems?, *Phys. Chem. Chem. Phys.*, 2020, **22**, 14433–14448.

212 B. Lier, P. Poliak, P. Marquetand, J. Westermayr and C. Oostenbrink, BuRNN: Buffer Region Neural Network Approach for Polarizable-Embedding Neural Network/Molecular Mechanics Simulations, *J. Phys. Chem. Lett.*, 2022, **13**, 3812–3818.

213 G. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, Clarendon Press, 1994.

214 D. Koner, R. J. Bemish and M. Meuwly, Dynamics on multiple potential energy surfaces: Quantitative studies of elementary processes relevant to hypersonics, *J. Phys. Chem. A*, 2020, **124**, 6255–6269.

215 D. Koner, O. T. Unke, K. Boe, R. J. Bemish and M. Meuwly, Exhaustive state-to-state cross sections for reactive molecular collisions from importance sampling simulation and a neural network representation, *J. Chem. Phys.*, 2019, **150**, 211101.

216 J. Arnold, J. C. San Vicente Veliz, D. Koner, N. Singh, R. J. Bemish and M. Meuwly, Machine learning product state distributions from initial reactant states for a reactive atom–diatom collision system, *J. Chem. Phys.*, 2022, **156**, 034301.

217 J. Arnold, D. Koner, S. Käser, N. Singh, R. J. Bemish and M. Meuwly, Machine Learning for Observables: Reactant to Product State Distributions for Atom-Diatom Collisions, *J. Phys. Chem. A*, 2020, **124**, 7177–7190.

218 F. Kaufman and L. J. Decker, *7th Symp. (Int.) Combustion*, 1959, vol. 57.

219 J. B. Livesey, A. L. Roberts and A. Williams, The Formation of Oxides of Nitrogen in some Oxy-Propane Flames, *Combust. Sci. Technol.*, 1971, **4**, 9–15.

220 W. Wilson, Rate constant for the reaction $N + O_2 \rightarrow NO + O$, *J. Chem. Phys.*, 1967, **46**, 2017–2018.

221 D. Koner, R. J. Bemish and M. Meuwly, The $C(^3P) + NO(X^2\Pi) \rightarrow O(^3P) + CN(X^2\Sigma^+)$, $N(^2D)/N(^4S) + CO(X^1\Sigma^+)$ reaction :

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Rates, branching ratios, and final states from 15 K to 20 000 K, *J. Chem. Phys.*, 2018, **149**, 094305.

222 J. C. San Vicente Veliz, D. Koner, M. Schwilk, R. J. Bemish and M. Meuwly, The $N(^4S) + O_2(X\text{-}^3\Sigma_g) \leftrightarrow O(^3P) + NO(X\ ^2\Pi)$ reaction: thermal and vibrational relaxation rates for the $^2A'$, $^4A'$ and $^2A''$ states, *Phys. Chem. Chem. Phys.*, 2020, **22**, 3927–3939.

223 J. C. S. V. Veliz, D. Koner, M. Schwilk, R. J. Bemish and M. Meuwly, The $C\ (^3P) + O_2\ (3\sum_g^-) \rightarrow CO_2 \leftrightarrow CO\ (^1\Sigma^+) + O\ (^1D)/O\ (^3P)$ reaction: thermal and vibrational relaxation rates from 15 K to 20000 K, *Phys. Chem. Chem. Phys.*, 2021, **23**, 11251–11263.

224 P. L. Houston, A. Nandi and J. M. Bowman, A machine learning approach for prediction of rate constants, *J. Phys. Chem. Lett.*, 2019, **10**, 5250–5258.

225 P. L. Houston, A. Nandi and J. M. Bowman, A Machine Learning Approach for Rate Constants. III. Application to the $Cl(^2P) + CH_4 \rightarrow CH_3 + HCl$ Reaction, *J. Phys. Chem. A*, 2022, **126**, 5672–5679.

226 I. R. Craig and D. E. Manolopoulos, A refined ring polymer molecular dynamics theory of chemical reaction rates, *J. Chem. Phys.*, 2005, **123**, 034102.

227 K. Kwac, H. Freedman and M. Cho, Machine Learning Approach for Describing Water OH Stretch Vibrations, *J. Chem. Theory Comput.*, 2021, **17**, 6353–6365.

228 S. Koutsoukos, F. Philippi, F. Malaret and T. Welton, A review on machine learning algorithms for the ionic liquid chemical space, *Chem. Sci.*, 2021, **12**, 6820–6843.

229 K. Töpfer, A. Pasti, A. Das, S. M. Salehi, L. I. Vazquez-Salazar, D. Rohrbach, T. Feurer, P. Hamm and M. Meuwly, Structure, Organization, and Heterogeneity of Water-Containing Deep Eutectic Solvents, *J. Am. Chem. Soc.*, 2022, **144**, 14170–14180.

230 G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller and O. A. Von Lilienfeld, Machine learning of molecular electronic properties in chemical compound space, *New J. Phys.*, 2013, **15**, 095003.

231 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, *Sci. Data*, 2014, **1**, 1–7.

232 J. S. Smith, R. Zubatyuk, B. Nebgen, N. Lubbers, K. Barros, A. E. Roitberg, O. Isayev and S. Tretiak, The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules, *Sci. Data*, 2020, **7**, 1–10.

233 J. S. Smith, O. Isayev and A. E. Roitberg, ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules, *Sci. Data*, 2017, **4**, 1–8.

234 J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio Jr and A. Tkatchenko, QM7-X, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules, *Sci. Data*, 2021, **8**, 1–11.

235 S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt and K.-R. Müller, Machine learning of accurate energy-conserving molecular force fields, *Sci. Adv.*, 2017, **3**, e1603015.

236 S. Chmiela, H. E. Sauceda, K.-R. Müller and A. Tkatchenko, Towards exact molecular dynamics simulations with machine-learned force fields, *Nat. Commun.*, 2018, **9**, 1–10.

237 C. Qu, P. L. Houston, R. Conte, A. Nandi and J. M. Bowman, MULTIMODE calculations of vibrational spectroscopy and 1d interconformer tunneling dynamics in Glycine using a full-dimensional potential energy surface, *J. Phys. Chem. A*, 2021, **125**, 5346–5354.

238 J. M. Bowman, C. Qu, R. Conte, A. Nandi, P. L. Houston and Q. Yu, The MD17 datasets from the perspective of datasets for gas-phase "small" molecule potentials, *J. Chem. Phys.*, 2022, **156**, 240901.

239 L. Zhang, S. Zhang, A. Owens, S. N. Yurchenko and P. O. Dral, VIB5 database with accurate *ab initio* quantum chemical molecular potential energy surfaces, *Sci. Data*, 2022, **9**, 1–10.

240 M. Ceriotti, Unsupervised machine learning in atomistic simulations, between predictions and understanding, *J. Chem. Phys.*, 2019, **150**, 150901.

241 A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé and A. Laio, Unsupervised learning methods for molecular simulation data, *Chem. Rev.*, 2021, **121**, 9722–9758.

242 M. Abraham, R. Apostolov, J. Barnoud, P. Bauer, C. Blau, A. M. Bonvin, M. Chavent, J. Chodera, K. Condic-Jurkic, L. Delemotte, *et al.*, Sharing data from molecular simulations, *J. Chem. Inf. Model.*, 2019, **59**, 4093–4099.

243 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, *et al.*, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 2016, **3**, 1–9.

244 N. Artrith, K. T. Butler, F.-X. Coudert, S. Han, O. Isayev, A. Jain and A. Walsh, Best practices in machine learning for chemistry, *Nat. Chem.*, 2021, **13**, 505–508.

245 J. Li and Y. Liu, Data quality, data sampling and data fitting: A tutorial guide for constructing full-dimensional accurate potential energy surfaces (PESs) of small molecular systems, *ChemRxiv*, 2022, preprint, ChemRxiv:10.26434/chemrxiv-2022-1jpvc.

246 M. Du, N. Liu and X. Hu, Techniques for interpretable machine learning, *Commun. ACM*, 2019, **63**, 68–77.

247 W. Samek and K.-R. Müller, *Explainable AI: interpreting, explaining and visualizing deep learning*; Springer, 2019, pp. 5–22.

248 W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl and B. Yu, Definitions, methods, and applications in interpretable machine learning, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 22071–22080.

249 R. Dybowski, Interpretable machine learning as a tool for scientific discovery in chemistry, *New J. Chem.*, 2020, **44**, 20914–20920.

250 S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek and K.-R. Müller, Unmasking Clever Hans predictors and assessing what machines really learn, *Nat. Commun.*, 2019, **10**, 1–8.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

251 T. Schnake, O. Eberle, J. Lederer, S. Nakajima, K. T. Schutt, K.-R. Mueller and G. Montavon, Higher-Order Explanations of Graph Neural Networks *via* Relevant Walks, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021, 1.

252 K. T. Schütt, M. Gastegger, A. Tkatchenko and K.-R. Müller, *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, Springer, 2019, pp. 311–330.

253 S. Letzgus, P. Wagner, J. Lederer, W. Samek, K.-R. Müller and G. Montavon, Toward Explainable Artificial Intelligence for Regression Models: A methodological perspective, *IEEE Signal Process. Mag.*, 2022, **39**, 40–58.

254 C. Babbage, *Passages from the Life of a Philosopher*; Cambridge Library Collection - Technology; Cambridge University Press, 2011.

255 W. G. Noid, in *Biomolecular Simulations: Methods and Protocols*, ed. L. Monticelli and E. Salonen, Humana Press, Totowa, NJ, 2013, pp. 487–531.

256 D. Reith, M. Pütz and F. Müller-Plathe, Deriving effective mesoscale potentials from atomistic simulations, *J. Comput. Chem.*, 2003, **24**, 1624–1636.

257 A. P. Lyubartsev and A. Laaksonen, Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1995, **52**, 3730.

258 M.-C. Chang, C.-H. Tung, S.-Y. Chang, J. M. Carrillo, Y. Wang, B. G. Sumpter, G.-R. Huang, C. Do and W.-R. Chen, A machine learning inversion scheme for determining interaction from scattering, *Commun. Phys.*, 2022, **5**, 1–8.

259 K. Low, M. L. Coote and E. I. Izgorodina, Inclusion of More Physics Leads to Less Data: Learning the Interaction Energy as a Function of Electron Deformation Density with Limited Training Data, *J. Chem. Theory Comput.*, 2022, **18**, 1607–1618.

260 J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, *et al.*, A survey of uncertainty in deep neural networks, *arXiv*, 2021, preprint, arXiv:2107.03342.

261 R. Krems, Bayesian machine learning for quantum molecular dynamics, *Phys. Chem. Chem. Phys.*, 2019, **21**, 13392–13410.

262 J. Cui and R. V. Krems, Efficient non-parametric fitting of potential energy surfaces for polyatomic molecules with Gaussian processes, *J. Phys. B: At., Mol. Opt. Phys.*, 2016, **49**, 224001.

263 D. Vieira and R. V. Krems, Rate Constants for Fine-structure Excitations in O–H Collisions with Error Bars Obtained by Machine Learning, *Astrophys. J.*, 2017, **835**, 255.

264 A. Malinin; S. Chervontsev; I. Provilkov and M. Gales, Regression prior networks, *arXiv*, 2020, preprint arXiv:2006.11590.

265 B. Settles, Active learning, *Synthesis lectures on artificial intelligence and machine learning 1*; Morgan & Claypool Publishers, 2012, vol. 6, pp. 1–114.

266 P. J. Haley and D. Soloway, Extrapolation limitations of multilayer feedforward neural networks, *Proceedings 1992 IJCNN international joint conference on neural networks*, 1992, pp. 25–30.

267 J. Behler, Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17930–17955.

268 M. Meuwly and J. M. Hutson, The potential energy surface and near-dissociation states of He-H_2^+, *J. Chem. Phys.*, 1999, **110**, 3418–3427.

269 P. Soldán and J. M. Hutson, On the long-range and short-range behavior of potentials from reproducing kernel Hilbert space interpolation, *J. Chem. Phys.*, 2000, **112**, 4415–4416.

270 D. Koner, J. C. S. V. Veliz, A. van der Avoird and M. Meuwly, Near dissociation states for H_2^+–He on MRCI and FCI potential energy surfaces, *Phys. Chem. Chem. Phys.*, 2019, **21**, 24976–24983.

271 C. Qu, Q. Yu and J. M. Bowman, Permutationally invariant potential energy surfaces, *Annu. Rev. Phys. Chem.*, 2018, **69**, 151–175.

272 Z. Xie, B. J. Braams and J. M. Bowman, *Ab initio* global potential-energy surface for $H_5^+ \rightarrow H_3^+ + H_2$, *J. Chem. Phys.*, 2005, **122**, 224307.

273 L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother and U. Köthe, Analyzing inverse problems with invertible neural networks, *arXiv*, 2018, preprint arXiv:1808.04730.

274 S. Arridge, P. Maass, O. Öktem and C.-B. Schönlieb, Solving inverse problems using data-driven models, *Acta Numer.*, 2019, **28**, 1–174.

275 K. Kothari, A. Khorashadizadeh, M. de Hoop and I. Dokmanić, Trumpets: Injective flows for inference and inverse problems, *Uncertainty in Artificial Intelligence*, 2021, pp. 1269–1278.

276 R. Rydberg, Graphische darstellung einiger bandenspektroskopischer ergebnisse, *Z. Angew. Math. Phys.*, 1932, **73**, 376–385.

277 O. Klein, Zur berechnung von potentialkurven für zweiatomige moleküle mit hilfe von spektraltermen, *Z. Angew. Math. Phys.*, 1932, **76**, 226–235.

278 A. Rees, The calculation of potential-energy curves from band-spectroscopic data, *Proc. Phys. Soc.*, 1947, **59**, 998.

279 D. J. Nesbitt and M. S. Child, Rotational-RKR inversion of intermolecular stretching potentials: Extension to linear hydrogen bonded complexes, *J. Chem. Phys.*, 1993, **98**, 478–486.

280 M. Raissi, P. Perdikaris and G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.*, 2019, **378**, 686–707.

281 L. G. Wright, T. Onodera, M. M. Stein, T. Wang, D. T. Schachter, Z. Hu and P. L. McMahon, Deep physical neural networks trained with backpropagation, *Nature*, 2022, **601**, 549–555.

282 Y. Tao, X. Zeng, Y. Fan, J. Liu, Z. Li and J. Yang, Exploring Accurate Potential Energy Surfaces *via* Integrating

Variational Quantum Eigensolver with Machine Learning, *J. Phys. Chem. Lett.*, 2022, **13**, 6420–6426, PMID: 35816117.
283 S. Portegies Zwart, The ecological impact of high-performance computing in astrophysics, *Nat. Astron.*, 2020, **4**, 819–822.
284 L. Lannelongue, J. Grealey and M. Inouye, Green algorithms: quantifying the carbon footprint of computation, *Adv. Sci.*, 2021, **8**, 2100707.
285 J. Grealey, L. Lannelongue, W.-Y. Saw, J. Marten, G. Méric, S. Ruiz-Carmona and M. Inouye, The carbon footprint of bioinformatics, *Mol. Biol. Evol.*, 2022, **39**, msac034.

Open Access Article. Published on 21 Dezember 2022. Downloaded on 01.05.26 18:20:33.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

© 2023 The Author(s). Published by the Royal Society of Chemistry