

PERSPECTIVE | NOVEMBER 01 2016

## Perspective: Machine learning potentials for atomistic simulations

Jörg Behler



*J. Chem. Phys.* 145, 170901 (2016)  
<https://doi.org/10.1063/1.4966192>



### Articles You May Be Interested In

Thermal conductivity prediction by atomistic simulation methods: Recent advances and detailed comparison

*J. Appl. Phys.* (December 2021)

Unsupervised machine learning in atomistic simulations, between predictions and understanding

*J. Chem. Phys.* (April 2019)

Atomistic defects as single-photon emitters in atomically thin MoS<sub>2</sub>

*Appl. Phys. Lett.* (August 2020)

Nanotechnology & Materials Science      Optics & Photonics      Impedance Analysis      Scanning Probe Microscopy      Sensors      Failure Analysis & Semiconductors

Unlock the Full Spectrum.  
From DC to 8.5 GHz.  
Your Application. Measured.

Find out more

Zurich Instruments

# Perspective: Machine learning potentials for atomistic simulations

Jörg Behler

*Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany*

(Received 18 August 2016; accepted 11 October 2016; published online 1 November 2016)

Nowadays, computer simulations have become a standard tool in essentially all fields of chemistry, condensed matter physics, and materials science. In order to keep up with state-of-the-art experiments and the ever growing complexity of the investigated problems, there is a constantly increasing need for simulations of more realistic, i.e., larger, model systems with improved accuracy. In many cases, the availability of sufficiently efficient interatomic potentials providing reliable energies and forces has become a serious bottleneck for performing these simulations. To address this problem, currently a paradigm change is taking place in the development of interatomic potentials. Since the early days of computer simulations simplified potentials have been derived using physical approximations whenever the direct application of electronic structure methods has been too demanding. Recent advances in machine learning (ML) now offer an alternative approach for the representation of potential-energy surfaces by fitting large data sets from electronic structure calculations. In this perspective, the central ideas underlying these ML potentials, solved problems and remaining challenges are reviewed along with a discussion of their current applicability and limitations. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4966192>]

## I. INTRODUCTION

Machine learning (ML) methods<sup>1</sup> have a long tradition in chemistry and physics.<sup>2,3</sup> While for many years they have been used almost exclusively as a classification tool to assist, e.g., the analysis of spectra,<sup>4</sup> the prediction of binding sites of biomolecules<sup>5</sup> or the derivation of quantitative structure-activity relationships (QSARs),<sup>6</sup> they have now started to enter the heart of theoretical chemistry and computational materials science in a manifold way. This process was initially slow, because of the understandable scepticism of theoretical chemists and physicists, who are used to well-controlled hierarchies of physical approximations, regarding these obscure black box methods. Further, only in recent years ML methods have reached a mature state mainly driven by commercial needs, resulting in a variety of new applications.

The border between ML, which is a subtopic of artificial intelligence, and other purely mathematical regression methods is difficult to define and is not possible without some degree of arbitrariness. Still, several definitions have been proposed in the literature over the years starting from the famous classical definition given by Arthur Samuel, a pioneer in ML research, in 1959: “[ML is a] Field of study that gives computers the ability to learn without being explicitly programmed,” to a modern definition by Tom M. Mitchell:<sup>1</sup> “A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with the experience  $E$ .”

Nowadays, ML affects everyday life in manifold ways, like character and voice recognition software, fingerprint identification, e-mail spam filtering, autonomously driving cars, computer game opponents, credit card fraud detection and too many others to mention. For these purposes, ML offers

a whole toolbox of techniques, like artificial neural networks,<sup>7</sup> Gaussian processes,<sup>8</sup> and support vector machines.<sup>9</sup>

Driven by the desire for a more rational design of materials, in recent years ML has also established a new trend in computational materials science,<sup>10,11</sup> which shows many similarities to the traditional field of QSAR and drug design. In both cases the goal is to identify systems, molecules or materials, exhibiting certain properties. To reach this goal the machinery of ML is used to search large databases for hidden relationships between the atomic structure, which can be encoded directly or indirectly using a huge number of possible descriptors, and the property of interest.<sup>12</sup>

The idea to employ information from large databases is not new, and repositories of experimental data like the Cambridge Structural Database<sup>13</sup> and the Protein Data Bank<sup>14</sup> have been existing for decades. New is the availability of enormous amounts of high quality theoretical data from electronic structure calculations, which are inaccessible to a manual inspection and analysis. Still, in spite of the high performance of modern supercomputers, it will remain impossible for a long time to screen all candidate structures directly by electronic structure calculations due to the exponentially increasing complexity of configuration space with chemical composition and system size. ML now offers a very exciting tool to fill this gap enabling the analysis of the available data, the discovery of hitherto unknown relationships and the identification of new promising materials.

The topic of this perspective is a different application of ML methods, i.e., the accurate representation of atomic interactions for applications in atomistic computer simulations of problems in chemistry, physics and materials science. In contrast to the classification and pattern recognition applications discussed above, the task to be performed for this purpose is the fitting of a complicated function using a set

of known function values. This function, the potential energy surface (PES), is a multidimensional real-valued function providing the potential energy of a system as a function of the atomic positions.

The concept of a PES is founded on the Born–Oppenheimer approximation<sup>15</sup> of quantum mechanics. If the atomic positions, the nuclear charges and the total charge are known, the potential energy of a system is fully defined by its electronic Hamiltonian. Electronic structure methods enable in principle the calculation of arbitrary atomic configurations corresponding to individual points on the PES, which can be either a ground state or excited state surface. For applications like molecular dynamics (MD) simulations, the energies and forces for a large number of atomic configurations are required, which can be obtained either on-the-fly, by performing the electronic structure calculations as needed while the trajectory evolves, or by evaluating an analytic expression, which provides a direct functional relation between the atomic configuration and the energy. This analytic expression, which has to be defined before running the simulation, is often called “interatomic potential,” “force field” or potential-energy surface.

Electronic structure calculations are computationally very demanding, even if relatively efficient methods like density functional theory (DFT) are used, and DFT-based “*ab initio* MD simulations”<sup>16</sup> are thus restricted to a few hundred atoms and simulation times significantly shorter than 1 ns. Analytic potentials, which provide a much simpler direct relation between the structure and its energy in closed form, can be calculated many orders of magnitude faster, but they are usually derived by introducing physical approximations. While in most cases, the essential features of the atomic interactions are still reasonably described, these potentials necessarily represent a compromise between efficiency and accuracy. Interatomic potentials employing ML techniques aim to combine the advantages of both approaches, the accuracy of first principles calculations and the efficiency of a simple functional form. Much progress has been made in recent years to reach this goal, but still this field is young and developing very dynamically.

In principle, the requirements to be met by ML potentials are very similar to conventional empirical potentials. The evaluation should be fast to enable large scale MD and Monte Carlo simulations. The computational effort to construct the potential should be significantly lower than the costs of a direct application of electronic structure methods. Minimum human effort and manual intervention should be needed to develop the potential and systematic improvements should be possible. Analytic energy gradients for the calculation of the forces, the stress tensor and other gradient-dependent quantities should be available. The information required to evaluate the potential should not include any classification of atoms into types beyond the specification of the nuclear charge, as chemical environments and bonding situations should be allowed to change in the course of simulations. Consequently, also the use of predefined atomic connectivities and bonds should be avoided to obtain a reactive potential enabling the making and breaking of bonds. Ideally, potentials should be generally applicable and not restricted to certain

types of atomic interactions, like covalent, ionic, or metallic bonding, and due to their purely mathematical unbiased form ML methods are particularly promising candidates to reach this goal. Finally, the most important requirement is the accuracy of the potential, i.e., the energies and forces should be as close as possible to the underlying first principles data. Aiming for predictive simulations, the most stringent test concerns the transferability to structures, which are very different from the data used in the construction of the potential.

Based on these requirements, a definition of ML potential can be given. A ML potential

- employs a ML method to construct a direct functional relation between the atomic configuration and its energy;
- does not contain any physical approximations apart from the chosen reference electronic structure method used in its construction;
- is developed using a consistent set of electronic structure data.

In particular, the last point is of high practical importance because the use of experimental data or the mixing of electronic structure data from different levels of theory would result in numerically inconsistent data, which would give rise to serious problems in the construction of the potential.

The application of ML potentials in computer simulations in chemistry, physics, and materials science involves several steps. When using a ML potential, e.g., in MD simulations, in the first step the atomic positions need to be transformed to a set of coordinates suitable as input for the ML method. The complexity of this crucial step has represented a significant challenge in the early days of ML potentials. This problem emerged rather unexpectedly since in most empirical potentials feeding the structure into the functional form is quite trivial, and often, like in classical force fields, readily available internal coordinates like interatomic distances, bond angles, and dihedral angles are well suited. Overcoming the problem of finding appropriate structural descriptors for ML potentials has slowed down their development substantially in the first ten years, but nowadays several options are available. After the transformation of the coordinates onto a suitable set of descriptors, which must have certain properties as discussed in Section II, a ML method is used to evaluate the energy and if required the atomic forces. In principle, they can further be validated to get an estimate for the reliability. They are then passed to the propagator of the MD code and a new atomic configuration is obtained, which then undergoes the same cycle again.

In this perspective, the construction and application of ML potentials for computer simulations with first principles accuracy is reviewed and discussed. A special focus is put on the two central components of any ML potential: the transformation of the atomic positions onto suitable descriptors and the subsequent assignment of an energy to this structure using a functional form provided by a ML method. Since there is no strict border between ML potentials and other available mathematical fitting approaches to represent PESs, the list of methods covered in this perspective cannot be complete, and several important related approaches like

splines,<sup>17</sup> modified Shepard interpolation,<sup>18</sup> interpolating moving least squares,<sup>19</sup> and permutation invariant polynomial fits,<sup>20</sup> which are very similar in spirit in that they do not employ a physically derived functional form and to some extent face similar challenges, will not be discussed here.

## II. STRUCTURAL DESCRIPTION

### A. The role of the descriptor

Many different names have been used in the literature for the effective coordinates, i.e., the descriptors, representing the structural inputs for the ML algorithms. The set of descriptors plays a central role, not only in the construction of PESs,<sup>21</sup> but also beyond in applications like structure-property relationships for property prediction in materials science<sup>22</sup> and also in the strongly related older field of QSAR.<sup>23</sup> Further applications with similar requirements regarding the structural description are the analysis of MD trajectories, the screening of large numbers of atomic configurations for certain features, and structure identification in general.<sup>24</sup> Even the collective variables used in simulation techniques like metadynamics have similar requirements.<sup>25</sup>

The substantial challenge of finding suitable descriptors for ML potentials has been recognized in the advent of the first NN potentials about 20 years ago,<sup>26</sup> which initially have been developed for rather low-dimensional systems like small molecules and diatomic molecules interacting with frozen surfaces. As NNs, like all other ML methods, process ordered vectors of numbers, the output is not invariant with respect to any permutation of the position of equivalent atoms in the structure. Further, directly available descriptors of the atomic positions like Cartesian coordinates are not invariant with respect to translation and rotation of the system, while the resulting ML potential must possess these properties. There have been satisfying customized solutions for specific molecular<sup>27</sup> and surface systems<sup>28,29</sup> also beyond the field of ML potentials, e.g., for permutation invariant polynomials.<sup>20</sup> Still, initially the lack of suitable input descriptors was the main obstacle to construct ML potentials for high-dimensional systems including hundreds or thousands of atoms.

In general, a transformation of the atomic coordinates onto a suitable set of descriptors must provide the same numerical descriptor values for any equivalent atomic configuration, including translation, rotation, permutation, and any applicable point group symmetry. If this is ensured, then all these equivalent representations of the system have the same effective coordinate values and consequently they are considered as the same structure with necessarily the same energy by the ML method. For simple potentials like classical force fields meeting these requirements is straightforward, because internal coordinates are translationally and rotationally invariant, and the total energy is constructed as a sum of very low-dimensional terms, which maintain permutation invariance. For more accurate potentials employing ML, many-body descriptors are needed, that are much more complicated than additive low dimensional functions.

It is of utmost importance that the descriptors do not contain any artificial symmetries because in this case two different structures would be mapped onto the same coordinates and become indistinguishable. Ideally there would be a one-to-one correspondence between the structure and the set of descriptor values, but this would only be guaranteed if there would be one descriptor per degree of freedom in the system. In this case the number of descriptors would depend on the system size, which should be avoided to obtain generally applicable ML potentials as discussed below. It should also be noted that the transformation from the atomic positions to the set of descriptors does not need to be inverted, and rather complicated functional forms can therefore be used.

Apart from these formal requirements there are also practical aspects concerning the choice of descriptors. Since the transformation onto the descriptors has to be carried out for every structure, the evaluation should be fast, and the descriptors need to be differentiable with respect to the atomic positions to enable the calculation of analytic gradients for the forces. To obtain very accurate ML potentials, the descriptors must provide a very detailed structural description and typically multidimensional functions are required to allow for fitting the many-body interactions in electronic structure calculations with high precision.

In summary, the choice of the descriptors is of vital importance because it is the set of descriptors to which the energy is assigned. The number of descriptors has an immediate impact on the feasibility of any ML potential, as it does not only restrict the efficiency of the energy evaluation but also determines the dimensionality of the coordinate space to be mapped by the training sets. On the other hand, the number of descriptors must be sufficiently large to enable an unambiguous distinction of different structures. In Secs. II B 1–II B 4 several descriptors which have been developed for ML potentials are discussed. This list is by far not complete, and there are plenty of new interesting approaches, which have been proposed recently, like moment tensors,<sup>30</sup> scattering transforms,<sup>31</sup> and many others.

## B. Descriptors for machine learning potentials

### 1. Atom centered symmetry functions

A ML approach applicable to high-dimensional systems containing large numbers of atoms has been proposed by Behler and Parrinello in 2007 employing NNs.<sup>32</sup> As in many conventional empirical potentials, the potential energy  $E$  is constructed as a sum of local atomic energies  $E_i$  of all atoms  $i$  in the system,

$$E = \sum_{i=1}^{N_{\text{atoms}}} E_i. \quad (1)$$

This locality of the atomic interactions is known as near-sightedness in quantum chemistry<sup>33</sup> and has also been exploited in a series of electronic structure methods for a long time. The chemical environments of the atoms are characterized by sets of atom-centered symmetry function (ACSF) descriptors, which depend on the positions of the neighboring atoms up to a cutoff radius  $R_c$ , at which the

atomic interactions are screened by the cutoff function,

$$f_c(R_{ij}) = \begin{cases} 0.5 \cdot \left[ \cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1 \right], & \text{for } R_{ij} \leq R_c \\ 0.0, & \text{for } R_{ij} > R_c \end{cases}. \quad (2)$$

This function is essentially the monotonously decreasing part of a cosine function of the distance  $R_{ij}$  to all neighboring atoms  $j$  inside the cutoff radius and zero otherwise. Similar cutoff functions have been used for many years in empirical potentials like the Tersoff potential.<sup>34</sup>

The positions of the atoms inside  $R_c$  are then described by many-body ACSFs, which simultaneously depend on the positions of all neighboring atoms. The most important ACSFs are the “radial function”

$$G_i^{\text{atom,rad}} = \sum_{j=1}^{N_{\text{atom}}} e^{-\eta(R_{ij}-R_s)^2} \cdot f_c(R_{ij}), \quad (3)$$

and the “angular function”

$$G_i^{\text{atom,ang}} = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{all}} (1 + \lambda \cos \theta_{ijk})^\zeta \cdot e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} \cdot f_c(R_{ij}) \cdot f_c(R_{ik}) \cdot f_c(R_{jk}). \quad (4)$$

The radial function in Eq. (3) is a sum of Gaussians of  $R_{ij}$ , which can be shifted by parameter  $R_s$  and are multiplied by  $f_c$  to ensure a smooth decay in value and slope to zero at  $R_c$ . The width of the Gaussians is controlled by the parameter  $\eta$  and typically a set of  $\eta$  values is used to obtain a radial fingerprint of the environment. Due to the summation over all neighbors in the cutoff spheres, the number of ACSFs is independent of the coordination of each atom  $i$ , which is important since NNs require a constant number of input descriptors irrespective of the coordination number that might change frequently in MD simulations.

The angular ACSF (Eq. (4)) provides an angular fingerprint of the environment using the angles  $\theta_{ijk}$  formed with each pair of neighbors  $j$  and  $k$ . The angular resolution can be controlled by the parameter  $\zeta$ , while  $\lambda = \pm 1$  defines the positions of the extrema of the cosine function.

A number of potentials have been constructed using these ACSFs,<sup>35</sup> and typically 50–100 symmetry functions are used per atom differing in the specific values of  $\eta$ ,  $\zeta$ ,  $R_s$  and  $\lambda$ . The cutoff  $R_c$  is a convergence parameter, which needs to be increased until all energetically relevant interactions are included, and in most cases values between 6 and 9 Å have been found to be sufficient. All ACSFs provide a rotationally and translationally invariant description of the environment because they depend on the internal coordinates  $R_{ij}$  and  $\theta_{ijk}$ . Further, because of the sum over all neighbors, they are invariant with respect to any permutation of chemically equivalent atoms in the environment. The number of ACSFs grows with the number of elements in the system because for each atom there is one radial function per element. The angular functions must be constructed for all possible element combinations involved in the angles  $\theta_{ijk}$ . Further details and a discussion of the properties of these and other ACSFs can be found elsewhere.<sup>36</sup>

It can easily be shown that Eq. (1) can be written as a sum of atom pair energies  $E_{ij}$ ,

$$E = \sum_{i=1}^{N_{\text{atoms}}} \sum_{j>i}^{N_{\text{atoms}}} E_{ij}. \quad (5)$$

Also this form can be used for the construction of ML potentials using the pair centered symmetry functions (PCSF)<sup>37</sup>

$$G_{ij}^{\text{pair,rad}} = f_c(R_{ij}) \cdot e^{-\eta \cdot R_{ij}^2} \cdot \left[ \sum_k f_c(R_{ik}) e^{-\eta \cdot R_{ik}^2} + \sum_k f_c(R_{jk}) e^{-\eta \cdot R_{jk}^2} \right] \quad (6)$$

and

$$G_{ij}^{\text{pair,ang}} = f_c(R_{ij}) \cdot e^{-\eta \cdot R_{ij}^2} \cdot 2^{1-\zeta} \sum_{\alpha} \left[ (1 + \lambda \cos \alpha)^\zeta \cdot e^{-\eta \cdot (R_{ik}^2 + R_{jk}^2)} \cdot f_c(R_{ik}) \cdot f_c(R_{jk}) \right], \quad (7)$$

which have been demonstrated to be equally suitable for obtaining high-quality potentials.

## 2. Bispectrum of the neighbor density

An alternative approach to describe the atomic environments based on Eq. (1) has been proposed by Csányi and coworkers in 2010<sup>38</sup> for constructing ML potentials for high-dimensional systems employing Gaussian processes. In this very systematic approach, the selection of parameters defining the spatial shape of the descriptors, that is required in case of ACSFs, is avoided by an expansion of the environment in a series of spherical harmonics.

The starting point for developing the descriptors for the atomic environments is the construction of a neighbor density  $\rho_i(\mathbf{R})$  at each point in space  $\mathbf{R}$  for each atom  $i$ . This is done by placing  $\delta$  functions at all positions of neighboring atoms  $j$  in the environment up to the cutoff and at the position of the reference atom itself as<sup>38</sup>

$$\rho_i(\mathbf{R}) = \delta(\mathbf{R}) + \sum_{R_{ij} < R_c} f_c(R_{ij}) \cdot w_j \cdot \delta(\mathbf{R} - \mathbf{R}_{ij}). \quad (8)$$

The dimensionless weights  $w_j$  enable a discrimination of different elements. The cutoff function  $f_c$ , which is the same as used for the ACSFs (Eq. (2)), ensures a smooth decay to zero at the cutoff  $R_c$ . The angular distribution of the neighbors is then obtained by projecting all neighbors onto a sphere to obtain a projected density  $\hat{\rho}(\theta, \phi)$ , which is then expanded in spherical harmonics  $Y_{lm}$ ,

$$\hat{\rho}(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_{lm} Y_{lm}(\theta, \phi). \quad (9)$$

While the radial distribution could be described in principle by a set of radial functions in full analogy to basis sets used in electronic structure calculations, instead they are converted to an additional angle  $\theta_0$  extending the spherical harmonics to four dimensions. The positions on the four-dimensional sphere, which are given by

$$(\phi, \theta, \theta_0) = (\tan^{-1}(y/x), \cos^{-1}(z/|\mathbf{r}|), |\mathbf{r}|/r_0) \quad (10)$$

are then expanded in hyperspherical harmonics  $U_{m,m'}^j$  as

$$\rho = \sum_{j=0}^{\infty} \sum_{m,m'=-j}^j c_{m,m'}^j U_{m,m'}^j. \quad (11)$$

The coefficients  $c_{m,m'}^j = \langle U_{m,m'}^j | \rho \rangle$  are used to build the bispectrum matrix,

$$B_{j_1,j_2,j} = \sum_{m_1,m'_1=-j_1}^{j_1} \sum_{m_2,m'_2=-j_2}^{j_2} \sum_{m,m'=-j}^j (c_{m,m'}^j) \\ \cdot C_{j_1 m_1 j_2 m_2}^{jm} \times C_{j_1 m'_1 j_2 m'_2}^{jm'} \cdot c_{m'_1 m_1}^{j_1} \cdot c_{m'_2 m_2}^{j_2}, \quad (12)$$

which has originally been developed for image and pattern recognition applications.<sup>39</sup> The  $C_{j_1 m_1 j_2 m_2}^{jm}$  are Clebsch-Gordon coefficients.

Like in the case of ACSFs, the structural description by the bispectrum has all required properties like translational and rotational as well as permutation invariance. An advantage is the possibility to refine the description of the environments very systematically by using more and more spherical harmonics.

### 3. Smooth overlap of atomic positions

A drawback of the neighbor density defined in Eq. (8) is that it employs  $\delta$ -functions and consequently even slight deviations of the positions between two environments result in strong numerical changes. In the SOAP (Smooth Overlap of Atomic Positions) approach,<sup>21</sup> the  $\delta$  functions have therefore been replaced by Gaussians centered at the central atom and at all  $N_{\text{env}}$  neighboring atoms in the environment. The neighbor density in SOAP is then given by

$$\rho_{\text{SOAP}}(\mathbf{R}) = \sum_{i=1}^{N_{\text{env}}} \exp(-\alpha|\mathbf{R} - \mathbf{R}_i|^2). \quad (13)$$

This expression is related to the radial ACSFs (Eq. (3)) in Section II B 1 in which a sum of distance-dependent Gaussians is used. But while for each radial symmetry function  $G_i^{\text{atom,rad}}$  only a single function value is obtained and the full environment is characterized by a set of functions,  $\rho_{\text{SOAP}}$  can be viewed as a three-dimensional generalization characterizing the full atomic environment at once. Consequently,  $\rho_{\text{SOAP}}$  cannot be rotationally invariant, and when comparing two different neighbor densities  $\rho_{\text{SOAP}}$  and  $\rho'_{\text{SOAP}}$ , one of the densities needs to be integrated over all possible rotations to incorporate this invariance.

The SOAP kernel for comparing two environments is therefore the overlap of the two corresponding neighbor densities, one of which has been integrated over all rotations  $\hat{R}$  as<sup>21</sup>

$$k(\rho_{\text{SOAP}}, \rho'_{\text{SOAP}}) = \int d\hat{R} \left| \int \rho_{\text{SOAP}}(\mathbf{r}) \rho'_{\text{SOAP}}(\hat{\mathbf{R}}\mathbf{r}) d\mathbf{r} \right|^{n_{\text{SOAP}}}. \quad (14)$$

The exponent  $n_{\text{SOAP}}$  is usually chosen as 2, since for  $n_{\text{SOAP}} = 1$  the rotational information is lost due to the interchangeability of the integrals. Finally, a normalization by the factor  $\frac{1}{\sqrt{k(\rho_{\text{SOAP}}, \rho_{\text{SOAP}})k(\rho'_{\text{SOAP}}, \rho'_{\text{SOAP}})}}$  ensures that the overlap

of any environment with itself is one. Finally, an expansion of  $\rho_{\text{SOAP}}$  using a set of normalized radial basis functions  $g_n(r)$  and spherical harmonics has been advocated<sup>21</sup> as it is computationally significantly more efficient resulting in the kernel

$$k(\rho_{\text{SOAP}}, \rho'_{\text{SOAP}}) = \sum_{n,n',l} p_{n,n',l} p'_{n,n',l} \quad (15)$$

with

$$p_{n,n',l} = \sum_m c_{n,n',l} (c_{n,n',l})^* \quad (16)$$

containing the expansion coefficients  $c_{n,n',l}$  of the basis functions and  $p_{n,n',l}$  corresponding to the power spectrum.

### 4. Coulomb matrix

In 2012 Rupp *et al.* have proposed a very different descriptor in form of the eigenvalues of a Coulomb matrix,<sup>40</sup> which is related to the concept of a distance matrix. The Coulomb matrix  $\mathbf{M}$  is defined as

$$M_{IJ} = \begin{cases} 0.5 \cdot Z_I^{2.4}, & \text{for } I = J \\ \frac{Z_I \cdot Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}, & \text{for } I \neq J \end{cases} \quad (17)$$

The diagonal elements represent a polynomial fit of the atomic energies to the nuclear charge  $Z_I$ . The off-diagonal elements take the form of the Coulomb repulsion between the nuclei, which can be viewed as scaled inverse distances. Apart from containing the full information about the pairwise distances of all atoms, the inclusion of the nuclear charges introduces numerical offsets for different element combinations allowing to encode information about the chemical composition without the introduction of additional descriptors, which however might also result in numerical difficulties in the fits due to the range of values over many orders of magnitude. While the original Coulomb matrix in Eq. (17) contains an overcomplete description of the system, which would be no problem for ML methods, the  $N_{\text{atoms}}$  permutation invariant eigenvalues represent only a subset of the information in the  $3 \cdot N_{\text{atoms}} - 6$  dimensional configuration space and nicely illustrate the frequent problem of simultaneously finding a complete descriptor while the required symmetry is still fully included. The size of the Coulomb matrix, which contains one line and one row per atom, is given by the largest system used in the reference data set, and for smaller systems the unnecessary parts of the matrix are filled with zeros. Like the other descriptors discussed before, the eigenvalues of the Coulomb matrix are invariant with respect to rotation, translation, and permutation.

Several of the drawbacks of the original Coulomb matrix have been removed by introducing local Coulomb matrices<sup>41</sup> depending on the neighboring atoms inside a cutoff radius. In this way the description of many-atom systems becomes possible, while the locality of many properties is exploited like in the other methods developed for high-dimensional systems. Still, access to the analytic derivatives is complicated, as the eigenvalues are not continuous and differentiable if the number of neighbors in the environment changes.<sup>21</sup>

To date, the Coulomb matrix has not been applied to the construction of PESs, but several energy-related applications have been reported, like the atomization energies of small organic molecules,<sup>40</sup> and atomic forces as a function of the environment.<sup>41</sup> Further, also new related approaches like the “bag of bonds” method have been proposed recently with promising first results.<sup>42</sup>

### III. A BRIEF SURVEY OF ML POTENTIALS

#### A. Overview

The development of ML potentials has taken place in two steps, which are intimately related to the progress made in the derivation of suitable structural descriptors. In the initial phase, ML potentials have been available only for low-dimensional systems, since ML approaches based on the decomposition of the total energy into atomic contributions according to Eq. (1) had not been devised. These early potentials almost exclusively relied on artificial neural networks<sup>43,44</sup> and have been applicable to systems containing up to about six atoms. Nevertheless, several interesting and quite successful attempts have also been made in this early stage to extend the applicability of NNPs to systems including larger numbers of atoms.<sup>45,46</sup> Starting with NNPs<sup>32</sup> in 2007, several different ML approaches suitable for high-dimensional systems containing thousands of atoms have now become available<sup>38,40</sup> and are still evolving rapidly.<sup>10,47,48</sup> An overview about the increasing use of different ML techniques to the construction of PESs is shown in Fig. 1.

In this section, several ML methods along with their use in the construction of PESs are outlined, which in principle could be combined almost arbitrarily with the descriptors discussed above. Still, currently there are a few dominant combinations of descriptors and ML methods, but this is not because of the compatibility or incompatibility of the different approaches, but it is mainly a consequence of the personal preferences of the developers.

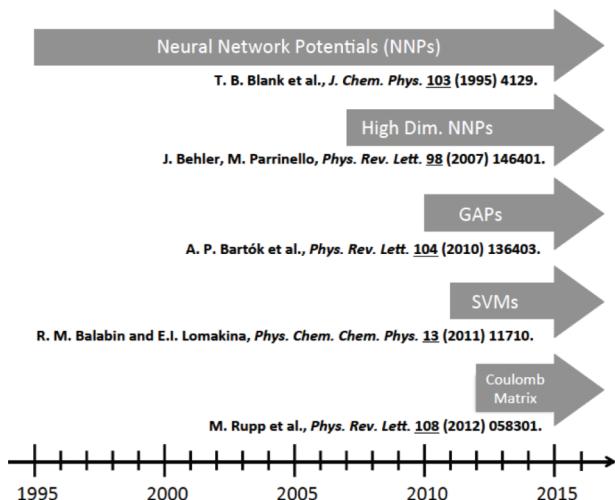


FIG. 1. Historic development of machine learning potentials.

#### B. Neural networks

Artificial neural networks are the oldest ML method<sup>49</sup> and they have first been developed to understand the signal processing in the brain. They also represent the first ML method which has been used to fit a DFT PES by Doren and coworkers in 1995,<sup>26</sup> and to date a large number of NN potentials have been developed for many molecular and condensed systems.<sup>27,28,50</sup>

The central component of all NN potentials (NNPs) are feed-forward neural networks (FFNNs) as shown in Fig. 2. The artificial neurons plotted as circles are arranged in one or more hidden layers and provide the functional flexibility to represent arbitrary functions. In the input layer a set of descriptors  $\{G_i\}$  is fed into the network, the output node on the right provides the potential energy  $E$ . Each node  $i$  in each layer  $k$  is connected to the nodes  $j$  in the next layer  $l = k + 1$  by weight parameters  $a_{ij}^{kl}$ , which are the fitting parameters of the network. Thus, the small FFNN in Fig. 2 corresponds to the analytic form

$$E = f \left( b_1^3 + \sum_{k=1}^3 a_{k1}^{23} \cdot f \left( b_k^2 + \sum_{j=1}^3 a_{jk}^{12} \cdot f \left( b_j^1 + \sum_{i=1}^2 G_i \cdot a_{ij}^{01} \right) \right) \right). \quad (18)$$

The bias weights  $b_i^j$  are acting as adjustable offsets for each node, and the functions  $f$  are non-linear activation functions like the hyperbolic tangent or a Gaussian function. The weight parameters are determined using energies and forces from electronic structure calculations in an iterative gradient-based optimization process and finally they contain all the information of the reference set. When applying the NN potential, Eq. (18) is directly evaluated without the need to know the reference set.

By using a FFNN for each atom in the system and employing Eq. (1), an extension to high-dimensional NN potentials applicable to very large systems has been proposed by Behler and Parrinello in 2007.<sup>32</sup> For this purpose, the atomic environments have been described by ACSFs. Further details about this method can be found elsewhere.<sup>35,51</sup> In 2013 it has been explicitly demonstrated by Geiger and Dellago<sup>52</sup> that this scheme, i.e., the combination of atomic NNs and ACSFs, can

<b>Input Layer</b>	<b>Hidden Layer 1</b>	<b>Hidden Layer 2</b>	<b>Output Layer</b>
--------------------	-----------------------	-----------------------	---------------------

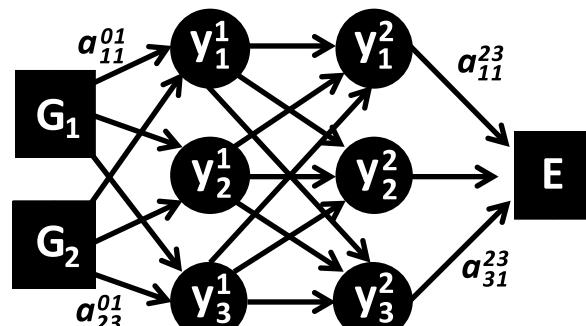


FIG. 2. Structure of a feed-forward neural network.<sup>7</sup>

also be used to accurately identify amorphous and crystalline structures in MD simulations. They found that this approach is superior to conventional methods for structure identification like Steinhardt parameters,<sup>53</sup> which supports the suitability of ACSFs for the distinction of different structural motifs. This application, which is very useful on its own, can be viewed as an intermediate step of the construction of the NN PES, since the proper characterization of the atomic environments is a mandatory ingredient for obtaining a reliable structure-energy relation in NNPs.

Finally, it should be noted that NNs are currently experiencing an enormous boost in development under the name “deep learning,” which is essentially based on very large NNs.

### C. Gaussian approximation potentials and kernel methods

Gaussian approximation potentials (GAPs)<sup>38,54</sup> are an important example for kernel-based ML potentials, which are obtained by combining a suitable structural descriptor and a kernel establishing the connection between structure and energy. In case of GAPs, the atomic energies  $E_i$  in Eq. (1) are interpolated in the bispectrum descriptor space using Gaussian process regression,<sup>8</sup>

$$\begin{aligned} E_i &= \sum_{n=1}^{N_{\text{ref}}} \alpha_n e^{-0.5 \cdot \sum_{l=1}^L [(b_l - b_{n,l}) / \theta_l]^2} \\ &= \sum_{n=1}^{N_{\text{ref}}} \alpha_n e^{G(\mathbf{b}, \mathbf{b}_n)}, \end{aligned} \quad (19)$$

where  $L$  is the number of components in the truncated bispectrum and the  $\theta_l$  are fitting parameters. Thus, the atomic energy is a weighted sum over the energies of the known atomic environments in the reference set, which has to be available when applying the potential. The determination of the parameters  $\alpha_n$  requires an  $O(N_{\text{ref}}^3)$  inversion of the covariance matrix, which is a rather demanding operation.<sup>55</sup> In earlier work, also Popelier and coworkers have used Gaussian processes, which are also called Kriging, for the fitting of electrostatic multipoles<sup>56</sup> for improving conventional force fields.

GAPs belong to the class of kernel methods, which have the general form

$$E_i(\mathbf{d}_i) = \sum_{j=1}^{N_{\text{basis}}} \alpha_j \chi_j(\mathbf{d}_i), \quad (20)$$

i.e., they are linear combination of some basis functions  $\chi_j$  depending on descriptor  $\mathbf{d}_i$  for the environment of  $i$ . The coefficient vector  $\alpha$  is independent of the specific environment. The energy is expanded in terms of the known data points using a nonlinear kernel function  $K(\mathbf{d}, \mathbf{d}^j)$  providing a measure for the similarity of any two structures or environments,<sup>54</sup>

$$E_i = \sum_{j=1}^{N_{\text{train}}} \alpha_j K(\mathbf{d}, \mathbf{d}^j). \quad (21)$$

GAPs use the squared exponential kernel

$$K_{\text{SE}}(\mathbf{d}, \mathbf{d}') = \exp \left( - \sum_j \frac{(d_j - d'_j)^2}{\theta_j} \right). \quad (22)$$

Kernel methods are very frequently used. Another example is the combination with a Coulomb matrix descriptor for representing the atomization energies of organic molecules by Rupp *et al.*<sup>40</sup> Here, the “distance” between two molecules is defined as

$$d(\mathbf{M}, \mathbf{M}') = d(\epsilon, \epsilon') = \sqrt{\sum_n |\epsilon_n - \epsilon'_n|^2}, \quad (23)$$

which is the Euclidean norm of the diagonalized Coulomb matrix.

### D. Support vector machines

Support Vector Machines (SVMs)<sup>9</sup> belong to the most important ML methods, but to date they have been rarely used for the construction of PES.<sup>57,58</sup> Originally, SVMs have been developed for the classification of data into two groups with different properties by introducing a multidimensional hyperplane separating the data. As this hyperplane usually has a very complicated non-linear shape, the data is processed by a kernel substitution, which is in essence the mapping of the original data onto a higher-dimensional feature space, in which a linear separation is possible (see Fig. 3). In order to assign the data to real-valued properties like energies, instead of the original SVMs, an extension in the form of support vector regression (SVR) is used.<sup>57</sup>

### E. Spectral neighbor analysis potential

The Spectral Neighbor Analysis Potential (SNAP) introduced by Thompson *et al.*<sup>59</sup> represents a linear version of GAPs and is based on the same bispectrum components, which are now assumed to be related linearly to the atomic energies. The atomic energy of atom  $i$  in SNAP is then given as a linear combination of the  $K$  projected bispectrum components by<sup>59</sup>

$$E_{i,\text{SNAP}} = \beta_0^{\alpha_i} + \sum_{k=1}^K \beta_k^{\alpha_i} \cdot B_k^i \quad (24)$$

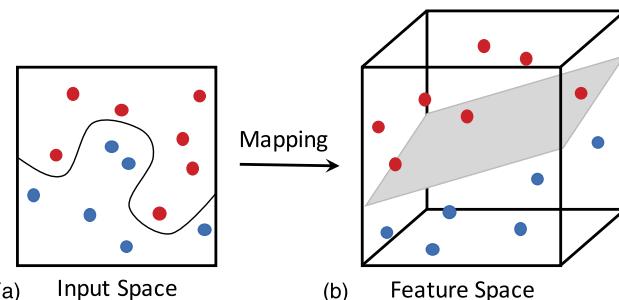


FIG. 3. Operation mode of support vector machines (SVMs).<sup>9</sup> The data to be classified into two categories cannot be separated linearly by a hyperplane (a). This linear separation becomes possible after the transformation onto a higher-dimensional space as shown in (b).

using the coefficients  $\beta_k^\alpha$ , which can be determined in a least-squares linear regression.  $\alpha_i$  specifies the atom type of atom  $i$ ,  $B_k^i$  is the bispectrum component  $k$  of atom  $i$ , and  $\beta_0^{\alpha_i}$  is a constant element-specific contribution. Also further properties related to the energy gradients, like the forces and the stress tensor, can be expressed as similar linear combinations of the corresponding derivative terms.

#### IV. DISCUSSION AND OUTLOOK

Starting from the two main components discussed in Secs. II and III, ML potentials can be constructed by combining a descriptor to convert the structure into a representation fulfilling the required invariances and a ML method to relate the set of descriptor values to an energy. In principle this approach is quite modular and many combinations of descriptors and ML techniques are possible, although only a few have been used so far.

The obtained ML potentials have a solid theoretical foundation. Concerning the physics this is the Born Oppenheimer approximation. Mathematically it has been proven for several ML approaches, e.g., for NNs,<sup>60,61</sup> that very accurate representations of the potential energy functions are possible, although for practical applications these proofs are of limited use. Therefore, the application of ML to PES fitting is more rigorous than the prediction of other physical properties, although in the end most properties are related to the quantum mechanical Hamiltonian, i.e., to the structure. This relation, however, might be very indirect and thus more complicated to identify for ML.

ML potentials have many advantages. They offer a very fast calculation of the energies and forces, although they are usually one or two orders of magnitude slower than very simple force fields. High-dimensional ML potentials employing Eq. (1) can easily be parallelized and scale close to linearly with system size. Therefore, they enable simulations of very large systems much beyond the realm of *ab initio* MD. They all provide very accurate energies close to the underlying electronic structure methods, but they cannot and must not provide better results than this chosen reference method because by themselves they are not relying on a physical functional form but must learn the physical shape of the PES from the provided example data.

Still, there are also some remaining challenges. The construction of ML potentials currently requires very large reference data sets from electronic structure calculations, making the construction computationally very demanding and time consuming. Reducing the size of the reference sets should therefore be one of the primary goals, and all available information about the PES, like forces, should also be used. The minimum size of the reference sets strongly depends on the complexity of the PES, which must be fully mapped, and research in this direction is just in its infancy. In any case, the identification of missing relevant structures is closely related to the analysis of structural similarities.<sup>24,62,63</sup> Alternatively, the flexibility of the ML methods can also be used to identify structures, which are too far away from the known reference set.<sup>51</sup> Another interesting approach is to test ML forces on-the-fly in MD simulations and to add more data if necessary

while running the simulations.<sup>55</sup> In any case, ML potentials require very careful testing and validation because they can fail spectacularly, if they have not been constructed properly. ML potentials are a tool to speed up simulations, but they are unable to provide new insights that have not been provided in the training set, but these insights might be complex and hidden and very difficult to extract without the use of ML techniques.

Another current limitation is caused by the exponentially growing complexity of configuration space with chemical composition, which drastically increases both, the number of descriptors and the amount of reference data. Therefore, many ML potentials are restricted to only a few chemical elements, a limitation that they share with many other potentials in materials science. Only if the configuration space is reduced, either by using only very low-dimensional terms like in classical force fields, or by the restriction to local minima of the PES,<sup>40</sup> a higher chemical complexity can be dealt with.

A future trend to overcome this limitation will certainly be the combination of ML methods with physically meaningful energy terms. The first step in this direction has already been taken for the most obvious choice, the electrostatic energy contribution. Already in 2007 Popelier and coworkers demonstrated that NNs can be used to represent atomic electrostatic multipoles<sup>64</sup> to obtain improved electrostatic energy terms in classical force fields. In 2011 environment-dependent atomic charges have been expressed as an extension to the short range energy part in high-dimensional NN potential to avoid the truncation of the atomic interactions at the cutoff radius defining the atomic environments.<sup>65</sup> Both approaches rely on reference charges from electronic structure calculations, which are mathematically well-defined but physically not unique. A possible solution to overcome this limitation in a ML framework has been proposed by Ghasemi *et al.*<sup>66</sup>

#### V. CONCLUSIONS

A lot of progress has been made in recent years in the development of interatomic potentials based on the combination of ML techniques and data from electronic structure calculations. For this purpose, a toolbox of structural descriptors and ML methods is now ready for use and several new techniques can be expected to enter the stage soon. First applications have demonstrated that potentials of very high accuracy can be obtained for realistic systems containing many atoms and complex atomic configurations. Apart from benchmark studies, state-of-the-art simulations can now be performed, which allow to solve problems in chemistry, physics, and materials science that have hitherto been inaccessible due to the lack of suitable potentials with first-principles accuracy.<sup>67–73</sup>

In spite of this success, the construction of ML potentials is still very demanding and the potentials need to be tested very carefully due to the lack of a physical basis of their functional form. Consequently, the distribution of potentials from developers to users is not as quickly as for other types of potentials, which is certainly one of the main reasons for the only slowly increasing usage of these methods. Further

efforts are also required regarding the applicability to systems containing more than a few elements. This bottleneck, which ML potentials share with many conventional potentials in materials science, is one of the most urgent problems to be addressed in the near future. Finally, it would be desirable to compare different approaches for well-defined benchmark cases to identify and possibly combine the advantages of these methods. To date only the tip of the iceberg of the field of ML has been exploited in the construction of interatomic potentials and many new approaches will be developed to further extend the applicability of ML-based potentials in atomistic simulations.

## ACKNOWLEDGMENTS

The author thanks the DFG for financial support (Heisenberg Fellowship No. Be3264/6-1). This work was supported by the Cluster of Excellence RESOLV (EXC 1069) funded by the Deutsche Forschungsgemeinschaft.

- <sup>1</sup>T. M. Mitchell, *Machine Learning* (McGraw-Hill, 1997).
- <sup>2</sup>B. G. Sumpter, C. Getino, and D. W. Noid, *Annu. Rev. Phys. Chem.* **45**, 439 (1994).
- <sup>3</sup>J. Gasteiger and J. Zupan, *Angew. Chem. Int. Ed.* **32**, 503 (1993).
- <sup>4</sup>J. U. Thomsen and B. Meyer, *J. Magn. Reson.* **84**, 212 (1989).
- <sup>5</sup>M. Keil, T. E. Exner, and J. Brickmann, *J. Comput. Chem.* **25**, 779 (2004).
- <sup>6</sup>D. K. Agrafiotis, W. Cedeno, and V. S. Lobanov, *J. Chem. Inf. Comput. Sci.* **42**, 903 (2002).
- <sup>7</sup>S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. (Prentice Hall, 2008).
- <sup>8</sup>C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, 2006).
- <sup>9</sup>N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (Cambridge University Press, 2000).
- <sup>10</sup>L. Ward and C. Wolverton, “Atomistic calculations and materials informatics: A review,” *Curr. Opin. Solid State Mater. Sci.* (in press).
- <sup>11</sup>A. Jain, G. Hautier, S. P. Ong, and K. Persson, *J. Mater. Res.* **31**, 977 (2016).
- <sup>12</sup>G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, *Sci. Rep.* **3**, 2810 (2013).
- <sup>13</sup>F. H. Allen, *Acta Crystallogr.* **B58**, 380 (2002).
- <sup>14</sup>F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *Arch. Biochem. Biophys.* **185**, 584 (1978).
- <sup>15</sup>M. Born and R. Oppenheimer, *Ann. Phys.* **389**, 457 (1927).
- <sup>16</sup>D. Marx and J. Hutter, *Ab initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, 2009).
- <sup>17</sup>W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Programming*, 3rd ed. (Cambridge University Press, 2007).
- <sup>18</sup>P. Lancaster and K. Salkauskas, *Curve and Surface Fitting: An Introduction* (Academic Press, 1986).
- <sup>19</sup>T. Ishida and G. C. Schatz, *Chem. Phys. Lett.* **314**, 369 (1999).
- <sup>20</sup>B. J. Braams and J. M. Bowman, *Int. Rev. Phys. Chem.* **28**, 577 (2009).
- <sup>21</sup>A. P. Bartók, R. Kondor, and G. Csányi, *Phys. Rev. B* **87**, 184115 (2013).
- <sup>22</sup>L. M. Ghiringhelli, J. Vyvial, S. V. Levchenko, C. Draxl, and M. Scheffler, *Phys. Rev. Lett.* **114**, 105503 (2015).
- <sup>23</sup>*Topological Indices and Related Descriptors in QSAR and QSPR*, edited by J. Deviller and A. T. Balaban (Gordon and Breach Science Publishers, 1999).
- <sup>24</sup>M. Valle and A. Oganov, *Acta Crystallogr., Sect. A: Found. Crystallogr.* **66**, 507 (2010).
- <sup>25</sup>A. Laio and M. Parrinello, in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, Lecture Notes in Physics Vol. 703, edited by M. Ferrario, G. Cicotti, and K. Binder (Springer, 2006), pp. 315–347.
- <sup>26</sup>T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, *J. Chem. Phys.* **103**, 4129 (1995).
- <sup>27</sup>H. Gassner, M. Probst, A. Lauenstein, and K. Hermansson, *J. Phys. Chem. A* **102**, 4596 (1998).
- <sup>28</sup>S. Lorenz, A. Groß, and M. Scheffler, *Chem. Phys. Lett.* **395**, 210 (2004).
- <sup>29</sup>J. Behler, S. Lorenz, and K. Reuter, *J. Chem. Phys.* **127**, 014705 (2007).
- <sup>30</sup>A. V. Shapeev, e-print arXiv:1512.06054.
- <sup>31</sup>M. Hirn, N. Poilvert, and S. Mallat, e-print arXiv:1502.02077.
- <sup>32</sup>J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
- <sup>33</sup>W. Kohn, *Phys. Rev. Lett.* **76**, 3168 (1996).
- <sup>34</sup>J. Tersoff, *Phys. Rev. B* **39**, 5566 (1989).
- <sup>35</sup>J. Behler, *J. Phys.: Condens. Matter* **26**, 183001 (2014).
- <sup>36</sup>J. Behler, *J. Chem. Phys.* **134**, 074106 (2011).
- <sup>37</sup>K. V. Jovan Jose, N. Artrith, and J. Behler, *J. Chem. Phys.* **136**, 194111 (2012).
- <sup>38</sup>A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.* **104**, 136403 (2010).
- <sup>39</sup>R. Kondor, e-print arXiv:cs/0701127 (2007).
- <sup>40</sup>M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Phys. Rev. Lett.* **108**, 058301 (2012).
- <sup>41</sup>M. Rupp, R. Ramakrishnan, and O. A. von Lilienfeld, *ACS Catal.* **6**, 1675 (2016).
- <sup>42</sup>K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, *J. Phys. Chem. Lett.* **6**, 2326 (2015).
- <sup>43</sup>C. M. Handley and P. L. A. Popelier, *J. Phys. Chem. A* **114**, 3371 (2010).
- <sup>44</sup>J. Behler, *Phys. Chem. Chem. Phys.* **13**, 17930 (2011).
- <sup>45</sup>S. Hobday, R. Smith, and J. Belbruno, *Modell. Simul. Mater. Sci. Eng.* **7**, 397 (1999).
- <sup>46</sup>S. Manzhos and T. Carrington, Jr., *J. Chem. Phys.* **125**, 84109 (2006).
- <sup>47</sup>C. M. Handley and J. Behler, *Eur. Phys. J. B* **87**, 152 (2014).
- <sup>48</sup>M. Rupp, *Int. J. Quantum Chem.* **115**, 1058 (2015).
- <sup>49</sup>W. McCulloch and W. Pitts, *Bull. Math. Biophys.* **5**, 115 (1943).
- <sup>50</sup>S. Manzhos, X. Wang, R. Dawes, and T. Carrington, Jr., *J. Phys. Chem. A* **110**, 5295 (2006).
- <sup>51</sup>J. Behler, *Int. J. Quantum Chem.* **115**, 1032 (2015).
- <sup>52</sup>P. Geiger and C. Dellago, *J. Chem. Phys.* **139**, 164105 (2013).
- <sup>53</sup>P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, *Phys. Rev. B* **28**, 784 (1983).
- <sup>54</sup>A. P. Bartók and G. Csányi, *Int. J. Quantum Chem.* **115**, 1051 (2015).
- <sup>55</sup>Z. Li, J. R. Kermode, and A. De Vita, *Phys. Rev. Lett.* **114**, 096405 (2015).
- <sup>56</sup>C. M. Handley, G. I. Hawe, D. B. Kell, and P. L. A. Popelier, *Phys. Chem. Chem. Phys.* **11**, 6365 (2009).
- <sup>57</sup>A. Vitek, M. Stachon, P. Krömer, and V. Snasel, in *IEEE 5th International Conference on Intelligent Networking and Collaborative Systems* (2013), p. 121.
- <sup>58</sup>R. M. Balabin and E. I. Lomakina, *Phys. Chem. Chem. Phys.* **13**, 11710 (2011).
- <sup>59</sup>A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, *J. Comput. Phys.* **285**, 316 (2015).
- <sup>60</sup>G. Cybenko, *Math. Control. Signals Syst.* **2**, 303 (1989).
- <sup>61</sup>K. Hornik, *Neural Networks* **4**, 251 (1991).
- <sup>62</sup>S. De, A. P. Bartok, G. Csanyi, and M. Ceriotti, *Phys. Chem. Chem. Phys.* **18**, 13754 (2016).
- <sup>63</sup>L. Zhu, M. Amsler, T. Fuhrer, B. Schaefer, S. Faraji, S. Rostami, S. A. Ghasemi, A. Sadeghi, M. Grauzintyte, C. Wolverton, and S. Goedecker, *J. Chem. Phys.* **144**, 034203 (2016).
- <sup>64</sup>S. Houlding, S. Y. Liem, and P. L. A. Popelier, *Int. J. Quantum Chem.* **107**, 2817 (2007).
- <sup>65</sup>N. Artrith, T. Morawietz, and J. Behler, *Phys. Rev. B* **83**, 153101 (2011).
- <sup>66</sup>S. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker, *Phys. Rev. B* **92**, 045131 (2015).
- <sup>67</sup>R. Z. Khalilullin, H. Eshet, T. D. Kühne, J. Behler, and M. Parrinello, *Nat. Mater.* **10**, 693 (2011).
- <sup>68</sup>G. C. Sosso, D. Donadio, S. Caravati, J. Behler, and M. Bernasconi, *Phys. Rev. B* **86**, 104301 (2012).
- <sup>69</sup>T. Morawietz, A. Singraber, C. Dellago, and J. Behler, *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8368 (2016).
- <sup>70</sup>M. Hellström and J. Behler, *J. Phys. Chem. Lett.* **7**, 3302 (2016).
- <sup>71</sup>N. Artrith and A. M. Kolpak, *Nano Lett.* **14**, 2670 (2014).
- <sup>72</sup>J. S. Elias, N. Artrith, M. Bugnet, L. Giordano, G. A. Botton, A. M. Kolpak, and Y. Shao-Horn, *J. Phys. Chem. Lett.* **6**, 3309 (2015).
- <sup>73</sup>S. Kondati Natarajan and J. Behler, *Phys. Chem. Chem. Phys.* **18**, 28704 (2016).