ADVANCED MATERIALS
www.advmat.de



# Machine Learning Interatomic Potentials as Emerging Tools for Materials Science

*Volker L. Deringer,* Miguel A. Caro, and Gábor Csányi*

**Atomic-scale modeling and understanding of materials have made remarkable progress, but they are still fundamentally limited by the large computational cost of explicit electronic-structure methods such as density-functional theory. This Progress Report shows how machine learning (ML) is currently enabling a new degree of realism in materials modeling: by "learning" electronic-structure data, ML-based interatomic potentials give access to atomistic simulations that reach similar accuracy levels but are orders of magnitude faster. A brief introduction to the new tools is given, and then, applications to some select problems in materials science are highlighted: phase-change materials for memory devices; nanoparticle catalysts; and carbon-based electrodes for chemical sensing, supercapacitors, and batteries. It is hoped that the present work will inspire the development and wider use of ML-based interatomic potentials in diverse areas of materials research.**

## 1. Introduction

Materials science is an interdisciplinary research field spanning many different length scales. The most fundamental one is that of individual atoms, and of the 3D structures that they form, driven by physical laws and chemical bonding. For example, the intercalation of Li and Na ions in electrode materials constitutes the atomistic mechanism for energy storage in next-generation batteries, and it is governed by a delicate interplay of physical forces[1] and the often highly complicated local structures involved.[2] Accurately describing, understanding, and ultimately controlling the atomic-scale structure of matter is therefore a central goal of materials science.[3–5]

Dr. V. L. Deringer, Prof. G. Csányi
Department of Engineering
University of Cambridge
Cambridge CB2 1PZ, UK
E-mail: vld24@cam.ac.uk

Dr. V. L. Deringer
Department of Chemistry
University of Cambridge
Cambridge CB2 1EW, UK

Dr. M. A. Caro
Department of Electrical Engineering and Automation and Department of Applied Physics
Aalto University
Espoo 02150, Finland

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adma.201902765.

**DOI: 10.1002/adma.201902765**

In pursuit of this goal, atomic-scale computer simulations have long been a central approach, and two major families of methods are routinely used today. On the one hand, there are quantum-mechanical simulations, in which we solve Schrödinger's equation for the electronic structure of molecular and periodic systems, most widely based on density-functional theory (DFT).[6–8] These methods provide (largely) reliable results for structural models of materials that normally contain a few tens or hundreds of atoms. State-of-the-art DFT methods can be applied to many material classes, and they are increasingly used for high-throughput screening and "in silico" (computer-based) design of materials: new compositions and previously unknown structures have been identified in DFT searches and subsequently experimentally realized.[5,9–11] On the other hand, interatomic potential models ("force fields"), parameterizing interactions between atoms with (relatively) simple functional forms, are widely used in materials science to describe matter in molecular dynamics (MD) simulations. These simulations grant access to larger time and length scales, reaching system sizes of up to hundreds of thousands of atoms.[12] In parameterizing these potentials, a certain physical form of the atomic interactions is assumed, often in terms of bond distances, angles, and so on, and physical properties such as equilibrium lattice parameters or elastic constants enter the fitting of the potential. For this reason, such potentials are often called "empirical." They are several orders of magnitude faster than DFT, but necessarily less accurate and less easily transferable.

In this Progress Report, we highlight recent developments in "machine-learned" interatomic potentials, which represent a rapidly growing field that promises to do away with the aforementioned trade-offs. Over the last year, there has been a surge of interest in machine learning (ML) methodology: part of it is due to the dramatic growth of ML throughout the scientific disciplines, and part of it is due to tangible success stories of ML-based interatomic potentials that are now beginning to emerge. We will argue that this is an exciting development with very practical implications, currently on the verge of moving from a somewhat specialized new technology to everyday applicability, poised to enhance and complement the communities' existing strengths in computational materials modeling. We will show selected applications of ML potentials to problems in materials science, discuss the current limitations (and possible pitfalls), and outline what we expect to be interesting directions for the development of the field in the coming years.

## 2. Essentials of Machine Learning Potentials

Machine learning is the science and craft of extracting information from large datasets, forming a subset of the broader field of "artificial intelligence." ML methods are currently making a tremendous impact in different areas of physics, chemistry, and materials science,[13–19] with research questions ranging from the accurate prediction of molecular energies[20–22] and exploration of molecular reaction mechanisms[23,24] to the high-throughput screening of functional materials by way of optimizing a target property (say, the electronic band gap for a solar-cell light absorber, or the elastic properties for a superhard material) as a function of structural and other descriptors.[25–27] In what follows, we will strictly focus on one of the many emerging applications of ML in materials science: namely, on the generation and application of ML-based interatomic potentials for condensed-matter systems, which help to understand the structures, reactivities, and physical properties of materials on the atomic scale.

ML potentials are a mathematical representation of the $3N$-dimensional potential-energy surface (PES), that is, of the total energy and interatomic forces for a given set of $N$ atomic positions. The concept resembles that of empirically fitted force fields, and ML potentials likewise make physically motivated assumptions, such as locality and smoothness of the PES. The crucial difference, however, is that ML models do not make any prior assumptions about the shape of the interatomic potential, that is, the specific functional form of the PES as a function of the atomic positions. Instead, any and all such information is extracted directly from a large set of input data, computed at an accurate and computationally much more expensive reference level (commonly, DFT). Once the potential has been fitted, it can be used to predict energies and forces for larger ensembles of atoms, without the need for additional reference data. It thereby gives the user access, e.g., to MD simulations that are orders of magnitude faster than with the reference method, bringing more complex structural problems within computational reach. In fact, access is gained to any property that is derived from the PES: for example, by computing interatomic force constants and vibrational (phonon) modes, one can predict thermal transport properties.[28] However, the ML potentials discussed herein do not provide an explicit description of the electronic structure, and therefore they give no electronic information, such as bandgaps or densities of states (DOS).

Three ingredients are needed to generate an ML potential for a given material (**Figure 1**): a database of reference structures and associated quantum-mechanical data (to which the potential will be fitted); a mathematical way to represent the atomic structure such that it can be fed into the ML algorithm; and finally, the regression or "learning" task itself. We will discuss these three ingredients in sequence.

### 2.1. Ingredient 1: Reference Databases

The starting point for an ML potential is a set of reference configurations for which accurate energies and forces on atoms can be computed, typically using DFT. These reference data are collected for a representative set of atomic environments ("representative" in the sense that the potential is expected to encounter similar environments at runtime), shown as points in the stylized PES in Figure 1a, and stored in a suitable database format. The quality of these datapoints, and the effort and cost

**Volker Deringer** studied chemistry at RWTH Aachen University, where he received his diploma in 2010 and his doctorate in 2014 under guidance of Richard Dronskowski. In 2015, he moved to the University of Cambridge, where he held fellowships from the Alexander von Humboldt Foundation (2015–2017) and the Leverhulme Trust (2017–2019). In September 2019, he will take up an Associate Professorship at the University of Oxford. His research combines quantum-mechanical approaches with machine learning, aiming to understand the connections of structure, bonding, and properties in complex solid-state materials.

**Miguel Caro** received a graduate degree in physics from University of La Laguna in 2009 and a PhD in condensed-matter physics from University College Cork in 2013. After that, he joined Aalto University as postdoctoral researcher. Since 2017 he is Academy of Finland Postdoctoral Researcher at the same institution. His interests include atomistic modeling of materials, carbon-based materials in particular, using simulation techniques from quantum mechanics to empirical force fields. He is currently starting as Principal Investigator with an Academy of Finland Early-Career grant, where he will lead a team working toward improving computational performance of machine learning interatomic potentials.

**Gábor Csányi** studied mathematics at Cambridge University and has a doctorate in physics from the Massachusetts Institute of Technology. He joined the faculty of Cambridge University in 2007 in the engineering laboratory, where he is now Professor of Molecular Modelling. He is interested in computer simulation of materials on the atomic scale, particularly fitting interatomic potentials using machine learning approaches, but also statistical mechanics, optimization, and sampling.

 

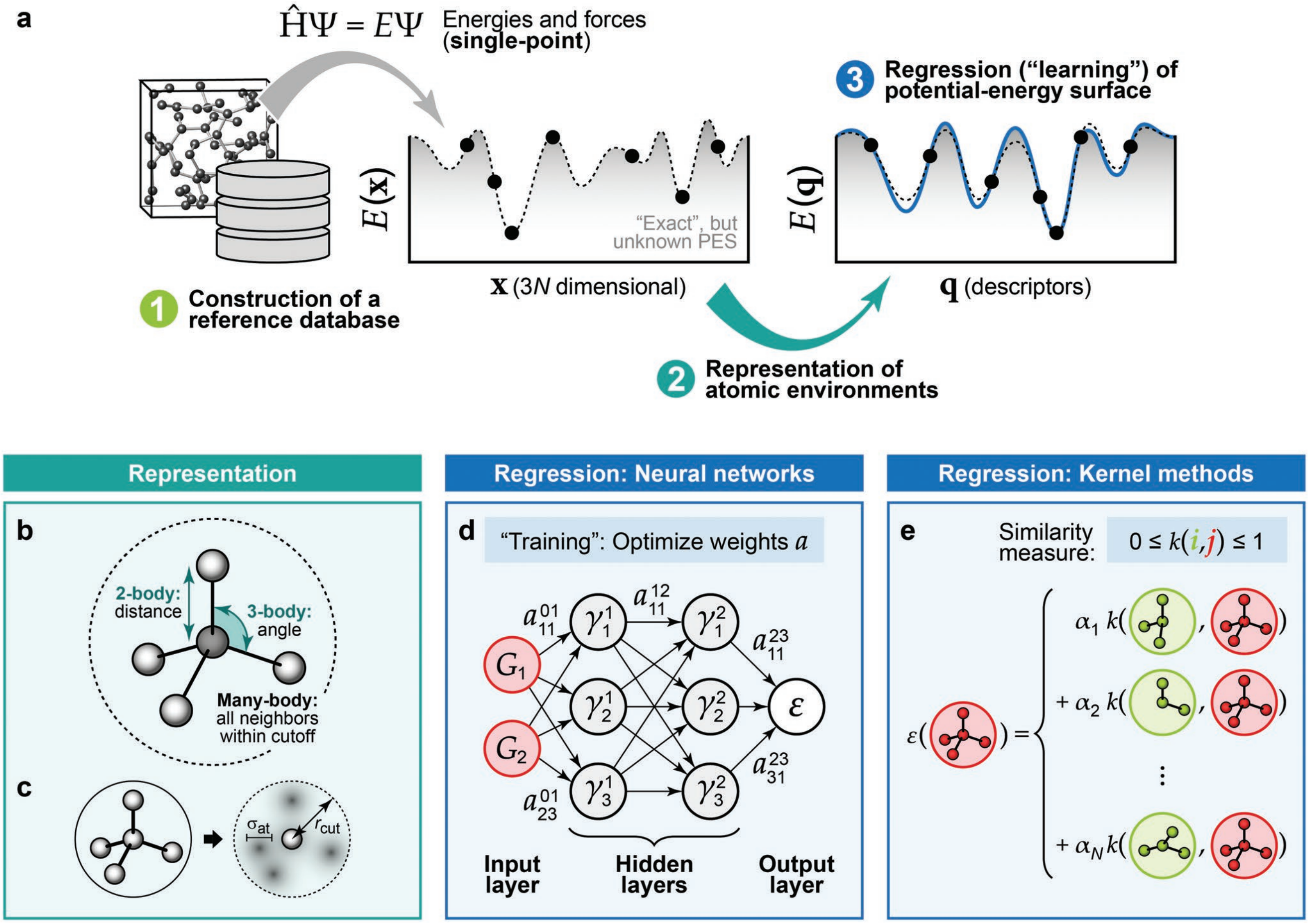


**Figure 1.** Machine-learning methodology for materials simulations. a) General overview of how ML-based interatomic potentials are constructed: from left to right, assembling a database of representative structural models, computing energies and forces using a reference quantum-mechanical method (typically, DFT), expressing the atomic structure in "machine-readable" form using descriptors, and finally regressing ("learning") the potential-energy surface. b) Different types of descriptors for atomic environments, as commonly used in empirical as well as ML interatomic potentials. c) The basic idea behind the SOAP representation,[29] in which the neighbor density of a given atom is expressed through Gaussian functions centered on atoms, controlled by the smoothness parameter, $\sigma_{at}$, and up to a specified cutoff, $r_{cut}$. Adapted with permission under the terms of the Creative Commons Attribution License.[30] Copyright 2019, the Authors. Published by Wiley-VCH. d) Sketch of a neural-network architecture for fitting interatomic potential models, where more information may be found. The atomic environment is represented by input functions, $G$, and the result is an atomic energy, $\varepsilon$. Adapted with permission under the terms of the Creative Commons Attribution License.[31] Copyright 2017, the Authors. Published by Wiley-VCH. e) Schematic of kernel methods to interpolate atomic properties (here, energies) by comparing an environment (red) with $N$ entries in the reference database (green), after ref. [32]. The approach is based on a similarity or kernel function $k$ that returns, for pairs of environments, a value between 0 (entirely dissimilar) and 1 (identical up to symmetry operations).

of acquiring them by quantum-mechanical computations are important aspects in the generation of any ML potential.

We emphasize that this database must contain higher-up regions of the PES, not only the local minima (e.g., experimentally known crystal structures). One example is to take structural snapshots from the liquid phase of a material, as sketched in Figure 1a. The aim in generating fitting databases is to achieve sufficient sampling with a finite number of structures, and one assumes that the smoothness of the PES will make it possible to interpolate between them—an assumption that is validated by experience. For example, an ML potential fitted only to a database of liquid and amorphous carbon configurations (hence containing no prior knowledge of crystal structures, but of course some local resemblance of "tetrahedral" carbon) was shown to be suitable for crystal-structure searching in the spirit of the Ab Initio Random Structure Searching (AIRSS)[33] technique, identifying a large number of known and new hypothetical carbon allotropes.[34]

Initially, reference databases for ML potentials have been constructed manually, which is still widely done today. As a starting point, one will extract the relevant crystal structures from data sources such as the Inorganic Crystal Structure Database (ICSD).[35] Rather than fully optimized cells, one will add structures in which the atoms are slightly distorted (which induces forces on atoms, and these forces provide valuable atomically resolved information about the PES). Furthermore, structural models for point defects, surfaces, etc. are constructed, as one would do in DFT-based modeling,[36] and again a sufficient number of these configurations must be added to the reference database. Of course, the point is not just to reproduce what would be possible with DFT anyhow, but to exploit the "near-sightedness" of atomic and electronic structure[37] to learn from small systems and then to access larger length scales. For example, concerning crystal surfaces, the important Si (111) surface forms intriguingly complex and stable $(2n+1) \times (2n+1)$ type reconstructions (of which the $7 \times 7$ reconstruction has the lowest energy), involving adatoms, stacking faults and rings of 10 atoms.[38] These complex surface structures can be approximated by a set of smaller motifs (such as the adatoms), and these in turn can be included in the reference database. Consequently, it was shown that the energetic preference for the $7 \times 7$ reconstruction can be captured by an ML potential, whereas state-of-the-art empirical potentials fail at the task (all of them incorrectly predicting the unreconstructed surface structure to be the most favorable).[39]

For liquid and amorphous structures in particular, iterative schemes have been used, in which an initial (coarse) potential is fitted to a limited DFT-MD dataset, and then simulations with this initial potential are run to create a more diverse set of structures. This way, the ML potential "steers" itself into regions of configuration space that need to be further explored, and large amounts of structures can be quickly assembled, much more quickly than would be possible with DFT-MD. This iterative training approach was suggested early on already, in the creation of the first ML potentials for silicon,[40] sodium,[41] and for the phase-change material GeTe,[42] the latter of which we will discuss in Section 4.1. Similar strategies were later used for amorphous carbon,[43] where hundreds of small structural snapshots could be generated in parallel runs using a computationally cheap interim potential, and the end points of these trajectories were evaluated with DFT and added to the database.[43]

A more pressing challenge concerns the sampling of general PESs, which is needed to ensure wide applicability of ML potentials—combining robustness and flexibility in the high-energy regions with sufficient accuracy in the low-energy regions (that is, for the stable and metastable crystal structures). A recently proposed strategy is to explore the PES using global searches[44–47] that are normally done in DFT-based crystal-structure prediction.[5] This idea can be taken a step further, by exploring and fitting structural space on the fly (and running all structural optimizations with interim potentials, requiring DFT only for single-point input data), thereby unifying ideas from ML potential fitting and crystal-structure searching.[46,48–50] Such approaches hold the long-term promise of discovering new materials on larger length scales than would be accessible to current state-of-the-art methods.[5,49]

### 2.2. Ingredient 2: Representations for Atomic Structure

With the reference data available, the second task is to describe the atomic structure in a suitable form as an input for the fit. In the ML jargon, the mathematical objects which encode the atomic structure (or, more generally, any information about the system) are referred to as "descriptors" (or "features"). Descriptors for local environments, centered on an atom and encoding information about its neighbors, range from simple two- or three-body terms all the way to complex "many-body" formalisms (Figure 1b). All these descriptors must fulfill various symmetry requirements, namely, permutational (regarding exchange of two atoms of the same kind), translational, and rotational invariance. For a descriptor to be usable in practice, it is normally confined to a local environment of the atom, up to a given cutoff radius (typically, 5 or 6 Å).

Albeit the choices of representation and regressor are not necessarily tied to each other, the development of (currently) widely used representations for ML potentials has been motivated by a specific fitting framework. Behler's atom-centered symmetry functions (ASCFs) were introduced in 2011,[51] and today they are popular choices for NN potentials. For kernel-based methods, these atomic descriptors are used to construct similarity measures between atomic environments, which are then employed for the regression task. A many-body descriptor initially developed for use with GAPs (see below) is the so-called "Smooth Overlap of Atomic Positions" (SOAP).[29] Within SOAP, Gaussian functions are placed on each atom inside the cutoff sphere (Figure 1c), the smoothness being controlled by the width of the Gaussian, $\sigma_{at}$. Two atomic environments can be compared by integration over positions and rotations; in practice, this is computed efficiently by expanding the density in a basis of radial functions and spherical harmonics, resembling how atomic orbitals are represented in quantum chemistry. SOAP can directly compare atomic environments with different and arbitrary numbers of neighbors, returning a value between zero (totally dissimilar) and unity (identical up to symmetry operations); it is invariant under all physical symmetries and has continuous values and derivatives.

Recently, the works of Ceriotti and co-workers[52] and Drautz[53] revealed that SOAP and many other representations,[54,55] including Behler's symmetry functions,[51] are in fact closely related and can be derived as projections of the atomic neighbor density onto variously chosen basis functions (Figure 1c). Work in the community is ongoing, and we expect that new and improved basis functions will emerge in the near future.

The relative merits of descriptors in terms of accuracy, speed, and the trade-off between the two, are currently being investigated;[56–58] the emerging conclusion seems to be that several many-body formalisms can be similarly suitable for practical use in ML potentials. There are many recent developments aside from the examples discussed above,[59,60] and we may refer the interested reader to a publicly available library of different descriptors.[57] We also stress that some approaches (such as the Coulomb matrix:[61] a scaled, pairwise table of inverse atomic distances) provide a global, rather than atom-centered representation of a system, which is more useful for gas-phase molecules than for solids.

### 2.3. Ingredient 3: Regression Tools and Implementations

Once the reference data and atomic representations are in place, the final step is to carry out the fit itself. In the taxonomy of ML approaches, this is a "supervised learning" problem, because the input data (structures) are labeled (have reference energies); more specifically, it represents a regression task, because a continuous range of output values (energies) is sought. Approaches currently employed in the field can be roughly divided into three categories, namely, artificial neural networks (NNs), kernel-based methods, and linear regression.

The description of potential-energy surfaces by artificial NNs builds on ideas from the 1990s[62] and 2000s[63–65] when such networks were constructed for specific problems with fixed numbers of atoms in a system (e.g., an $H_2$ molecule on a surface[63]). Over the last few years, NNs have gained enormous interest—driven by the idea to "learn" local, atomically short-sighted energies, which are then summed up to yield the total energy.[40] This approach makes it possible to access (much) larger system sizes than those encountered at the fitting stage, and it can make the resulting potentials more general, less dependent on which particular problem is to be studied. This idea of locality forms the basis of ML potential development as of today.

NNs are a set of mathematical functions that aim at resembling the functionality of neurons in the brain. Neurons, referred to as "nodes" in the context of a NN, accumulate and weigh input signals and "fire" (giving an output) once a certain threshold value is reached. This is achieved by a nonlinear mathematical activation function; normally, several consecutive "hidden" layers (functions) are coupled, in which case one speaks of a "deep" NN. A simple example for such an architecture[31] is sketched in Figure 1d: the atomic structure is encoded by descriptors that form the "input layer," from which information is passed to (in this example) two hidden layers and finally to the output; the hidden layers thereby carry no physical meaning on their own. To generate an NN potential, one will define the architecture (the number and shape of the nodes, the number of layers, etc.) and then "train" the model by optimizing the weights, tuning them such that the network's error on a set of known data is minimized. A more detailed introduction to NN-based ML potentials and the associated methodology may be found in refs. [31] and [66].

An alternative approach is given by kernel methods. In these, an atomic property is interpolated as a linear combination of kernel functions, the latter measuring how similar a new configuration's descriptor is to those of the reference data. The property is typically a local energy, or a force acting on an atom, and the kernels can be understood as similarity measures (on a scale from zero to one) between the new environment and those contained in the database, both of which are represented by the descriptor. The regression coefficients that weigh each kernel basis function are computed during the fitting using simple linear algebra. Kernel ridge regression (KRR) and Gaussian Process Regression (GPR) are two currently employed techniques, differing only slightly in how these coefficients are computed. These methods rely on a carefully designed kernel function that is appropriate for the problem (such as the SOAP kernel in the present case), whereas NNs are more general—but often need more data to balance out their extra flexibility.

With the field evolving so rapidly, it is difficult to provide an accurate snapshot of its development. We collect in **Table 1** several approaches that combine a theoretical framework and a practical implementation. Beyond the two major categories discussed herein (NNs and kernel methods), this notably includes emerging methods based on linear fitting, such as the SNAP[55] and Moment Tensor Potential[54] approaches; a complete review of all advances in the field is not possible within the scope of this Progress Report, but we stress that these methods are under active development as well. All methods referenced in Table 1 are primarily focused on condensed phases, as opposed to molecules (for which increasingly powerful ML potentials are being developed just as well[77–79]), and they are implemented in freely available simulation software, such that they can be directly used by anyone.

We conclude this methodological overview by stressing that all methods discussed here lead to "offline-trained" ML potentials: one generates (and extensively validates) a suitable potential for a specific material, and then applies it without further modification. We also mention on-the-fly or "online-trained" ML potentials:[80–82] these start with a quantum-mechanical (again, normally DFT-based) MD simulation and train an ML potential while the simulation is being run. Numerical measures are put in place to monitor the degree of extrapolation (that is: do we need additional data at this point?), and after a while, DFT evaluations are only needed when the potential runs outside its "comfort zone." The resulting potentials will be specific to the problem being studied:, e.g., an on-the-fly fitted ML potential for $TiO_2$ will not be expected (or required) to describe all the different crystal structures of elemental Ti correctly. In turn, such techniques promise an even much broader availability of ML-driven atomistic modeling in materials research. As a direct consequence, it becomes more and more important for the wider community to be aware of the nature, benefits, and limitations of such potentials.

**Table 1.** Selected examples of current methods and available implementations for ML-based interatomic potentials that aim primarily at condensed-phase materials modeling. We do not add web addresses, as these may be subject to change, but they are easily retrieved using a web search. The fitting methods ("regressors") include neural networks (NN), kernel methods such as Gaussian process regression (GPR) and kernel ridge regression (KRR), and linear fits.

| Method | Refs. | | Regressor | Implementation |
|---|---|---|---|---|
| Artificial neural networks (Behler) | [40] | (2007) | NN | Standalone ("RuNNer"); LAMMPS interface[67] |
| Gaussian approximation potentials (Bartók and Csányi) | [68] | (2010) | GPR | GAP code (custom); LAMMPS interface |
| Spectral neighbor analysis potential (SNAP) (Thompson) | [55,69] | (2015) | Linear fit | LAMMPS interface |
| Adaptive, generalizable, and neighborhood informed (AGNI) force fields (Ramprasad) | [70–72] | (2015) | KRR | LAMMPS interface |
| aenet (Artrith) | [73] | (2016) | NN | Standalone ("aenet") |
| Amp (Korshidi and Peterson) | [74] | (2016) | NN | Standalone ("amp"); LAMMPS interface |
| Moment tensor potentials (Shapeev) | [54,75] | (2016) | Linear fit | LAMMPS interface |
| DeePMD (E) | [76] | (2018) | NN | Standalone ("DeePMD-kit"); LAMMPS interface |

## 3. High Accuracy for Crystalline and Amorphous Materials: The Case of Silicon

We begin by illustrating the accuracy that ML potentials can reach in atomistic simulations for materials science. As a case in point, we choose elemental silicon, which has traditionally been a prominent test system for materials modeling. In terms of empirical interatomic potentials, the foundational models by Stillinger and Weber[83] and Tersoff[84] were introduced for silicon, as was, a decade later, the environment-dependent interatomic potential (EDIP) family.[85] Revised and further improved empirical silicon potentials are continuing to be reported, such as one very recently optimized for the graphene-analogous 2D structure of "silicene."[86] On the DFT side, having been able to describe the aforementioned silicon (111) surface was a major success at the time.[87] Far beyond the computational relevance and the deep academic interest, e.g., in the nature of its amorphous (noncrystalline) phase,[88] silicon is a centrally important material for applications: in crystalline form, it currently forms the backbone of the electronics industry; in amorphous form, it is used in solar cells,[89] thin-film transistors,[90] or as an anode material for batteries.[91]

Silicon has been the verification and application case of choice for ML potentials as well, since the first high-dimensional NN potential was developed in 2007 by Behler and Parrinello: in their initial work, the neighbor statistics in the liquid phase were computed, and shown to outperform various high-quality empirical potentials.[40] Directly following, in 2008, Behler et al. applied their NN potential to the phase transitions in silicon under pressure,[92,93] using a technique known as metadynamics to jump over barriers in the PES (from one local minimum to the other).[94] In these simulations, the *Cmca* phase of silicon had not been included in the fit but was readily discovered by the ML-potential-driven metadynamics simulations, which already points toward the applicability for crystal-structure searching. Further early developments in the field were tested and benchmarked for silicon: Bartók et al. computed elastic properties of diamond-type silicon (as well as other semiconductors) with the first GAPs in 2010,[68] and Li et al. used the material to test their on-the-fly ML scheme in 2015.[80]

Subsequently, silicon served to explore the question as to whether a "general-purpose" ML potential can be created for a given element.[95] Such a general potential should be able to accurately describe a very wide range of atomic configurations, including multiple phases, surfaces, and defects, for which the training database is explicitly "designed," but at the same time give sensible results for completely new kinds of configurations (e.g., structures found during systematic random structure search, or more complicated structural defects that were not included in the training but may occur in experiment). Following this logic, a systematic database of reference configurations was constructed for silicon (**Figure 2**a).[95]

The first necessary test for the resulting potential is the performance in regard to the "usual" quality indicators, which includes standard material properties that are relevant for practical applications (such as the bulk and shear modulus). Running these tests and comparing their outcome to the arsenal of high-quality empirical potentials available (Figure 2b) allows one to quickly appreciate the usefulness of the method. It should be noted, as always when considering ML potentials, that these are normally orders of magnitude slower than empirical potentials (yet orders of magnitude faster than DFT).

Although such validation against DFT-computed properties is an important step, the most important application for ML potentials is in those simulation scenarios that are completely out of reach for DFT. A typical example is the formation and propagation of cracks (Figure 2c), a microscopic effect that has tremendous importance for the macroscopic behavior of materials. In crack simulations, a large piece of material is "cleaved" across a certain crystallographic direction by applying strain on either side, and one monitors whether a pre-formed crack tip propagates (brittle failure) or blunts (plastic deformation). Because of the ongoing breaking and making of bonds during fracture, this is a very challenging problem for atomistic simulations. Indeed, empirical interatomic potentials may have trouble correctly describing these mechanisms even qualitatively, and it was shown how suitable corrections such as an environment-dependent cutoff can help to restore the correct materials behavior.[99] The ML potential described in ref. [95] also correctly captures this qualitative behavior—but what is more, it recovers even subtle bond-rotation mechanisms that have previously been only seen in extremely expensive QM/MM simulations (mixing a quantum-mechanical method with a force field), using DFT at the crack tip.[100] Figure 2c also illustrates an intrinsic feature of GPR, the ML method used for this particular potential: the "predicted error" measures how far the potential is from configurations that it has encountered during training (indicating the degree of extrapolation), which here arrives at a few meV per atom, given by color coding in Figure 2c.[95] We note that tests for extrapolation are also done for NN potentials:, e.g., by fitting several networks in parallel and determining the variance between them.[31]

Beyond the crystalline phases, their phase transitions and engineering-relevant properties, the amorphous phase of silicon is also of wide interest, and similarly a challenging case for atomistic simulations. Traditionally, DFT-based studies have reported calculations for small melt-quenched unit cells, but such simulations cannot normally reach below $\approx 10^{12}$ K s$^{-1}$ cooling rates, corresponding to roughly one million timesteps. In contrast, using an ML potential, it was possible to reduce the quench rate to $10^{11}$ K s$^{-1}$,[97] later even to $10^{10}$ K s$^{-1}$.[30] One cannot probe the local structural environments directly by imaging techniques, but one can investigate them by comparing to a set of different experimental observables. One such tool is NMR, where shifts can be computed from first principles (Figure 2d).[101]

Very recently, simulations using an NN potential were reported for the (structurally even more complex) liquid phase of silicon and the nucleation of crystals.[102] In these studies, the authors emphasize the importance of having a flexible potential that can treat the structurally very different local environments that occur during the formation and growth of crystallites in the disordered phase: seed nuclei might look very different from the corresponding bulk phases that grow out of them. In a more general sense, understanding crystal nucleation is a very important goal for chemistry and materials science.[103]

We summarize the present section in **Table 2**. We expect that the accurate modeling of silicon using ML potentials will continue to be an important benchmark for methods development, and for linking ML potentials to other (high-level) simulation techniques. For example, the melting point of Si was very

 

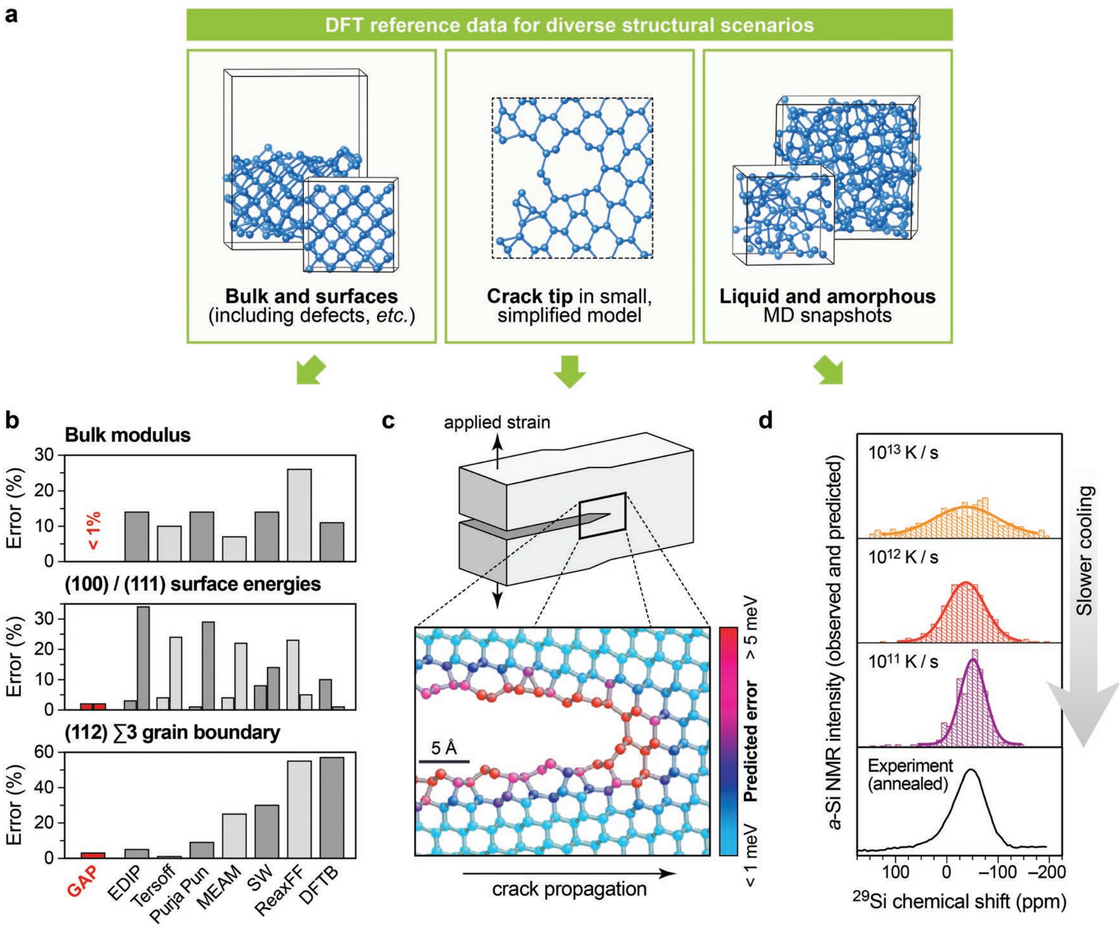


**Figure 2.** Highly accurate materials modeling using ML-based potentials, illustrated for the example of a recently developed silicon potential (see Table 2 for broader context). a) Representative structures from a large reference database, for which DFT energies have been computed.[95] Structures were visualized using VESTA.[96] b) Numerical errors for key engineering properties, comparing the ML potential (GAP; red) to several state-of-the-art empirical potentials and to density-functional tight-binding (DFTB), which are widely used to study materials on large length scales. Unsigned errors are given relative to DFT results. Data are taken from ref. [95], where a much more detailed set of benchmarks may be found. c) Crack propagation in crystalline silicon simulated by GAP. Color coding shows the prediction error, which is intrinsically obtained in the GAP framework, and allows for uncertainty quantification. Adapted with permission under the terms of the Creative Commons Attribution License.[95] Copyright 2018, the Authors. Published by American Physical Society. d) NMR fingerprints of amorphous silicon (*a*-Si) samples created in ML-driven melt-quench simulations with progressively slower quench rates and increasingly high structural order; Adapted with permission under the terms of the Creative Commons Attribution License.[97] Copyright 2018, the Authors. Published by American Chemical Society. Experimental NMR data are shown for comparison. Adapted with permission.[98] Copyright 1990, American Physical Society.

recently described by random phase approximation (RPA) computations, arriving within a few percent of the experimental reference.[104] Looking toward the future, it seems promising to fit to such RPA data (given that gradients and thus force data are available), rather than DFT. Once the fit is done, the cost of using the ML potential is independent of the level of electronic

**Table 2.** Selected milestones in the modeling of elemental silicon using ML potentials.

| | | |
|---|---|---|
| 2007 | First neural-network (NN) potential for bulk silicon and tests for the liquid phase | Behler and Parrinello[40] |
| 2008 | Description of high-pressure phase transitions in silicon using NN potentials and metadynamics | Behler et al.[92,93] |
| 2010 | First Gaussian approximation potential (GAP) model for diamond-type silicon and tests for elastic properties | Bartók et al.[68] |
| 2015 | Development of an on-the-fly learning scheme using the example of silicon | Li et al.[80] |
| 2017 | Correct energy ranking of the silicon (111) and (100) surface reconstructions using a GAP | Bartók et al.[39] |
| 2018 | Description of surface energies, dislocations, cracks, etc. with a general-purpose GAP | Bartók et al.[95] |
| 2018 | Description of amorphous silicon with high accuracy by GAP-driven slow quenching; validation against experimental probes (calorimetry, diffraction, $^{29}$Si NMR) | Deringer et al.[97] |
| 2019 | Description of liquid silicon and crystal nucleation from the liquid, using a DeePMD potential | Bonati and Parrinello[102] |

 

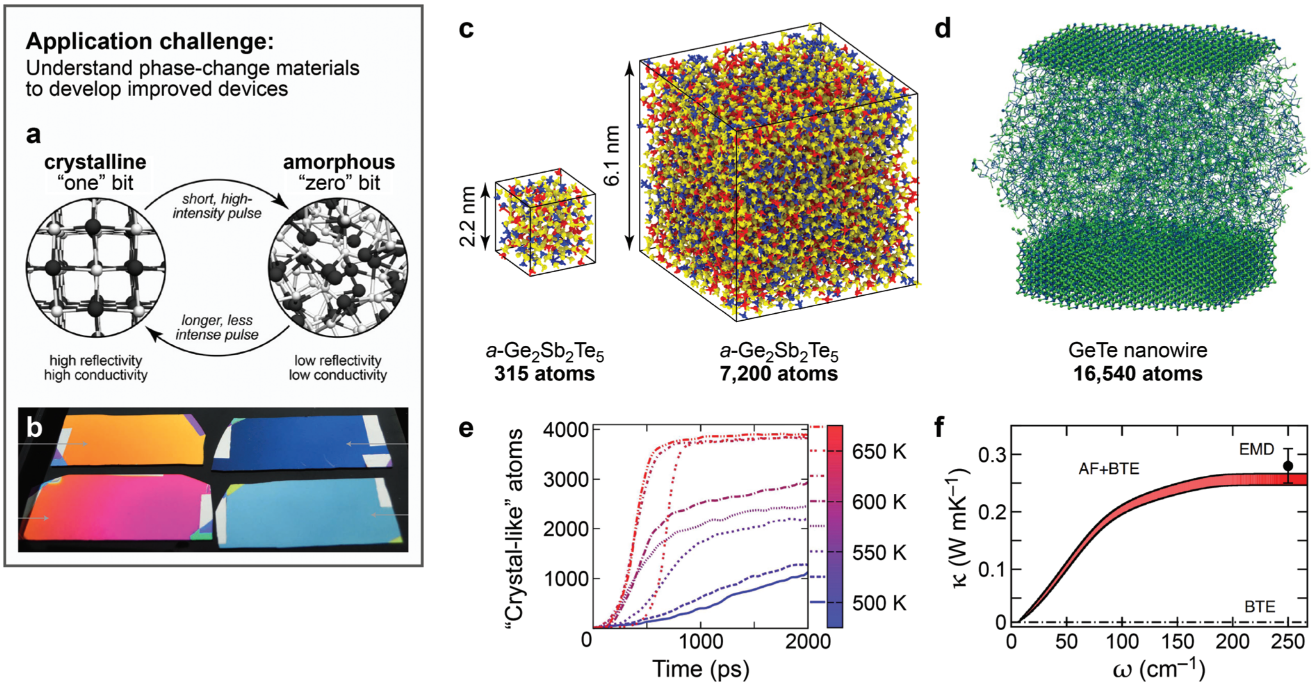


**Figure 3.** Phase-change materials (PCMs) for data storage, photonics, and thermoelectrics, and their atomistic modeling using ML-based interatomic potentials. a) The functional principle of PCMs, which are switched between a crystalline and an amorphous phase with notably different physical properties;[105] Adapted with permission.[118] Copyright 2015, Wiley-VCH. b) An example application in switchable, flexible displays, where crystalline (top) and amorphous (bottom) layers lead to a change in colors due to their different optical properties.[111] Reproduced with permission.[111] Copyright 2014, Springer Nature. c) An illustration of the relevant system sizes, here for the ternary PCM $Ge_2Sb_2Te_5$, emphasizing the capabilities of ML potentials.[129] The small cell is one that is accessible to DFT-MD; the larger is one that is accessible to ML potentials; both are drawn to scale. Reproduced with permission.[129] Copyright 2018, American Chemical Society. d) Simulation of a GeTe nanowire, reaching device-size models: in this case, the center of the hexagonal wire has been liquefied by heating whereas the bottom part was kept fixed in the crystalline structure during the simulation.[124] Reproduced with permission.[124] Copyright 2017, American Chemical Society. e) Crystallization simulations of GeTe, starting from a fully amorphous phase at various temperatures, as described in nanosecond-long MD simulations by an NN-type potential.[130] Adapted with permission.[130] Copyright 2013, American Chemical Society. f) Thermal conductivity in GeTe;[131] this property can now be simulated using ML potentials.[28] Reproduced with permission.[131] Copyright 2012, American Physical Society.

structure method that was used to produce the training data. This promises access to the modeling of liquid, amorphous, and nanocrystalline materials at computational levels of quality that are currently completely out of reach for such systems.

## 4. Applications to Materials-Science Problems

The long-term goal for ML potentials must be to solve challenging problems, in a way akin to DFT and empirical potentials, but addressing new challenges that neither of the established approaches can handle. The focus, thereby, shifts away from the numerically accurate description of energies and forces—which is very much needed, of course, but which does not solve an application problem on its own. Instead, one also needs flexibility and transferability, to be able to explore previously unknown scenarios and structures: this is not a trivial task for ML potentials, because they lack inherent physically informed functional forms. Truly predictive potentials will need to describe diverse atomic environments as they may occur in manufacturing and practical applications of materials: during melting and solidification, mechanical shearing, or deposition and growth of thin films, to name just a few examples. We highlight here three selected areas of current research interest where applications of ML potentials to materials-science problems have already become reality.

### 4.1. Structure, Dynamics, and Function in Phase-Change Materials

Phase-change materials (PCMs) for data-storage applications are a group of chalcogenides, most prominently Ge—Sb—Te alloys, that can be rapidly and reversibly switched between crystalline and amorphous phases with a peculiar contrast in reflectivity and electrical resistivity (**Figure 3**a).[105] This property contrast can be exploited to encode digital "ones" and "zeroes" in memory devices.[105] The need for new data-storage and data-processing technologies is obvious, nowadays more than ever: ongoing efforts to optimize the switching mechanisms are bringing sub-nanosecond memory operations within reach,[106–108] and even the use of PCMs in brain-inspired computing has been proposed.[109] Extending their range of applications further, PCMs are also rapidly garnering interest as candidate materials for photonic devices such as switchable displays (Figure 3b)[110–112] or thermoelectrics for waste heat recovery (e.g., by alloying GeTe with guest atoms to modify the atomic and electronic structure).[113–115]

While the structures of crystalline PCMs have long been known and can be accurately studied using diffraction experiments,[116–118] a similar understanding of the structures and properties of the amorphous phases has remained a major challenge. DFT-MD simulations are therefore one of the primary approaches for obtaining insight into amorphous PCMs.[119,120] The challenge is exacerbated by the fact that the amorphous phases are structurally very complex—with coexisting tetrahedral and octahedral-like fragments,[121] and with a substantial number of homopolar Ge—Ge bonds[120] that are relevant to aging (resistivity changes over time) in amorphous PCMs.[122–124] This large structural diversity makes it very difficult to create empirical potential models for PCMs.

The relevant system sizes are illustrated in Figure 3c,d. Typical DFT-MD simulations in this area encompass on the order of 300–400 atoms,[107,125,126] having very recently reached up to 900 atoms in one instance,[127] but this is only possible with fast supercomputers and sophisticated algorithms.[128] In contrast, an ML potential readily makes system sizes of many thousand atoms accessible, as we show for a structural model of bulk amorphous (*a*-) $Ge_2Sb_2Te_5$, containing 7200 atoms (Figure 3c),[129] and of a partially melted GeTe nanowire, described by an even more complex structural model with over 16 000 atoms (Figure 3d).[124]

An NN potential for GeTe was introduced by Sosso et al. in 2012 already.[42] The first and most central application was in modeling crystallization (the atomistic process leading from a "zero" to a "one" in digital memories).[130] MD simulations, initially carried out in a system containing 4096 atoms, showed how at low temperature a single nucleation event induces crystal growth, whereas at high temperature multiple nuclei emerge. Figure 3e illustrates how at a temperature of 650 K, the number of "crystal-like" atoms (as measured by a suitable structural order parameter) quickly approaches the number of atoms in the system. The process is completed in less than a nanosecond at high temperature (red) but takes much longer at lower temperature (blue).

GeTe has also served as an example for the prediction of thermal properties (Figure 3f): the thermal conductivity, $\kappa$, was computed using the above-mentioned NN potential.[131] In that work, the thermal conductivity has been studied by different approaches, namely, Allen–Feldman plus Boltzmann transport equation (AF+BTE) and also by equilibrium molecular dynamics (EMD). More details of the methodology and the specific technical questions pertaining to GeTe have been reviewed elsewhere.[28]

Open questions in applying future generations of ML potentials to PCMs fall into two categories, defined by the chemical composition. On the one hand, ongoing work is concerned with the Ge–Sb–Te system, where further optimizations of scalability and switching speed are needed, and the recent development of a memory device based only on Sb ("monatomic" phase-change memory)[132] attests to the continued relevance of this material system. The nature of defect states in amorphous $Ge_2Sb_2Te_5$ has been studied very recently using a large number of structural models generated in ML-driven simulations.[133] Besides the bulk phases (Figure 3c) and nanowires (Figure 3d) discussed in this Progress Report, it would be highly valuable to understand more thoroughly the atomistic structure and behavior of superlattice structures,[134] continuing along the lines of existing DFT studies[135] but now on larger length scales. We expect that further developments in ML potentials for the ternary Ge–Sb–Te system will pay good dividends in the long run. On the other hand, active research in the PCM field focuses on the systematic selection of guest atoms and dopants,[136] as recently demonstrated for Sc-alloyed $Sb_2Te_3$.[107] These systems will require careful optimization of ML potentials due to the (even) higher chemical and structural complexity involved. However, once a successful "recipe" has been established, it could be applied to various types of dopants (e.g., different transition metals), and even to other material classes.

### 4.2. Nanoparticles for Catalysis

Nanoparticles (NPs) are another vast research field where knowledge of the atomic-scale structure is challenging but mandatory. Chemical synthesis techniques allow for an increasing degree of control over the shapes and sizes of NPs, and this has been going hand in hand with increased computational capacity.[137] The surface composition of crystalline materials as a function of their chemical environments[138] and reaction mechanisms on pristine and defective crystal surfaces have been widely studied with DFT. On the other hand, nanoparticle materials are often structurally different from the corresponding bulk phases, and their modeling and understanding has remained an open challenge—with which ML potentials promise to help.

**Figure 4**a introduces the problem by way of an example: a TEM image with sufficiently high resolution to allow for discerning individual atoms.[139] It is evident that the center of the particle exhibits crystalline order whereas the outer region is strongly disordered. The disordered outer region is very difficult to fully quantify with imaging and other structure-determination methods. Furthermore, it is not enough to study one single particle but one needs an entire ensemble of those.

Nanoparticles are used as catalysts, and therefore understanding their structure is only the first step; one then needs to model the behavior of guest atoms and adsorbates as well. We stress, in this context, that very early work in the field (before high-dimensional ML potentials with finite cutoff spheres were introduced) already dealt with molecules on surfaces (in that case, with crystalline order and a fixed number of atoms).[63–65] Two examples that have recently been explored with high-dimensional ML potentials are the CO oxidation on mixed Cu/$CeO_2$ NPs (Figure 4b)[139] and the $CO_2$ reduction to give useful feedstock molecules which is performed, for example, on mixed gold/copper metal NPs (Figure 4c).[140]

Surface energies as computed from DFT can be directly translated into a so-called Wulff construction,[141] giving information about the equilibrium shape of the NP. Figure 4d shows that ML potentials can reproduce these predicted shapes with good accuracy. The major challenge, however, is to describe NPs beyond their idealized surfaces, and beyond what is accessible to DFT-driven simulations.

ML potentials can be combined with any suitable simulation tool that requires the evaluation of energies as a function of structure: the majority of this Progress Report focuses on

 

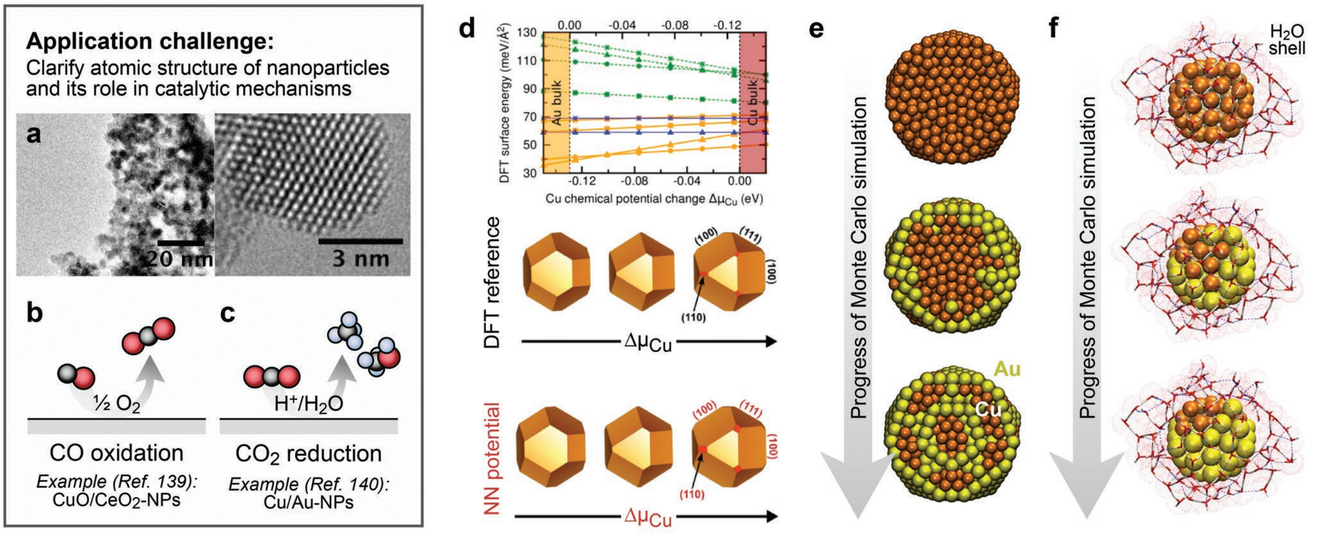


**Figure 4.** Nanoparticles and their reactivity as a key application perspective for ML potentials. a) Example TEM images of as-synthesized nanoparticles: the atomic positions can be visualized with modern techniques, but which species are located where (and is there a preferred surface species, which would be relevant for catalysis)? Reproduced with permission.[139] Copyright 2016, American Chemical Society. b,c) Two important reactions in nanoparticle catalysis, both of which have been exemplarily studied using ML potentials;[139,140] schematics after ref. [139]. d) The Wulff construction, translating a computed surface phase diagram (top) into equilibrium shapes that minimize the overall surface energy (middle, DFT; bottom, NN potential; for a general overview, see ref. [137]). e) Monte Carlo simulations on Au/Cu nanoparticles, driven by an NN potential.[140] Exemplary snapshots are shown that illustrate how the "onion"-like structure gradually emerges. f) Same for an Au/Cu nanoparticle where the ML potential includes four species, adding also H and O and, therefore, a description of explicit water molecules around the particle. Snapshots have been taken from the Supporting Information of ref. [140], where entire videos are provided. Panels d–f) adapted with permission.[140] Copyright 2014, American Chemical Society.

MD simulations, but longer length and time scales are accessible to other approaches, such as Monte Carlo simulations. Such a combined approach has been recently used for Cu/Au NPs.[140] Figure 4e shows how as the simulation progresses, the "onion-like" structure of the NP is recovered. This analysis is not restricted to the pristine (gas-phase) particles, but it may include other species such as water molecules, which are important when the particle is in solution (Figure 4f). Future work may now deal with more complex entities, such as ligands coming from a chemical synthesis route. ML potentials for modeling these molecules are certainly available.[77,78]

Other, recent work in this area includes the structural modeling of NPs on support surfaces, using various flavors of ML potentials. Kolsbjerg et al. studied NPs on a surface and used the resulting structural models to further study the adsorption of CO molecules with DFT methods.[47] Furthermore, a "multitribe" evolutionary search was conducted where information from a given particle size is transferred to a larger or smaller size intermittently during the iterations.[142] Both studies exemplify the usefulness of fast and flexible ML potentials being combined with global exploration. Such methods could, in the future, assist with the large-scale screening for NPs with desirable properties and the targeted synthesis in the laboratory.

### 4.3. Carbon Nanomaterials

Carbon is among the structurally most diverse elements, which goes back to its unique ability to form twofold ("sp"), threefold ("sp$^2$"), and fourfold ("sp$^3$") coordinated atomic environments. This structural diversity is reflected in many known and hypothetical crystalline structures,[34,143] metastable over an unusually wide stability window,[144] and in a plethora of disordered phases with very diverse properties.[145,146] Between the ideally ordered structures of diamond and graphite on the one hand, and fully disordered amorphous structures on the other hand (**Figure 5**a), there are partly ordered, "porous" and "hard" carbon materials,[147–150] a few examples of which are shown in Figure 5b. These materials are of utmost importance for practical applications, for example, for energy storage and conversion.[151]

From the viewpoint of simulations, carbon materials have been widely described using empirical potentials.[152–154] They have also been of interest very early on for ML-type potentials: an initial NN-type model was developed for the diamond-graphite coexistence line[155] and applied to simulate the mechanisms by which diamond forms from graphite under compression.[156] Other phases that might be synthesized from cold compression of graphite are of interest as ultrahard materials,[157] and could be explored in the future using such ML potentials.

We will focus here on more disordered, amorphous carbons, which encompass a wide field with different structural and chemical properties, e.g., "tetrahedral" amorphous (*ta*-C)[158] and porous carbons.[147–150] In 2017, a GAP model for carbon was introduced and validated for properties that include DFT-predicted surface energies and experimentally measured elastic moduli.[43] Shortly after, Caro et al. performed large-scale deposition simulations using this potential—which, for the first time, reproduced the 90% sp$^3$ content in dense *ta*-C films and clarified the microscopic growth mechanism of the material, which had been under debate for three decades (Figure 5c).[159] *ta*-C has important applications as protective coating for friction management. More recently, both high- and low-density forms

 

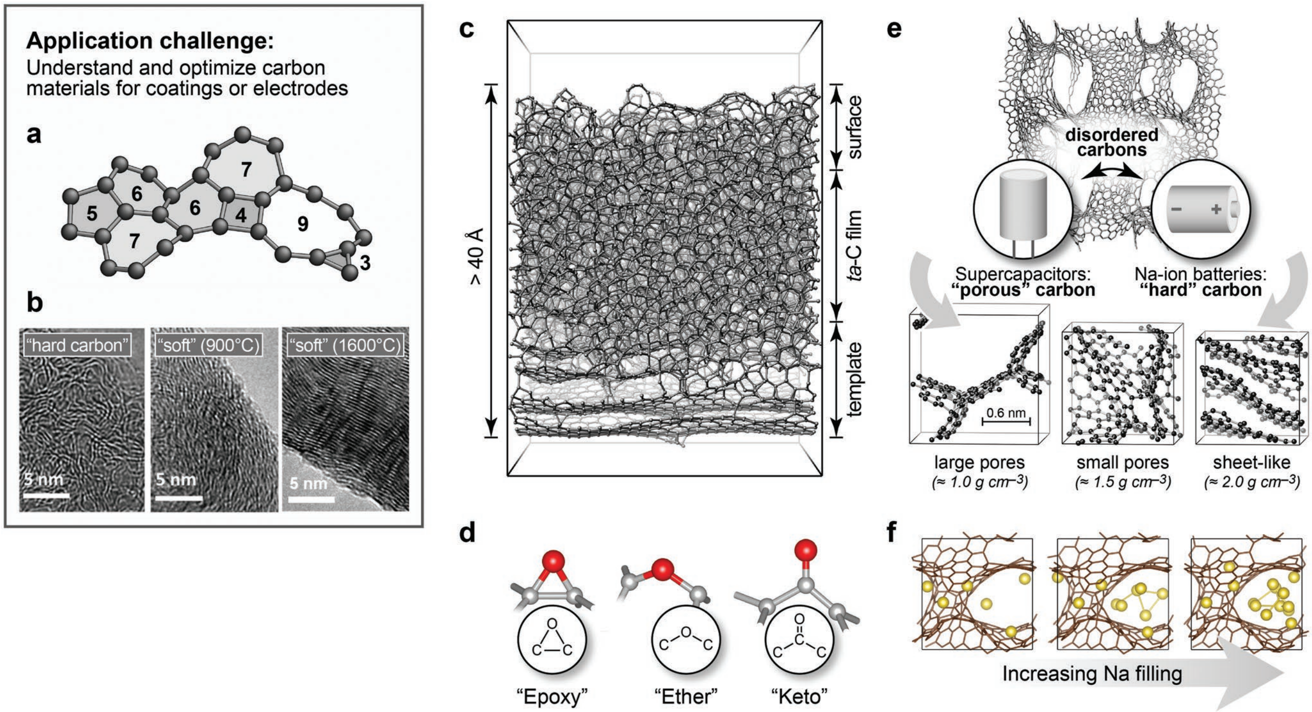


**Figure 5.** A Gaussian approximation potential (GAP) for carbon nanomaterials and example applications. a) Fragment from a GAP-generated structural model of *a*-C, emphasizing the coexistence of various ring fragments.[43] Reproduced with permission.[43] Copyright 2017, American Physical Society. b) TEM micrographs of three different carbon materials as used in electrochemistry, differing in their structural order ("graphite-like-ness") which increases from left to right.[150] Adapted with permission.[150] Copyright 2017, American Chemical Society. c) Slab model of *ta*-C generated in explicit growth simulations using the carbon GAP, modeling the impact of one atom at a time. Structural drawing created with data from ref. [159] d) The most relevant surface motifs as found in DFT-MD simulations based on a library of GAP-generated surface models.[160] Adapted with permission.[160] Copyright 2018, American Chemical Society. e) Structural models of disordered, sp$^2$-rich carbon materials, generated in GAP-MD annealing simulations from amorphous precursors. f) Intercalation of Na atoms, modeling the atomistic processes pertinent to Na-ion batteries. Panels e) and f) adapted under the terms and conditions of the Creative Commons Attribution 3.0 Unported Licence.[163] Copyright 2018, The Authors, published by The Royal Society of Chemistry.

of *a*-C have begun to be employed as electrode materials in electrochemistry, in particular, in the context of in vivo analytical electrochemistry, owing to its superior biocompatibility.[146] The surface chemistry of *a*-C is complex, due in part to the disordered nature of the material but also due to the many different types of O-, H-, and N-containing functional groups that can become attached to the surface, either during the manufacturing process or when the materials are put in contact with an electrolyte.[146]

Enabled by the computational efficiency of ML-based interatomic potentials, the large variety of structural motifs and (surface) functional groups that are present in *a*-C materials can be explored with a level of detail that is out of reach for simulations using DFT.[160,161] These computational atomistic models can then be compared to experimental data, such as X-ray or Raman spectroscopy, to elucidate the surface chemistry of the materials and ultimately make the connection between atomic structure and application performance.

The range of applicability of ML-based potentials to study interfacial chemistry is of course not limited to carbon materials. We expect that, in the near future, these studies will be extended to understand the surface chemistry of other materials used in electrocatalysis and electrochemistry. A bit further down the line, we also expect that, as accurate ML-based models of water become available,[162] we can gain an unprecedented level of understanding about the structure of aqueous/solid interfaces, where the effects of solvation are crucial in determining reaction mechanisms and free energy barriers for chemical reactions. An important application of ML potentials in this regard will be the systematic computational discovery of new efficient catalysts for $CO_2$ reduction and $H_2$ production.

With regard to applications in battery materials, disordered and "porous" carbons have been studied using the aforementioned carbon potential,[163] following ideas initially pursued with empirical interatomic potentials.[153,164] Building on a computational description of pure disordered carbon, it was also shown how the insertion of Li ions in carbon can be described by a difference potential.[165] In this context, it is interesting to mention two separate studies that develop NN models for Li in amorphous silicon anodes:[166,167] these works provide a thematic link to the wide importance of silicon (Section 3) and underline the possibilities of ML potentials especially for energy materials modeling.

## 5. Challenges and Future Directions

The usefulness of ML-based interatomic potentials has now been firmly established, through examples shown in this

 

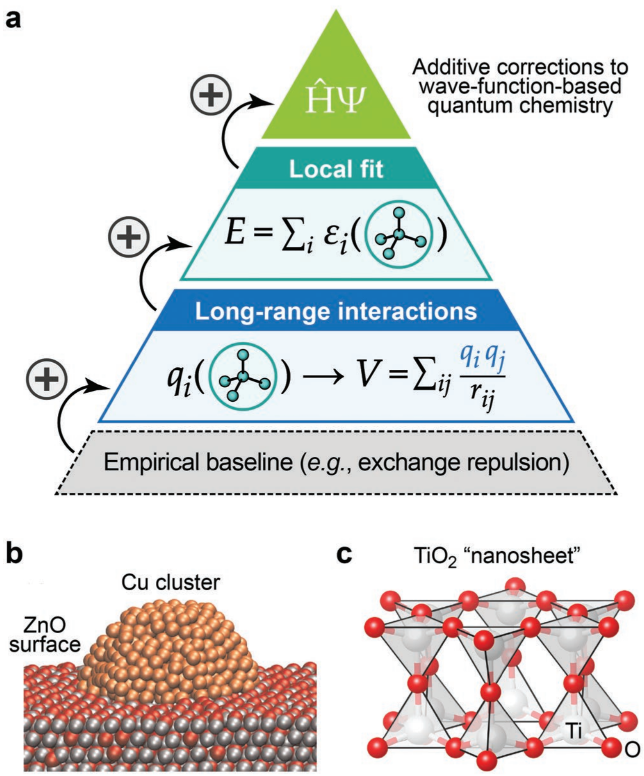


**Figure 6.** Hierarchical ML models for interatomic potentials as one perspective for possible developments in the coming years. a) Schematic of a hierarchical approach, starting with an empirical baseline (e.g., the exchange repulsion between atoms that come very close to one another), to which "machine-learned" terms are then gradually added. b) Example of a simulation carried out with an NN model that combines short-range bonding and long-range electrostatics, fitted with separate NNs and then summed up, applied to a metallic Cu cluster at an ionic ZnO surface.[171] Reproduced with permission.[171] Copyright 2013, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. c) A $TiO_2$ nanosheet, identified in structure searches accelerated by a charge-equilibration model, relying directly on an NN representation of atomic charges.[172] Structural model drawn with data from ref. [172].

Progress Report and in the rapidly evolving body of literature in the field, combining new methodology with practical research questions. We discuss here selected challenges, leading to a perspective for a hierarchical approach to ML potential fitting (**Figure 6**a), and thereby to our vision of how we expect that ML potentials can enable a new level of realism in the modeling and understanding of materials in the future.

Current ML potentials are "fragile": if they are fitted to be highly accurate for a certain narrow range of configurations, they are not very accurate elsewhere. This is an intrinsic feature of fitting functions in high dimensions (referred to as the "curse of dimensionality"). At the same time, they may fail—even catastrophically—when taken to a region of configuration space that they have not encountered before (generalization). Sometimes, it is clear a priori where those regions are: a simple example is the strong exchange repulsion between two atoms that come very close to one another. Rather than fitting to the reference quantum-mechanical data as they are, one may start by subtracting an empirical baseline from the input data and adding it back only once the ML prediction is complete; this has been done, e.g., in the construction of the silicon potential discussed in Section 3.[95] Other baselines could include, for example, simple fixed-charge models for electrostatics[68] or even semiempirical quantum mechanical models, which brings us to the next issue.

Current ML potentials are "short-sighted": typical cutoff radii enclose the first few coordination shells, but not more. Experience has shown that this is a very successful approach for predominantly covalent and metallic materials;[40,41] however, it reaches its limits when long-range interactions become important, most centrally in ionic solids but also in those (ubiquitous) systems that are dominated by van der Waals dispersion. Fortunately, developments in the field suggest that the parameters of electrostatics models (such as partial charges, local dipoles or polarizabilities) can be fitted using ML with local descriptors.[168–170] These could now be made part of a hierarchical (and more general) ML potential fitting framework.

Again, previous work suggests this to be a promising direction: successful ML potentials for simple ionic systems have been constructed that "learn" electrostatic information as a function of atomic structure. Behler and co-workers showed how the electrostatic energy in solids can be described by environment-dependent charges that are fitted through a separate NN, later summing up the short- and long-range terms,[169,171] similar in spirit to how an Ewald summation is done. Figure 6b shows an example for such a simulation: a metallic Cu NP rests atop an ionic ZnO surface—again stressing the importance of nanostructures discussed in Section 4.2. More recently, Goedecker and co-workers proposed an environment-dependent charge equilibration scheme for ionic systems, initially introduced for NaCl clusters in the gas phase,[170] constructing a simple force field just from electrostatic terms. Emerging applications of this method deal with surface reconstructions of fluorite-structured materials[173] and the structural prediction of a stable $TiO_2$ nanosheet (Figure 6c), found in global exploration as accelerated by the ML potential.[172] The latter structure appears to be interesting due to its electronic band gap, which is much larger than that of known (3D) $TiO_2$ polymorphs—which might make the material a candidate for photocatalytic hydrogen production, given that a way can be found to synthesize it (e.g., through epitaxial growth on a suitable substrate).[172] However, a framework that would "learn" all long-range terms and hierarchically combine this with a short-range fit, thereby making it generally applicable to all kinds of solids (in particular weakly ionic ones), is lacking so far.

Current ML potentials deal with inorganic materials or molecules—but not normally with both: In terms of transferability, DFT has been successfully applied to study molecular systems as well as solids, ranging from gas-phase chemistry all the way to condensed-matter physics. On the contrary, while there are connections between the two worlds, for example in the description of bulk liquid water,[162] a truly unified ML potential framework that would at the same time treat a wide range of inorganic solids and organic molecules is not currently available. One of the challenges here is linked to the localized (delocalized) nature of electrons in molecules (periodic solids), respectively, and how this manifests in the different physical models and simplifications that can be made for them. For example, a thorough treatment of electrostatics in extended crystal structures requires complex position-dependent polarizability and screening to be described, whereas molecules can often be treated "in vacuo" to first approximation. In turn, molecules are often treated at much higher (post-DFT) levels of theory, reaching further improved accuracy. Being able to incorporate such high-level quantum-chemistry method into fitting frameworks for materials could aid the construction of ML potentials strongly: by adding corrections on top of the (DFT-fitted) long- and short-range parts, thereby contributing the final step of the hierarchical approach outlined here (top of the pyramid in Figure 6a).

Current ML potentials deal with specific compositions: Another aspect where ML potentials are still severely lacking transferability is in moving from one material system to another. At the moment, ML potentials are not transferable in alchemical space (that is, when atoms of different species are included). If one wants to study the alloying of, say, Si and C, the existing ML potentials for both elements as discussed in the previous sections cannot be used on their own. Instead, a new potential would need to be constructed that is able to handle both elements, either by extending the existing elemental potentials (and databases) to incorporate the Si–C interactions, or by training a Si–C potential completely from scratch. In this case, a more automatic recipe for database generation and potential training would be highly beneficial, since it would allow us to generate new potentials as the need for them arises, without the large amounts of expert human time that is currently needed for this task. This would significantly increase the number of applications suitable for ML potentials and their wide accessibility to the community, similar to the current widespread use of DFT across many fields. In an effort to quantify this transferability, standardized tests for both robustness and accuracy of ML potentials would be highly beneficial—for instance, in the spirit of a recent community-wide benchmark of many DFT codes that are currently in use.[8]

## 6. Conclusions

Atomic-scale simulations are routinely used to clarify and understand the structure of matter, ultimately aiming to design new and improved materials for applications. ML-driven simulation methods are currently emerging as a powerful addition to the toolbox of materials modeling, and they are quickly becoming accurate and flexible enough to be applied to real-world materials-science problems. Due to their intrinsically physically and chemically agnostic nature, they must be used with greater care than in the case of empirically fitted interatomic potentials—but there are ways to accommodate that, such as uncertainty quantification. We see the future of ML-driven materials modeling in two ways: on the one hand, in routine applications where the potential may be generated "on the fly" or in an otherwise automated fashion, and on the other hand, in general-purpose potentials that take longer to develop but can then be used off-the-shelf, on their own, for a wide range of materials problems. We have presented a few applications, and we are excited to see what will follow in the years to come.

## Acknowledgements

The authors thank their colleagues and collaborators for ongoing useful discussions. V.L.D. acknowledges a Leverhulme Early Career Fellowship and support from the Isaac Newton Trust. M.A.C. thanks Lauri Himanen for useful discussions on atomic descriptors and acknowledges personal funding from the Academy of Finland under project #310574. The authors are thankful for generous allocation of computational resources on the ARCHER UK National Supercomputing Service (EPSRC grants EP/K014560/1 and EP/P022596/1) and by CSC - IT Center for Science, Finland, which supported some of the work discussed herein. V.L.D. and M.A.C. are grateful for mutual HPC-Europa3 exchange visits (funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 730897), during one of which this manuscript was finalized.

## Conflict of Interest

G.C. is listed as an inventor on a patent filed by Cambridge Enterprise Ltd. related to SOAP and GAP (PCT/GB2009/001414, filed on 5 June 2009 and published on 23 September 2014). The other authors declare no conflict of interest.

## Keywords

amorphous solids, atomistic modeling, big data, force fields, molecular dynamics

Received: April 30, 2019
Revised: July 26, 2019
Published online: September 5, 2019

[1] Y. Liu, B. V Merinov, W. A. Goddard, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3735.
[2] K. J. Griffith, K. M. Wiaderek, G. Cibin, L. E. Marbella, C. P. Grey, *Nature* **2018**, *559*, 556.
[3] M. Jansen, *Adv. Mater.* **2015**, *27*, 3229.
[4] A. Jain, Y. Shin, K. A. Persson, *Nat. Rev. Mater.* **2016**, *1*, 1.
[5] A. R. Oganov, C. J. Pickard, Q. Zhu, R. J. Needs, *Nat. Rev. Mater.* **2019**, *4*, 331.
[6] R. Dronskowski, *Computational Chemistry of Solid State Materials*, Wiley-VCH, Weinheim, Germany **2005**.
[7] K. Burke, *J. Chem. Phys.* **2012**, *136*, 150901.
[8] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. De Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Dułak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iuşan, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Küçükbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G. M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. Van Setten, V. Van Speybroeck, J. M. Wills, J. R. Yates, G. X. Zhang, S. Cottenier, *Science* **2016**, *351*, aad3000.
[9] A. Zakutayev, X. Zhang, A. Nagaraja, L. Yu, S. Lany, T. O. Mason, D. S. Ginley, A. Zunger, *J. Am. Chem. Soc.* **2013**, *135*, 10048.
[10] R. Gautier, X. Zhang, L. Hu, L. Yu, Y. Lin, T. O. L. Sunde, D. Chon, K. R. Poeppelmeier, A. Zunger, *Nat. Chem.* **2015**, *7*, 308.

[11] Y. Hinuma, T. Hatakeyama, Y. Kumagai, L. A. Burton, H. Sato, Y. Muraba, S. Iimura, H. Hiramatsu, I. Tanaka, H. Hosono, F. Oba, *Nat. Commun.* **2016**, *7*, 11962.
[12] R. L. C. Vink, G. T. Barkema, M. A. Stijnman, R. H. Bisseling, *Phys. Rev. B* **2001**, *64*, 2452141.
[13] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *npj Comput. Mater.* **2017**, *3*, 54.
[14] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547.
[15] J. E. Gubernatis, T. Lookman, *Phys. Rev. Mater.* **2018**, *2*, 120301.
[16] A. Mosquera, B. Fu, K. L. Kohlstedt, G. C. Schatz, M. A. Ratner, *ACS Energy Lett.* **2018**, *3*, 155.
[17] A. Seko, K. Toyoura, S. Muto, T. Mizoguchi, S. Broderick, *MRS Bull.* **2018**, *43*, 690.
[18] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, A. Fazzio, *J. Phys.: Mater.* **2019**, *2*, 032001.
[19] M. Ceriotti, *J. Chem. Phys.* **2019**, *150*, 150901.
[20] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, O. A. Von Lilienfeld, *J. Chem. Theory Comput.* **2017**, *13*, 5255.
[21] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, A. Tkatchenko, *Nat. Commun.* **2017**, *8*, 13890.
[22] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, A. E. Roitberg, *Nat. Commun.* **2019**, *10*, 2903.
[23] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, *ACS Cent. Sci.* **2018**, *4*, 268.
[24] P. S. Gromski, A. B. Henson, J. M. Granda, L. Cronin, *Nat. Rev. Chem.* **2019**, *3*, 119.
[25] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D. G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams, A. Aspuru-Guzik, *Nat. Mater.* **2016**, *15*, 1120.
[26] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, *Nat. Commun.* **2017**, *8*, 15679.
[27] A. Mansouri Tehrani, A. O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T. D. Sparks, J. Brgoch, *J. Am. Chem. Soc.* **2018**, *140*, 9844.
[28] G. C. Sosso, V. L. Deringer, S. R. Elliott, G. Csányi, *Mol. Simul.* **2018**, *44*, 866.
[29] A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, *87*, 184115.
[30] N. Bernstein, B. Bhattarai, G. Csányi, D. A. Drabold, S. R. Elliott, V. L. Deringer, *Angew. Chem., Int. Ed.* **2019**, *58*, 7057.
[31] J. Behler, *Angew. Chem., Int. Ed.* **2017**, *56*, 12828.
[32] M. A. Caro, *Arkhimedes* **2018**, *3*, 21.
[33] C. J. Pickard, R. J. Needs, *Phys. Rev. Lett.* **2006**, *97*, 045504.
[34] V. L. Deringer, G. Csányi, D. M. Proserpio, *ChemPhysChem* **2017**, *18*, 873.
[35] A. Belkly, M. Helderman, V. L. Karen, P. Ulkch, *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 364.
[36] W. J. Szlachta, A. P. Bartók, G. Csányi, *Phys. Rev. B* **2014**, *90*, 104108.
[37] W. Kohn, *Phys. Rev. Lett.* **1996**, *76*, 3168.
[38] G. Binnig, H. Rohrer, C. Gerber, E. Weibel, *Phys. Rev. Lett.* **1983**, *50*, 120.
[39] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, M. Ceriotti, *Sci. Adv.* **2017**, *3*, e1701816.
[40] J. Behler, M. Parrinello, *Phys. Rev. Lett.* **2007**, *98*, 146401.
[41] H. Eshet, R. Z. Khaliullin, T. D. Kühne, J. Behler, M. Parrinello, *Phys. Rev. B* **2010**, *81*, 184107.
[42] G. C. Sosso, G. Miceli, S. Caravati, J. Behler, M. Bernasconi, *Phys. Rev. B* **2012**, *85*, 174103.
[43] V. L. Deringer, G. Csányi, *Phys. Rev. B* **2017**, *95*, 094203.
[44] P. E. Dolgirev, I. A. Kruglov, A. R. Oganov, *AIP Adv.* **2016**, *6*, 085318.
[45] S. Hajinazar, J. Shao, A. N. Kolmogorov, *Phys. Rev. B* **2017**, *95*, 014114.
[46] V. L. Deringer, C. J. Pickard, G. Csányi, *Phys. Rev. Lett.* **2018**, *120*, 156001.
[47] E. L. Kolsbjerg, A. A. Peterson, B. Hammer, *Phys. Rev. B* **2018**, *97*, 195424.
[48] Q. Tong, L. Xue, J. Lv, Y. Wang, Y. Ma, *Faraday Discuss.* **2018**, *211*, 31.
[49] V. L. Deringer, D. M. Proserpio, G. Csányi, C. J. Pickard, *Faraday Discuss.* **2018**, *211*, 45.
[50] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, A. R. Oganov, *Phys. Rev. B* **2019**, *99*, 064114.
[51] J. Behler, *J. Chem. Phys.* **2011**, *134*, 074106.
[52] M. J. Willatt, F. Musil, M. Ceriotti, *J. Chem. Phys.* **2019**, *150*, 154110.
[53] R. Drautz, *Phys. Rev. B* **2019**, *99*, 014104.
[54] A. Shapeev, *Multiscale Model. Simul.* **2016**, *14*, 1153.
[55] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, G. J. Tucker, *J. Comput. Phys.* **2015**, *285*, 316.
[56] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, S. P. Ong, *arXiv* **2019**, arXiv:1906.08888.
[57] L. Himanen, M. O. J. Jäger, E. V Morooka, F. F. Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, A. S. Foster, *arXiv* **2019**, arXiv:1904.08875.
[58] M. A. Caro, *Phys. Rev. B* **2019**, *100*, 024112.
[59] A. Seko, H. Hayashi, K. Nakayama, A. Takahashi, I. Tanaka, *Phys. Rev. B* **2017**, *95*, 144110.
[60] H. Huo, M. Rupp, *arXiv* **2018**, arXiv:1704.06439.
[61] M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. von Lilienfeld, *Phys. Rev. Lett.* **2012**, *108*, 058301.
[62] T. B. Blank, S. D. Brown, A. W. Calhoun, D. J. Doren, *J. Chem. Phys.* **1995**, *103*, 4129.
[63] S. Lorenz, A. Groß, M. Scheffler, *Chem. Phys. Lett.* **2004**, *395*, 210.
[64] S. Lorenz, M. Scheffler, A. Gross, *Phys. Rev. B* **2006**, *73*, 115431.
[65] J. Behler, S. Lorenz, K. Reuter, *J. Chem. Phys.* **2007**, *127*, 014705.
[66] J. Behler, *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930.
[67] A. Singraber, J. Behler, C. Dellago, *J. Chem. Theory Comput.* **2019**, *15*, 1827.
[68] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, *Phys. Rev. Lett.* **2010**, *104*, 136403.
[69] M. A. Wood, A. P. Thompson, *J. Chem. Phys.* **2018**, *148*, 241721.
[70] V. Botu, R. Ramprasad, *Phys. Rev. B* **2015**, *92*, 094306.
[71] T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, R. Ramprasad, *npj Comput. Mater.* **2017**, *3*, 37.
[72] V. Botu, R. Batra, J. Chapman, R. Ramprasad, *J. Phys. Chem. C* **2017**, *121*, 511.
[73] N. Artrith, A. Urban, *Comput. Mater. Sci.* **2016**, *114*, 135.
[74] A. Khorshidi, A. A. Peterson, *Comput. Phys. Commun.* **2016**, *207*, 310.
[75] E. V. Podryabinkin, A. V. Shapeev, *Comput. Mater. Sci.* **2017**, *140*, 171.
[76] H. Wang, L. Zhang, J. Han, W. E, *Comput. Phys. Commun.* **2018**, *228*, 178.
[77] J. S. Smith, O. Isayev, A. E. Roitberg, *Chem. Sci.* **2017**, *8*, 3192.
[78] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, K. Müller, *Sci. Adv.* **2017**, *3*, e1603015.
[79] K. Yao, J. E. Herr, D. W. Toth, R. McKintyre, J. Parkhill, *Chem. Sci.* **2018**, *9*, 2261.
[80] Z. Li, J. R. Kermode, A. De Vita, *Phys. Rev. Lett.* **2015**, *114*, 096405.
[81] R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, M. Bokdam, *Phys. Rev. Lett.* **2019**, *122*, 225701.
[82] R. Jinnouchi, F. Karsai, G. Kresse, *Phys. Rev. B* **2019**, *100*, 014105.
[83] F. H. Stillinger, T. A. Weber, *Phys. Rev. B* **1985**, *31*, 5262.
[84] J. Tersoff, *Phys. Rev. B* **1988**, *37*, 6991.
[85] M. Z. Bazant, E. Kaxiras, J. Justo, *Phys. Rev. B* **1997**, *56*, 8542.
[86] G. P. Purja Pun, Y. Mishin, *Phys. Rev. B* **2017**, *95*, 224103.

 

[87] K. D. Brommer, M. Needels, B. E. Larson, J. D. Joannopoulos, *Phys. Rev. Lett.* **1992**, *68*, 1355.
[88] R. Xie, G. G. Long, S. J. Weigand, S. C. Moss, T. Carvalho, S. Roorda, M. Hejna, S. Torquato, P. J. Steinhardt, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13250.
[89] D. E. Carlson, C. R. Wronski, *Appl. Phys. Lett.* **1976**, *28*, 671.
[90] R. A. Street, *Adv. Mater.* **2009**, *21*, 2007.
[91] L.-F. Cui, R. Ruffo, C. K. Chan, H. Peng, Y. Cui, *Nano Lett.* **2009**, *9*, 491.
[92] J. Behler, R. Martonák, D. Donadio, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*, 185501.
[93] J. Behler, R. Martoňák, D. Donadio, M. Parrinello, *Phys. Status Solidi B* **2008**, *245*, 2618.
[94] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562.
[95] A. P. Bartók, J. R. Kermode, N. Bernstein, G. Csányi, *Phys. Rev. X* **2018**, *8*, 041048.
[96] K. Momma, F. Izumi, *J. Appl. Crystallogr.* **2011**, *44*, 1272.
[97] V. L. Deringer, N. Bernstein, A. P. Bartók, M. J. Cliffe, R. N. Kerber, L. E. Marbella, C. P. Grey, S. R. Elliott, G. Csányi, *J. Phys. Chem. Lett.* **2018**, *9*, 2879.
[98] W.-L. Shao, J. Shinar, B. C. Gerstein, F. Li, J. S. Lannin, *Phys. Rev. B* **1990**, *41*, 9491.
[99] L. Pastewka, P. Pou, R. Pérez, P. Gumbsch, M. Moseler, *Phys. Rev. B* **2008**, *78*, 161402.
[100] J. R. Kermode, T. Albaret, D. Sherman, N. Bernstein, P. Gumbsch, M. C. Payne, G. Csányi, A. De Vita, *Nature* **2008**, *455*, 1224.
[101] C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azaïs, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri, C. J. Pickard, *Chem. Rev.* **2012**, *112*, 5733.
[102] L. Bonati, M. Parrinello, *Phys. Rev. Lett.* **2018**, *121*, 265701.
[103] J. Anwar, D. Zahn, *Angew. Chem., Int. Ed.* **2011**, *50*, 1996.
[104] F. Dorner, Z. Sukurma, C. Dellago, G. Kresse, *Phys. Rev. Lett.* **2018**, *121*, 195701.
[105] M. Wuttig, N. Yamada, *Nat. Mater.* **2007**, *6*, 824.
[106] D. Loke, T. H. Lee, W. J. Wang, L. P. Shi, R. Zhao, Y. C. Yeo, T. C. Chong, S. R. Elliott, *Science* **2012**, *336*, 1566.
[107] F. Rao, K. Ding, Y. Zhou, Y. Zheng, M. Xia, S. Lv, Z. Song, S. Feng, I. Ronneberger, R. Mazzarello, W. Zhang, E. Ma, *Science* **2017**, *358*, 1423.
[108] W. Zhang, R. Mazzarello, M. Wuttig, E. Ma, *Nat. Rev. Mater.* **2019**, *4*, 150.
[109] C. D. Wright, Y. Liu, K. I. Kohary, M. M. Aziz, R. J. Hicken, *Adv. Mater.* **2011**, *23*, 3408.
[110] A.-K. U. Michel, D. N. Chigrin, T. W. W. Maß, K. Schönauer, M. Salinga, M. Wuttig, T. Taubner, *Nano Lett.* **2013**, *13*, 3470.
[111] P. Hosseini, C. D. Wright, H. Bhaskaran, *Nature* **2014**, *511*, 206.
[112] M. Wuttig, H. Bhaskaran, T. Taubner, *Nat. Photonics* **2017**, *11*, 465.
[113] M. N. Schneider, T. Rosenthal, C. Stiewe, O. Oeckler, *Z. Kristallogr.* **2010**, *225*, 463.
[114] J. Li, Z. Chen, X. Zhang, H. Yu, Z. Wu, H. Xie, Y. Chen, Y. Pei, *Adv. Sci.* **2017**, *4*, 1700341.
[115] M. Hong, J. Zou, Z. G. Chen, *Adv. Mater.* **2019**, *31*, 1807071.
[116] J. Goldak, C. S. Barrett, D. Innes, W. Youdelis, *J. Chem. Phys.* **1966**, *44*, 3323.
[117] T. Rosenthal, M. N. Schneider, C. Stiewe, M. Döblinger, O. Oeckler, *Chem. Mater.* **2011**, *23*, 4349.
[118] V. L. Deringer, R. Dronskowski, M. Wuttig, *Adv. Funct. Mater.* **2015**, *25*, 6343.
[119] J. Akola, R. O. Jones, *Phys. Status Solidi B* **2012**, *249*, 1851.
[120] W. Zhang, V. L. Deringer, R. Dronskowski, R. Mazzarello, E. Ma, M. Wuttig, *MRS Bull.* **2015**, *40*, 856.
[121] S. Caravati, M. Bernasconi, T. D. Kühne, M. Krack, M. Parrinello, *Appl. Phys. Lett.* **2007**, *91*, 171906.
[122] J.-Y. Raty, W. Zhang, J. Luckas, C. Chen, R. Mazzarello, C. Bichara, M. Wuttig, *Nat. Commun.* **2015**, *6*, 7467.
[123] S. Gabardi, S. Caravati, G. C. Sosso, J. Behler, M. Bernasconi, *Phys. Rev. B* **2015**, *92*, 054201.
[124] S. Gabardi, E. Baldi, E. Bosoni, D. Campi, S. Caravati, G. C. Sosso, J. Behler, M. Bernasconi, *J. Phys. Chem. C* **2017**, *121*, 23827.
[125] J. Kalikka, J. Akola, R. O. Jones, *Phys. Rev. B* **2014**, *90*, 184109.
[126] T. H. Lee, S. R. Elliott, *Adv. Mater.* **2017**, *29*, 1700814.
[127] I. Ronneberger, W. Zhang, R. Mazzarello, *MRS Commun.* **2018**, *8*, 1018.
[128] T. D. Kühne, M. Krack, F. R. Mohamed, M. Parrinello, *Phys. Rev. Lett.* **2007**, *98*, 066401.
[129] F. C. Mocanu, K. Konstantinou, T. H. Lee, N. Bernstein, V. L. Deringer, G. Csányi, S. R. Elliott, *J. Phys. Chem. B* **2018**, *122*, 8998.
[130] G. C. Sosso, G. Miceli, S. Caravati, F. Giberti, J. Behler, M. Bernasconi, *J. Phys. Chem. Lett.* **2013**, *4*, 4241.
[131] G. C. Sosso, D. Donadio, S. Caravati, J. Behler, M. Bernasconi, *Phys. Rev. B* **2012**, *86*, 104301.
[132] M. Salinga, B. Kersting, I. Ronneberger, V. P. Jonnalagadda, X. T. Vu, M. Le Gallo, I. Giannopoulos, O. Cojocaru-Mirédin, R. Mazzarello, A. Sebastian, *Nat. Mater.* **2018**, *17*, 681.
[133] K. Konstantinou, F. C. Mocanu, T.-H. Lee, S. R. Elliott, *Nat. Commun.* **2019**, *10*, 3065.
[134] R. E. Simpson, P. Fons, A. V Kolobov, T. Fukaya, M. Krbal, T. Yagi, J. Tominaga, *Nat. Nanotechnol.* **2011**, *6*, 501.
[135] J.-J. Wang, J. Wang, H. Du, L. Lu, P. C. Schmitz, J. Reindl, A. M. Mio, C.-L. Jia, E. Ma, R. Mazzarello, M. Wuttig, W. Zhang, *Chem. Mater.* **2018**, *30*, 4770.
[136] J. Hegedus, S. R. Elliott, *Phys. Status Solidi A* **2010**, *207*, 510.
[137] A. S. Barnard, *Rep. Prog. Phys.* **2010**, *73*, 086502.
[138] K. Reuter, M. Scheffler, *Phys. Rev. B* **2001**, *65*, 35406.
[139] J. S. Elias, N. Artrith, M. Bugnet, L. Giordano, G. A. Botton, A. M. Kolpak, Y. Shao-Horn, *ACS Catal.* **2016**, *6*, 1675.
[140] N. Artrith, A. M. Kolpak, *Nano Lett.* **2014**, *14*, 2670.
[141] G. Wulff, *Z. Krystallograph. Miner.* **1901**, *34*, 449.
[142] S. Hajinazar, E. D. Sandoval, A. J. Cullo, A. N. Kolmogorov, *Phys. Chem. Chem. Phys.* **2019**, *21*, 8729.
[143] R. Hoffmann, A. A. Kabanov, A. A. Golov, D. M. Proserpio, *Angew. Chem., Int. Ed.* **2016**, *55*, 10962.
[144] M. Aykol, S. S. Dwaraknath, W. Sun, K. A. Persson, *Sci. Adv.* **2018**, *4*, eaaq0148.
[145] J. Robertson, *Adv. Phys.* **1986**, *35*, 317.
[146] T. Laurila, S. Sainio, M. A. Caro, *Prog. Mater. Sci.* **2017**, *88*, 499.
[147] R. Dash, J. Chmiola, G. Yushin, Y. Gogotsi, G. Laudisio, J. Singer, J. Fischer, S. Kucheyev, *Carbon* **2006**, *44*, 2489.
[148] Z. Zhang, R. Brydson, Z. Aslam, S. Reddy, A. Brown, A. Westwood, B. Rand, *Carbon* **2011**, *49*, 5049.
[149] A. C. Forse, C. Merlet, P. K. Allan, E. K. Humphreys, J. M. Griffin, M. Aslan, M. Zeiger, V. Presser, Y. Gogotsi, C. P. Grey, *Chem. Mater.* **2015**, *27*, 6848.
[150] Z. Jian, C. Bommier, L. Luo, Z. Li, W. Wang, C. Wang, P. A. Greaney, X. Ji, *Chem. Mater.* **2017**, *29*, 2314.
[151] D. Kong, Y. Gao, Z. Xiao, X. Xu, X. Li, L. Zhi, *Adv. Mater.* **2018**, 1804973.
[152] J. Tersoff, *Phys. Rev. Lett.* **1988**, *61*, 2879.
[153] C. de Tomas, I. Suarez-Martinez, N. A. Marks, *Carbon* **2016**, *109*, 681.
[154] T. B. Shiell, D. G. McCulloch, D. R. McKenzie, M. R. Field, B. Haberl, R. Boehler, B. A. Cook, C. De Tomas, I. Suarez-Martinez, N. A. Marks, J. E. Bradby, *Phys. Rev. Lett.* **2018**, *120*, 215701.
[155] R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler, M. Parrinello, *Phys. Rev. B* **2010**, *81*, 100103.
[156] R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler, M. Parrinello, *Nat. Mater.* **2011**, *10*, 693.
[157] Q. Zhu, A. R. Oganov, M. A. Salvadó, P. Pertierra, A. O. Lyakhov, *Phys. Rev. B* **2011**, *83*, 193410.

 

[158] J. Robertson, *Mater. Sci. Eng., R* **2002**, *37*, 129.
[159] M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, G. Csányi, *Phys. Rev. Lett.* **2018**, *120*, 166101.
[160] V. L. Deringer, M. A. Caro, R. Jana, A. Aarva, S. R. Elliott, T. Laurila, G. Csányi, L. Pastewka, *Chem. Mater.* **2018**, *30*, 7438.
[161] M. A. Caro, A. Aarva, V. L. Deringer, G. Csányi, T. Laurila, *Chem. Mater.* **2018**, *30*, 7446.
[162] B. Cheng, E. A. Engel, J. Behler, C. Dellago, M. Ceriotti, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1110.
[163] V. L. Deringer, C. Merlet, Y. Hu, T. H. Lee, J. A. Kattirtzi, O. Pecher, G. Csányi, S. R. Elliott, C. P. Grey, *Chem. Commun.* **2018**, *54*, 5988.
[164] R. C. Powles, N. A. Marks, D. W. M. Lau, *Phys. Rev. B* **2009**, *79*, 075430.
[165] S. Fujikake, V. L. Deringer, T. H. Lee, M. Krynski, S. R. Elliott, G. Csányi, *J. Chem. Phys.* **2018**, *148*, 241714.
[166] N. Artrith, A. Urban, G. Ceder, *J. Chem. Phys.* **2018**, *148*, 241711.
[167] B. Onat, E. D. Cubuk, B. D. Malone, E. Kaxiras, *Phys. Rev. B* **2018**, *97*, 094106.
[168] M. G. Darley, C. M. Handley, P. L. A. Popelier, *J. Chem. Theory Comput.* **2008**, *4*, 1435.
[169] N. Artrith, T. Morawietz, J. Behler, *Phys. Rev. B* **2011**, *83*, 153101.
[170] S. Faraji, S. A. Ghasemi, S. Rostami, R. Rasoulkhani, B. Schaefer, S. Goedecker, M. Amsler, *Phys. Rev. B* **2017**, *95*, 104105.
[171] N. Artrith, B. Hiller, J. Behler, *Phys. Status Solidi B* **2013**, *250*, 1191.
[172] H. A. Eivari, S. A. Ghasemi, H. Tahmasbi, S. Rostami, S. Faraji, R. Rasoulkhani, S. Goedecker, M. Amsler, *Chem. Mater.* **2017**, *29*, 8594.
[173] G. Fisicaro, M. Sicher, M. Amsler, S. Saha, L. Genovese, S. Goedecker, *Phys. Rev. Mater.* **2017**, *1*, 033609.