Machine learning for the modeling of interfaces in energy storage and conversion materials

Nongnuch Artrith

Department of Chemical Engineering, Columbia University, New York, NY 10027, United States of America

E-mail: nartrith@atomistic.net

Keywords: machine learning, energy storage, energy conversion, interfaces, atomistic simulations

Abstract

The properties and atomic-scale dynamics of interfaces play an important role for the performance of energy storage and conversion devices such as batteries and fuel cells. In this topical review, we consider recent progress in machine-learning (ML) approaches for the computational modeling of materials interfaces. ML models are computationally much more efficient than first principles methods and thus allow to model larger systems and extended timescales, a necessary prerequisite for the accurate description of many interface properties. Here we review the recent major developments of ML-based interatomic potentials for atomistic modeling and ML approaches for the direct prediction of materials properties. This is followed by a discussion of ML applications to solid–gas, solid–liquid, and solid–solid interfaces as well as to nanostructured and amorphous phases that commonly form in interface regions. We then highlight how ML has been used to obtain important insights into the structure and stability of interfaces, interfacial reactions, and mass transport at interfaces. Finally, we offer a perspective on the current state of ML potential development and identify future directions and opportunities for this exciting research field.

1. Introduction

The performance of energy storage and conversion devices depends to a significant extent on materials interfaces and their properties: heterogeneous catalysis at solid–liquid and solid–gas interfaces drives the interconversion between thermal [1] electrical [2, 3] or electromagnetic [4, 5] energy and chemical energy. Typical degradation pathways in batteries are decomposition reactions at the cathode/electrolyte solid–liquid interface [6]. Similarly, the durability of solid state batteries [7–13], in which the liquid electrolyte is replaced by a solid ionic conductor, is currently limited by the instability of the electrode/electrolyte solid–solid interfaces. All of these factors, namely the catalytic activity, the surface transport/impedance, and the stability of the electrode/electrolyte interfaces are also crucial for the performance of solid oxide fuel cells [14–16].

A precise understanding of the structure, the chemical reaction dynamics, and the transport processes at such interfaces on the atomic scale is needed to identify specific performance bottlenecks or deficiencies that subsequently can be addressed in the next generation of materials. However, owing to the narrow widths of interface regions and to their often non-crystalline structures, in situ experimental characterization of interfaces with atomic resolution is extremely challenging.

Atomistic simulations are, in principle, an ideal tool to investigate dynamical phenomena on the atomic scale. First principles methods, especially density-functional theory (DFT) calculations [17, 18], have become a standard tool for the modeling of catalytic reactions at solid–gas interfaces also including strong metal-support interactions [19, 20]. But realistic models of solid–solid and solid–liquid interface structures that exhibit disorder and crystallographic defects typically require simulating several hundred to thousands of atoms. Such system sizes are beyond the capabilities of conventional DFT methods while conventional interatomic potentials are often not reliable for interfaces, especially when two very different materials are involved (e.g. at metal–insulator boundaries).
During the last years, new computational approaches based on machine-learning (ML) techniques have demonstrated promising capabilities for materials modeling [21–24]. ML is used both to accelerate atomistic simulations with near DFT accuracy and to avoid atomistic simulations altogether by making predictions based on information extracted from either computational or experimental databases. ML models may also aid in interpreting interface characterization measurements, for example, by modeling the electrochemical response in impedance spectroscopy [25, 26].

In this review, we discuss progress that has been made in the modeling of interfaces and related structures using different ML approaches. In the next section, we will briefly consider milestones in the methodological developments of ML for materials simulations. We then review applications of ML for research questions related to surfaces and interfaces before offering a perspective on the remaining challenges and possible future directions.

2. Progress in machine learning methods for materials simulations

In the last decade, machine learning has experienced renewed interest in materials science, and the recent perspectives by Sanchez-Lengeling and Aspuru-Guzik [23] and by Butler et al [21] provide an overview of state-of-the-art ML techniques for materials science applications. However, ML techniques, especially artificial neural networks (ANN), have a long history in materials science and chemistry and have been used since the early 1990s to aid, e.g. in the interpretation of spectroscopic data [27] and for general pattern detection [28]. Mitchell recently reviewed a wide range of ML techniques that have been applied to different problems in cheminformatics and drug discovery covering the last 30 years [29], and Venkatasubramanian reviewed machine learning applications in chemical engineering [30]. Early applications of ANNs as replacement for conventional interatomic potentials have been reported already in the 1990s. For example, already in 1992 Sumpter et al employed ANNs to predict molecular force fields from vibrational spectroscopy [31], and in 1997 No, Scheraga et al used ANNs to interpolate the potential energy surface of the water dimer from quantum-chemistry calculations [32]. Another early example of an ANN potential is the work by Gassner et al on solvated aluminum ions from 1998 [33].

These early techniques did not find widespread adoption as they were specific to atomic structures and owing to a lack of available free and open-source tools. The sudden progress in ML techniques for materials applications during the last decade has been enabled mainly by two factors: First, advanced ML techniques such as deep learning [34] have significantly improved the predictive power of ML for classification and have resulted in the release of user-friendly industry-backed software packages such as Google’s TensorFlow [35] and Facebook’s PyTorch [36]. And second, new approaches for the training of transferable machine-learning potentials (MLP) for atomistic simulations have been developed in the late 2000s [37, 38].

The new software libraries have made state-of-the-art ML methods accessible to researchers, resulting in numerous exciting applications in the area of materials science in the past few years. A significant body of work has demonstrated that ML models can be trained to predict the results of first principles DFT calculations without the need for actual computationally intensive computations beyond the calculation of the reference data set. To name just a few examples, Rupp et al used kernel ridge regression (KRR) to train a model for the prediction of DFT molecular atomization energies [39], Jong et al used a gradient boosting approach to train DFT elastic moduli [40], Faber et al used a KRR model to predict DFT formation energies of Elpasolite (ABC₃Dₓ) compositions [41], and Ye et al trained a deep ANN to predict DFT crystal stability [42]. Isayev et al proposed a model trained to DFT references with the gradient boosting decision tree technique to predict various properties of crystal structures including band gaps, elastic properties, and heat capacities [43]. Chen et al developed a graph network model that can also be trained to predict various properties including band gaps, and elastic moduli [44].

The list in the previous paragraph is not exhaustive, and other publications along the same line have been published during the last years. Apart from the ML technique (ANN, KRR, decision trees, etc), the methods mainly differ in their choice of descriptor used to extract feature vectors as input for the ML model. Here, the challenge is to capture both the structure and composition of a material in a constant-size vector that is invariant with respect to rotation, translation, and the exchange of equivalent atoms. Many descriptors used in ML models for materials applications have recently been reviewed by Reveil and Clancy [45].

Other exciting recent work has made use of ML for predicting the structure and composition of new materials based on the properties of known materials. This is the inverse of what can be done with first principles calculations, as here the structure and composition is the output and the materials properties are the input. In 2010, Hautier et al employed a probabilistic inference approach based on a database of known inorganic materials to predict new likely stable ternary compositions and their crystal structures, proposing a list of 209 new materials [46]. In a similar spirit, Meredig et al constructed a ML model based on ensembles of decision trees
and trained to a large database of DFT reference calculations to predict novel ternary compositions and their properties, identifying 4500 new potentially stable compounds in 2014 [47]. A very different approach was taken more recently by Raccuglia et al who constructed a support vector machine ML model that is able to predict the likely success or failure of solid-state syntheses based on historic data about successful and failed synthesis attempts [48]. These applications are very powerful, as ML here does a job that was previously done by humans, namely to predict new materials by extrapolation from previous experience.

The second area of research that has recently seen tremendous advancement is the development of MLP as alternative to conventional interatomic potentials or reactive force fields. The idea behind MLPs is to train ML models to energies and potentially interatomic forces from first principles reference calculations of small structures (containing typically fewer than 200 atoms), and then to use the more efficient ML model instead of the computationally demanding first principles method for large-scale (thousands of atoms) simulations. In 2007, Behler and Parrinello (BP) proposed an MLP for silicon based on an ANN that is constructed by learning atomic energy contributions from the DFT energy of entire structures [37]. In the BP approach, the total energy \( E(σ) \) of an atomic structure \( σ \) is the sum of atomic energies, \( E(σ) = \sum_i E_i(σ_i) \), and the atomic energy of an atom \( i \) is an ANN that takes as input a descriptor \( σ_i \) of the local atomic environment of atom \( i \). The BP approach allowed, for the first time, to re-use trained MLPs for the simulation of structures with compositions and sizes different from the ones used in the original training set. The BP ANN potential approach was subsequently extended to multiple chemical species by Artrith et al [49] and was successfully applied to a wide variety of materials including metals, transition metal oxides, composites, and nanostructures [50].

In parallel, other groups have explored the use of different ML techniques instead of ANNs and different types of descriptors for the local atomic environment. A very successful example is the Gaussian Approximation Potential (GAP) by Bartok et al that is based on Gaussian process regression [38]. Various other MLP approaches and implementations have been developed in recent years [51–59].

3. Applications of machine learning for the modeling of interfaces

In this section we consider applications of ML to the modeling of materials systems that are relevant in the context of interfaces involving at least one solid phase, i.e. solid–gas, solid–liquid, and solid–solid interfaces, as those are relevant for energy applications. Modeling realistic solid–solid interfaces poses a formidable challenge because of the complex, often amorphous phases that may form at the boundary region. In addition, predictions for solid–solid interfaces are often hard to verify experimentally. Considering the only recent emergence of ML techniques for materials modeling, it is therefore not surprising that only few reports related specifically to solid–solid interfaces have been published so far. On the other hand, the computational heterogeneous catalysis community has been particularly active in the adoption of ML approaches, and therefore the body of literature concerned with surface properties and molecule–surface interaction (solid–gas and solid–liquid interfaces) is significantly larger. Many of the features of solid–solid interfaces, such as amorphous phases, nanostructures, and metal–insulator boundaries (in form of metal-decorated oxide surfaces) have been successfully modeled with ML approaches, and the same techniques would in principle be applicable to solid–solid interfaces as well.

The present section is therefore subdivided not by type of interface but rather by materials properties ranging from structural features in sections 3.1 and 3.2 over reactions at interfaces in section 3.3 to (mass) transport (section 3.4).

3.1. Nanostructured and amorphous phases

Nanostructured and amorphous phases are often formed at interfaces and are challenging to model as they cannot be approximated with small unit cells and periodic boundary conditions but instead require the explicit treatment of a large number of atoms (e.g. there are around 2000 atoms in a 4 nm gold nanoparticle). While simple amorphous phases and nanostructures can be modeled using conventional interatomic potentials, MLPs have proven versatile for the modeling of diverse materials including systems with mixed bonding contributions.

Artrith and Kolpak employed ANN potentials to investigate the atomic ordering in CuAu nanoalloys with up to around 4000 atoms [60]. Using a simulated annealing approach, the authors found that small CuAu nanoparticles up to diameters of \( \sim 3.5 \) nm exhibit core–shell structures with gold atoms at the surface, while particles with diameters of \( \sim 6 \) nm approach the solid solution of the bulk alloys. This approach is not limited to metallic nanoparticle, and Elias et al also used an ANN potential for simulated annealing simulations to determine the distribution of Cu dopants in oxygen–deficient ceria nanoparticles, an efficient low-temperature catalyst for CO oxidation [61]. The ANN potential simulations predicted a segregation of the Cu dopant atoms to the \{ 100 \} facets in particles with a diameter of \( \sim 3.5 \) nm (\( \sim 1300 \) atoms) which is supported by electron microscopy, allowing insights into the likely nature of the catalytically active sites. The same authors also
investigated the change of the structure and composition of CuAu nanoparticles with temperature using grand-canonical molecular dynamics (MD) simulations [62]. Another example of nanoparticle modeling is the work by Chiriki et al who investigated the phase-transition dynamics of small gold nanoparticles near the melting point using ANN-potential MD simulations [63]. This work did not make any assumptions regarding the shape of the nanoparticles and was therefore limited to smaller numbers of atoms for which an exhaustive search for ground-state structures is still feasible. The simulations reproduced previously reported ground-state structures and identified a new ground-state geometry for Au34 clusters, showing that carefully constructed MLPs are suitable for the discovery of structures that were not included in the training set used for the potential fit.

Ibarra-Hernández et al also made use of ANN potentials for structural search, in this case for Mg–Ca alloy crystal structures using two different global structure optimization techniques [64]. By combining the minima-hopping method [65] and the firefly swarm optimization algorithm [66], the authors were able to identify several previously unknown candidate structures that are predicted to be stable at elevated temperatures and pressures. While the work by Ibarra-Hernández et al is not related to amorphous or nanostructured phases, it further evidences the versatility of MLPs for structure sampling with a wide range of techniques.

Simulating amorphous phases is an equally daunting task as simulating nanostructures because the absence of long-range order prevents the use of small periodic unit cells. In addition, the local structural motifs of amorphous phases are not usually known beforehand, so that extensive sampling is required to determine realistic stable or metastable structures. Owing to their high accuracy at significantly reduced computational cost, DFT-based MLPs are a good option to extend the predictive power of DFT to the required system sizes.

Deringer et al developed a GAP to model surface structures of tetrahedral amorphous carbon (ta-C) [67]. The authors employed a melt-quench protocol to generate amorphous structures and subsequently characterized the local atomic environment of each carbon atom using the smooth overlap of atomic positions structural fingerprint [68]. Results of this analysis (here reproduced in figure 1) showed, in agreement with chemical intuition, that the environment of the surface carbon atoms in ta-C is more similar to graphene (i.e. sp2) whereas the structure in the bulk is more diamond-like (i.e. sp3).

Melt-quench simulation protocols model the formation of an amorphous phase from a melt or at high temperatures. Artrith et al developed a methodology that couples a genetic algorithm (GA) with a specialized ANN potential for the efficient construction of phase diagrams of electrochemical amorphization [69]. This methodology was applied to amorphous Li/Si alloys that are prospective high-capacity anode materials for Li-ion batteries. The authors compared the amorphous structures obtained from computational delithiation (electrochemical amorphization) with those obtained from melt-quench MD simulations (thermal amorphization) and found slight differences in the pair distribution functions which they attributed to the quench-rate in the MD simulations. In subsequent work, the same group applied the GA-ANN formalism to amorphous N-substituted lithium phosphate (LiPON), a solid electrolyte for Li-ion batteries, to generate amorphous structure models as input for ab initio MD simulations in order to investigate the effect of nitrogen doping on the local atomic structure and the Li diffusivity [70].

Figure 1. Deringer et al performed melt-quench molecular dynamics simulations using a Gaussian approximation potential (GAP) to generate the amorphous carbon structures shown in the figure [67]. The color code indicates the similarity of the local atomic environment of each carbon atom with those in (a) diamond and (b) graphene. (Reproduced with permission from [67]. Copyright (2018), American Chemical Society.)
3.2. Structure and stability of surfaces and interfaces

Much of the previous section’s discussion on structure sampling in amorphous and nanostructured phases transfers directly to the modeling of surface and interface structures. Interfaces of very different materials, such as interfaces between metals and oxides, are yet more challenging to model because of the mixed bonding that has to be accounted for. Conventional interatomic potentials are typically tailored to one class of chemical bonding so that more general potentials such as MLPs are needed.

Copper nanoparticles supported on zinc oxide surfaces are the reactive component of the industrial catalyst for methanol synthesis [20, 71]. Artrith et al constructed a DFT-trained ANN potential for the Cu/ZnO system to investigate the dynamic structure changes of the catalyst at 1000 K [72]. By comparing the predicted forces acting on atoms at the interface to DFT, the authors confirmed that the ANN potential is able to accurately describe the metal/oxide boundary and can be used in predictive simulations.

While the initial structure of large supported nanoparticles can be estimated based on Wulff constructions [74] followed by thermal equilibration, the structure of smaller supported metal clusters is not known a priori. Kolsbjerg et al therefore employed an ANN potential approach coupled with GA sampling for the search for the most stable structures of MgO-supported Pt13 clusters (figure 2) [73]. The authors observed that the catalytic activity depends strongly on the cluster structure, and proposed to consider thermal ensembles of cluster geometries instead of a static cluster model.

A related example of global structure optimization with MLPs is the work by Sun and Sautet who used an ANN potential coupled with a GA to search for low-energy structures of catalytic Pt clusters in hydrogen atmosphere [75]. These simulations showed that, in agreement with the findings by Kolsbjerg et al [73], a static ground-state structure may not always be an appropriate model for catalyst nanoparticles at operation conditions, as dynamical structure changes of the Pt clusters play a major role for the catalytic activity.

In the case of supported nanoparticles or clusters, the area of the solid–solid interface region is small and only one of the two materials in contact is extended. Modeling the boundary of two crystalline materials is technically significantly more involved, as the crystal periodicity of both materials has to be accommodated. Tamura et al approached this challenge with a multi-step procedure for the modeling of grain boundary structures in aluminum (schematic reproduced in figure 3) [76]: in the first step, small grain boundary model structures are optimized with a conventional interatomic potential; the optimized structure models are then...
evaluated with DFT, and the DFT energy is partitioned into atomic contributions; the resulting data is used to train a LASSO (least absolute shrinkage and selection operator) model for the prediction of atomic energies that can then be used to predict the DFT energy of large grain-boundary structures that would otherwise be inaccessible to DFT.

Thus, the approach by Tamura et al builds on the idea to use small reference structures from first principles to train MLPs for large-scale simulations. But additionally it leverages conventional interatomic potentials for the sampling of the training set structures, which is possible if reliable potentials are available.

In the case of solid–liquid interfaces only one of the materials in contact has long-range periodicity, but the properties of liquids, such as the short-range structure, are temperature dependent. Computational studies employ simulation techniques such as MD or Monte Carlo (MC) to account for temperature effects. Artrith and Kolpak used ANN potentials to investigate the atomic ordering of CuAu nanoparticles and surface models in water using a lattice MC approach [60]. In this initial work employing a static solvation shell, the authors observed that Cu atoms segregate at the interface whereas the surfaces of CuAu alloys in vacuum are Au terminated (reproduced in figure 4). This difference has a strong impact on the electrocatalytic properties of the material and demonstrates that solvent effects cannot be neglected in electrocatalyst models.

Obtaining reliable thermodynamic averages of liquid phases from MD simulations requires sampling over long time scales that are not usually accessible by DFT. Natarajan and Behler employed ANN potentials trained on DFT data to perform nanosecond-long MD simulations of copper/water interface structures with up to 2000 atoms at temperatures between 300 and 800 K [77]. The authors analyzed radial distribution functions and observed different interaction strengths between water and the three low-index Cu surfaces. The copper/water interaction is strongest for the Cu(110) surface, followed by the Cu(100) and Cu(111) surfaces.

Solid–liquid interfaces also occur during the phase transition from the solid to the melt or vice versa. Phase transitions are slow processes compared to the femtosecond time steps of MD simulations, and thus efficient potentials are needed that allow reaching nanosecond timescales. Morawietz et al developed efficient ANN potentials to study the melting behavior of ice based on large-scale simulations of the solid–liquid ice/water interface [78]. In these simulations, the interface was stabilized using an interface pinning technique [80]. Owing to the efficiency of the ANN potentials, the authors were able to simulate the large unit cells containing 2304
water molecules (i.e. 6912 atoms) and long simulations times (15 ns) that are required to reliably estimate the melting temperature of ice (figure 5(a)). The study revealed that van der Waals (vdW) interactions are essential for a correct description of the anomalous melting behavior of water, and only when vdW forces are accurately accounted for, liquid water is denser than ice at coexistence.

The reverse process, crystallization from the melt, cannot usually be observed on the time-scales accessible by direct MD simulation since nucleation is a rare event, and hence biased sampling techniques are required. Bonati and Parrinello performed accelerated MD simulations of the crystallization of silicon from the melt within the well-tempered metadynamics formalism [81, 82] using a deep ANN potential with 5 hidden layers.
trained to DFT reference data [79]. Remarkably, a single collective variable derived from the Debye structure factor was sufficient to steer the crystallization (figure 5(b)). The approach was able to predict the entropy and enthalpy change during the phase transition in excellent agreement with experiment. Sosso et al [83] and Gabardi et al [84] investigated the fast crystallization of GeTe, a phase-change compound, using ANN-potential MD simulations. Sosso et al studied the crystallization from the supercooled liquid and identified the atomic-scale mechanism responsible for the rapid nucleation rate. Gabardi et al observed the nucleation of crystalline GeTe in 3 ns long melt-quench MD simulations near the glass transition temperature and discuss the impact of the quench rate on the nucleation mechanism, finding that a similar nucleation behavior as the one in supercooled liquids can be achieved under appropriate conditions.

Another example of the simulation of nucleation and growth is the work by Botu et al who employed an ANN potential trained to first principles atomic forces to simulate the adatom ripening on the Al(111) surface [85]. The authors demonstrated that direct MD simulations with the MLP could reproduce the experimentally observed temperature dependence of the growth mechanism that changes from dispersed small one-dimensional islands at below room temperature to large nucleated islands at room temperature.

The utility of ML for the construction of interface models is not limited to the use of MLPs in atomistic simulations. An entirely different approach towards the modeling of solid–liquid interface structures was taken by Wilson and Barnard who used regression trees to predict stable interface structures within a data set of ZnO/water-bilayer structures [86]. Within this approach, the space of metastable atomic configurations generated using a conventional interatomic potential is partitioned into sub-spaces with common structural features that are small enough to be sampled with first principles techniques.

3.3. Chemical reactions at interfaces

Two materials that are in contact at an interface region may react chemically with each other. An obvious example is heterogeneous catalysis where such a reaction is desired. On the other hand, there are unwanted interface reactions, as is the case in all-solid batteries where they constitute a major degradation pathway. Most ML applications to interface reactions that have been reported so far have been related to molecules interacting with surfaces.

Already in 2004, Lorenz et al trained an ANN to interpolate the potential energy surface of H2 dissociation over a potassium-covered Pd catalyst surface [87]. The ANN potential by Lorenz is not an atomic-energy potential but rather a system-specific interpolated potential energy surface akin to the early work by Gassner et al. discussed above [33]. However, Lorenz introduced a coordinate transform from atomic positions to symmetry-adapted coordinates so that the ANN is invariant with respect to the symmetries of the potential energy surface. A similar ANN approach, but with direct atomic distances as input features (i.e. no exploitation of symmetries), was also employed by Shen et al in 2015 to describe the 15-dimensional potential energy surface of methane dissociation over Ni(111) [88]. Kolb et al investigated the interaction of HCl molecules with the Au(111) surface using an ANN potential to provide predictions of HCl scattering from the Au surface [89].

Instead of explicitly modeling the substrate-adsorbate interactions, ML can also be used to predict interaction energies without simulation. Ma et al trained an ANN model on DFT reference data for the prediction of CO chemisorption energies on multi-metallic alloys for CO2 reduction [90]. The ANN model was able to predict adsorption energies on alloys with a wide variety of compositions with an accuracy of ~0.1 eV (figure 6). The same authors later refined the use of ML for chemisorption and developed a model that takes easily accessible features and intrinsic properties of the adsorption sites as input (geometry, coordination, electronegativity, ionic potential, electron affinity) and can predict CO adsorption energies on the surfaces of a diverse set of different alloys with a comparable accuracy (~0.1 eV) [91]. Following a similar approach, Ulissi...
et al employed an ANN potential trained to DFT to identify likely catalytically active sites for CO₂ reduction over bimetallic NiGa alloys, which were subsequently confirmed with DFT calculations [92]. This ML-assisted screening led to the discovery of a previously unconsidered active site at a dramatically reduced computational cost (figure 7). Ulissi et al. also trained a Gaussian process model on DFT adsorption energies to predict the most likely reaction pathways of syngas over Rh(111) [93], demonstrating that ML models are efficient to narrow down the reaction network to the most relevant pathways.

Although adsorption energies can often be correlated with catalytic activity, ultimately chemical reaction barriers are what governs the kinetic performance of catalysts. Singh et al. recently explored the prediction of reaction barriers using several different ML models including conventional and gradient-boosted random forest, a Gaussian process model, and an ANN model [94]. The best ML model (the ANN) achieves an accuracy of 0.22 eV, which is still too large an uncertainty for quantitative catalyst design but around 50% better than linear regression.

3.4. Transport at surfaces, interfaces, and in amorphous phases
Mass transport along and across interfaces is performance limiting in many applications. For example, in batteries and fuel cells, transport across the often amorphous electrode/electrolyte interfaces can become rate limiting as outlined in the introduction. Another example of technological relevance is lateral diffusion of reactants and products along catalyst surfaces.

Depending on the diffusion constants of the involved species, accurate simulations of transport properties often require timescales that are not accessible by direct first principles methods. Additionally, many relevant solid ionic conductors are prepared in amorphous phases, modeling of which necessitates the use of large simulation cells as discussed in section 3.1. ML methods have the potential to overcome these challenges for the modeling of transport phenomena, and successful applications have been already discussed in previous sections. In this section, we will consider work that explicitly focuses on transport properties.

Several interesting studies of mass transport near surfaces in contact with water have been published by Behler and coworkers employing DFT-trained ANN potentials. Natarajan and Behler investigated different room-temperature transport phenomena at the copper/water interface using metadynamics-steered nanosecond-long MD simulations [96]. The authors evaluated the free energy barriers for Cu adatom and vacancy diffusion and quantified the structural response of the water near the defects. Quaranta, Hellström, and Behler employed MD simulations to understand the mechanism of proton transfer at the water-ZnO(1010) interface, which led to the identification of a presolvation mechanism previously observed in highly basic solutions [97]. The same authors discovered that the proton-transfer mechanism over the ZnO(1120) surface
differs from that over ZnO(1010) [98] and also investigated in further detail the structure of the water near ZnO surfaces [99].

MLP-based simulations have also proven useful for the modeling of ionic transport in disordered and amorphous solids, especially Li ion transport in Li-ion battery electrode and solid electrolyte materials. Amorphous lithium phosphate (LiP04) is a Li ion conductor with potential applications as solid electrolyte in all-solid batteries [100]. Li, Watanabe et al employed ANN potentials trained on DFT data to investigate Li transport in amorphous LiP04 [95]. The study considered both small DFT-generated structure models of stoichiometric LiP04 and large models (up to ~1000 atoms) with slightly Li-deficient compositions directly from ANN-potential calculations (figure 8), all of which were obtained from melt-quench simulations. The activation energies for Li diffusion were obtained from MD simulations using the ANN potential and were estimated to be ~0.55 eV, in good agreement with experiment [101, 102] and ab initio MD simulations [70]. Li et al also used an ANN potential to model Cu diffusion in amorphous Ta2O5, though the potential in this later work was trained only on the energy differences caused by Cu intercalation, thereby reducing the complexity of the potential energy surface [103]. A similar approach was taken by Fujikake et al who augmented a GAP model with a pair potential to account for the energy differences arising from Li intercalation into carbon-based materials including graphene, graphite, and disordered carbon nanostructures [104]. These two proof-of-concept studies show that the atomic interactions that govern transport properties often have a lower complexity than the full system comprised of the diffusing species and the surrounding matrix, a fact that can be exploited in the construction of MLPs.

It is also feasible to model highly complex structure-composition-conductivity relationships: Artrith et al investigated Li transport in LiSi nanoparticles including the amorphization and the change of the Li diffusivity upon Li extraction [105]. These simulations employed a DFT-trained ANN potential for nanosecond-long MD simulations and an iterative delithiation protocol starting with crystalline Li15Si4 coupled with the GA approach of [69]. The simulations revealed that during the initial stages of delithiation, Si clusters into chains which is beneficial for Li diffusion, though the formation of extended Si networks at lower Li contents eventually becomes limiting for Li transport. The authors therefore suggest that doping with species that strongly interact with Si could enhance Si clustering and thereby improve Li diffusivity in amorphous LiSi alloys.

Figure 8. Structure models of lithium and oxygen deficient amorphous lithium phosphate (LiP04) generated in melt-quench simulations using DFT (smallest size only) and an ANN potential. To accommodate for the lack of oxygen, some of the phosphate groups (P04 tetrahedra) share one oxygen atom, resulting in characteristic P2O7 local structural motifs that are found in all of the generated structures. (Reproduced from [95], with the permission of AIP Publishing.)
4. Remaining challenges and outlook

As reviewed in the previous section, the adoption of ML techniques for the modeling of interfaces and complex energy materials and for the direct prediction of related materials properties has rapidly increased over the past years. Nevertheless, both practical as well as technical limitations remain that will have to be addressed in future research to remove remaining barriers.

In the case of MLPs for atomic-scale simulations, it is our opinion that three factors related to the infrastructure are currently impeding progress in both method development and method adoption (figure 9):

(i) Standardization of parameter formats, the definition of the ML method, and the descriptors will be needed, so that MLP models generated with one specific training software can be used with a number of sampling (e.g. MD or MC) software. Currently, most MLP implementations provide interfaces with specific codes, so that multiple distinct MLP interfaces with popular MD software exist that differ in efficiency and features. Unified standards would facilitate the reuse and sharing of published potentials. Since one of the main differences of the various MLP approaches is the descriptor that is used to capture local atomic environments, we emphasize the importance of transferable and well-tested descriptor libraries [45].

(ii) Protocols for the automated training and testing of MLPs could also be standardized and shared. The construction of MLPs requires large reference data sets that are typically generated using automated (high-throughput) DFT calculations. The choice of reference structures depends critically on the application. For example, for interface MLPs it is important that both the individual materials in contact, as well as the boundary region, are well represented in the reference data set (figure 10). The process of reference data compilation and the potential training for different applications could be automated with standard tools such as AFLow [106], Pymatgen [107], AiiDA [108], ASE [109], etc so that training protocols can be exchanged. Protocols for rigorous potential tests could also be developed to ensure that MLPs are compared on an equal footing.
(iii) Centralized repositories would facilitate the exchange of MLP models and protocols. Ideally, the MLP community would leverage existing repositories for interatomic potentials such as the open knowledgebase of interatomic models (OpenKIM) [110] or the NIST interatomic potentials repository [111] which would have to be extended towards compatibility with MLPs.

There are also technical challenges related to MLPs that have to be overcome with further methodological development. Most pressing, in our opinion, is to address the complexity explosion with increasing number of chemical species, especially considering that real-world energy materials and their interfaces often involve compounds with diverse and complex chemistry (e.g. metal oxides and sulfides interacting with organic phases). We have recently proposed a method that removes the scaling of the descriptor dimension with the number of chemical species, so that complex compositions with many chemical species (we tested up to 11 chemical elements) become accessible for MLP techniques [112]. However, the configurational space also increases with the number of species owing to the combinatorial explosion, and as a consequence the size of reference data sets rapidly grows for complex compositions, unless sampling is limited to specific phases and applications. New ideas will be needed to overcome this limitation. One possible avenue might be including additional physical interactions such as charge-density terms in the next generation of MLPs [113, 114].

Finally, the direct DFT-quality prediction of materials properties with ML, such as the prediction of adsorption energies and reaction pathways discussed in section 3.3, has great potential also for the prediction of other interface properties. However, in our opinion, the greatest untapped potential of ML for interface properties lies in the inverse models discussed in section 2 that can answer questions that are impossible to address directly with atomistic modeling, such as the prediction of interfaces structures and compositions with specific properties. It will be exciting to see the emergence of applications in this area over the coming years, especially related to the properties of interfaces in energy conversion and storage.

Acknowledgments

I would like to thank the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation under Grant No. ACI-1053575, for supporting the development of the machine learning potential framework. I am grateful to Joerg Behler (back then at Ruhr-University Bochum), Alexie Kolpak (MIT), and Gerbrand Ceder (UC Berkeley) for helpful discussions during the past years. Finally, I would like to thank Tobias Morawietz (Stanford University) and Alexander Urban (Columbia University) for commenting on the manuscript.

ORCID iDs

Nongnuch Artrith https://orcid.org/0000-0003-1153-6583

References

[1] Schögl R 2015 Angew. Chem. Int. Ed. 54 3465–520
[2] Hwang J, Rao R R, Giordano L, Katayama Y, Yu Y and Shao-Horn Y 2017 Science 358 751–6
[3] Shao M, Chang Q, Dodelet J P and Chenitz R 2016 Chem. Rev. 116 3594–657
[4] Liu J, Liu Y, Liu N, Han Y, Zhang X, Huang H, Lifshitz Y, Lee S T, Zhong J and Kang Z 2015 Science 347 970–4
[5] Chen S, Takata T and Dom en K 2017 Nat. Rev. Mater. 2 17050
[6] Kim H, Kim M G, Jeong H Y, Nam H and Cho J 2015 Nano Lett. 15 2111–9
[7] Luntz A C, Voss J and Reuter K 2015 J. Phys. Chem. Lett. 6 4599–604
[8] Ma C, Cheng Y, Yin K, Luo J, Sharafi A, Sakamoto J, Li J, Moore K L, Dudney N J and Chi M 2016 Nano Lett. 16 7030–6
[9] Richards W D, Miara I J, Wang Y, Kim J C and Ceder G 2016 Chem. Mater. 28 266–73
[10] Vardar G et al 2018 Chem. Mater. 30 6529–76
[11] Zekoll S et al 2018 Energy Environ. Sci. 11 183–201
[12] Canepa P, Dawson J A, Sai Gautam G, Statham J M, Parker S C and Islam M S 2018 Chem. Mater. 30 3019–27
[13] Tian Y, Sun Y, Hannah D C, Xiao Y, Liu H, Chapman K W, Bo S H and Ceder G 2019 Joule 3 1037–50
[14] Graves C, Ebbesen S D, Jensen S H, Simonsen S B and Mogensen M B 2015 Nat. Mater. 14 239–44
[15] Irvine J T S, Neagu D, Verbraeken M C, Chatzichristodoulou C, Graves C and Mogensen M B 2016 Nat. Energy 1 13014
[16] Tsvetkov N, Li Q, Sun L, Crumlin E J and Yildiz B 2016 Nat. Mater. 15 1010–6
[17] Burke K 2012 J. Chem. Phys. 136 150901
[18] Becke AD 2014 J. Chem. Phys. 140 18A301
[19] Behrens M et al 2012 Science 336 893–7
[20] Kattel S, Ramírez P J, Chen J G, Rodríguez I A and Liu P 2017 Science 355 1296–9
[21] Butler K T, Davies D W, Cartwright H, Isayev O and Walsh A 2018 Nature 559 547–55
[22] Kitchin JR 2018 Nat. Catal. 1 230–2
[23] Sanchez-Lengeling B and Aspuru-Guzik A 2018 Science 361 360–5
[24] Ong SP 2019 Comput. Mater. Sci. 161 143–50
[25] Belayadi A, Mougari A and Zabat M 2019 J. Solid State Electrochem. 23 195–204
[95] Li W, Ando Y, Minamitani E and Watanabe S 2017 J. Chem. Phys. 147 214106
[96] Kondati Natarajan S and Behler J 2017 J. Phys. Chem. C 121 4368–83
[97] Quaranta V, Hellström M and Behler J 2017 J. Phys. Chem. Lett. 8 1476–83
[98] Hellström M, Quaranta V and Behler J 2019 Chem. Sci. 10 1232–43
[99] Quaranta V, Behler J and Hellström M 2019 J. Phys. Chem. C 123 1293–304
[100] Bates J, Dudney N J, Gruzalski G R, Zuhr R A, Choudhary A, Luck C F and Robertson J D 1992 Solid State Ion. 53–56 647–54
[101] Kuwata N, Iwagami N, Tanji Y, Matsuda Y and Kawamura J 2010 J. Electrochem. Soc. 157 A521–7
[102] Kuwata N, Lu X, Miyazaki T, Iwai Y, Tanabe T and Kawamura J 2016 Solid State Ion. 294 59–66
[103] Li W, Ando Y and Watanabe S 2017 J. Phys. Soc. Japan 86 104004
[104] Fujikake S, Deringer V L, Lee T H, Krynski M, Elliott S R and Csányi G 2018 J. Chem. Phys. 148 241714
[105] Artrith N, Urban A, Wang Y and Ceder G 2019 arXiv:1901.09272 [Cond-Mat Physics]
[106] Curtarolo S et al 2012 Comput. Mater. Sci. 58 218–26
[107] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 Comput. Mater. Sci. 68 314–9
[108] Pizzi G, Cepellotti A, Sabatini R, Marzari N and Kozinsky B 2016 Comput. Mater. Sci. 111 218–30
[109] Larsen A H et al 2017 J. Phys.: Condens. Matter 29 273902
[110] Tadmor E B, Elliott R S, Sethna J P, Miller R E and Becker C A 2011 JOM 63 17–17
[111] Becker C A, TavaZZa F, Trautzt Z T and Buarque de Macedo R A 2013 Curr. Opin. Solid State Mater. Sci. 17 277–83
[112] Artrith N, Urban A and Ceder G 2017 Phys. Rev. B 96 014112
[113] Ghasemi S A, Hofstetter A, Saha S and Goedecker S 2015 Phys. Rev. B 92 045131
[114] Faraji S, Ghasemi S A, Rostami S, Rasoulkhani R, Schaefer B, Goedecker S and Amsler M 2017 Phys. Rev. B 95 104105