Modelling and understanding battery materials with machine-learning-driven atomistic simulations

Volker L Deringer

Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, Oxford OX1 3QR, United Kingdom
E-mail: volker.deringer@chem.ox.ac.uk

Keywords: battery materials, energy storage, lithium-ion batteries, machine learning, materials modelling

Abstract

The realistic computer modelling of battery materials is an important research goal, with open questions ranging from atomic-scale structure and dynamics to macroscopic phenomena. Quantum-mechanical methods offer high accuracy and predictive power in small-scale atomistic simulations, but they quickly reach their limits when complex electrochemical systems are to be studied—for example, when structural disorder or even fully amorphous phases are present, or when reactions take place at the interface between electrodes and electrolytes. In this Perspective, it is argued that emerging machine learning based interatomic potentials are promising tools for studying battery materials on the atomistic and nanometre length scales, affording quantum-mechanical accuracy yet being many orders of magnitude faster, and thereby extending the capabilities of current battery modelling methodology. Initial applications to solid-state electrolyte and anode materials in lithium-ion batteries are highlighted, and future directions and possible synergies with experiments are discussed.

1. Introduction

Batteries are complex, highly specialised electrochemical reaction systems with multiple interdependent parts. To a large extent, their functionality originates on the atomic scale, and it is therefore important to correctly determine the atomistic structure and reactivity of all materials involved. Today, advanced experimental methods can give detailed structural insight into batteries during operation (operando); this includes x-ray and neutron diffraction, electron microscopy down to atomistic resolution, as well as nuclear magnetic resonance (NMR) and x-ray spectroscopy which both probe the local atomic environments [1–4]. As is often the case with structurally complex solids, each of these techniques adds a specific piece of the puzzle—but only their combination can provide the full picture, requiring time and expert knowledge. Even if the structure of a given battery material is accurately known, it is still more challenging to establish the connections between primarily structural aspects (e.g. the local environments of Li or Na ions intercalated in carbon-based anodes), the resulting electrochemical properties (manifested in the voltage profile during charging and discharging), and the material’s macroscopic behaviour (fatigue, failure, etc). And in addition to the materials themselves, their solid–solid or solid–electrolyte interfaces are just as important, because they can have a direct impact on macroscopic properties—say, in the development of all-solid-state batteries, where a loss of physical contact between different components due to structural inhomogeneity can lead to a loss of performance [5–7].

To complement the wide range of experimental characterisation tools in battery materials research, the atomistic structure and reactivity of these materials have been frequently studied with the help of computer simulations. The current primary methods of choice are based on density-functional theory (DFT) for simulation cells containing a few hundred or at most a few thousand atoms (corresponding to the Ångström and few-nm length scales), and on empirically fitted interatomic potential models (‘force fields’) for atomistic simulations of substantially larger systems. A general introduction to atomistic battery materials modelling may be found in Ref. [8], and a very recent overview has been provided, for example, in Ref. [9]. The aim of such computational studies is two-fold: first, to develop an accurate structural model of a given cathode,
electrolyte, or anode material (with ‘accurate’ here taken to mean ‘consistent with available experimental data’, including static and dynamic properties); second, to use this structural knowledge to understand the results of experiments, such as the Li-ion conductivity of a solid-state electrolyte or the voltage profile during charging and discharging of a candidate battery. Quantum-mechanically based DFT methods largely offer the required accuracy, to within a few kilojoules per mole or better, and they are predictive, in that they may suggest new experiments and research directions. Prominent examples include DFT predictions of a variety of previously unknown, stable and metastable crystal structures, which were subsequently realised by experiment in many cases [10]. Yet, the computational cost of DFT is so large that one remains restricted to small simulation cells and often to idealised model systems. For example, the intercalation mechanisms of Li and Na ions in graphite, and the question why Na intercalates so poorly, have been widely studied by DFT computations on small systems, including an in-depth analysis of the various competing energetic contributions [11]. In contrast, similar studies of disordered hard-carbon anodes require much larger model systems owing to the lack of long-range order, and therefore such simulations are still in their infancy.

The aim of this short Perspective is to highlight an emerging, complementary approach in atomistic materials modelling which is thought to be of interest for battery research: namely, the creation of fast and accurate interatomic potentials by machine learning (ML) from quantum-mechanical reference data. With a strict focus on batteries, this work naturally cannot cover all the exciting developments in the field more generally (nor can it make reference to all of the growing body of literature), and the interested reader is referred to recent reviews of ML potentials [12–14] and to more general overviews concerned with ML for materials science [15–17] and with materials design [18]. There are also recent reviews of diverse emerging ML methods being increasingly applied to energy materials, including, but not restricted to, battery research [19–22]. It is hoped that the present work will highlight capabilities of ML-driven interatomic potentials (and possible future directions) to researchers in the wider field of energy storage, including both experimentally and computationally focused readers—and that it might thereby encourage new applications of ML potentials and enable further insight into the structures and properties of battery materials.

2. What are ML potentials?

Machine learning aims to extract information from (very) large datasets, and many refer to it as the ‘fourth paradigm’ of science, following on from the earlier empirical, model-based, and computational paradigms [23]. Whilst the earliest ML approaches date back to the middle of the 20th century, the availability of immense amounts of data has led to a spectacular rise in interest in the recent years. ML-based interatomic potentials are an example of a supervised learning problem, more specifically, an applied regression task: given a dataset of accurately computed energies and forces for selected points on a potential-energy surface (PES), make the best possible fit to these data, without assuming a specific functional form of the PES. The last point marks the principal difference compared to existing, empirically fitted interatomic potential models which have been used for decades. Empirical potentials rely on simple physical models (Coulomb, Lennard–Jones, Buckingham, …) which parameterise the PES as a function of bond lengths, angles, and so on; they are therefore very fast to evaluate, and have been widely used to run large-scale molecular-dynamics (MD) simulations even for complex battery chemistries [8]. In contrast, ML potentials do not pre-define the shape of the interaction—so that even the simplest features of the PES (say, the repulsion between atoms at short distance) must be ‘learned’ from suitable reference data or otherwise included in the model construction [24]. A more general view of this emerging field was given, e.g. in a recent overview article which includes applications of ML potentials in diverse areas of materials research [13].

Figure 1 provides an overview of the most central concepts [13], using here the example of the aforementioned carbon anode materials: ‘learning’ from small unit cells which sample the PES at properly chosen points, such as a distorted graphite structure and more disordered supercells based on it, ML potentials enable the study of much larger systems—here, a 1000-atom cell, which describes a material that resembles graphite but is strongly disordered [25]. Indeed, systems with tens of thousands of atoms are within reach for ML potentials, and their simulation-cell lengths approach the size of nanostructures in devices: experiment and highly accurate simulation are now looking at the same thing, in terms of length scales [26].

The construction of the reference database, highlighted in yellow in figure 1, is one of the three central tasks (and often, one of the barriers) in developing any ML potential, alongside the representation of atomic structure and the regression itself [13]. So far, reference databases have been widely built by hand: for example, running MD simulations with evolving versions of a potential and ‘feeding back’ structural snapshots for single-point energy and force evaluation, thereby gradually growing the database [29, 30]. More recent, ongoing research in the community is concerned with the question how such databases might be built and optimised in a more automated fashion [31–34], which would promise an increased usefulness.
Figure 1. Machine-learning-based interatomic potentials. Top: The general idea and the main required components. First, a reference database of small structural models (here, represented by a distorted graphite structure) is constructed and single-point quantum-mechanical computations are carried out; the database is then typically grown and curated over time. Having sampled the high-dimensional potential-energy surface (PES) at selected points and represented the atomic environments in a suitable mathematical form, the fit itself is carried out—in most cases, this is based on the assumption that the total energy, $E$, can be decomposed into atomic contributions, $\varepsilon_j$ [27, 28]. This way, much larger systems can be studied, exemplified here by a disordered graphite-like structure containing $\approx 1000$ atoms (drawn with data from Ref. [25]). Bottom: Three major families of regression methods that are currently used for fitting ML potentials, as described in the text. Adapted from Ref. [13] with permission. Original figure copyright © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

(through wider availability of high-quality databases) to many fields of applied research. The challenges of database construction, with a specific focus on energy materials including batteries, have been emphasised in a recent Topical Review in the present journal [20]: highly complex interfaces must be represented by small structural models which nonetheless need to adequately cover all parts of the ‘real-world’ system, and an improved standardisation of fitting databases and protocols would benefit developments across the field [20].

The three families of ML regression methods that are currently most widely used for fitting potentials are sketched in the lower three panels in figure 1. Artificial neural networks (ANNs) [12] are sets of mathematical functions that mimic the network of neurons in the brain; the ‘training’ of an ANN model consists of adjusting the weights, $w$, between these functions by adapting them to a set of known data (the reference database). In 2007, Behler and Parrinello showed how ANN potentials for high-dimensional (condensed-phase) systems can be constructed, by using a suitable representation of the local atomic structure and by writing the total energy as a sum of ‘machine-learned’ local energies [27], the latter approximation now being central to other fitting methods as well. Today, a wide variety of developments based on ANNs has been reported, including, e.g. a broadly applicable ML potential for organic molecules [35] or a recent class of ‘deep’ neural network models [36]. The second family of methods used to make ML potentials is based on kernel (similarity) functions, which compare given known and new atomic environments. In 2010, Bartók et al introduced the Gaussian approximation potential (GAP) [28] framework, which utilises Gaussian process regression and a suitable choice of representation [37]; other methodological developments related to kernel methods include gradient-domain ML [38] and automated ‘on-the-fly’ potentials coupled to DFT-MD [39]. The third family of methods is based on linear fitting: the idea here is to express an atomic structure by appropriately chosen basis functions, $B$, which enter the energy model using (only) linear terms; this includes the spectral neighbor analysis potential (SNAP) [40] and moment tensor potential (MTP) [41] approaches. All three classes of fitting methods sketched in figure 1 can yield potentials with high performance, as shown in a recent benchmark study across different implementations [42]; all three are under continuous improvement, each in specific ways. All three classes of methods and derived ML potentials have begun to be used for battery-related studies very recently, and these emerging applications are discussed in the following section.
ML potentials are on the verge of being widely used—and openly available source codes and potential parameters, as well as integration with existing simulation software, are all poised to accelerate their uptake in the community. A notable example are interfaces to the widely used (and similarly freely available) LAMMPS MD code [43], which have now been provided for several classes of ML potentials; therefore, an experienced user might need to do no more than swap out a few lines of input to run their existing simulation setup with a new ML potential. It is emphasised, again, that these potentials are still much more computationally costly than their empirically fitted counterparts. This trade-off was recently discussed by de Tomas et al [44] in a systematic benchmark study of carbon potentials (focusing on disordered, partly graphitised forms, which are relevant to battery research as well); among the methods tested, an ML potential showed high performance but was also by far the most expensive [44], and a good empirical potential may be a much faster option in many cases, as shown by previous successful applications of such potentials to graphitic-like carbon (see, e.g. [45]). Indeed, battery research includes larger-scale simulation tasks which will be out of reach for many currently available ML potentials, but accessible to fast empirical models. On the other hand, there are complex atomic-scale problems (e.g. bond breaking and making during thin-film growth [46]) for which an ML potential might well be worth the additional cost. It is about the right choice of computational tools, as always.

3. What has been done?

The ML-driven atomistic modelling of battery materials is at a very early stage, and it has largely focused on selected prototype materials so far. Figure 2(a) provides a (highly simplified) overview of different components in a battery. Blue circles indicate research directions which have already been pursued with ML potentials; green circles show areas where, as of this writing and to the author’s best knowledge, no directly battery-related applications of ML potentials have been published, but where the required methodology or large parts of it are principally available. Walking from left to right through the schematic in figure 2(a), anode materials have been studied with several different types of ML potential frameworks; of particular interest here are those materials which have too complex structures to be easily described by DFT, such as disordered carbon- and silicon-based anodes. Whilst previous ML-driven simulations have focused on bulk anode phases, e.g. Li,Si [47, 48], the associated electrode–electrolyte interfaces are similarly important, as mentioned, e.g. by Huang et al [49], and accurate atomistic models of such interfaces are often plainly out of reach for DFT. For the electrolyte itself, there are two primary classes, liquid and solid—of which the latter garnered some attention from the ML potential community, as discussed below. Finally, the cathode materials are typically strongly ionic solids, ranging in structural complexity from the iconic LiCoO$_2$ [50] to nanostructured metal oxides [51], all of which pose challenges of their own to ML potential development (see, e.g. Ref. [52] for an early study of a prototypical metal oxide). The present section will focus on currently reported applications to battery materials (i.e. on the blue circles in figure 2(a)), whereas the following section will give an outlook on possible future directions.

Figure 2(b) shows an example of the model solid-state electrolyte, Li$_3$PO$_4$, in which the Li$^+$ ions move between rigid [PO$_4$]$^{3-}$ tetrahedra, representative of many more complex electrolyte materials. The material was studied using MD simulations with a neural-network type potential [53], illustrating how ML potentials enable the DFT-accurate treatment of much larger simulation cells (often large enough that the prediction of a property of interest does not change substantially when the cell size is increased further [55]). This example is also representative of how ML potentials can improve not only the length but also the simulation time scales: sometimes, the challenge is not in the mere number of atoms, but in being able to perform many millions (and more) of individual MD steps—e.g. to create high-quality amorphous silicon structures by simulated slow quenching from the melt [56], or to obtain accurate diffusion coefficients in solid-state electrolytes, as studied in Ref. [53] for Li$_3$PO$_4$. The authors also provide an example for how simulation results are typically validated in the ML potential community: by comparison to small-scale DFT simulations, for example, in terms of computed radial distribution functions (which provide a simplified but instructive ‘fingerprint’ for the local structure) [53].

A similar validation strategy, although with a different ML fitting framework and application case, was followed by Fujikake et al [57] who introduced a GAP model for Li diffusion in disordered carbon structures. In that case, the emphasis was on how one can build difference potentials by subtracting previously known interactions (those between carbon atoms) and fitting a separate ML model for the energy and force differences resulting from Li intercalation into the host structure. That said, the validation of ML potentials, not only in terms of fitting accuracy and computed observables, but also in terms of how resilient they are to failure, remains an outstanding challenge—especially because of the absence of a physically motivated functional form (as would be present in empirical interatomic potentials).
Figure 2. Modelling battery materials with ML-based potentials. (a) A highly simplified sketch of different components of a battery, each requiring different atomistic modelling approaches. (b) Simulation cells representing structurally disordered Li$_3$PO$_4$, indicating the step up in system size that is possible by moving from DFT to ML potentials [53]. Reprinted from [53], with the permission of AIP Publishing. (c) Computational prediction of Li-ion diffusivity in selected candidate cathode coating (or electrolyte) materials [54], each indicated by a different colour. The plot compares the authors’ results of DFT-driven ab initio MD (open symbols) to those of an MTP-driven ‘learn on the fly’ scheme (filled symbols); the latter give access to much faster simulations, and therefore allow the study of diffusion at lower temperatures. Adapted with permission from [54] Copyright 2020 American Chemical Society. (d) Global exploration of Li–Si alloys using neural-network potential modelling [47]: here, an ANN scheme is coupled to two different computational approaches, a genetic algorithm (top) and MD (bottom), and the results are then translated into simulated voltage–filling curves with the help of DFT, which may be compared to experimental data, as shown on the right-hand side (see [47] and references therein). Plots reprinted from [47], with the permission of AIP Publishing.

The long-term value of ML potentials in battery materials research hinges on their ability to predict not only structural properties, but also those properties that are more directly tied to application. An example of this was recently demonstrated for cathode coating materials (many of which correspond to solid electrolytes, although the specific requirements are different in detail) [54]. Existing high-throughput studies aiming to find such materials, e.g. in Ref. [58], are typically concerned with materials properties as a function of the chemical composition, and they do not normally take the atomistic structure into account explicitly (this is well justified when isostructural materials are studied, but may pose challenges for new compounds whose structure is a priori unknown). One of the reasons is that a fully structure-specific treatment, by running long...
DFT simulations in an adequately sized supercell, measuring the atomic displacements over time, and fitting diffusivities to the resulting data, is often impracticable. Figure 2(c) now shows an ML-driven approach to this problem, for a range of materials from very good ($\beta$-Li$_3$PS$_4$, activation barrier of 0.23 eV) to rather poor ion conductors (Li$_2$GeO$_3$, 1.14 eV), all treated in the same computational framework either by DFT-driven ('ab initio') molecular dynamics (AIMD), or with a 'learn-on-the-fly' (LOTF) ML model [54]. Having access to simulation tools which are orders of magnitude faster than previously available makes it possible to study lower temperatures, at which atomic 'jumps' occur less frequently and therefore longer simulations are needed to predict the diffusivity. For example the temperature range in which simulation results are shown in figure 2(c) ends at 800 K for AIMD, but reaches down to 300 K for the LOTF MD (the specific range for these and other materials is given in Ref. [54]). The authors reported a total simulation time of more than 1300 nanoseconds across a variety of materials that they studied, corresponding to hundreds of millions of individual simulation steps. The work in Ref. [54] was carried out using the MTP approach [41], but the authors' general findings are likely transferable to other suitable ML potential fitting methods.

On the anode side (figure 2(a)), the intercalation of Li ions into graphite is a 'textbook' reaction in the context of batteries—and somewhat more generally, the intercalation of Li into various nanostructured (nanotubes etc) and disordered forms of carbon ('hard' carbons) has been widely studied by DFT [9]. Such studies have also been extended to other alkaline and alkaline-earth ions, because hard carbons are (for example) a promising contender for Na-ion battery anodes [59]. It was previously shown that disordered forms of carbon can be accurately described by ML potentials, with evidence including various structural aspects [30, 44] and the demonstrated ability to describe the growth of thin films in quantitative agreement with experimentally observed 'sp$^3$' content [46]. Among a range of emerging applications, GAP-driven simulations have been used to generate small-scale structural models of porous carbon at different densities [25] and to survey alkali metal intercalation in such structures, based on a subsequent combination with DFT energy and charge analyses (which will be discussed in the section below) [49].

ML potentials are widely coupled to MD simulations, but they may be used in combination with any method which requires energies as a function of atomic coordinates. In the context of battery anodes, Artrith and colleagues demonstrated the combination of ML (in that case, ANN) potentials with different ways to explore and quantify the local structures of amorphous Li–Si phases (figure 2(d)) [47]. In addition to MD, this included sampling using a genetic algorithm based on a neural-network energy prediction, and both sampling methods allowed the authors to probe many candidate structures at a given composition (x, Li filling, shown on the horizontal axis in the panels in figure 2(d)). Once the most favourable structures at a given x and their respective energies are known, the associated voltage–filling curves can be computed [9], providing one of the primary quantities of interest for battery materials. In another study of Li–Si phases, which introduced a more general 'implanted' neural-network methodology for the different parts of compositional space, the diffusivity was investigated and tested against experiments [48]—which is directly relevant to charging and discharging.

It is likely that elemental anodes, due to their relatively simple compositions but complex atomistic structures, will continue to be valuable targets for ML-driven modelling, extending previous successful studies based on DFT-driven structure exploration [60, 61] whilst reaching larger simulation length scales than previously available (exemplified recently, e.g. in Ref. [62]). It is emphasised, again, that the present state of the field in figure 2(a) refers specifically to atomistic modelling—other ML methods, such as the generation of simulated microstructures from generative adversarial networks (GANs) are beginning to emerge as well [63], but are outside of the scope of the present Perspective.

4. What is next?

ML potentials are an increasingly popular approach to atomistic simulations, but compared to the established methodologies in battery modelling, they are only just getting started. In the future, there might be two directions for the more general uptake of ML potentials in the wider community [13]. On the one hand, carefully curated fitting databases and potentials are beginning to be provided online, and such 'general-purpose' potentials [24, 64] would be expected to be applicable (in the sense of reaching close-to DFT accuracy) for a wide range of relevant problems and to be at least robust (not showing unphysical behaviour) for any reasonable atomistic structure at a given composition. The most broadly useful potentials will be those which deal with widely studied material systems—in the context of batteries, this is expected to include relatively simple chemical systems which nonetheless continue to pose open questions, such as the most prototypical oxide cathode materials. It is also interesting to mention the entirely automated ('on-the-fly') learning of interatomic potentials, which was recently coupled to a widely used DFT code [39]. This approach gives the user less control over the parameterisation of the potential, but in turn requires no (or almost no) input from their side, and it does not require the transferability of a general-purpose ML
potential—instead allowing the user to speed up a simulation of a given, perhaps very specific, atomistic system which has been ‘kick-started’ by DFT.

New developments in ML potential fitting methodology are expected to substantially accelerate research in the field (quite literally, through faster algorithms and codes) and to enable applications to a wider range of battery-related problems. Most of the discussion within this short Perspective has focused on condensed phases, although molecular modelling is similarly beginning to benefit very strongly from ML potentials, often reaching far beyond what standard DFT approaches can do [64, 65]. In the context of batteries, this promises a more accurate description of liquid electrolytes, and in particular of diverse chemical reactions taking place at solid–electrolyte interfaces, which involve the making and breaking of covalent bonds and are therefore not always easily accessible to existing simulation methods. One specific example are those adverse reactions that must be kept under control in lithium–air batteries [66].

In terms of materials to be studied, one may now anticipate the step up towards more chemically complex compositions, and ultimately to the discovery of new materials before they have been synthesised. Looking back at figure 2(a), one area that has so far been less widely explored with ML potentials is that of bulk and nanostructured cathode materials; indeed, not many ML-driven studies of complex oxides have been carried out, partly due to structural and compositional complexity, partly due to the presence of long-range electrostatic interactions. An example where all these challenges come together are a class of recently reported low-symmetric block and bronze-like phases for lithium-ion storage [67]: will ML potentials be able to treat such materials in the future? In terms of technical ability, early methodological developments have been made—including, for example, the ‘learning’ of environment-dependent charges with separate ANNs [52], or the construction of primarily charge-based force fields for strongly ionic materials [68]. In terms of the chemical complexity, the coupling to existing materials databases and high-throughput workflows may prove highly beneficial [18, 69]—perhaps accelerated further by ML predictions of electrode voltages [70] or other application-relevant properties, which might narrow down the range of candidate compositions and structures.

In terms of methodological developments, it is also interesting to envision new synergies between established DFT methods and experimental data, both being enabled by ‘bridging’ through ML potentials. Figure 3(a) sketches a few connections of this type which are beginning to emerge, and which are expected to be relevant in the future. One recurring theme is the generation of systematic libraries of small, representative structural models which can then be studied further with regard to electronic structure and bonding (figure 3(b)) [25, 55]. A recent example in the wider field of energy materials, showcasing the usefulness of combining ML-based and existing simulation methods, is the computational modelling of CO₂ reduction catalysts based on chemically modified carbonaceous materials [71]. Candidate structures were explored in large-scale DFT computations [71], starting from an existing library of pure amorphous carbon surface models, the latter having been generated previously with an ML potential [55]. Another energy-related application case is a study of porous carbons in supercapacitors and specifically the effect of pore size on diffusivity [72], again building partly on ML-generated structural models [25] and moving on to much more complex chemical systems.

Turning to a possible direct synergy with experiments (figure 3(a)), it would now be desirable to carry out systematic experimental benchmark studies: for example, one might synthesise silicon anodes with various degrees of local order in the laboratory, allowing the community to test ML-driven predictions with regard to both the pure phases [56] and the Li-intercalated ones [47, 48]. More ‘tailored’ synthetic carbon materials, with parameters (such as the annealing temperature) tuned in fine steps, have begun to be reported very recently [75], and studies of this type would be expected to be highly valuable to the developers of ML potentials (and other atomistic simulation methods). A somewhat similar request for systematic experimental benchmark studies has been formulated recently for molecular quantum chemistry [76].

A clear long-term perspective is the connection of DFT, ML potentials, and experiments to achieve a combined structural solution, and thereby to arrive at a more holistic understanding of the atomistic and nanometre structure of batteries. Local structural probes including pair distribution function curves, as exemplified in figure 3(c) for Na-ion batteries [74], can be compared to the outcome of accurate simulations. Comparison with solid-state NMR data has been used as a means to validate ML-computed structures [56], and this technique is very widely employed in battery research as well [2]. For example, NMR experiments have been used to trace the chemical nature of Li [77] and Na [74] ions during insertion in porous carbons, both ex situ and operando (the latter being exemplified in figure 3(d)), and these measurements have begun to be linked to ML-driven computational work and DFT-based bonding analyses [25]. In the future, the tripartite combination of accurate input structures (from ML-driven simulations), accurate methodology for predicting the NMR parameters themselves (based on DFT methods, as reviewed in Ref. [78]), and ‘real-world’ experimental data is expected to become a powerful tool for battery research. In this context, one may mention recent progress towards predicting x-ray spectroscopy (XAS and XPS) fingerprints using
Machine-learned potentials as a new addition to the community’s toolkit for battery materials research. (a) A perspective for the interplay of ML-based potentials, established DFT methods, and experimental techniques in the quest for characterising and understanding battery materials on the atomic scale. (b) A recent example of how atomistic modelling driven by ML potentials can be interfaced to DFT bonding analyses. After a library of disordered carbon structures has been generated in long ML-driven MD runs, DFT computations are carried out to study the intercalation behaviour of Li ions. The plot is a distance map based on structural similarities, as detailed in Ref. [49] and references therein—colour-coded by the charges on individual atomic environments (from Li⁺ almost down to metallic Li⁰), which can be computed using DFT. (c) Pair distribution functions, plotted with data from Ref. [74]. Arrows indicate the appearance of subtle features on the nanometre scale due to Na⁺ (de-) intercalation. (d) Operando $^{23}$Na NMR measurements for the same system, showing the cyclic appearance and disappearance of a more metallic-like signal in the measured chemical shift during operation. Adapted from Ref. [74]—Published by The Royal Society of Chemistry. These experimental probes for the local structure can now be compared with the outcome of accurate, ML-driven atomistic simulations—both regarding structural models (panel c) and the physical property predictions that may be based on it (panel d).

The 2019 Nobel Prize in Chemistry to Goodenough, Whittingham, and Yoshino was awarded for fundamental studies that later led to an industrial-scale breakthrough. Many of the fundamental studies in battery research have been concerned with structure and reaction mechanisms on the atomistic scale. We now have more accurate atomistic insight into materials than ever, with modelling and experiment going hand in hand, and ML becoming applicable to ‘real-world’ problems in the field. With accurate ML-driven potentials and simulations in reach, there is a hope to thoroughly understand not only idealised crystal structures but grains and their interfaces, as well as partly or fully amorphous phases, all of which are of primary importance for batteries. Without doubt, it is exciting to envision that in the coming years, ML potentials may become a firmly integrated part of the battery modellers’ toolkit—and to see what this will mean for the coming decade of energy materials development.
Acknowledgments

The author thanks J. George and O. Pecher for helpful comments on the manuscript and useful discussions, and the Leverhulme Trust for an Early Career Fellowship. The structural drawings in figure 1 were created using VESTA [82].

ORCID iD

Volkert L. Deringer  @https://orcid.org/0000-0001-6873-0278

References

[1] Lu J, Wu T and Amine K 2017 State-of-the-art characterization techniques for advanced lithium-ion batteries Nat. Energy 2 17011
[2] Pecher O, Carretero-González J, Griffith K J and Grey C P 2017 Materials’ methods: NMR in battery research Chem. Mater. 29 213–42
[3] Liu D et al 2019 Review of recent development of in situ/operando characterization techniques for lithium battery research Adv. Mater. 31 1806620
[4] Liang G, Didier C, Guo Z, Pang W K and Peterson V K 2020 Understanding rechargeable battery function using in operando neutron powder diffraction Adv. Mater. 32 1904528
[5] Janek J and Zeier W G 2016 A solid future for battery development Nat. Energy 1 16141
[6] Famprikis T, Canepa P, Dawson J A, Islam M S and Masquelier C 2019 Fundamentals of inorganic solid-state electrolytes for batteries Nat. Mater. 18 1278–91
[7] Hatzell K B et al 2020 Challenges in lithium metal anodes for solid-state batteries ACS Energy Lett. 5 922–34
[8] Islam M S and Fisher C A J 2014 Lithium and sodium battery cathode materials: computational insights into voltage, diffusion and nanostructural properties Chem. Soc. Rev. 43 185–204
[9] Van der Ven A, Deng Z, Banerjee S and Ong S P 2020 Rechargeable alkali-ion battery materials: theory and computation Chem. Rev. 120 6977–7019
[10] Oganov A R, Pickard C J, Zhu Q and Needs R J 2019 Structure prediction drives materials discovery Nat. Rev. Mater. 4 331–48
[11] Liu Y, Merinov B V and Goddard W A 2016 Origin of low sodium capacity in graphite and generally weak substrate binding of Na and Mg among alkali and alkaline earth metals Proc. Natl Acad. Sci. USA 113 3735–9
[12] Behler J 2017 First principles neural network potentials for reactive simulations of large molecular and condensed systems Angew. Chem., Int. Ed. 56 12828–40
[13] Deringer V L, Caro M A and Csányi G 2019 Machine learning interatomic potentials as emerging tools for materials science Adv. Mater. 31 1902765
[14] Noé F, Tkatchenko A, Müller K-R and Clementi C 2020 Machine learning for molecular simulation Annu. Rev. Phys. Chem. 71 361–90
[15] Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A and Kim C 2017 Machine learning in materials informatics: recent applications and prospects Npj Comput. Mater. 3 54
[16] Butler K T, Davies D W, Cartwright H, Isayev O and Walsh A 2018 Machine learning for molecular and materials science Nature 559 547–55
[17] Schmidt J, Marques M R G, Botti S and Marques M A L 2019 Recent advances and applications of machine learning in solid-state materials science Npj Comput. Mater. 5 83
[18] Alberi K et al 2019 The 2019 materials by design roadmap J. Phys. D: Appl. Phys. 52 013001
[19] Gu G H, Noh J, Kim I and Jung Y 2019 Machine learning for rechargeable alkali-ion battery materials: theory and computation Chem. Rev. 120 6977–7019
[20] Arthir N 2019 Machine learning for the modeling of interfaces in energy storage and conversion materials J. Phys. Energy 1 032002
[21] Chen C, Zuo Y, Ye W, Li X, Deng Z and Ong S P 2020 A critical review of machine learning of energy materials Adv. Energy Mater. 10 1903242
[22] Akyol M, Herring P and Anapolsky A 2020 Machine learning for continuous innovation in battery technologies Nat. Rev. Mater. 5 725–7
[23] Agrawal A and Choudhary A 2016 Perspective: materials informatics and big data: realization of the “fourth paradigm” of science in materials science Adv. Mater. 28 403208
[24] Bartók A P, Kermode J, Bernstein N and Csányi G 2018 Machine learning a general-purpose interatomic potential for silicon Phys. Rev. X 8 041048
[25] Deringer V L, Merlet C, Hu Y, Lee T H, Kattirtzi J A, Pecher O, Csányi G, Elliott S R and Grey C P 2018 Towards an atomistic understanding of disordered carbon electrode materials Chem. Commun. 54 5988–91
[26] Gabardi S, Baldi E, Bosoni E, Campi D, Caravati S, Sossos G C, Behler J and Bernasconi M 2017 Atomistic simulations of the crystallization and aging of GeTe nanowires J. Phys. Chem. 121 23827–38
[27] Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces Phys. Rev. Lett. 98 146401
[28] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons Phys. Rev. Lett. 104 136403
[29] Sossos G C, Miceli G, Caravati S, Behler J and Bernasconi M 2012 Neural network interatomic potential for the phase change material GeTe Phys. Rev. B 85 174103
[30] Deringer V L and Csányi G 2017 Machine learning based interatomic potential for amorphous carbon Phys. Rev. B 95 094203
[31] Botu V, Batra R, Chapman J and Ramprasad R 2017 Machine learning force fields: construction, validation, and outlook J. Phys. Chem. 121 511–22
[32] Guhaev K, Podryabinkin E V, Hart G L W and Shapeev A V 2019 Accelerating high-throughput searches for new alloys with active learning of interatomic potentials Comput. Mater. Sci. 156 148–56
[33] Bernstein N, Csányi G and Deringer V L 2019 De novo exploration and self-guided learning of potential-energy surfaces Npj Comput. Mater. 5 99
[34] Zhang L, Lin D-Y, Wang H, Car R and E W 2019 Active learning of uniformly accurate interatomic potentials for materials simulation Phys. Rev. Mater. 3 023804
[35] Smith J S, Isayev O and Roitberg A E 2017 ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost Chem. Sci. 8 3192–203
[36] Zhang L, Han J, Wang H, Car R and E W 2018 Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics Phys. Rev. Lett. 120 143001
[37] Bartók A P, Kondor R and Csányi G 2013 On representing chemical environments Phys. Rev. B 87 184115
[38] Chmiela S, Tkatchenko A, Sauceda H E, Poltarsky I, Schütt K T and Müller K R 2017 Machine learning of accurate energy-conserving molecular force fields Sci. Adv. 3 e1603015
[39] Jinnouchi R, Lahneiter J, Karsai F, Kresse G and Bokdam M 2019 Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with bayesian inference Phys. Rev. Lett. 122 225701
[40] Thompson A R, Swier L R, Trott C R, Foiles S M and Tucker G J 2015 Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials J. Comput. Phys. 285 316–30
[41] Shapere A V 2016 Moment tensor potentials: a class of systematically improvable interatomic potentials Multiscale Model. Simul. 14 1155–73
[42] Zuo Y et al 2020 Performance and cost assessment of machine learning interatomic potentials J. Phys. Chem. 124 731–45
[43] Plimpton S 1995 Fast parallel algorithms for short-range molecular dynamics J. Comput. Phys. 117 1–19
[44] de Tomás C, Aghajamali A, Jones J L, Lim D J, López M J, Suarez-Martínez I and Marks N A 2019 Transferability in interatomic potentials for carbon Carbon 155 624–34
[45] Shellie T B et al 2018 Graphitization of glassy carbon after compression at room temperature Phys. Rev. Lett. 120 215701
[46] Caro M A, Deringer V L, Koskinen J, Laurila T and Csányi G 2018 Growth mechanism and origin of high sp³ content in tetrahedral amorphous carbon Phys. Rev. Lett. 120 166101
[47] Artrith N, Urban A and Ceder G 2018 Constructing first-principles phase diagrams of amorphous Li$_x$Si using machine-learning-assisted sampling with an evolutionary algorithm J. Chem. Phys. 148 241711
[48] Onat B, Cubuk E D, Maloney B D and Kaxiras E 2018 Implanted neural network potentials: application to Li-Si alloys Phys. Rev. B 97 094106
[49] Huang J-X, Csányi G, Zhao J-B, Cheng J and Deringer V L 2019 First-principles study of alkali-metal intercalation in disordered carbon materials J. Mater. Chem. A 7 19070–80
[50] Mizushima K, Jones P C, Wiseman P J and Goodenough J B 1980 Li$_x$CoO$_2$ (0<x≤1): a new cathode material for batteries of high energy density Mater. Res. Bull. 15 783–9
[51] Wang Y and Cao G 2008 Developments in nanostructured cathode materials for high-performance lithium-ion batteries Adv. Mater. 20 2251–69
[52] Artrith N, Morawietz T and Behler J 2011 High-dimensional neural-network potentials for multicomponent systems: applications to zinc oxide Phys. Rev. B 83 153101
[53] Li W, Ando Y, Minamitani E and Watanabe S 2017 Study of Li atom diffusion in amorphous Li$_x$PO$_4$ with neural network potential J. Chem. Phys. 147 214106
[54] Wang C, Aoyagi K, Wisesa P and Mueller T 2020 Lithium ion conduction in cathode coating materials from on-the-fly machine learning Chem. Mater. 32 3741–52
[55] Deringer V L, Caro M A, Jana R, Aarva A, Elliott S R, Laurila T, Csányi G and Pustelka L 2018 Computational surface chemistry of tetrahedral amorphous carbon by combining machine learning and density functional theory Chem. Mater. 30 7438–45
[56] Deringer V L, Bernstein N, Bartók A P, Cliffe M J, Kerber R N, Marbella L E, Grey C P, Elliott S R and Csányi G 2018 Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics J. Phys. Chem. Lett. 9 2879–83
[57] Fujikake S, Deringer V L, Lee T H, Krynski M, Elliott S R and Csányi G 2018 Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures J. Chem. Phys. 148 241714
[58] Akyol M, Kim S, Hegde V I, Snysdacker D, Lu Z, Hao S, Kirklin S, Morgan D and Wolverton C 2016 High-throughput computational design of cathode coatings for Li-ion batteries Nat. Commun. 7 13779
[59] Irizarri E, Ponrouch A and Palacin M R 2015 Review—hard carbon negative electrode materials for sodium-ion batteries J. Electrochem. Soc. 162 A2476–82
[60] Mayo M, Griffith K J, Pickard C J and Morris A J 2016 Ab initio study of phosphorus anodes for lithium- and sodium-ion batteries Chem. Mater. 28 2011–21
[61] See A et al 2014 Ab initio structure search and in situ 3Li NMR studies of discharge products in the Li-S battery system J. Chem. Soc. Chem. Commun. 136 16368–77
[62] Deringer V L, Pickard C J and Proserpio D M 2020 Hierarchically atomized allotropes of phosphorus from data-driven exploration Angew. Chem., Int. Ed. 59 15880–8
[63] Gayon-Lombardo A, Moser L, Brandon N P and Cooper S J 2020 Pores for thought: generative adversarial networks for stochastic reconstruction of 3D multi-phase electrode microstructures with periodic boundaries npj Comput. Mater. 6 82
[64] Smith J S, Neben B T, Zabatyuk R, Lubbers N, Devereux C, Barros K, Trettia S, Isayev O and Roitberg A E 2019 Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning Nat. Commun. 10 29032
[65] Chmiela S, Sauceda H E, Müller K R and Tkatchenko A 2018 Towards exact molecular dynamics simulations with machine-learned force fields Nat. Commun. 9 3887
[66] Aurbach D, McCloskey B D, Nazar L F and Bruce P G 2016 Advances in understanding mechanisms underpinning lithium–air batteries Nat. Energy 1 16128
[67] Griffith K J, Wiaderek K M, Gibin G, Marbella L E and Grey C P 2018 Niobium tungsten oxides for high-rate lithium-ion energy storage Nature 559 536–43
[68] Ghasedi S A, Hofstetter A, Saha S and Goedecker S 2015 Interatomic potentials for ionic systems with density functional accuracy based on cluster densities obtained by a neural network Phys. Rev. B 92 045121
[69] George J, Hautier G, Bartók A P, Csányi G and Deringer V L 2020 Combining phonon accuracy with high transferability in Gaussian approximation potential models J. Chem. Phys. 153 044104
[70] Joshi R P, Eichkolt J, Li L, Fornari M, Barone V and Peralta J E 2019 Machine learning the voltage of electrode materials in metal-ion batteries ACS Appl. Mater. Interfaces 11 18494–503
[71] Wang X et al 2020 Efficient electrically powered CO$_2$-to-ethanol via suppression of deoxygenation Nat. Energy 5 478–86
[72] Lahrar E H, Belhoub A, Simon P and Merlet C 2020 Ionic liquids under confinement: from systematic variations of the ion and pore sizes toward an understanding of the structure and dynamics in complex porous carbons *ACS Appl. Mater. Interfaces* 12 1789–98

[73] Ertural C, Steinberg S and Dronskowski R 2019 Development of a robust tool to extract Mulliken and Löwdin charges from plane waves and its application to solid-state materials *RSC Adv.* 9 29821–30

[74] Stratford J M, Allan P K, Pecher O, Chater P A and Grey C P 2016 Mechanistic insights into sodium storage in hard carbon anodes using local structure probes *Chem. Commun.* 52 12430–3

[75] Kubota K *et al* 2020 Structural analysis of sucrose-derived hard carbon and correlation with the electrochemical properties for lithium, sodium, and potassium insertion *Chem. Mater.* 32 2961–77

[76] Mata R A and Suhm M A 2017 Benchmarking quantum chemical methods: are we heading in the right direction? *Angew. Chem., Int. Ed.* 56 11011–8

[77] Letellier M, Chevallier F, Clinard C, Frackowiak E, Rouzaud J-N, Béguin F, Morcrette M and Tarascon J-M 2003 The first in situ Li nuclear magnetic resonance study of lithium insertion in hard-carbon anode materials for Li-ion batteries *J. Chem. Phys.* 118 6038–45

[78] Bonhomme C *et al* 2012 First-principles calculation of NMR parameters using the gauge including projector augmented wave method: a chemist's point of view *Chem. Rev.* 112 5733–79

[79] Aarva A, Deringer V L, Sainio S, Laurila T and Caro M A 2019 Understanding x-ray spectroscopy of carbonaceous materials by combining experiments, density functional theory, and machine learning. Part I: fingerprint spectra *Chem. Mater.* 31 9243–55

[80] Aarva A, Deringer V L, Sainio S, Laurila T and Caro M A 2019 Understanding x-ray spectroscopy of carbonaceous materials by combining experiments, density functional theory, and machine learning. Part II: quantitative fitting of spectra *Chem. Mater.* 31 9256–67

[81] NobelPrize.org 2019 *The Nobel Prize in Chemistry* [www.nobelprize.org/prizes/chemistry/2019/press-release/](http://www.nobelprize.org/prizes/chemistry/2019/press-release/) (Accessed 06 June 2020)

[82] Momma K and Izumi F 2011 VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data *J. Appl. Crystallogr.* 44 1272–6