

Machine Learning Force Fields: Recent Advances and Remaining Challenges

Igor Poltavsky and Alexandre Tkatchenko*



Cite This: *J. Phys. Chem. Lett.* 2021, 12, 6551–6564



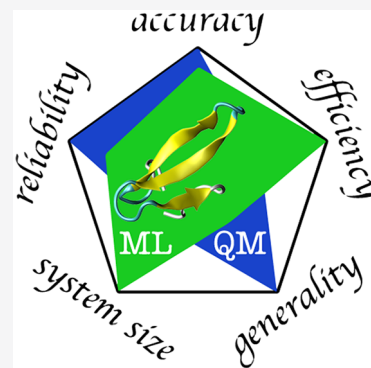
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: In chemistry and physics, machine learning (ML) methods promise transformative impacts by advancing modeling and improving our understanding of complex molecules and materials. Each ML method comprises a mathematically well-defined procedure, and an increasingly larger number of easy-to-use ML packages for modeling atomistic systems are becoming available. In this Perspective, we discuss the general aspects of ML techniques in the context of creating ML force fields. We describe common features of ML modeling and quantum-mechanical approximations, so-called global and local ML models, and the physical differences behind these two classes of approaches. Finally, we describe the recent developments and emerging directions in the field of ML-driven molecular modeling. This Perspective aims to inspire interdisciplinary collaborations crossing the borders between physical chemistry, chemical physics, computer science, and data science.



A central part of any problem in physics and chemistry is a model, a set of equations, plus initial and boundary conditions. The traditional scientific approach consists of writing down the model, solving it using some (hopefully reasonable) approximations, analyzing the calculations, and verifying the obtained results with reference data from higher-level calculations or experiments. Unfortunately, when the system under study is too complicated and our knowledge and intuition are insufficient to formulate reliable models, such an approach is obviously inapplicable. An alternative way was provided with the development of sophisticated machine learning (ML) methods. The first step of model formulation is avoided because ML models are data-driven rather than based on human intuition. However, predicting the behavior of a given system is far from the coveted goal of achieving a mechanistic understanding of the important processes that underlie the phenomenon of interest. Detectable scientific progress typically arises from a practical combination of ML and human (chemical and physical) intuition.

Many problems in chemistry and physics that eluded practical solutions for a long time have now been solved or are in the process of being solved by including ML models into the simulator's toolbox. These problems include the design of new materials,^{5–13} large-scale simulations,^{1,2,14} predicting new chemical reactions,^{15–17} and understanding a complex interplay of interatomic and intermolecular interactions.^{18,19} Among this large variety of applications, the construction of accurate and efficient ML force fields (MLFFs) for molecules and materials is of particular interest.^{20–28}

The combination of chemistry, physics, and ML is a true symbiosis. Increasingly powerful ML methods help in the

The combination of chemistry, physics, and ML is a true symbiosis. Increasingly powerful ML methods help in the development of more accurate and more transferable MLFFs. Similarly, we learn how to develop better and more data-efficient ML methods by incorporating prior knowledge into the construction of these sophisticated models.

development of more accurate and more transferable MLFFs. Similarly, we learn how to develop better and more data-efficient ML methods by incorporating prior knowledge into the construction of these sophisticated models. The majority of traditional ML methods were developed to deal with extended data arrays containing different types of properties: continuous and discrete, static and updated in real-time, measured in

Received: April 14, 2021

Accepted: July 5, 2021



ACS Publications

© XXXX American Chemical Society

6551

<https://doi.org/10.1021/acs.jpclett.1c01204>
J. Phys. Chem. Lett. 2021, 12, 6551–6564

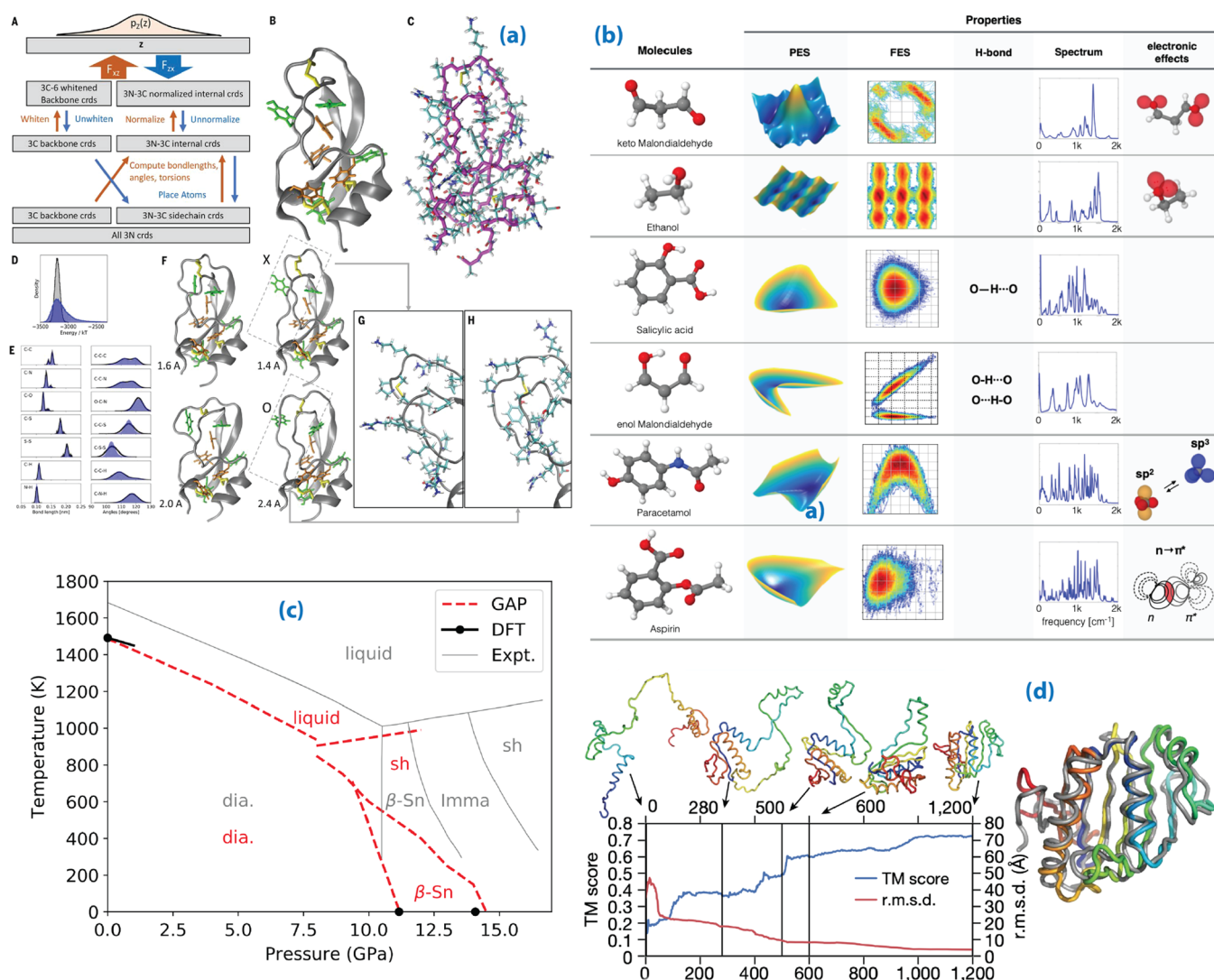


Figure 1. Examples of practical modeling applications for molecules and solids which would be computationally prohibitive without employing ML methods. Examples include an accurate sampling of the atomic configurations¹ as well as simulation of the folding process of proteins² (a and d), constructing and using accurate coupled-cluster potential energy surfaces for small molecules³ (b), and reproducing the phase diagram of silicon with density-functional theory accuracy⁴ (c). Images used with permission from refs 1–4. Copyright 2019 (AAAS), 2020 (Springer Nature), 2019 (AIP Publishing), and 2018 (the authors), respectively.

numerous units, possessing varying scales, and so on. The availability of data led to the development of powerful but “heavy-weighted” methods, such as deep neural networks (NN).²⁹ As typical examples we can mention the NNs developed for playing video games^{30,31} or speech and text recognition.^{32–35} The common feature of these approaches is a complex architecture with millions of training parameters, which require enormous reference data sets for their adjustment. Collecting such data sets in natural sciences, where each training point is a result of computationally expensive *ab initio* calculations or challenging experimental measurements, is often impossible. Hence, the data efficiency of the ML models becomes a crucial factor. This led to the emergence of a novel direction in ML, where the basic physical laws or knowledge are directly implemented into the architecture of ML models.^{20–24} Such models demonstrate superior performance when using limited reference data sets compared to the ML methods relying solely on brute force (deep architectures). Also, knowledge-based ML models can provide insights into complex physical

and chemical processes, becoming truly novel scientific approaches in their own right.^{4,18,19,36}

Combining ML methods with analytical and computational physics/chemistry has allowed a substantial extension in the range of accessible applications, and we highlight some of the prominent applications in Figure 1:

- 1. Accelerated statistical sampling:** ML-driven Boltzmann generators were developed to obtain unbiased equilibrium samples of condensed-matter systems and proteins, which allow accurate computation of free-energy differences without prior knowledge of reaction coordinates and avoiding time-demanding simulations (Figure 1a).¹
- 2. Essentially exact MLFFs for small molecules:** The data-efficient sGDML MLFF approach enables the construction of essentially exact molecular force fields by learning atomic forces at the coupled-cluster level of quantum mechanics for molecules with up to dozens of atoms (Figure 1b).³

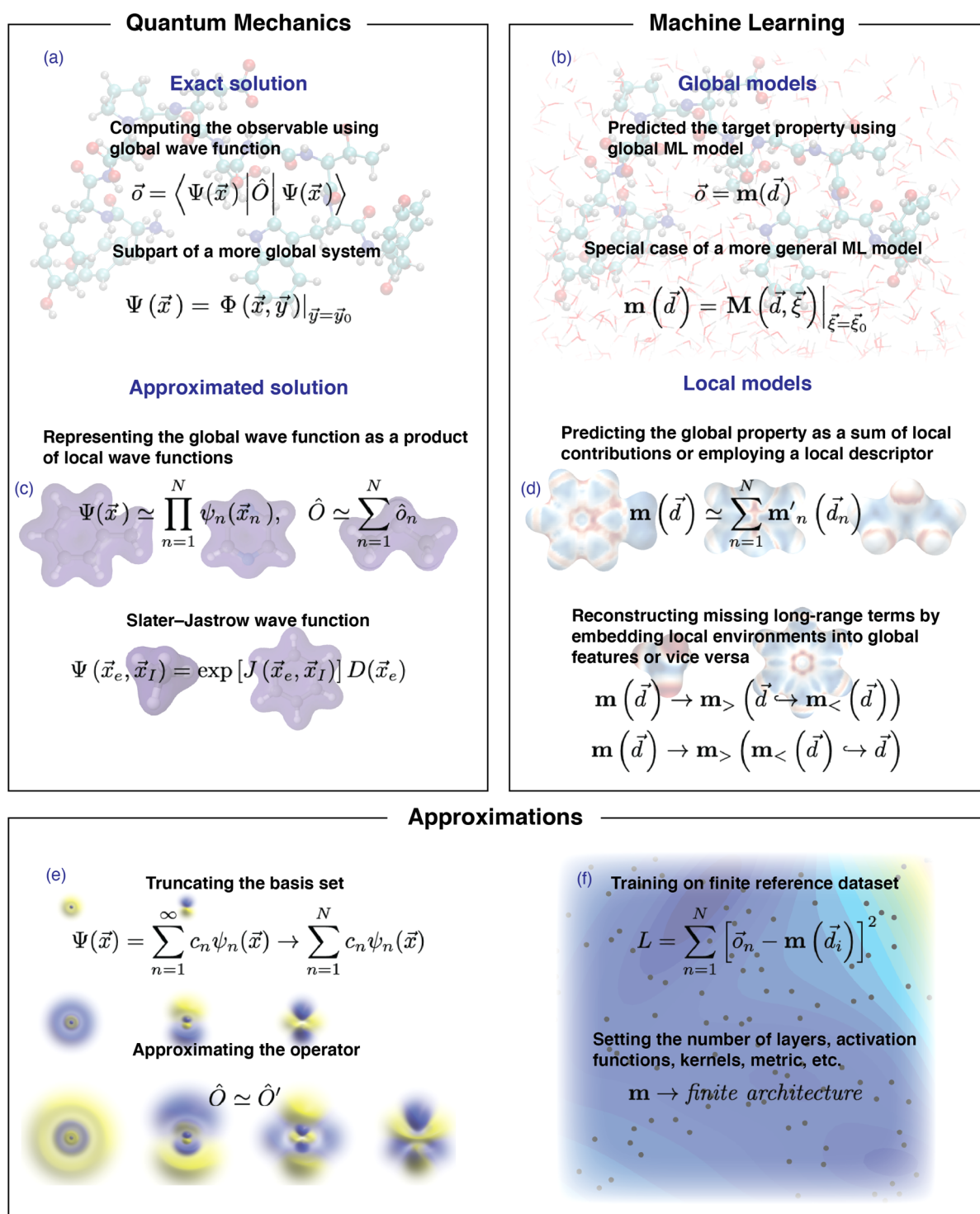


Figure 2. Similarity between methods in quantum mechanics and machine learning. Images from left to right from top to bottom: Chignolin protein⁴⁷ without (a) and with (b) the water environment, visualized total charge densities (c) and local chemical potentials (d) obtained using the SchNet model,³⁶ wave functions for hydrogen atom (e) and Müller-Brown potential (f). Images used with permission from refs 36 and 47. Copyright 2017 (the authors) and 2011 (AAAS).

- Accurate phase diagrams from a unified FF:** A single ML model for silicon accurately reproduces experimental reference results for a wide range of observable properties, including crystal, liquid, and amorphous bulk phases (Figure 1c).⁴
- Data-driven protein folding:** A deep ML model is able to predict folded conformations for complex proteins from

amino acid sequences providing potential insights into the function and malfunction of proteins (Figure 1d).²

All these applications were considered to be out of reach for molecular simulation methods just a decade ago. However, the sheer variety of existing ML models for constructing MLFFs may be overwhelming even for specialists. Widely used examples include sGDML,²¹ SchNet,²² GAP,²⁴ DPMD,²⁷ HDNN,²⁶

PhysNet,²⁵ and ANI,^{37,38} and one can continue this list with their favorite options. At this point, understanding the basic concepts behind these models as well as their limitations becomes essential to advance the field of ML-enabled computational chemistry and physics.

We remark that this Perspective is not a detailed review of ML methods but is rather aimed at providing a concise perspective on recent advances and remaining challenges for MLFF methods. One can find many more details and comprehensive discussions about state-of-the-art ML models in recent review articles (see refs 39–46). Also, we heavily refer to the sGDML model to demonstrate different challenges throughout the text. This by no way means that the discussed challenges are inherent problems of only the sGDML model. The choice is dictated by the fact that we try to be as self-critical as possible when discussing fundamental challenges that many MLFF models face.

ML models often employ some knowledge from the relevant application domain. Therefore, understanding similarities between traditional methods and ML models is the key to develop efficient and accurate ML-enabled modeling techniques. Below we describe our perspective on the similarity between modern ML methods used in computational chemistry and physics and traditional approximate techniques widely employed to study large quantum-mechanical systems in practice (see Figure 2). We start with the direct similarity between computing an observable using the corresponding operator and many-particle wave function and predicting this property using a global ML model based on some global descriptor

$$\bar{O} = \langle \Psi(\vec{x}) | \hat{O} | \Psi(\vec{x}) \rangle \Leftrightarrow \bar{O} = \mathbf{m}(\vec{d})$$

Here \hat{O} is the corresponding quantum-mechanical operator; Ψ is the system wave function depending on all degrees of freedom (DOFs) \vec{x} ; \mathbf{m} is the ML model, and \vec{d} is the descriptor employed to encode the DOFs.

The global model and descriptor mean that the entire system is treated as a single object without approximating the range of interactions or partitioning the system into independent building blocks. Formally, global ML models should be capable of learning any mapping between the descriptor \mathcal{D} and target \mathcal{P} spaces, whenever enough reference data are provided and the model contains sufficient number of adjustable parameters. The drawback of global models is that as they are trained for specific systems, global models lose their suitability whenever the compound, size, or environment of these systems undergo any noticeable changes. Because any real system is a smaller part of another more extended system, any global ML model is formally a special case of a more general ML model. Mathematically this means that the function $\mathbf{m}(\vec{d})$ is just a special case of an extended function $\mathbf{M}(\vec{d}, \vec{\xi})$, with $\vec{\xi}$ being a descriptor for environment variables, additional particles, *etc.* Knowing the function \mathbf{m} is not enough to predict the results of the function \mathbf{M} . For this reason, the transferability and scalability of global ML models is an open challenge requiring novel ideas and breakthroughs.

To make ML models transferable and scalable, the assumption of locality was introduced. Local ML models divide a system into subregions by introducing either explicit cutoff radii or direct partitioning into building blocks. Then, instead of solving the entire problem in one go, each ML model deals with a simplified learning task. The global properties are reconstructed by summation over the contribution from individual ML models. This idea is very similar to the

approximated solution of the Schrödinger equation, when the full wave function is approximated by a product of “local” wave functions or orbitals

$$\Psi(\vec{x}) = \prod_n \psi_n(\vec{x}_n) \Leftrightarrow \mathbf{m}(\vec{d}) = \sum_n \mathbf{m}_n(\vec{d}_n)$$

Here \vec{x}_n describes only a part of the quantum system (local descriptor \vec{d}_n), and the wave functions ψ_n (models \mathbf{m}_n) depend solely upon a corresponding subset of DOFs. A well-known example of such an approach is the Born–Oppenheimer approximation where the nuclei and electronic variables are decoupled. In principle, the assumption of locality should make ML models applicable to systems of varying size and composition. In practice, these aspects require careful assessment for each class of systems of interest.

Similar to corresponding quantum-mechanical approximations, which can be quite reliable in many cases, local ML models also not only demonstrate high accuracy but also may provide direct insights into the physics of the studied systems. In Figure 2, one can see the direct similarity between the total charge densities reproduced using density-functional calculations with FHI-aims software package⁴⁸ using the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional⁴⁹ with tight settings and the Tkatchenko–Scheffler (TS) method⁵⁰ for a set of small molecules and the local chemical potentials computed using the SchNet ML model.³⁶ To achieve such results, state-of-the-art ML models must embed global information into local descriptors and employ intelligent ways to combine local models (beyond simple summation). This can be likened to approximately solving the Schrödinger equation when the trial wave functions are dressed by additional terms accounting for missing DOFs (starting with a mean-field approach, adding a Jastrow correlation factor, *etc.*). Unfortunately, local ML models also inherit the weaknesses of approximate quantum-mechanical solutions. Namely, local ML models have problems describing long-range interactions that naturally violate the partitioning scheme. Moreover, employing local descriptors leads to an incomplete representation of the system’s quantum-mechanical states. This can affect the resulting ML models’ predictive capabilities as much as or even more than the choice of the regression method.^{51,52}

Finally, some approximations are unavoidable when solving quantum equations in practice using both QM and ML methods. The first and the most obvious one is truncating the wave function basis to perform practical calculations, which is similar to choosing a finite training set. Here, the main factor defining the accuracy of the resulting solution is not just the size of the basis (training) set, but the particular choice of the wave functions (training points). For instance, plane-wave basis sets, Gaussian basis functions, or numeric atom-centered basis functions could be employed to solve density-functional theory (DFT) equations with a desired performance and accuracy depending on the type of systems under study. Similarly, a particular choice of the training set may substantially improve or worsen the ML model predictions, everything else being equal.⁵³

Another common approximation in quantum mechanics is related to simplifying the potential-energy operator. An example is the dipole approximation to describe interactions between atoms and radiation. It essentially consists of neglecting the spatial variation of the electromagnetic field over the atom; instead, one uses the field at the location of the nucleus. Such approximations substantially simplify the mathematics and computations but restrict the capability to describe a certain

class of effects. This is analogous to specifying a particular architecture of the NN or defining the kernel and the metric for ridge-regression methods in ML. As soon as such choices are made, the maximum complexity of the mapping function between the descriptor and target spaces is predefined. This limits the potential predictive power of the ML models. While

While ML models can be employed for predicting any chemical or physical property, creating efficient and accurate models requires accounting for the specific nature of a target property.

ML models can be employed for predicting any chemical or physical property, creating efficient and accurate models requires accounting for the specific nature of a target property. Therefore, before discussing the challenges in existing MLFFs, we briefly review the main features of interatomic interactions. In quantum mechanics, a system of interacting particles is described by a wave function $\Psi(\vec{r}_1, \dots, \vec{r}_n, \vec{R}_1, \dots, \vec{R}_n)$, which depends on the position of all electrons (\vec{r}) and nuclei (\vec{R}). As a result, the interaction energy between any two atoms a and b

$$U_{ab}(\vec{R}_1, \dots, \vec{R}_n) = \langle \Psi(\vec{r}_1, \dots, \vec{r}_n, \vec{R}_1, \dots, \vec{R}_n) | \hat{U}(\vec{r}_a, \vec{r}_b, \vec{R}_a, \vec{R}_b) | \Psi(\vec{r}_1, \dots, \vec{r}_n, \vec{R}_1, \dots, \vec{R}_n) \rangle$$

also depends on the coordinates of all nuclei even when the interaction potential is a pairwise operator, like the Coulomb potential. Indeed, after integrating out the electronic DOFs, the dependence on all atomic positions emerges through the wave functions. This global nature of the effective potential energy experienced by the atoms is responsible for nonlocal many-body effects present in molecules and materials. Because the Schrödinger equation is exactly solvable in practice only for a small number of particles, introducing approximations for the wave function is an unavoidable step. Subsequently, approximate wave functions are amenable to perturbative decompositions in terms of interatomic or interelectronic distances, which allow the definition of different chemical and physical forces. Understanding various types of forces and the approximations which they involve is crucial for developing chemical and physical intuition and accurately modeling realistic molecular systems. The same idea holds true for the development of MLFFs. Here, the understanding of different interactions relevant for the system of interest is key for developing accurate and data-efficient ML models.

There is no unique classification of interactions as short- or long-ranged, as such definition greatly depends on the particular problem under study. For instance, short-range interactions at a nanometer level can be considered long-range interactions from the perspective of atomistic simulations. In this work, we take a chemical perspective in terms of interaction ranges and define short-range interactions as those that decay exponentially and become negligible at distances a few times larger than the average equilibrium interatomic distances between the nearest covalently bonded neighbors. Such examples are covalent bonds involving sharing of electron pairs between atoms or a single hydrogen bond, describing the attraction between a hydrogen atom, which is covalently bound to a more electronegative atom or group, and another electronegative atom bearing a lone pair of

electrons. The bond lengths in both cases are within a few angstroms, and such interactions become negligible at distances larger than 5–7 Å. In contrast, the screened electrostatic interactions in ionic liquids, despite their exponential decay, cannot be considered as short-ranged, because the corresponding effective Debye length is a few nanometers, substantially exceeding covalent-bond or hydrogen-bond distances. On the other hand, electrostatic interactions in semiconductors are effectively short-ranged given the relatively high dielectric constant of these materials.

Accounting only for short-range interactions is rarely sufficient for a reliable description of molecules and materials at the atomistic level. Both electrostatic and electrodynamic (e.g., van der Waals dispersion) interactions often play an important role in defining structural, dynamic, thermodynamic, and optical properties.⁵⁷ A general feature of long-range interactions is their nonadditivity, which leads to a plethora of distinct effects, both thermodynamic and kinetic, such as slow relaxation or broken ergodicity.⁵⁸ Complex wave functions of systems containing hundreds and thousands of atoms result in nontrivial functional forms of long-range forces. The many-body nature of long-range interactions also leads to a strong dependence of these interactions upon the environment of interacting objects. In contrast to pairwise forces, long-range interactions involve feedback effects, with a cascade of distance scales implicated in a given interaction. Hence, long-range forces can propagate to distances much larger than one would expect based on their asymptotic functional form for two objects. For instance, in biomolecules surrounded by water, the contribution of vdW dispersion interactions to the solvation energy extends to at least 25 Å from the protein surface involving multiple shells of water molecules (see Figure 3).⁵⁵ At the nanoscale, the range of vdW interactions can become even larger, reaching hundreds of angstroms with a very peculiar dependence of their functional form on the dimensionality and topology of interacting objects, substantially deviating from the expected pairwise-additive asymptote (see Figure 4).⁵⁶ The same is true for all other types of long-range forces. Classical electrostatic interactions can become nontrivial in systems such as organic semiconductors and organic solar cells;⁵⁹ induced electrostatics or polarization interactions demonstrate peculiar behavior in the presence of surfaces or ionic liquids,⁶⁰ and so on. In summary, while short-range forces are well-described and understood, accounting for long-range interactions is still an open challenge in both theory and simulations. In order to accurately describe the plethora of long-range interactions, the physical or ML models must have significant flexibility without employing strictly predefined functional forms or imposing characteristic length scales.

The concept of global and local models plays a crucial role in MLFF developments. This forms the basis for how ML models treat interactions in the studied system. On the one hand, the widely quoted nearsightedness approximation⁶¹ tells us that only neighbors within some cutoff radius may significantly contribute to the forces acting on a given atom. This allows splitting the learning of the interatomic interactions into blocks, introducing localized descriptors used by local ML models. Whenever such an approach is applicable (and its applicability can be tested only *a posteriori*), it provides many practical benefits. Namely, as a result of the approximation of locality, the ML models can be trained using small reference molecules or a set of reference structures representing various possible local configurations without sampling the entire configurational space (CS). After training, the obtained MLFFs are applicable to

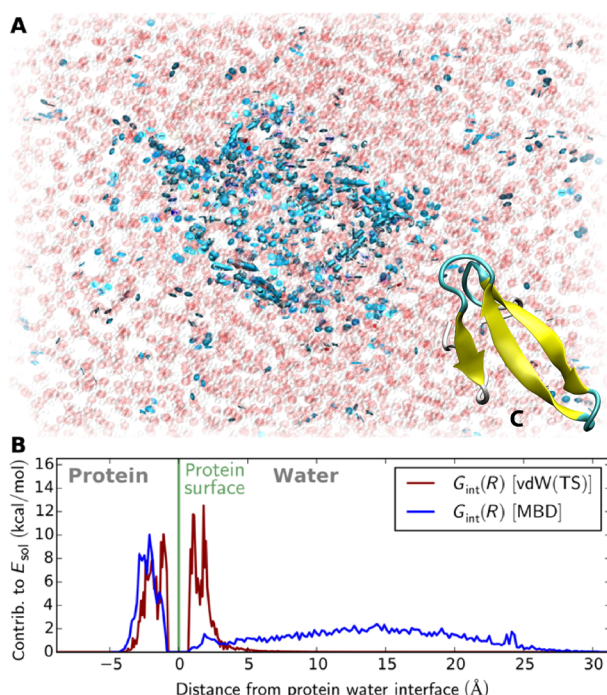


Figure 3. Highly nonlocal contribution to van der Waals (vdW) dispersion interaction in biomolecules surrounded by water. (A) Illustration of low-frequency plasmon-like dispersion fluctuations in solvated Fip35-WW protein, which show the largest contribution to the protein–water vdW interactions (solvent shown in atomistic detail: oxygen, red; hydrogen, white). The arrows (blue) depict the direction of simultaneous electron density deformations (eigenmode of the electron density). If no arrow is shown, the given atom does not contribute notably to the eigenmode. (B) Contributions to vdW solvation energy within the pairwise vdW(TS) approach⁵⁰ and the many-body dispersion (MBD) formalism⁵⁴ as radial distribution functions. (C) Structure of the Fip35-WW protein. Image used with permission from ref 55. Copyright 2019 (the authors).

reproduce multiple complex processes under different conditions or even for the systems with different size and composition (scalability and transferability). For an example, see Figure 1c, where the entire phase diagram of silicon is obtained from a single ML model, or ref 28, where the dihedral rotation energy profiles of a simple peptide bond and glycine molecule are reproduced without having these molecules in the training set. The transferability and scalability of local ML models make them very attractive for large-scale applications. As a result, the majority of state-of-the-art MLFFs employ the nearsightedness approximation in one form or another. More details about local ML models can be found in ref 44.

On the other hand, as clearly discussed in the seminal work of Prodan and Kohn,⁶¹ the nearsightedness approximation is formally fulfilled only for ideal metals. In insulators and semiconductors, nonlocal (long-range) interactions often render the interatomic forces complex nonlinear functions of atomic positions, which then exhibit many-atom and multiscale nature. Obviously, short-range forces remain important. No local ML model with any reasonable cutoff radius, typically between 5–10 Å to avoid high computational costs, can accurately describe both local and nonlocal interactions on equal footing. Hence, additional steps are needed to extend the applicability of local ML models to describe long-range interactions. Alternatively, one should switch to global ML models removing the nearsightedness approximation.

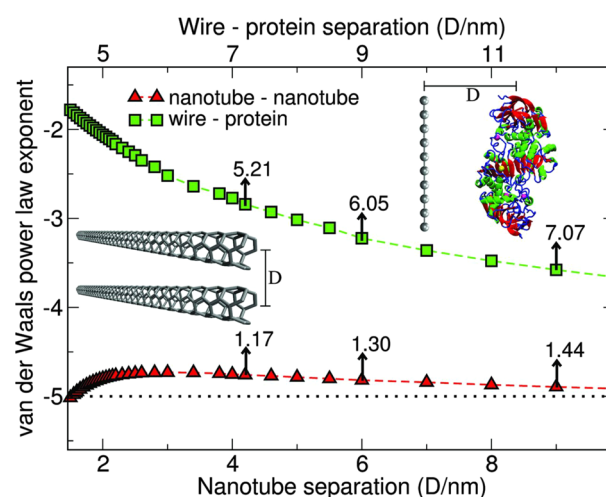


Figure 4. Long-ranged nature and nontrivial functional forms of vdW interactions in nanosystems. Interaction energy power-law exponents computed using the many-body dispersion (MBD) method⁵⁴ to account for vdW interactions for two parallel carbon nanotubes and a wire-protein nanostructure (depicted in insets) as a function of the respective center-of-mass separations D . The pairwise-additive asymptote of D^{-5} is denoted by the dotted line.⁵⁶ The numbers above the curves indicate the growing ratios between MBD and pairwise dispersion energies as a function of D . Image used with permission from ref 56. Copyright 2016 (AAAS).

Unfortunately, switching to a global ML model is not a silver bullet. The first issue global models inevitably face is the multiscale nature of real interactions. Namely, optimal hyperparameters and the choice of a descriptor for a ML model are different for short- and long-range interactions. Mixing both types of interactions within one MLFF requires a compromise to be made, limiting the accuracy of the predictions. Figure 5 shows the results for the predicted close-to-equilibrium potential energy surface (PES) of a water tetramer. The reference data set has been generated using ORCA⁶² 4.2.0 *ab initio* molecular dynamics trajectories of nine different runs starting from the same geometry (equilibrium structure) with different initial velocities. The simulation temperatures were 100, 300, and 500 K (three trajectories per temperature with different random initial velocities). The time step was set to 1 fs with 100 steps per trajectory. Second-order Møller–Plesset perturbation theory⁶³ calculations with def2-TZVP basis set were employed to compute energy and forces. sGDML models were created using 300 data points for training and validation. First, we train a single global model to reproduce the entire PES of a water tetramer (see Figure 5a). Second, we constructed a multiscale ML model using four sGDML models. To do so, we started with a sGDML model, M_1 , for a single water molecule. Then we computed the 2-body intermolecular interaction forces, $\vec{F}_{2\text{-body}}$, as a difference between the forces, \vec{F}_{dimer} , acting in the dimers constituting the tetramer (the forces were recomputed for each dimer configuration to avoid the effect from the other two water molecules) and the predictions of the M_1 model for each water molecule in the dimer

$$\vec{F}_{2\text{-body}} = \vec{F}_{\text{dimer}} - M_1(\text{water}_1) - M_1(\text{water}_2)$$

A new sGDML model, M_2 , is trained afterward to reproduce $\vec{F}_{2\text{-body}}$ forces. A similar procedure is repeated for 3-body and 4-body intermolecular interactions, resulting in four sGDML models used to predict the forces shown in Figure 5b. One sees

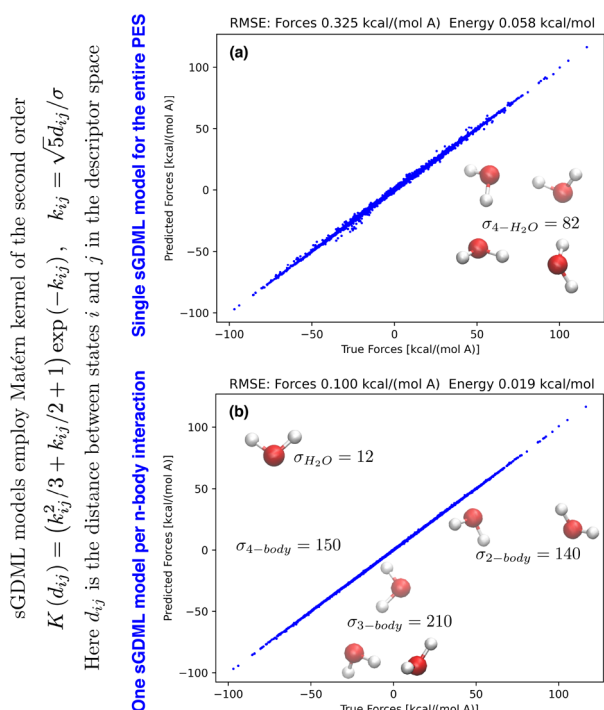


Figure 5. Multiscale nature of realistic interactions. Comparison between the predictions for the close-to-equilibrium part of the PES of four water molecules obtained within a single sGDML model (a) and by consequentially learning n -body interaction from a single molecule and up to the tetramer with individual sGDML models (b). The optimal value of the hyperparameter σ , defining the model's scale in the descriptor space, is shown for each case.

that the hyperparameter σ , defining the range in the descriptor space of the sGDML model, is noticeably different for intra- and intermolecular forces. A single global model attempts to compromise the small value of 12 for a single water molecule and an order of magnitude larger values of 140–210 for intermolecular interactions. As a result, the predictions of a global model are noisier, and it yields a 3-fold larger root-mean-squared error (RMSE) for both energy and forces compared to the multiscale approach. Note that the generated data set covers only close-to-equilibrium parts of the PES, making the learning problem quite simple. This explains the very high accuracy, with RMSE below 0.06 kcal/mol, of the models with only 300 training points. Consequently, the difference between the predictions of the global and the multiscale ML models in Figure 5 may not look impressive. However, achieving a 3-fold decrease in the RMSE without increasing the size of the training set is important in practice and is a nontrivial challenge for such accurate predictions.

The problem with multiscale interactions is not only relevant for kernel methods and global models. Figure 6 shows the difficulties a set of ML models faces in predicting energy profiles of the PES of cumulenes ($C_{2+n}H_4$) as a function of a dihedral angle between the terminal methylene groups. Here the energetically favorable overlap of π -orbitals along the carbon chain, which is broken when the methylene groups are rotated against each other, leads to a highly nonlocal interaction inside the molecule (see ref 44 for more details). This clearly demonstrates that the multiscale nature of interactions is not limited to systems only with competing intra- and intermolecular forces (Figure 3) or nanosystems (Figure 4). It is a general phenomenon present even within single molecules. The

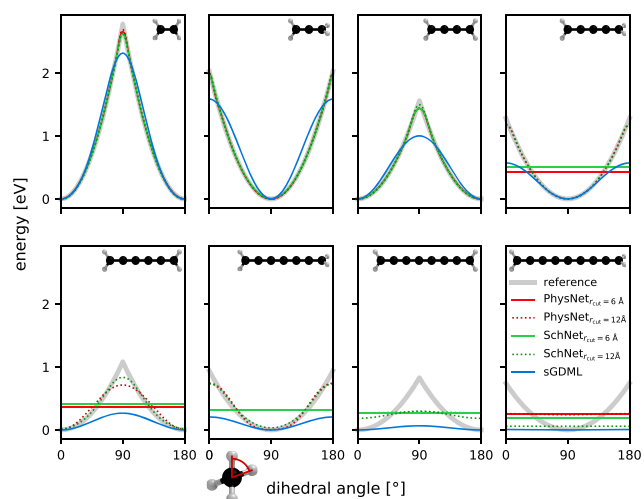


Figure 6. Long-ranged interactions within a molecule. Energy profiles of different ML-based PESs for a rotation of the dihedral angle between the terminal methylene groups of cumulenes ($C_{2+n}H_4$) of different sizes ($0 \leq n \leq 7$).⁴⁴ Image reprinted from ref 44. Copyright 2021 (American Chemical Society).

global kernel-based sGDML model, as well as the local deep NNs SchNet and PhysNet, fail in reproducing the correct PES profile in such a case. Thereby, an accurate description of the corresponding long-range interactions is missing in all three models.

Importantly, despite the failure in reproducing the energy as a function of a dihedral angle between the terminal methylene groups, the overall RMSEs of all the models shown in Figure 6 is small. This shows an inherent problem in the training procedure. A small overall RMSE does not guarantee good predictions for all parts of the PES. One may have entire clusters of configurations poorly predicted by a given ML model. Figure 7 shows a detailed error analysis for sGDML, SchNet, and GAP/SOAP models trained for ethanol, uracil, toluene, and salicylic acid molecules.⁵³ One can see that for all examples and all methods, there are clusters of configurations with similar geometry and energy where the RMSEs are much higher than those for the best-predicted regions of the PES. Because the relative population of such clusters is small in the reference data set, the large errors in those parts of the PES do not contribute considerably to the overall prediction accuracy. Nevertheless, regions of the PES with large errors in ML models may correspond to important rare events in molecular systems, and such events might be entirely wrongly predicted by MLFF models. For instance, this can lead to an incorrect description of an energy profile for a hydrogen atom shared between the carboxylic and hydroxyl groups in salicylic acid (see ref 53 for more details).

The challenges shown in Figures 5–7 originate from different sources, despite the same fundamental reason—the multiscale nature of interatomic (interelectronic) interactions. Figure 5 demonstrates the limitation of a single kernel hyperparameter to describe simultaneously inter- and intramolecular interactions. In contrast, the errors in Figures 6 and 7 come mainly from the limitations of the employed descriptors to describe short- and long-range neighborhoods equally accurately. Because the short-range forces are typically much stronger than long-range ones, relevant descriptors tend to emphasize the nearest neighbors, diminishing the importance of mutual correlations between separated atoms. This happens with both physical descriptors

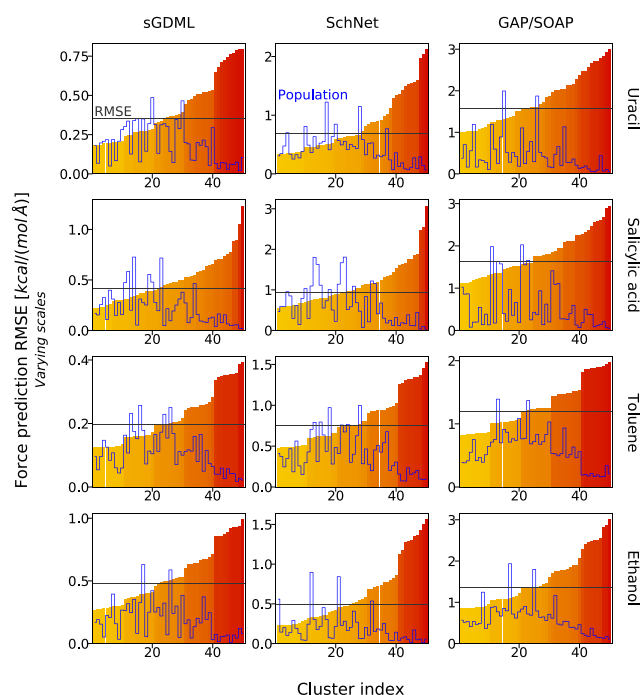


Figure 7. Detailed analysis of the accuracy of ML models trained for small molecules. Force prediction RMSE for sGDML, SchNet, and GAP/SOAP (with 12 radial and 6 angular functions) models on the same ethanol, uracil, toluene, and salicylic acid data sets (y-axis, scale adapted for each model for better visibility), split into 50 clusters of similar configurations (x-axis) ordered by ascending error. RMSE (bars) is given on a per-cluster basis in contrast to the RMSE over the entire data set (solid horizontal black line). Relative cluster populations are also indicated (solid blue line, arbitrary units).⁵³ Image used with permission from ref 53. Copyright 2021 (AIP Publishing).

(sGDML) and descriptors constructed by ML models (SchNet and PhysNet). Indeed, an MLFF training process starts by fitting the parameters of the model to capture the main contributions to the PES in terms of the descriptor components. On an atomistic level, these are typically local contributions that depend upon distances and angles between neighboring atoms. When the amount of reference data is not excessively large, ML models cannot build arbitrarily complex mappings between the employed descriptors and target spaces. Hence, ML models benefit from having a descriptor consisting of the leading DOFs for the PES function. On the other hand, the overemphasizing of local DOFs in descriptors restricts the ML models' capability to learn realistic PESs. As an example, we consider here the inverse distances, the default descriptor for the sGDML model. Figure 8a shows the distribution of the Euclidean distances in the descriptor space between different geometries of an azobenzene molecule within trans- (trans–trans) or cis- (cis–cis) and between cis- and trans-like configurations.⁶⁴ Because we employ the inverse interatomic distances as a molecular descriptor, the distances in the descriptor space are measured in \AA^{-1} . As can be seen on the plot, any stationary kernel model based on this descriptor and employing Euclidean metrics would start mixing cis- and trans-like configurations at a distance of 0.62 \AA^{-1} . This corresponds to small average changes in individual interatomic distances, at least an order of magnitude smaller compared to thermal fluctuations at ambient conditions. Hence, even when the number of reference points is sufficient to represent the molecule's dynamics, we may have issues with reproducing the PES because of mixing qualitatively different molecular states.

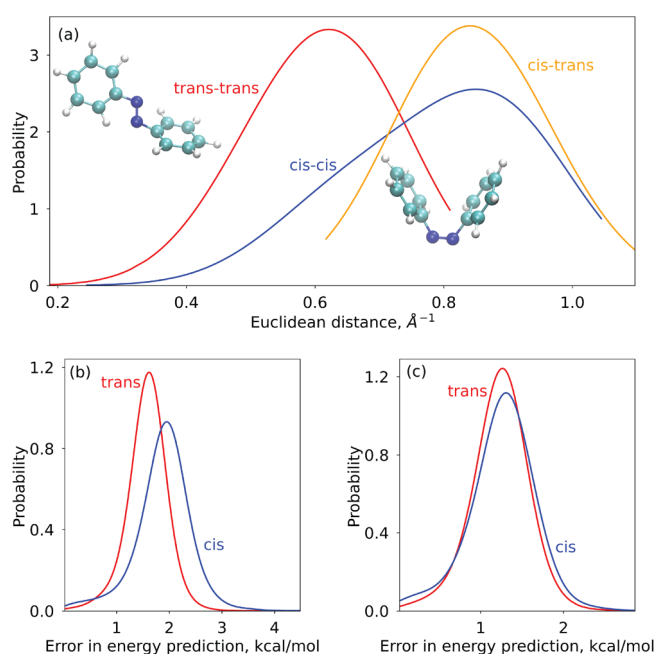


Figure 8. Mixing physically different states within a global ML model. (a) Distribution of the Euclidean distances in the inverse distance descriptor space between different geometries of an azobenzene molecule within trans- (trans–trans), cis- (cis–cis), and between cis- and trans-like configurations. The data sets are the results of classical MD simulations at 500 K. The energy prediction error distribution in sGDML models trained on 500 (b) and 700 (c) molecular geometries.

Given this analysis, the errors for cis-like configurations should be noticeably larger than those for trans-like, which is supported by Figure 8b. An increase in the density of training points in the descriptor space (larger training sets) will obviously solve the problem (see Figure 8c), but the resulting models will be data-inefficient. Instead of learning the PES, one feeds the ML model with data to avoid overlapping between cis- and trans-like configurations in our predictions. Indeed, the improvement of predictions for cis-like configurations with increasing the training set size from 500 to 700 points is larger than that for the trans-like geometries. With an increase in system size and flexibility, the problem described above will become increasingly pronounced. The definition of similarity in a high-dimensional space is an open challenge affecting most ML tasks, including regression, classification, clustering, and nearest-neighbor search algorithms. For constructing MLFFs, the situation is complicated even further because of the PESs' complex structures containing multiple wells, saddle points, and peaks of different scale and shape. Consequently, the similarity in the descriptor space, characterizing how similar the geometry of molecular configurations is, may not represent the similarity in the property space (energy, forces, dipole moment, polarizability, *etc.*). Namely, whenever the difference in geometry is finite, similar geometry does not always mean similar forces and energy. For instance, a small change of a bond length in the repulsion region may result in a large change in the potential energy. In contrast, the same geometrical deformation in the close-to-equilibrium region can be negligible in terms of energy change.

Finally, the major drawback of existing global ML models is that their applicability range is limited because they are trained for specific systems. Whenever the compound, size, or environment of these systems undergoes any noticeable

changes, the global models must be retrained. This raises the question, can the global ML models be useful for large-scale simulations, or is their applicability limited to only small- or moderate-size systems? At this point, we come back to the beginning of this section, where the local models were discussed. Strictly speaking, none of the realistic systems is a closed one. Any molecule, nanosystem, solid, liquid, *etc.*, dwells in some environment. Hence, all global ML models are, in fact, local models with a large cutoff radius. This radius is indirectly introduced to generate the reference data when we ignore the interaction of the model system with the rest of the world. Hence, both local and global ML models' advancement is necessary to develop a unified and efficient approach for accurate large-scale simulations.

Multiple research efforts aim to make accurate simulations of systems containing hundreds and thousands of atoms as trivial as those for small rigid molecules or solids with simple bonding patterns. Schematically, one can loosely subdivide existing challenges into the four groups shown in Figure 9.

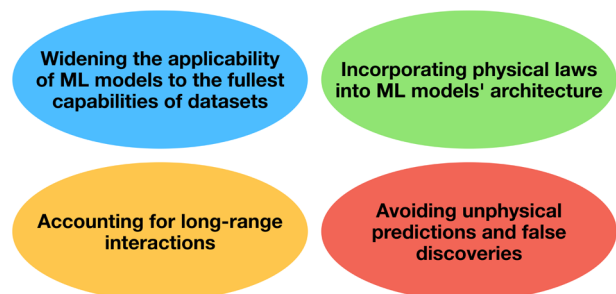


Figure 9. Important research directions to be undertaken to extend the applicability of existing MLFF models to large systems and long time scales.

Widening the Applicability of ML Models to the Fullest Capabilities of Data Sets. The backbone of any ML model is the reference data in general and the training set in particular. In fact, a good training set may be more critical than the ML model by itself. With the increasing size and complexity of the studied systems, generating the reference data and choosing an optimal training set becomes a challenging task. The situation is further complicated because the reference data are often collected using multiple different methods, for example, MD simulations, metadynamics, or sampling of normal modes.⁶⁵ These extensive samplings are performed using low-cost methods of computational chemistry/physics, generating hundreds of thousands of data-points. While novel adaptive sampling techniques substantially improve the situation, the final data sets are still massive and nonoptimal. Before training ML models, one should choose a much smaller subset of training and validation configurations and then recompute them using more accurate computational chemistry methods. Thereby, efficient techniques for optimizing the reference data sets become of particular interest. One example is a recently developed self-consistent training set optimization scheme, flattening the prediction error across the configuration space (Figure 10).³³ This approach not only makes ML models equally reliable for all parts of the PES covered by the reference data set but also stabilizes the behavior of the MLFFs in actual simulations (see ref 53 for more details). Still, more work is needed to make such methods applicable to arbitrary large systems, which require novel approaches for handling the “curse of dimensionality” problem. Indeed, locating

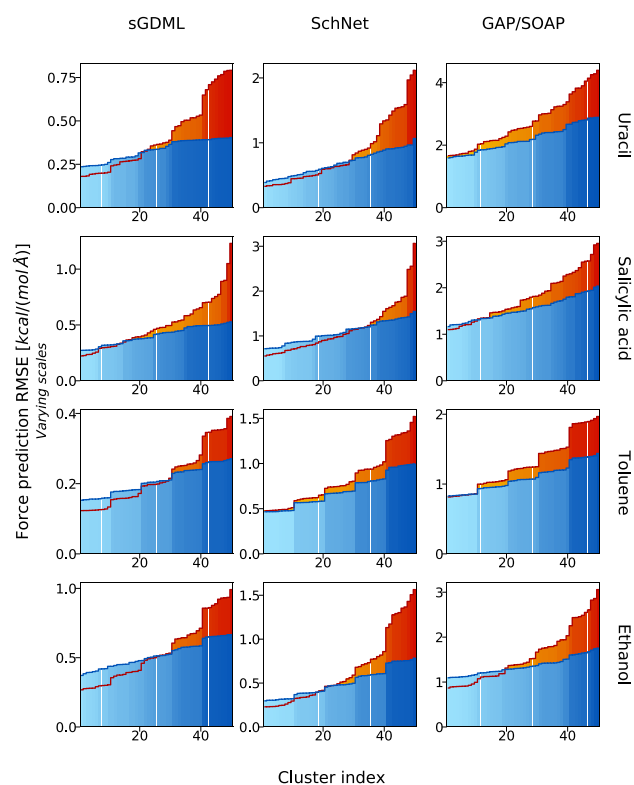


Figure 10. Flattening the prediction errors of ML models throughout the PES by optimizing the training set. Force prediction RMSE for sGDML, SchNet, and GAP/SOAP default models (as implemented in corresponding software packages) compared to the models trained on optimized training sets as explained in ref 53 (orange/blue bars, y-axis scale adapted for each model for better visibility). RMSE was computed on a per-cluster basis on ethanol, uracil, toluene, and salicylic acid data sets, split into 50 clusters of similar configurations (*x*-axis) ordered by ascending error. Image used with permission from ref 53. Copyright 2021 (AIP Publishing).

representative points of the PES and distinguishing between different system geometries become more and more problematic with the increasing size of the system's descriptor space. Any advances in this direction will have a significant impact on ML models in computational chemistry/physics and other fields.

Incorporating Physical Laws into ML Models' Architecture. The main advantage of ML models over traditional methods of computational chemistry is their reduced computational cost. The specific value of this speed-up highly depends on the reference quantum-mechanical method, but also the ML models by themselves can demonstrate a rather varied computational efficiency and hardware requirements. As an example, for predicting the forces shown in Figures 10 and 11, GAP/SOAP models need approximately an order of magnitude more time than sGDML or SchNet models, using the same hardware. Whereas for molecules containing only a few tens of atoms such a difference is irrelevant considering how fast the MLFFs are, going to large-scale simulations requires not only highly accurate but also computationally efficient ML models. These are mutually exclusive requirements. To achieve better accuracy, one needs more complex architectures for ML models. To be faster, one should use simpler architectures. To be fast and accurate simultaneously, state-of-the-art MLFFs went beyond traditional ML approaches and directly incorporated physical laws into their architecture. Contrary to the general-purpose ML models, where the data can be of any nature, physical systems

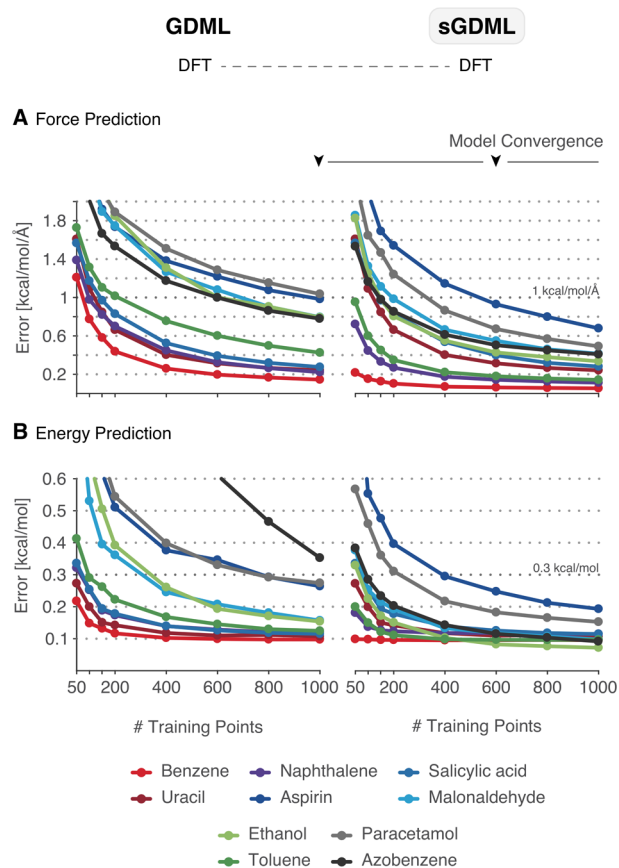


Figure 11. Data efficiency gains by using spatial symmetries in building ML models. Energy and force prediction accuracy (in terms of the mean absolute error (MAE)) as a function of training set size of sGDML and GDML models trained on DFT forces: the gain in efficiency and accuracy is directly linked to the number of symmetries in the system.⁶⁶ Image used with permission from ref 66. Copyright 2018 (the authors).

and corresponding ML models obey specific laws, which are the same for any molecule or material. For instance, the permutation of identical atoms cannot change the system properties. The same is true for the rotation and translation of the system as a whole in the absence of external fields: interatomic forces must be conservative, the total energy of two noninteracting molecules must be additive, and many more. Incorporating all such features on a model's architecture or descriptor level substantially improves the accuracy of the ML model, without making its architecture computationally demanding. Examples are sGDML⁶⁶ and GDML²⁰ models. The sGDML model, where spatial symmetries are implemented as part of the similarity measure, demonstrates significantly reduced requirements for the number of training points compared to the same model without these symmetries (GDML) (see Figure 11). The number of training points is crucial for kernel-based models, defining their evaluation speed and computational costs for both training and prediction. Thus, sGDML models are potentially applicable to larger systems as compared to GDML ones. In fact, including correct physical constraints and symmetries directly into ML models became a must-have feature for any modern MLFF.

Unfortunately, the benefits coming from employing symmetries are often lost while learning a complex global PES within a single ML model. Multiple local symmetries that a system possesses often disappear on the global scale. This suggests that the learning problem should be split into pieces. Just as one introduces a cutoff radius in local models to effectively reduce the complexity of system configurations, one can split the CS into domains holding the same symmetries to simplify the learning problem for each particular ML model, thereby substantially improving the overall data-efficiency.⁶⁴

Accurately Accounting for Long-Range Interactions. While for small rigid molecules and solids with simple bonding patterns, many modern ML models demonstrate excellent results, the situation changes drastically when noncovalent interactions play a role in the phenomena of interest. Obviously, local models based on partitioning schemes are not able to reproduce long-

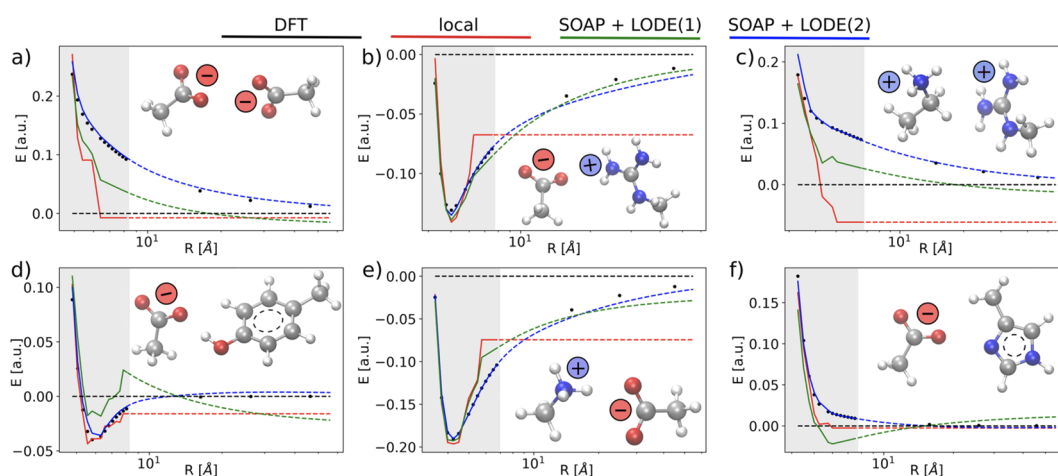


Figure 12. Improving the description of long-ranged interactions by introducing nonlocal representations of a system into local ML models. Comparison of reference and predicted binding curves of six molecular dimers:⁶⁷ (black dots) DFT reference calculations, (red lines) local SOAP predictions, (green lines) combined SOAP and LODE(1) predictions, and (blue lines) combined SOAP and LODE(2) predictions. Full lines and the shaded background represent the range of distances that is comparable to the geometries included in the training set. Dashed lines refer to predictions carried out in an extrapolative (long-range) regime. Panels a and c correspond to repulsive charge–charge interactions; panels b and e correspond to attractive charge–charge interactions; panel d corresponds to an attractive charge–dipole interaction, and panel e corresponds to repulsive charge–dipole interaction. Images reused with permission from ref 67. Copyright 2019 (AIP Publishing).

range forces. Fixing this weakness is a hot topic in modern research. One approach is the long-distance equivalent (LODE) framework,⁶⁷ which substantially extends the applicability of the smooth overlap of atomic positions (SOAP) descriptor for charged organic molecular dimers and liquid water (see Figure 12). As an alternative, one can introduce a multiscale approach, that is, splitting the problem into “local” and “nonlocal” parts, by, for instance, combining density-functional tight-binding with a deep tensor NN for the localized many-body interatomic repulsive energy,²⁸ building the NN PES based on environment-dependent atomic energies and charges,⁶⁸ combining a charge equilibration scheme employing environment-dependent atomic electronegativities with accurate atomic energies,⁶⁹ or combining a high-dimensional NN potential with the D3 scheme⁷⁰ to account for nonlocal vdW interactions.⁷¹ All in all, the development of a general and reliable scheme to describe long-range interactions within the ML framework is a critical element required to construct accurate and efficient MLFFs for nanoscale applications.

The development of a general and reliable scheme to describe long-range interactions within the ML framework is a critical element required to construct accurate and efficient MLFFs for nanoscale applications.

Avoiding Unphysical Predictions and False Discoveries. A risk that all MLFF models face when being employed to perform atomistic simulations is reaching their extrapolation regime. This can happen when a system undergoes structural changes and drifts away from the CS regions included in the training set. In this case, the predictions become unreliable and can lead to “novel” false discoveries. In this context, the estimation of the reliability of MLFF models is crucial. This task is difficult to achieve for systems containing thousands of atoms, when taking into account the high-dimensionality of their descriptor spaces. Importantly, not only the dimensionality of the CS but also the complexity of the PES is essential. As we have shown in Figure 8, the commonly used efficient inverse-distance descriptor becomes inadequate in distinguishing between cis- and trans-like configurations of azobenzene molecule, which contains only 24 atoms and has just two minima on its PES. For more complex systems, such as proteins, appropriate similarity measures need to be worked out. Indeed, with increasing system size, Euclidean distance measures become unreliable. This problem is actively investigated within the field of adaptive sampling, where the ability to detect qualitatively new states is a crucial aspect for the development of new methods.^{72,73} Other studies that would benefit from improved distance measures are dimensionality reduction techniques aimed at analyzing or enhancing atomistic simulations.^{74,75} Here, the aim is to maximize the sampling of CS beyond the already visited regions represented in the reference data set. Similar techniques should become a common practice also for MLFFs, where the reliability of predictions should be estimated and reported for each simulation step (see ref 4 as an example).

Another useful way to make ML models stable for practical applications is the incorporation of known asymptotic terms, which enforce physically relevant behavior in extrapolation

regions, such as for covalent-bond breaking. Such an approach is intensively employed for adding proper treatment of long-range interactions to local ML models,²⁵ within range-separation methods,²⁸ or when applying Δ -learning techniques.^{76,77} Nevertheless, enforcing specific asymptotic behavior does not guarantee overall correct performance. For example, kernel-based ML models yield, by construction, a constant interaction energy at a sufficiently large distance from all training points. This value can be set to zero, giving the correct result for noninteracting systems. However, at intermediate distances from the training points in the descriptor space (for finite values that are larger than those in the training set), one can obtain artificial minima or maxima with large energy gradients. This can lead to wrong bond-breaking behavior and qualitatively incorrect simulation results. Therefore, implementing estimators for the reliability prediction of MLFFs is essential even when the correct asymptotic behavior is fulfilled.

Finally, we would like to mention a direction in the development of MLFFs that goes beyond learning of observables. This includes approaches where ML models are used as architectures for solving physical equations. Examples include the FermiNet⁷⁸ and PauliNet⁷⁹ approaches, where deep NNs are used as variational wave functions for solving the many-electron Schrödinger equation. In principle, these wave functions could provide access to calculating different quantum-mechanical observables. They can also be used as input for variational quantum Monte Carlo calculations, able to approximate the exact solution of the Schrödinger equation. The computational cost of FermiNet or PauliNet models scales as N^4 (N being the number of explicit electrons), in contrast to highly accurate methods of quantum chemistry having N^7 or steeper scaling. Achieving such efficiency requires physically motivated ML architectures, accounting for the antisymmetric nature and correct asymptotic behavior of valid electronic wave functions.

Summarizing our discussion, we can state that recent advances in the development of MLFFs in physics and chemistry have substantially expanded the horizons of atomistic modeling. The state-of-the-art MLFF models can be simultaneously data-efficient, accurate, and applicable to various classes of molecules and materials and can be evaluated with a rather low computational cost. Moreover, recent developments have concentrated on incorporating an increasing amount of physical covariances and invariances, as well as conservation laws, and knowledge about intramolecular and intermolecular interactions. Fulfilling these requirements transforms ML techniques into unique and novel approaches, contributing to the advancement of both natural sciences and the ML models themselves.

Despite substantial progress in the field of MLFFs, many challenges remain to be addressed on the path toward broadly applicable models and universally accurate predictions. Multiple aspects of MLFF models should be revised and further improved. For example, we need rigorous estimation of applicability ranges of different MLFFs, moving from good overall predictions to models equally reliable for all the relevant parts of the configuration space, improving the training-data selection schemes, incorporating many different types of long-range interactions, improving atomistic representations, understanding the metric in descriptor spaces, refining loss functions, and developing novel dimensionality reduction schemes. Of particular importance are anomaly detection techniques to prevent extrapolations and a stronger embedding of ML models into existing computational protocols. Solutions to these and

many more remaining challenges require creative methods that combine expertise in physics, chemistry, and machine learning.

AUTHOR INFORMATION

Corresponding Author

Alexandre Tkatchenko – Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg; orcid.org/0000-0002-1012-4854; Email: alexandre.tkatchenko@uni.lu

Author

Igor Poltavsky – Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg; orcid.org/0000-0002-3188-7017

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpclett.1c01204>

Notes

The authors declare no competing financial interest.

Biographies

Igor Poltavsky is a senior researcher at the University of Luxembourg. He received his Ph.D. in Theoretical Physics from B. Verkin Institute for Low Temperature Physics Engineering in 2009. His research interests include statistical physics, imaginary-time path integral methods, nuclear quantum effects, ab initio simulations, and machine learning.

Alexandre Tkatchenko is a professor at the Department of Physics and Materials Science (and head of this department since January 2020) at the University of Luxembourg, where he holds a chair in Theoretical Chemical Physics. Tkatchenko also holds a distinguished visiting professor position at the Technical University of Berlin. His group develops accurate and efficient first-principles computational models to study a wide range of complex materials, aiming at qualitative understanding and quantitative prediction of their structural, cohesive, electronic, and optical properties at the atomic scale and beyond. Tkatchenko serves on the editorial boards of *Science Advances* and *Physical Review Letters* and has received several awards, including APS Fellow from the American Physical Society; Gerhard Ertl Young Investigator Award of the German Physical Society; Dirac Medal from the World Association of Theoretical and Computational Chemists (WATOC); van der Waals prize of ICNI-2021; and three flagship grants from the European Research Council: a Starting Grant in 2011, a Consolidator Grant in 2017, and Proof-of-Concept Grant in 2020.

ACKNOWLEDGMENTS

We acknowledge Prof. John Keith and Mr. Alex Maldonado for permission to use unpublished results for Figure 5. We acknowledge financial support from the Luxembourg National Research (FNR) (Grant C19/MS/13718694/QML-FLEX) and the European Research Council (ERC-CoG grant BeStMo).

REFERENCES

- (1) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **2019**, *365*, eaaw1147.
- (2) Senior, A. W.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710.
- (3) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, 114102.
- (4) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Phys. Rev. X* **2018**, *8*, 041048.
- (5) Lookman, T.; Eidenbenz, S.; Alexander, F.; Barnes, C. *Materials Discovery and Design by Means of Data Science and Optimal Learning*; Springer International Publishing: Basel, 2018.
- (6) Ryan, K.; Lengyel, J.; Shatruk, M. Crystal Structure Prediction via Deep Learning. *J. Am. Chem. Soc.* **2018**, *140*, 10158–10168.
- (7) Graser, J.; Kauwe, S. K.; Sparks, T. D. Machine Learning and Energy Minimization Approaches for Crystal Structure Predictions: A Review and New Horizons. *Chem. Mater.* **2018**, *30*, 3601–3612.
- (8) Balachandran, P. V.; Kowalski, B.; Sehirlioglu, A.; Lookman, T. Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nat. Commun.* **2018**, *9*, 1668.
- (9) Li, W.; Jacobs, R.; Morgan, D. Predicting the thermodynamic stability of perovskite oxides using machine learning models. *Comput. Mater. Sci.* **2018**, *150*, 454–463.
- (10) Faber, F. A.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R. Machine Learning Energies of 2 Million Elpasolite (ABC_2D_6) Crystals. *Phys. Rev. Lett.* **2016**, *117*, 135502.
- (11) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **2019**, *5*, 83.
- (12) Kang, Y.; Li, L.; Li, B. Recent progress on discovery and properties prediction of energy materials: Simple machine learning meets complex quantum chemistry. *J. Energy Chem.* **2021**, *54*, 72–88.
- (13) Sun, W.; Zheng, Y.; Yang, K.; Zhang, Q.; Shah, A. A.; Wu, Z.; Sun, Y.; Feng, L.; Chen, D.; Xiao, Z.; Lu, S.; Li, Y.; Sun, K. Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials. *Sci. Adv.* **2019**, *5*, eaay4275.
- (14) Brunton, S. L.; Noack, B. R.; Koumoutsakos, P. Machine Learning for Fluid Mechanics. *Annu. Rev. Fluid Mech.* **2020**, *52*, 477–508.
- (15) Fooshee, D.; Mood, A.; Gutman, E.; Tavakoli, M.; Urban, G.; Liu, F.; Huynh, N.; Van Vranken, D.; Baldi, P. Deep learning for chemical reaction prediction. *Mol. Syst. Des. Eng.* **2018**, *3*, 442–452.
- (16) Walker, E.; Kammeraad, J.; Goetz, J.; Robo, M. T.; Tewari, A.; Zimmerman, P. M. Learning To Predict Reaction Conditions: Relationships between Solvent, Molecular Structure, and Catalyst. *J. Chem. Inf. Model.* **2019**, *59*, 3645–3654.
- (17) Schwaller, P.; Vaucher, A. C.; Laino, T.; Reymond, J.-L. Prediction of chemical reaction yields using deep learning. *Mach. Learn.: Sci. Technol.* **2021**, *2*, 015016.
- (18) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. In *Machine Learning Meets Quantum Physics*; Schütt, K. T., Chmiela, S., von Lilienfeld, O. A., Tkatchenko, A., Tsuda, K., Müller, K.-R., Eds.; Springer International Publishing: Cham, 2020; pp 277–307.
- (19) Sauceda, H. E.; Vassilev-Galindo, V.; Chmiela, S.; Müller, K.-R.; Tkatchenko, A. Dynamical strengthening of covalent and non-covalent molecular interactions by nuclear quantum effects at finite temperature. *Nat. Commun.* **2021**, *12*, 442.
- (20) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3*, e1603015.
- (21) Chmiela, S.; Sauceda, H. E.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. sGDML: Constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* **2019**, *240*, 38–45.
- (22) Schütt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Müller, K.-R. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15*, 448–455.
- (23) Ceriotti, M.; Willatt, M. J.; Csányi, G. In *Handbook of Materials Modeling: Methods: Theory and Modeling*; Andreoni, W., Yip, S., Eds.; Springer International Publishing: Cham, 2020; pp 1911–1937.
- (24) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

- (25) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.
- (26) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (27) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.
- (28) Stöhr, M.; Medrano Sandomas, L.; Tkatchenko, A. Accurate Many-Body Repulsive Potentials for Density-Functional Tight Binding from Deep Tensor Neural Networks. *J. Phys. Chem. Lett.* **2020**, *11*, 6835–6843.
- (29) Sze, V.; Chen, Y.-H.; Yang, T.-J.; Emer, J. S. Efficient processing of deep neural networks. *Synth. Lectures Comput. Archit.* **2020**, *15*, 1.
- (30) Vinyals, O.; et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* **2019**, *575*, 350–354.
- (31) Badia, A. P.; Piot, B.; Kapturowski, S.; Sprechmann, P.; Vitvitskiy, A.; Guo, Z. D.; Blundell, C. Agent57: Outperforming the Atari Human Benchmark. *Proceedings of the 37th International Conference on Machine Learning* **2020**, 507–517.
- (32) Afouras, T.; Chung, J. S.; Senior, A.; Vinyals, O.; Zisserman, A. Deep Audio-visual Speech Recognition. *IEEE Trans* **2019**, *1*.
- (33) Chung, J. S.; Senior, A.; Vinyals, O.; Zisserman, A. *Lip Reading Sentences in the Wild*. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2017**, 3444–3453.
- (34) Donahue, J.; Dieleman, S.; Bińkowski, M.; Elsen, E.; Simonyan, K. *End-to-End Adversarial Text-to-Speech*; 2020 (<https://arxiv.org/abs/2006.03575>).
- (35) Bahdanau, D.; Chorowski, J.; Serdyuk, D.; Brakel, P.; Bengio, Y. *End-to-end attention-based large vocabulary speech recognition*. 2016 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* **2016**, 4945–4949.
- (36) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.
- (37) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 2903.
- (38) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
- (39) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning For Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361–390.
- (40) Schütt, K. T.; Chmiela, S.; von Lilienfeld, O. A.; Tkatchenko, A.; Tsuda, K.; Müller, K.-R. *Machine Learning Meets Quantum Physics*; Springer, 2020.
- (41) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, And Outlook. *J. Phys. Chem. C* **2017**, *121*, 511–522.
- (42) Behler, J. First Principles Neural Network Potentials For Reactive Simulations Of Large Molecular And Condensed Systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828–12840.
- (43) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials As Emerging Tools For Materials Science. *Adv. Mater.* **2019**, *31*, 1902765.
- (44) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, DOI: 10.1021/acs.chemrev.0c01111.
- (45) Huang, B.; von Lilienfeld, O. A. Ab initio machine learning in chemical compound space. *arXiv* **2020**, 2012.07502v3.
- (46) Musil, F.; Grisafi, A.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M. Physics-inspired structural representations for molecules and materials. *arXiv* **2021**, 2101.04673.
- (47) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.
- (48) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.
- (49) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (50) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.
- (51) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13*, 5255–5264.
- (52) Goscinski, A.; Fraux, G.; Imbalzano, G.; Ceriotti, M. The role of feature space in atomistic learning. *Mach. learn.: sci. technol.* **2021**, *2*, 025028.
- (53) Fonseca, G.; Poltavsky, I.; Vassilev-Galindo, V.; Tkatchenko, A. Improving molecular force fields across configurational space by combining supervised and unsupervised machine learning. *J. Chem. Phys.* **2021**, *154*, 124102.
- (54) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.
- (55) Stöhr, M.; Tkatchenko, A. Quantum mechanics of proteins in explicit water: The role of plasmon-like solute-solvent interactions. *Sci. Adv.* **2019**, *5*, eaax0024.
- (56) Ambrosetti, A.; Ferri, N.; DiStasio, R. A.; Tkatchenko, A. Wavelike charge density fluctuations and van der Waals interactions at the nanoscale. *Science* **2016**, *351*, 1171–1176.
- (57) French, R. H.; et al. Long range interactions in nanoscale science. *Rev. Mod. Phys.* **2010**, *82*, 1887–1944.
- (58) Gupta, S.; Ruffo, S. The World of Long-Range Interactions: A Bird's Eye View. *Memorial Volume on Abdus Salam's 90th Birthday*; **2017**; pp 303–345.
- (59) D'Avino, G.; Muccioli, L.; Castet, F.; Poelking, C.; Andrienko, D.; Soos, Z. G.; Cornil, J.; Beljonne, D. Electrostatic phenomena in organic semiconductors: fundamentals and implications for photovoltaics. *J. Phys.: Condens. Matter* **2016**, *28*, 433002.
- (60) Gebbie, M. A.; Smith, A. M.; Dobbs, H. A.; Lee, A. A.; Warr, G. G.; Banquy, X.; Valtiner, M.; Rutland, M. W.; Israelachvili, J. N.; Perkin, S.; Atkin, R. Long range electrostatic forces in ionic liquids. *Chem. Commun.* **2017**, *53*, 1214–1224.
- (61) Prodan, E.; Kohn, W. Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 11635–11638.
- (62) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1327.
- (63) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (64) Vassilev-Galindo, V.; Fonseca, G.; Poltavsky, I.; Tkatchenko, A. Challenges for machine learning force fields in reproducing potential energy surfaces of flexible molecules. *J. Chem. Phys.* **2021**, *154*, 094119.
- (65) Herr, J. E.; Yao, K.; McIntyre, R.; Toth, D. W.; Parkhill, J. Metadynamics for training neural network model chemistries: A competitive assessment. *J. Chem. Phys.* **2018**, *148*, 241710.
- (66) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.
- (67) Grisafi, A.; Ceriotti, M. Incorporating long-range physics in atomic-scale machine learning. *J. Chem. Phys.* **2019**, *151*, 204105.
- (68) Morawietz, T.; Sharma, V.; Behler, J. A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges. *J. Chem. Phys.* **2012**, *136*, 064103.
- (69) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12*, 398.
- (70) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion

correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(71) Morawietz, T.; Behler, J. A Density-Functional Theory-Based Neural Network Potential for Water Clusters Including van der Waals Corrections. *J. Phys. Chem. A* **2013**, *117*, 7356–7366.

(72) Botu, V.; Ramprasad, R. Adaptive machine learning framework to accelerate ab initio molecular dynamics. *Int. J. Quantum Chem.* **2015**, *115*, 1074–1083.

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

(74) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Demonstrating the Transferability and the Descriptive Power of Sketch-Map. *J. Chem. Theory Comput.* **2013**, *9*, 1521–1532.

(75) Wales, D. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*; Cambridge Molecular Science; Cambridge University Press, 2004.

(76) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. Δ -machine learning for potential energy surfaces: A PIP approach to bring a DFT-based PES to CCSD(T) level of theory. *J. Chem. Phys.* **2021**, *154*, 051102.

(77) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.

(78) Pfau, D.; Spencer, J. S.; Matthews, A. G. D. G.; Foulkes, W. M. C. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Phys. Rev. Research* **2020**, *2*, 033429.

(79) Hermann, J.; Schätzle, Z.; Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **2020**, *12*, 891–897.