Check for updates

Extending machine learning beyond interatomic potentials for predicting molecular properties

Nikita Fedik[1,2,3], *Roman Zubatyuk*[4], *Maksim Kulichenko*[1,3], *Nicholas Lubbers*[5], *Justin S. Smith*[1,6], *Benjamin Nebgen*[1], *Richard Messerly*[1], *Ying Wai Li*[5], *Alexander I. Boldyrev*[3], *Kipton Barros*[1,2], *Olexandr Isayev*[4] *and Sergei Tretiak*[1,2,7]✉

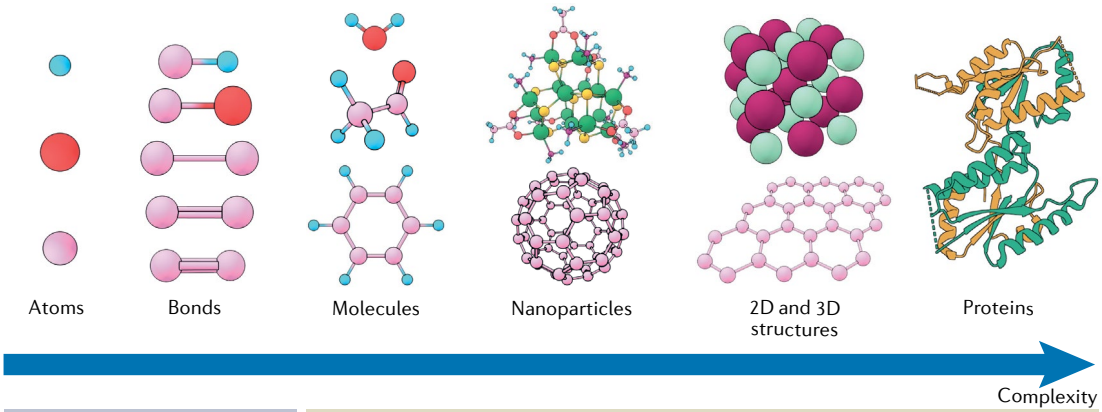
Abstract | Machine learning (ML) is becoming a method of choice for modelling complex chemical processes and materials. ML provides a surrogate model trained on a reference dataset that can be used to establish a relationship between a molecular structure and its chemical properties. This Review highlights developments in the use of ML to evaluate chemical properties such as partial atomic charges, dipole moments, spin and electron densities, and chemical bonding, as well as to obtain a reduced quantum-mechanical description. We overview several modern neural network architectures, their predictive capabilities, generality and transferability, and illustrate their applicability to various chemical properties. We emphasize that learned molecular representations resemble quantum-mechanical analogues, demonstrating the ability of the models to capture the underlying physics. We also discuss how ML models can describe non-local quantum effects. Finally, we conclude by compiling a list of available ML toolboxes, summarizing the unresolved challenges and presenting an outlook for future development. The observed trends demonstrate that this field is evolving towards physics-based models augmented by ML, which is accompanied by the development of new methods and the rapid growth of user-friendly ML frameworks for chemistry.

Chemistry encompasses an enormous body of knowledge gained over several centuries. Chemical knowledge is many-faceted, as are the techniques that can be used to obtain and analyse chemical data. When analysing a specific property of a chemical structure, it is necessary to consider various scales. Some properties can be attributed to specific regions of a structure, such as fragments, bonds or even atoms, and such local properties can provide insight into the global properties and functionality of the structure. Often, the overall structural motif is dissected to try to comprehend the individual local contributions. This picture is complicated by the presence of long-range phenomena such as electrostatic interactions and charge transfer, which can be dominant factors in determining many chemical properties. Accordingly, the chemical mindset evenly considers both the local and the global properties and usually classifies atomistic objects by their complexity (FIG. 1).

Local properties on the atomic scale (such as atomic charges and hybridization) and global properties on the molecular scale (such as dipole moments and ground-state and excited-state energies) have become central to the chemical mindset and the practical vocabulary for the description of fundamental concepts and design applications. Subsequently, these properties are the primary targets for experimental and theoretical studies. Potentially, all of these properties could be inferred from first principles electronic structure calculations that computationally solve the Schrödinger equation. However, in practice, exact solutions are rarely numerically tractable. As a result, an extensive lineup of methods with varying fidelities has been developed, ranging from extremely accurate wavefunction methods (such as coupled cluster techniques[1]) to the practical and widely used density functional theory (DFT)[2,3] and to lower-accuracy semiempirical approaches[4]. As a rule of thumb, greater accuracy comes with a steep increase in computational demand, with up to an exponential scaling with respect to the number of electrons[5]. Even after decades of development, the applicability of conventional electronic structure theory is still limited by the fundamental scaling of the underlying numerical methods. The advent of machine learning (ML) in computational chemistry has offered a new approach that

✉*e-mail:* serg@lanl.gov

https://doi.org/10.1038/s41570-022-00416-3



Local properties		Global properties			
Atomic energy	Atomic charge	Total energy	Orbital energies	Electron density	Dipoles
Bond order		Reactivity	Bandgap	Optical response	Transport

Fig. 1 | **Coarse atomistic scale of matter from the perspective of a chemist.** Local properties are related to the essential structural elements such as atoms, bonds or fragments, whereas global properties are attributed to the entire system. Two-dimensional (2D) and three-dimensional (3D) structures are usually treated as periodic; thus, proteins as a bulk non-periodic example conclude the scale of increasing complexity.

can provide rapid solutions for large-scale systems[6–8]. Advances in ML for chemistry clearly indicate that atomic and molecular properties could be 'learnable' by machines, making it possible to overcome the conventional limitations and numerical barriers outlined above. For example, the development of ML interatomic potentials[6,7,9,10] has enabled massive molecular dynamics studies at a computational cost almost as low as that of classical force fields without sacrificing the accuracy of quantum mechanics, by bypassing the explicit solutions of the quantum mechanical (QM) approaches. These advances have notable implications in fields such as drug design[11–13] and materials development[14,15]. The ability of ML to establish an efficient and accurate surrogate model for mapping different variables (both theoretical and experimental) has quickly made it a method that is well adapted to uncovering chemical structure–property relationships (BOX 1, FIG. 2).

This Review is written with a few specific goals in mind to augment the large body of available literature on the use of ML for chemical discovery. First, this Review is centred on machine learnable chemical properties rather than algorithms and their evolution. A general theme of the narrative is how local properties contribute to non-local and global molecular properties, and eventually to properties of supramolecular complexes and bulk materials. The structure of this Review is dictated by the hierarchy of the intuitive length scales and is not necessarily aligned with the history of specific ML algorithms. When possible, figures and discussions follow a bottom-up scheme, permitting an investigation of the same phenomena from both local and global perspectives. Second, many of the properties outlined (such as dipole moments) are physical observables, which implies that ML models can be validated with (or perhaps even trained on) experimental data, inviting a fruitful collaboration between experimentalists and computational chemists. We link the learned chemical properties with true physical quantities whenever possible. Third, we overview a rapidly emerging set of reduced quantum chemical methods, such as density-functional tight-binding (DFTB), and indicate how ML can directly improve electronic structure calculations. We refrain from reviewing the use of ML interatomic potentials for predicting total energies or running dynamics of chemical systems, because these advances have been well described elsewhere[16–18]. Other relevant areas that have been recently reviewed include the application of ML to investigate electronically excited states[19,20], chemical reactivity[21,22], catalysis[23], drug discovery[11,12,24], battery materials[15] and the control of organic synthesis[25,26]. Given the large variety of ML approaches available for chemical discovery applications, neural networks (NNs) are among the most ubiquitous and powerful ML algorithms. This Review focuses primarily on predictive models based on NNs. Other useful and promising models such as the Gaussian approximation potential (GAP)[27], spectral neighbour analysis potential (SNAP)[28], moment tensor potentials (MTPs)[29] and symmetric gradient domain learning (sGDML)[30,31] are reviewed elsewhere.

In the following sections, we highlight several selected NN architectures and discuss how they are employed to train extensible and transferable ML models that elucidate diverse chemical properties of interest.

Author addresses

[1]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA.
[2]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA.
[3]Department of Chemistry and Biochemistry, Utah State University, Logan, UT, USA.
[4]Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA, USA.
[5]Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA.
[6]NVIDIA, Santa Clara, CA, USA.
[7]Center for Integrated Nanotechnologies, Los Alamos National Laboratory, Los Alamos, NM, USA.

Architectures of atomistic neural networks

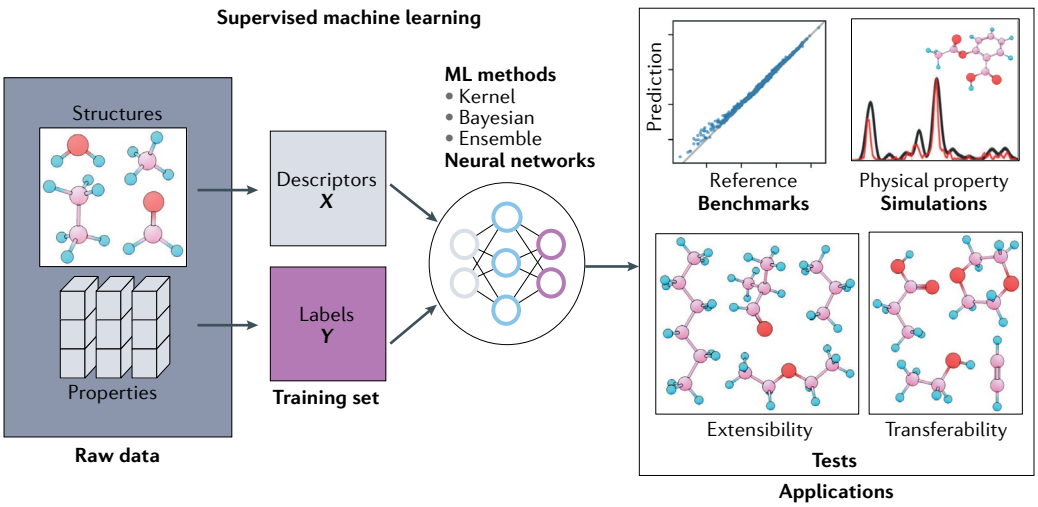
One of the main characteristics that determines the accuracy, computational performance and applicable domain of atomistic ML models is the method used for the molecular geometry representation or encoding. Following common chemical intuition, most models are based on the modelling assumption of spatial locality, which only permits interactions between nearby atoms. The local decomposition ansatz builds on the principle of locality, whereby the reference global property, P^{ref}, which is attributed to the entire system, can be approximated by P, and partitioned into learnable local contributions P_i. The system property P is then reconstructed as the sum of the local (atomic) properties: $P^{ref} \approx P = \sum_{i=1}^{N_{atoms}} P_i$, where N_{atoms} is the number of atoms in the system[32–34]. The essence of such an expression is the introduction of a special representation of the local atomic environment within a cut-off sphere (usually 5–8 Å in radius). Such representations are many-body functions of the atomic positions inside the

Box 1 | **Machine learning for chemical properties**

In general, machine learning (ML) approaches aim to formulate a surrogate model by establishing an implicit relationship between chemical structures and one or more chemical properties[42,167]. Typically, crafting a predictive ML model with supervised learning[168] involves iterative training on a reference 'ground-truth' dataset that is composed of quantum mechanical (QM) simulation data of the target properties (or labels *Y*) derived for all structures[61,69,169]. A properly trained model can be used as a black-box tool. The training and application of ML models require an input of chemical structures (such as molecular specifications) that can be represented as arrays of atomic numbers and the respective positions of the atoms. These arrays are then converted to a machine-readable form (a descriptor *X*) (BOX 2). Using a chemical structure as an input (*X*) and targeting energy as an output (*Y*) underpins the design of ML interatomic potentials (see the figure)[17,170]. This practice is analogous to the design of classical force fields[171,172]. In contrast with the development of classical force fields, which requires tedious manual reparameterization[172], the training of ML models relies on the automatic adjustment of parameters to achieve the best accuracy for a given dataset with no or minimal human intervention[173]. This automatic adjustment of parameters is done by minimizing a cost function that measures the error of the model using the difference between a reference value Y^{ref} and the ML-predicted *Y* averaged across all training points (*N*). A simple and popular choice of cost function is a root-mean-squared error (RMSE) expressed as $\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - Y_i^{ref})^2}$.

Supervised ML approaches include linear models with carefully designed descriptors, kernel methods such as Gaussian process regression, and artificial neural networks (NNs)[98,174]. NNs are universal approximators[175,176], capable of identifying complex nonlinear relationships between data attributes, which may be far beyond human intuition. Although NNs can 'learn' almost anything, caution should be exercised with how they are 'taught'. ML models are very data-sensitive, which implies that data collection, generation and refinement are of paramount importance[38,60,61]. An essential area of research is active learning, in which new data is iteratively collected according to the observed shortcomings of the model[32,60,144]. Best practices for data collection are beyond the scope of this Review and are discussed elsewhere[14,32,60,61,143,144].

Highly flexible ML models benefit from data-driven approaches. When the training data is adequately representative and the training is successful, ML can make accurate predictions outside the training and test sets[52,60,90]. Here, we distinguish between extensibility and transferability. Extensibility is the applicability of the model to much larger systems than those used during the training. For instance, an extensible model, trained on configurations of small alcohols (such as methanol (CH_3OH)) should be able to accurately predict the properties of much larger homologues, such as C_3H_7OH. Transferability suggests that the model is accurate for systems that are structurally different to the reference (having different dimensionality, shape or number of functional groups). Taking the previous example, a transferable model trained on methanol should be able to accurately reproduce other molecular families comprising the same atoms (H, C and O) and functional groups (–OH), such as diols or pure hydrocarbons. Importantly, transferability is always limited[177,178]. In the example above, the prediction of aromatic species, such as benzene, would probably fail because the ML model was not aware of carbon hybridization and electronic delocalization effects. Nevertheless, NNs could potentially evaluate some properties beyond the labels *Y*, paving the way towards predictions of global properties using only local property information. For example, training to atomic charges can enable the prediction of molecular dipole moments (FIG. 2).



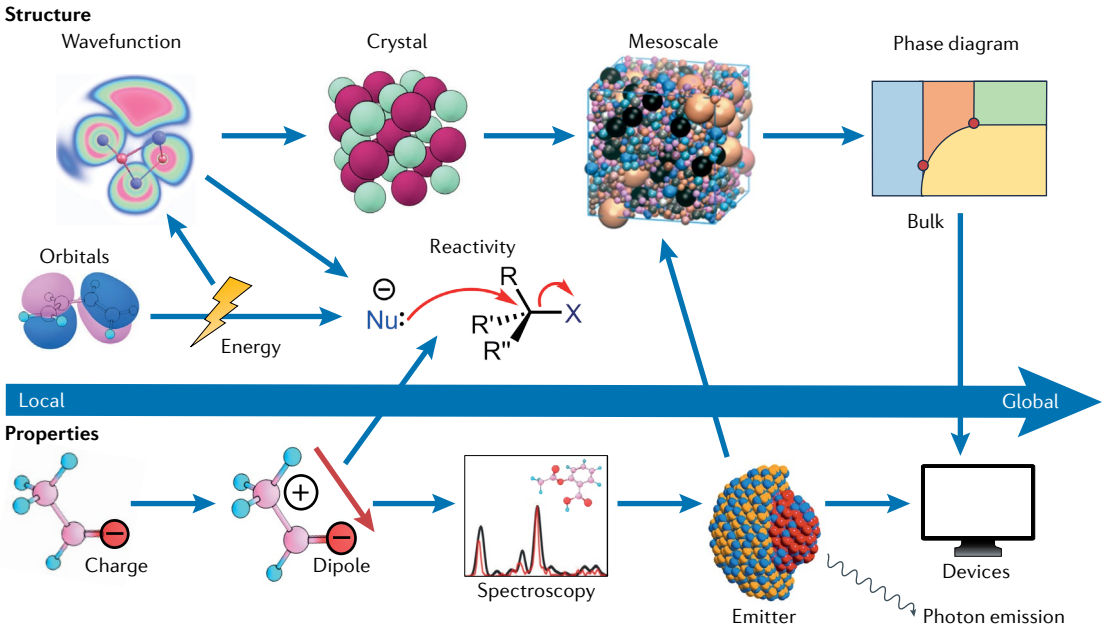


Fig. 2 | **Relationships between chemical structure and properties from local and global perspectives.** Some correlations are intuitive rather than rigorous. For example, charges and dipoles are defined by the quantum-mechanical wavefunction. The dynamics of these systems are reflected in the measurable spectra and underpin desirable properties such as the emission of photons.

Extensive properties
Properties that intrinsically grow with system size. Examples include enthalpies of formation, which scale with the number of chemical bonds in a system, and entropies of solvation.

Extensibility
The applicability of a ML model to systems substantially larger than those in the training set. An example of extensibility is a ML model trained to small organic molecules that can accurately simulate a protein.

Interaction layers
Groups of NN operations that transfer information between atoms. Typically expressed as a local graph convolution, this operation creates new features for each atom based on the features and distances of local neighbouring atoms. Also called message-passing.

Charge equilibration scheme
An approach for predicting charge distributions in molecules or solids using tabulated or ML-predicted atomic properties such as atomic electronegativity and hardness, or bond polarizability.

Transferability
The applicability of a ML model to systems not originally included in the training set. Transferable models exhibit high accuracy for chemical systems that are structurally different from the ones in the training set.

cut-off region (BOX 2). The local decomposition ansatz is very effective at approximating extensive properties (for example, total molecular energy) and closely resembles classical many-body potentials (such as embedded atom models)[35]. However, in contrast to many-body expansion, NNs are not rigid functional expressions but flexible ML models. Models that are based on the locality assumption have computational costs that scale linearly with respect to the system size, and consistent prediction accuracies that facilitate the extensibility of the models.

The locality assumption fails to capture the long-range effects that can arise from charge transfer, polarization and electrostatic or dispersion interactions[17,34]. The importance of including these long-range interactions is becoming more pronounced as the capabilities of ML models continue to increase. In the context of the interplay of local and global properties spanning various length scales (FIGS. 1 and 2), the lack of long-range interactions limits the overall model quality. Here, we outline the general features of ML models (including the introduction of interaction layers, self-consistent field (SCF)-like updates of local atomic descriptors, and charge equilibration schemes) and showcase their performance for chemical problems with non-local phenomena. Although these approaches can account for long-range interactions to some extent, cases with fully delocalized electronic wavefunctions and non-local excited-state properties are more challenging. One possible approach is to retain minimalistic QM treatment using effective model Hamiltonians, with parameters that are tuned to obtain observables that are comparable with those obtained with more accurate calculations. This task can also be accomplished with NN architectures, as illustrated in the selected examples that follow.

Behler and Parrinello pioneered the idea of modelling interatomic potentials as a sum of per-atom NN predicted contributions[33,36–38] (FIG. 3a). In the Behler–Parrinello architecture, atomic coordinates are transformed into atom-centred symmetry functions (ACSFs)[37] using interatomic distances with a smooth distance cut-off (BOX 2). Summation over all the neighbouring atoms incorporates radial and angular information, leading to useful descriptors that account for the local environment of the central atom. In the Behler–Parrinello formalism, each chemical element is described using its own high-dimensional neural network (HDNN), and the outputs of the atomic NNs are summed to obtain a total energy. The assumption of locality is central to Behler–Parrinello architectures for the prediction of extensive properties, such as total molecular energies. Behler–Parrinello architectures have been substantially improved across four generations of development[17,39]. In particular, the latest model can assess non-local effects such as long-range electron transfer[39]. This development is achieved by using environment-dependent atomic electronegativities and an additional charge equilibration scheme in the NN term. The Behler–Parrinello approach has been used to perform highly accurate simulations of a wide range of inorganic materials, and an exhaustive overview of Behler–Parrinello architectures is reported elsewhere[17].

However, the original ACSFs were constructed for each pair of elements, meaning that the transferability of the techniques was limited. A modification of the ACSFs, to tune the angular part of the descriptor and account for the atomic numbers of neighbouring atoms, led to the development of the accurate neural network engine for molecular energies (ANI). ANI models have demonstrated good transferability over a large chemical

One-hot scheme
A binary vector representation for unranked categorical values. For example, three atom types C, H and O are represented by the vectors (1, 0, 0), (0, 1, 0) and (0, 0, 1), respectively.

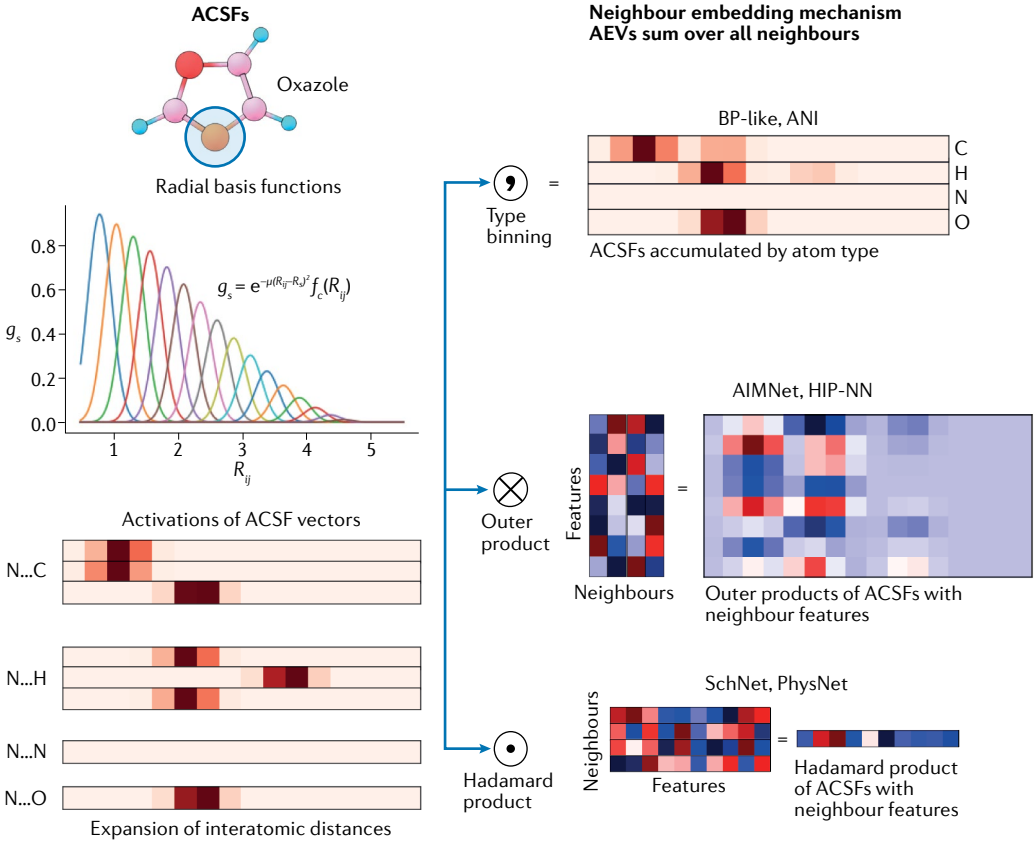
space of organic molecules[40]. Nevertheless, the ANI architecture still has the other inherent limitations of the Behler–Parrinello model; for example, the size of the descriptor increases quadratically with the number of chemical elements. Also, reparameterization for a new set of chemical elements requires training from scratch. Reported models include up to nine (H, C, N, O, F, S, Cl, Br and I) chemical elements[41].

The recent generations of ML models, such as the hierarchically interacting particle neural network (HIP-NN)[6], atoms-in-molecules network (AIMNet)[42], SchNet[43,44], PhysNet[45] and DimeNet[46,47] address the problem of extending the model to a large number of chemical species by introducing a learnable atomic representation (or atomic embedding vectors) for different atom types. These representations are either initialized randomly or according to a one-hot scheme for each atom. To form a final descriptor, also called the atomic environment vector (AEV), embedding should incorporate structural information about the neighbours.

Box 2 | Atomic environment descriptors and architectures

Neural network (NN) potential architectures frequently use interatomic distances, in the basis of atom-centred symmetry functions (ACSFs), to obtain descriptors, which serve as inputs to the NN. This collection of descriptors is the atomic environment vector (AEV). Beyond radial ACSFs, NN models include higher-order many-body descriptors, such as angular[37,42,90], vector or tensor[179,180], or spherical harmonics[181]. The optimal functional form of the AEV descriptor might be specific to the model, dataset or task, such as the atom-pair symmetry functions in AP-Net[79] that target intermolecular interactions.

In the original Behler–Parrinello architecture and the ANI model, the AEV is the concatenation of sums of ACSFs for the chemical type of each neighbouring atom. For example, in the oxazole molecule (see the figure), the AEV for the nitrogen atom contains the ACSF vector for the neighbouring oxygen atom, the sum of the ACSFs for three carbon atoms, the sum of the ACSFs for three hydrogen atoms, and an empty vector for the neighbouring nitrogen atoms (as there are none). The AEV size, and the weights of the NN model, strongly depend on the number of chemical elements in the parametrization.

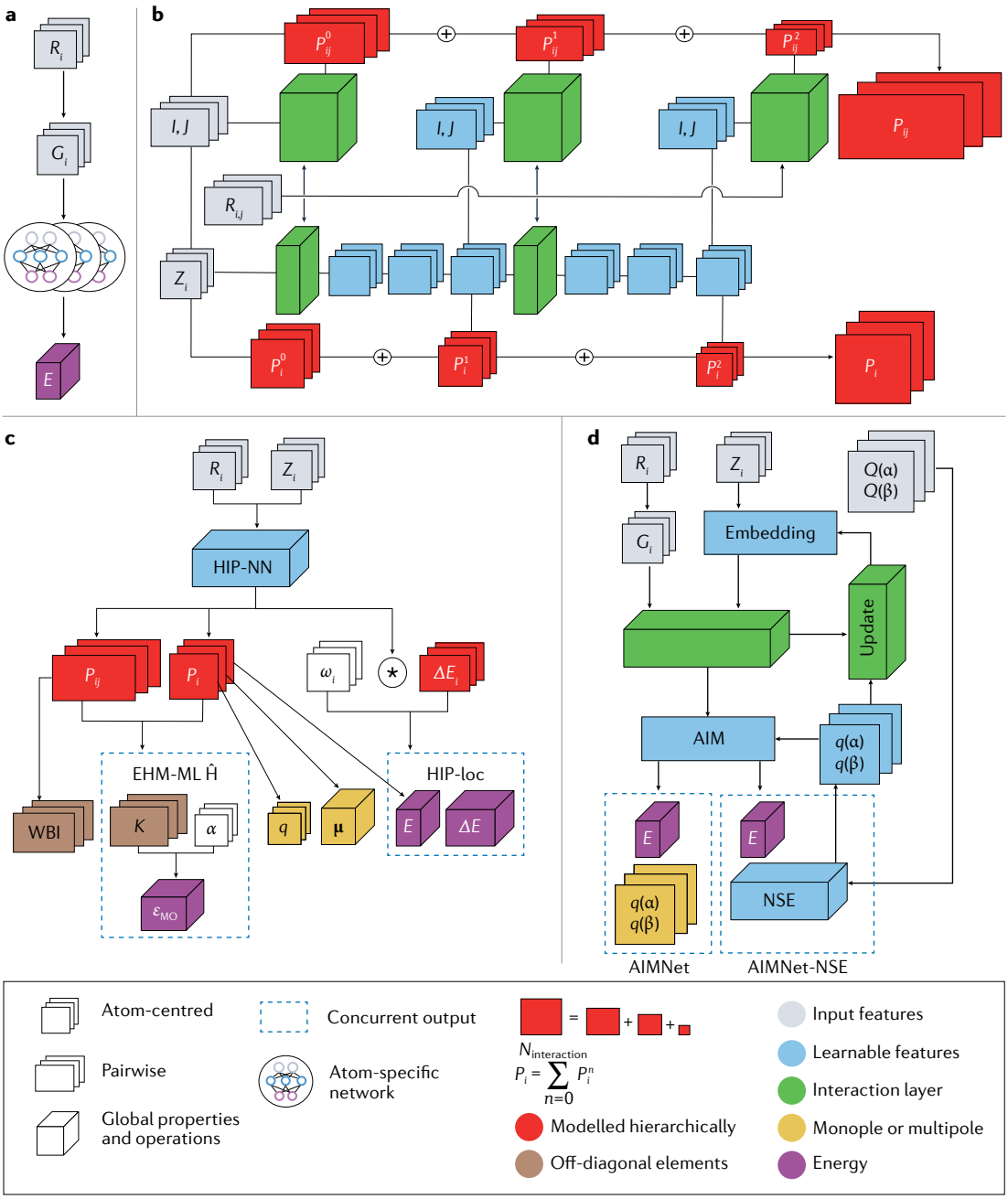


When encoding the environment of the nitrogen atom in the oxazole molecule a set of 16 Gaussian ACSFs (g_s) with shifts (R_s) and width parameter (μ) in the range 0.8–4.8 Å applied to all the pairwise distances (R_{ij}) from the nitrogen atom to the neighbours, results in $\mathbb{R}^{16}$ vectors with one or more elements activated that correspond to R_{ij} close to R_s. f_c is a cut-off function that ensures locality and smoothly zeros the function beyond the cut-off radius. This spatial information is combined with the atom types, either by concatenation or by multiplication with learnable atomic embedding vectors, to form the AEV.

As mentioned above, many models[42,43,45,50] introduce a learnable atomic embedding. The final AEV is a product of the atomic embeddings and Behler–Parrinello-style ACSFs and does not depend on the number of elements. The atoms-in-molecules network (AIMNet) model calculates the outer product of the atomic embedding vector with a vector of ACSFs, then constructs a sum of the products of the neighbouring atoms. For an atom with M neighbours, with Gaussian basis functions for ACSFs of length N and an atomic embedding vector of length K, the matrix of ACSFs $\mathbf{G} \in \mathbb{R}^{(M,N)}$ and atom feature matrix $\mathbf{A} \in \mathbb{R}^{(M,K)}$, the AEV ($\mathbf{X}$) in the AIMNet model is: $X_{nk} = \Sigma_m G_{nm} A_{mk}$ or $\mathbf{X} = \mathbf{G} \otimes \mathbf{A}$. Similarly, hierarchically interacting particle NN (HIP-NN) models provide an outer product of the atomic features with ACSFs that are based on pairwise distances.

The SchNet and PhysNet models construct learnable ACSFs, G_{km}, one for each feature in the environment, using an inner NN. The output features X_k are generated using element-wise (Hadamard) multiplication with the features of the neighbouring atoms: $X_k = \Sigma_m G_{km} A_{mk}$ or $\mathbf{X} = \mathbf{G} \odot \mathbf{A}$. These Hadamard-type models lead to layers with fewer parameters than outer product models, but each layer considers each type of feature in the environment separately, relying on further layers to mix information between the channels. Outer-product models typically use two or three interaction layers, and Hadamard-product models often use five or more interactions. Thus, Hadamard-product featurization produces models with a larger total cut-off range.

Another aspect is the complexity of the symmetry functions used to build NN potential architectures. The HIP-NN, SchNet and PhysNet models use only radial ACSFs that encode interatomic distances, but do not directly contain any angular information. However, angular information is indirectly available through a message-passing approach: the angular sensitivity is encoded by iteratively updating the atomic feature vectors with the distribution of the neighbouring atoms. By contrast, the ANI and AIMNet models explicitly incorporate the angular symmetry functions.



The AEV is obtained as a matrix product of the atomic embedding and the corresponding ACSFs. This procedure is usually known as graph convolution and is accomplished using an architecture-specific interaction layer (in SchNet this interaction layer is called cfconv; in PhysNet, an interaction block; in AIMNet, an embedding block; and in HIP-NN, it is referred to simply as an interaction layer). In these interaction layers, edge features are defined in terms of interatomic distances and node features are given by atomic embeddings. We note that in the common terminology, which is used in this Review, atomic embedding does not directly incorporate the local environment information (unlike AEV); however, in the original AIMNet notation[42], the final AEV augmented by structural information is also called an embedding. Although local approximations with ACSFs lead to high computational efficiencies and in many cases high or reasonable accuracy, a simple extension of the architecture makes it possible to capture semi-local effects. With the message-passing approach[48,49], the atomic features depend not only on the positions of the immediate neighbours of the atom, but also on the positions of neighbours of the neighbours. Most modern NN potential architectures use message passing to increase the accuracy of prediction.

HIP-NN (FIG. 3b) extends the notion of locality by including intermediate-range contributions (BOX 2). Mathematically, in the HIP-NN formalism, P_i is further decomposed into hierarchical contributions of order n: $P_i = \sum_{n=0}^{N_{\text{interaction}}} P_i^n$ and typically $|P_i^n| > |P_i^{n+1}|$, where $N_{\text{interaction}}$ is the number of interaction layers. The term P_i^n represents a contribution at the n^{th} interaction

Atomic embedding
A fixed-length vector representation of an atom with no information about its environment. Embedding is a learnable map between the atomic number and a high-dimensional feature space. Can incorporate additional atomic physicochemical properties, such as electronegativity.

◀ Fig. 3 | **Modern architectures of neural networks for learning local and global properties. a** | Behler–Parrinello high-dimensional neural network (HDNN) for molecular energy (E) prediction. Inferences are obtained as the sum of local atomic contributions. Initial Cartesian coordinates (R_i) are converted into atom-centred symmetry functions (G_i) for each atom, which are treated with an individual atomic NN to construct an extensible framework. **b** | The hierarchically interacting particle neural network (HIP-NN) architecture. The initial network input is a one-hot encoded vector of atomic species {Z}. The interaction layers (green boxes) are responsible for mixing information between atoms while incorporating pairwise distances R_{ij} between them. The first interaction layer collects information about the radial distribution of atom types in the environment, and later interaction layers collect information about the radial distribution of the atomic features computed by the previous layers. Between interaction layers are on-site layers (blue boxes), which are standard nonlinear feedforward layers that act on each atom without mixing information between atoms. These operations explicitly encode radial information while implicitly encoding angular information. The hierarchical ansatz constructs the atomic property P_i as a sum of P_i^n contributions where P_i^0 is the best-fit energy for each atomic species without environmental dependence, P_i^1 accounts for the first level of interactions, and P_i^2 accounts for more complex interactions. P_i^n hierarchical contributions are output by linear layers (stacked red squares) from on-site layers. Similarly, properties $P_{i,j}$, related to only pairs of atoms I, J (such as bond orders) are modelled linearly in terms of features and nonlinearly in distance. **c** | HIP-NN variants for learning different atomic and molecular properties. The models and their associated learned properties from left to right: the bond order model learns the Wiberg bond indices (WBI); the extended Hückel model machine learning (EHM-ML) Hamiltonian ($\hat{H}$) learns the off-diagonal (K) and diagonal (α) Hamiltonian parameters, and molecular orbital energies (ε_{MO}); HIP-NN trained on atomic charges learns the partial atomic charges (q); affordable charge assignment learns the dipole moments ($\boldsymbol{\mu}$); and HIP-loc (HIP-NN with a localization layer enabled by learnable localization weights ω_i) learns the total molecular energy (E) and electronic excitation energies (ΔE). **d** | Variants of the atoms-in-molecules network (AIMNet) architecture. In AIMNet, the atomic descriptor is a learnable representation (embedding), which combines information about {Z} and {G}. The final AIM representation (blue block) accounts for non-local effects through a series of self-consistent updates (green update block), which introduces the mean-field effects of neighbouring atoms and reaches beyond the initial cut-off radius. The original AIMNet can be trained for a particular state of the molecule, that is, neutral or charged. The AIMNet neural spin equilibration (AIMNet-NSE) architecture is conditioned on total molecular spin charges $Q(\alpha)$ and $Q(\beta)$. Globally normalized spin-polarized atomic charges $q(\alpha)$ and $q(\beta)$ are iteratively updated during training, so that AIMNet-NSE is capable of predicting non-local spin charges in both neutral and charged states.

layer of the network. This hierarchical decomposition reflects the intuition that very local contributions dominate predictions, and longer-range contributions play a smaller, corrective part in determining the full prediction; in a hierarchically trained model, it is frequently observed that P_i^n rapidly decays as n increases. Although P_i remains atom-centred, it can capture information at distances beyond the cut-off radius by using repeated message passing between atomic neighbours.

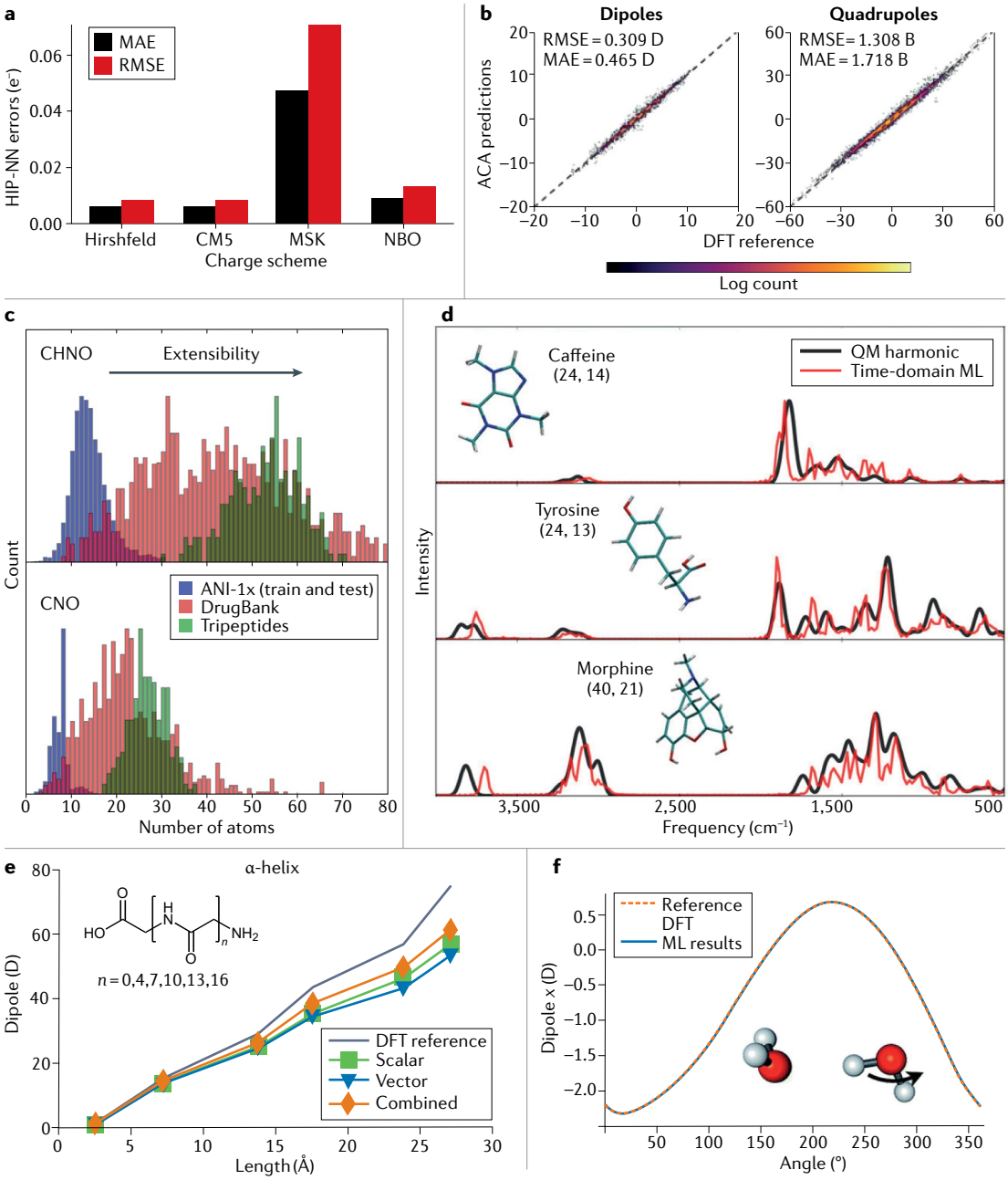
Message passing is achieved by using the learnable interaction layers to transmit information between atoms within a local neighbourhood. This approach can capture many-body effects in a perturbative fashion. Notably, the HIP-NN model can be trained using various quantities, such as total energy[50], partial atomic charges[51] and dipole moments[52] (FIG. 3c). The modularity of HIP-NN models makes it possible to combine both atomic and pairwise features to learn bond orders[53] and parametrize Hamiltonian models[54]. This is an example of multitask learning[55]. Additionally, a localized HIP-NN variant denoted HIP-loc[56], which is augmented by including localization layers, allows for the treatment of non-extensive properties[57]. Under the local decomposition ansatz all atoms have non-zero contributions to the total molecular energy; however, with the HIP-loc approach atoms are allowed to have zero contributions to the excitation energy. This development is achieved by introducing learnable weights associated with specific atoms[56]. The applications of HIP-NN are discussed further in the following sections.

Another example of a NN architecture for learning atomic and molecular properties is AIMNet[42] (FIG. 3d). The name arises from Bader's theory of atoms in molecules[58], which tracks the changes in atomic densities upon the formation of a molecule. AIMNet treats non-local phenomena by introducing SCF-like updates of atomic features (BOX 2). These updates bypass the constraints of locality (dictated by the cut-off radius) and ensure that interactions between both close and distant neighbours are accounted for. The first AIMNet implementation used multitask learning and was trained to six atomic and molecular properties simultaneously (including energy, atomic volumes and atomic charges)[42]. The AIMNet neural spin equilibration (AIMNet-NSE)[59] model extends the atomic features by embedding spin-polarized charges $q(\alpha)$ and $q(\beta)$ and correcting them iteratively (FIG. 3d). Hence, this model can predict partial α and β spin charges given only the total molecular charge and multiplicity. Conceptually, the NSE module serves as a neural charge and spin equilibration scheme by redistributing spin charges through the iterative procedure and making energy predictions based on the distribution of α and β spin densities.

Although a detailed overview of existing datasets[18] and methodologies[60,61] for data generation is beyond the scope of this Review, we briefly describe the datasets that are frequently referred to throughout this paper. ANI-1 is a library containing about 20 million non-equilibrium conformations derived by normal-mode sampling of about 57,000 small organic molecules (up to eight C, N and O atoms)[62]. All of the properties are calculated at the DFT level. ANI-1x, which is a portion[60,63] of the original ANI-1 set, comprises around 5 million entries selected by active learning to cover the most unique regions of chemical space. Comprehensive machine-learning potential (COMP6) is a more structurally diverse set[60] (compared with ANI-1 and ANI-1x) and contains organic molecules that comprise up to 312 atoms. COMP6 was designed as an extensibility and transferability benchmark and includes six subsets, three of which are of interest here: ANI-MD (systems comprising 20–312 atoms sampled by molecular dynamics and selectively recalculated using DFT), Tripeptide (248 tripeptides) and Drugbank (around 14,000 conformations of about 800 small drug molecules). All the datasets mentioned above include only closed-shell neutral molecules. The training dataset, Ions-12 (about 6.4 million structures that comprise up to 12 non-hydrogen atoms) includes open-shell cations and anions as well as closed-shell neutral systems, all of which are taken from the UNICHEM database[64] and sampled using molecular dynamics[59]. The test set Ions-16 contains 300,000 structures that are larger than those in the Ions-12 dataset (comprising up to 16 non-hydrogen atoms) to probe transferability[59]. More extensive discussions on datasets for ML-assisted chemistry are available in the literature[7,16,65–69].

Multitask learning
When one model is concurrently trained on multiple labels (for example, total molecular energies and atomic forces). Leverages useful information contained in multiple related tasks to improve the general performance of all tasks.

Non-extensive properties
Properties that do not scale directly with system size. Examples include characteristics of localized electronic states, electronic excitation energies, vibrational frequencies, electron affinities and ionization potentials.



In the next sections we illustrate how the aforementioned NN architectures can be applied to the prediction of various chemical properties, paying particular attention to benchmarking, extensibility, transferability and applications.

Charge distribution

Molecular charge distribution is a continuous spatial distribution of electronic density, which is rigorously defined by the respective QM wavefunction, and can be imaged using modern microscopic techniques[70,71]. This concept is intrinsic to the language of chemistry, enabling the use of chemical intuition and simple rules to make powerful predictions about chemical bonding and reactivity. The total molecular charge allows us to differentiate between neutral and charged species. Dipoles, quadrupoles and other multipoles reflect a non-uniform distribution of positive and negative charges throughout a molecule. A deeper understanding requires the elucidation of the charges belonging to specific regions or even atoms. Partitioning the total charge into individual atom-centred quantities is a convenient coarse-grained way to represent the distribution of electronic density. Typically, such compact charge assignments require complete QM calculations, followed by subsequent population analyses. Such routes quickly become numerically expensive for large systems or long molecular dynamics runs, calling for the use of ML techniques.

Partial atomic charges. Although the spatial charge density distribution is well defined by the electronic wavefunction, density partitioning of atomic charges is not unique. This uncertainty gives rise to numerous charge assignment schemes, such as the Hirshfeld[72],

◀ Fig. 4 | **Machine learning prediction of atomic charges, vibrational spectra, dipoles and quadrupoles.** The figure is organized in ascending order from local (partial charges) to global (dipoles) properties. **a** | The mean absolute error (MAE) and root mean squared error (RMSE) of hierarchically interacting particle neural network (HIP-NN) charge predictions when trained with the test set (ANI-1x) to reproduce the partial atomic charges of various charge assignment schemes: Hirshfeld, charge-model 5 (CM5), Merz–Singh–Kollman (MSK) and natural bond orbital (NBO). CM5 and Hirshfeld charges are the most consistent and suitable for ML training, whereas the MSK scheme performs poorly. **b** | Performance of HIP-NN affordable charge assignment (ACA) model on the Tripeptides set when trained only on ANI-1x dipoles, compared with the density functional theory (DFT) reference. The left and right panels show the prediction of dipole moments (in Debye) and quadrupole moments (in Buckingham), respectively. Low errors in modelling the predicted dipole moments for molecules beyond the ANI-1x set demonstrate the strong extensibility of the model, whereas the accurate prediction of quadrupole moments confirms model transferability. The colour scheme for each histogram is normalized by its maximum bin count. **c** | Distribution of molecules by size in the training and extensibility sets for the ACA model. The top panel counts the total number of atoms (C, H, N, O) per molecule, whereas the bottom panel counts the number of non-hydrogen atoms (C, N, O) per molecule. ACA is applicable to systems that are several times larger (such as tripeptides in the Tripeptides set) than those in the training set (ANI-1x). **d** | ACA simulation of the infrared spectra of selected bioactive molecules. The total and non-hydrogen number of atoms are given in parentheses. The morphine molecule is two times larger than any system in the training set, yet the predicted infrared spectrum is accurate compared with the results of the quantum mechanical (QM) harmonic simulations. **e** | Machine learning (ML) predictions of the dipole moment of polyglycine in the α-helix conformation with different chain lengths (shown in the inset), compared with a DFT reference. The scalar model is based on partial atomic charges, the vector model is based on local (atomic) dipoles, and the combined model uses both approaches concurrently. **f** | Dynamics of the dipole moment projected to the *x* axis upon rotation of an O–H bond in a water molecule within a dimer. Rotation leads to the breaking of an intramolecular hydrogen bond (inset). Part **a** is adapted with permission from REF.[51], ACS. Parts **b**,**c** and **d** are adapted with permission from REF.[52], ACS. Part **e** is adapted with permission from REF.[87], AIP Publishing. Part **f** is adapted with permission from REF.[88], RSC.

ChargeModel 5 (CM5)[73], Merz–Singh–Kollman (MSK)[74] and natural bond orbital (NBO) methods[75], which are suitable for various chemical predictions or models. For example, Hirshfeld charges are frequently used for designing classical force fields[76]. Notably, atomic partial charges are atom-centred scalar quantities, so they could be directly used as an input to NNs that have been designed for energy predictions based on local approximations. In fact, the original HIP-NN architecture adopted for charge learning[51] exhibits an outstanding performance in reproducing Hirshfeld, CM5 and NBO charge schemes (FIG. 4a). Errors for all the tested charge models are remarkably small, except for MSK charges, and do not exceed $0.01e^-$. Most probably, the inferior performance of MSK charges is related to the locality of the NN descriptor, whereas the MSK scheme was designed to reproduce molecular (non-local) dipole moments. Training on the diverse ANI-1x dataset[60,63] of DFT results (only neutral molecular species) accentuates the applicability of the model across diverse chemical space. Indeed, the HIP-NN model for charge prediction shows excellent accuracy when tested on the COMP6 dataset[60], emphasizing the extensibility and transferability of the model.

These results bring us one step closer to the assignment of charges in large biomolecules and proteins, which is of paramount importance for modelling their activity. HIP-NN accurately reproduces[51] partial charges on the mini-proteins Chignolin (Protein Data Bank, PDB) identifier: 1UAO)[77] and Trp-Cage (PDB ID:1L2Y)[78]. Benchmark charge assessment in glucoamylase (PDB ID: 1AYX)[79], which contains more than 4,000 atoms, takes only 2 minutes to compute with HIP-NN[51], whereas it is prohibitively expensive using DFT. Efficient scaling of the HIP-NN charge model makes accessible the biochemistry and proteomics that are not easily reachable with QM methods. Overall, this example highlights that charge distribution, which is naively represented as an array of scalar atomic-centred features, can be efficiently learned using NNs that embed the underlying physics defined by quantum mechanics.

The usefulness and ubiquity of atomic and molecular charges in chemical discovery and vocabulary makes them priority targets for ML models. Although we are not able to cover all the exciting advances[80–82] in this field here, we would like to direct interested readers to other ML predictors of atomic charges such as the Atom-Path-Descriptor model[83], ContraDRG[84] and PhysNet[45], among others.

Dipoles and quadrupoles. The above examples demonstrate that ML can achieve QM accuracy in charge prediction following a particular scheme, yet partial atomic charges are poorly defined[85,86]. There is no unique way of charge partitioning, simply because the electron density is shared across the whole molecule, and atomic density basins are very approximate, if not artificial. By contrast, dipole moments are system-level quantities and are therefore observable. Dipole moments are accessible experimentally and are usually predicted very well by QM methods. Previous efforts constructed statically parameterized charge-partitioning models for reproducing dipole moments, for example, MSK[74] and CM5 (REF.[73]) schemes. The logical continuation of these endeavours is to replace the static parameters with learnable features: rather than training NNs to local atomic charges, one can train ML models to the dipoles to predict the underlying atomic charges. Such a ML model optimizes the local charge distribution by reconstructing a global and observable quantity — the molecular dipole. In HIP-NN, this model was termed affordable charge assignment (ACA)[52].

Once the NN is trained to the ANI-1x dataset[60] of DFT dipoles, ACA charges become uniquely defined and can efficiently predict the dipole moments of unknown molecules beyond those in ANI-1x. For example, ACA achieves remarkable accuracy in the prediction of the dipole moments of tripeptides in the Tripeptide subset from COMP6 (REF.[60]) (FIG. 4b, left), confirming the extensibility and transferability of the model, as emphasized by the size distribution histogram (FIG. 4c). Moreover, the ACA model can handle higher electric moments and can predict quadrupoles (FIG. 4b, right), with only a moderate decrease in accuracy compared with dipole predictions. We emphasize that with no human intervention, ACA not only automatically restores the dipoles from learned atomic contributions but also reproduces new properties, of which it was unaware during the training. These observations confirm that the ACA scheme implicitly learns the underlying physics of charge distribution. Interestingly, ACA closely mimics the results of the CM5 charge assignment scheme (also conditioned on the dipoles) but is orders of magnitude faster, enabling the immediate assignment of multipole moments. Low-cost assessment of dipoles using ACA enables simulations of infrared

spectra based on short molecular dynamics trajectories (100 ps with a 0.1-fs timestep) using the ANI-1x ML potential[60] followed by HIP-NN ACA modelling, without the need to use a QM calculator. Given the absence of the anharmonicity correction, ACA-derived infrared spectra for selected bioactive molecules (FIG. 4d) are in excellent agreement with the QM results used as a reference.

ML also enables the assessment of dipole moments in much larger molecules, including some polymers. Comparing the ML predictions of the dipole moment of polyglycine in the α-helix conformation obtained by modelling the scalar (local charges) and vector (local dipoles) properties and their combined variants, reveals good agreement with the reference DFT dipole (FIG. 4e). The large size of the α-helix compared with that of the bioactive molecules modelled by the ACA approach demonstrates that the accurate ML prediction of charges and dipoles is feasible for nanoscale objects (such as polymers), which are very expensive to model using conventional QM approaches. Interestingly, simultaneous learning of scalar and vector properties can yield higher accuracy than a single scheme[87].

Intermolecular interactions. The results outlined above were mostly obtained for isolated molecules; however, the accurate description of intermolecular interactions still poses a serious challenge. This difficulty is imposed by the locality of the atomic environment (typically featuring a 5–7-Å cut-off radius), which is often not large enough to capture all the intermolecular forces. However, in some cases, approaches based on the Behler–Parrinello approximation still yield reasonable results. One such example is the Tensormol[88] architecture, which employs separate Behler–Parrinello-type NNs conditioned to the total energy and molecular dipoles to capture long-range physics. Tensormol precisely evaluates the dipole moment in water dimers upon the rotation of one water molecule about the O–H bond, which leads to the breaking of a hydrogen bond (FIG. 4f). This study emphasizes the use of ML to investigate intermolecular properties, which are essential for modelling dynamics and reactivity. Another study[89] employing the ANI potential[90] (Behler–Parrinello-like descriptor) assessed π-stacking interactions between various heteroaromatic rings while including explicit solvation effects. The trained model was then used to perform a conformational search to rank favourable and unfavourable π-stacking interactions relevant to drug discovery[91]. A more sophisticated approach has been proposed that aims to extend Behler–Parrinello symmetry functions to specifically model intermolecular interactions. In brief, the original Behler–Parrinello ACSFs sum overall the atoms, but a more rigorous way to treat intermolecular interactions is to relate this summation to the interacting fragments and to introduce atom-pair symmetry functions[92]. Such simple and clever modifications enable the targeting of energetic components of symmetry-adapted perturbation theory[93], which is a powerful tool for elucidating intermolecular interactions. This approach[92] shows remarkable accuracy for hydrogen-bonded complexes (mean absolute error of 1.2 kcal mol^{-1}) for an extensive conformational space that spans complexes with binding energies of up to 20 kcal mol^{-1}. This approach was subsequently encoded into the atomic-pairwise neural network (AP-Net) architecture[94], resulting in a reduction in the binding energy error by a factor of five compared with that of traditional Behler–Parrinello descriptors.

Non-local effects and electron delocalization

In this section, we focus on important electron localization and delocalization metrics, such as ML analogues of spin density and chemical bonding. The above examples dealt mostly with local charge distributions. For NN architectures based on the local environment, features outside the cut-off radius are usually invisible to the central atom and do not participate in training. In other words, long-range effects are not captured. However, these effects are ubiquitous, even in rather simple organic compounds. For example, charged species might feature strong non-local effects that are emphasized by electron delocalization. Likewise, even in neutral systems, changing a spin state can affect the entire molecule or its fragment. Chemical bonding patterns can be limited to an atomic pair or spatially distributed, as is the case for aromatic compounds. All of these electron delocalization effects deserve scrutiny, as they are responsible for determining chemical reactivity, the propensity to excitations and ultimately define the functionality of a material across broad application areas.

Non-local self-consistent charges. SCF-like updates in AIMNet[42] overcome the issues associated with non-locality and account for long-range interactions that occur beyond the local atomic environment. An illustrative example of this approach is the calculation of the partial charge on the sulfur atom for a series of thioaldehydes with various electron-withdrawing and electron-donating substituents located at a distance beyond the cut-off radius (FIG. 5a). With a single iterative pass ($t = 1$), AIMNet incorrectly predicts that the sulfur atom has the same charge in all molecules in the series; however, consequent iterations propagate the long-range interactions and converge to the self-consistent state of the atoms-in-molecules representation ($t = 3$).

Spin charges. So far, we have discussed only closed-shell neutral species (that is, species with zero net molecular charge). Given the ubiquity and importance of charged systems (cations and anions), further advances in ML are impossible without considering charged species. For example, consider the electron transfer in the different charge states of 4-amino-4′-nitrobiphenyl[59] (FIG. 5b). In the anionic form, simple chemical logic suggests that the electron-accepting nitro-group would withdraw electron density from the π system, which is reflected by the NBO spin density and charges. Conversely, in the cationic form, the electron-donating amino group adds charge to the conjugated π-system to compensate for the overall positive charge. AIMNet-NSE, trained on the Ions-12 dataset, can simulate charged open-shell systems and redistribute spin densities iteratively to capture the long-range effects[59]. Prior to the NSE equilibration, the spin densities are equivalent for both the anion

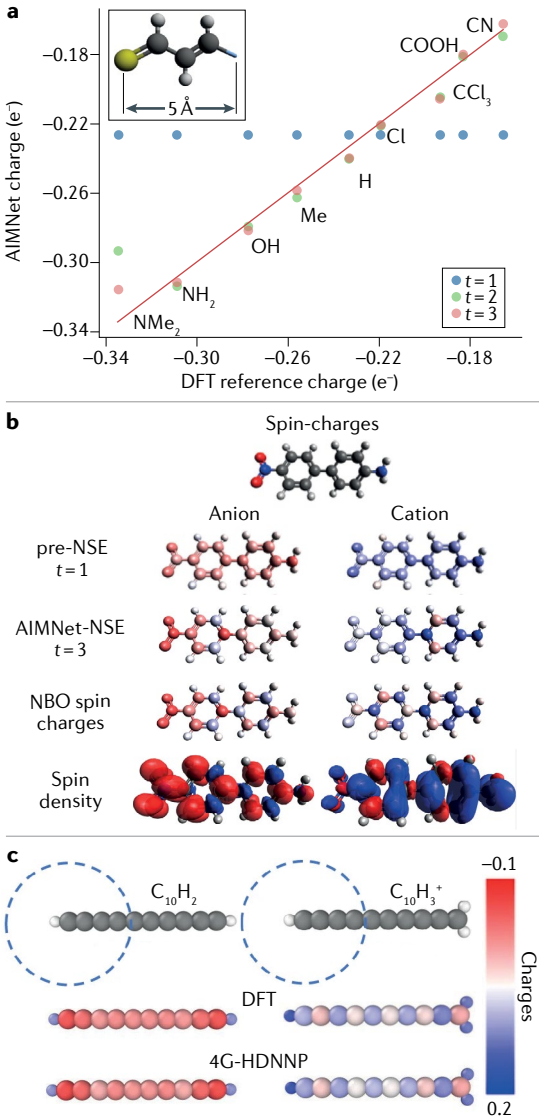
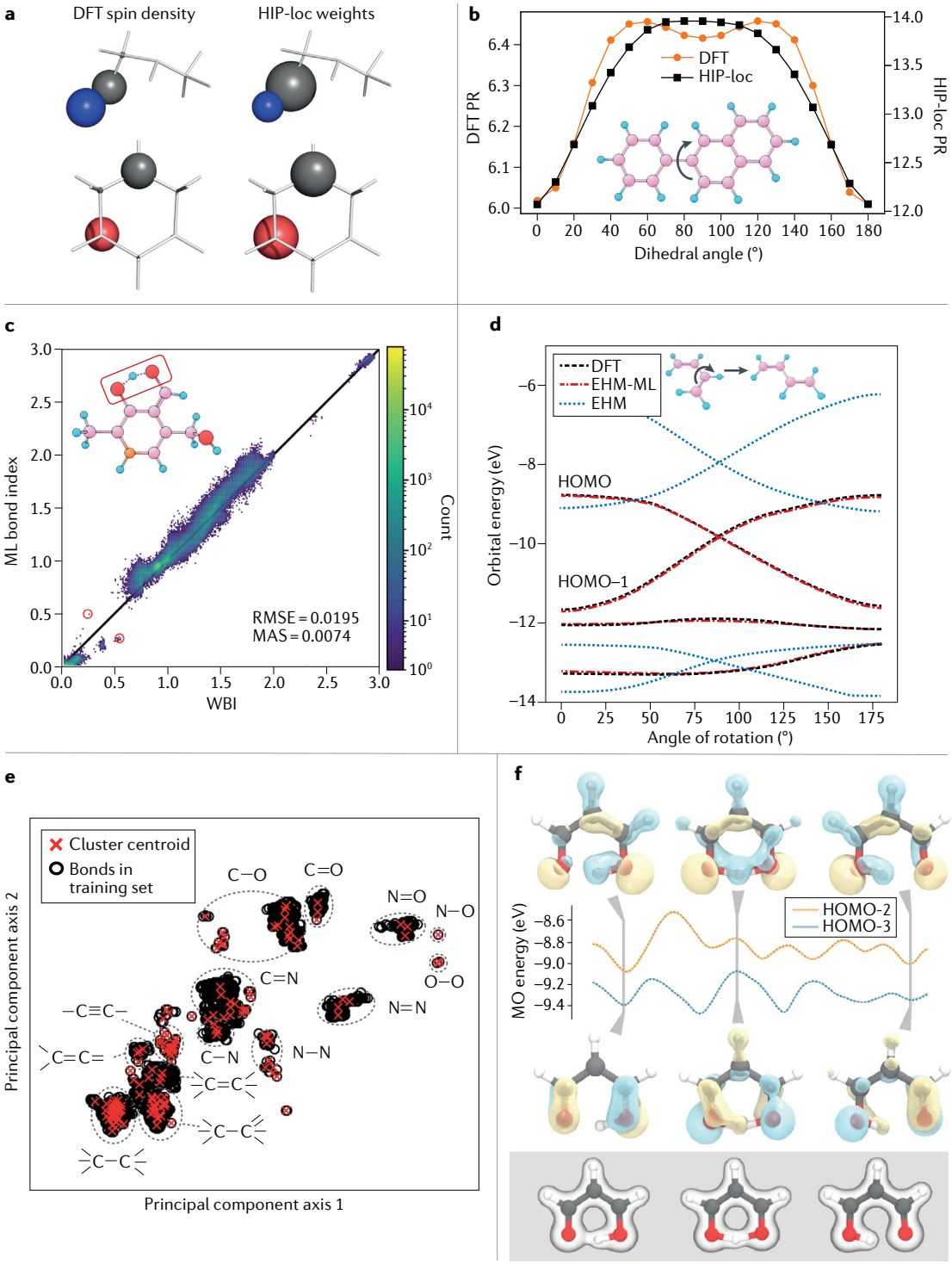


Fig. 5 | **Machine learning prediction of spin-polarized charges and total electron density. a** | Atomic charges on a sulfur atom in a series of substituted thioaldehydes, as predicted by atoms-in-molecules network (AIMNet) trained on the ANI-1x dataset augmented by molecules with fluorine, sulfur and chlorine atoms. The cut-off radius is set to 4.6 Å, although the distance between the sulfur atom and the group R (such as NMe_2 or OH) is about 5 Å (as shown in the inset). The predicted AIMNet charges evolve through iterative updates (*t*) as the neural network (NN) gains more information beyond the cut-off radius. At $t = 1$, the performance of the model is usually that of a locally aware NN (namely, the Behler–Parrinello and accurate NN engine for molecular energies (ANI) models). At $t = 3$, the long-range effects of the terminal substituents are recognized by AIMNet and, thus, the model reproduces the density functional theory (DFT) charges. **b** | Iterative updates in AIMNet neural spin equilibration (AIMNet-NSE) make it possible to learn the α and β electron densities concurrently. At pre-equilibration step $t = 1$, spin densities are the same for both the cation and anion although different in sign; at $t = 3$, the spin charges are equilibrated and are almost indistinguishable from the reference natural bond orbital (NBO) charges. The anion is the spin electron atomic charge or density ($\alpha - \beta$). The cation is the spin hole density ($\beta - \alpha$). Red and blue indicate negative and positive spin charges respectively. **c** | Performance of a fourth-generation (4G) high-dimensional neural network (HDNN) potential in simulating non-local charge transfer. The top, middle and bottom rows show the molecular structures, DFT charges and 4G-HDNN potential charges, respectively. The 4G-HDNN potential recognizes long-range electron transfer. Part **a** is adapted with permission of AAAS from REF.[42]. © The Authors, some rights reserved; exclusive licensee AAAS. Distributed under a CC BY-NC 4.0 License. Part **b** is adapted with permission from REF.[59], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Part **c** is adapted with permission from REF.[39], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

and the cation (although different in sign; FIG. 5b). Subsequent NSE updates promote the effects of distant substituents, ultimately converging to the results of DFT simulations[59]. The application to 4-amino-4′-nitrobiphenyl, which is a system outside the training and validation sets, demonstrates that the AIMNet-NSE model is transferable across open-shell systems with pronounced long-range effects. The accuracy of these results is maintained for the Ions-16 dataset, which is a diverse test set of charged species. Moreover, joint training to neutral and charged states also enables the assessment of different reactivity indices, such as the Fukui functions or electronegativity following the conceptual DFT formalism[95]. These developments bring ML computations one step closer to simulating truly reactive chemistry, in which open-shell charged intermediates have a vital role.

Partial atomic charges and long-range charge transfer could also be identified on the basis of environment-dependent atomic electronegativities, as implemented in the fourth-generation Behler–Parrinello HDNN potentials[17,39]. This model relies on charge equilibration, which takes into account different charge states. The results of charge assignment by this network closely follow those of the reference DFT Hirshfeld charges for both neutral and cationic systems (FIG. 5c). With these remarkable results, this pioneering fourth-generation Behler–Parrinello architecture is one of the most advanced ML frameworks and is able to capture non-local phenomena. Clearly, the ML models for charge prediction are currently undergoing exciting advances. Starting with the simple task of learning scalar charges, ML models have become sensitive enough to capture long-range charge redistribution upon the addition or removal of electrons[96]. Traditionally, we expect such accuracy only from advanced electronic structure calculations, yet properly crafted surrogate models can also capture these complicated transitions.

Electron density. Aside from atomic charges, the electron density is another central quantity in the chemical paradigm. It is possible to represent the total electron density as the corrected sum of proatomic densities (that is, the densities of isolated atoms), resulting in an extensible model[97]. This development was achieved by employing the symmetry-adapted Gaussian process regression (SA-GPR) model[97]. Although this Review focuses on chemical NNs, we believe that investigating a GPR model[27,98,99] is beneficial here, as it shows that other



models can provide viable alternatives to NN potentials. Furthermore, the molecular representation is far more important than the specific algorithm (GPR or NN) used for parameter optimization[100]. The resulting SA-GPR model trained on the electron densities of small molecules accurately reproduces basins of molecular electron densities in molecules that are approximately two times larger than those used in training. Interestingly, SA-GPR also shows remarkable accuracy for non-covalent interactions[101]. We point interested readers to a review[102] that summarizes developments in GPR frameworks and presents the learnable chemical properties. Similarly to SA-GPR, electron density could be a learnable target for NNs. Another suggested methodology, called the anisotropic analytical model of density (A2MD)[103], uses Behler–Parrinello symmetry functions to construct an extensible and transferable density model. Notably, the A2MDnet architecture enables linear scaling, making it feasible to model the reactivity of large biomolecules and even proteins. Finally, among other work in the area, we would like to mention message-passing architecture, which has been used for electron density learning[104], and the work on

◀ Fig. 6 | **Machine learning prediction of spin-density, bond orders and effective Hamiltonian models. a** | Comparison of density functional theory (DFT) spin density and visualized hierarchically interacting particle-localized (HIP-loc) localization weights for selected molecules. HIP-loc effectively learns local domains that contribute to the electronic transitions. The radii of the spheres are proportional to the magnitude of the weights. Colours are atomic-specific, where grey is carbon, blue is nitrogen and red is oxygen. **b** | Changes in the participation ration (PR) when scanning over the C–C dihedral angle in a polyaromatic molecule (inset). PR quantifies the spatial electron spin distribution as inferred from DFT atom-centred spin density and HIP-loc weights. **c** | Performance of the HIP-NN model for bond orders trained on a portion of the ANI-1x set and applied to the Drugbank set; HIP-NN (ML) bond orders against the Wiberg bond indices (WBIs) are shown. The low root mean squared error (RMSE) and mean absolute error (MAE) suggest that the model is transferable. Outliers are indicated by the red circles. The inset shows an example of outliers, in particular, a hydrogen bond for which the accuracy of machine learning (ML) predictions is low. **d** | The frontier molecular orbitals (MOs) swap during *cis–trans* isomerization of 1,3-butadiene, as shown in the inset. The ML-parameterized tight-binding Hamiltonian (EHM-ML, where EHM is the extended Hückel model) provides eigenvalues that quantitatively reproduce DFT results, in particular, recognizing the crossing of the HOMO and HOMO-1. **e** | Clusters with different bond topologies in the training set used to fit bond-specific repulsive potentials in the density functional tight-binding (DFTB) framework. The picture is based on a principal component analysis as a part of an unsupervised ML application. The black circles are bonds in the training set, which are mapped to the nearest cluster centroid, shown as a red cross, to divide the bonds into groups for the fitting of repulsive potentials. The grey dashed circles highlight the coarser categorization of bonds based only on a local topology. **f** | Proton transfer in malondialdehyde as a series of snapshots from a molecular dynamics simulation. The top panel shows the energies and contours of the selected frontier molecular orbitals (HOMO-2 and HOMO-3), which reflect changes that occur during the proton transfer from the right oxygen atom to the left one. The central configuration is conceptually similar to the transition state. Total electron densities for each configuration are given at the bottom. Parts **a** and **b** are adapted with permission from REF.[56], RSC. Part **c** is adapted with permission from REF.[53], AIP. Part **d** is adapted with permission from REF.[54], AIP. Part **e** is adapted with permission from REF.[127], ACS. Part **f** is adapted with permission from REF.[118], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

learning electron densities in periodic systems to extend ML models beyond the molecular domain[105].

Spin density. Until now we have focused only on learnable ground state properties; however, many relevant processes also involve electronically excited states, which require a more sophisticated description and pose a challenge for ML models. The spin excitation that occurs when a molecule changes its spin multiplicity is an example of an electronically excited state (ML advances for excited states are reviewed elsewhere[19]). From a practical perspective, transitions between spin states are used across many applications, ranging from light emitters[106] to photovoltaic[107] and spintronic devices[108]. QM methods that inherently account for spin can accurately predict spin localization. However, the incorporation of spin in a ML model is difficult owing to non-local effects (BOX 2).

Changes in spin density due to the electronic excitations (transition density) can be evaluated using HIP-loc[56]. In the localization layers, a local contribution is combined with atomic weights which are globally normalized. Prior to normalization, the weights are predicted in the same way as any local quantity, such as charge. Post-normalization, these atomic weights sum to one, and are used to re-weight the local contributions to create a prediction that only includes contributions from a subset of the atoms in a system. We emphasize that these weights are not merely abstract auxiliary quantities: they highlight the atoms that are responsible for an electronic transition in the same way as spin-polarized density in QM methods[109] (FIG. 6a). This analogy is especially prominent given that HIP-loc is completely unaware of the spin density or the wavefunction. In fact, molecular energies and forces of the singlet and triplet states are the only inputs used to train the HIP-loc model, which nevertheless enables the recovery of an approximate spatial distribution of the wavefunction (FIG. 6a). This is another example wherein ML captures the underlying quantum mechanics, which was not incorporated during training.

Regions of concentrated spin density evolve following conformational changes in a molecule. This effect can be illustrated by examining the rotational profile of a molecule that comprises several aromatic rings using the participation ratio (FIG. 6b). The participation ratio ranges from 1 (fully localized on one atom) to the number of atoms N, which denotes delocalization over the entire system. Rotation around the dihedral angle (inset in FIG. 6b) affects the participation ratio by modifying the electronic delocalization across conjugated rings. At 90°, the excitation becomes more delocalized, and the qualitative agreement between the participation ratios computed with ML and DFT indicates that HIP-loc recognizes this trend[56]. This example emphasizes the ability of the HIP-loc model to capture the dynamic changes in the spin density distribution following changes in the molecular conformation.

Chemical bonding. Another intuitive and guiding concept of chemistry is chemical bonding, which explains how and why atoms are held together. For instance, the reactivity of unsaturated aliphatic and aromatic species (such as the propensity to electrophilic addition versus substitution) could easily be inferred from their distinct bonding. From a QM point of view, the bond order of an atomic pair is related to the electron density concentrated between the atoms; thus, an assessment of the bonding requires a full solution of the Schrödinger equation, which slows high-throughput bonding analysis of large amounts of data. To unify chemical bonding and ML, a rigorously defined bonding metric is needed that can be calculated in a batch and does not require the manual selection of bonding elements. Bonding information is usually organized into an $N \times N$ matrix for a system comprising N atoms; this matrix is a derivative of the electronic density matrix. The off-diagonal elements are the bond indices between different atoms in a pairwise manner. Similar to the case of atomic charges, approaches to describing chemical bonding are not unique and several schemes have been proposed[110,111]. For example, the Wiberg bond index (WBI) is well defined for each pair of atoms[112]. Such a matrix representation is very attractive as a label for NNs as the architecture might be adopted to learn not only atom-centred properties but also pairwise features, such as those implemented with a HIP-NN[53] (FIG. 3c).

To demonstrate this approach, the HIP-NN was trained on off-diagonal WBI matrix elements on a portion of the ANI-1x set[60,63]. When tested on the DrugBank set, there was good correlation between the ML bond index and the WBI with very little spread (FIG. 6c), where the outliers are mostly associated with the presence of non-covalent interactions or hydrogen bonds, which

Wiberg bond index (WBI). A quantitative bonding model that expresses the bond order between pairs of atoms. WBI measures the electron population overlap between atoms and frequently numerically aligns with chemical intuition, for example, the WBI of the C=C bond in ethane C_2H_4 is expected to be close to two.

are challenging to describe even with the reference QM methods. The accuracy of the predictions for the extensibility set (ANI-MD) is comparable with that of the DrugBank predictions. In general, the HIP-NN trained on WBI matrices produces transferable and extensible models that can be applied to a wide variety of small and large organic molecules and reproduce bond orders with QM precision. It is worth noting that the disagreement between ML bond orders and WBIs is even smaller than that between different QM bonding schemes[12]. Thus, ML predictors for chemical bonding provide reliable and fast alternatives to the traditional QM schemes.

Remarkably, the NN-based strategy to classify compounds according to aromaticity criteria was suggested more than a decade ago[113,114]. Even so, reports on bridging ML and chemical bonding are extremely scarce and account for few searchable articles[53,113–116]. Given the usefulness of chemical bonding paradigms, this might be surprising. However, one should keep in mind that the localization of bonds to regions of high electron density is artificial and lacks a rigorous physical basis. As such, many preferable bonding schemes require user attention and intuition, delaying ML featurization. Moreover, the absence of comprehensive datasets focused on bonding information complicates the issue. Therefore, we believe that chemical bonding beyond the WBI metric still awaits proper featurization by ML models. Perhaps this development should be preceded by the elaboration of reliable black-box bonding schemes.

Parameterization of Hamiltonians

All of the ML methods discussed so far are surrogate models that predict electronic features (such as charges, dipoles, bond orders and energy) for a given molecular geometry. Such ML models explicitly neglect electrons and the underlying quantum mechanics. Although these approaches are successful, the total abandonment of quantum mechanics may not be advisable in the case of strongly non-local interactions and electronic delocalizations. For example, the description of abrupt changes in the wavefunction following electronic excitations, spatially delocalized molecular orbitals and the addition or removal of charges are still very challenging for ML. One popular approach is to retain some reduced QM description augmented by ML[54,117–119]. For example, efforts were made to use ML to parametrize effective Hamiltonian models to fully or partially treat electron behaviour, replacing the majority of the terms and integrals with only a few parameters.

Semi-empirical models. A family of so-called tight-binding Hamiltonian models provide an intuitive minimalistic description of many materials and chemical systems. These models retain the form of a $N \times N$ matrix, where N refers to the number of atoms or atomic orbitals. As in the case of chemical bonding, the elements are defined by pairwise features, which partially incorporate complex Coulombic interactions, such as electron–electron repulsion. For example, the extended Hückel model (EHM) offers simplified, yet insightful QM treatment of different chemical systems. Here, the diagonal elements are related to the ionization energy of the valence states of the isolated atoms, whereas off-diagonal elements describe pairwise interactions between the atoms. Although the fixed parameterization of the EHM[120,121] is highly intuitive, it does not provide quantitative accuracy. A ML enhancement for EHM was proposed that relies on the dynamical parametrization of the matrix[54], where the Hamiltonian matrix elements become environment-dependent parameters that are a smooth function of the molecular geometry. This constitutes a much more flexible representation than that achieved with the HIP-NN parametrized on the ANI-1x set.

The resulting dynamically parameterized EHM-ML model was trained on DFT frontier molecular orbital energies, allowing for a fast assessment of these quantities[54]. The accuracy of these predictions matches that of more expensive reference DFT-level calculations with deviations of less than 0.2 eV per orbital between the two approaches. Moreover, the environment-dependent nature of the generated EHM-ML parameters greatly improves transferability and makes it possible to track the evolution of molecular orbitals under dynamic conditions (FIG. 6d). For example, for the internal rotation of 1,3-butadiene around a single C–C bond, DFT predicts the crossing of the highest occupied molecular orbital (HOMO) and HOMO–1 upon *cis–trans* isomerization, and EHM-ML precisely captures this phenomenon.

Beyond the simple EHM-ML model, ML was also used[122] to substantially increase the accuracy of more advanced semi-empirical models such as parametric method 3 (PM3)[123]. The ML-PM3 Hamiltonian is trained on DFT reference data and incorporates the environmental awareness through the HIP-NN[50] framework. By allowing the ML model to dynamically adjust PM3 parameters according to the local atomic environment, the model energy prediction errors are reduced by up to 70% compared with the results obtained using the original PM3 parameterization or the regular HIP-NN interatomic potential. Moreover, since the use of ML here only corrects the Hamiltonian parameters and does not aim to learn the full surrogate structure-energy relationship, ML-PM3 can be efficiently trained on smaller datasets than interatomic potentials. For example, the highly transferable ML-PM3 model was trained only on about 60,000 organic molecules[122] while accurate interatomic potentials are often trained to millions of data points[60,62]. Additionally, ML-PM3 excels in structure optimization, which is a common goal of computational chemistry. Among the approximately 650 structures tested, ML-PM3 converged all the systems to the true minima providing accurate estimations of the vibrational frequencies, compared with the pure HIP-NN model that failed to converge to optimal solutions for around 400 of the approximately 650 molecules tested[122]. Although surrogate ML models can be accurate for static properties, environment-aware parametrization of Hamiltonians enables the assessment of more complex dynamic computations, including structural optimization and reactivity. These results exemplify the potential of the dynamic parametrization of physical models by ML algorithms. ML can increase the accuracy of a model such that it becomes comparable with that of more advanced levels of theory without sacrificing the simplicity of the model.

Δ-ML
A composite approach in which baseline values, obtained from cheap QM methods (usually DFT), are corrected towards a target line, calculated by a more sophisticated level of theory, for example, the coupled-cluster methods of perturbation theory.

Density functional tight-binding. Another popular simplified Hamiltonian model arises from DFTB[124,125] theory. Briefly, instead of solving the Kohn–Sham equations self-consistently for all electrons in the system, one can obtain QM solutions based on the linear combination of atomic orbitals for isolated atoms in an electrostatic potential that is applied to minimize the diffuseness of atomic-centred density in the tight-binding framework. Therefore, the electronic Hamiltonian includes only one-centre and two-centre parameters (as in EHM). Diagonal elements represent the energies of one-electron atomic orbitals, and the off-diagonal elements are responsible for pairwise interactions. Additionally, a repulsive potential is imposed to treat interatomic interactions in a force-field manner. This DFTB repulsive potential is subject to fitting, which is traditionally done by optimizing flexible expressions such as splines or polynomials[126], to match more sophisticated DFT results. This optimization of repulsive potentials can greatly profit from the application of ML techniques.

Traditionally, a repulsive potential depends on the atom type. The incorporation of neighbourhood awareness into a repulsive potential starts with accounting for bond types rather than atom types[127]. This approach closely resembles the idea of a local chemical environment, as introduced above, and drastically reduces error in energy estimations, from approximately 7.6 kcal mol^{-1} to 2.6 kcal mol^{-1}. Of course, replacing the atom-specific repulsive potential with a bond-specific one requires rigorous classification of bond types. This classification can benefit from an unsupervised[128] ML approach to enable fast and robust clustering. Clustering is the ML task of associating data items (in this case, bonds) with discrete categories that are not pre-specified for training. For example, clustering was used to divide bonds into groups (FIG. 6e) that were not specified by humans but by the data itself. The large circled clusters of coarse bond types have been manually annotated post hoc. It is apparent that clustering limits the number of bond-aware repulsive potentials that are subject to fitting.

However, the bond type clustering approach[127] described above does not modify the DFTB Hamiltonian itself. Alternatively, static Hamiltonian parameters can be replaced with dynamically generated ones when a NN is attached[117]. Such an approach noticeably reduces the prediction error of target properties (such as energies per atom, atomic charges and dipoles) by almost an order of magnitude compared with those of conventional parameterizations[124]. The accuracy of the predictions of DFTB becomes similar to those of the parent DFT methods, which is remarkable given the approximately thousand times increase in speed. These achievements suggest that ML can restore the accuracy of 'top-down' approximations, such as EHM and DFTB, making them comparable with a superior reference theory.

Quantum mechanical methods. In this section, we touch on a remarkable effort in the community devoted to ML-driven advances in QM methods themselves, such as developments in DFT approximations that will potentially bring the method closer to the exact answer. Obtaining a DFT solution for a system starts with the computation of the kinetic and potential energies, and the Coulomb energy of non-interacting electrons. The challenge is to approximate the other electron interactions by an accurate exchange–correlation functional. Once the latter is known, solutions of the Kohn–Sham equations provide the desired energy as the functional of the density. Recent reports prove that this stage can be bypassed, and that the mapping of the potential and density (or energy and density) can be achieved directly with ML, yielding highly accurate results[129–132]. The development of ab initio methods can also benefit from data-driven approaches. For instance, the Δ-ML[133] approach was used to increase the accuracy of Møller–Plesset perturbation theory (MP2) energies by training the model to the difference between the correlation energies at coupled-cluster and MP2 levels[134]. Similarly, DFT dipole polarizabilities were corrected to mimic the coupled-cluster results[135]. A review of the past and present challenges for using ML for electronic structure calculations can be found elsewhere in a collaborative roadmap[136].

Wavefunction. The Hamiltonian and the wavefunction can both be parameterized[118,137]. Parameterized wave functions retain access to a diverse set of molecular properties, including not only orbital energies but also the other features that can be derived directly from a wavefunction. This non-trivial idea was implemented in the SchNet for Orbitals (SchNOrb) framework, which was used to study the dynamics of the frontier molecular orbitals in malonaldehyde during internal proton transfer[118] (FIG. 6f). The dynamic picture obtained illustrates the expected redistribution of electron densities associated with the chemical reaction. With access to the learnable wavefunction, SchNOrb reproduces the reactivity in terms of quantum chemical phenomena. Moreover, SchNOrb enables inverse ML-assisted design of materials when a desired property (such as a specific gap between the HOMO and the lowest unoccupied molecular orbital (LUMO)) dictates the structure, by varying parameters such as bond distances. Notably, when injected into classical QM codes, SchNOrb wavefunctions provide excellent guesses for iterative procedures that greatly accelerate convergence.

Conclusions and outlook

The rapid development of accurate ML-based predictors in chemistry marks a new era in which the classical physics-motivated treatment of chemical systems shares the space with ML models. Historically, interatomic potentials based on NNs were among the first ML models used in chemistry. However, ground-state energy, which is the primary target of the interatomic potential NNs, gives only a small fraction of the desired information about the chemical system. Intuitive and guiding properties such as atomic charges, dipole moments, electronic densities, excitation energies and bond orders form the core of the chemical mindset and deserve no less attention than the ground-state energy.

The application of ML has been extended to predict various chemical properties beyond energy. The numerical cost of the discussed ML models scales linearly with the system size and these models are extensible

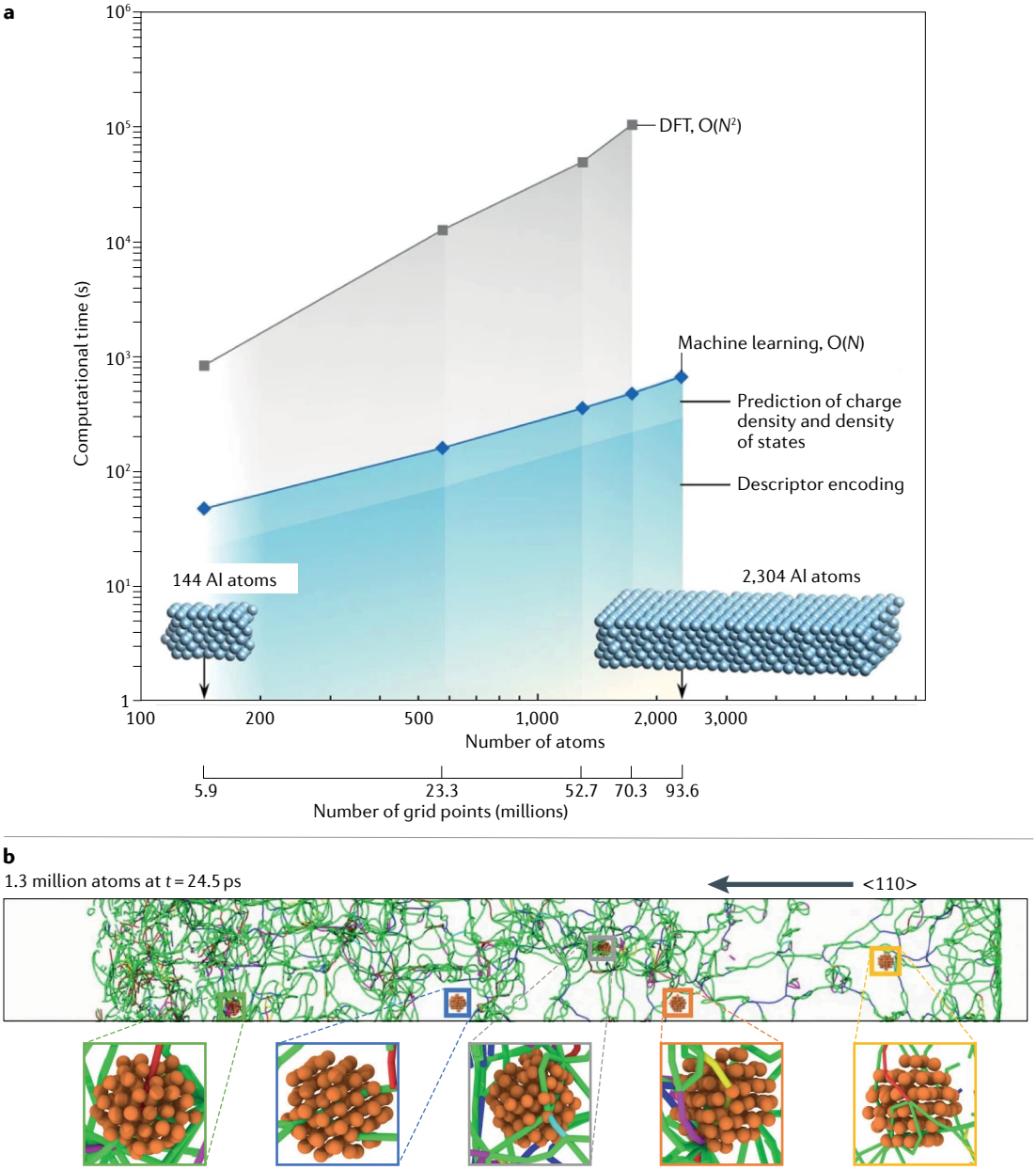


Fig. 7 | **Large-scale molecular simulations enabled by machine learning. a** | Scaling of density functional theory (DFT) compared with machine learning (ML) for bulk materials of increasing size (where *N* is the number of atoms). DFT scaling is almost quadratic, $O(N^2)$, whereas the ML model exhibits linear scaling, $O(N)$, and is an order of magnitude faster. **b** | Dislocated structure of bulk aluminium at 24.5 ps after application of a shock of 1.5 km s^{-1}, simulated using the ANI for elemental aluminium (ANI-Al) potential. The shock was initiated from the right along the ⟨110⟩ crystallographic direction. The boxes show randomly selected magnified regions. Part **a** is reprinted with permission from REF.[139], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Part **b** is adapted with permission from REF.[140], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

owing to the partitioning of the molecular structure into local chemical environments, which can be summed to reproduce a global property. Both system-level quantities (observables), such as electric moments or electron densities, and theoretical properties, such as atomic charges or bond orders, can be predicted with desirable 'chemical accuracy'.

Notably, some learnable features of ML models closely resemble quantities from the underlying quantum mechanics. An example of this is the excitation weights in the HIP-loc architecture, which have a spatial localization that correlates with the DFT-derived spin density that contributes to the electronic excitation[56]. Another example is the accurate prediction of quadrupoles when atomic charges were conditioned using only dipoles[52]. This example demonstrates that a properly constructed ML model can describe physical phenomena and achieve transferability. Alternatively, ML can dynamically parametrize QM models (Hamiltonians or even wavefunctions) as shown by the examples of

EHM-ML[54], DFTB[117] and the SchNOrb[118] framework. As a rule of thumb, on-the-fly parametrization greatly increases the model accuracy in a diverse chemical space. It is apparent that ML might benefit from encoding as much physics as possible into the architecture[42,138].

The full power of ML models in chemistry can be demonstrated by studying the scaling of ML models for bulk materials[139] (FIG. 7a). DFT-level simulations of systems with thousands of atoms are relatively common these days. However, such calculations still require considerable computational power. If an accurate pre-trained ML model is available, the same results can be obtained within minutes on a regular laptop. Each year, the community pushes boundaries by simulating larger chemical systems and materials. A study in 2021 reported the simulation of shock propagation in a crystalline aluminium sample containing 1.3 million atoms[140] (FIG. 7b). Additionally, ML-accelerated simulations of a system comprising more than 100 million atoms have been conducted[141]. Without ML, computational chemists would not be able even to imagine simulations on such scales accompanied by near quantum-chemical accuracy. In the realm of classical force-field dynamics, the record simulation size is currently 1.6 billion atoms[142]. Nevertheless, we foresee that in the near future, ML-augmented modelling will achieve or even break this limit, hopefully bringing QM-level accuracy to cellular-scale investigations.

Despite great progress, many challenges and problems are yet to be addressed. Data-driven approaches greatly benefit from good-quality data, especially when it can be gathered or generated in a reasonable time, perhaps days or weeks. Active learning[32,60,143,144] techniques present promising strategies for the optimal collection of datasets for interatomic potentials without losing their transferability. These methods and smart ML-assisted sampling[145–148] for properties beyond energy and forces are yet to be implemented. Currently, many architectures either train NNs on one target property or invoke separate networks, each responsible for its own label. We foresee that multitask learning will become a method of choice, giving concurrent access to an array of essential chemical properties[42,138,149]. Moreover, we expect that simplified physical models, such as effective Hamiltonians[54,117,127,150], will also receive increasing attention as their accuracy could be greatly increased by ML-parametrization. We also envision notable improvements for ab initio methods. For example, ML offers an automated protocol for selecting the active orbital space for configurational approaches near and at the bond-breaking limit, as has been shown for the dissociation curves of diatomics[151].

To conclude, with a computational cost only ten times greater than that of molecular mechanics[152] and an accuracy between that of DFT and coupled-cluster approaches, ML models grant scientists access to timescales and system sizes that have not previously been accessible. One trend is clear: ML methods are becoming indispensable tools at the workbenches of computational scientists. Publicly available solutions (such as, SchetPack[44], TensorMol[88], aenet[153], MLatom[154], AMP[155], PROPHet[156], DeePMD-kit[157], TorchANI[158], DScribe[159] ChemML[160], SimpleNN[161] and PiNN[162]) for designing and training ML models are currently contributing to the accelerated development and application of data science in chemistry. Several commercial distributions also follow this trend. The Amsterdam Density Functional code[163], for example, already includes interfaces to several popular ML interatomic potentials (see information from Software for Chemistry & Materials). In the powerful molecular modelling kit Atomic Simulation Environment[164], custom and existing ML models can be invoked to calculate molecular properties. Additionally, molecular dynamics packages such as LAMMPS and Tinker recently added interfaces[165] for ML interatomic potentials, lowering the entrance threshold for anyone interested in augmenting their research with ML techniques. Altogether, ML is becoming an important part of time-proven QM codes. For instance, the latest release of ORCA 5.0 introduced ML-optimized DFT integration grids[166], which are now the default for the majority of DFT computations. Benchmarks explicitly demonstrate improvements in the assessment of thermochemistry, reaction barriers, non-covalent interactions and vibrational frequencies, exemplifying how pure electronic structure calculations can benefit from data science[166]. We believe that these trends will continue and the use of ML will expand to other chemical problems and properties.

Published online 25 August 2022

1. Purvis, G. D. & Bartlett, R. J. A full coupled-cluster singles and doubles model: the inclusion of disconnected triples. *J. Chem. Phys.* **76**, 1910–1918 (1982).
2. Burke, K. Perspective on density functional theory. *J. Chem. Phys.* **136**, 150901 (2012).
3. Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **115**, 2315–2372 (2017).
4. Thiel, W. Semiempirical quantum–chemical methods. *WIREs Comput. Mol. Sci.* **4**, 145–157 (2014).
5. Ratcliff, L. E. et al. Challenges in large scale quantum mechanical calculations. *WIREs Comput. Mol. Sci.* **7**, e1290 (2017).
6. von Lilienfeld, O. A., Müller, K.-R. & Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **4**, 347–358 (2020).
7. Keith, J. A. et al. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* **121**, 9816–9872 (2021).
8. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
9. Dral, P. O. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* **11**, 2336–2347 (2020).
10. Musil, F. et al. Physics-inspired structural representations for molecules and materials. *Chem. Rev.* **121**, 9759–9815 (2021).
11. Vamathevan, J. et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).
12. David, L., Thakkar, A., Mercado, R. & Engkvist, O. Molecular representations in AI-driven drug discovery: a review and practical guide. *J. Cheminform.* **12**, 56 (2020).
13. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, eaap7885 (2018).
14. Pollice, R. et al. Data-driven strategies for accelerated materials design. *Acc. Chem. Res.* **54**, 849–860 (2021).
15. Guo, H., Wang, Q., Stuke, A., Urban, A. & Artrith, N. Accelerated atomistic modeling of solid-state battery materials with machine learning. *Front. Energy Res.* **9**, 265 (2021).
16. Kulichenko, M. et al. The rise of neural networks for materials and chemical dynamics. *J. Phys. Chem. Lett.* **12**, 6227–6243 (2021).
17. Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).
18. Gokcan, H. & Isayev, O. Learning molecular potentials with neural networks. *WIREs Comput. Mol. Sci.* **12**, e1564.
19. Dral, P. O. & Barbatti, M. Molecular excited states through a machine learning lens. *Nat. Rev. Chem.* **5**, 388–405 (2021).
20. Westermayr, J. & Marquetand, P. Machine learning for electronically excited states of molecules. *Chem. Rev.* **121**, 9873–9926 (2021).
21. Jorner, K., Tomberg, A., Bauer, C., Sköld, C. & Norrby, P.-O. Organic reactivity from mechanism to machine learning. *Nat. Rev. Chem.* **5**, 240–255 (2021).
22. Gallegos, L. C., Luchini, G., St. John, P. C., Kim, S. & Paton, R. S. Importance of engineered and learned

molecular representations in predicting organic reactivity, selectivity, and chemical properties. *Acc. Chem. Res.* **54**, 827–836 (2021).
23. Toyao, T. et al. Machine learning for catalysis informatics: recent applications and prospects. *ACS Catal.* **10**, 2260–2297 (2020).
24. Yang, X., Wang, Y., Byrne, R., Schneider, G. & Yang, S. Concepts of artificial intelligence for computer-assisted drug discovery. *Chem. Rev.* **119**, 10520–10594 (2019).
25. Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).
26. Gao, H. et al. Using machine learning to predict suitable conditions for organic reactions. *ACS Cent. Sci.* **4**, 1465–1476 (2018).
27. Bartók, A. P. & Csányi, G. Gaussian approximation potentials: a brief tutorial introduction. *Int. J. Quantum Chem.* **115**, 1051–1057 (2015).
28. Thompson, A. P., Swiler, L. P., Trott, C. R., Foiles, S. M. & Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).
29. Novikov, I. S., Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. The MLIP package: moment tensor potentials with MPI and active learning. *Mach. Learn. Sci. Technol.* **2**, 025002 (2021).
30. 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.* **240**, 38–45 (2019).
31. Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **9**, 3887 (2018).
32. Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. Machine learning of molecular properties: locality and active learning. *J. Chem. Phys.* **148**, 241727 (2018).
33. Behler, J. Perspective: machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
34. Behler, J. & Csányi, G. Machine learning potentials for extended systems: a perspective. *Eur. Phys. J. B* **94**, 142 (2021).
35. Daw, M. S., Foiles, S. M. & Baskes, M. I. The embedded-atom method: a review of theory and applications. *Mater. Sci. Rep.* **9**, 251–310 (1993).
36. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
37. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
38. Behler, J. Constructing high-dimensional neural network potentials: a tutorial review. *Int. J. Quantum Chem.* **115**, 1032–1050 (2015).
39. 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.* **12**, 398 (2021).
40. Ruddigkeit, L., van Deursen, R., Blum, L. C. & Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **52**, 2864–2875 (2012).
41. Devereux, C. et al. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.* **16**, 4192–4202 (2020).
42. Zubatyuk, R., Smith, J. S., Leszczynski, J. & Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* **5**, eaav6490 (2019).
43. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet — a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
44. Schütt, K. T. et al. SchNetPack: a deep learning toolbox for atomistic systems. *J. Chem. Theory Comput.* **15**, 448–455 (2019).
45. Unke, O. T. & Meuwly, M. PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
46. Gasteiger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2003.03123 (2020).
47. Gasteiger, J., Giri, S., Margraf, J. T. & Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2011.14115 (2020).
48. Mueller, T., Hernandez, A. & Wang, C. Machine learning for interatomic potential models. *J. Chem. Phys.* **152**, 050902 (2020).
49. Glick, Z. L., Koutsoukas, A., Cheney, D. L. & Sherrill, C. D. Cartesian message passing neural networks for directional properties: fast and transferable atomic multipoles. *J. Chem. Phys.* **154**, 224103 (2021).
50. Lubbers, N., Smith, J. S. & Barros, K. Hierarchical modeling of molecular energies using a deep neural network. *J. Chem. Phys.* **148**, 241715 (2018).
51. Nebgen, B. et al. Transferable dynamic molecular charge assignment using deep neural networks. *J. Chem. Theory Comput.* **14**, 4687–4698 (2018).
52. Sifain, A. E. et al. Discovering a transferable charge assignment model using machine learning. *J. Phys. Chem. Lett.* **9**, 4495–4501 (2018).
53. Magedov, S., Koh, C., Malone, W., Lubbers, N. & Nebgen, B. Bond order predictions using deep neural networks. *J. Appl. Phys.* **129**, 064701 (2021).
54. Zubatiuk, T. et al. Machine learned Hückel theory: interfacing physics and deep neural networks. *J. Chem. Phys.* **154**, 244108 (2021).
55. Caruana, R. Multitask learning. *Mach. Learn.* **28**, 41–75 (1997).
56. Sifain, A. E. et al. Predicting phosphorescence energies and inferring wavefunction localization with machine learning. *Chem. Sci.* **12**, 10207–10217 (2021).
57. Tretiak, S. & Mukamel, S. Density matrix analysis and simulation of electronic excitations in conjugated and aggregated molecules. *Chem. Rev.* **102**, 3171–3212 (2002).
58. Bader, R. F. W. *Atoms in Molecules: a Quantum Theory* (Clarendon Press, 1994).
59. Zubatyuk, R., Smith, J. S., Nebgen, B. T., Tretiak, S. & Isayev, O. Teaching a neural network to attach and detach electrons from molecules. *Nat. Commun.* **12**, 4870 (2021).
60. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more: sampling chemical space with active learning. *J. Chem. Phys.* **148**, 241733 (2018).
61. Miksch, A. M., Morawietz, T., Kästner, J., Urban, A. & Artrith, N. Strategies for the construction of machine-learning potentials for accurate and efficient atomic-scale simulations. *Mach. Learn. Sci. Technol.* **2**, 031001 (2021).
62. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules. *Sci. Data* **4**, 170193 (2017).
63. Smith, J. S. et al. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Sci. Data* **7**, 134 (2020).
64. Chambers, J. et al. UniChem: a unified chemical structure cross-referencing and identifier tracking system. *J. Cheminform.* **5**, 3 (2013).
65. Wu, Z. et al. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **9**, 513–530 (2018).
66. Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).
67. Nakata, M. & Shimazaki, T. PubChemQC project: a large-scale first-principles electronic structure database for data-driven chemistry. *J. Chem. Inf. Model.* **57**, 1300–1308 (2017).
68. Curtarolo, S. et al. AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
69. Pinheiro, G. A. et al. Machine learning prediction of nine molecular properties based on the SMILES representation of the QM9. *J. Phys. Chem. A* **124**, 9854–9866 (2020).
70. Wießner, M. et al. Complete determination of molecular orbitals by measurement of phase symmetry and electron density. *Nat. Commun.* **5**, 4156 (2014).
71. Gao, W. et al. Real-space charge-density imaging with sub-ångström resolution by four-dimensional electron microscopy. *Nature* **575**, 480–484 (2019).
72. Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **44**, 129–138 (1977).
73. Marenich, A. V., Jerome, S. V., Cramer, C. J. & Truhlar, D. G. Charge Model 5: an extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* **8**, 527–541 (2012).
74. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).
75. Glendening, E. D., Landis, C. R. & Weinhold, F. Natural bond orbital methods. *WIREs Comput. Mol. Sci.* **2**, 1–42 (2012).
76. Pérez de la Luz, A., Aguilar-Pineda, J. A., Méndez-Bermúdez, J. G. & Alejandre, J. Force field parametrization from the hirshfeld molecular electronic density. *J. Chem. Theory Comput.* **14**, 5949–5958 (2018).
77. Honda, S., Yamasaki, K., Sawada, Y. & Morii, H. 10 residue folded peptide designed by segment statistics. *Structure* **12**, 1507–1518 (2004).
78. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Mol. Biol.* **9**, 425–430 (2002).
79. Ševčík, J. et al. Structure of glucoamylase from *Saccharomycopsis fibuligera* at 1.7 Å resolution. *Acta Cryst. D.* **54**, 854–866 (1998).
80. Bleiziffer, P., Schaller, K. & Riniker, S. Machine learning of partial charges derived from high-quality quantum-mechanical calculations. *J. Chem. Inf. Model.* **58**, 579–590 (2018).
81. Wang, X. & Gao, J. Atomic partial charge predictions for furanoses by random forest regression with atom type symmetry function. *RSC Adv.* **10**, 666–673 (2020).
82. Kato, K. et al. High-precision atomic charge prediction for protein systems using fragment molecular orbital calculation and machine learning. *J. Chem. Inf. Model.* **60**, 3361–3368 (2020).
83. Wang, J. et al. Fast and accurate prediction of partial charges using atom-path-descriptor-based machine learning. *Bioinformatics* **36**, 4721–4728 (2020).
84. Martin, R. & Heider, D. ContraDRG: automatic partial charge prediction by machine learning. *Front. Genet.* **10**, 990 (2019).
85. Cioslowski, J. & Surján, P. R. An observable-based interpretation of electronic wavefunctions: application to "hypervalent" molecules. *J. Mol. Struc. THEOCHEM* **255**, 9–33 (1992).
86. Francl, M. M., Carey, C., Chirlian, L. E. & Gange, D. M. Charges fit to electrostatic potentials. II. Can atomic charges be unambiguously fit to electrostatic potentials? *J. Comput. Chem.* **17**, 367–383 (1996).
87. Veit, M., Wilkins, D. M., Yang, Y., DiStasio, R. A. & Ceriotti, M. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *J. Chem. Phys.* **153**, 024113 (2020).
88. Yao, K., Herr, J. E., Toth, D. W., Mckintyre, R. & Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* **9**, 2261–2269 (2018).
89. Loeffler, J. R. et al. Conformational shifts of stacked heteroaromatics: vacuum vs. water studied by machine learning. *Front. Chem.* https://doi.org/10.3389/fchem.2021.641610 (2021).
90. 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.* **8**, 3192–3203 (2017).
91. McGaughey, G. B., Gagné, M. & Rappé, A. K. π-Stacking interactions: alive and well in proteins. *J. Biol. Chem.* **273**, 15458–15463 (1998).
92. Metcalf, D. P. et al. Approaches for machine learning intermolecular interaction energies and application to energy components from symmetry adapted perturbation theory. *J. Chem. Phys.* **152**, 074103 (2020).
93. Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *WIREs Comput. Mol. Sci.* **2**, 254–272 (2012).
94. Glick, Z. L. et al. AP-Net: an atomic-pairwise neural network for smooth and transferable interaction potentials. *J. Chem. Phys.* **153**, 044112 (2020).
95. Geerlings, P., De Proft, F. & Langenaeker, W. Conceptual density functional theory. *Chem. Rev.* **103**, 1793–1874 (2003).
96. Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. General-purpose machine learning potentials capturing nonlocal charge transfer. *Acc. Chem. Res.* **54**, 808–817 (2021).
97. Grisafi, A. et al. Transferable machine-learning model of the electron density. *ACS Cent. Sci.* **5**, 57–64 (2019).
98. Glielmo, A., Sollich, P. & De Vita, A. Accurate interatomic force fields via machine learning with covariant kernels. *Phys. Rev. B* **95**, 214302 (2017).
99. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of

quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
100. Nguyen, T. T. et al. Comparison of permutationally invariant polynomials, neural networks, and Gaussian approximation potentials in representing water interactions through many-body expansions. *J. Chem. Phys.* **148**, 241725 (2018).
101. Fabrizio, A., Grisafi, A., Meyer, B., Ceriotti, M. & Corminboeuf, C. Electron density learning of non-covalent systems. *Chem. Sci.* **10**, 9424–9432 (2019).
102. Deringer, V. L. et al. Gaussian process regression for materials and molecules. *Chem. Rev.* **121**, 10073–10141 (2021).
103. Cuevas-Zuviría, B. & Pacios, L. F. Analytical model of electron density and its machine learning inference. *J. Chem. Inf. Model.* **60**, 3831–3842 (2020).
104. Cuevas-Zuviría, B. & Pacios, F. Machine learning of analytical electron density in large molecules through message-passing. *J. Chem. Inf. Model.* **61**, 2658–2666.
105. Lewis, A. M., Grisafi, A., Ceriotti, M. & Rossi, M. Learning electron densities in the condensed phase. *J. Chem. Theory Comput.* **17**, 7203–7214 (2021).
106. Zou, S.-J. et al. Recent advances in organic light-emitting diodes: toward smart lighting and displays. *Mater. Chem. Front.* **4**, 788–820 (2020).
107. Nayak, P. K., Mahesh, S., Snaith, H. J. & Cahen, D. Photovoltaic solar cell technologies: analysing the state of the art. *Nat. Rev. Mater.* **4**, 269–285 (2019).
108. Hirohata, A. et al. Review on spintronics: principles and device applications. *J. Magn. Magn. Mater.* **509**, 166711 (2020).
109. Tretiak, S., Chernyak, V. & Mukamel, S. Localized electronic excitations in phenylacetylene dendrimers. *J. Phys. Chem. B* **102**, 3310–3315 (1998).
110. Zhao, L., Pan, S., Holzmann, N., Schwerdtfeger, P. & Frenking, G. Chemical bonding and bonding models of main-group compounds. *Chem. Rev.* **119**, 8781–8845 (2019).
111. Mayer, I. Bond order and valence indices: a personal account. *J. Comput. Chem.* **28**, 204–221 (2007).
112. Wiberg, K. B. Application of the Pople–Santry–Segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane. *Tetrahedron* **24**, 1083–1096 (1968).
113. Alonso, M. & Herradón, B. Neural networks as a tool to classify compounds according to aromaticity criteria. *Chem. Eur. J.* **13**, 3913–3923 (2007).
114. Alonso, M., Miranda, C., Martín, N. & Herradón, B. Chemical applications of neural networks: aromaticity of pyrimidine derivatives. *Phys. Chem. Chem. Phys.* **13**, 20564–20574 (2011).
115. Ferreira, A. R. Chemical bonding in metallic glasses from machine learning and crystal orbital hamilton population. *Phys. Rev. Mater.* **4**, 113603 (2020).
116. Matlock, M. K., Dang, N. L. & Swamidass, S. J. Learning a local-variable model of aromatic and conjugated systems. *ACS Cent. Sci.* **4**, 52–62 (2018).
117. Li, H., Collins, C., Tanha, M., Gordon, G. J. & Yaron, D. J. A density functional tight binding layer for deep learning of chemical Hamiltonians. *J. Chem. Theory Comput.* **14**, 5764–5776 (2018).
118. Schütt, K. T., Gastegger, M., Tkatchenko, A., Müller, K.-R. & Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019).
119. Wang, Z. et al. Machine learning method for tight-binding Hamiltonian parameterization from ab-initio band structure. *npj Comput. Mater.* **7**, 11 (2021).
120. Hoffmann, R. An extended Hückel theory. I. Hydrocarbons. *J. Chem. Phys.* **39**, 1397–1412 (1963).
121. Grabill, L. P. & Berger, R. F. Calibrating the extended Hückel method to quantitatively screen the electronic properties of materials. *Sci. Rep.* **8**, 10530 (2018).
122. Zhou, G., Lubbers, N., Barros, K., Tretiak, S. & Nebgen, B. Deep learning of dynamically responsive chemical Hamiltonians with semiempirical quantum mechanics. *Proc. Natl Acad. Sci. USA* **119**, e2120333119 (2022).
123. Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **10**, 209–220 (1989).
124. Elstner, M. et al. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58**, 7260–7268 (1998).
125. Gaus, M., Cui, Q. & Elstner, M. Density functional tight binding: application to organic and biological molecules. *WIREs Comput. Mol. Sci.* **4**, 49–61 (2014).
126. Panosetti, C., Engelmann, A., Nemec, L., Reuter, K. & Margraf, J. T. Learning to use the force: fitting repulsive potentials in density-functional tight-binding with Gaussian process regression. *J. Chem. Theory Comput.* **16**, 2181–2191 (2020).
127. Kranz, J. J., Kubillus, M., Ramakrishnan, R., von Lilienfeld, O. A. & Elstner, M. Generalized density-functional tight-binding repulsive potentials from unsupervised machine learning. *J. Chem. Theory Comput.* **14**, 2341–2352 (2018).
128. Hastie, T., Tibshirani, R. & Friedman, J. *Elements Of Statistical Learning: Data Mining, Inference, And Prediction* 2nd edn (Springer, 2009).
129. Snyder, J. C., Rupp, M., Hansen, K., Müller, K.-R. & Burke, K. Finding density functionals with machine learning. *Phys. Rev. Lett.* **108**, 253002 (2012).
130. Li, L. et al. Understanding machine-learned density functionals. *Int. J. Quantum Chem.* **116**, 819–833 (2016).
131. Brockherde, F. et al. Bypassing the Kohn–Sham equations with machine learning. *Nat. Commun.* **8**, 872 (2017).
132. Hollingsworth, J., Baker, T. E. & Burke, K. Can exact conditions improve machine-learned density functionals? *J. Chem. Phys.* **148**, 241743 (2018).
133. 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.* **11**, 2087–2096 (2015).
134. McGibbon, R. T. et al. Improving the accuracy of Møller–Plesset perturbation theory with neural networks. *J. Chem. Phys.* **147**, 161725 (2017).
135. Wilkins, D. M. et al. Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc. Natl Acad. Sci.* **116**, 3401–3406 (2019).
136. Kulik, H. et al. Roadmap on machine learning in electronic structure. *Electron. Struct.* https://doi.org/10.1088/2516-1075/ac572f (2022).
137. Gastegger, M., McSloy, A., Luya, M., Schütt, K. T. & Maurer, R. J. A deep neural network for molecular wave functions in quasi-atomic minimal basis representation. *J. Chem. Phys.* **153**, 044123 (2020).
138. Zubatiuk, T. & Isayev, O. Development of multimodal machine learning potentials: toward a physics-aware artificial intelligence. *Acc. Chem. Res.* **54**, 1575–1585 (2021).
139. Chandrasekaran, A. et al. Solving the electronic structure problem with machine learning. *npj Comput. Mater.* **5**, 22 (2019).
140. Smith, J. S. et al. Automated discovery of a robust interatomic potential for aluminum. *Nat. Commun.* **12**, 1257 (2021).
141. Jia, W. et al. Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2005.00223 (2020).
142. Jung, J. et al. New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems. *J. Comput. Chem.* **42**, 231–241 (2021).
143. Jinnouchi, R., Miwa, K., Karsai, F., Kresse, G. & Asahi, R. On-the-fly active learning of interatomic potentials for large-scale atomistic simulations. *J. Phys. Chem. Lett.* **11**, 6946–6955 (2020).
144. Zhang, L., Lin, D.-Y., Wang, H., Car, R. & E, W. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Mater.* **3**, 023804 (2019).
145. Noé, F., Olsson, S., Köhler, J. & Wu, H. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 (2019).
146. Ribeiro, J. M. L., Bravo, P., Wang, Y. & Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **149**, 072301 (2018).
147. Wang, Y., Ribeiro, J. M. L. & Tiwary, P. Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nat. Commun.* **10**, 3573 (2019).
148. Gebauer, N. W. A., Gastegger, M. & Schütt, K. T. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1906.00957 (2020).
149. Kuenneth, C. et al. Polymer informatics with multi-task learning. *Patterns* **2**, 100238 (2021).
150. Krämer, M. et al. Charge and exciton transfer simulations using machine-learned hamiltonians. *J. Chem. Theory Comput.* **16**, 4061–4070 (2020).
151. Jeong, W. et al. Automation of active space selection for multireference methods via machine learning on chemical bond dissociation. *J. Chem. Theory Comput.* **16**, 2389–2399 (2020).
152. Qiao, Z., Welborn, M., Anandkumar, A., Manby, F. R. & Miller, T. F. OrbNet: deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **153**, 124111 (2020).
153. Artrith, N. & Urban, A. An implementation of artificial neural-network potentials for atomistic materials simulations: performance for TiO_2. *Comput. Mater. Sci.* **114**, 135–150 (2016).
154. Dral, P. O. et al. MLatom 2: an integrative platform for atomistic machine learning. *Top. Curr. Chem.* **379**, 27 (2021).
155. Khorshidi, A. & Peterson, A. A. Amp: a modular approach to machine learning in atomistic simulations. *Computer Phys. Commun.* **207**, 310–324 (2016).
156. Kolb, B., Lentz, L. C. & Kolpak, A. M. Discovering charge density functionals and structure-property relationships with PROPhet: a general framework for coupling machine learning and first-principles methods. *Sci. Rep.* **7**, 1192 (2017).
157. Wang, H., Zhang, L., Han, J. & E, W. DeePMD-kit: a deep learning package for many-body potential energy representation and molecular dynamics. *Computer Phys. Commun.* **228**, 178–184 (2018).
158. Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J. S. & Roitberg, A. E. TorchANI: a free and open source pytorch-based deep learning implementation of the ANI neural network potentials. *J. Chem. Inf. Model.* **60**, 3408–3415 (2020).
159. Himanen, L. et al. DScribe: library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020).
160. Haghighatlari, M. et al. ChemML: a machine learning and informatics program package for the analysis, mining, and modeling of chemical and materials data. *WIREs Comput. Mol. Sci.* **10**, e1458 (2020).
161. Lee, K., Yoo, D., Jeong, W. & Han, S. SIMPLE-NN: an efficient package for training and executing neural-network interatomic potentials. *Comput. Phys. Commun.* **242**, 95–103 (2019).
162. Shao, Y., Hellström, M., Mitev, P. D., Knijff, L. & Zhang, C. PiNN: a Python library for building atomic neural networks of molecules and materials. *J. Chem. Inf. Model.* **60**, 1184–1193 (2020).
163. Velde, Gte et al. Chemistry with ADF. *J. Comput. Chem.* **22**, 931–967 (2001).
164. Larsen, A. H. et al. The atomic simulation environment — a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
165. Chen, M. S., Morawietz, T., Mori, H., Markland, T. E. & Artrith, N. AENET–LAMMPS and AENET–TINKER: interfaces for accurate and efficient molecular dynamics simulations with machine learning potentials. *J. Chem. Phys.* **155**, 074801 (2021).
166. Neese, F. Software update: the ORCA program system — version 5.0. *WIREs Comput. Mol. Sci.* https://doi.org/10.1002/wcms.1606 (2022).
167. Cova, T. F. G. G. & Pais, A. A. C. C. Deep learning for deep chemistry: optimizing the prediction of chemical patterns. *Front. Chem.* **7**, 809 (2019).
168. Bzdok, D., Krzywinski, M. & Altman, N. Machine learning: supervised methods. *Nat. Methods* **15**, 5–6 (2018).
169. Shaidu, Y. et al. A systematic approach to generating accurate neural network potentials: the case of carbon. *npj Comput. Mater.* **7**, 52 (2021).
170. Botu, V., Batra, R., Chapman, J. & Ramprasad, R. Machine learning force fields: construction, validation, and outlook. *J. Phys. Chem. C* **121**, 511–522 (2017).
171. Senftle, T. P. et al. The ReaxFF reactive force-field: development, applications and future directions. *npj Comput. Mater.* **2**, 15011 (2016).
172. Leach, A. R. *Molecular Modelling: Principles and Applications* 2nd edn, Ch. 7 (Pearson, 2001)
173. Unke, O. T. et al. Machine learning force fields. *Chem. Rev.* **121**, 10142–10186 (2021).
174. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
175. Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 (1989).
176. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**, 251–257 (1991).
177. Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Edn* **56**, 12828–12840 (2017).
178. Benoit, M. et al. Measuring transferability issues in machine-learning force fields: the example of gold–iron interactions with linearized potentials. *Mach. Learn. Sci. Technol.* **2**, 025003 (2021).

179. Anderson, B., Hy, T.-S. & Kondor, R. Cormorant: Covariant Molecular Neural Networks. Preprint at *Arxiv* https://arxiv.org/abs/1906.04015 (2019).
180. Jackson, R., Zhang, W. & Pearson, J. TSNet: predicting transition state structures with tensor field networks and transfer learning. *Chem. Sci.* **12**, 10022–10040 (2021).
181. Kocer, E., Mason, J. K. & Erturk, H. A novel approach to describe chemical environments in high-dimensional neural network potentials. *J. Chem. Phys.* **150**, 154102 (2019).

Acknowledgements
The work at Los Alamos National Laboratory (LANL) was supported by the LANL Directed Research and Development Funds (LDRD) and performed in part at the Center for Nonlinear Studies (CNLS) and the Center for Integrated Nanotechnologies (CINT), a US Department of Energy (DOE) Office of Science user facility at LANL. N.F. and M.K. acknowledge financial support from the Director's Postdoctoral Fellowship at LANL funded by LDRD. K.B. and S.T. acknowledge support from the US DOE, Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under Triad National Security (Triad) contract grant number 89233218CNA000001 (FWP: LANLE3F2). This research used resources provided by the LANL Institutional Computing Program. LANL is managed by Triad National Security for the US DOE's NNSA, under contract number 89233218CNA000001. A.I.B. acknowledges the R. Gaurth Hansen Professorship. O.I. acknowledges support from the National Science Foundation (NSF) grants CHE-1802789 and CHE-2041108. The work performed by O.I. and R.Z. in part was made possible by the Office of Naval Research (ONR) through support provided by the Energetic Materials Program (MURI grant number N00014-21-1-2476).

Author contributions
N.F., R.Z., M.K., J.S.S., B.N., R.M., Y.W.L., A.I.B., K.B., O.I. and S.T. researched data for the article. N.F., R.Z., M.K., J.S.S., B.N., K.B., O.I. and S.T. contributed substantially to discussion of the content. N.F., R.Z., M.K., N.L., O.I. and S.T. wrote the article. All authors reviewed and/or edited the manuscript before submission.

Competing interests
Authors declare no competing interests.

Peer review information
Nature Reviews Chemistry thanks Eric Bittner, Hao Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

RELATED LINKS
Software for Chemistry & Materials — Machine Learning Potentials: https://www.scm.com/product/machine-learning-potentials/

© Springer Nature Limited 2022, corrected publication 2022