Applied Physics Reviews REVIEW pubs.aip.org/aip/are

On the design space between molecular mechanics and machine learning force fields

Cite as: Appl. Phys. Rev. **12**, 021304 (2025); doi: 10.1063/5.0237876
Submitted: 7 September 2024 · Accepted: 18 March 2025 ·
Published Online: 2 April 2025

Yuanqing Wang,[1,2,3,a] Kenichiro Takaba,[4] Michael S. Chen,[1] Marcus Wieder,[5] Yuzhi Xu,[1,6] Tong Zhu,[6] John Z. H. Zhang,[6,7] Arnav Nagle,[8] Kuang Yu,[9] Xinyan Wang,[10] Daniel J. Cole,[11] Joshua A. Rackers,[12] Kyunghyun Cho,[2,3,12] Joe G. Greener,[13] Peter Eastman,[14] Stefano Martiniani,[1,3,15] and Mark E. Tuckerman[1,3]

AFFILIATIONS

[1]Simons Center for Computational Physical Chemistry, Department of Chemistry, New York University, New York, New York 10003, USA
[2]Center for Data Science, New York University, New York, New York 10004, USA
[3]Courant Institute of Mathematical Sciences, New York University, New York, New York 10003, USA
[4]Asahi Kasei Pharma Corporation, 632-1 Mifuku, Izunokuni, Shizuoka 410-2321, Japan
[5]Open Molecular Software Foundation, Davis, California 95616, USA
[6]NYU-ECNU Center for Computational Chemistry and Shanghai Frontiers Science Center of Artificial Intelligence and Deep Learning, NYU Shanghai, Shanghai 200062, People's Republic of China
[7]Faculty of Synthetic Biology, Shenzhen University of Advanced Technology, Shenzhen 518055, People's Republic of China
[8]University of Washington, Seattle, Washington 98195, USA
[9]Institute of Materials Research, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, People's Republic of China
[10]DP Technology, Beijing 100089, People's Republic of China
[11]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom
[12]Prescient Design, Genentech, New York, New York 10004, USA
[13]Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom
[14]Department of Chemistry, Stanford University, Stanford, California 94305, USA
[15]Center for Soft Matter Research, Department of Physics, New York University, New York, New York 10003, USA

[a)]**Author to whom correspondence should be addressed:** wangyq@wangyq.net

ABSTRACT

A force field as accurate as quantum mechanics (QMs) and as fast as molecular mechanics (MMs), with which one can simulate a biomolecular system efficiently enough and meaningfully enough to get quantitative insights, is among the most ardent dreams of biophysicists—a dream, nevertheless, not to be fulfilled any time soon. Machine learning force fields (MLFFs) represent a meaningful endeavor in this direction, where differentiable neural functions are parametrized to fit *ab initio* energies and forces through automatic differentiation. We argue that, as of now, the utility of the MLFF models is no longer bottlenecked by accuracy but primarily by their speed, as well as stability and generalizability—many recent variants, on limited chemical spaces, have long surpassed the *chemical accuracy* of 1 kcal/mol—the empirical threshold beyond which realistic chemical predictions are possible—though still magnitudes slower than MM. Hoping to kindle exploration and design of faster, albeit perhaps slightly less accurate MLFFs, in this review, we focus our attention on the technical design space (the speed-accuracy trade-off) between MM and ML force fields. After a brief review of the building blocks (from a machine learning-centric point of view) of force fields of either kind, we discuss the desired properties and challenges now faced by the force field development community, survey the efforts to make MM force fields more accurate and ML force fields faster, and envision what the next generation of MLFF might look like.

TABLE OF CONTENTS

I. INTRODUCTION: THE PAST, PRESENT, AND FUTURE OF FORCE FIELDS

Computational chemists and biophysicists are interested in characterizing the *energy landscape* of many-body systems—the distribution of coordinates $\mathbf{x} \in \mathbb{R}^{N\times 3}$ at a certain state with energy $U(\mathbf{x})$, which adopts a Boltzmann form[1]

$$p(\mathbf{x}) \propto \exp\left(-\frac{U(\mathbf{x})}{kT}\right), \tag{1}$$

where N is the number of particles in a many-body system, k is the Boltzmann constant, and T is the temperature. In a simulation, since the true (reference) energy U is almost always inaccessible, and *ab initio* methods are usually prohibitively expensive, one resorts to a surrogate model, as an *ersatz* for the true energy landscape, as a function of the coordinates $\mathbf{x}$, the identity of the particles in the system $\mathbf{h} \in \mathbb{R}^N$, and a set of parameters Φ

$$\hat{p} \propto \exp\left(-\frac{\hat{U}(\mathbf{x}; h, \Phi)}{kT}\right). \tag{2}$$

Evidently, the closer U and $\hat{U}$ are, the smaller the divergence between the true and simulated probability distribution p and $\hat{p}$ will be. We call this parametrized scalar field $\hat{U} : \mathbb{R}^{N\times 3} \to \mathbb{R}^1$ a *force field* (FF). Molecular dynamics (MD) based on Monte Carlo simulations is usually employed to generate samples from this distribution.

Dating back to Levitt and Lifson,[2] Hagler *et al.*,[3] and McCammon *et al.*[4] since 1969, molecular mechanics (MMs) force fields have been curated, using structural and quantum mechanic (QM) data, to capture the qualitative behavior of biomolecular systems[5–12] to power the *in silico* modeling of all aspects of chemistry, from drug discovery to material sciences. The blessing and the curse of MM force fields both lie in their simple functional forms [Eq. (9)]. On the one hand, these terms afford linear $\mathcal{O}(N)$ runtime complexity and can be further aggressively optimized in modern compute hardware, namely, graphics processing units (GPUs), simulating more than hundreds of nanoseconds per day for many biomolecular drug targets[13–16] while still achieving useful accuracy for tasks such as predicting protein-ligand binding free energies.[17–19] At the same time, the limited *expressiveness* of this functional form dictates that it is impossible to fit the QM energies and forces well, especially in the high-energy region (which is of importance when it comes to simulating events such as reactions)—see Fig. 2 for a comparison of QM and MM energies of a very simple ethanol molecule, although they recover the *position* of QM minima relatively well.[20] Worse still, even within the limits of this functional form, there is no guarantee that maximal expressiveness has been achieved, as the assignment of parameters to the chemical motifs (atoms, bonds, angles, and torsions) in a MM force field has traditionally been relying on a human-derived, labor-intensive, and inextendable scheme termed *atom typing*, where atoms of distinct nature are forced to share

parameters. Takaba *et al.*[21] show that, on very limited chemical space and low energy region, the energy disagreement between legacy force fields and QM is far beyond the chemical accuracy of 1 kcal/mol—the empirical threshold under which we believe that the qualitative characterization of a many-body system is possible. Moreover, even when coupled with a trainable, flexible parameterization engine, the *training* accuracy still cannot exceed the chemical accuracy, suggesting a limitation in both the *functional form* and *parameterization* steps of MM force fields.

Another line of fruitful research[23–34] focuses on developing machine learning force fields (MLFFs) with flexible functional forms to fit the energy (and force) landscape of *ab initio* calculations. Often, the energy is predicted by a (typically, equivariant or invariant) neural network, whereas the forces are given by using automatic differentiation w.r.t. the positions of particles. We argue that the energy and force accuracy of machine learning force fields is no longer a limiting factor for the widespread applications thereof, as most such models achieve an energy error well below the chemical accuracy of 1 kcal/mol. In addition to the lack of stability (the guarantee that the simulation would not yield unphysical events after sufficient time steps) and generalizability (the ability to simulate diverse chemical systems different from the training sets), the speed of MLFFs is what prevents them from wide applications—although they are usually faster by magnitudes than QM calculations (and scale linearly w.r.t. the size of the system), they are still hundreds of times slower than MM force fields. Note that, very recently, Kabylda *et al.*[35] achieved superior speed with merely a 40-fold slow down compared to classical MM. For small molecule systems up to 10^2 atoms, some of the fastest MLFFs[31] still take around 1 ms per energy and force evaluation on an A100 GPU, compared to less than 0.005 ms for MM force fields. As such, to simulate any biomolecular system of considerable size for a reasonably long time frame would usually have a prohibitive computational cost, significantly limiting its deployment to further our understanding of biologically relevant systems.

In this review, we direct our attention to the design space (in terms of the speed-accuracy trade-off) between ML and MM force fields. Hoping that this would inspire the design of a class of MLFF incorporating MM philosophy that is significantly faster, more stable, more interpretable, and more generalizable than current state-of-the-art MLFFs albeit slightly less accurate/expressive, we organize this review as follows. First, in Sec. II, we outline the *desiderata* of force fields. Second, in Secs. III and IV, we briefly review the building blocks of both MM and ML force fields and recent advances in the functional forms and parametrizations thereof, to inspire the design of novel models starting from building blocks of either school. Third, in Sec. V A, we discuss the datasets and protocols for curating MM and ML force fields to encourage the field to converge on standard best practices. Finally, we envision the shape of the next generation of ultra-fast MLFF with high utility in biomolecular modeling in Sec. VIII. We summarize (albeit with considerable overgeneralization) the key theoretical and practical properties discussed in this paper in Table I, where the blueprint of a general-purpose, high-utility force field is also outlined.

II. DESIDERATA: THE BALANCE BETWEEN SPEED AND ACCURACY, AND BEYOND

The speed and accuracy are two natural axes in which the community is interested (Fig. 1)—a faster and more accurate force field is almost always desired. With this in mind, we review the desired properties of a force field, regardless of its type, for the applications in physical modeling.

A. Invariance

Symmetries are inherent in all physical systems. To exploit these symmetries in model construction and parameterization enhances data efficiency as it avoids unnecessary replicated manifestations of the same piece of data and avoids unphysical model interpretations. In other words, the law that two *symmetrical* data points would yield the same prediction does not need to be *learned* if it were *encoded*, and the physically unreasonable scenario where these two data points manifesting different predictions is not possible. Nevertheless, we also point out that encoding such symmetry is not the only avenue toward

TABLE I. Molecular mechanics (MM) vs machine learning force fields (MLFF).

	Molecular mechanics (MM)	Machine learning force fields (MLFF)	Envisioned general purpose FF (Sec. VIII)
Genesis (Sec. I)	Levitt and Lifson[2]	Behler and Parrinello[36]	
Runtime Complexity	$\mathcal{O}(N)$	$\mathcal{O}(N)$	$\mathcal{O}(N)$
Speed (Sec. VII)	$> 1\mu$s/day	on the magnitude of 1 ns/day	$\approx 1\mu$s/day
Invariance (Equation 4)	$E(3)$	$E(3)$	$E(3)$
Universality (Lemma 3)	Impossible	Possible	Guaranteed
Accuracy	$>$ 1 kcal/mol	$\ll$ 1 kcal/mol for small molecules	$\approx$ 1 kcal/mol
Stability	Usually guaranteed	Not guaranteed	Encoded
Topology (Sec. VIII A)	Usually required	Usually not required	Optional
Force differentiation	Analytical	Autograd	Analytical
Long-range interactions	Modeled	Usually ignored	Modeled self-consistently
Parameterization	Traditionally human-derived	Automated	Automated
Customization	Difficult	Trivial	Trivial
Platform	Specialized	Tensor-accelerating frameworks	T.B.D. (Sec. VIII D)

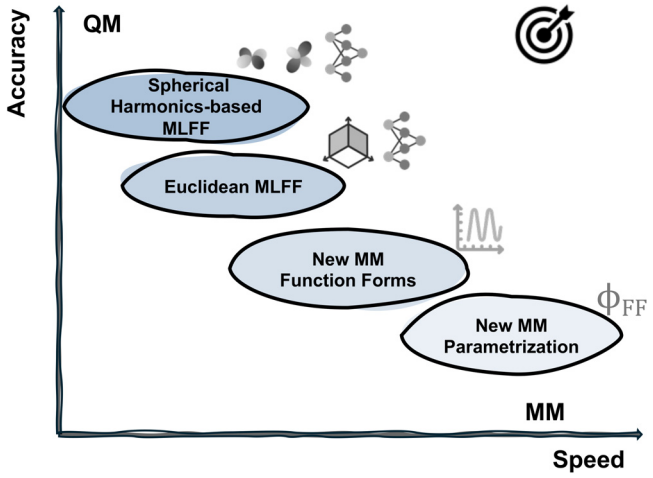

FIG. 1. Overview of the design space between molecular mechanics (MM) and machine learning (ML) force fields.

performant models,[37–39] and data-intensive models such as AlphaFold3[40] do not hard-code such symmetries.

Formally, we use *equivariance* and *invariance* to describe the symmetry in functional space.

Definition 1 (Equivariance and invariance).

A function $f : \mathscr{X} \to \mathscr{Y}$ is said to be *equivariant* to a symmetry group G if

$$f(T_g(\mathbf{x})) = S_g(f(\mathbf{x})), \tag{3}$$

and *invariant* if

$$f(T_g(\mathbf{x})) = f(\mathbf{x}), \tag{4}$$

with $\forall g \in G$ and some equivalent transformations on the two spaces, respectively, $T_g : \mathscr{X} \to \mathscr{X}$ and $S_g : \mathscr{Y} \to \mathscr{Y}$.

In the three-dimensional space we happen to dwell in, we consider functions with input space being $\mathscr{X} = \mathbb{R}^3$, and output being either $\mathscr{Y} = \mathbb{R}^3$ (for equivariant functions such as atomic forces) or scalar $\mathscr{Y} = \mathbb{R}^1$. For energies and forces, we can model transformations such as the rotations ($T_g(\mathbf{x}) = \mathbf{x}R$, where $R \in \mathbb{R}^{3\times3}$ is a rotation matrix $RR^T = I$), translations ($T_g(\mathbf{x}) = \mathbf{x} + \Delta\mathbf{x}$), and reflections ($T_g(\mathbf{x}) = \mathrm{Ref}_\theta + (\mathbf{x})$). These hold since the coordinate system on which the Euclidean coordinates are based is artificial and arbitrary. Similarly, since the indexing of the particles is arbitrary, permutation equivariance and invariance are also crucial: $T_g(\mathbf{x}) = S_g(\mathbf{x}) = P\mathbf{x}$, where P is the associated permutation matrix, here we omit that the particle identity representation should also permutate with P.

B. Linear runtime complexity

One of the key applications of force fields in biomolecular simulations is modeling the energy landscapes of heterogeneous biomolecules, such as protein-ligand complexes. In such cases, the number of atoms of such systems can significantly exceed 10^3, making the training and inference unattainably expensive if the runtime complexity is anywhere higher than quadratic $\mathscr{O}(N^2)$. At first glance, it might not be intuitive that the popular force fields, including MM [Eq. (9)], which have pairwise interactions, scale linearly. In practice, however, not all N^2 pairwise interactions are present in the energy function, as a *cutoff* function is almost always employed to mask off any interaction between two particles with coordinates $\mathbf{x}_{1,2}$ with a distance longer than a threshold L

$$\lambda(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} \Lambda(||\mathbf{x}_1 - \mathbf{x}_2||), & ||\mathbf{x}_1 - \mathbf{x}_2|| < L, \\ 0, & \texttt{elsewhere}, \end{cases} \tag{5}$$

with smooth boundary condition $\lim_{||\mathbf{x}_1-\mathbf{x}_2||\to L} \Lambda = 0$. For ML force fields, it is typically implemented in the form of a cosine annealing, such as $\Lambda = \cos(||\mathbf{x}_1 - \mathbf{x}_2||\pi/2L)$; for MM force fields, a reaction field method[41,42] for Coulomb interactions, and a polynomial switching function for van der Waals interactions are most typical. Meanwhile, particle mesh Ewald[43] approximation can ensure linear runtime complexity even without specifying cutoffs. This goes hand-in-hand with the assumption of constant sparsity: on average, the number of interactions that fall within the cutoff boundary for each particle stays constant. (Note that this is a strong assumption that would not stand without prior knowledge, i.e., $p(x)$ adopts a uniform distribution.)

C. Energy conservation

Since the potential energy $U(x)$ is a state function, there is zero work associated with moving the system through a trajectory starting and ending at the same place[22]

$$W \equiv \oint_C \vec{F} \cdot \mathrm{d}\vec{r} = 0. \tag{6}$$

This is satisfied if the force is always implemented as the negative gradient of the energy

$$F = -\nabla_x U(\mathbf{x}). \tag{7}$$

Numerically, Eq. (6) always stands accurately for simple functional forms such as an MM force field but less so for ML force fields with sophisticated functional forms.

D. Differentiability

In the previous paragraph, we have discussed the differentiation of the energy $\hat{U}(\mathbf{x}; h, \Phi)$ w.r.t. $\mathbf{x}$, which yields the force prediction. In MLFF, to ensure that at least the first- and second-order derivatives are smooth and further differentiable, neural activation function choices are limited to those with continuous first- and second-order derivatives, such as the ELU[44]-family. Meanwhile, if one wishes to optimize the MM force field parameters (see Sec. VIII D) Φ_{FF}, it is also crucial to evaluate the parameter derivative $\partial\hat{U}(\mathbf{x}; h, \Phi_{\mathrm{FF}})/\partial\Phi_{\mathrm{FF}}$.

E. Universality

Typically, there is little prior knowledge to be encoded in the inductive bias about the *ab initio* energy landscape except for the fact that when particles are separated far enough, the energy approaches zero. As such, at least in a limited region, the force field should be expressive enough to represent a diverse set of functions, or ideally *universal*, modulo the group invariance.

Definition 2 (Universality of invariant functions).

A parametrized invariant function $f(\cdot;\Phi)$ is said to be universal w.r.t. a symmetry group G on a space $\mathscr{X}$ if for all transformations in that symmetry group $\forall T_g \in G$, and for all invariant functions h satisfying $h(T_g(\mathbf{x})) = h(\mathbf{x})$, there exists a set of parameters Φ such that f can be approximated arbitrarily well

$$||h(\mathbf{x}) - f(\mathbf{x},\Phi)|| < \varepsilon, \quad \forall \mathbf{x} \in \mathscr{X}, \tag{8}$$

for arbitrarily small $\varepsilon > 0$.

In other words, such functions can approximate any invariant functions arbitrarily well, enabling the modeling of flexible energy landscapes.

F. Stability

Fu *et al.*[45] have illustrated that for many MLFF models, the simulation would collapse into unphysical regions after a certain number of integration steps. From a data-centric perspective, this largely results from not having substantial high-energy samples in training. Tackling this problem with sampling[46] on the high-energy regions, possibly with active learning,[25,47,48] is a data-driven way to counteract this instability, although this might lead to significantly more data as the high-energy regions (in a Tolstoyan way) are much more diverse. On the other hand, MM force fields are almost intrinsically stable since the functional forms are designed to be restrictive. The balance between expressiveness and stability resembles that between bias and variance, and given the same amount of data, can only be achieved with carefully designed inductive biases.

G. How fast is fast enough?

To be able to simulate 1 ms per day[15] of macromolecular behavior is a target that modern GPU-accelerated MM frameworks strive to achieve. As such, mechanisms such as protein folding, whose timescale is milliseconds,[49–53] can be simulated within a reasonable time frame. For small molecules, current ML force fields are about 10^2–10^3 slower than MM.[31,54] Since they both scale linearly, it is not unreasonable to estimate that this ratio is similar for large systems.

When it comes to using these force fields to simulate biomolecular systems in a drug discovery setting, it is also worth paying attention to the economics of simulations.[55] Since the cost of one GPU hour is around $1, whereas the cheapest wet lab assay, in a high-throughput screening setting, might cost less than $100, it is highly likely that to reliably estimate an observable *in silico* using ML force fields might cost significantly more than measuring it in a wet lab, provided that the acquisition of the molecule and the biological system is not extraordinarily expensive.

Finally, it is worth noting that most computing resources are spent on *force* rather than *energy* evaluations for both MM and ML force fields, and *energy* evaluations can even be skipped for intermediate steps. With MM, however, the forces are computed with analytical gradients rather than automatic differentiation,[56] which can potentially be more efficient. See more discussion in Sec. VIII D.

H. How accurate is accurate enough?

The *chemical accuracy* of 1 kcal/mol has long been an empirical standard beyond which the estimation of key chemical properties consistent with experiments is qualitatively possible. In a Nobel lecture,[57] *Pople* set this threshold as the target accuracy for the estimation of formation and ionization potentials. We also note that the discrepancies among various levels of theory are higher than this threshold, especially in high-energy regions. Moreover, the error of density functional theory (DFT) calculations also routinely exceed this number.[58]

For biomolecular applications, if one were to apply a force field to calculate the binding free energies of a protein-ligand system, Mobley and Klimovich[59] have demonstrated that a 1 kcal/mol noise in the *free energy* would roughly translate to making a drug discovery campaign five times faster if the goal is to increase the association constant K_a by tenfold (admittedly, the noise in the free energy is not the same as the noise in the potential energy, but we provide this finding here as an empirical gauge).

In some limited chemical spaces, the state-of-the-art ML potentials have long surpassed this threshold. Take the benchmark study[29,60] on (r)MD17,[22,61] for example, almost all competitive models have error well below 10^{-1} kcal/mol on all molecules, and the differences among the top three models usually come down to the scale of 10^{-2} kcal/mol. On a more diverse dataset containing organic small molecules,[62,63] to achieve accuracy significantly higher than the quantum chemical threshold of 1 kcal/mol has long been possible.[64,65] See Poltavsky *et al.*[66] for a large-scale benchmark on biomolecular and material systems.

III. MOLECULAR MECHANICS FORCE FIELDS: SIMPLE, CRUDE, BUT PRACTICAL FOR WIDE APPLICATIONS

Molecular mechanics (MMs) force fields[67,68] are empirical models representing atomic point masses interacting through parametrized local or pairwise functions of atomic coordinates. These functions characterize the potential energy of a system via valence (bond, angle, and torsion) and nonbonded terms, typically expressed as the sum of polynomials and truncated Fourier series. Here, we briefly review the functional form used in MM force fields, to demonstrate one extreme of the speed-accuracy trade-off. The most popular and widely used MM force field in biomolecular modeling and simulation is the class I MM force field, primarily due to its computational efficiency, arising from its simplified functional form, which can be typically expressed as

$$\begin{aligned} U_{\mathrm{MM}}(\mathbf{x};\Phi_{\mathrm{FF}}) = &\sum_{\text{bond}} \frac{K_r}{2}(r_{ij}-r_0)^2 + \sum_{\text{angle}} \frac{K_\theta}{2}(\theta_{ijk}-\theta_0)^2 \\ &+ \sum_{\text{torsion}} \sum_{n=1}^{n_{\max}} K_{\phi,n}\left[1+\cos(n\phi_{ijkl}-\phi_0)\right] \\ &+ \sum_{\text{Coulomb}} \frac{1}{4\pi\varepsilon_0}\frac{q_i\,q_j}{r_{ij}} \\ &+ \sum_{\text{van der Waals}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \end{aligned} \tag{9}$$

where the total potential energy U_{MM} as a function of the coordinates of the system $\mathbf{x}$ and the collection of force field parameters (also see Sec. VI B) $\Phi_{\mathrm{FF}} = \{K_r, K_\theta, r_0, \theta_0, K_{\phi,n}, \phi_0, q, \sigma, \varepsilon\}_i$ is modeled as the sum of bond, angle, torsion, and nonbonded energy. The bond, angle, and torsion force constants are represented as K_r, K_θ, and $K_{\phi,n}$, with their equilibrium values and phases denoted as r_0, θ_0, and ϕ_0, respectively. The atomic point charges are represented by q, while ε and radii σ parametrize the Lennard–Jones energy well. $r_{i,j}$, $\theta_{i,j,k}$, $\phi_{i,j,k,l}$—which represents the distance between covalently bonded atoms i, j, the angle among $i-j-k$, and the dihedral angle between the planes formed by

i, j, k and j, k, l, respectively—are simple functions extracted from the coordinates $\mathbf{x}_{i,j,k,l}$. In practice, such operations can be implemented as

$$\begin{aligned} r_{i,j} &= ||\vec{\mathbf{x}}_{ij}|| = ||\mathbf{x}_i - \mathbf{x}_j||; \\ \theta_{i,j,k} &= \text{atan2}\left(||\vec{\mathbf{x}}_{ij} \times \vec{\mathbf{x}}_{jk}||, \vec{\mathbf{x}}_{ij} \cdot \vec{\mathbf{x}}_{jk}\right); \\ \phi_{i,j,k,l} &= \text{atan2}\Big(\left((\vec{\mathbf{x}}_{ij} \times \vec{\mathbf{x}}_{jk}) \times (\vec{\mathbf{x}}_{jk} \times \vec{\mathbf{x}}_{kl})\right) \cdot \frac{\vec{\mathbf{x}}_{jk}}{||\vec{\mathbf{x}}_{jk}||}, \\ &\quad \left((\vec{\mathbf{x}}_{ij} \times \vec{\mathbf{x}}_{jk}) \cdot (\vec{\mathbf{x}}_{jk} \times \vec{\mathbf{x}}_{kl})\right)\Big). \end{aligned} \tag{10}$$

An out-of-plane term, known as *improper torsion*, can also be introduced with the torsion term to enhance the molecular planarity and prevent chiral inversions. In theory, multipole expansion—such as dipole and quadrupole moments—are necessary to accurately represent the quantum mechanical electrostatic potential. However, empirical force fields try to approximate this multipole expansion by assigning point charges localized at the nuclei of atoms, sometimes with virtual sites to model lone pairs and σ-holes,[69–71] to reproduce the same electrostatic potential that would be given by the true electronic structure and electron density distribution. The van der Waals interaction combines repulsive and attractive forces, typically represented by a 12–6 Lennard–Jones potential. The Lorentz–Berthelot[72] combining rules can be employed to determine σ and ε between different atom types, though alternative combination rules are possible.[73] The r^{-12} and r^{-6} terms represent London dispersion interactions, with r^{-12} also accounting for the short-range repulsion due to Pauli exclusion, preventing atom collapse.

This subsection and Table II have surveyed only the simplest and most traditional MM force fields, the efforts to enhance the expressiveness via additional terms, which is highly non-trivial, are discussed in Sec. VIII B.

IV. THE TOOLBOX OF COMPOSABLE OPERATIONS FOR MACHINE LEARNING FORCE FIELDS

Having reviewed MM force fields in Sec. III, we now direct our attention to the MLFFs, which have currently focused more on accuracy rather than speed. While there has been a plethora of the previous work focused on developing and applying system-specific MLFFs to homogeneous media and benchmarking various MLFF architectures on datasets of small biologically relevant organic molecules,[22,28,32,87–91] there have been relatively few applications to date employing MLFFs in lieu of traditional MM models for extended MD simulations of large biomolecular systems.

Part of the challenge in developing MLFFs for large biomolecular systems stems from the computational challenge of constructing datasets of *ab initio* energies and forces for these large heterogeneous systems. Unlike homogeneous systems such as liquid water where one

TABLE II. Representative intramolecular and intermolecular potentials.[74] r_0, θ_0, and ϕ_0 represent equilibrium bond distance, bond angle, and dihedral angle, respectively.

Potential	Expression	Description
Bond		
Harmonic potential	$U(r) = K(r - r_0)^2$	K: spring constant
Morse potential[75]	$U(r) = D(1 - e^{-\alpha(r-r_0)})^2$	D: well depth, α: width of the potential
Angle		
Harmonic potential	$U(\theta) = K(\theta - \theta_0)^2$	K: bending constant
CHARMM potential[76]	$U(\theta) = K(\theta - \theta_0)^2 + K_{UB}(r - r_{UB})^2$	K_{UB}: constant for Urey-Bradley term r_{UB}: equilibrium bond length between first and third atoms
Cosine squared potential[77]	$U(\theta) = \frac{1}{2}K[\cos(\theta) - \cos(\theta_0)]^2$	K: bending constant
Torsion		
Cosine potential	$U(\phi) = K[1 + \cos(n\phi - \phi_0)]$	K: bending constant n: multiplicity representing the periodicity
OPLS potential[78]	$U(\phi) = K_1[1 + \cos(\phi)] + \frac{1}{2}K_2[1 - \cos(2\phi)] + \frac{1}{2}K_3[1 + \cos(3\phi)] + \frac{1}{2}K_4[1 - \cos(4\phi)]$	K_1, K_2, K_3, K_4: bending constants
van der Waals		
Lennard-Jones (12-6) potential[79]	$U(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$	ε: well depth, σ: distance when $U(r) = 0$
Buckingham potential[80,81]	$U(r) = Ae^{-\frac{r}{\rho}} - \frac{C}{r^6}$	A and C: well depth, ρ: width of the potential
Morse potential[75,82,83]	$U(r) = D[e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)}]$	D: well depth, α: reciprocal length
Double exponential potential[84–86]	$U(r) = \varepsilon\left[\frac{\beta e^{\alpha}}{\alpha - \beta}\exp\left(-\alpha\frac{r}{r_0}\right) - \frac{\alpha e^{\beta}}{\alpha - \beta}\exp\left(-\beta\frac{r}{r_0}\right)\right]$	ε: well depth α and β: steepness of the repulsive and attractive interactions, respectively

can simply train a MLFF on smaller periodic boxes and clusters of water and readily apply the trained MLFF to simulations for larger periodic boxes,[92] the chemical heterogeneity and the importance of long-range interactions in proteins[93] renders their decomposition into smaller, computationally more tractable training structures less straightforward. Alternatively, one could simply construct datasets encompassing the entirety of large biomolecular systems using energies and forces obtained from QM/MM evaluations. However, in addition to the QM/MM evaluations still potentially being expensive depending on the target QM method and size of the QM region, expending training resources on such large structures where most of the interactions are probably described purely at the MM-level seems wasteful.

Recently, several MLFFs trained to diverse datasets encompassing tens of thousands of molecules (see Table III and Sec. V A) have demonstrated the potential of bottom-up approaches to developing general-purpose MLFFs that can be readily applied to large biomolecules not wholly represented in the respective training sets.[24,63,64,94–98] Some of the earlier general-purpose MLFFs were limited in applicability due to their training set and model architectures not explicitly accounting for charged species,[24,63] but were still usable for conducting stable MD simulations for smaller protein systems[99] or in mixed-level ML/MM simulations.[100–105] However, recently, general-purpose MLFFs based on the GEM[90,98] model have been successfully used to generate MD trajectories of the solvated crambin protein (18 000–25 000 atoms) for which the simulated THz-region vibrational modes, characteristic of slow collective protein motions, seem to show better agreement with the experimental spectrum than what is obtained using AmberFF.[64] Perhaps the largest demonstration to date of applying a general-purpose MLFF is the use of an Allegro[32] model, trained to the SPICE dataset,[62] for an all-atom MD simulation of the solvated HIV capsid consisting of 44- $\times 10^6$ atoms that achieves 8.7 timesteps/s when employing 5120 A100 GPUs.[97]

On the other hand, system-specific bottom-up approaches leveraging molecular fragmentation schemes, specifically electrostatically embedded generalized molecular fractionation with conjugate caps (EE-GMFCCs),[106–108] generalized energy-based fragmentation (GEBF),[109] and residue-based systematic molecular fragmentation (rSMF),[110] have also been used to develop datasets to train MLFFs for specific proteins.[30,99,110–112] As a proof of concept, Cheng *et al.*[99] demonstrate that the GEBF method can be applied to specific proteins (1XQ8, 1013 atoms) to create relatively small datasets of fragments (5020 configurations across 65 fragments) that can be used to train accurate MLFFs for MD simulations. Unsurprisingly, their MLFFs trained to this dataset give more accurate energy and force predictions, as compared to the reference method (DFT using the ω B97X-D functional), for the target protein system than the general-purpose ANI-1x.[24]

Here, we briefly review the composable operations frequently used in the construction of MLFF through the lens of the desired properties discussed in Sec. II.

A. Atomistic decomposition of energy

Most, if not all MLFF models decompose the total energy into a sum of *per-atom* energies

$$U = \sum_i U_i, \tag{11}$$

where each U_i is a function of the local atomic environments within a spatial cutoff radius. Although widely adopted in MLFFs, this approach lacks a rigorous physical foundation, as the concept of

TABLE III. Popular datasets used to curate force fields. For a more comprehensive list, see Ullah and Dral.[142]

Dataset	Elements	Chemical spaces	# Molecules	# Conformers	Level of theory	Sampling
ine MD17[22,61]	C,H,O	Small molecules up to 21 atoms	10	2.7 m	DFT	Path-integral MD
MD22[87]	C, H, O, N	Small molecule w/ 42–270 atoms	7	22 k	PBE+MBD[143,144]	MD
QM9[145]	C, H, O, N, F	Small molecule	134 k	134 k	B3LYP/6-31G(2df,p)	Minimized
QM7-X[146]	C, H, N, O, S, Cl	Small molecule	7 k	42M	PBE0+MBD[143,144]	DFT-optimized + normal mode
ANI-1[147]	C, N, O, F	Small molecules up to 11 heavy atoms	57 462	24 687 809	DFT	DFT-optimized + normal mode
ANI-2[63]	H, C, N, O, S, F, Cl	Small molecules	13 405	4 695 707	wB97X/631Gd	Dimer, normal mode, torsion sampling, and active learning
OrbNet Denali[148,149]	H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, I	Organic molecules	15 000	2.3×10^6	DFT	Normal mode + MD
SPICE[62]	Li, C, N, O, F, Na, Mg, P, S, Cl, K, Ca, Br, I	Small molecules peptides, ion pairs	19 k	1.1 m	ω B97M-D3(BJ) /def2-TZVPPD	MD + Minimization
ine						

per-atom energy is not well-defined in many-body quantum systems. Nonetheless, this treatment is consistent with the graph-level readout that graph neural networks (GNNs) use for graph-level regression and is trivial to implement in machine learning programs.

B. Equivariant and invariant features

It has been outlined in Sec. II that an ideal force field $\hat{U}(\mathbf{x})$ should be an $E(3)$-invariant function of $\mathbf{x}$. It has been empirically illustrated in Unke and Meuwly,[89] Thölke and Fabritiis,[113] Schütt *et al.*,[114] Wang and Chodera,[31] Unke *et al.*,[90] and Frank *et al.*[115]; however, that intermediate *equivariant* representations can boost the expressiveness and performance of invariant models. Concretely, w.r.t. a group G, an invariant function $y = f(\mathbf{x})$ can be constructed as the composition of an invariant function g_{I} and an equivariant function g_{E} as

$$y = f(\mathbf{x}) = g_{\mathrm{I}}(g_{\mathrm{E}}(\mathbf{x})). \tag{12}$$

The equivariant-invariant mapping g_{I} is also called a *scalarization*. In practice, a modern MLFF usually keeps track of both invariant and equivariant features and updates them simultaneously, somewhat analogous to the discretized integration of speed and position in MCMC. So the functional signature of a G-equivariant MLFF layer operating on both invariant (containing information such as the atomic identity and *chemical* environment of the particle) $h \in \mathscr{H}$ and equivariant (containing information such as the *geometric* environment of the particle) $\mathbf{x} \in \mathscr{X}$ (w.r.t. the same space) features $f : \mathscr{H} \oplus \mathscr{X} \to \mathscr{H} \oplus \mathscr{X}$ can be written as

$$\mathbf{x}', h' = f(\mathbf{x}, h), \tag{13}$$

where we have

$$T_g(\mathbf{x}'), h' = f(T_g(\mathbf{x}), h), \quad \forall T_g \in G. \tag{14}$$

C. Message passing

Following the *atomistic decomposition* scheme [Eq. (11)] outlined above, atoms would need to exchange information among themselves to come up with meaningful geometric and semantic representations. The message passing steps in graph neural networks (GNNs), which have originally been designed for social modeling, become the go-to choice to satisfy this requirement. Most *spatial* GNNs[116–126] on topological graphs (without geometry features) adopt a message passing framework. Following the framework from Xu *et al.*[119] and Battaglia *et al.*,[120] the kth layer of a GNN could be written as two steps—*neighborhood aggregation*

$$a_v^{(k)} = \rho^{(k)}(h_u^{(k-1)}, u \in \mathscr{N}(v)), \tag{15}$$

and *node update*

$$h_v^{(k)} = \phi^{(k)}(h_v^{(k-1)}, a_v^{(k)}), \tag{16}$$

where h_v^k is the feature of node v at kth layer, $h_v^0 = \mathbf{x}_v$ and $\mathscr{N}(\cdot)$ denotes the operation to return the multiset of neighbors of a node. More concisely, omitting the nonlinear transformation step ϕ ubiquitous in all neural models and assuming a convolutional *aggregate* function $\rho = \mathrm{SUM}$, or $\rho = \mathrm{MEAN}$, a graph neural network layer is characterized by the aggregation/convolution operation that pools representations from neighboring nodes, forming an intermediary representation $\mathbf{X}'$, which on a global level, with activation function σ and weights W, can be written as

$$\mathbf{X}' = \sigma(\hat{A}\mathbf{X}W). \tag{17}$$

The difference among GNN architectures, apart from the subsequent treatment of the resulting intermediate representation $\mathbf{X}'$, typically amounts to the choices of transformations $(\hat{A})$ of the original adjacency matrix (A)—the normalized Laplacian for graph convolutional networks (GCNs),[117] a learned, sparse stochastic matrix for graph attention networks (GATs),[127] powers of the graph Laplacian for simplifying graph networks (SGNs),[128] and the matrix exponential thereof for graph neural diffusion (GRAND).[129] To expand these frameworks to incorporate geometry information, the layers to incorporate coordinates surveyed in this section can be plugged into Eqs. (15) and (16). We also note that the convolution operator in Eq. (17) is the root of a plethora of performance pathologies including over-smoothing,[130,131] over-squashing,[132,133] and limited expressiveness.[119,134,135] Therefore, alternative forms of GNNs[136] might be needed to address them to make them more expressive and robust.

D. Gaussian smearing

Distances among particles are arguably the simplest and most crucial invariant feature—with a distance matrix, the coordinates of the particles can be reconstructed modulo the $E(3)$-equivariance, and all invariant functions can be approximated arbitrarily well. Nevertheless, to naïvely throw the distances into a neural network yields only highly correlated representations, since a neural network after initialization is close to linear, as discussed in detail in Schütt *et al.*[137] Since Behler and Parrinello,[36] the Gaussian-smeared distances have been used widely as a radial, invariant feature. The most generic form of Gaussian smearing resembles the radial basis function (RBF) kernel

$$K(x, x') = \exp\left(-\frac{\| x - x' \|^2}{2\sigma^2}\right). \tag{18}$$

By carefully choosing a set of evenly spaced $x'_0, x'_1, \ldots, x'_i, \ldots x'_k$, Eq. (18) yields a k-dimensional vector with maximal convoluted signal at x'_i, resolution defined by $x'_{i+1} - x'_i$, and sensitivity controlled by σ.

E. Angular symmetry function

Also proposed in Behler and Parrinello,[36] the angular symmetry function takes the angles among triplets of atoms [which can be calculated using Eq. (10)] directly in a neural network. This term closely resembles how MM encodes angular environment, albeit with a more expressive function. Many popular MLFFs, most notably Smith, Isayev, and Roitberg,[24] incorporate this feature locally to account for angular environments.

F. Linear combination of equivariant features

Ignoring the translational equivariance again and considering only the $O(3)$ (or $O(n)$) group, the linear combination of equivariant features is naturally equivariant, and the linear combination of equivariant *maps* $(\{f_i\})$

$$F(\mathbf{x}) = \sum_i \lambda_i f_i(\mathbf{x}), \tag{19}$$

where λ_i are constants, are also equivariant. In practice, this summation can be implemented as a linear transform, or a single-layer neural network without biases or activation, operated on the last dimension of a $\mathbb{R}^{N,3,D}$ vector representation. Moreover, the coefficients $\{\lambda_i\}$ can also be calculated from the invariant representations.

G. Spherical harmonics

Batzner *et al.*,[29] Musaelian *et al.*,[32] Batatia *et al.*,[60] and Thomas *et al.*[138] construct $SO(3)$-equivariant (removing both the translation and reflection transformations from $E(3)$) convolution filters to be products of radial (invariant) functions and spherical harmonics

$$F(\vec{r}) = R(r) Y_m^{(l)}(\vec{r}), \tag{20}$$

which can be contracted with the Clebsch–Gordan coefficients to form higher-order tensor products. Currently, the most accurate MLFFs mostly rely on these representations, although it has been outlined in Lemma 3 that they are not required for universality.

H. Dot product scalarization

Finally, we review what we believe would be one of the most crucial building blocks for the class of fast and accurate force fields we plan to construct—the dot product scalarization, $V^T V$, which reduces equivariant representations to invariant representations, showing the same *functional signature* as the force field in general,.

We demonstrate that this operation is highly expressive, and universal in cases. If we center our view around each particle and do not consider the translation invariance, the group we consider would be $O(3)$ rather than $E(3)$, and all $O(3)$-invariant functions can be universally approximated using the dot products among the inputs. In such case, we realize that all functions can be universally approximated by dot products among the edge vectors, or formally.

Lemma 3 (First Fundamental Theorem[139] for $O(n)$).

If f is an $O(n)$-invariant scalar function of vector inputs $v_1, \ldots, v_n \in \mathbb{R}^D$, then $f(v_1, v_2, \ldots, v_n)$ can be written as a function of only the scalar products of the v_i. That is, there is a function $g(\cdot)$ such that

$$f(v_1, v_2, \ldots, v_n) = g(V^\top V) = g((v_i^\top v_j)_{i,j=1}^n). \tag{21}$$

This lemma is the theoretical cornerstone of a number of locally universal (i.e., each particle being able to realize its own atomic environment even without message-passing; see Wang and Chodera[31]) MLFF models.[31,89,113,114] It is worth mentioning, however, that function g in Lemma 3 is not necessarily permutation invariant, and extra care is needed to design $O(3)$-invariant, permutationally invariant, and universal functions. Wang and Chodera,[31] for example, neurally parametrizes a series of edge vectors prior to the dot product. In terms of the engineering, we also remark that thanks to the popularity of attention[140]-based architectures, dot product operations have been drastically accelerated, and parallelized, flexible implementations are also available via operations such as Einstein summation (`einsum`).[141]

V. BEST PRACTICES AND PITFALLS FOR CURATING MM AND ML FORCE FIELDS

Thus far, we have been discussing the functional form of force fields $\hat{U}(\mathbf{x}; h, \Phi)$, but to design a practical $\hat{U}$, one needs to find a set of optimal parameters by maximizing the likelihood of some particular force and energy data. Hoping to inspire the community to converge on a set of standard protocols for designing and producing force fields, in this section, we briefly review the popular datasets and common procedures used in the curation of MM and ML force fields and the best practices for training and evaluation.

A. Datasets

1. A note on the QM target

There has been little consensus, when it comes to biomolecular applications, regarding which QM levels of theory might correspond best to experimental measurements.[150] As a result, popular datasets (see Table III) are curated with various levels of theory and sampling strategies, making the merging of the data difficult, demanding meta-learning solutions.[151] Before further evidence would emerge, if the field were to agree on a single level of theory for a community-wide effort to push for high-quality, high-volume data for MLFF training, cheaper targets might be more appealing.

When developing novel MLFFs, it is also worth reminding ourselves that all we are trying to fit is a known function that can be solved analytically. One can view the QM energy function not as a reservoir of data but a *surrogate*, from which repeated acquisition is possible. Active learning techniques, such as Smith *et al.*,[25] Wang *et al.*,[47] and Schwalbe-Koda *et al.*,[48] present a useful avenue to gather data in a rationally parsimonious fashion.

2. Chemical and conformational diversity

Conceptually, to accurately fit energies and forces on *all* chemical spaces is no different than having the MLFF model able to solve Schrödinger's equation, which seems impossible. To this end, QM datasets are always curated with biases, in terms of the coverage on chemical spaces and conformational landscape.

Within the chemical space of (bio)organic molecules, MD17[22] is among the most popular datasets used for MLFF benchmark. Consisting of DFT calculations of ten small organic molecules (benzene, uracil, naphthalene, aspirin, salicylic acid, malonaldehyde, ethanol, toluene, paracetamol, and azobenzene) with fewer than 1×10^6 snapshots each, it is limited in chemical diversity. Christensen and von Lilienfeld[152] further revised this dataset to reduce the noise. There are also high-quality QM datasets focusing on single molecules such as 3BPA[153] and AcAc.[154] ISO17,[137] on the other hand, is slightly more chemically diverse as it samples 129 isomers of $C_7O_2H_{10}$ with 5000 snapshots each. In contrast, QM9,[145] like its predecessors, is rich in chemical diversity but not conformational diversity—it contains more than 134 k small molecules, although all in low-energy state.

Moving on to larger datasets emphasizing *utility* beyond just proof-of-concept, ANI1[147] and ANI2[27] dataset, containing 20×10^6 off-equilibrium snapshots, is among the most popular datasets that are simultaneously chemically and conformationally diverse, albeit no QM forces were annotated, which would be more information-rich than energies. SPICE[62,155] also displays a vast chemical and conformational

diversity on 2×10^6 conformations of small molecules and peptides, with forces annotated. The SPICE[62] dataset, along with many diverse datasets for (bio)molecules, has been generated on the QCFractal platform,[156] which is being used to actively run QM calculations to curate the next generation of datasets for MLFF fitting.

Another community-driven project, ColabFit[157] Exchange curates an open-source, diverse database containing a large collection of systematically organized datasets from multiple chemistry/materials domains that are especially designed for ML atomistic model development (providing physical units standardization, a unified data standard, integrated data loaders, etc.). To date, the database contains 372 datasets containing more than 180M atomic configurations and 500M properties, spanning over 100 thousand different chemistries, and it is constantly expanding.

In principle, topology-free MLFFs are more naturally suited for simulating chemical bond-breaking and forming events than standard FFs, and would enable the simulation of fundamental biochemical processes like enzymatic reactions for which *ab initio* accuracy and FF efficiencies would ideally be employed. However, in practice, the curation of datasets to train MLFFs to accurately model chemical reactions presents additional challenges,[158] the same age-old challenges associated with sampling rare events, as compared to developing models for equilibrium sampling of nonreactive systems. To efficiently sample higher-energy training configurations representative of reaction pathways, various techniques have been employed ranging from simply running unbiased MD simulations at elevated temperatures[159,160] and identifying minimum energy paths[161–163] to other approaches incorporating biased MD simulations leveraging enhanced sampling algorithms in active learning protocols.[164–167]

Finally, we note that training on condensed-phase properties has only been very recently[168] possible. MM force fields are, on the other hand, usually successful in faithfully reproducing these properties.[169] These largely depend upon the *inter*molecular interactions, whereas long-range interactions are traditionally neglected in MLFFs (prior to SPICE,[62] the datasets used to curate MLFFs only include intramolecular interactions).

B. Training and evaluation

We observe that, in recent literature on MLFFs, the community is converging on a set of practices for efficient training and fair comparison of MLFF models. Certain practices, such as extremely small batch size, exponentially decaying learning rate, and parsimonious use of normalization, are common in the curation of highly performant MLFFs. Since the forces and energies, and thereby force errors $\mathscr{L}_F$ and energy errors $\mathscr{L}_U$, for instance measured in mean-squared-error, are of different units, one has to apply a set of perhaps physically meaningless constants to combine them as the loss function $\mathscr{L}$

$$\mathscr{L}_U = ||\hat{U} - U||^2, \tag{22}$$

$$\mathscr{L}_F = \frac{1}{3N}\sum_{i=1}^{N}\sum_{\alpha=1}^{3}\left\| -\frac{\partial \hat{U}}{\partial r_{i,\alpha}} - F_{i,\alpha}\right\|^2, \tag{23}$$

$$\mathscr{L} = \lambda_U \mathscr{L}_U + \lambda_F \mathscr{L}_F. \tag{24}$$

Empirically, when using atomic units, $\mathscr{L}_F/\mathscr{L}_U$ within the range 100–1000 usually yields the best results.

It is worth recalling that, while the MSE [Eq. (23)] and RMSE error on forces are $E(3)$-invariant, the MAE loss on forces is *not*, and is dependent upon the choice of the coordinate systems

$$\mathscr{L}_{\mathrm{F,MAE}} = \frac{1}{3N}\sum_{i=1}^{N}\sum_{\alpha=1}^{3}\left| -\frac{\partial \hat{U}}{\partial r_{i,\alpha}} - F_{i,\alpha}\right|. \tag{25}$$

In a sense, this error has a bias to favor the conformations more aligned with the axes. It is alarming to see that this biased and arguably erroneous metric has been used widely in both the training and evaluation stages of MLFF models incorporating force matching.

The aforementioned error is a typical example of how error-prone the implementation of MLFF modules is—some intuitively benign operations might break the symmetry without catching the eyes of a seasoned engineer-researcher. Thus, we recommend that an equivariance/invariance unit test be included in all modules of MLFF implementation. An example test suite for a function f that works on both $SE(3)$ invariant (h) and equivariant (x) representations [see Eq. (13)] can be implemented in five lines, using NumPy,[170] for example:

```
def test_equivariant_and_invariance(f, h, x):
  # random translation
  import numpy as np
  T = np.random.randn(1, 3)
  # random rotation
  R = np.linalg.qr(
     np.random.randn(3, 3))[0]
  # random transformation
  F = lambda x: x @ R + T
  # assert h changed
  # and x transformed accordingly
  h, x = f(h, x); h1, x1 = f(h, F(x))
  assert (
     np.allclose(h, h1)
     & np.allclose(F(x), x1)
  )
```

VI. MAKING MM MORE ACCURATE

Having reviewed the standard building blocks and procedures in constructing both MM and ML force fields, we start to march toward the desired spot on the design space (upper-right corner in Fig. 1) by looking at approaches to making MM more accurate and MLFFs faster.

A. Functional forms: More, but only slightly more, than harmonics and Fourier series

The choice and design of functional forms (see Table II) greatly limit the flexibility and expressiveness of the force field. For instance, the Lennard–Jones 12–6 potential, described in Eq. (9), was developed decades ago and has been widely adopted since, despite the existence of alternative approaches[73,85] which demonstrates greater flexibility or stability. Enriching the complexity of the MM functional form has long been an area of intensive research, though we also notice that the

reluctance toward the newer modalities can be attributed more toward the conservatism of this field rather than the inferior properties of the innovations.

Notably, Class II MM force fields[171–173] replace the harmonic terms in Eq. (9) with more much more flexible terms such as follows.

- *Higher order polynomials:* The harmonic terms for bonds and angles can be rewritten to incorporate higher k-order polynomials in the form of [reusing the notation from Eq. (9)]:

$$\sum_{\text{bond}} \sum_{k} K_k (r_{ij} - r_k)^k; \quad \sum_{\text{angle}} \sum_{k} K_k (\theta_{ij} - \theta_k)^k. \tag{26}$$

  Since the geometry of a system is uniquely defined by the interatomic distances, given sufficiently high-order polynomials and dense enough bond connections, this functional form can be made universal.

- *Coupling terms:* Terms like bond displacement, angle displacement, and Fourier series for (proper and improper) torsions can be combined multiplicatively to form bond-bond coupling (shown here as an example),

$$\sum_{\text{bond}} \sum_{\text{bond}} K_{\text{bond,bond}} (r_{i_0 j_0} - r_0)(r_{i_1 j_1} - r_1), \tag{27}$$

  angle-angle coupling, bond-angle coupling, bond-torsion coupling, angle-torsion coupling, or torsion-angle-angle coupling.

In a similar spirit, Xie *et al.*[174] have constructed ultra-fast machine learning potentials using B-splines and parametrizations using machine learning approaches. On a simple test system, this force field has a favorable trade-off between speed and accuracy and is guaranteed to be smooth.

On the other hand, polarizable force fields, such as Drude[175,176] and AMOEBA force fields,[177] incorporate the ability to dynamically adjust the distribution of atomic charges in response to the local electrostatic environment and generally involve a self-consistent calculation of induced dipoles. These can be augmented with neural networks to account for missing contributions.[178] Reactive force fields, such as ReaxFF,[179,180] can handle bond breaking and forming during simulations by dynamically updating bond orders, based upon interatomic distances, for every simulated MD frame and making the relevant energy terms dependent on those bond orders. Empirical valence bond (EVB) models[181] also enable the use of standard MM force fields for reactive simulations,[182–185] but require the parameterization of a coupling term between the reactant and product states for a particular system. These have traditionally been system-specific, but recent work has focused on developing transferable parametrizations of EVB models across different reactive systems.[186] However, improving force field accuracy by introducing more expressive and additional functional terms comes with increased computational costs, which may not justify the trade-off between accuracy and speed, and significantly increases the complexity of force field parameterization, which can be automated using approaches described in the following section.

B. Parameterization: From engineer-years to GPU-days

Another challenge in developing a reliable, robust, and extensible MM force field is the parameterization scheme—the assignment of the parameter set Φ_{FF}—which must ensure comprehensive chemical coverage across the broad and heterogeneous chemical space relevant to biomolecular systems.

The determination of Φ_{FF} has traditionally been reliant on a human labor-intensive, inflexible, and inextensible rule-based scheme named *atom typing*—it classifies atoms into discrete categories representing distinct chemical environments. This classification enables MM parameters to be subsequently assigned from a tabulated table of relevant atomic, bond, angle, and torsion parameters. For example, in the case of small molecules, atom types are determined by the attributes of the atom, such as element identity, hybridization, and aromaticity, as well as the attributes of the neighboring atoms and their connected bonds, and the number of neighboring atoms. For amino acids, the atom types are traditionally assigned according to the residues. Most of these atom types have a receptive field (the radius within which changes are reflected in the type) of two or three bonds, and chemical motifs outside this receptive field are generally not realized. After the atom types are determined, the bond, angle, and torsion types are determined simply by composing (using the operation and dictionary lookup) atom types, and as a result, K atom types can naively lead to K^4 torsion types without simplification. The van der Waals interactions, on the other hand, are usually described with Lennard–Jones 12–6 potentials using the Lorentz-Berthelot[72] combining rules to determine σ and ε between different atom types.

The force field parameters can be further optimized in a systematic manner using ensemble reweighting method[187–189] and machine-learning methods.[190,191] However, co-optimizing the discrete chemical perception defined by the rule-based atom types and continuous force field parameters remains intractable. In general, the force field accuracy is constrained by the resolution of chemical perception. Efforts to improve the accuracy by increasing the number of atom types lead to a combinatorial explosion of required types. Note that the substructure matching-based approach, such as Mobley *et al.*,[192] Wang *et al.*[193] do not suffer from such symptoms.

Although there are efforts to automate the development and parameterization process,[194–198] human expertise remains essential, introducing challenges in adjusting existing parameters to accommodate new ones, particularly when extending the force field to new chemical domains of interest. In addition, biomolecular systems are inherently heterogeneous, making MM force field optimization challenging. The popular AmberTools 23 package, for instance, combines independently developed force fields for chemical subspaces including proteins,[199] DNA,[200,201] RNA,[202] water,[203–205] monovalent[206,207] and divalent[208–210] counterions, lipids,[211] carbohydrates,[212] glycoconjugates,[213,214] small molecules,[215,216] post-translational modifications,[217] and nucleic acid modifications[218]—crystallized from more than 100 engineer-years of effort. Nevertheless, there is no guarantee that the optimized solution for each class will constitute the globally optimized solution.

1. Substructure pattern matching

Substructure pattern matching approaches[192,219,220] represent another class of force field parameterization schemes that focus more directly on chemical perception than the traditional atom-typing methods. The Open Force Field Initiative,[193] for instance, has developed an ecosystem of toolkits[221] and force field releases[20] that use standard SMARTS-based chemical substructure queries[192] to assign entire sets of valence parameters (atoms, bonds, angles, and torsions) in a

hierarchical manner. These approaches help mitigate the combinatorial explosion of parameters and significantly reduce atom type redundancy while maintaining the force field accuracy. For example, the latest Sage force field (openff-2.2.0[222]) from the Open Force Field Initiative contains 187 torsion parameters, approximately 800 times fewer than the OPLS3e force field,[223] which relies on atom-typing. Furthermore, the reduced complexity of substructure-based approaches facilitates the automated fitting of parameters for specific force fields, such as through the use of ForceBalance.[188] However, determining and refining substructure patterns for more reliable force fields still requires a human-in-the-loop approach, combining human expertise with automated[224] procedures.

2. Graph-based chemical perception: Molecular topology as a graph

Applying graph neural networks (GNNs, Sec. IV) for more robust and extensible MM force field parameterization is another emerging area.[21,191,225–232] For example, Wang *et al.*,[225] Takaba *et al.*[21] have demonstrated the ability to replace traditional rule-based discrete atom-typing schemes with continuous atomic representations generated by neural networks operating directly on chemical graphs using an end-to-end differentiable framework. The neural network parameters are optimized directly through standard machine learning frameworks to fit quantum chemical and/or experimental data. These approaches enable the co-optimization of chemical perceptions, represented as continuous atom embeddings, alongside continuous force field parameters. For example, the latest Espaloma force field,[21] trained in less than one GPU day on a vast chemical space (comprising 17 k molecules and over 1×10^6 QM snapshots) consisting of small molecules, peptides, proteins, and RNAs, has shown the capability to accurately predict not only energy and forces but also, when used in MD simulations, NMR observables, and protein-ligand binding free energies. This demonstrates a promising path forward for the flexible and efficient curation of molecular mechanics (MMs) force fields. Nevertheless, Leonard-Jones parameters are *not* learned in this framework, which is required to complete a force field.

VII. MAKING ML FORCE FIELDS FASTER

As of now, most of the fastest ML force fields use dot-product scalarization (Lemma 3) to compute the energies and automatic differentiation to compute the forces. Here, we review general strategies for accelerating MLFFs—both for accelerating spherical harmonics-based equivariant representations as well as making dot-product-based methods even faster.

A. Accelerating SO(3) convolutions

Although the spherical harmonics-based SO(3)-equivariant representations have been shown to be performant and data-efficient in constructing MLFFs,[29,32,60,138] an L-degree tensor convolution would naively require $\mathcal{O}(L^6)$ complexity. Passaro and Zitnick[233] address this issue by reducing SO(3) convolutions to SO(2) without losing information. Luo, Chen, and Krishnapriyan[234] also achieve such cubic complexity by performing the convolution in the Fourier space and using Gaunt coefficient rather than the Clebsch–Gordan coefficients. Cheng,[235] on the other hand, performs the *atomic cluster expansion*[153] directly in the Cartesian coordinate system. Frank *et al.*[115] uncouple geometric information from atomic features and conduct attention separately, achieving superior data and parameter efficiency. Interestingly, the works cited here have all been published in the years 2023 and 2024, as this is a very active frontier of research.

B. Machine learning for coarse-graining

Coarse-graining (CG)[236–239] refers to the technique to group atoms into larger particles termed *beads*, whose interactions are used to approximate the interaction energy among atoms. This enables the simulation of slow, collective motions while *averaging out* fast, local movements. For N atoms with coordinates $\mathbf{X} \in \mathbb{R}^{N\times 3}$ and n beads $\mathbf{x} \in \mathbb{R}^{n\times 3}$, the CG operator can be written as

$$\mathbf{x} = P\mathbf{X}, \tag{28}$$

where P is right-stochastic ($\sum_j P_{ij} = 1$). While P has traditionally been pre-defined and discrete, it can be made continuous and optimizable with machine learning.[240–242] This operator has several interesting properties. First of all, hence the translation and rotation equivariance discussed in Sec. II are naturally satisfied. Second, Eq. (28) also closely resembles the linear projection step in Linformer,[243] an approach to reduce the complexity of transformer[140] models from quadratic to linear. So, this represents a practical solution to modeling the long-range interactions in a self-consistent manner, as discussed in Sec. VIII B.

VIII. THE PATH FORWARD: WHAT WOULD THE NEXT GENERATION OF FORCE FIELDS LOOK LIKE?

Having reviewed the building blocks and recent trends of the development of both MM and ML force fields, in this section, we discuss a few key questions in the curation of force fields moving forward. We hope that this would serve as a blueprint for the development of a general-purpose, high-utility force field for biomolecular modeling in the next decade.

A. The dilemma of topology

It is debatable to what degree chemical bonds (and angles and torsions) are *real* and not artificial constructs. Classical biomolecular MM force fields usually require a *topology* (exceptions include Gale *et al.*[244]) and to define bond and angle energy accordingly, which is equivalent to putting a (very strong) prior on the probability density $\hat{p}$ in Eq. (2) to restrict $\mathbf{x}$ in a limited region. In practice, this is done by defining a topology graph where atoms are connected by chemical bonds. Since the chemical bonds have the harmonic functional form [Eq. (9)], any conformation that stretches the bonds too far will have $\hat{U} \to +\infty, \hat{p} \to 0$. Conversely, specifying $\hat{U}_{\text{bonds}} < U$ within a threshold defines a narrow region where the $\mathbf{x}$ is allowed to occupy. Such intense biases in a simulation ensure its stability and interpretability—clashes and distortions in the geometry will have near-zero probability under such formulation. At the same time, it also prohibits the force field model from incorporating reactive species and transition states.

Anecdotally, during the implementation of an MD simulation, a significant portion of time and effort of a researcher is typically spent on defining the *topology*—protonations and tautomers (namely, the location of the protonation and electrons)—of the system. Domain knowledge is also crucial in setting up an MD simulation, as even a

protonation state error can drastically change the entire energy landscape of a biomolecular system.[245]

MLFF, on the other hand, usually does not have the notion of *topology* entirely (exceptions exist, such as Eastman *et al.*,[246] which uses the molecular topology to define a set of pre-computed partial charges but does not explicitly restrain bonds and angles). Faithful to Eq. (2), it typically takes only the geometry $\mathbf{x}$ and the element identity h as input without restricting $\mathbf{x}$ on any subspace. Naturally, this mitigates the need for a carefully designed topological graph and can, in theory, simulate transition states and reactive species. There is no guarantee, however, that the simulation will stay stable and interpretable,[45] especially in high-energy regions.

To be topology-free can mean that the topological information needs to be re-realized every time. One can think that each forward pass of an MLFF model entails both the *topology realization stage* and the *inference* stage in an MM force field. As such, even for small perturbations on a limited conformational space, the computation always starts from scratch—with the exception of some very new foundation models.[35] We are interested in studying whether the realization of crude topology can be cached for similar conformations.

Being free of topology also means that there are no inductive biases for a model to avoid apparently unfavorable regions. Ideally, we envision that a force field should be able to incorporate constraints or restraints when specified, but remain free and flexible otherwise. We outline two promising directions to achieve this. First, the constraint can be encoded as an additional velocity field, as implemented in.[247] This incorporates the physics-based prior and restricts key geometries from adopting unreasonable conformations. Based on the modeler's domain knowledge, such priors can be placed on bonds that are not expected to break, or torsion angles not expected to change. Alternatively, one can also adopt a Δ-learning[248] approach to learn a small-magnitude, regularized deviation from an MM force field to keep key geometries in place.

B. What functional terms are here to stay?

Dot-product scalarization. We anticipate that this will be the main skeleton of the next generation of fast and accurate MLFFs. To reiterate the promising properties of this operator outlined in Sec. IV, dot-product scalarization is universally approximative, simple to implement, fast to execute, and has the same functional signature as the force field in general,. As examples of Lemma 3, all terms in Eq. (9) can be written as functions of dot products of edge (chemical bond) vectors, as one can trivially verify. For example, a self-dot product recovers the distance, and angles can be calculated as the ratio between dot products and distances. Since it is already universal, all invariant functions approximated by spherical harmonics can also be approximated by dot product-based algorithms, and spherical harmonics-based methods only afford the model with extra (physically inspired) inductive biases so that they are more data-efficient.

Long-range interactions. Recall the atomistic decomposition of energy [Eq. (11)]—this approach inherently assumes that the local structure is the primary determinant of interaction energy, and long-range interactions are either ignored or broken down into per-atom terms. This approximation limits the applicability of the developed models to local interactions.[249] Practically, most, if not all, MLFF models adopt a cutoff function [Eq. (5)] and interactions between particles more than 5 or 10 angstroms apart are all masked out. This means that crucial interactions are highly meaningful in key biomolecular processes such as protein folding, which has long been modeled using MM force fields,[250] cannot even be realized by MLFFs without non-local interactions. Here, we present two classes of solutions to this problem.

The first, more established solution entails modeling the non-bonded term separately using classical terms. For instance, van der Waals and Coulomb terms can be directly added to MLFFs.[90,249,251,252] The parameters of these terms can either be adopted from an existing force field (Sec. III), where a Δ-learning is carried out to learn the local interactions. Alternatively, the parameters can be jointly learned based on the chemical or geometrical environment, where the charges follow the charge equilibration scheme.[230,253] Free from compatibility issues in traditional MMs, this is also an excellent scenario where alternative forms to the 12–6 terms, such as the Buckingham temrs,[80,86] can be applied.

Despite the conceptual simplicity, this formulation still requires $\mathcal{O}(N^2)$ complexity if used without the PME approximation,[43] even though the quadratic term is very cheap to evaluate. Moreover, the long-range interactions will not have the flexibility and expressiveness as the short-range interactions afforded by the highly sophisticated neural networks. Thus, we provide a novel, alternative formulation using the coarse-graining operator [Eq. (28)]. Specifically, the coarse-graining matrix P, if made continuous, can contain faraway fine-grain atoms in coarse-grain beads. Thus, the treatment of the distances among coarse-grained beads will entail geometric information among faraway atoms, providing a unified and self-consistent solution to modeling both short- and long-term interactions.

C. Mixing MM with ML potential

Apart from incorporating pairwise functional forms for the long-range interactions, we can also further mix components of MM and ML potential models.

1. Mixing energy functions

ML/MM approaches, in direct analogy to QM/MM, can in principle help to close this timescale divide, and several recent works have demonstrated proof of principle on how such approaches could be employed to more efficiently model large biomolecular systems[100–105,254] (see Sec. VIII C). Recognizing the speed difference between MM and ML force fields and assuming that its ratio will remain near constant in the next generations of ML and MM force fields (which is a realistic assumption) leads to the question: Is it necessary to simulate the entire molecular system with an ML force field when the region involved in the event of interest is often limited to a small subsystem (reaction center, interactions between specific amino acids and a small molecule)? This is further motivated by the insight that generating an MM force field that generalizes well is challenging.

The *subtractive* scheme or *mechanical* embedding is the simplest and easiest to implement approach,[100,255,256] and is similar to the separate treatment of the long-range interactions detailed in the previous paragraph. In such a formalism, the ML force field describes the intramolecular energetics V_{ML} of the ligand $\vec{r}_L$ and the MM force field is responsible for the interaction between the ligand and the environment E

$$V_{MM/ML} = V_{MM}(\vec{r}_E, \vec{r}_L) + V_{ML}(\vec{r}_L) - V_{MM}(\vec{r}_L). \quad (29)$$

The main advantage is simplicity—no explicit MM-ML coupling terms are needed. On the other hand, the disadvantage is that the interaction between the MM and ML regions is handled entirely at the MM level; the limitations of classical force fields apply to non-bonded interactions.

This approach has been applied successfully and improved binding affinities,[101,102] yet there is some evidence that for solvation free energies it yields less convincing results.[257] In one particularly promising demonstration, Galvelis *et al.*[102] show that through a combination of well-optimized software and a mechanical embedding ML/MM scheme, they are able to perform MD simulations of large solvated protein-ligand complexes (∼30–60 thousand atoms), where the ligand (48–75 atoms) is treated at the level of the MLFF (ANI-2x[63]) with computational efficiencies just 1 order of magnitude slower than when solely MM is employed. It remains to be seen if the bespoke fitting of small molecule MM force fields to ML force fields can deliver the same results with increased speed.[197]

An improvement of the coupling (i.e., mutual polarization) of the charge densities between the MM and ML region is called *electrostatic* embedding, in which the interaction of the polarizable ML potential with the rigid MM densities (note that the charges in the MM regions are still fixed) are learned.[105,258] Nevertheless, improvements in computed free energies are yet to be demonstrated over the simpler mechanical embedding approach.

2. Mixing time scales

One general approach to speeding up MD simulations is to employ multiple time step (MTS) algorithms,[259–264] whereby the slower motions in a system are integrated less frequently than the faster motions. For standard FFs, this timescale separation is typically made with regard to the non-bonded vs bonded interactions, thereby allowing one to employ a larger time step between the more computationally expensive evaluations of the system's non-bonded interactions. For *ab initio* MD (AIMD) simulations, the same timescale separation cannot be as cleanly made. Instead, AIMD simulations typically adopt an MTS scheme where two levels of treatment for interactions in the system are employed, the target *ab initio* electronic structure method and another more computationally affordable model for the system's potential energy surface, such that the difference in forces between the two levels is slowly varying and can be integrated with a larger *outer* time step while the cheaper model is evaluated every *inner* time step.[265–268] In essence, the cheaper model needs to be accurate with respect to the target *ab initio* method for the faster short range interactions. Similar MTS schemes can be used to accelerate both MLFF and ML/MM simulations, and recently Jaffrelot Inizan *et al.*[104] demonstrated that for their benchmark modeling of benzene in water they were able to speed up their ML/MM (ANI-2x/AMOEBA) simulations approximately eightfold by employing MTS, with ML/MM and AMOEBA evaluations performed every 2 and 0.5 fs, respectively.

D. Ecosystems of molecular dynamics simulations, unite?

While the preceding sections have shown that the conceptual or mathematical formulation of MM and ML force fields is a continuous spectrum, with components not fundamentally different, they are typically implemented in completely different software platforms. Currently, the MM infrastructure and simulation platforms are typically segregated from the tensor-accelerating frameworks that are ubiquitous in all schools of machine learning and scientific computing. The reason can be attributed to the particularity of the functional forms and, consequently, the highly specialized kernels designed for these functions. The intensive requirements for domain knowledge might also justify the need for domain-specific software. As such, optimizing MM force fields requires reimplementation thereof and to blend MM force fields in an MLFF is also highly non-trivial.

1. Differentiable simulation

The conceptually simplest approach for controlling the course of the simulation and enabling the use of gradient-based methods to optimize the force field parameters (see Sec. II), apart from *post-hoc* reweighting-based techniques[169,187] that introduce additional error, is to make an MD simulation differentiable,[191,269,270] so that the evaluation of derivatives $\partial\mathscr{O}/\partial\theta$ some variables based upon the trajectory $\mathscr{O}(\mathbf{x})$ w.r.t. some parameters dictating the energy landscape $U(\cdot|\theta)$ is feasible.

The technical cornerstone for achieving this objective is the *adjoint sensitivity* method for taking derivatives across ordinary or stochastic differential equations (ODE/SDE),[271–273] which allows the differentiation of the loss function $\mathscr{L}(\mathbf{x}_T)$, dependent upon a later state $\mathbf{x}_T$ in a trajectory $\mathbf{x}_{0\cdots t\cdots T}$, to be evaluated in constant memory and linear time. Specifically, in an ODE setting where the dynamics are controlled by a parametrized function $\mathrm{d}\mathbf{x}/\mathrm{d}t = f(\mathbf{x}; t, \theta)$, for example, the *adjoint*, defined as $a(t) = \partial\mathscr{L}/\partial\mathbf{x}_t$, can be calculated by another ODE

$$\mathrm{d}a(t)/\mathrm{d}t = a(t)^T \partial f/\partial\mathbf{x}. \quad (30)$$

The gradient w.r.t. to the parameters can then be calculated as

$$\mathrm{d}L/\mathrm{d}\theta = -\int \mathrm{d}t a^T(t)\partial f/\partial\theta, \quad (31)$$

where $a\partial f/\partial\mathbf{x}$ and $a^T\partial f/\partial\theta$ can be efficiently computed as vector-Jacobian products.

Using this formalism, *reimplementing* MD simulation introduces a slowdown by a few factors, whereas the differentiation across the SDE introduces further speed bump and instability. It is worth considering whether treating the MD simulation as a black box and using gradient estimation methods might be desirable. For example, a common algorithm used in reinforcement learning is the policy gradient method,[274,275] where the gradient of parameters Θ can be approximated as

$$\nabla_\theta\mathscr{O} = \mathop{\mathbb{E}}_{\theta\sim p}[\log p(\theta)\mathscr{O}(\theta)], \quad (32)$$

if we assume the parameter Θ comes from a distribution p. Here, the SDE integration/MD simulation is simply treated as an oracle, just like an Atari game console that emits signals following a set of parametrized actions. As such, we can take the parameter derivatives of any target, including physical properties and binding free energies, although the variance of the Monte Carlo estimates might be potentially large.

2. MM re-implementation in tensor-accelerating frameworks

Schoenholz and Cubuk[276] and Doerr *et al.*[277] strive to overcome this barrier by implementing the MM energy functions and sampling

strategies in JAX[278] and PyTorch[279] to enable the end-to-end differentiation of MM energy functions. To efficiently do so is met with a multitude of challenges: to start with, tensor-accelerating frameworks are designed to run on a wide variety of hardware platforms, with different support characteristics and capabilities. For example, the support for different or mixed precisions differs for each piece of hardware, which is not detrimental to ML applications due to its over-parametrized nature,[280] but is crucial for MM simulations to run efficiently. More fundamentally, tensor-accelerating frameworks focus on general applicability and being able to efficiently evaluate a wide range of energy functional forms, with emphasis especially on highly parallel linear algebra operations (which, again, makes the dot-product-based operations more desirable) (Sec. VIII B), whereas MM platforms have more aggressively optimized the few local and pairwise terms. Efforts[281] to push toward accelerating MM implementation in tensor-accelerating frameworks usually end up re-writing all lower-level kernels for energy evaluation.

3. ML plug-in in MM platforms

Conversely, Eastman *et al.*[16] and Galvelis *et al.*[102] offer functionalities to plug MLFF into traditional MD simulations to streamline the inference and sampling of MLFF on biologically relevant systems. We envision that this will be the drive of the next generation of hardware-specific (such as Anton[282,283]) revolution.

The speed bottleneck in these implementations usually lies in the frequent transfer of coordinates and forces between the MM platform and the tensor-accelerating framework. At the very least, zero-copy is required to avoid duplicating the data or carrying the data from GPUs to hosts.

E. Foundation models for force fields and more

Many practitioners in the molecular modeling community are interested in characterizing one type of system, although in the case of force fields, the physical law governing the energy landscape—Schrödinger's equation—is singular across systems. This calls for foundation models—an approach to representation learning that involves training a large artificial neural network on extremely large amounts of heterogeneous, multi-modal, and easily available data.[284] The learned representation is then leveraged and fine-tuned for various downstream tasks. Training on large and diverse datasets allows the model to generalize well across tasks and domains. Foundation models have enabled truly spectacular achievements in natural language processing and computer vision, and have begun to penetrate the physical sciences as demonstrated by recent advances in protein structure prediction.

Recent work in the field of ML atomistic models has begun to recognize the need for more scalable approaches resulting in the generation of larger datasets, e.g., the Open Catalyst (OC20),[285] Materials Project,[286] and Open Materials[287] datasets, attempts to establish neural scaling laws for chemical models (on single data sources and, typically, on rather small scales) as well as universal models capable of few-shot, or even zero-shot, learning. Kabylda *et al.*[35] represents a meaningful step toward a general model for atomistic simulation of larger biomolecular systems. Although far from the successes of natural language processing and computer vision, these results suggest that by scaling these domain-specific models, we should be achievable comparable successes to those of more traditional computer science domains.

The high dimensionality of the chemical space and the scarcity and computational cost of generating training data, as well as the inconsistencies among available data, represent major obstacles for developing foundation models that are predictive over a broad range of materials and properties as needed for molecular and materials discovery. To fulfill this vision, critical limitations need to be overcome, such as *data scarcity* and *data redundancy*, the design of novel training strategies that can deal with inconsistencies across datasets, the exploration of adaptive batching and curriculum learning strategies, the development of scalable physics-informed models, as well as the characterization of scaling laws for domain-specific neural architectures.

IX. CONCLUSIONS

Currently, the MM force fields are arguably fast enough but not sufficiently accurate; the ML force fields are accurate enough but painfully slow. We anticipate that a force field faster than current state-of-the-art MLFF and significantly more accurate than MM, will show the greatest utility in the *in silico* modeling of biomolecular systems. With that in mind, we have herein surveyed the recent advances in either school in search of opportunities to bridge the design spaces. The key observations are summarized as follows.

- Current state-of-the-art MM force fields are very inaccurate when it comes to energy landscapes (Fig. 2).
- Current ML force fields are magnitudes slower than MM force fields (Sec. VII).
- The datasets used in the curation of all schools of force fields are diverse in terms of sampling strategy, chemical space coverage, and levels of theory. Some erratic training and evaluation metrics are still in use. (Table III, Sec. V A).

Building from the components from both MM and MLFFs, we envision that the next generation of MLFFs with the desired balance between speed and accuracy will have the following properties.

- It will be composed of simple yet universal equivariant operators, such as the dot-product scalarization (Lemma 3).
- Physically inspired inductive biases will be encoded to ensure its stability and smoothness, and topology can be additionally encoded as constraints (Secs. VIII B and VIII A).

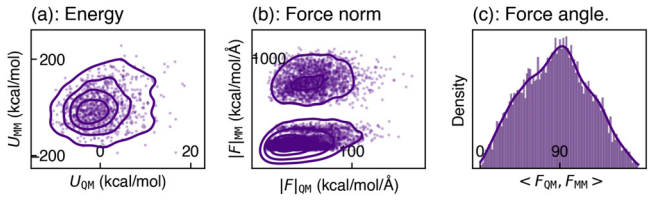

FIG. 2. Between MM and QM energies and forces, there is little correlation. Scatter plots and kernel density estimate (KDE) of: (a) MM energy (U_{MM}, mean-subtracted) plotted against QM energy (U_{QM}, mean-subtracted); (b) MM force magnitude ($|F|_{MM}$) plotted against QM force magnitude ($|F|_{QM}$); (c) distribution of deviation of angles between QM and MM forces. QM energies refer to the CCSD(T) computation of the ethanol molecule in MD17[22] dataset. MM energies and forces are re-calculated using the state-of-the-art `openff-2.0.0`[20] force field.

- This model will likely still rely on automatic differentiation and will be implemented in general tensor-accelerating frameworks (Sec. VIII D).

Finally, community-driven efforts will be the key to determining the chemical and conformational space coverage.

X. A NIHILIST EPILOGUE: BUT DO We REALLY NEED a FORCE FIELD?

Echoing the opening paragraph of this review, computational chemists and biophysicists, on most occasions, are interested in *sampling from* the Boltzmann distribution [Eq. (1)] rather than just knowing its exact value up to a normalizing constant. It has quickly emerged to become a focus of cutting-edge research to generate samples directly on the coordinate space for biomolecular systems of interest—from small molecule conformer generation,[288–292] folded protein structure prediction,[40,293,294] to trajectory forecasting.[31,295] More fascinatingly, a class of machine learning models known as *Boltzmann generators*[296–301] sample the Boltzmann distribution in an asymptotically unbiased fashion in one shot, without relying on simulation.

Sampling and *force field* have long been two orthogonal axes in computational chemistry research, with little cross-disciplinary communications. The theoretical and experimental advances in energy-based models[302,303] (EBM) and diffusion models[304–308] should remind us that these can be the same thing. The loss function in probabilistic generative modeling is usually some rendering of the energy function—in the context of learning by examples, the loss function can be viewed as the local Gaussian or Laplacian extrapolation around the learning set. When training a generative model, one might get a force field for free.[309]

ACKNOWLEDGMENTS

This work was carried out on the high-performance computing resources at Memorial Sloan Kettering Cancer Center and the Washington Square and Abu Dhabi campuses of New York University—we are grateful to the technical supporting teams.

Many ideas in this paper are heavily inspired by if not merely a transcription of, conversations between YW and John D. Chodera (MSKCC), Yutong Zhao (Relay Therapeutics), Xiang Fu (Meta, Inc.), and Sheng Gong (ByteDance, Inc.), toward whom YW is thankful.

YW acknowledges the support from the Schmidt Science Fellowship in partnership with the Rhodes Trust. YW and MSC acknowledge the support from the Simons Center for Computational Physical Chemistry at New York University. SM acknowledges the NSF (Award No. 2311632) and the Simons Center for Computational Physical Chemistry for financial support. PE acknowledges the support from the NIH (Award No. R01GM140090).

YW has limited financial interest in Flagship Pioneering, Inc. and its subsidiaries.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Yuanqing Wang: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Kenichiro Takaba:** Writing – original draft (equal). **Michael S. Chen:** Writing – original draft (equal). **Marcus Wieder:** Writing – original draft (equal). **Yuzhi Xu:** Writing – original draft (equal). **Tong Zhu:** Writing – original draft (equal). **John Z. H. Zhang:** Writing – original draft (equal). **Arnav Nagle:** Writing – original draft (equal). **Kuang Yu:** Writing – original draft (equal). **Xinyan Wang:** Writing – original draft (equal). **Daniel J. Cole:** Writing – original draft (equal). **Joshua A. Rackers:** Writing – original draft (equal). **KyungHyun Cho:** Writing – original draft (equal). **Joe G. Greener:** Writing – original draft (equal). **Peter Eastman:** Writing – original draft (equal). **Stefano Martiniani:** Writing – original draft (equal). **Mark E. Tuckerman:** Funding acquisition (equal); Writing – original draft (equal).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1] L. Boltzmann, "Studien uber das gleichgewicht der lebenden kraft," Wissenschafiliche Abhandlungen **1**, 49–96 (1868).
[2] M. Levitt and S. Lifson, "Refinement of protein conformations using a macromolecular energy minimization procedure," J. Mol. Biol. **46**, 269–279 (1969).
[3] A. Hagler, E. Huler, and S. Lifson, "Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals," J. Am. Chem. Soc. **96**, 5319–5327 (1974).
[4] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," Nature **267**, 585–590 (1977).
[5] J. W. Ponder and D. A. Case, "Force fields for protein simulations," in *Advances in Protein Chemistry* (Elsevier, 2003), Vol. 66, pp. 27–85.
[6] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "GROMACS: Fast, flexible, and free," J. Comput. Chem. **26**, 1701–1718 (2005).
[7] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The amber biomolecular simulation programs," J. Comput. Chem. **26**, 1668–1688 (2005).
[8] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with NAMD," J. Comput. Chem. **26**, 1781–1802 (2005).
[9] F. E. Calculations, *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, Springer Series in Chemical Physics Vol. 86 (Springer, 2007).
[10] C. Wang, G. Pilania, S. Boggs, S. Kumar, C. Breneman, and R. Ramprasad, "Computational strategies for polymer dielectrics design," Polymer **55**, 979–988 (2014).
[11] C. Li and A. Strachan, "Molecular scale simulations on thermoset polymers: A review," J. Polym. Sci, Part B: Polym. Phys. **53**, 103–122 (2015).
[12] H. Sun, Z. Jin, C. Yang, R. L. Akkermans, S. H. Robertson, N. A. Spenley, S. Miller, and S. M. Todd, "Compass II: Extended coverage for polymer and drug-like molecule databases," J. Mol. Model. **22**, 47 (2016).
[13] M. J. Harvey, G. Giupponi, and G. D. Fabritiis, "ACEMD: Accelerating biomolecular dynamics in the microsecond time scale," J. Chem. Theory Comput. **5**, 1632–1639 (2009).
[14] R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. Le Grand, and R. C. Walker, "Routine microsecond molecular dynamics simulations with AMBER on

GPUs. 2. Explicit solvent particle mesh Ewald," J. Chem. Theory Comput. **9**, 3878–3888 (2013).
[15]P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern *et al.*, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," PLoS Comput. Biol. **13**, e1005659 (2017).
[16]P. Eastman, R. Galvelis, R. P. Peláez, C. R. Abreu, S. E. Farr, E. Gallicchio, A. Gorenko, M. M. Henry, F. Hu, J. Huang *et al.*, "OpenMM 8: Molecular dynamics simulation with machine learning potentials," J. Phys. Chem. B **128**, 109–116 (2023a).
[17]L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood *et al.*, "Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field," J. Am. Chem. Soc. **137**, 2695–2703 (2015).
[18]C. E. Schindler, H. Baumann, A. Blum, D. Bose, H.-P. Buchstaller, L. Burgdorf, D. Cappel, E. Chekler, P. Czodrowski, D. Dorsch *et al.*, "Large-scale assessment of binding free energy calculations in active drug discovery projects," J. Chem. Inf. Model. **60**, 5457–5474 (2020).
[19]V. Gapsys, D. F. Hahn, G. Tresadern, D. L. Mobley, M. Rampp, and B. L. de Groot, "Pre-exascale computing of protein–ligand binding free energies with open source software for drug design," J. Chem. Inf. Model. **62**, 1172–1177 (2022).
[20]S. Boothroyd, P. K. Behara, O. C. Madin, D. F. Hahn, H. Jang, V. Gapsys, J. R. Wagner, J. T. Horton, D. L. Dotson, M. W. Thompson *et al.*, "Development and benchmarking of open force field 2.0.0: The sage small molecule force field," J. Chem. Theory Comput. **19**, 3251 (2023).
[21]K. Takaba, A. J. Friedman, C. E. Cavender, P. K. Behara, I. Pulido, M. M. Henry, H. MacDermott-Opeskin, C. R. Iacovella, A. M. Nagle, A. M. Payne, M. R. Shirts, D. L. Mobley, J. D. Chodera, and Y. Wang, "Machine-learned molecular mechanics force fields from large-scale quantum chemical data," Chem. Sci. **15**, 12861–12878 (2024).
[22]S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields," Sci. Adv. **3**, e1603015 (2017a).
[23]O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, "Machine learning force fields," Chem. Rev. **121**, 10142–10186 (2021a).
[24]J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost," Chem. Sci. **8**, 3192–3203 (2017a).
[25]J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, "Less is more: Sampling chemical space with active learning," J. Chem. Phys. **148**, 241733 (2018).
[26]J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, "Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning," Nat. Commun. **10**, 1–8 (2019).
[27]C. Devereux, J. S. Smith, K. K. Davis, K. Barros, R. Zubatyuk, O. Isayev, and A. E. Roitberg, "Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens," J. Chem. Theory Comput. **16**, 4192–4202 (2020).
[28]K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet—A deep learning architecture for molecules and materials," J. Chem. Phys. **148**, 241722 (2018).
[29]S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," Nat. Commun. **13**, 2453 (2022a).
[30]Y. Han, Z. Wang, Z. Wei, J. Liu, and J. Li, "Machine learning builds full-QM precision protein force fields in seconds," Briefings Bioinf. **22**, bbab158 (2021).
[31]Y. Wang and J. D. Chodera, "Spatial attention kinetic networks with e(n)-equivariance," arXiv:2301.08893 (2023).
[32]A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky, "Learning local equivariant representations for large-scale atomistic dynamics," Nat. Commun. **14**, 579 (2023).
[33]G. D. Fabritiis, "Machine learning potentials: A roadmap toward next-generation biomolecular simulations," arXiv:2408.12625 (2024).
[34]S. Barnett and J. D. Chodera, "Neural network potentials for enabling advanced small-molecule drug discovery and generative design," GEN Biotechnol. **3**, 119–129 (2024).
[35]A. Kabylda, J. T. Frank, S. S. Dou, A. Khabibrakhmanov, L. M. Sandonas, O. T. Unke, S. Chmiela, K.-R. Muller, and A. Tkatchenko, "Molecular simulations with a pretrained neural network and universal pairwise force fields," ChemRxiv (2024).
[36]J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Phys. Rev. Lett. **98**, 146401 (2007).
[37]S. Thais and D. Murnane, "Equivariance is not all you need: Characterizing the utility of equivariant graph neural networks for particle physics tasks," arXiv:2311.03094 (2023).
[38]O. Puny, M. Atzmon, H. Ben-Hamu, I. Misra, A. Grover, E. J. Smith, and Y. Lipman, "Frame averaging for invariant and equivariant network design," arXiv:2110.03336 (2022).
[39]A. Duval, V. Schmidt, A. H. Garcia, S. Miret, F. D. Malliaros, Y. Bengio, and D. Rolnick, "FaeNet: Frame averaging equivariant GNN for materials modeling," arXiv:2305.05577 (2023).
[40]J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick *et al.*, "Accurate structure prediction of biomolecular interactions with AlphaFold 3," Nature **630**, 493 (2024).
[41]J. A. Barker and R. O. Watts, "Monte Carlo studies of the dielectric properties of water-like models," Mol. Phys. **26**, 789–792 (1973).
[42]R. Watts, "Monte Carlo studies of liquid water," Mol. Phys. **28**, 1069–1083 (1974).
[43]J. Kolafa and J. W. Perram, "Cutoff errors in the Ewald summation formulae for point charge systems," Mol. Simul. **9**, 351–368 (1992).
[44]D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," arXiv:1511.07289 (2016).
[45]X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli, and T. Jaakkola, "Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations," arXiv:2210.07237 (2023).
[46]S. Stocker, J. Gasteiger, F. Becker, S. Günnemann, and J. T. Margraf, "How robust are modern graph neural network potentials in long and hot molecular dynamics simulations?," Mach. Learn: Sci. Technol. **3**, 045010 (2022).
[47]W. Wang, T. Yang, W. H. Harris, and R. Gómez-Bombarelli, "Active learning and neural network potentials accelerate molecular screening of ether-based solvate ionic liquids," Chem. Commun. **56**, 8920–8923 (2020a).
[48]D. Schwalbe-Koda, A. R. Tan, and R. Gómez-Bombarelli, "Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks," Nat. Commun. **12**, 5104 (2021).
[49]K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," Science **334**, 517–520 (2011).
[50]J. Kubelka, J. Hofrichter, and W. A. Eaton, "The protein folding 'speed limit'," Curr. Opin. Struct. Biol. **14**, 76–88 (2004).
[51]W. A. Eaton, "Modern kinetics and mechanism of protein folding: A retrospective," J. Phys. Chem. B **125**, 3452–3467 (2021).
[52]B. Manavalan, K. Kuwajima, and J. Lee, "PFDB: A standardized protein folding database with temperature correction," Sci. Rep. **9**, 1588 (2019).
[53]J.-C. Horng, V. Moroz, and D. P. Raleigh, "Rapid cooperative two-state folding of a miniature α–β protein and design of a thermostable variant," J. Mol. Biol. **326**, 1261–1270 (2003).
[54]Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, and T. F. Miller, "OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features," J. Chem. Phys. **153**, 124111 (2020).
[55]M. Retchin, Y. Wang, K. Takaba, and J. D. Chodera, "DrugGym: A testbed for the economics of autonomous drug discovery," bioRxiv:2024.05.28.596296 (2024).
[56]A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: A survey," J. Mach. Learn. Res. **18**, 1–43 (2018).

[57]J. A. Pople, "Nobel lecture: Quantum chemical models," Rev. Mod. Phys. **71**, 1267 (1999).
[58]M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, "Quantum chemical accuracy from density functional approximations via machine learning," Nat. Commun. **11**, 5223 (2020).
[59]D. L. Mobley and P. V. Klimovich, "Perspective: Alchemical free energy calculations for drug discovery," J. Chem. Phys. **137**, 230901 (2012).
[60]I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csányi, "MACE: Higher order equivariant message passing neural networks for fast and accurate force fields," in Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2022a), Vol. 35, pp. 11423–11436.
[61]A. S. Christensen and O. A. von Lilienfeld, "On the role of gradients for machine learning of molecular energies and forces," arXiv:2007.09593 (2020a).
[62]P. Eastman, P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr, J. T. Horton, Y. Mao, J. D. Chodera, B. P. Pritchard, Y. Wang *et al.*, "Spice, a dataset of drug-like molecules and peptides for training machine learning potentials," Sci. Data **10**, 11 (2023b).
[63]J. S. Smith, R. Zubatyuk, B. Nebgen, N. Lubbers, K. Barros, A. E. Roitberg, O. Isayev, and S. Tretiak, "The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules," Sci. Data **7**, 134 (2020).
[64]D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, V. Kapil, W. C. Witt, I.-B. Magdău, D. J. Cole, and G. Csányi, "Mace-off23: Transferable machine learning force fields for organic molecules," arXiv:2312.15211 (2023a).
[65]G. Simeon and G. De Fabritiis, "TensorNet: Cartesian tensor representations for efficient learning of molecular potentials," in Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2024), Vol. 36.
[66]I. Poltavsky, M. Puleva, A. Charkin-Gorbulin, G. C. Fonseca, I. Batatia, N. J. Browning, S. Chmiela, M. Cui, J. T. Frank, S. Heinen *et al.*, "Crash testing machine learning force fields for molecules, materials, and interfaces: Molecular dynamics in the TEA challenge 2023," Chem. Sci. **16**, 2931 (2025).
[67]P. Dauber-Osguthorpe and A. T. Hagler, "Biomolecular force fields: Where have we been, where are we now, where do we need to go and how do we get there?," J. Comput.-Aided. Mol. Des. **33**, 133–203 (2019).
[68]A. T. Hagler, "Force field development phase II: Relaxation of physics-based criteria… or inclusion of more rigorous physics into the representation of molecular energetics," J. Comput.-Aided. Mol. Des. **33**, 205–264 (2019).
[69]C. Ringrose, J. T. Horton, L.-P. Wang, and D. J. Cole, "Exploration and validation of force field design protocols through QM-to-MM mapping," Phys. Chem. Chem. Phys. **24**, 17014–17027 (2022).
[70]X. C. Yan, M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen, "Improved description of sulfur charge anisotropy in OPLS force fields: Model development and parameterization," J. Phys. Chem. B **121**, 6626–6636 (2017).
[71]M. H. Kolar and P. Hobza, "Computer modeling of halogen bonds and other σ-hole interactions," Chem. Rev. **116**, 5155–5187 (2016).
[72]J. Delhommelle and P. Millié, "Inadequacy of the Lorentz-Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation," Mol. Phys. **99**, 619–625 (2001).
[73]T. A. Halgren, "The representation of van der Waals (vdw) interactions in molecular mechanics force fields: Potential form, combination rules, and VDW parameters," J. Am. Chem. Soc. **114**, 7827–7843 (1992).
[74]M. Abraham, A. Alekseenko, V. Basov, C. Bergh, E. Briand, A. Brown, M. Doijade, G. Fiorin, S. Fleischmann, S. Gorelov, G. Gouaillardet, A. Grey, E. I. M. F. Jalalypour, J. Jordan, C. Kutzner, J. A. Lemkul, M. Lundborg, P. Merz, V. Miletic, D. Morozov, J. Nabet, S. Pall, A. Pasquadibisceglie, M. Pellegrino, H. Santuz, R. Schulz, T. Shugaeva, A. Shvetsov, A. Villa, S. Wingbermuehle, B. Hess, and E. Lindahl, GROMACS User Manual, version 2024.2 (2024). Available at https://manual.gromacs.org/2024.2/manual-2024.2.pdf.
[75]P. M. Morse, "Diatomic molecules according to the wave mechanics. II. Vibrational levels," Phys. Rev. **34**, 57–64 (1929).
[76]K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, Jr., "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," J. Comput. Chem. **31**, 671–690 (2010).
[77]S. L. Mayo, B. D. Olafson, and W. A. Goddard, "Dreiding: A generic force field for molecular simulations," J. Phys. Chem. **94**, 8897–8909 (1990).
[78]M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen, "Improved peptide and protein torsional energetics with the OPLS-AA force field," J. Chem. Theory Comput. **11**, 3499–3509 (2015).
[79]J. E. Jones, "On the determination of molecular fields—II. From the equation of state of a gas," Proc. R. Soc. London, Ser. A **106**, 463–477 (1924).
[80]R. A. Buckingham, "The classical equation of state of gaseous helium, neon and argon," Proc. R. Soc. London, Ser. A **168**, 264–283 (1938).
[81]J. J. K. Chung, M. L. Brown, and P. L. A. Popelier, "Transferability of Buckingham parameters for short-range repulsion between topological atoms," J. Chem. Theory Comput. **128**, 4561–4572 (2024).
[82]J. R. Hart and A. K. Rappé, "van der Waals functional forms for molecular simulations," J. Chem. Phys. **97**, 1109–1115 (1992).
[83]L. Yang, L. Sun, and W.-Q. Deng, "Combination rules for Morse-based van der Waals force fields," J. Phys. Chem. A **122**, 1672–1677 (2018).
[84]X. Wu and B. R. Brooks, "A double exponential potential for van der Waals interaction," AIP Adv. **9**, 065304 (2019).
[85]V. H. Man, X. Wu, X. HeXiang-Qun, X. R. Brooks, and J. Wang, "Determination of van der Waals parameters using a double exponential potential for nonbonded divalent metal cations in tip3p solvent," J. Chem. Theory Comput. **17**, 1086–1097 (2021).
[86]J. T. Horton, S. Boothroyd, P. K. Behara, D. L. Mobley, and D. J. Cole, "A transferable double exponential potential for condensed phase simulations of small molecules," Digital Discovery **2**, 1178–1187 (2023).
[87]S. Chmiela, V. Vassilev-Galindo, O. T. Unke, A. Kabylda, H. E. Sauceda, A. Tkatchenko, and K.-R. Müller, "Accurate global machine learning force fields for molecules with hundreds of atoms," Sci. Adv. **9**, eadf0873 (2023).
[88]D. P. Kovács, I. Batatia, E. S. Arany, and G. Csányi, "Evaluation of the MACE force field architecture: From medicinal chemistry to materials science," J. Chem. Phys. **159**, 044118 (2023b).
[89]O. T. Unke and M. Meuwly, "PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges," J. Chem. Theory Comput. **15**, 3678–3693 (2019).
[90]O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda, and K.-R. Müller, "SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects," Nat. Commun. **12**, 7273 (2021b).
[91]F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, "Machine learning for molecular simulation," Annu. Rev. Phys. Chem. **71**, 361–390 (2020).
[92]M. S. Chen, J. Lee, H.-Z. Ye, T. C. Berkelbach, D. R. Reichman, and T. E. Markland, "Data-efficient machine learning potentials from transfer learning of periodic correlated electronic structure methods: Liquid water at AFQMC, CCSD, and CCSD(T) accuracy," J. Chem. Theory Comput. **19**, 4510–4519 (2023).
[93]M. Rossi, W. Fang, and A. Michaelides, "Stability of complex biomolecular structures: Van der Waals, hydrogen bond cooperativity, and nuclear quantum effects," J. Phys. Chem. Lett. **6**, 4233–4238 (2015).
[94]K. Yao, J. E. Herr, D. Toth, R. Mckintyre, and J. Parkhill, "The tensormol-0.1 model chemistry: A neural network augmented with long-range physics," Chem. Sci. **9**, 2261–2269 (2018).
[95]T. Plé, L. Lagardère, and J.-P. Piquemal, "Force-field-enhanced neural network interactions: From local equivariant embedding to atom-in-molecule properties and long-range effects," Chem. Sci. **14**, 12554–12569 (2023).
[96]D. M. Anstine, R. Zubatyuk, and O. Isayev, "AIMNet2: A neural network potential to meet your neutral, charged, organic, and elemental-organic needs," ChemRxiv (2023).
[97]B. Kozinsky, A. Musaelian, A. Johansson, and S. Batzner, "Scaling the leading accuracy of deep equivariant models to biomolecular simulations of realistic size," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (ACM, 2023), pp. 1–12.
[98]O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. M. Sandonas, J. T. Berryman, A. Tkatchenko, and K.-R. Müller, "Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments," Sci. Adv. **10**, eadn4397 (2024).
[99]Z. Cheng, J. Du, L. Zhang, J. Ma, W. Li, and S. Li, "Building quantum mechanics quality force fields of proteins with the generalized energy-based fragmentation approach and machine learning," Phys. Chem. Chem. Phys. **24**, 1326–1337 (2022).

[100]S. L. J. Lahey and C. N. Rowley, "Simulating protein-ligand binding with neural network potentials," Chem. Sci. **11**, 2362–2368 (2020).
[101]D. A. Rufa, H. E. Bruce Macdonald, J. Fass, M. Wieder, P. B. Grinaway, A. E. Roitberg, O. Isayev, and J. D. Chodera, "Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning/molecular mechanics potentials," bioRxiv:2020.07.29.227959 (2020).
[102]R. Galvelis, A. Varela-Rial, S. Doerr, R. Fino, P. Eastman, T. E. Markland, J. D. Chodera, and G. De Fabritiis, "NNP/MM: Accelerating molecular dynamics simulations with machine learning potentials and molecular mechanics," J. Chem. Inf. Model. **63**, 5701–5708 (2023).
[103]F. Sabanés Zariquiey, R. Galvelis, E. Gallicchio, J. D. Chodera, T. E. Markland, and G. De Fabritiis, "Enhancing protein–ligand binding affinity predictions using neural network potentials," J. Chem. Inf. Model. **64**, 1481–1485 (2024).
[104]T. Jaffrelot Inizan, T. Plé, O. Adjoua, P. Ren, H. Gökcan, O. Isayev, L. Lagardère, and J.-P. Piquemal, "Scalable hybrid deep neural networks/polarizable potentials biomolecular simulations including long-range effects," Chem. Sci. **14**, 5438–5452 (2023).
[105]K. Zinovjev, L. Hedges, R. Montagud Andreu, C. Woods, I. Tuñón, and M. W. van der Kamp, "emle-engine: A flexible electrostatic machine learning embedding package for multiscale molecular dynamics simulations," J. Chem. Theory Comput. **20**, 4514–4522 (2024).
[106]D. W. Zhang and J. Z. H. Zhang, "Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy," J. Chem. Phys. **119**, 3599–3605 (2003).
[107]X. He and J. Z. H. Zhang, "The generalized molecular fractionation with conjugate caps/molecular mechanics method for direct calculation of protein energy," J. Chem. Phys. **124**, 184703 (2006).
[108]X. Wang, J. Liu, J. Z. H. Zhang, and X. He, "Electrostatically embedded generalized molecular fractionation with conjugate caps method for full quantum mechanical calculation of protein energy," J. Phys. Chem. A **117**, 7149–7161 (2013).
[109]W. Li, S. Li, and Y. Jiang, "Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules," J. Phys. Chem. A **111**, 2193–2199 (2007).
[110]H. Wang and W. Yang, "Toward building protein force fields by residue-based systematic molecular fragmentation and neural network," J. Chem. Theory Comput. **15**, 1409–1417 (2019).
[111]Z. Wang, Y. Han, J. Li, and X. He, "Combining the fragmentation approach and neural network potential energy surfaces of fragments for accurate calculation of protein energy," J. Phys. Chem. B **124**, 3027–3035 (2020b).
[112]Y. Han, Z. Wang, A. Chen, I. Ali, J. Cai, S. Ye, and J. Li, "An inductive transfer learning force field (ITLFF) protocol builds protein force fields in seconds," Briefings Bioinf. **23**, bbab590 (2022).
[113]P. Thölke and G. D. Fabritiis, "TorchMD-NET: Equivariant transformers for neural network based molecular potentials," arXiv:2202.02541 (2022).
[114]K. Schütt, O. Unke, and M. Gastegger, "Equivariant message passing for the prediction of tensorial properties and molecular spectra," in *Proceedings of the International Conference on Machine Learning* (PMLR, 2021), pp. 9377–9388.
[115]J. T. Frank, O. T. Unke, K.-R. Müller, and S. Chmiela, "A Euclidean transformer for fast and stable machine learned force fields," Nat. Commun. **15**, 6539 (2024).
[116]D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," in *Proceedings of the Advances in Neural Information Processing Systems* (2015), pp. 2224–2232.
[117]T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv:1609.02907 (2016).
[118]J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proceedings of the International Conference on Machine Learning* (PMLR, 2017), pp. 1263–1272.
[119]K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?," arXiv:1810.00826 (2018).
[120]P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner *et al.*, "Relational inductive biases, deep learning, and graph networks," arXiv:1806.01261 (2018).
[121]J. Du, S. Zhang, G. Wu, J. M. F. Moura, and S. Kar, "Topology adaptive graph convolutional networks," arXiv:1710.10370 (2018).
[122]F. Wu, T. Zhang, A. H. d. Souza, Jr., C. Fifty, T. Yu, and K. Q. Weinberger, "Simplifying graph convolutional networks," arXiv:1902.07153 (2019a).
[123]M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, Y. Gai *et al.*, "Deep graph library: A graph-centric, highly-performant package for graph neural networks," arXiv:1909.01315 (2019a).
[124]Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," ACM Trans. Graph. **38**, 1–12 (2019b).
[125]C. K. Joshi, C. Bodnar, S. V. Mathis, T. Cohen, and P. Lio, "On the expressive power of geometric graph neural networks," in *Proceedings of the International Conference on Machine Learning* (PMLR, 2023), pp. 15330–15355.
[126]Y. Wang and T. Karaletsos, "Stochastic aggregation in graph neural networks," arXiv:2102.12648 (2021).
[127]P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," arXiv:1710.10903 (2018).
[128]F. Wu, A. Souza, Jr., T. Zhang, C. Fifty, T. Yu, and K. Q. Weinberger, "Simplifying graph convolutional networks," arXiv:1902.07153 (2019b).
[129]B. P. Chamberlain, J. Rowbottom, M. I. Gorinova, S. Webb, E. Rossi, and M. M. Bronstein, "GRAND: Graph neural diffusion," arXiv:2106.10934 (2021).
[130]C. Cai and Y. Wang, "A note on over-smoothing for graph neural networks," arXiv:2006.13318 (2020).
[131]T. K. Rusch, M. M. Bronstein, and S. Mishra, "A survey on oversmoothing in graph neural networks," arXiv:2303.10993 (2023).
[132]U. Alon and E. Yahav, "On the bottleneck of graph neural networks and its practical implications," arXiv:2006.05205 (2020).
[133]J. Topping, F. D. Giovanni, B. P. Chamberlain, X. Dong, and M. M. Bronstein, "Understanding over-squashing and bottlenecks on graphs via curvature," arXiv:2111.14522 (2022).
[134]G. Corso, L. Cavalleri, D. Beaini, P. Liò, and P. Veličković, "Principal neighbourhood aggregation for graph nets," arXiv:2004.05718 (2020).
[135]V. K. Garg, S. Jegelka, and T. S. Jaakkola, "Generalization and representational limits of graph neural networks," arXiv:2002.06157 (2020).
[136]Y. Wang and K. Cho, "Non-convolutional graph neural networks," arXiv:2408.00165 (2024).
[137]K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "SchNet: A continuous-filter convolutional neural network for modeling quantum interactions," in *Proceedings of the Advances in Neural Information Processing Systems* (2017), Vol. 30.
[138]N. Thomas, T. E. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, "Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds," arXiv:1802.08219 (2018).
[139]S. Villar, D. W. Hogg, K. Storey-Fisher, W. Yao, and B. Blum-Smith, "Scalars are universal: Equivariant machine learning, structured like classical physics," in *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)* (2021), Vol. 34, pp. 28848–28863.
[140]A. Vaswani, "Attention is all you need," in *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)* (2017).
[141]E. Albert, W. Perrett, and G. Jeffery, "The foundation of the general theory of relativity," Ann. Phys. **354**, 769 (1916).
[142]A. Ullah and P. O. Dral, "Molecular quantum chemical data sets and databases for machine learning potentials," arXiv:2408.12058 (2024).
[143]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865 (1996).
[144]A. Tkatchenko, R. A. DiStasio, Jr., R. Car, and M. Scheffler, "Accurate and efficient method for many-body van der Waals interactions," Phys. Rev. Lett. **108**, 236402 (2012).
[145]R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," Sci. Data **1**, 1–7 (2014).
[146]J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio, and A. Tkatchenko, "QM7-X, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules," Sci. Data **8**, 43 (2021).

[147]J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules," Sci. Data **4**(1), 8 (2017c).
[148]A. S. Christensen, S. K. Sirumalla, Z. Qiao, M. B. O'Connor, D. G. A. Smith, F. Ding, P. J. Bygrave, A. Anandkumar, M. Welborn, F. R. Manby, and T. F. Miller III, "OrbNet Denali training data" (2021a).
[149]A. S. Christensen, S. K. Sirumalla, Z. Qiao, M. B. O'Connor, D. G. A. Smith, F. Ding, P. J. Bygrave, A. Anandkumar, M. Welborn, F. R. Manby, and T. F. Miller, "OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy," J. Chem. Phys. **155**, 204103 (2021b).
[150]C. N. Cavasotto, "Binding free energy calculation using quantum mechanics aimed for drug lead optimization," in *Quantum Mechanics in Drug Discovery* (Humana, 2020), pp. 257–268.
[151]A. E. A. Allen, N. Lubbers, S. Matin, J. Smith, R. Messerly, S. Tretiak, and K. Barros, "Learning together: Towards foundational models for machine learning interatomic potentials with meta-learning," arXiv:2307.04012 (2023).
[152]A. S. Christensen and O. A. von Lilienfeld, "On the role of gradients for machine learning of molecular energies and forces," arXiv:2007.09593 (2020b).
[153]D. P. Kovács, C. v d Oord, J. Kucera, A. E. Allen, D. J. Cole, C. Ortner, and G. Csányi, "Linear atomic cluster expansion force fields for organic molecules: Beyond RMSE," J. Chem. Theory Comput. **17**, 7696–7711 (2021).
[154]I. Batatia, Batzner, D. P. Kovács, A. Musaelian, G. N. C. Simm, R. Drautz, C. Ortner, B. Kozinsky, and G. Csányi, "The design space of E(3)-equivariant atom-centered interatomic potentials," arXiv:2205.06643 (2022b).
[155]P. Eastman, P. K. Behara, D. Dotson, R. Galvelis, J. Herr, J. Horton, Y. Mao, J. Chodera, B. Pritchard, Y. Wang, G. De Fabritiis, and T. Markland, "Spice 2.0.1" (2024a).
[156]D. G. Smith, D. Altarawy, L. A. Burns, M. Welborn, L. N. Naden, L. Ward, S. Ellis, B. P. Pritchard, and T. D. Crawford, "The MoLSSI QCArchive project: An open-source platform to compute, organize, and share quantum chemistry data," WIREs. Comput. Mol. Sci. **11**, e1491 (2021).
[157]J. A. Vita, E. G. Fuemmeler, A. Gupta, G. P. Wolfe, A. Q. Tao, R. S. Elliott, S. Martiniani, and E. B. Tadmor, "ColabFit exchange: Open-access datasets for data-driven interatomic potentials," arXiv:2306.11071 (2023).
[158]Y. Yang, S. Zhang, K. D. Ranasinghe, O. Isayev, and A. E. Roitberg, "Machine learning of reactive potentials," Annu. Rev. Phys. Chem. **75**, 371–395 (2024).
[159]U. Rivero, O. T. Unke, M. Meuwly, and S. Willitsch, "Reactive atomistic simulations of Diels-Alder reactions: The importance of molecular rotations," J. Chem. Phys. **151**, 104301 (2019).
[160]S. Zhang, M. Z. Makoś, R. B. Jadrich, E. Kraka, K. Barros, B. T. Nebgen, S. Tretiak, O. Isayev, N. Lubbers, R. A. Messerly *et al.*, "Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential," Nat. Chem. **16**, 727–734 (2024).
[161]M. Gastegger and P. Marquetand, "High-dimensional neural network potentials for organic reactions and an improved training algorithm," J. Chem. Theory Comput. **11**, 2187–2198 (2015).
[162]M. Schreiner, A. Bhowmik, T. Vegge, J. Busk, and O. Winther, "Transition1x—A dataset for building generalizable reactive machine learning potentials," Sci. Data **9**, 779 (2022).
[163]T. A. Young, T. Johnston-Wood, H. Zhang, and F. Duarte, "Reaction dynamics of Diels–Alder reactions from machine learned potentials," Phys. Chem. Chem. Phys. **24**, 20820–20827 (2022).
[164]X. Pan, J. Yang, R. Van, E. Epifanovsky, J. Ho, J. Huang, J. Pu, Y. Mei, K. Nam, and Y. Shao, "Machine-learning-assisted free energy simulation of solution-phase and enzyme reactions," J. Chem. Theory Comput. **17**, 5745–5758 (2021).
[165]T. Devergne, T. Magrino, F. Pietrucci, and A. M. Saitta, "Combining machine learning approaches and accurate ab initio enhanced sampling methods for prebiotic chemical reactions in solution," J. Chem. Theory Comput. **18**, 5410–5421 (2022).
[166]M. Yang, L. Bonati, D. Polino, and M. Parrinello, "Using metadynamics to build neural network potentials for reactive events: The case of urea decomposition in water," Catal. Today **387**, 143–149 (2022).
[167]Z. Benayad, R. David, and G. Stirnemann, "Prebiotic chemical reactivity in solution with quantum accuracy and microsecond sampling using neural network potentials," Proc. Natl. Acad. Sci. U. S. A. **121**, e2322040121 (2024).
[168]I.-B. Magdău, D. J. Arismendi-Arrieta, H. E. Smith, C. P. Grey, K. Hermansson, and G. Csányi, "Machine learning force fields for molecular liquids: Ethylene carbonate/ethyl methyl carbonate binary solvent," npj Comput. Mater. **9**, 146 (2023).
[169]S. Boothroyd, L.-P. Wang, D. L. Mobley, J. D. Chodera, and M. R. Shirts, "Open force field evaluator: An automated, efficient, and scalable framework for the estimation of physical properties from molecular simulation," J. Chem. Theory Comput. **18**, 3566–3576 (2022).
[170]C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," Nature **585**, 357–362 (2020).
[171]J. R. Maple, M.-J. Hwang, T. P. Stockfisch, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler, "Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules," J. Comput. Chem. **15**, 162–182 (1994a).
[172]M. J. Hwang, T. Stockfisch, and A. Hagler, "Derivation of class II force fields. 2. Derivation and characterization of a class II force field, CFF93, for the alkyl functional group and alkane molecules," J. Am. Chem. Soc. **116**, 2515–2525 (1994).
[173]J. Maple, M.-J. Hwang, T. Stockfisch, and A. Hagler, "Derivation of class II force fields. III. Characterization of a quantum force field for alkanes," Isr. J. Chem. **34**, 195–231 (1994b).
[174]S. R. Xie, M. Rupp, and R. G. Hennig, "Ultra-fast interpretable machine-learning potentials," npj Comput. Mater. **9**, 162 (2023).
[175]J. A. Lemkul, J. Huang, B. Roux, and A. D. J. MacKerell, "An empirical polarizable force field based on the classical Drude oscillator model: Development history and recent applications," Chem. Rev. **116**, 4983–5013 (2016).
[176]A. P. Jones, J. Crain, V. P. Sokhan, T. W. Whitfield, and G. J. Martyna, "Quantum Drude oscillator model of atoms and molecules: Many-body polarization and dispersion interactions for atomistic simulation," Phys. Rev. B **87**, 144103 (2013).
[177]Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, "Polarizable atomic multipole-based amoeba force field for proteins," J. Chem. Theory Comput. **9**, 4046–4063 (2013).
[178]A. Illarionov, S. Sakipov, L. Pereyaslavets, I. V. Kurnikov, G. Kamath, O. Butin, E. Voronina, I. Ivahnenko, I. Leontyev, G. Nawrocki, M. Darkhovskiy, M. Olevanov, Y. K. Cherniavskyi, C. Lock, S. Greenslade, S. K. Sankaranarayanan, M. G. Kurnikova, J. Potoff, R. D. Kornberg, M. Levitt, and B. Fain, "Combining force fields and neural networks for an accurate representation of chemically diverse molecular interactions," J. Am. Chem. Soc. **145**, 23620–23629 (2023).
[179]A. C. T. v Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "ReaxFF: A reactive force field for hydrocarbons," J. Phys. Chem. A **105**, 9396–9409 (2001).
[180]M. C. Kaymak, A. Rahnamoun, K. A. O'Hearn, A. C. T. v. Duin, K. M. J. Merz, and H. M. Aktulga, "JAX-ReaxFF: A gradient-based framework for fast optimization of reactive force fields," J. Chem. Theory Comput. **18**, 5181–5194 (2022).
[181]A. Warshel and R. M. Weiss, "An empirical valence bond approach for comparing reactions in solutions and in enzymes," J. Am. Chem. Soc. **102**, 6218–6226 (1980).
[182]J. Lobaugh and G. A. Voth, "The quantum dynamics of an excess proton in water," J. Chem. Phys. **104**, 2056–2069 (1996).
[183]D. E. Sagnella and M. E. Tuckerman, "An empirical valence bond model for proton transfer in water," J. Chem. Phys. **108**, 2073–2083 (1998).
[184]U. W. Schmitt and G. A. Voth, "Multistate empirical valence bond model for proton transport in water," J. Phys. Chem. B **102**, 5547–5551 (1998).
[185]J. Åqvist and A. Warshel, "Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches," Chem. Rev. **93**, 2523–2544 (1993).
[186]A. E. A. Allen and G. Csányi, "Toward transferable empirical valence bonds: Making classical force fields reactive," J. Chem. Phys. **160**, 124108 (2024).

[187]L.-P. Wang, J. Chen, and T. Van Voorhis, "Systematic parametrization of polarizable force fields from quantum chemistry data," J. Chem. Theory Comput. **9**, 452–460 (2013).
[188]L.-P. Wang, T. J. Martinez, and V. S. Pande, "Building force fields: An automatic, systematic, and reproducible approach," J. Phys. Chem. Lett. **5**, 1885–1891 (2014).
[189]S. Thaler and J. Zavadlav, "Learning neural network potentials from experimental data via differentiable trajectory reweighting," Nat. Commun. **12**, 6884 (2021).
[190]B. J. Befort, R. S. DeFever, G. M. Tow, A. W. Dowling, and E. J. Maginn, "Machine learning directed optimization of classical molecular modeling force fields," J. Chem. Inf. Model. **61**, 4400–4414 (2021).
[191]X. Wang, J. Li, L. Yang, F. Chen, Y. Wang, J. Chang, J. Chen, W. Feng, L. Zhang, and K. Yu, "DMFF: An open-source automatic differentiable platform for molecular force field development and molecular dynamics simulation," J. Chem. Theory Comput. **19**, 5897–5909 (2023).
[192]D. L. Mobley, C. C. Bannan, A. Rizzi, C. I. Bayly, J. D. Chodera, V. T. Lim, N. M. Lim, K. A. Beauchamp, D. R. Slochower, M. R. Shirts *et al.*, "Escaping atom types in force fields using direct chemical perception," J. Chem. Theory Comput. **14**, 6076–6092 (2018).
[193]L. Wang, P. K. Behara, M. W. Thompson, T. Gokey, Y. Wang, J. R. Wagner, D. J. Cole, M. K. Gilson, M. R. Shirts, and D. L. Mobley, "The open force field initiative: Open software and open science for molecular modeling," J. Phys. Chem. B **128**, 7043 (2024a).
[194]K. B. Koziara, M. Stroet, A. K. Malde, and A. E. Mark, "Testing and validation of the automated topology builder (ATB) version 2.0: Prediction of hydration free enthalpies," J. Comput.-Aided. Mol. Des. **28**, 221–233 (2014).
[195]R. M. Betz and R. C. Walker, "Paramfit: Automated optimization of force field parameters for molecular dynamics simulations," J. Comput. Chem. **36**, 79–87 (2015).
[196]E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight *et al.*, "OPLS3: A force field providing broad coverage of drug-like small molecules and proteins," J. Chem. Theory Comput. **12**, 281–296 (2016).
[197]J. T. Horton, S. Boothroyd, J. Wagner, J. A. Mitchell, T. Gokey, D. L. Dotson, P. K. Behara, V. K. Ramaswamy, M. Mackey, J. D. Chodera, J. Anwar, D. L. Mobley, and D. J. Cole, "Open force field bespokefit: Automating bespoke torsion parametrization at scale," J. Chem. Inf. Model. **62**, 5622–5633 (2022).
[198]A. Kumar and A. D. J. MacKerell, "FFParam-v2.0: A comprehensive tool for CHARMM additive and Drude polarizable force-field parameter optimization and validation," J. Phys. Chem. B **128**, 4385–4395 (2024).
[199]J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb," J. Chem. Theory Comput. **11**, 3696–3713 (2015).
[200]M. Zgarbová, J. Sponer, M. Otyepka, T. E. Cheatham, I. I. I. R. Galindo-Murillo, and P. Jurecka, "Refinement of the sugar–phosphate backbone torsion beta for amber force fields improves the description of z-and b-dna," J. Chem. Theory Comput. **11**, 5723–5736 (2015).
[201]R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, and T. E. Cheatham III, "Assessing the current state of amber force field modifications for DNA," J. Chem. Theory Comput. **12**, 4114–4127 (2016).
[202]M. Zgarbová, M. Otyepka, J. Šponer, A. Mládek, P. Banáš, T. E. Cheatham, and I. I. I. P. Jurečka, "Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles," J. Chem. Theory Comput. **7**, 2886–2902 (2011).
[203]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926–935 (1983).
[204]H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew," J. Chem. Phys. **120**, 9665–9678 (2004).
[205]S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building water models: A different approach," J. Phys. Chem. Lett. **5**, 3863–3871 (2014).
[206]I. S. Joung and T. E. Cheatham III, "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations," J. Phys. Chem. B **112**, 9020–9041 (2008).
[207]I. S. Joung and T. E. Cheatham III, "Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters," J. Phys. Chem. B **113**, 13279–13290 (2009).
[208]P. Li, B. P. Roberts, D. K. Chakravorty, and K. M. Merz, Jr., "Rational design of particle mesh Ewald compatible Lennard-Jones parameters for+ 2 metal cations in explicit solvent," J. Chem. Theory Comput. **9**, 2733–2748 (2013).
[209]P. Li and K. M. Merz, Jr, "Taking into account the ion-induced dipole interaction in the nonbonded model of ions," J. Chem. Theory Comput. **10**, 289–297 (2014).
[210]P. Li, L. F. Song, and K. M. Merz, Jr., "Parameterization of highly charged metal ions using the 12-6-4 lj-type nonbonded model in explicit water," J. Phys. Chem. B **119**, 883–895 (2015).
[211]C. J. Dickson, R. C. Walker, and I. R. Gould, "Lipid21: Complex lipid membrane simulations with AMBER," J. Chem. Theory Comput. **18**, 1726–1736 (2022).
[212]K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, and R. J. Woods, "GLYCAM06: A generalizable biomolecular force field. carbohydrates," J. Comput. Chem. **29**, 622–655 (2008).
[213]M. L. DeMarco and R. J. Woods, "Atomic-resolution conformational analysis of the G_{M3} ganglioside in a lipid bilayer and its implications for ganglioside–protein recognition at membrane surfaces," Glycobiology **19**, 344–355 (2008).
[214]M. L. DeMarco, R. J. Woods, J. H. Prestegard, and F. Tian, "Presentation of membrane-anchored glycosphingolipids determined from molecular dynamics simulations and NMR paramagnetic relaxation rate enhancement," J. Am. Chem. Soc. **132**, 1334–1338 (2010).
[215]J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," J. Comput. Chem. **25**, 1157–1174 (2004).
[216]J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," J. Mol. Graph. Modell. **25**, 247–260 (2006).
[217]G. A. Khoury, J. P. Thompson, J. Smadbeck, C. A. Kieslich, and C. A. Floudas, "Forcefield_PTM: Ab initio charge and AMBER forcefield parameters for frequently occurring post-translational modifications," J. Chem. Theory Comput. **9**, 5653–5674 (2013).
[218]R. Aduri, B. T. Psciuk, P. Saro, H. Taniga, H. B. Schlegel, and J. SantaLucia, "AMBER force field parameters for the naturally occurring modified nucleosides in RNA," J. Chem. Theory Comput. **3**, 1464–1475 (2007).
[219]M. Stroet, B. Caron, M. S. Engler, J. v. d. Woning, A. Kauffmann, M. v. Dijk, M. El-Kebir, K. M. Visscher, J. Holownia, C. Macfarlane, B. J. Bennion, S. Gelpi-Dominguez, F. C. Lightstone, T. v. d. Storm, D. P. Geerke, A. E. Mark, and G. W. Klau, "OFraMP: A fragment-based tool to facilitate the parametrization of large molecules," J. Comput.-Aided. Mol. Des. **37**, 357–371 (2023).
[220]J. D. Yesselman, D. J. Price, J. L. Knight, and C. L. Brooks III, "Match: An atom-typing toolset for molecular mechanics force fields," J. Comput. Chem. **33**, 189–202 (2012).
[221]J. Wagner, M. Thompson, D. L. Mobley, J. Chodera, C. Bannan, A. Rizzi, Trevorgokey, D. L. Dotson, J. A. Mitchell, Jaimergp, Camila, P. Behara, C. Bayly, JoshHorton, I. Pulido, L. Wang, V. Lim, S. Sasmal, SimonBoothroyd, A. Dalke, D. Smith, B. Westbrook, J. Horton, L.-P. Wang, R. Gowers, Z. Zhao, C. Davel, and Y. Zhao, "openforcefield/openff-toolkit: 0.15.2 minor feature release" (2024).
[222]A. R. McIsaac, P. K. Behara, T. Gokey, C. Cavender, J. Horton, L. Wang, B. R. Westbrook, M. W. Thompson, M. Osato, H. M. Baumann, H. Jang, J. Wagne, D. Cole, C. Bayly, and D. Mobley, "openforcefield/openff-forcefields" (2024).
[223]K. Roos, C. Wu, W. Damm, M. Reboul, J. M. Stevenson, C. Lu, M. K. Dahlgren, S. Mondal, W. Chen, L. Wang, R. Abel, R. A. Friesner, and E. D. Harder, "Opls3e: Extending force field coverage for drug-like small molecules," J. Chem. Theory Comput. **15**, 1863–1874 (2019).
[224]T. Gokey and D. L. Mobley, "Hierarchical clustering of chemical space using binary-encoded SMARTS for building data-driven chemical perception models," J. Chem. Theory Comput. **19**, 3251–3275 (2023).
[225]Y. Wang, J. Fass, B. Kaminow, J. E. Herr, D. Rufa, I. Zhang, I. Pulido, M. Henry, H. E. Bruce Macdonald, K. Takaba, and J. D. Chodera, "End-to-end

differentiable construction of molecular mechanics force fields," Chem. Sci. **13**, 12016–12033 (2022).
[226]G. Chen, T. Jaffrelot Inizan, T. Plé, L. Lagardère, J.-P. Piquemal, and Y. Maday, "Advancing force fields parameterization: A directed graph attention networks approach," J. Chem. Theory Comput. **20**, 5558–5569 (2024).
[227]L. Seute, E. Hartmann, J. Stühmer, and G. Frauke, "Grappa—A machine learned molecular mechanics force field," arXiv:2404.00050 (2024).
[228]M. Thurlemann, L. Boselt, and S. Riniker, "Regularized by physics: Graph neural network parametrized potentials for the description of intermolecular interactions," J. Chem. Theory Comput. **19**, 562–579 (2023).
[229]M. T. Lehner, P. Katzberger, N. Maeder, C. C. Schiebroek, J. Teetz, G. A. Landrum, and S. Riniker, "DASH: Dynamic attention-based substructure hierarchy for partial charge assignment," J. Chem. Inf. Model. **63**, 6014–6028 (2023).
[230]Y. Wang, I. Pulido, K. Takaba, B. Kaminow, J. Scheen, L. Wang, and J. D. Chodera, "EspalomaCharge: Machine learning-enabled ultrafast partial charge assignment," J. Phys. Chem. A **128**, 4160–4167 (2024b).
[231]Y. Wang, J. Fass, C. D. Stern, K. Luo, and J. Chodera, "Graph nets for partial charge prediction," arXiv:1909.07903 (2019c).
[232]Y. Wang, *Graph Machine Learning for (Bio) Molecular Modeling and Force Field Construction* (Weill Medical College of Cornell University, 2023).
[233]S. Passaro and C. L. Zitnick, "Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs," arXiv:2302.03655 (2023).
[234]S. Luo, T. Chen, and A. S. Krishnapriyan, "Enabling efficient equivariant operations in the Fourier basis via gaunt tensor products," arXiv:2401.10216 (2024).
[235]B. Cheng, "Cartesian atomic cluster expansion for machine learning interatomic potentials," npj Comput. Mater. **10**, 157 (2024).
[236]B. Smit, P. Hilbers, K. Esselink, L. Rupert, N. Van Os, and A. Schlijper, "Computer simulations of a water/oil interface in the presence of micelles," Nature **348**, 624–625 (1990).
[237]M. Müller, K. Katsov, and M. Schick, "Biological and synthetic membranes: What can be learned from a coarse-grained description?," Phys. Rep. **434**, 113–176 (2006).
[238]S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries, "The martini force field: Coarse grained model for biomolecular simulations," J. Phys. Chem. B **111**, 7812–7824 (2007).
[239]P. C. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. Bruininks, T. A. Wassenaar *et al.*, "Martini 3: A general purpose force field for coarse-grained molecular dynamics," Nat. Methods **18**, 382–388 (2021).
[240]J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. De Fabritiis, F. Noé, and C. Clementi, "Machine learning of coarse-grained molecular dynamics force fields," ACS Cent. Sci. **5**, 755–767 (2019d).
[241]W. Wang and R. Gómez-Bombarelli, "Coarse-graining auto-encoders for molecular dynamics," npj Comput. Mater. **5**, 125 (2019).
[242]S. Yang and R. Gómez-Bombarelli, "Chemically transferable generative backmapping of coarse-grained proteins," arXiv:2303.01569 (2023).
[243]S. Wang, B. Z. Li, M. Khabsa, H. Fang, and H. Ma, "Linformer: Self-attention with linear complexity," arXiv:2006.04768 (2020c).
[244]J. D. Gale, L. M. LeBlanc, P. R. Spackman, A. Silvestri, and P. Raiteri, "A universal force field for materials, periodic GFN-FF: Implementation and examination," J. Chem. Theory Comput. **17**, 7827–7849 (2021).
[245]M. R. Gunner, T. Murakami, A. S. Rustenburg, M. Işık, and J. D. Chodera, "Standard state free energies, not pKas, are ideal for describing small molecule protonation and tautomeric states," J. Comput.-Aided. Mol. Des. **34**, 561–573 (2020).
[246]P. Eastman, B. P. Pritchard, J. D. Chodera, and T. E. Markland, "Nutmeg and spice: Models and data for biomolecular machine learning," arXiv:2406.13112 (2024b).
[247]W. Huang, J. Han, Y. Rong, T. Xu, F. Sun, and J. Huang, "Equivariant graph mechanics networks with constraints," arXiv:2203.06442 (2022).
[248]C. Qu, Q. Yu, R. Conte, P. L. Houston, A. Nandi, and J. M. Bomwan, "A δ-machine learning approach for force fields, illustrated by a CCSD(T) 4-body correction to the MB-pol water potential," Digital Discovery **1**, 658–664 (2022).
[249]T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, "A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer," Nat. Commun. **12**, 1–11 (2021).
[250]S. M. Larson, C. D. Snow, M. Shirts, and V. S. Pande, "Folding@home and Genome@home: Using distributed computing to tackle previously intractable problems in computational biology," arXiv:0901.0866 (2009).
[251]M. Thürlemann, L. Böselt, and S. Riniker, "Learning atomic multipoles: Prediction of the electrostatic potential with equivariant graph neural networks," J. Chem. Theory Comput. **18**, 1701–1710 (2022).
[252]R. E. Duke, O. N. Starovoytov, J.-P. Piquemal, and G. A. Cisneros, "GEM∗: A molecular electronic density-based force field for molecular dynamics simulations," J. Chem. Theory Comput. **10**, 1361–1365 (2014).
[253]A. K. Rappe and W. A. Goddard III, "Charge equilibration for molecular dynamics simulations," J. Phys. Chem. **95**, 3358–3363 (1991).
[254]L. Böselt, M. Thürlemann, and S. Riniker, "Machine learning in qm/mm molecular dynamics simulations of condensed-phase systems," J. Chem. Theory Comput. **17**, 2641–2658 (2021).
[255]H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems," Angew. Chem. Int. Ed. Engl. **48**, 1198–1229 (2009).
[256]D. J. Cole, L. Mones, and G. Csányi, "A machine learning based intramolecular potential for a flexible organic molecule," Faraday Discuss. **224**, 247–264 (2020).
[257]J. Karwounopoulos, Z. Wu, S. Tkaczyk, S. Wang, A. Baskerville, K. Ranasinghe, T. Langer, G. P. Wood, M. Wieder, and S. Boresch, "Insights and challenges in correcting force field based solvation free energies using a neural network potential," J. Phys. Chem. B **128**, 6693–6703 (2024).
[258]B. Lier, P. Poliak, P. Marquetand, J. Westermayr, and C. Oostenbrink, "BuRNN: Buffer region neural network approach for polarizable-embedding neural network/molecular mechanics simulations," J. Phys. Chem. Lett. **13**, 3812 (2022).
[259]A. W. H. Grubmüller, H. Heller, and K. Schulten, "Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions," Mol. Simul. **6**, 121–142 (1991).
[260]M. Tuckerman, B. J. Berne, and G. J. Martyna, "Reversible multiple time scale molecular dynamics," J. Chem. Phys. **97**, 1990–2001 (1992).
[261]D. T. W. B. Streett and G. Saville, "Multiple time-step methods in molecular dynamics," Mol. Phys. **35**, 639–648 (1978).
[262]P. Minary, M. E. Tuckerman, and G. J. Martyna, "Long time molecular dynamics for enhanced conformational sampling in biomolecular systems," Phys. Rev. Lett. **93**, 150201 (2004).
[263]P. F. Batcho, D. A. Case, and T. Schlick, "Optimized particle-mesh Ewald/multiple-time step integration for molecular dynamics simulations," J. Chem. Phys. **115**, 4003–4018 (2001).
[264]C. R. A. Abreu and M. E. Tuckerman, "Hamiltonian based resonance-free approach for enabling very large time steps in multiple time-scale molecular dynamics," Mol. Phys. **119**, e1923848 (2021).
[265]M. Guidon, F. Schiffmann, J. Hutter, and J. VandeVondele, "Ab initio molecular dynamics using hybrid density functionals," J. Chem. Phys. **128**, 214104 (2008).
[266]R. P. Steele, "Communication: Multiple-timestep ab initio molecular dynamics with electron correlation," J. Chem. Phys. **139**, 011102 (2013).
[267]N. Luehr, T. E. Markland, and T. J. Martínez, "Multiple time step integrators in ab initio molecular dynamics," J. Chem. Phys. **140**, 084116 (2014).
[268]Z. Long, A. O. Atsango, J. A. Napoli, T. E. Markland, and M. E. Tuckerman, "Elucidating the proton transport pathways in liquid imidazole with first-principles molecular dynamics," J. Phys. Chem. Lett. **11**, 6156–6163 (2020).
[269]W. Wang, S. Axelrod, and R. Gómez-Bombarelli, "Differentiable molecular simulations for control and learning," arXiv:2003.00868 (2020).
[270]J. G. Greener and D. T. Jones, "Differentiable molecular simulation can learn all the parameters in a coarse-grained force field for proteins," PLoS One **16**, e0256990 (2021).
[271]P. Kidger, R. T. Q. Chen, and T. J. Lyons, "'hey, that's not an ode': Faster ode adjoints via seminorms," in Proceedings of the International Conference on Machine Learning (2021).
[272]R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," in *Proceedings of the Advances in Neural Information Processing Systems* (2018).
[273]L. S. Pontryagin, *Mathematical Theory of Optimal Processes* (Routledge, 2018).
[274]R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," Mach. Learn. **8**, 229–256 (1992).

[275]J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv:1707.06347 (2017).
[276]S. S. Schoenholz and E. D. Cubuk, "JAX M.D. A framework for differentiable physics," in *Proceedings of the Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2020), Vol. 33.
[277]S. Doerr, M. Majewski, A. Pérez, A. Kramer, C. Clementi, F. Noe, T. Giorgino, and G. De Fabritiis, "TorchMD: A deep learning framework for molecular simulations," J. Chem. Theory Comput. **17**, 2355–2363 (2021).
[278]J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Neubig, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: Composable transformations of Python+NumPy programs," GitHub repository (2018). Available at https://github.com/google/jax.
[279]A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proceedings of the Adv*ances in *Neural Information Processing Systems* (2019), Vol. 32.
[280]P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh, and H. Wu, "Mixed precision training of deep neural networks," NVIDIA Developer Blog (2018). Available at https://developer.nvidia.com/blog/mixed-precision-training-deep-neural-networks/.
[281]See https://github.com/proteneer/timemachine for Y. Zhao, "timemachine" (2024).
[282]R. O. Dror, C. Young, and D. E. Shaw, "Anton, a special-purpose molecular simulation machine" (2011).
[283]D. E. Shaw, P. J. Adams, A. Azaria, J. A. Bank, B. Batson, A. Bell, M. Bergdorf, J. Bhatt, J. A. Butts, T. Correia *et al.*, "Anton 3: Twenty microseconds of molecular dynamics simulation before lunch," in Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (2021), pp. 1–11.
[284]R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, E. Brynjolfsson, S. Buch, D. Card, R. Castellon, N. Chatterji, A. Chen, K. Creel, J. Q. Davis, D. Demszky, C. Donahue, M. Doumbouya, E. Durmus, S. Ermon, J. Etchemendy, K. Ethayarajh, L. Fei-Fei, C. Finn, T. Gale, L. Gillespie, K. Goel, N. Goodman, S. Grossman, N. Guha, T. Hashimoto, P. Henderson, J. Hewitt, D. E. Ho, J. Hong, K. Hsu, J. Huang, T. Icard, S. Jain, D. Jurafsky, P. Kalluri, S. Karamcheti, G. Keeling, F. Khani, O. Khattab, P. W. Koh, M. Krass, R. Krishna, R. Kuditipudi, A. Kumar, F. Ladhak, M. Lee, T. Lee, J. Leskovec, I. Levent, X. L. Li, X. Li, T. Ma, A. Malik, C. D. Manning, S. Mirchandani, E. Mitchell, Z. Munyikwa, S. Nair, A. Narayan, D. Narayanan, B. Newman, A. Nie, J. C. Niebles, H. Nilforoshan, J. Nyarko, G. Ogut, L. Orr, I. Papadimitriou, J. S. Park, C. Piech, E. Portelance, C. Potts, A. Raghunathan, R. Reich, H. Ren, F. Rong, Y. Roohani, C. Ruiz, J. Ryan, C. Ré, D. Sadigh, S. Sagawa, K. Santhanam, A. Shih, K. Srinivasan, A. Tamkin, R. Taori, A. W. Thomas, F. Tramèr, R. E. Wang, W. Wang, B. Wu, J. Wu, Y. Wu, S. M. Xie, M. Yasunaga, J. You, M. Zaharia, M. Zhang, T. Zhang, X. Zhang, Y. Zhang, L. Zheng, K. Zhou, and P. Liang, "On the opportunities and risks of foundation models," arXiv:2108.07258 (2022).
[285]L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, and Z. Ulissi, "Open catalyst 2020 (OC20) dataset and community challenges," ACS Catal. **11**, 6059–6072 (2021).
[286]A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder *et al.*, "Commentary: The materials project: A materials genome approach to accelerating materials innovation," APL Mater. **1**, 011002 (2013).
[287]L. Barroso-Luque, M. Shuaibi, X. Fu, B. M. Wood, M. Dzamba, M. Gao, A. Rizvi, C. L. Zitnick, and Z. W. Ulissi, "Open materials 2024 (omat24) inorganic materials dataset and models," arXiv:2410.12771 (2024).
[288]C. Zang and F. Wang, "Moflow: An invertible flow model for generating molecular graphs," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery; Data Mining, KDD'20* (ACM, 2020).
[289]W. Jin, R. Barzilay, and T. Jaakkola, "Junction tree variational autoencoder for molecular graph generation," arXiv:1802.04364 (2019).
[290]C. Shi, M. Xu, Z. Zhu, W. Zhang, M. Zhang, and J. Tang, "GraphAF: A flow-based autoregressive model for molecular graph generation," arXiv:2001.09382 (2020).
[291]E. Mansimov, O. Mahmood, S. Kang, and K. Cho, "Molecular geometry prediction using a deep generative graph neural network," Sci. Rep. **9**, 20381 (2019).
[292]E. Hoogeboom, V. G. Satorras, C. Vignac, and M. Welling, "Equivariant diffusion for molecule generation in 3D," arXiv:2203.17003 (2022).
[293]J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko *et al.*, "Highly accurate protein structure prediction with alphafold," Nature **596**, 583–589 (2021).
[294]G. Ahdritz, N. Bouatta, C. Floristean, S. Kadyan, Q. Xia, W. Gerecke, T. J. O'Donnell, D. Berenberg, I. Fisk, N. Zanichelli *et al.*, "OpenFold: Retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization," Nat. Methods **21**, 1514 (2024).
[295]L. Klein, A. Y. K. Foong, T. E. Fjelde, B. Mlodozeniec, M. Brockschmidt, S. Nowozin, F. Noé, and R. Tomioka, "Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics," arXiv:2302.01170 (2023).
[296]F. Noé, S. Olsson, J. Köhler, and H. Wu, "Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning," Science **365**, eaaw1147 (2019).
[297]M. E. Tuckerman, "Machine learning transforms how microstates are sampled," Science **365**, 982–983 (2019).
[298]L. Klein and F. Noé, "Transferable boltzmann generators," arXiv:2406.14426 (2024).
[299]J. Köhler, L. Klein, and F. Noé, "Equivariant flows: Exact likelihood generative learning for symmetric densities," in *Proceedings of the International Conference on Machine Learning* (PMLR, 2020), pp. 5361–5370.
[300]L. I. Midgley, V. Stimper, J. Antorán, E. Mathieu, B. Schölkopf, and J. M. Hernández-Lobato, "SE(3) equivariant augmented coupling flows," arXiv:2308.10364 (2024).
[301]V. G. Satorras, E. Hoogeboom, F. B. Fuchs, I. Posner, and M. Welling, "E(N) equivariant normalizing flows," arXiv:2105.09016 (2022).
[302]Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, F. Huang *et al.*, "A tutorial on energy-based learning," Predicting structured data **1** (2006).
[303]Y. Song and D. P. Kingma, "How to train your energy-based models," arXiv:2101.03288 (2021).
[304]F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," IEEE Trans. Pattern Anal. Mach. Intell. **45**, 10850–10869 (2023).
[305]L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion models: A comprehensive survey of methods and applications," ACM Comput. Surv. **56**, 1–39 (2023).
[306]Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," arXiv:2011.13456 (2021).
[307]J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," arXiv:2006.11239 (2020).
[308]K. Kahouli, S. S. P. Hessmann, K.-R. Müller, S. Nakajima, S. Gugler, and N. W. A. Gebauer, "Molecular relaxation by reverse diffusion with time step prediction," Mach. Learn.: Sci. Technol. **5**, 035038 (2024).
[309]M. Arts, V. G. Satorras, C.-W. Huang, D. Zuegner, M. Federici, C. Clementi, F. Noé, R. Pinsler, and R. van den Berg, "Two for one: Diffusion models and force fields for coarse-grained molecular dynamics," arXiv:2302.00600 (2023).