Molecular force fields with gradient-domain machine learning (GDML): Comparison and synergies with classical force fields (EP)

Huziel E. Sauceda (iD), Michael Gastegger (iD), Stefan Chmiela (iD), Klaus-Robert Müller (iD), and Alexandre Tkatchenko (iD)

COLLECTIONS

Note: This paper is part of the JCP Special Topic on Machine Learning Meets Chemical Physics.

(EP) This paper was selected as an Editor's Pick

View Online   Export Citation   CrossMark

Molecular force fields with gradient-domain machine learning (GDML): Comparison and synergies with classical force fields ⓔⓟ

Huziel E. Sauceda,[1,2,3,a] ⓘⒹ Michael Gastegger,[2,3,4] ⓘⒹ Stefan Chmiela,[2] ⓘⒹ Klaus-Robert Müller,[2,5,6,7,a] ⓘⒹ
and Alexandre Tkatchenko[1,a] ⓘⒹ

AFFILIATIONS

[1] Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg, Luxembourg
[2] Machine Learning Group, Technische Universität Berlin, 10587 Berlin, Germany
[3] BASLEARN, BASF-TU Joint Lab, Technische Universität Berlin, 10587 Berlin, Germany
[4] DFG Cluster of Excellence "Unifying Systems in Catalysis" (UniSysCat), Technische Universität Berlin, 10623 Berlin, Germany
[5] Department of Artificial Intelligence, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, South Korea
[6] Max Planck Institute for Informatics, Stuhlsatzenhausweg, 66123 Saarbrücken, Germany
[7] Google Research, Brain Team, Berlin, Germany

Note: This paper is part of the JCP Special Topic on Machine Learning Meets Chemical Physics.
[a] **Authors to whom correspondence should be addressed:** sauceda@tu-berlin.de; klaus-robert.mueller@tu-berlin.de;
and alexandre.tkatchenko@uni.lu

ABSTRACT

Modern machine learning force fields (ML-FF) are able to yield energy and force predictions at the accuracy of high-level *ab initio* methods, but at a much lower computational cost. On the other hand, classical molecular mechanics force fields (MM-FF) employ fixed functional forms and tend to be less accurate, but considerably faster and transferable between molecules of the same class. In this work, we investigate how both approaches can complement each other. We contrast the ability of ML-FF for reconstructing dynamic and thermodynamic observables to MM-FFs in order to gain a qualitative understanding of the differences between the two approaches. This analysis enables us to modify the generalized AMBER force field by reparametrizing short-range and bonded interactions with more expressive terms to make them more accurate, without sacrificing the key properties that make MM-FFs so successful.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0023005

I. INTRODUCTION

Computational studies of materials and molecular systems at the atomic level of detail constitute a central tool in contemporary physics, biology, materials science, and chemistry research. Standard electronic-structure theories [e.g., density-functional theory (DFT) or wavefunction-based methods] are ultimately limited in the size of systems and time scales they can treat due to the high computational cost associated with modeling of the electronic interactions. Conventional molecular mechanics force fields (MM-FFs) overcome these limitations by introducing a range of computationally efficient approximations, circumventing the need to solve the electronic problem explicitly.[1–8] MM-FFs are developed to be transferable for a restricted set of homologous molecules and to reproduce condensed-phase properties under given thermodynamic conditions. Hence, their design is top-down rather than bottom-up. This has made it possible to simulate biologically relevant systems (such as proteins), gaining insights into their innermost workings at an atomistic level, culminating in a Nobel prize in chemistry awarded to Karplus, Levitt, and Warshel. However, the high computational efficiency of MM-FFs comes at the cost of accuracy compared to electronic-structure methods. Moreover,

the basic form of the physical approximations employed restricts their applicability to certain classes of problems, e.g., most of them are unable to model chemical reactions involving bond breaking/formation or do not include crucial intramolecular interactions[9] or many-body effects.[10]

Recently, modern machine learning methods (ML) have emerged as an elegant solution to this dilemma.[11–13] Due to the high functional flexibility of ML methods, they can be used to construct force fields (FFs) with the accuracy of electronic-structure calculations while retaining high computational efficiency albeit still behind conventional classical force fields. A plethora of ML based FFs has been reported within the past few years, demonstrating the general utility of this approach. In particular, a great amount of work has been done in FFs based on neural networks (NN)[14–24] and kernel-based models.[9,25–33] In order to understand better how these models work and to improve their performance, other ML studies have been done on molecular representations,[34–48] data sampling,[49–52] explanation methods,[21,53] and inference of molecular properties,[25,27,31,32,54–63] as well as software development.[64–66] Yet, although it is generally assumed that modern machine learning force fields (ML-FFs) have much better accuracy than MM-FFs, no in-depth comparisons [e.g., potential-energy surfaces (PES) and molecular dynamics predictions] have been performed up to date. Furthermore, such comparison raises the question of how the ML-FFs and MM-FFs could be combined to complement each other.

The goal of the present work is to perform a detailed investigation of the differences between both approaches based on a set of small molecules exhibiting different quantum–mechanical phenomena. Even though the accuracy of intra-molecular interactions is not the primary purpose of MM-FFs, understanding their successes and limitations at the molecular level offers a perspective for further refinement. Based on these results, different alternatives are explored for improving the data generation process and their applicability for expediting the force-field learning procedure. Furthermore, improvement of the accuracy for MM-FFs is studied by reparameterizing them based on more accurate reference data and testing their limits and functional form flexibility. For this task, we use the symmetric Gradient Domain Machine Learning (sGDML) framework[31,32] as the ML-FF of choice, as it is able to efficiently reconstruct the potential-energy surfaces (PES) of medium sized molecules. The investigated systems are the molecules ethanol, the keto form of malondialdehyde (keto-MDA), and acetylsalicylic acid (aspirin). In the context of these systems, we study the performance of MM-FFs and sGDML derived FFs based on the overall reliability of the generated PESs, as well as effects arising from chemical phenomena such as interactions between molecular orbitals. Although we restrict ourselves to the sGDML approach, it can nevertheless be expected that the results found here are equally valid for ML-FFs in general.

II. THEORY

Learning theory provides a framework to create a diverse collection of universal approximators capable of reconstructing any function. In practice, reconstructing PESs from highly accurate reference calculations such as coupled cluster theory (CC) is a challenging task given that the amount of available data is limited

by computational resources that such a level of theory requires. Consequently, mathematically constraining the solution space of a ML approximator by enforcing universal physical laws is highly advantageous. In this manner, a data-efficient model capable of delivering physically meaningful predictions can be obtained. Some of the basic desirable properties of high accuracy ML-FFs from the view of physics and chemistry are as follows:

A global description of the system. The nature of the quantum interactions resulting from the Schrödinger equation (within the Born–Oppenheimer approximation) $\mathcal{H}\Psi = V_{BO}\Psi$ and from the evaluation of the Hellmann–Feynman forces $-\mathbf{F} = \langle\Psi^*|\partial\mathcal{H}/\partial\mathbf{x}|\Psi\rangle$ is inherently many-body. In practice, such global description of the system has to be enforced by avoiding the non-unique partitioning of the total energy V_{BO} in atomic contributions.

Time invariance of the Hamiltonian. This constraint requires a time-invariant Hamiltonian $\mathcal{H} = \mathcal{T} + \hat{f}_E$, and hence, energy conservation must be enforced in the ML model, $\hat{\mathbf{f}}_\mathbf{F} = -\nabla\hat{f}_E$, yielding conservative forces.

Permutational invariance of identical atoms. In quantum mechanics, the square of the wave function is invariant to permuting two identical atoms in a molecule, a property that is carried over to the energy of the system: $V_{BO}(\ldots, \vec{x}_i, \ldots, \vec{x}_j, \ldots) = V_{BO}(\ldots, \vec{x}_j, \ldots, \vec{x}_i, \ldots)$. Fulfilling this spatial symmetry by the ML models often represents a big challenge, as well as in MM-FFs, since a key ingredient of ML models is to invariantly compare objects, here molecules. Thus, if no invariance can be guaranteed, then permutations typically give rise to noisy comparisons hampering consistent learning. The above mentioned physical properties constitute a set of mathematical constraints that narrow down the space of solution of the general ML approximator, thereby increasing the data efficiency and accuracy in the reconstruction of the original data generator. However, all these considerations, and the desire for a global model in particular, place limitations to the computational performance of such an approach.

A completely different strategy is pursued by molecular mechanical force fields, whose design follows chemical intuition and known physical approximations to intra- and intermolecular interactions. In this case, the globality of the model is replaced by a more scalable partitioning of the interactions in terms of interatomic pair distances, angles, and dihedral angles in local domains. The three-body (angles) and four-body (dihedral angles) terms are the ones responsible for the permutational invariance violation in MM-FFs, a problem that becomes less and less relevant as the size of the system increases from 100s to 1000s or even millions of atoms. On the other hand, these force fields are energy conserving by construction, since the molecular forces are derived as the derivative of a potential in the form of the interaction energy. This strategy has been proven to be highly successful over the years in many domains of molecular modeling.[2]

Before comparing these two contrasting methodologies, we briefly review both frameworks, which will be important for the interpretation and further improvement of both approaches.

A. sGDML model

The requirements mentioned above served as a basis for the *s*ymmetric *G*radient *D*omain *M*achine *L*earning (sGDML) FF.[32,67]

This framework consists in the use of a permutationally symmetrized Hessian of the covariance function, which then satisfies energy conservation and atomic indistinguishability by construction. The global nature of the interactions is retained by using a global molecular descriptor.

In practice, the sGDML model consisting of the gradient of a kernel ridge regression (KRR) estimator is trained only on forces $\mathbf{F}$, and the use of the Hessian matrix of the kernel κ as a covariance structure naturally gives an analytically integrable force estimator $\hat{\mathbf{f}}_F$, $\hat{f}_E = -\int \hat{\mathbf{f}}_F \cdot d\mathbf{x} + c$.[31,32] Then, the training of the model consists in solving the normal equation of the KRR estimator in the gradient domain for the parameter-vectors $\vec{\alpha}$,[31,32]

$$\left(\mathbf{K}_{\mathrm{Hess}(\kappa)} + \lambda \mathbb{I}\right)\vec{\alpha} = \nabla V_{BO} = -\mathbf{F}, \tag{1}$$

where $\mathbf{K}_{\mathrm{Hess}(\kappa)}$ is the kernel matrix and $\mathbb{I}$ is the identity matrix weighted by the regularization parameter λ. An extensive description of the sGDML framework can be found in Refs. 32, 67, and 64.

The atomic indistinguishability is imposed by adding permutational symmetry (i.e., rigid space group and fluxional symmetries) constraints on $K_{\mathrm{Hess}(\kappa)}$.[32] Extracting those symmetries from the molecular structure requires physical and chemical intuition about the system under study, e.g., knowing if a molecular rotor could jump the energetic rotational barriers at a given temperature, which is impractical in general. This problem is circumvented by implementing a data-driven multi-partite matching procedure that automates the discovery of permutation matrices $\mathbf{P}(\tau)$ corresponding to the index permutation τ embedded in the training data, e.g., molecular dynamics simulations trajectories, followed by a synchronization procedure. The permutation group recovered by the multi-partite matching procedure, also known as the *Longuet–Higgins group*,[68] has the particular advantage that it limits the symmetry recovery to only energetically feasible permutational configurations. This avoids considering unfeasible permutations such as the permutation of two random atoms of the same species, which would not contribute any valuable information to the symmetrized kernel. This translates into a strong reduction of computational efforts in evaluating the model.

The final form of the FF estimator trained on M reference geometries for a molecule with N atoms and a Longuet–Higgins group containing S symmetry transformations takes the form

$$\hat{\mathbf{f}}_F(\vec{x}) = \sum_i^M \sum_q^S \{(\mathbf{P}_q \vec{\alpha}_i) \cdot \nabla\} \nabla \kappa(\vec{x}, \mathbf{P}_q \vec{x}_i). \tag{2}$$

The corresponding energy $\hat{f}_E$ predictor describing the underlying PES is obtained by integrating $\hat{\mathbf{f}}_F$ with respect to the Cartesian coordinates,

$$-\hat{f}_E(\vec{x}) = \int \hat{\mathbf{f}}_F \cdot d\mathbf{x} = \sum_i^M \sum_q^S \{(\mathbf{P}_q \vec{\alpha}_i) \cdot \nabla\} \kappa(\vec{x}, \mathbf{P}_q \vec{x}_i). \tag{3}$$

Figure 1 gives a general perspective of the sGDML model's training process from sampling the molecular dynamics trajectory and extracting the permutational symmetries and a reconstruction process of the underlying PES.

Recently, we have systematically demonstrated that sGDML models trained on only few 100s of reference structures are able to reconstruct molecular PESs with a mean absolute error (MAE) of less that 0.06 kcal mol^{-1} for small molecules with up to 15 atoms and less than 0.16 kcal mol^{-1} for molecules as complex as aspirin, paracetamol, and azobenzene.[9,32,64,69,70] These results will be of particular importance when directly comparing with less flexible MM-FFs in Sec. IV E.

B. Molecular mechanics force fields

Molecular mechanics force fields are designed with the goal to render simulations possible, which are intractable with conventional electronic-structure methods, either due to the system size or simulation time scales. This is achieved by substituting quantum–mechanical interactions with simpler, computationally efficient physical approximations.

In order to construct these approximations, most common force fields explicitly consider the bonds between the atoms in a molecule. Hence, based on the seminal work by Bixon and Lifson,[1] the total energy predicted by a FF can be separated into a sum of bonded and non-bonded interactions,

$$E_{\mathrm{FF}} = E_{\mathrm{bonded}} + E_{\mathrm{non\text{-}bonded}}. \tag{4}$$

Bonded interactions describe the energy contributions due to the interactions of atoms with their closest neighbors and are typically expressed in terms of internal coordinates (bonds, angles, and dihedral angles). Bond energies are modeled after experimentally
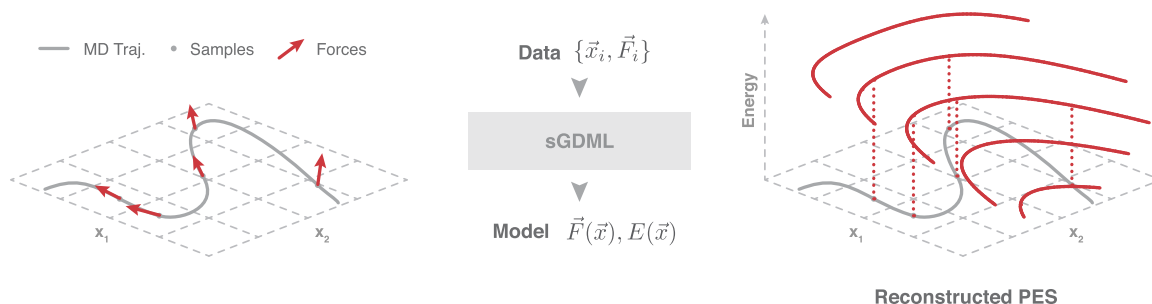
FIG. 1. Construction of the sGDML model. Left panel: from a molecular dynamics trajectory (MD Traj.), a random set of data points, $\{\mathbf{x}_i, \mathbf{F}_i\}_{i=1}^M$, is collected ($\mathbf{x}$: gray dots and $\mathbf{F}$: red arrows). Middle panel: the $\{\mathbf{x}_i, \mathbf{F}_i\}_{i=1}^M$ data are then the input for the sGDML software, which the provides the reconstructed force field and its potential-energy surface (PES; right panel).

observed dissociation curves for diatomic molecules via truncated Taylor series. For the sake of computational efficiency, these series are truncated after the quadratic term, leading to the harmonic expression,

$$E_{\text{bond}}(r_{ij}) = \frac{1}{2} k_{ij} \left(r_{ij} - r_{ij}^{(\text{eq})} \right)^2, \tag{5}$$

where r_{ij} is the distance between atoms i and j, $r_{ij}^{(\text{eq})}$ is the equilibrium bond length, and k_{ij} is the force constant modulating the strength of the bond. k_{ij} and $r_{ij}^{(\text{eq})}$ differ depending on the chemical elements involved in the bonds, as well as their chemical environments (e.g., hybridization state), and are determined based on experimental observations or electronic-structure calculations.

Angles are modeled in a similar manner,

$$E_{\text{angle}}(\alpha_{ijk}) = \frac{1}{2} k_{ijk} \left(\alpha_{ijk} - \alpha_{ijk}^{(\text{eq})} \right)^2. \tag{6}$$

Here, α_{ijk} is the angle between the bonded atoms i, j, and k, while $\alpha_{ijk}^{(\text{eq})}$ and k_{ijk} are the angle at equilibrium and the associated restitution force constant, respectively. Interactions arising from dihedral angles are modeled via Fourier series in order to account for the correct periodicity of the associated PES,

$$E_{\text{dihedral}}(\theta_{ijkl}) = \frac{1}{2} \sum_{\gamma} V_{ijkl}^{(\gamma)} \tag{7}$$

$$\times \left[1 + (-1)^{\gamma+1} \cos(\gamma \theta_{ijkl} + \phi_{ijkl}^{(\gamma)}) \right], \tag{8}$$

where θ_{ijkl} is the dihedral angle between atoms i, j, k, and l. $\phi_{ijkl}^{(\gamma)}$ is a phase shift and $V_{ijkl}^{(\gamma)}$ is an amplitude controlling the interaction strength. The total bonded energy is then obtained as a sum over all these terms arising from to the predetermined connectivity of the investigated system,

$$E_{\text{bonded}} = \sum_{\text{bonds}} E_{\text{bond}}(r_{ij}) \tag{9}$$

$$+ \sum_{\text{angles}} E_{\text{angle}}(\alpha_{ijk}) \tag{10}$$

$$+ \sum_{\text{dihedral}} E_{\text{dihedral}}(\theta_{ijkl}). \tag{11}$$

Non-bonded interactions—defined as all interactions of atoms separated by more than two neighbors—are most commonly treated in the form of pairwise potentials in conventional FFs. Electrostatic interactions are modeled as a Coulomb potential,

$$E_{\text{Coulomb}}(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \tag{12}$$

where ϵ_0 is a dielectric constant and q_i and q_j are the partial charges associated with the interacting atoms. These properties are once again free parameters of the FF model and need to be determined based on the identity of the interaction partners. Finally, long range van der Waals type interactions are typically accounted for via a

Lennard-Jones potential[71] due to its correct decay behavior and computationally efficient evaluation,

$$E_{\text{vdW}}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right]. \tag{13}$$

Here, ϵ_{ij} controls the depth of the potential well and the parameters A_{ij} and B_{ij} modulate its overall shape. As was the case with bonded interactions, all non-bonded terms are combined to yield

$$E_{\text{non-bonded}} = \sum_{i,j>i} E_{\text{Coulomb}}(r_{ij}) \tag{14}$$

$$+ \sum_{i,j>i} E_{\text{vdW}}(r_{ij}). \tag{15}$$

Although the number of bonded energy contributions grows linearly with the number of atoms present in the system, non-bonded interactions grow quadratically due to their pairwise nature. In order to overcome this $\mathcal{O}(N^2)$ scaling behavior, a range of highly efficient algorithms for treating long range effects has been developed, such as methods based on Ewald summation[72,73] or fast multipole methods[74] (see Ref. 75 for an overview). Interestingly, there is a growing number of ML-FF approaches, which can make direct use of these algorithms to improve performance for extended systems in a manner analogous to MM-FFs (see e.g., Refs. 24, 42, 76, and 77).

At the core of every force field lies a set of parameters tabulated for the above interactions, which are determined by fitting to experimental and theoretical data. There is a wide range of conventional force fields, differing in the class of target systems and the types of atom environments considered (e.g., AMBER,[2] CHARMM,[3] and GROMOS[5]). All of these excel at high computational speeds, making possible to simulate systems containing hundreds of thousands of atoms, e.g., biologically relevant systems such as proteins, DNA strands, and cell membranes.[78] However, the price MM-FFs pay for this efficiency is a loss of flexibility and accuracy. For example, harmonic expressions for bond energies [Eq. (5)] are only valid close to the equilibrium and fail to account for anharmonicities in bonds, as well as situations involving formation and breaking of bonds.

III. COMPUTATIONAL METHODS

All sGDML models in this study were trained on subsets of 1000 points sampled from the respective bulk datasets of reference calculations according to the Boltzmann distribution. Smaller independent validation subsets were used to determine the best hyperparameter-configuration for each model. The remainders of each dataset served as test splits to estimate the generalization performance of each model. This process is fully automated by the sGDML software package (available at www.sgdml.org).[64] The molecular datasets used in this study were generated using DFT at the generalized gradient approximation (GGA) level of theory with the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional,[79] and the van der Waals interactions were incorporated using the Tkatchenko–Scheffler (TS) method[80] as implemented in the FHI-aims package.[81] For the keto-MDA,

enol-MDA, and ethanol molecules, we also computed datasets using all-electron CCSD(T) and all-electron CCSD for aspirin using the Psi4 software.[82–84]

PES scans for the different MM-FFs were performed with the Atomic Simulation Environment (ASE)[85] using the built-in calculator for generalized AMBER force field (GAFF) and custom calculators for UFF,[86] MMFF94,[4] and Ghchemical[87] based on their implementation in Open Babel.[88] For GAFF computations, the AMBER 18 force field package[89] was used and the charges of all molecules were derived from RESP charges[90] at the HF/6-31G* level of theory[91,92] computed with Gaussian.[93] Molecular dynamics simulations were carried out using ASE. A time step of 0.5 fs was used to propagate the system, and temperatures were kept constant using a Langevin thermostat.[94] sGDML simulations were performed under similar conditions using the sGDML-ASE calculator implemented in the sGDML software package.[64]

IV. RESULTS AND DISCUSSION

A. Comparison of MM-FFs and ML-FFs

ML-FFs are able to construct efficient and accurate surrogate models of high-level electronic-structure methods. As a consequence, they can be used to directly compare CC-level PES with the predictions of conventional MM-FFs, offering insights into the quality and potential shortcomings of these methods at a fundamental level. This in turn can pave the way toward improved MM-FF approaches.

Here, we demonstrate the utility of such a comparison based on ethanol and keto-malondialdehyde as example systems. These molecules are very useful to exemplify two ubiquitous phenomena in chemistry and biology, the effects of electron lone-pairs and purely quantum–mechanical effects (e.g., $n \rightarrow \sigma^*$ transitions).[9,95,96] As a consequence, the PES of both molecules exhibit highly complex features, which even DFT approximations struggle to capture.[9,32]

Figure 4 compares two-dimensional cuts through the PESs of both molecules for a sGDML model trained on CCSD(T) data with the predictions of different MM-FFs. The Generalized Amber Force Field (GAFF)[8] is a representative of the Bixon–Lifson type FFs introduced in Sec. II B. The other MM-FFs, the Merck Molecular Force Field (MMFF94),[4] Ghemical,[87] and the Universal Force Field (UFF),[86] are extensions to this framework and introduce more complex functional forms in order to improve generalization and predictive accuracy. All MM-FFs studied here have been selected due to their performance for small to medium sized organic molecules.

In the case of the ethanol molecule, a qualitative comparison reveals that GAFF and MMFF94 yield good approximations to the sGDML@CCSD(T) reference PES. However, they miss the coupling between the hydroxyl and methyl functional groups caused by the interactions of the oxygen lone pair electrons with the methyl hydrogens. This coupling induces an asymmetric behavior in the PES of both angle coordinates, which can be observed for the sGDML model. While the absence of this effect is to be expected in GAFF, since there is no explicit description of lone pairs, MMFF94 also fails to account for the interaction despite its more elaborate functional form. Ghemical and UFF, on the other hand, both manage to reproduce the rotor coupling qualitatively but completely misrepresent the remainder of the PES.

The keto-MDA molecule is a particularly interesting system with two symmetrical main degrees of freedom in the form of its aldehyde rotors. The PES associated with these rotors displays a wide variety of features due to complex quantum–mechanical interactions.[9] Hence, accounting for these effects is nontrivial for conventional FF approaches as well as ML techniques. In Fig. 2, we observe that the default version of UFF is insufficiently parametrized to describe the interaction between the two aldehyde groups in general. GAFF reproduces the global minima qualitatively well, mainly in the form of electrostatic interactions between both rotors. Yet, as would be expected from a Brixon–Lifson FF, it misses the
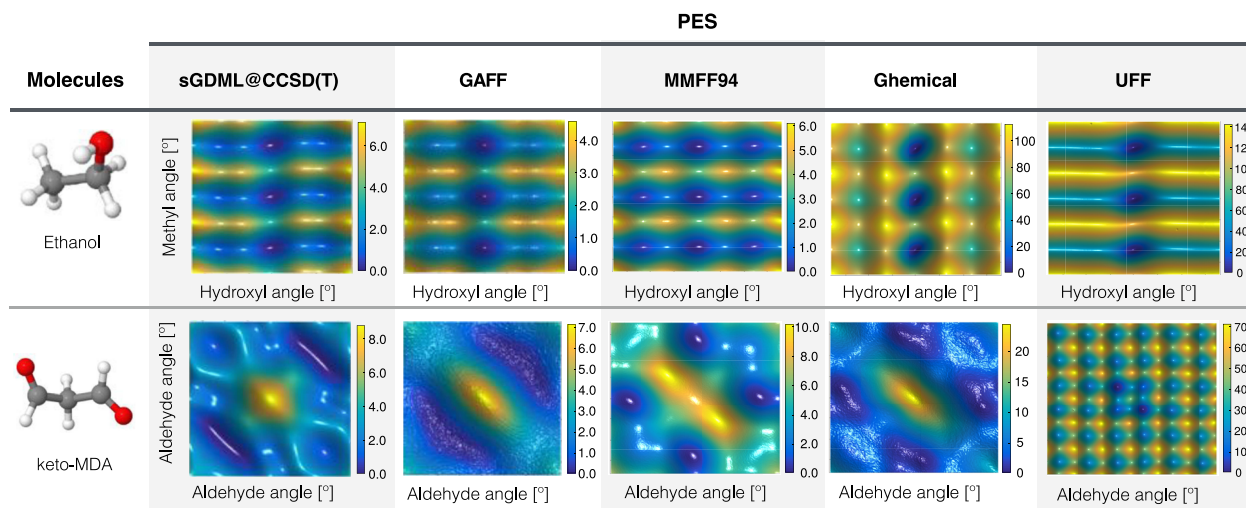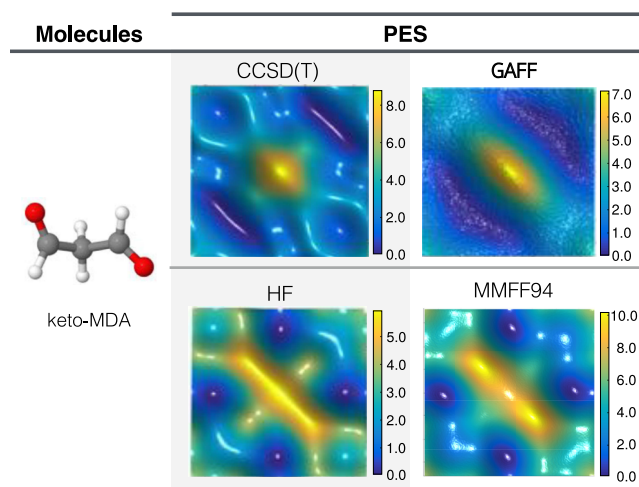

FIG. 2. Comparison of PES generated using several MM-FFs for ethanol and keto-MDA compared to the sGDML@CCSD(T) ML-FF. All energies are given in kcal mol^{-1}.

FIG. 3. Comparison of keto-MDA's PES generated from sGDML@CCSD(T), GAFF, sGDML@HF, and MMFF94. All energies are given in kcal mol^{-1}.

two local minima generated by non-classical electronic interactions. Interestingly, both MMFF94 and Ghemical FFs generate a PES that bears only remote resemblance to the PES predicted by sGDML@CCSD(T) or GAFF. The reason for this behavior is that both FFs were primarily parametrized using data at the Hartree–Fock (HF) level of theory.[4] For MMFF94 in particular, the agreement with the reference HF PES as depicted in Fig. 3 is remarkable. This is an important result, since it demonstrates that the more complex functional form of the MMFF94 FF could, in principle, be reparametrized using a higher level of theory data [generated with the sGDML@CCSD(T) model] to yield high quality PES.

In the following, we perform a more detailed analysis of the sGDML@CCSD(T) model and GAFF as a representative for MM-FFs in order to gain a better understanding of the strengths and shortcomings of both approaches. Such an in-depth understanding is tantamount for devising potential improvements to both kinds of FFs. GAFF is chosen as the primary MM-FF due to its general reliability (see above), as well as its simple analytical form, which facilitates the interpretation of results. Figure 4 shows detailed PESs of ethanol and keto-MDA for sGDML@CCSD(T) and GAFF, together with the differences between both methods.
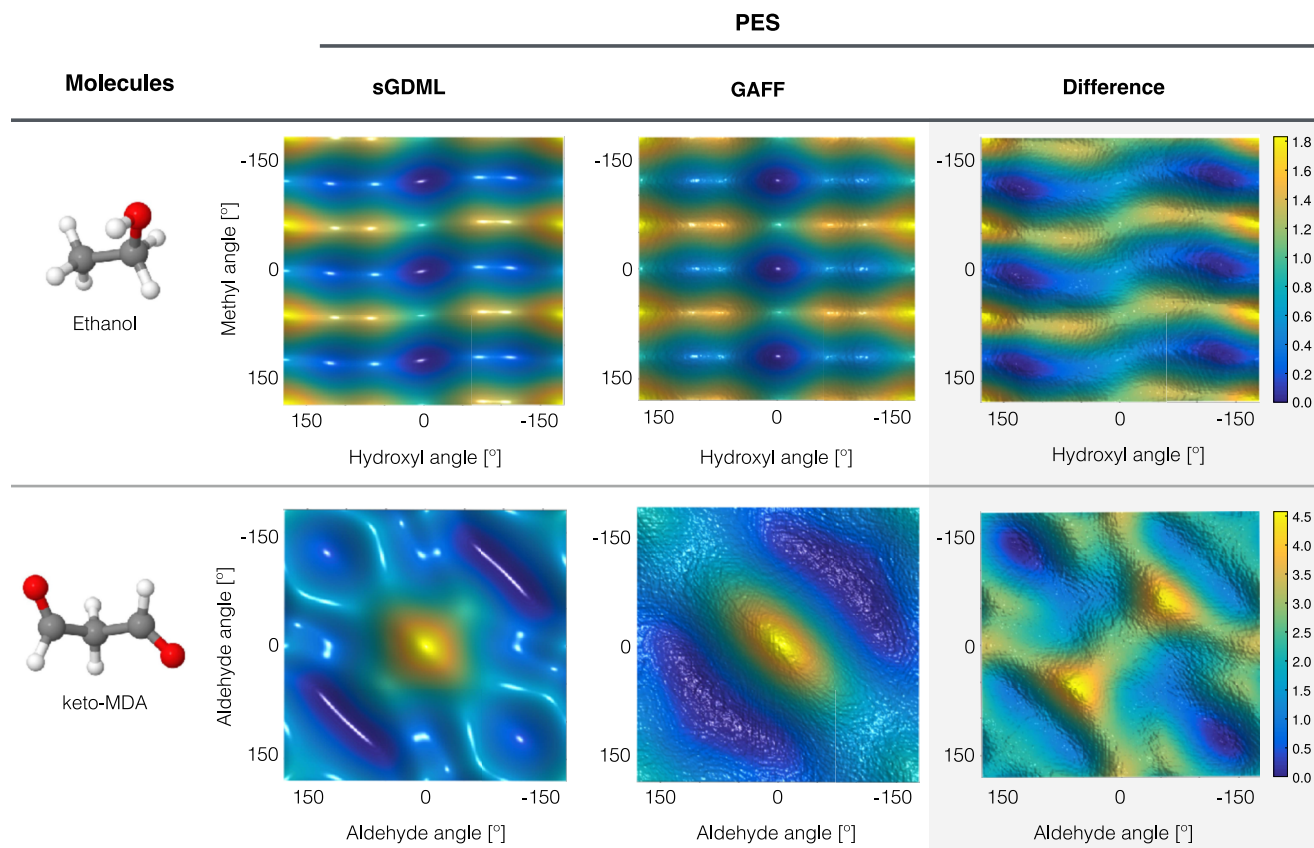
FIG. 4. PESs generated from sGDML@CCSD(T)/cc-pVTZ, GAFF for ethanol and keto-MDA, and their respective difference PES$_{sGDML}$–PES$_{GAFF}$. Energy differences are given in kcal mol^{-1}.
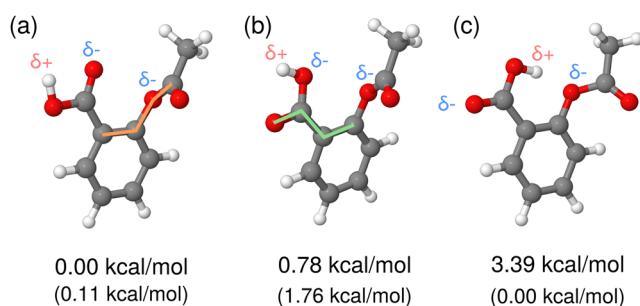
FIG. 5. [(a)–(c)] Three lowest minima of aspirin computed with CCSD/cc-pVDZ. Energy differences relative to the minimum in kcal mol^{-1} are shown for the sGDML@CCSD model and GAFF (in parentheses).

In the case of the ethanol molecule (Fig. 4, first row), the deviations between both approaches appear to be only minor when comparing the PES directly. However, the difference plot reveals the details missing in the GAFF description in the form of the coupling between the two rotors in ethanol. The overall differences are more obvious for keto-MDA, where only the global minima and the central peak in the PES are reproduced satisfyingly by GAFF (see Fig. 4, second row). The fact that, in general, GAFF describes the reference PESs qualitatively well for these two molecules (particularly their global minimum) opens the opportunity to use it for creating a reliable sampling of their PES via molecular dynamics. This would be an alternative to its more computationally expensive counterpart, DFT molecular dynamics, as we will explore in Sec. IV C.

In addition to ethanol and keto-MDA, we also study the performance of MM-FFs for the aspirin molecule. Here, we use GAFF to predict the ordering of the three lowest lying equilibrium configurations of aspirin identified with a sGDML@CCSD model in a previous study.[32] The associated structures, along with the relative sGDML@CCSD energy differences, are depicted in Fig. 5. Contrary to the cases of ethanol and keto-MDA, GAFF strongly misrepresents the ordering of the minima on the aspirin PES (Fig. 5, parentheses). In aspirin, the global minimum shown in Fig. 5(a) is stabilized by the $n \to \pi^*$ attractive interaction between the electron lone pair n of the oxygen atom in the carboxylic acid group and the anti-bonding π^* orbital located in the carbon atom of the ester group.[95] If we remove this interaction from the equation and just consider the electrostatic contributions such in the GAFF case, the geometric configuration Fig. 5(a) is not longer stable due to the very close proximity of the three negatively charged oxygen atoms. In this case, Fig. 5(c) is the most stable configuration, where two the oxygen atoms are sharing a positive hydrogen atom. In the CCSD level of theory, this structure is also a local minimum, but it is 3.39 kcal mol^{-1} higher than the global minimum.

B. Anharmonicities and vibrational spectra

To supplement the previous analysis based on a static picture, we want to briefly discuss some of the dynamical implications of using conventional FFs in general. We performed a series of long MD trajectories of the ethanol molecule for different temperatures

(100 K–400 K) to compute the vibrational density of states (VDOS). In Fig. 6, we show the results for GAFF and for sGDML@CCSD(T). First, we can see that GAFF gives a good prediction of the VDOS in the fingerprint region (500 cm^{-1}–1500 cm^{-1}), since, in general, it is parametrized to perform well in this frequency range. On the other hand, it poorly describes the bond stretching vibrations involving hydrogen atoms (~3000 cm^{-1}). Moreover, if we analyze the evolution of the VDOS with temperature, we can see that there is no major change or shift in the GAFF frequencies with increasing temperature. This a direct consequence of the MM-FFs functional form, where bond and angle interactions [Eqs. (5) and (6)] are approximated as harmonic terms, neglecting higher order contributions. However, in reality, both interactions cannot be expressed purely in terms of harmonic oscillators, with bond dissociation being one of the best counterexamples. These anharmonic contributions give rise to the temperature dependent shifts absent in the middle to high frequency regimes of the MM-FF VDOS (>500 cm^{-1}), but modeled correctly by the ML-FF. For low frequency vibrations, anharmonicities can still be captured by a MM-FF through the interplay of dihedral terms [Eq. (8)] and long range interactions [Eqs. (12) and (13)], which are only computed between atoms at least three bonds removed in classical Bixon–Lifson FFs. This behavior can be observed in the VDOS regions of the hydroxyl and methyl rotors (<500 cm^{-1}), which both show a temperature induced redshift. The dynamics obtained using the sGDML@CCSD(T), on the other hand, exhibit non-trivial shifts in the frequencies and VDOS populations across the whole spectrum, demonstrating the advantage of the less restricted functional forms used in ML-FFs.

It is worth noting that by analyzing the vibrational normal modes, we can see that the two rotors are completely decoupled in the case of GAFF, while in the sGDML case, the well-known coupling is present. Based on this, we conclude that the lack of anharmonicities in GAFF will certainly affect the sampling of the PES, as will be explored in the following.

C. Generating reference data with MM-FFs

For machine learning methods, the construction of suitable reference databases is a crucial step, given that any model trained or fitted to a given database can only generate predictions as accurate as the data itself. For example, a model trained on DFT data can at most give results as accurate as DFT. Therefore, the level of theory and methodology used to generate the data has to be chosen carefully, trying to maximize the amount of data accessible at a theory level that accurately encodes the quantum–mechanical effects of interest utilizing the computational resources at hand. In this regard, DFT reference data generated by running molecular dynamics simulations at a high temperature (500 K) have been used to generate large and reliable molecular datasets, known as MD17.[9,31,32] The software used to generate the MD17 dataset was the FHI-aims package[81] using the generalized gradient approximation (GGA) with the Perdew–Burke–Ernzerhof (PBE)[79] exchange–correlation functional combined with the Tkatchenko–Scheffler (TS) method[80] to incorporate van der Waals interactions.

While a direct sampling approach is feasible at the DFT level, it quickly become intractable at higher levels of theory. To circumvent
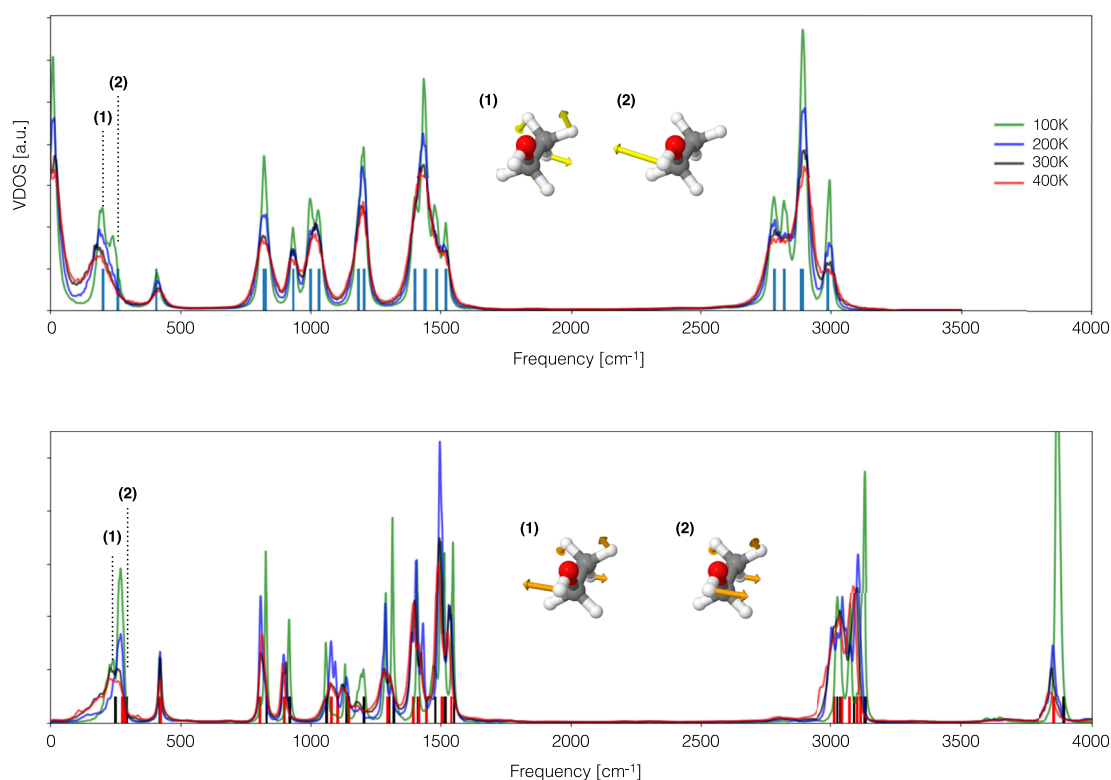
FIG. 6. Comparison of the VDOSs generated with sGDML@CCSD(T) and GAFF force fields for ethanol. (a) GAFF. (b) sGDML.

this limitation, a two-step data generation approach can be used to create more accurate datasets (see Ref. 97). In a recent example, such a procedure was used to generate a CCSD(T) level MD17 dataset by randomly subsampling some of the molecular dynamics trajectories in the MD17 dataset and recomputing energies and forces using the CCSD(T) level of theory. This methodology successfully generated highly accurate sGDML@CCSD(T) FFs able to replicate experimental results giving further insights about the importance of electron correlation and nuclear quantum effects.[9,32] The origin of the success of the two-step process is that the PES_{DFT} sampled in the MD17 trajectories contains already most of the general features in the $PES_{CCSD(T)}$, and therefore, the sGDML model can use the same molecular configurations as the basis to reconstruct both PESs.

This naturally raises the question on the validity of using datasets generated by MD simulations with conventional FFs as the basis for a similar two-step reference database construction. Such a procedure has long been suggested, since it would offer the opportunity to create vast amounts of molecular configurations to be used as inputs in machine learning methods. Nevertheless, to our knowledge, no study has ever analyzed the reliability of the set of configurations generated in such a manner. As stated above, the success of a two step procedure would depend on how similar the MM-FF free energy is to the ones generated by the reference electronic-structure method, in our case GAFF and DFT or CCSD(T), respectively. In Secs. IV A and IV B, we

already analyzed some of the differences, advantages, disadvantages, regimes of applicability, and possible problems for certain types of molecules.

With this information at hand, we now explore how these effects influence the sampling behavior of GAFF and sGDML, with a particular focus on the implications this has for using GAFF to construct a MM-FF based version of the MD17 dataset. In Fig. 7, we present a comparison between the sampling generated by DFT and GAFF molecular dynamics for ethanol, keto-MDA, and aspirin at 500 K. From Fig. 7(a), it is straightforward to see that even in this 2D representation of the data, these two methodologies generate different samplings of the configurational space. As mentioned before, in the case of ethanol, one of the key ingredients missing in GAFF is the coupling between methyl and hydroxyl rotors and the fact that the methyl rotor is not symmetric due to its parametrization. Of course, this is a projection of the data in their two main degrees of freedom, and therefore, in Fig. 7(b), we show the radial distribution function comparing the interatomic distances obtained in each case. Clearly, it was not expected for GAFF to generate the same results as DFT, but the sole purpose of this section is to highlight the different dynamics generated by each approximation. From Fig. 7(b), we can see that in general, GAFF performs with reliable accuracy in predicting the interatomic distances. The second row of Fig. 7(a) is the demonstration to what was discussed earlier in Sec. IV A in terms of the PES of keto-MDA, where the missing local minima will influence the sampling process. The most
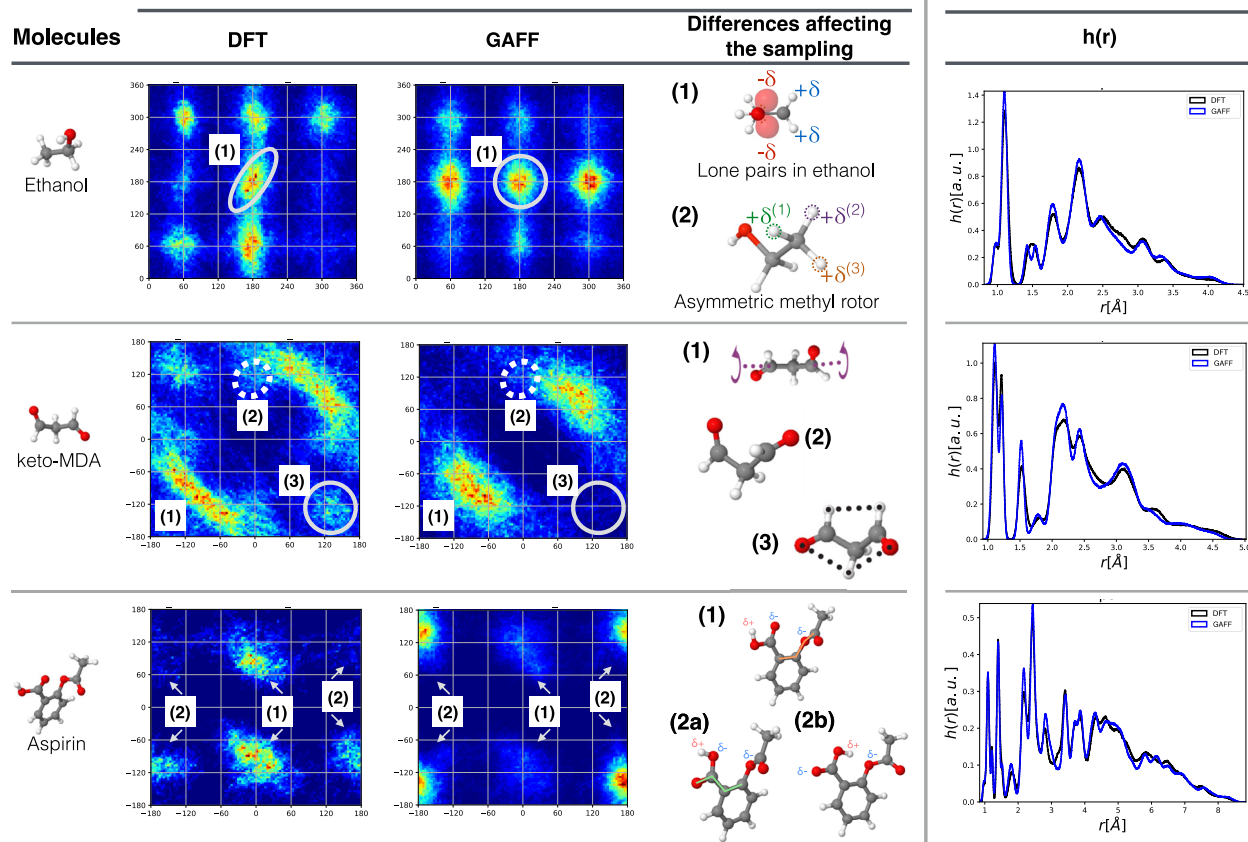
FIG. 7. Sampling generated by comparison of sampling generated from DFT and GAFF-force field for ethanol, keto-MDA, and aspirin. The overlap between the sampling is shown in the rightmost column as a pair distribution function of interatomic distances. (a) PES sampling at 500 K. (b) Pair distribution function.

radical case is aspirin, where GAFF preferably samples two of its local minima instead of the global one, as shown in the bottom of Fig. 7(a).

Now, from Fig. 7, we have evidence that DFT and GAFF, in general, sample different parts of the configuration space. Therefore, with this information at hand, the next question to answer is whether the GAFF sampled configurations can still be used to reconstruct the DFT PES. It should also be noted at this point that the goal of the learning process is to generate a function able to generalize, i.e., to reconstruct, the underlying level of theory used to generate the database and not to learn or memorize the database itself.

D. Cross comparison between MM-FF and DFT sampling

Creating highly accurate *ab initio* [e.g., CCSD(T)] data directly from MD simulations is infeasible for all but the smallest molecules. Instead, it is customary to run the simulations at a lower and hence more affordable level of theory, such as DFT (e.g., PBE) combined

with a small basis set. A fundamental hypothesis is that the PES generated by DFT methods is close enough to a higher level of theory PES, such as CCSD(T). Therefore, by sampling the PES using DFT based molecular dynamics simulations, one would expect the generated molecular configurations to be a good approximation to sampling the corresponding CCSD(T) PES directly. This in turn makes it possible to accurately reconstruct the CCSD(T) PES from MD@DFT trajectories. This has been empirically shown to be a valid assumption,[9,32,64,70] since DFT functionals incorporate a considerable amount of exchange and correlation to describe the quantum system.

It tempts us to use the same approach to generate data from MD@ML-FFs to reconstruct DFT PESs, but from Figs. 2 and 7, we see that the GAFF and PBE + TS PESs exhibit fundamental differences. Hence, running MD simulations on each surface will generate distinct populations of the configuration space.[32] Consequently, there is a high chance that the PES reconstructed by the sGDML model based on MD@FFs data will generate good training errors, but will not be able to generalize properly. In order to verify this assertion, we randomly sampled 2000 geometries (1000
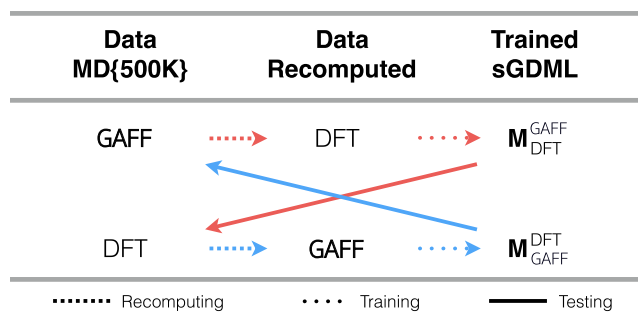
FIG. 8. Scheme describing the generation of datasets and models used for the cross comparison experiments.

for training and 1000 for testing) from the GAFF simulations used in Fig. 7, which were recomputed using DFT (PBE + TS; see the supplementary material for more details). These data, which we refer to as the MD17^{GAFF} dataset, were then used to train a set of sGDML models (see Fig. 8). The results of this experiment are reported in Table I.

Based on these results, we find that for the considered molecules, the sGDML models trained and validated on GAFF data (sGDML@MD17^{GAFF}) perform slightly better than the sGDML models trained and validated on DFT data (sGDML@MD17^{DFT}). Nevertheless, when validating the sGDML@MD17^{GAFF} model on MD@DFT sampled data, the errors increase almost by a factor of two. We would like to highlight that even when validating the sGDML@MD17^{GAFF} model on unseen data from another distribution, it manages to generate good force MAEs below 0.9 kcal mol^{-1} Å^{-1}, which is a remarkable achievement. This behavior hints at a potential use in conjunction with adaptive sampling approaches,[20] where an initial model trained on MM-FF sampled data could serve as a basis for further iterative refinement.

The analysis presented here gives a better perspective on the use of conventional MM-FFs to generate representative databases to higher levels of theory for ML reconstruction of the molecular PES.

TABLE I. Prediction accuracy for interatomic forces of the GAFF sampled datasets at 500 K and recomputed using PBE + TS. Force errors are reported in kcal mol^{-1} Å^{-1}.

Dataset	Sampling		Force prediction	
	Training	Validation	MAE	RMSE
Keto-MDA	GAFF	GAFF	0.29	0.44
	DFT	DFT	0.41	0.62
	GAFF	DFT	0.54	0.82
Ethanol	GAFF	GAFF	0.27	0.42
	DFT	DFT	0.33	0.49
	GAFF	DFT	0.45	0.71

E. Augmenting MM-FFs with ML

Up to this point, we have focused on GAFF, a Bixon, and Lifson MM-FF with the predetermined functional form and parametrization scheme. However, as can be seen in Fig. 4, extending the functional form or changing the set of parameters can lead to significant differences in the behavior of force fields. This hints at potential applications where the synergy between MM-FFs and ML approaches can be exploited in order to improve upon conventional FFs. Here, we study aspects of two possible MM/ML-FF setups, where (a) ML methods might be used to generate additional data for parametrizing a MM-FF at a higher level of theory or (b) augment the functional form of a force field, incorporating interactions due to quantum effects.

In a first series of experiments, we explore to what extent MM-FFs can profit from additional data as well as changes to their functional makeup. To this end, we construct GAFF type force fields for ethanol and aspirin based on the electronic-structure energies and forces of 110, 1100, and 11 000 reference structures taken from the MD17 database. As a baseline MM-FF, we introduce a PyTorch[98] implementation of the GAFF. This allows us to reoptimize all GAFF parameters using molecular energies and forces, as would be done in a typical ML-FF scenario. Moreover, in order to study the effect of changing the functional form, we introduce two extensions to this basic GAFF. In the first variant, all harmonic bond potentials are replaced by Morse potentials and the number of terms in the dihedral Fourier series is doubled (see Sec. II B). The second variant introduces additional flexibility by modeling all bonded interactions via feed-forward neural networks (NNs) instead of the conventional physical approximations (see the supplementary material for more details). These models will be refered to as GAFF-Morse and GAFF-NN, respectively. The test set errors achieved for the differently sized training sets, molecules, and GAFF variants are reported in Table II. Shown are the energy and force RMSEs obtained as the average over three independent runs for each model.

In general, significantly higher errors are obtained for the GAFF variants compared to pure ML-FFs, such as sGDML. This is caused by the less flexible, physically motivated form of MM-FFs compared to their ML counterpart. However, the potential advantage of constructing force fields in such a manner becomes apparent when looking at the saturation of the error with respect to the amount of reference data used during training. Even with only 110 data points, MM-FFs exhibit good generalization performance, which changes little upon the inclusion of more data. This is especially remarkable in the case of flexible molecules such as aspirin. These observations indicate that the refitted GAFF models for ethanol and aspirin studied here are already saturated with respect to the training data, and little performance can be gained upon providing additional data without fundamentally changing the functional form of classical FFs. Moving on to GAFF-Morse, we find the above conclusions confirmed. The increase in functional flexibility allows the model to converge to slightly better RMSEs compared to the unmodified versions and utilize additional data in a more effective manner. It is also interesting to note that forces appear to profit from the extended functional form to a larger extent than the energies. The reason for this effect is the presence of anharmonic contributions to the bond energies, also observed in the VDOSs studied in Sec. IV B. As would be expected from its more flexible

TABLE II. Classical and modified GAFFs trained on MD17 reference data for ethanol and aspirin. GAFF denotes the unmodified GAFF force field, while GAFF-Morse and GAFF-NN corresponds to the variants using Morse potentials and NNs. Energy and force RMSEs on the test set are given in kcal mol^{-1} and kcal mol^{-1} Å^{-1}, respectively. All errors are obtained as the average over three independent training runs. Where available, we have added the sGDML results.

Molecule	Datapoints	Energy				Forces			
		sGDML	GAFF	GAFF-Morse	GAFF-NN	sGDML	GAFF	GAFF-Morse	GAFF-NN
Ethanol	110	. . .	1.27	2.40	1.37	. . .	6.53	15.38	5.00
	1100	0.07	1.21	2.51	1.16	0.34	6.37	14.59	4.78
	11 000	. . .	1.20	1.15	1.11	. . .	6.38	4.80	4.66
Aspirin	110	. . .	2.18	2.94	2.28	. . .	7.91	10.29	6.72
	1100	0.19	2.20	2.37	2.19	0.68	7.46	7.77	6.15
	11 000	. . .	2.05	2.10	1.98	. . .	7.35	6.28	5.95

NN based structure, GAFF-NN continues the trends observed for data saturation and predictive accuracy. Energies and force predictions, in particular, appear to benefit greatly from the more flexible functional form provided by the NNs. Interestingly, even in the case of using only 110 data points, GAFF-NN outperforms the other variants with respect to force predictions. This suggests, that MM methods can profit from ML based terms even in low data regimes.

In a final experiment, we perform a proof of principle study on how ML approaches can be used to formulate extensions to MM-FFs, compensating for missing interaction terms. To this end, we target the $n \rightarrow \pi^*$ interaction in the aspirin molecule as a test application. The inability to account for this quantum–mechanical effect is the reason for GAFF yielding the wrong ordering of minima (Fig. 5) and exhibiting qualitatively different sampling (see Fig. 7). We introduce a sGDML based extension parametrized to correctly capture the orbital–orbital interaction (see the supplementary material for details). A GAFF model augmented with this ML term, $\hat{\mathbf{f}}_{\mathbf{F}}^{(n \rightarrow \pi^*)}$, is then used to repeat the sampling experiments performed in Sec. IV C. Figure 9 shows the different sampling behavior of sGDML@CCSD, pure GAFF, and the extended GAFF model (GAFF-ML).

As can be seen, even a relative simple augmentation is enough to recover the sampling behavior of the reference method to a large extent. The global minimum dominated by the $n \rightarrow \pi^*$ interaction (1) is now visited preferably by the GAFF + ML model, which agrees with results from sGDML@CCSD.[70] The relative sampling frequency of visiting the different minima is reproduced as well. This result is particularly impressive, considering the restrictive functional form of the basic MM-FF. Even more promising is the procedure by which this term was introduced, exemplifying a potential synergistic approach to developing force fields, where missing terms can be identified via ML-FF comparisons, condensed into a ML model and then deployed with MM-FFs in a highly systematic manner.

V. CONCLUSIONS AND OUTLOOK

Machine learning based force fields (ML-FFs) show excellent accuracy in modeling a wide range of chemical phenomena. Using
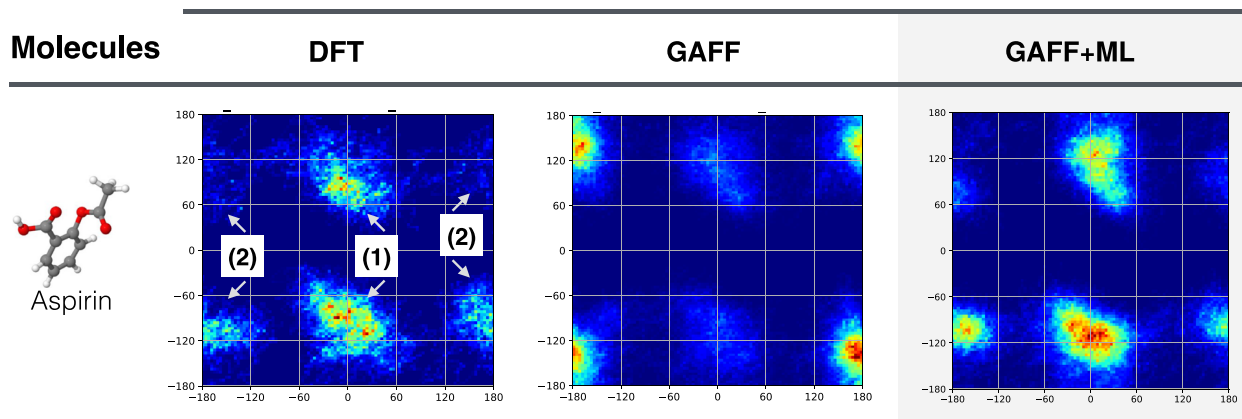
FIG. 9. Comparison of sampling of the three aspirin minima by a sGDML@DFT model, a standard GAFF, and a GAFF model corrected for the $n \rightarrow \pi^*$ interaction term.

specialized architectures, such as sGDML studied here, extremely reliable surrogate models of high-level electronic-structure methods can be constructed. These models are able to capture a wide range of the underlying quantum many-body effects while, at the same time, only incurring a fraction of the computational cost of the original method. However, due the involved functional structure used, the computational speed of ML-FFs still pales in comparison to classical molecular mechanics force fields (MM-FFs). MM-FFs employ a set of highly efficient approximations, making them capable of simulating systems containing many thousands of atoms. However, this efficiency comes at the cost of general accuracy, rendering MM-FFs less reliable or even incapable of modeling complex chemical phenomena. The focus of the present study is to identify differences and synergies between MM-FFs and ML-FFs and how the resulting knowledge can be used to combine both approaches.

An important aspect of such a synergistic approach is the analysis of models and identifications of their shortcomings. Here, we could show that accurate ML-FFs offer access to high-quality PES at the CCSD(T) level of theory, which can be used to probe the performance of MM-FFs in a very efficient manner. In this way, insight can be gained into situations where models perform reliably and cases where they fail completely. Moreover, ML-FFs can not only be used for a general analysis but also even allow pinpointing physical phenomena systematically missing from the description of MM-FFs, thus guiding the development of new, improved approaches. Another aspect explored in this study is the use of MM-FFs for generating reference datasets to be used in constructing ML-FFs. Due to their high computational efficiency, MM-FFs constitute a powerful tool for sampling complex potential-energy surfaces (PESs). It was found that MM-FFs can be used to directly sample reference configurations for ML models parametrized at a high electronic-structure level, provided that a sufficiently high overlap between the PES of both methods is present. However, even in situations where this assumption does not hold, MM-FFs show acceptable performance, which makes them highly promising in the early stages of constructing ML-FFs as they can provide an initial set of reference configurations, which can then be further refined with adaptive sampling.[20] ML approaches can also be used to directly augment MM-FFs. Here, we have investigated how changes to the functional form influence their capacity to leverage data in varying amounts of reference data. Increasing the flexibility of MM-FFs by modeling all bonded interactions via neural networks yielded a model, which exhibits improved accuracy over conventional schemes already in small data regimes and consistently improves upon addition of more data. This suggests that MM-FFs can profit greatly from functional features of modern ML algorithms. These can either be used in a targeted fashion as corrections to include additional interactions not captured by the basic force field or tightly integrated into a MM-FFs structure. The more flexible functional forms allow MM-FF models to capture a wider range of phenomena and leverage additional information without compromising the basic design paradigm of MM-FFs. In addition, the experiments in this study also demonstrated how ML components can in turn benefit from being embedded in a MM-FF framework. The principled physical structure and decomposition of interactions has the potential to further improve data efficiency and generalization behavior of ML-FFs.

All these findings point toward the general potential of integrating force fields and ML approaches. Efficient hybrid approaches offering high performance could be constructed, where ML-FF terms restricted to local regions of a system could be introduced to MM-FFs, accounting for physical phenomena not fully described by the purely classical interaction, e.g., orbital based effects. This could be achieved by restricting the molecular descriptor in sGDML to a set of atoms, similar to how bonded interactions are treated in Bixon–Lifson force fields. Another important phenomenon where terms derived from ML-FFs could prove highly advantageous is the treatment of long range van der Waals interactions. These are typically accounted for in an additive pairwise manner. However, there is mounting evidence that many-body effects play an important role in these interactions, which might even be crucial in describing the macroscopic structure of soft matter.[99] Other potential applications where both—MM-FF and ML-FF—could be combined to great effect is the construction of ML-FFs for extended systems, such as proteins or solutions. There, generating high-level reference data via sampling driven purely by electronic-structure theory is infeasible. MM-FFs, on the other hand, could provide efficient access to large parts of the configuration space. Missing regions could then be sampled iteratively.

In summary, classical MM-FFs and modern ML-FFs both exhibit particular strengths that complement each other. Here, we have investigated various facets of how both approaches can be integrated with each other, taking a first step toward systematically combining their strengths.

SUPPLEMENTARY MATERIAL

See the supplementary material for additional information on the reference data generation used for training the ML-FFs, information about the neural network variant of GAFF, its training details, and the tables with prediction accuracies, some supplementary notes regarding the $n \rightarrow \pi^*$ interaction in aspirin and the generated data for training, and the sGDML code and documentation that is available at http://quantummachine.org/gdml/.

ACKNOWLEDGMENTS

DATA AVAILABILITY

The training data and code base used in this study are available from the authors upon reasonable request.

REFERENCES

[1] M. Bixon and S. Lifson, Tetrahedron **23**, 769 (1967).

[2] P. K. Weiner and P. A. Kollman, J. Comput. Chem. **2**, 287 (1981).

[3] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

[4] T. A. Halgren, J. Comput. Chem. **17**, 490 (1996).

[5] T. A. Soares, P. H. Hünenberger, M. A. Kastenholz, V. Kräutler, T. Lenz, R. D. Lins, C. Oostenbrink, and W. F. van Gunsteren, J. Comput. Chem. **26**, 725 (2005).

[6] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[7] M. W. Mahoney and W. L. Jorgensen, J. Chem. Phys. **112**, 8910 (2000).

[8] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).

[9] H. E. Sauceda, S. Chmiela, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, J. Chem. Phys. **150**, 114102 (2019).

[10] M. Stöhr and A. Tkatchenko, Sci. Adv. **5**, eaax0024 (2019).

[11] K. T. Schütt, S. Chmiela, O. A. von Lilienfeld, A. Tkatchenko, K. Tsuda, and K.-R. Müller, in *Machine Learning Meets Quantum Physics* (Springer Lecture Notes in Physics, 2020), Vol. 968.

[12] O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, Nat. Rev. Chem. **4**, 347 (2020).

[13] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, Annu. Rev. Phys. Chem. **71**, 361 (2020).

[14] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[15] J. Behler, S. Lorenz, and K. Reuter, J. Chem. Phys. **127**, 014705 (2007).

[16] J. Behler, J. Chem. Phys. **134**, 074106 (2011).

[17] J. Behler, Phys. Chem. Chem. Phys. **13**, 17930 (2011).

[18] K. V. J. Jose, N. Artrith, and J. Behler, J. Chem. Phys. **136**, 194111 (2012).

[19] J. Behler, J. Chem. Phys. **145**, 170901 (2016).

[20] M. Gastegger, J. Behler, and P. Marquetand, Chem. Sci. **8**, 6924 (2017).

[21] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Nat. Commun. **8**, 13890 (2017).

[22] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017), Vol. 30, pp. 991–1001.

[23] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018).

[24] O. T. Unke and M. Meuwly, J. Chem. Theory Comput. **15**, 3678 (2019).

[25] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).

[26] A. P. Bartók and G. Csányi, Int. J. Quantum Chem. **115**, 1051 (2015).

[27] V. Botu and R. Ramprasad, Phys. Rev. B **92**, 094306 (2015).

[28] Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. **114**, 096405 (2015).

[29] E. V. Podryabinkin and A. V. Shapeev, Comput. Mater. Sci. **140**, 171 (2017).

[30] A. Glielmo, P. Sollich, and A. De Vita, Phys. Rev. B **95**, 214302 (2017).

[31] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Sci. Adv. **3**, e1603015 (2017).

[32] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, Nat. Commun. **9**, 3887 (2018).

[33] J. Wang, S. Chmiela, K.-R. Müller, F. Noé, and C. Clementi, J. Chem. Phys. **152**, 194106 (2020).

[34] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).

[35] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[36] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, J. Chem. Theory Comput. **9**, 3404 (2013).

[37] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, J. Phys. Chem. Lett. **6**, 2326 (2015).

[38] M. Rupp, R. Ramakrishnan, and O. A. von Lilienfeld, J. Phys. Chem. Lett. **6**, 3309 (2015).

[39] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. **18**, 13754 (2016).

[40] N. Artrith, A. Urban, and G. Ceder, Phys. Rev. B **96**, 014112 (2017).

[41] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Sci. Adv. **3**, e1701816 (2017).

[42] K. Yao, J. E. Herr, and J. Parkhill, J. Chem. Phys. **146**, 014106 (2017).

[43] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, J. Chem. Theory Comput. **13**, 5255 (2017).

[44] M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat, and L. Thiry, J. Chem. Phys. **148**, 241732 (2018).

[45] A. Glielmo, C. Zeni, and A. De Vita, Phys. Rev. B **97**, 184307 (2018).

[46] A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, Phys. Rev. Lett. **120**, 036002 (2018).

[47] Y.-H. Tang, D. Zhang, and G. E. Karniadakis, J. Chem. Phys. **148**, 034101 (2018).

[48] W. Pronobis, A. Tkatchenko, and K.-R. Müller, J. Chem. Theory Comput. **14**, 2991 (2018).

[49] P. O. Dral, A. Owens, S. N. Yurchenko, and W. Thiel, J. Chem. Phys. **146**, 244108 (2017).

[50] A. Mardt, L. Pasquali, H. Wu, and F. Noé, Nat. Commun. **9**, 4443 (2018).

[51] F. Noé, S. Olsson, J. Köhler, and H. Wu, Science **365**, eaaw1147 (2019).

[52] H. Chan, B. Narayanan, M. J. Cherukara, F. G. Sen, K. Sasikumar, S. K. Gray, M. K. Y. Chan, and S. K. R. S. Sankaranarayanan, J. Phys. Chem. C **123**, 6941 (2019).

[53] M. Meila, S. Koelle, and H. Zhang, "A regression approach for explaining manifold embedding coordinates," arXiv:1811.11891 (2018).

[54] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. Anatole von Lilienfeld, New J. Phys. **15**, 095003 (2013).

[55] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Nat. Commun. **8**, 872 (2017).

[56] T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, and R. Ramprasad, npj Comput. Mater. **3**, 37 (2017).

[57] T. Bereau, R. A. DiStasio, Jr., A. Tkatchenko, and O. A. Von Lilienfeld, J. Chem. Phys. **148**, 241706 (2018).

[58] N. Lubbers, J. S. Smith, and K. Barros, J. Chem. Phys. **148**, 241715 (2018).

[59] K. Kanamori, K. Toyoura, J. Honda, K. Hattori, A. Seko, M. Karasuyama, K. Shitara, M. Shiga, A. Kuwabara, and I. Takeuchi, Phys. Rev. B **97**, 125124 (2018).

[60] T. S. Hy, S. Trivedi, H. Pan, B. M. Anderson, and R. Kondor, J. Chem. Phys. **148**, 241745 (2018).

[61] J. S. Smith, O. Isayev, and A. E. Roitberg, Chem. Sci. **8**, 3192 (2017).

[62] J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. De Fabritiis, F. Noé, and C. Clementi, ACS Cent. Sci. **5**, 755 (2019).

[63] R. Winter, F. Montanari, F. Noé, and D.-A. Clevert, Chem. Sci. **10**, 1692 (2019).

[64] S. Chmiela, H. E. Sauceda, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, Comput. Phys. Commun. **240**, 38 (2019).

[65] K. Yao, J. E. Herr, D. W. Toth, R. Mckintyre, and J. Parkhill, Chem. Sci. **9**, 2261 (2018).

[66] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, J. Chem. Theory Comput. **15**, 448 (2019).

[67] S. Chmiela, H. E. Sauceda, A. Tkatchenko, and K.-R. Müller, "Accurate molecular dynamics enabled by efficient physically constrained machine learning approaches," in *Machine Learning Meets Quantum Physics*, edited by K. T. Schütt, S. Chmiela, O. A. von Lilienfeld, A. Tkatchenko, K. Tsuda and K.-R. Müller (Springer International Publishing, 2020), pp. 129–154.

[68] H. C. Longuet-Higgins, Mol. Phys. **6**, 445 (1963).

[69] H. E. Sauceda, S. Chmiela, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, "Construction of machine learned force fields with quantum chemical accuracy: Applications and chemical insights," in *Machine Learning Meets Quantum Physics*, edited by K. T. Schütt, S. Chmiela, O. A. von Lilienfeld, A. Tkatchenko, K. Tsuda and K.-R. Müller (Springer International Publishing, 2020), pp. 277–307.

[70]H. E. Sauceda, V. Vassilev-Galindo, S. Chmiela, K.-R. Müller, and A. Tkatchenko, "Dynamical strengthening of covalent and non-covalent molecular interactions by nuclear quantum effects at finite temperature," arXiv:2006.10578 (2020).

[71]J. E. Lennard-Jones, Proc. Phys. Soc. **43**, 461 (1931).

[72]P. P. Ewald, Ann. Phys. **369**, 253 (1921).

[73]R. W. Hockney and J. W. Eastwood, *Computer Simulations Using Particles* (McGraw-Hill, 1981).

[74]L. Greengard and V. Rokhlin, J. Comput. Phys. **73**, 325 (1987).

[75]A. Arnold and C. Holm, in *Advanced Computer Simulation Approaches for Soft Matter Sciences II* (Springer, 2005), pp. 59–109.

[76]N. Artrith, T. Morawietz, and J. Behler, Phys. Rev. B **83**, 153101 (2011).

[77]A. Grisafi and M. Ceriotti, J. Chem. Phys. **151**, 204105 (2019).

[78]W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman, Annu. Rev. Biophys. Biomol. Struct. **30**, 211 (2001).

[79]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[80]A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).

[81]V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Comput. Phys. Commun. **180**, 2175 (2009).

[82]J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, and T. D. Crawford, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 556 (2012).

[83]R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Theory Comput. **13**, 3185 (2017).

[84]D. G. A. Smith, L. A. Burns, D. A. Sirianni, D. R. Nascimento, A. Kumar, A. M. James, J. B. Schriber, T. Zhang, B. Zhang, A. S. Abbott, E. J. Berquist, M. H. Lechner, L. A. Cunha, A. G. Heide, J. M. Waldrop, T. Y. Takeshita, A. Alenaizan, D. Neuhauser, R. A. King, A. C. Simmonett, J. M. Turney, H. F. Schaefer, F. A. Evangelista, A. E. DePrince, T. D. Crawford, K. Patkowski, and C. D. Sherrill, J. Chem. Theory Comput. **14**, 3504 (2018).

[85]A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus *et al.*, J. Phys.: Condens. Matter **29**, 273002 (2017).

[86]A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).

[87]T. Hassinen and M. Peräkylä, J. Comput. Chem. **22**, 1229 (2001).

[88]N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, J. Cheminf. **3**, 33 (2011).

[89]R. Salomon-Ferrer, D. A. Case, and R. C. Walker, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 198 (2013).

[90]R. J. Woods and R. Chappelle, J. Mol. Struct.: THEOCHEM **527**, 149 (2000).

[91]C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[92]G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley, and J. Mantzaris, J. Chem. Phys. **89**, 2193 (1988).

[93]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09 Revision E.01, Gaussian Inc., Wallingford, CT, 2009.

[94]G. Bussi and M. Parrinello, Phys. Rev. E **75**, 056707 (2007).

[95]A. Choudhary, K. J. Kamer, and R. T. Raines, J. Org. Chem. **76**, 7933 (2011).

[96]R. W. Newberry and R. T. Raines, Acc. Chem. Res. **50**, 1838 (2017).

[97]J. Behler, Int. J. Quantum Chem. **115**, 1032 (2015).

[98]A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox and R. Garnett (Curran Associates, Inc., 2019), Vol. 32, pp. 8024–8035.

[99]J. Hermann, R. A. DiStasio, Jr, and A. Tkatchenko, Chem. Rev. **117**, 4714 (2017).