Learning from the Density to Correct Total Energy and Forces in First Principle Simulations

Sebastian Dick and Marivi Fernandez-Serra

Physics and Astronomy Department, Stony Brook University,
Stony Brook, New York 11794-3800, United States   and
Institute for Advanced Computational Science, Stony Brook University,
Stony Brook, New York 11794-3800, United States

We propose a new molecular simulation framework that combines the transferability, robustness and chemical flexibility of an ab initio method with the accuracy and efficiency of a machine learned force field. The key to achieve this mix is to use a standard density functional theory (DFT) simulation as a pre-processor for the atomic and molecular information, obtaining a good quality electronic density. General, symmetry preserving, atom-centered electronic descriptors are then built from this density to train a neural network to correct the baseline DFT energies and forces. These electronic descriptors encode much more information than local atomic environments, allowing a simple neural network to reach the accuracy required for the problem of study at a negligible cost. The balance between accuracy and efficiency is determined by the baseline simulation. This is shown in results where high level quantum chemical accuracy is obtained for simulations of liquid water at standard DFT cost, or where high level DFT-accuracy is achieved in simulations with a low-level baseline DFT calculation, at a significantly reduced cost.

The development of atomistic simulation methods that make use of machine learning (ML) techniques to achieve close or beyond ab initio level accuracy at a highly reduced computational cost is proving to be an efficient and robust alternative to traditional force field approaches [1–7]. This high level of activity is facilitated and supported by the availability and continuous generation of larger and improved data bases of highly accurate calculations or experiments both for molecular [8–10] and solid systems [11–13]. While in many cases, newly proposed methodologies and force-fields are only at the proof of concept stage, some of the most promising methods have already produced results that provide new physical insight into challenging problems [14]. Beyond predicting atomic structures of complex materials [15], they have been used to address problems previously inaccessible to ab initio level calculations such as the structure of the surface of water, or its temperature dependent dynamical properties using the MB-pol model [16–20]. These ab-initio quality force fields are however limited in their scope, incapable of being generalized and often, by design [18], cannot treat reactive environments. At the same time, it is also true that density functional theory (DFT)-based methods within their lower rungs of approximation to the exchange and correlation (XC) energy struggle in simulating systems like water or ionic solutions [21–24]. The use of more expensive XC approximations, such as hybrid functionals [25], may deliver the required accuracy, however the size of systems for which these simulations are feasible is limited. This trade-off between accuracy, flexibility and computational cost seems to lie at the heart of many problems the molecular science community is presently facing.

Within DFT, some progress has been made in the development of machine-learned density functionals. Apart from promising efficient orbital-free DFT calculations by learning the kinetic energy functional [26–28], these methods have been shown to allow to skip self-consistent calculations altogether by directly learning the map from atomic potential to electron density [29, 30]. However, applications still appear to be limited to one-dimensional systems and small molecules and, to our knowledge, none of the mentioned approaches is currently capable of treating condensed systems.

Faced with these challenges, we have developed a new method called Machine Learned Correcting Functionals (MLCFs) which, rather than replacing DFT, complements it by adding a machine learned functional of the electron density. The principal concept of this new approach could be termed "Informed Machine Learning." We argue that all the available physical insight into a problem should be used to help the ML-model make its predictions and minimize its generalization error.

In MLCFs, this translates to DFT assuming the role of a data-preprocessor that maps the atomic coordinates to a self-consistent electron density which is input into an artificial neural network (ANN). The ANN is then fitted to the difference in energy and forces between the baseline method which was used to obtain the electron density and a higher-level reference method of choice. Thus MLCFs draw upon and optimize the strengths of DFT, which is effective in calculating mean field electrostatic interactions, but often fails on accurately describing non local interactions rooted in quantum mechanics. At the same time, DFT provides a solid foundation which the ML-model can fall back on when faced with unseen scenarios in which it will likely fail to make reliable predictions. With this method, we seek to mitigate some of the challenges more traditional ML force fields are facing, namely the explicit treatment of Coulomb interactions in periodic systems and the need for vast datasets to cover the relevant configuration space [1].

Charge density representation.—As in any machine learning application, the representation of the data used

	Oxygen	Hydrogen
# radials	2	2
l_{max}	2	1
r_i	0.05	0.0
r_o	1.0	1.5

TABLE I. Hyperparameters used to create the charge density descriptors. Radii are given in units of Angstrom.

for input (aka, the features), should encode as much information in as few variables as possible, to avoid redundancies and make sure that the ANN model is quick to train and not prone to overfitting. Furthermore, we require the number of descriptors to scale linearly with the number of atoms in the system to guarantee size extensivity of the energy. To ensure that the model respects rotational symmetry we further impose the constraint that descriptors need to transform covariantly under $SO(3)$ rotations.

One way to create such a descriptor is to project the electron density in real space onto a set of atom-centered orbitals, which were in inspired by Bartok et al.'s work on representing chemical environments [31]. While completing this manuscript, the authors became aware of the work by Grifasi et al. [30] which uses a very similar approach.

The angular and radial basis functions are given by spherical harmonics $Y_l^m(\theta, \phi)$ and

$$\tilde{\zeta}_n(r) = \begin{cases} \frac{1}{N}(r - r_i)^2(r_o - r)^{n+2} & \text{for } r > r_i \text{ and } r < r_o \\ 0 & \text{else} \end{cases}$$
(1)

respectively. Apart from an outer cutoff radius r_o and a normalization factor N, the latter being determined by numerical integration, we also found it beneficial to include an inner cutoff radius r_i in order to disregard the core area. It should be noted that even though larger cutoff radii will generally lead to a better fit of the training data, they also cause the transferability of the obtained model to decrease.

After orthogonalizing the radial functions with the transformation

$$\zeta_n(r) = \sum_{n'} S_{nn'}^{-1/2} \tilde{\zeta}_n(r), \quad \text{where } S_{nn'} = \int dr \tilde{\zeta}_n(r)\tilde{\zeta}_{n'}(r).$$
(2)

the full basis is given as $\psi_{nlm}(\vec{r}) = Y_l^m(\theta, \phi)\zeta_n(r)$.

The descriptors $c_{nlm}^{\alpha, I}$ for atom I of species α at position $\vec{r}_I$ can be obtained by projecting the electron density ρ onto the corresponding basis functions ψ_{nlm}^α.

$$\tilde{c}_{nlm}^{\alpha, I} = \int_{\vec{r}} \rho(\vec{r} - \vec{r}_{\alpha, I})\psi_{nlm}^{*\alpha}(\vec{r}).$$
(3)

Introduction of an atomic species label α has made it possible to have different bases for distinct atomic species.
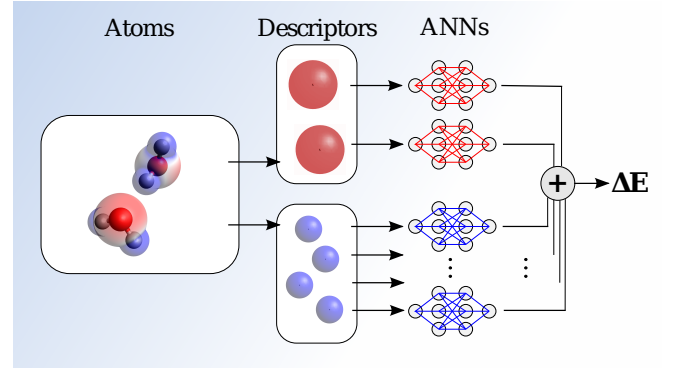

FIG. 1. Architecture of a Behler-Parrinello type neural network. Starting with a given configuration the self-consistent electron density is computed with the baseline DFT. The density is then encoded in descriptors that are fed into artificial neural networks (ANNs) which calculate atomic contributions to the total energy correction ΔE

To ensure rotational invariance, the basis should be aligned with a uniquely defined local coordinate system (LCS). Knowing the Euler angles $\{\alpha, \beta, \gamma\}$ that relate the global coordinate system (GCS) to the LCS, the rotated descriptors are given as

$$c_{nlm} = \sum_{m'} (D^l(\alpha, \beta, \gamma))^{-1}_{mm'} \tilde{c}_{nlm'},$$
(4)

with D^l being the Wigner D-matrix for angular momentum l [32].

The LCS of a given atom can be defined trough the position of atoms in its immediate environment. Even though this approach has been proven to work for machine learned force fields [3], it would be desirable to have a definition of the LCS that is completely independent of the atomic coordinates and only relies on electronic information.

Recognizing that the descriptors associated with the p-orbitals $\tilde{c}_{n1m}$ transform like vectors under $SO(3)$ rotations, we can use these vectors to define a local coordinate system that only depends on the electronic structure (see SI for more details).

Aligning the descriptors independently of structural variables has the major advantage that conformers that are close in their electron densities are described by similar features. This is particulary important for MLCFs that only correct the exchange correlation energy E_{xc}, as these conformers will necessarily exhibit the same error in E_{xc}. Conversely, a MLCF that goes beyond local correction, e.g. one that can interpolate between different basis sets, will possibly profit from additional structural information that is implicitly added by using the nearest-neighbor rule for alignment.

The ideal number of radial and angular basis functions as well as the inner and outer cutoff radii can be determined by cross-validation. Doing so, the hyperparameters listed in Tab. I were found to provide a good compromise between accuracy and computational efficiency.
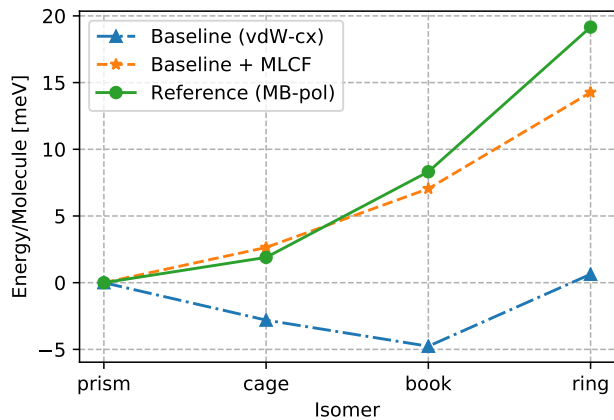
FIG. 2. Energies of water hexamers with respect to the prism isomer, which is correctly determined to be the most stable structure by the MLCF.
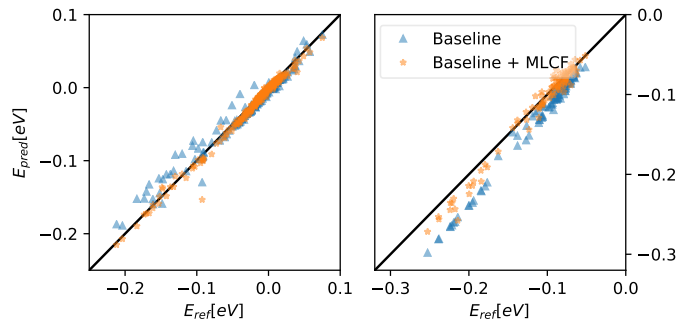

FIG. 3. Predicted vs. reference two (left) and three (right) body energies for the datasets taken from ref. 38. For the two body energies the RMSE is 4.20 for the baseline method and 2.80 for the MLCF corrected method. For the three body energies the errors are 7.10 and 3.56 respectively. Errors are given in meV/Molecule.

Machine learning models.—Inspired by the well-established Behler-Parrinello networks [33], the energy functional for a given number of molecules was represented as an artificial neural network (ANN) consisting of a sum of smaller atomic networks, each of which only sees the local charge density around the atom it is associated with (see Fig.1). Furthermore, networks belonging to atoms of the same species are identical to guarantee permutational invariance. Separate models were trained directly on the force corrections to be used in molecular dynamics simulations. Details about the force models are provided in the supplementary information.

Both models were trained using the Adam [34] optimizer with training rate $\alpha = 0.001$ and decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The number and size of layers as well as the training rate was determined using cross-validation with early stopping. A network with three layers of 8 nodes and one with three layers of 16 nodes were determined to have the optimal accuracy for energy and force predictions respectively, however satisfying results were obtained with a wide variety of model sizes. This suggests that even small networks are able to extract all relevant information from the input data and that the residual error cannot be explained by correlations in the local charge density. The sigmoid function was chosen as activation

$$f(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

as it was determined to be less prone to over-fitting than other possible choices such as the hyperbolic tangent.

Performance testing.—We tested our method on water clusters of varying size. The baseline calculations were conducted using SIESTA [35] with a quadruple zeta doubly polarized basis set and a van der Waals density functional of Dion et al. [36] with GGA exchange modified by K. Berland and P. Hyldgaard (vdW-cx) [37]. This XC functional has been shown to be as good as other vdW-type functionals in describing liquid water[38]. For the

reference energies, we use the MB-pol force field [16–18] which is fitted to highly accurate quantum chemical calculations of water monomers and clusters. It is important to emphasize that MB-pol has been shown to accurately reproduce the structural and thermodynamic properties of condensed phases of water, hence it is a superior model to any vdW or GGA XC functional.

The MLCF was trained to interpolate between the baseline and reference method, i.e. the target values used for fitting are given as the differences in energy $\Delta E = E_{ref} - E_{base}$.

The dataset comprised about 400 water monomers, 2000 dimers and 1500 trimer configurations that were all sampled from the data that was used to fit MB-pol [16–18] and an additional 300 monomers with a uniform distribution of bond lengths and angles. The data was split into 80 percent training and 20 percent hold-out set. We further created samples of larger water clusters with $n = 4, 5, 8, 16$ molecules (50 samples each) for testing the size extensivity of the network. In Tab. II the errors in relative energies of isomers are split up into root mean square (RMS) error, mean absolute error (MAE) and maximum (absolute) error.

No. Molecules	RMSE	MAE	max. Error
1	2.29 (53.19)	1.25 (43.13)	14.71 (151.19)
2	4.33 (40.17)	2.91 (31.28)	31.03 (136.91)
3	2.89 (28.25)	2.16 (22.29)	12.27 (75.20)
4	2.79 (9.69)	2.15 (7.93)	7.70 (24.95)
5	2.64 (11.24)	2.19 (8.97)	6.23 (36.69)
8	3.43 (9.26)	2.72 (7.34)	8.05 (22.67)
16	2.75 (6.28)	2.15 (5.03)	6.19 (17.15)

TABLE II. Model performance on test sets containing clusters of different sizes. Reported are the root mean square error (RMSE), mean absolute error (MAE) and, maximum absolute error (max. Error). All errors are normalized by the number of molecules and given in meV. The baseline method errors are given in parenthesis.

As can be seen in Tab. II, in spite of only being trained

on systems with up to three molecules, our model performs well for all cluster sizes and significantly decreases errors in relative energies of isomers, all RMS errors being below 5 meV/Molecule and mean absolute errors below 3 meV/Molecule. Moreover, Fig. 2 shows that the overall energy ordering of hexamer structures is correctly reproduced by the MLCF, a task that has so far proven largely elusive to GGA-level DFT calculations[38].

To gain further insights, we can follow the approach of previous works [38, 39] and split up the total energy of water clusters into their n-body contributions. This was done on the same datasets that were used in ref. 38 to optimize the GGA exchange enhancement factor for water systems.

Figure 3, shows that our model not only corrects the predominant errors in the one body energy (Tab. II) but also improves two and three body energies. This might seem surprising as the MLCF is only provided with the local charge density around each atom. However, the model seems to learn how to infer higher order interaction energies from correlations in the local charge density. This indicates that long range, and possibly non local interactions are encoded within local density spatial correlations. This enables us to use radial cutoffs for the descriptors that are very short, comparable to the nearest neighbor distance (see tableI). In order to obtain similar accuracy using only structural information, much larger cutoffs are needed[31, 39].

Having proven successful for gas-phase calculations, we went on to assess the method's performance in molecular dynamics simulations on liquid water, in particular its ability to correctly reproduce thermodynamic averages given by pair correlation functions. For reasons of practicality we opted to replace the previously employed van der Waals functional with the faster Perdew-Burke-Ernzerhof (PBE) [40] functional. We trained a MLCF to interpolate between this baseline method and MB-pol, which has been show to produce pair correlation functions close to experimental results. Further details about the training dataset used are given in the SI.

Figure 4 was produced by running DFT-based molecular dynamics (MD) simulations of 128 water molecules at experimental density in a periodic box using a Nose-Hoover thermostat [41] at a temperature of 300 K. The simulation length is 10 ps with a timestep of 0.5 fs using 4 ps for equilibration and 6 ps to create the correlation functions. We see that our MLCF provides a significant improvement over PBE and even vdW-cx, which are both known to over-structure liquid water. The MCLF approach also corrects the dynamical properties of the baseline calculation, bringing them to the level of the reference method. This can be seen in the vibrational density of states plots shown in the SI.

Beyond correcting the density functional approximation, MLCFs can also be used to extrapolate between two given basis sets. Thus, extremely fast calculations can be performed with minimal (single zeta) basis-sets that with the help of MLCFs are then brought to the
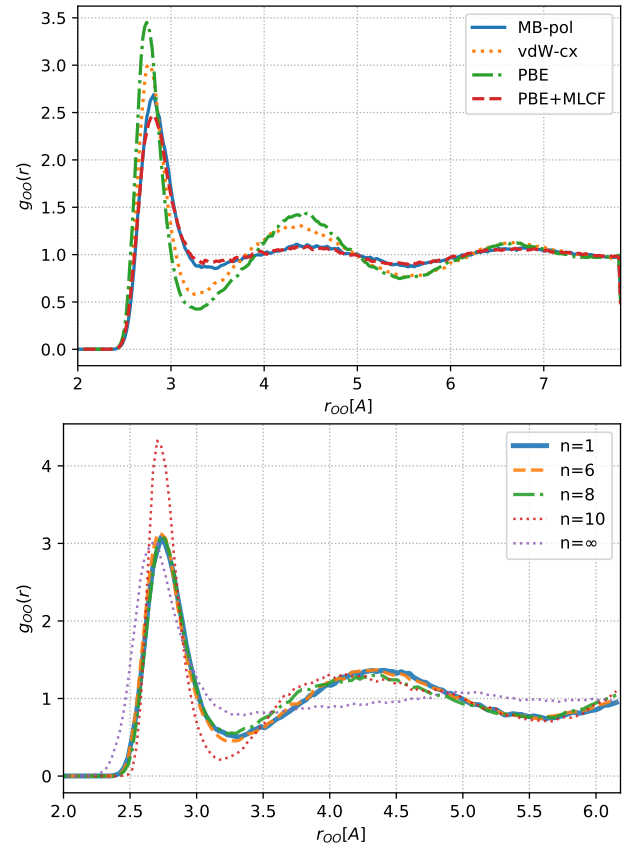

FIG. 4. Oxygen-oxygen radial distribution functions $g_{OO}(r)$ created with different density functional approximations. Top: The MLCF is used to interpolate between PBE and MB-pol. Bottom: Using the mixing method proposed by Anglada et al. [42] the radial distribution function remain unchanged for $n <= 6$ and reasonable results can be achieved up to $n <= 8$. Other pair correlation functions as well as the vibrational density of states can be found in the SI.

accuracy of more conservative basis sizes.

Here it should be noted that MLCFs rely upon the quality of the electron density used as input. Following an approach by Kim et al. [43], the error that the MLCF is trained to correct can be decomposed as

$$\Delta E = E_{ref}[n_{ref}] - E_{base}[n_{base}] = \Delta E_F + \Delta E_D \quad (6)$$

with $E_F = E_{ref}[n_{base}] - E_{base}[n_{base}]$ and $E_D = E_{ref}[n_{ref}] - E_{ref}[n_{base}]$ being the functional and density driven error respectively. By design, MLCFs are only effective at correcting functional driven errors. This generally does not pose any serious problems, as most density functional approximations yield a good quality electron density. Using very sparse basis sets, however, imposes an upper limit on the accuracy of the density, making density driven errors become dominant.

It turns out that these shortcomings can be circumvented by using a method developed by Anglada et al. [42]. In their paper, they propose to alternate fast steps calculated with a cheaper, less accurate method with slow

steps obtained with a well converged method. The forces used to integrate the equations of motion with a time step Δt are given as $\vec{F}(t) = \vec{F}_{fast}(t) + \Delta \vec{F}(t)$ where the correction term $\Delta \vec{F}(t)$ is defined as

$$\Delta \vec{F} = \begin{cases} n\delta \vec{F} & \text{if } (t/\Delta t \mod n = 0) \\ 0 & \text{else} \end{cases} \quad (7)$$

with $\delta \vec{F} = \vec{F}_{slow} - \vec{F}_{fast}$. The motivation behind this idea is that the error $\delta \vec{F}$ is small compared to the absolute forces and relatively smooth over time. Therefore, it is sufficient to only correct for this error every n steps, the limiting value of n being dependent on both the system and the accuracy of the fast method.

To test this, we ran AIMD simulations of 64 water molecules, at 300K and experimental density for 10 ps. The fast steps were calculated using the PBE exchange-correlation functional with a single zeta basis set and subsequently corrected with an adequately trained MLCF. For the slow, well-converged steps we used the vdW-cx functional and a double zeta polarized basis set. It is important to point out that the upper limit for the number of subsequent fast steps n_{lim} is strongly system-

dependent and rigorous testing should be conducted before using this mixing scheme. However, if carefully used, this method promises significant speed-ups that are roughly equal to n_{lim} for large system sizes (see SI for detailed analysis).

In conclusion, we have presented and tested a new molecular simulation framework where the electronic density is used to train a neural network to correct baseline DFT energies and forces to the accuracy provided by a higher level method. To do so we defined a set of electronic density descriptors, which encode much richer information than provided by local atomic environments. In particular, we have shown that real space semilocal correlations of the electronic density contain information about longer range effects. These can be corrected through the use of a neural network trained on purely on local data, an observation that might support the potential for developing machine learned density functionals using local density descriptors [30] and total energy targets.

[1] Jörg Behler. Perspective: Machine learning potentials for atomistic simulations. *The Journal of chemical physics*, 145(17):170901, 2016.

[2] Thuong T Nguyen, Eszter Székely, Giulio Imbalzano, Jörg Behler, Gábor Csányi, Michele Ceriotti, Andreas W Götz, and Francesco Paesani. Comparison of permutationally invariant polynomials, neural networks, and gaussian approximation potentials in representing water interactions through many-body expansions. *The Journal of Chemical Physics*, 148(24):241725, 2018.

[3] Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and E Weinan. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical review letters*, 120(14):143001, 2018.

[4] Stefan Chmiela, Alexandre Tkatchenko, Huziel E Sauceda, Igor Poltavsky, Kristof T Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science advances*, 3(5):e1603015, 2017.

[5] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, pages 991–1001, 2017.

[6] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O Anatole Von Lilienfeld, Klaus-Robert Muller, and Alexandre Tkatchenko. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *The journal of physical chemistry letters*, 6(12):2326–2331, 2015.

[7] Justin S Smith, Olexandr Isayev, and Adrian E Roitberg. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chemical science*, 8(4):3192–3203, 2017.

[8] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

[9] Mohammad M Ghahremanpour, Paul J Van Maaren, and David Van Der Spoel. The alexandria library, a quantum-chemical database of molecular properties for force field development. *Scientific data*, 5:180062, 2018.

[10] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1:140022, 2014.

[11] Scott Kirklin, James E Saal, Bryce Meredig, Alex Thompson, Jeff W Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *npj Computational Materials*, 1:15010, 2015.

[12] Kirsten Winther, Max J Hoffmann, Osman Mamun, Jacob R Boes, Jens K Nørskov, Michal Bajdich, Thomas Bligaard, Michal Bajdich, et al. Catalysis-hub. org: An open electronic structure database for surface reactions. 2018.

[13] Andriy Zakutayev, John Perkins, Marcus Schwarting, Robert White, Kristin Munch, William Tumas, Nick Wunder, and Caleb Phillips. High throughput experimental materials database. Technical report, National Renewable Energy Laboratory-Data (NREL-DATA), Golden, CO (United , 2017.

[14] Matti Hellström, Vanessa Quaranta, and Jörg Behler. One-dimensional vs. two-dimensional proton transport processes at solid–liquid zinc-oxide–water interfaces. *Chemical Science*, 2019.

[15] Volker L Deringer, Noam Bernstein, Albert Bartok, Matthew J Cliffe, Rachel N Kerber, Lauren E Marbella, Clare P Grey, Stephen R Elliott, and Gábor Csányi. Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics. *The journal of physical chemistry letters*, 2018.

[16] Volodymyr Babin, Claude Leforestier, and Francesco Paesani. Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient. *Journal of Chemical Theory and Computation*, 9(12):5395–5403, 2013.

[17] Volodymyr Babin, Gregory R. Medders, and Francesco Paesani. Development of a "first principles" water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters. *Journal of Chemical Theory and Computation*, 10(4):1599–1607, 2014.

[18] Gregory R. Medders, Volodymyr Babin, and Francesco Paesani. Development of a first principles water potential with flexible monomers. III: Liquid phase properties. *Journal of Chemical Theory and Computation*, page 140701133607002, 2014.

[19] Shumei Sun, Fujie Tang, Sho Imoto, Daniel R. Moberg, Tatsuhiko Ohto, Francesco Paesani, Mischa Bonn, Ellen H. G. Backus, and Yuki Nagata. Orientational distribution of free o-h groups of interfacial water is exponential. *Phys. Rev. Lett.*, 121:246101, Dec 2018.

[20] Sandeep K Reddy, Daniel R Moberg, Shelby C Straight, and Francesco Paesani. Temperature-dependent vibrational spectra and structure of liquid water from classical and quantum simulations with the mb-pol potential energy function. *The Journal of Chemical Physics*, 147(24):244504, 2017.

[21] Robert A DiStasio Jr, Biswajit Santra, Zhaofeng Li, Xifan Wu, and Roberto Car. The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *The Journal of chemical physics*, 141(8):84502, 2014.

[22] Michael J Gillan, Dario Alfè, and Angelos Michaelides. Perspective: How good is DFT for water? *The Journal of chemical physics*, 144(13):130901, 2016.

[23] Yi Yao and Yosuke Kanai. Free energy profile of nacl in water: First-principles molecular dynamics with scan and ωb97x-v exchange–correlation functionals. *Journal of chemical theory and computation*, 14(2):884–893, 2018.

[24] Mohan Chen, Hsin-Yu Ko, Richard C Remsing, Marcos F Calegari Andrade, Biswajit Santra, Zhaoru Sun, Annabella Selloni, Roberto Car, Michael L Klein, John P Perdew, et al. Ab initio theory and modeling of water. *Proceedings of the National Academy of Sciences*, 114(41):10846–10851, 2017.

[25] Axel D Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of chemical physics*, 98(7):5648–5652, 1993.

[26] Junji Seino, Ryo Kageyama, Mikito Fujinami, Yasuhiro Ikabata, and Hiromi Nakai. Semi-local machine-learned kinetic energy density functional with third-order gradients of electron density. *The Journal of Chemical Physics*, 148(24):241705, 2018.

[27] John C Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. Finding density functionals with machine learning. *Physical review letters*, 108(25):253002, 2012.

[28] Kun Yao and John Parkhill. Kinetic energy of hydrocarbons as a function of electron density and convolutional neural networks. *Journal of chemical theory and computation*, 12(3):1139–1147, 2016.

[29] Felix Brockherde, Leslie Vogt, Li Li, Mark E Tuckerman, Kieron Burke, and Klaus-Robert Müller. Bypassing the kohn-sham equations with machine learning. *Nature communications*, 8(1):872, 2017.

[30] Andrea Grisafi, David M Wilkins, Benjamin AR Meyer, Alberto Fabrizio, Clemence Corminboeuf, and Michele Ceriotti. A transferable machine-learning model of the electron density. *arXiv preprint arXiv:1809.05349*, 2018.

[31] Albert P Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Physical Review B*, 87(18):184115, 2013.

[32] Eugene Wigner. On unitary representations of the inhomogeneous lorentz group. *Annals of mathematics*, pages 149–204, 1939.

[33] Jörg Behler and Michele Parrinello. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters*, 98(14):146401, apr 2007.

[34] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[35] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The siesta method for ab initio order-n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.

[36] Max Dion, Henrik Rydberg, Elsebeth Schröder, David C Langreth, and Bengt I Lundqvist. Van der waals density functional for general geometries. *Physical review letters*, 92(24):246401, 2004.

[37] Kristian Berland and Per Hyldgaard. Exchange functional that tests the robustness of the plasmon description of the van der waals density functional. *Physical Review B*, 89(3):035412, 2014.

[38] Michelle Fritz, Marivi Fernández-Serra, and José M Soler. Optimization of an exchange-correlation density functional for water. *The Journal of chemical physics*, 144(22):224101, 2016.

[39] Albert P Bartók, Michael J Gillan, Frederick R Manby, and Gábor Csányi. Machine-learning approach for one- and two-body corrections to density functional theory: Applications to molecular and condensed water. *Physical Review B*, 88(5):054104, 2013.

[40] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[41] Glenn J Martyna, Michael L Klein, and Mark Tuckerman. Nosé–hoover chains: The canonical ensemble via continuous dynamics. *The Journal of chemical physics*, 97(4):2635–2643, 1992.

[42] Eduardo Anglada, Javier Junquera, and José M Soler. Efficient mixed-force first-principles molecular dynamics. *Physical Review E*, 68(5):055701, 2003.

[43] Min-Cheol Kim, Eunji Sim, and Kieron Burke. Understanding and reducing errors in density functional calculations. *Physical review letters*, 111(7):073003, 2013.

Learning from the Density to Correct Total Energy and Forces in First Principle Simulations - Supplementary Information

Sebastian Dick and Marivi Fernandez-Serra

Physics and Astronomy Department, Stony Brook University,
Stony Brook, New York 11794-3800, United States and
Institute for Advanced Computational Science, Stony Brook University,
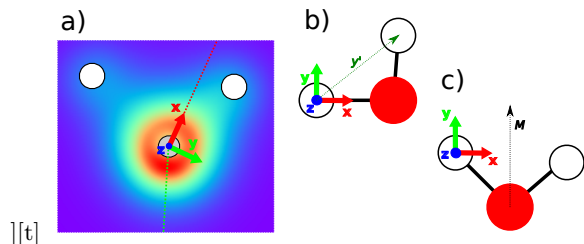*Stony Brook, New York 11794-3800, United States*

FIG. 1. Definition of the a) electronic, b) nearest-neighbor and c) molecular local coordinate system

I. COORDINATE SYSTEMS

In this work, several definitions of a local coordinate system (LCS) are used.

a. Electronic Given a set of descriptors $\{\tilde{c}_{nlm}\}$ for a given atom the descriptors associated with the p-orbitals $\tilde{c}_{n1m}$ transform like vectors under $SO(3)$ rotations. Starting with $n = 1$ and $l = 1$ the corresponding descriptor vector $\tilde{c}_{11m}$, after being brought into its real form and normalized, is used as the first local axis. Increasing n allows one to find a second vector of $l = 1$ descriptors that, using Gram-Schmidt orthogonalization, defines the second local axis.

In the example given in Fig.1a), the first vector $\tilde{c}_{11}$ (dashed red line) picks up on the asymmetry of the covalent bonds and is bent towards the closer Hydrogen atom, whereas the second vector $\vec{\tilde{c}}_{21}$ (dashed green line) points towards the lone pair. Problems may arise if all p-descriptors are collinear, as is the case for highly symmetric systems such as a water monomer with covalent bonds of equal length. In cases like this, one has to revert to using the nearest neighbor rule as described below to determine the second axis. However, extended systems are rarely this symmetrical and an increase in the number of radial functions is usually enough to resolve issues of linear dependence and find a set of p-descriptors that span a coordinate system.

b. Nearest Neighbor To determine the LCS around a given atom, one can also use structural information about the atom's immediate environment. The first axis is defined as the direction to the atom's nearest neighbor. The second axis is simply given by the direction to the next nearest neighbor orthogonalized with respect to axis number one. If all three atoms happen to lie on a line, more distant atoms are taken into consideration until non-colinearity is achieved

c. Molecular For non-reactive models, the molecular axes can be used as a local coordinate system. The axes assignment is hereby arbitrary but following convention we define the y-axis as the bisector (M) of the HOH-angle, the x-axis as the one parallel to the molecular plane pointing inwards, and the z-axis orthogonal to it (see Fig.1c).

II. FORCE MODEL

Given that the mapping from atomic configuration to electron density is obtained by conducting a full self-consistent DFT calculation, the forces cannot be derived from the energy functional, as the derivatives $\partial \Delta E / \partial r_{i,\alpha} = \Delta F_{i,\alpha}$ are unknown.

We therefore have to train a model to directly predict the force corrections $\Delta \vec{F}$, where again each atomic species is treated with a separate model. This has the advantage that additional errors that may arise when taking the derivative of the energy can be avoided and models trained directly on forces generally fair better in their prediction than the ones trained on energies. However, as there is no coupling between energy and force models, the resulting method does not exactly preserve denergy and momentum. These effects are due to and thus comparable in magnitude to the fitting error of the force MLCF which in our case is of the order of 0.01 eV/Å. As an ad-hoc solution for our molecular dynamics simulations we manually set the mean force acting on the system to zero at every time step, making sure that the relative acceleration between Oxygen and Hydrogen atoms remained unaltered. Ongoing research is being conducted into how energy and momentum conservation can be included in the fitting procedure.

III. MOLECULAR DYNAMICS

To test our method's performance in molecular dynamics simulations we trained an MLCF to interpolate between PBE [1] (baseline) and MB-pol [2–4] (reference). The previously employed quadruple zeta doubly polarized was replaced by a faster double zeta polarized (DZP) basis set and the real space mesh cutoff energy was reduced from 300 Ry to 200 Ry. This real space cutoff can be interpreted as the upper limit for the kinetic energy of
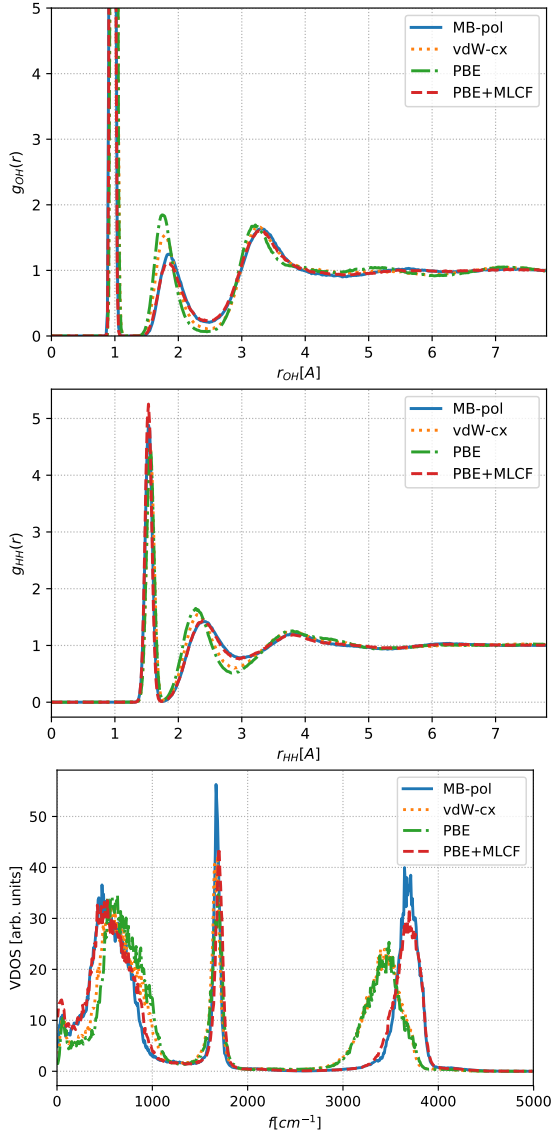
FIG. 2. Radial distribution functions and vibrational density of states using an MLCF with PBE as baseline and MB-pol as reference method
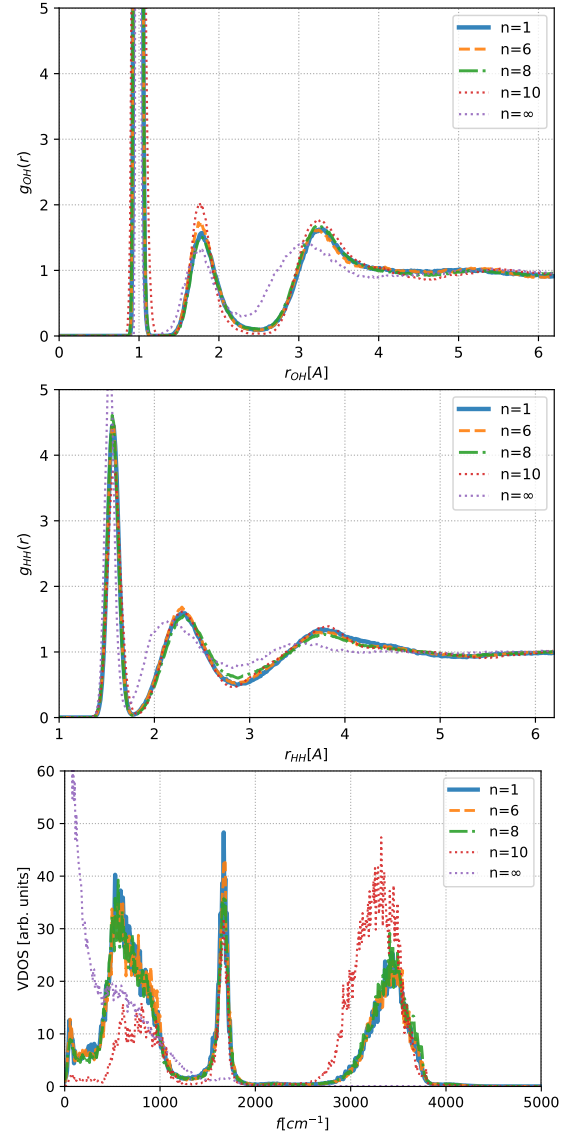

FIG. 3. Mixing method where fast steps are calculated with PBE and a single zeta (SZ) basis set and slow steps are calculated with vdW-cx and a double zeta polarized (DZP) basis set.

plane waves that can be represented on this grid without aliasing.

Moving to condensed phase systems, the electron density and thereby the descriptors undergo significant changes, making an MLCF trained on only small clusters inaccurate. The most obvious remedy is to simply train the MLCF on condensed phase data. This, however, has the drawback that reference calculations become increasingly costly. We avoided this issue by training the model on electron densities of $16(H_2O)$ clusters that were electrostatically embedded in liquid water modeled by TIP4P/2005 [5], opening the possibility of using more expensive, wave function based methods to perform reference calculations in the future. In detail, this means that a dataset was created by sampling a random molecule and its 15 nearest neighbors from snapshots

of an MD simulation and treating all remaining water molecules in the snapshot with TIP4P/2005. Doing so, the number of hydrogen bonds that the central molecule forms with its neighbors was used as a stratifying parameter during the sampling. For the descriptor alignment we chose to employ the aforementioned nearest-neighbor rule. To obtain the target energies and forces, the embedding was removed and calculations were performed on clusters only, thereby restricting the correction to short-range effects. Subsequently, only the central molecule was used in fitting the force corrections. It should be noted that this embedding scheme is only possible because MLCFs explicitly use the local electron density as descriptors and it is therefore not applicable for most other machine-learned force-fields that only depend on
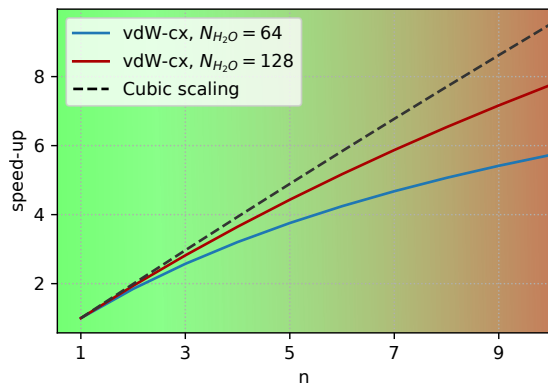
FIG. 4. Speed-up obtained by mixing PBE(SZ) with vdW-cx(DZP) for finite size systems compared to the speed-up that would be achieved for an algorithm that scales strictly cubic in the number of oribtals. The color gradient indicates the method's reliability to reproduce the reference method for a given value of n.

structural information.

For the MD simulations that used a single zeta (SZ) basis as baseline method we found that a model trained on smaller clusters (monomers, dimers and trimers) exhibited sufficient accuracy. Including larger clusters or using embedding did not lead to any signifant improvement. We accredit this to the 'stiffness' of the SZ basis set which is not capable of describing bonding and polarization effects, making electron densities of small clusters and condensed phases practically indistinguishable. We found, however, that using molecular local coordinate systems (as opposed to nearest neighbor or electronic ones) lead to a slight improvement in model performance.

In addition to the Oxygen-Oxygen pair correlation function which can be found in the main text, Fig.2 and Fig. 3 show the OH and HH correlation functions and the vibrational density of states (VDOS) which was obtained by taking the fourier-transform of the velocity autocorrelation function.

IV.   SPEED-UP

The speed-up that can be achieved with the mixing method by Anglada et al. [6] depends crucially on both the mixing parameter n and the ratio between the time it takes to propagate the system with the fast method t_f and the time it takes with the slower, accurate method t_s. Alternating (n-1) fast steps with one slow step, the speed-up η that can be obtained is given by

$$\eta = \frac{nt_s}{(n-1)t_f + t_s}. \tag{1}$$

Fig. 4 shows the speed-up obtained if PBE with SZ basis and vdW-cx [7] with DZP are mixed for which $t_f = 7s$ and $t_s \approx 75s$. The values are obtained for calculations on 16 Intel(R) Xeon(R) CPU E5-2620 CPUs and compared to the theoretical speed-up for an algorithm that is strictly cubic scaling in the number of orbitals. We see that for large enough system sizes $\eta \approx n$.

[1] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[2] Volodymyr Babin, Claude Leforestier, and Francesco Paesani. Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient. *Journal of Chemical Theory and Computation*, 9(12):5395–5403, 2013.

[3] Volodymyr Babin, Gregory R. Medders, and Francesco Paesani. Development of a "first principles" water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters. *Journal of Chemical Theory and Computation*, 10(4):1599–1607, 2014.

[4] Gregory R. Medders, Volodymyr Babin, and Francesco Paesani. Development of a first principles water potential with flexible monomers. III: Liquid phase properties. *Journal of Chemical Theory and Computation*, page 140701133607002, 2014.

[5] Jose LF Abascal and Carlos Vega. A general purpose model for the condensed phases of water: Tip4p/2005. *The Journal of chemical physics*, 123(23):234505, 2005.

[6] Eduardo Anglada, Javier Junquera, and José M Soler. Efficient mixed-force first-principles molecular dynamics. *Physical Review E*, 68(5):055701, 2003.

[7] Kristian Berland and Per Hyldgaard. Exchange functional that tests the robustness of the plasmon description of the van der waals density functional. *Physical Review B*, 89(3):035412, 2014.