

Density Functionals with Quantum Chemical Accuracy: From Machine Learning to Molecular Dynamics

Mihail Bogojeski, [Leslie Vogt-Maranto](#), Mark E. Tuckerman, Klaus-Robert Mueller, Kieron Burke

Submitted date: 03/05/2019 • Posted date: 06/05/2019

Licence: CC BY-NC-ND 4.0

Citation information: Bogojeski, Mihail; Vogt-Maranto, Leslie; Tuckerman, Mark E.; Mueller, Klaus-Robert; Burke, Kieron (2019): Density Functionals with Quantum Chemical Accuracy: From Machine Learning to Molecular Dynamics. ChemRxiv. Preprint.

Kohn-Sham density functional theory (DFT) is a standard tool in most branches of chemistry, but accuracies for many molecules are limited to 2-3 kcal/mol with presently-available functionals. Ab initio methods, such as coupled-cluster, routinely produce much higher accuracy, but computational costs limit their application to small molecules. We create density functionals from coupled-cluster energies, based only on DFT densities, via machine learning. These functionals attain quantum chemical accuracy (errors below 1 kcal/mol). Moreover, density-based Δ -learning (learning only the correction to a standard DFT calculation, Δ -DFT) significantly reduces the amount of training data required. We demonstrate these concepts for a single water molecule, and then illustrate how to include molecular symmetries with ethanol. Finally, we highlight the robustness of Δ -DFT by correcting DFT simulations of resorcinol on the fly to obtain molecular dynamics (MD) trajectories with coupled-cluster accuracy. Thus Δ -DFT opens the door to running gas-phase MD simulations with quantum chemical accuracy, even for strained geometries and conformer changes where standard DFT is quantitatively incorrect.

File list (2)

BogojeskiVogt_mldft.pdf (1.52 MiB)

[view on ChemRxiv](#) • [download file](#)

BogojeskiVogt_mldft_esi.pdf (3.61 MiB)

[view on ChemRxiv](#) • [download file](#)

Density functionals with quantum chemical accuracy: From machine learning to molecular dynamics

Mihail Bogojeski^{*1}, Leslie Vogt-Maranto^{*2}, Mark E. Tuckerman^{†2,3,4},
Klaus-Robert Müller^{†1,5,6}, and Kieron Burke^{†7,8}

¹Machine Learning Group, Technische Universität Berlin, Marchstr. 23, 10587 Berlin, Germany

²Department of Chemistry, New York University, New York, NY 10003, USA

³Courant Institute of Mathematical Science, New York University, New York, NY 10012, USA

⁴NYU-ECNU Center for Computational Chemistry at NYU Shanghai, 3663 Zhongshan Road North, Shanghai 200062, China

⁵Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 02841, Korea

⁶Max-Planck-Institut für Informatik, Stuhlsatzenhausweg, 66123 Saarbrücken, Germany

⁷Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA

⁸Department of Chemistry, University of California, Irvine, CA 92697, USA

Abstract

Kohn-Sham density functional theory (DFT) is a standard tool in most branches of chemistry, but accuracies for many molecules are limited to 2-3 kcal/mol with presently-available functionals. *Ab initio* methods, such as coupled-cluster, routinely produce much higher accuracy, but computational costs limit their application to small molecules. We create density functionals from coupled-cluster energies, based only on DFT densities, via machine learning. These functionals attain quantum chemical accuracy (errors below 1 kcal/mol). Moreover, density-based Δ -learning (learning only the correction to a standard DFT calculation, Δ -DFT) significantly reduces the amount of training data required. We demonstrate these concepts for a single water molecule, and then illustrate how to include molecular symmetries with ethanol. Finally, we highlight the robustness of Δ -DFT by correcting DFT simulations of resorcinol on the fly to obtain molecular dynamics (MD) trajectories with coupled-cluster accuracy. Thus Δ -DFT opens the door to running gas-phase MD simulations with quantum chemical accuracy, even for strained geometries and conformer changes where standard DFT is quantitatively incorrect.

*Contributed equally.

†To whom correspondence should be addressed.

M.E.T. (email: mark.tuckerman@nyu.edu),

K.-R.M. (email: klaus-robert.mueller@tu-berlin.de),

K.B. (email: kieron@uci.edu)

A long-standing question of overwhelming scientific importance is how to bridge the gap between *ab initio* quantum chemistry and molecular dynamics simulations of molecules over long time scales. Quantum chemistry provides highly accurate energetics for (almost) any configuration of the nuclei, but at considerable computational cost. At the other extreme, classical molecular dynamics (MD) employing relatively simple force fields is many orders of magnitude faster, but limited to configurations where the force field is accurate, such as close to equilibrium positions. A proper description of the dynamics of rare events, such as conformational changes or the making and breaking of chemical bonds, requires both computational efficiency and chemical accuracy. Over thirty years ago, Car and Parrinello [1] made a breakthrough by combining Kohn-Sham [2] density functional theory (KS-DFT) calculations with MD simulations, allowing an on-the-fly quantum treatment of electrons in an MD simulation. More recently, DFT-based Born-Oppenheimer MD simulations have found widespread use for systems in the gas phase [3–10], despite these calculations being much more expensive than classical force fields and less accurate than *ab initio* methods when using standard density functionals. These applications would clearly benefit from using more accurate electronic structure methods, but due to the high computational overhead, it has generally not been possible to perform such “on-the-fly” studies at a high level of *ab initio* theory such as coupled-cluster (CC).

At the heart of this problem is the severe incompatibility of the formalisms of DFT and *ab initio* methods. Both aim to deliver the ground-state energy of a molecule as a function of its nuclear coordinates. *Ab initio* methods directly solve the Schrödinger equation in a systematic, controllable fashion. But Kohn-Sham DFT buries all the quantum complexity into an unknown functional of the density, the exchange-correlation energy, which must be approximated [11, 12]. Myriad different forms for such KS-DFT approximations exist. Unfortunately, no practical route currently exists for converting an approximation in one formalism to an approximation in the other.

The present work provides one pragmatic approach to bridging the gap between these two disparate schools. By training on DFT densities and using CC energies as targets, a machine-learned density functional can produce CC energies at DFT (or lower) cost. We illustrate the importance of this framework by running gas-phase MD with this machine-learned functional, finding qualitatively different trajectories due to the higher accuracy of the novel functional with comparatively moderate computational effort.

This apparent miracle depends on insights from several different fields. Using age-old formalism in the DFT literature, one can construct functionals of densities other than the exact density, but that nonetheless yield the exact energy [13]. This bypasses the need for constructing coupled-cluster densities to train upon.

With the recent rise in popularity of machine-learning methods, there have been many advances in the prediction of properties of atomistic systems across chemical space [14–35] and the construction of accurate force fields [36–50] for ML-based MD simulations. Unlike these methods, our model uses the density directly, building on the successes of DFT over fifty years, and yielding densities as well as energies. The basic principles of constructing density functionals using kernel ridge regression (KRR) have been developed over several years [51–56]. One of the more recent and successful applications avoided the need to solve the KS equations by machine learning both the Hohenberg-Kohn map (ML-HK) from potentials to densities and the energy as a functional of the density [57]. In that case, MD simulations were performed using maps trained on densities and energies for PBE, a widely used density functional approximation [58]. Once trained, the ML maps find the solution of the KS equations for arbitrary positions of the nuclei (within the training manifold) without the need to solve them. The results merely reproduce the results of a PBE-MD simulation (albeit in a fraction of the time). While there have been other recent machine learning methods for the prediction of electron densities or density functionals [59–66], the ML-HK map uniquely facilitates the use of both machine-learned densities and density functionals for MD simulations.

Since the ML recipe is identical no matter what the source of inputs, one could readily imagine

training on a set of CC densities and energies. In practice, few quantum chemical packages yield the CC electronic density, as it is not something that is needed to find a CC energy. We show that the same basic result can be achieved more easily, by demonstrating a sequence of approximations that simplify the calculation, both in human terms and in computational cost. The first such step is to consider the CC energy as a functional of an approximate DFT density, such as PBE. This yields a paradox, as an accurate energy functional usually determines an accurate density via the Euler equation [67]. In fact, there is no requirement in approximate DFT that a self-consistent calculation must always be performed, as shown by the recent successes of density-corrected DFT [68–70]. Moreover, the density in this case is only serving as a (very useful) label for the different potentials and is an extremely useful representation for learning other properties, such as the energy itself. The ML algorithm automatically trains to find the CC energy as a functional of the approximate density. In the results section, we show that it is as easy to train to extract the CC energy from the DFT density as it is to train the DFT energy itself, i.e., the self-consistency of an approximate DFT calculation confers no special advantage in learning the map.

Furthermore, drawing from decades-old knowledge in machine learning [71], we found that one can learn the *difference* between a DFT and a CC energy much more efficiently than one can learn either separately. Using this we take a second step, in which we learn only the *difference* between the DFT and CC energies as a functional of the DFT density. As CC energies are far more accurate than standard DFT methods for the weakly correlated systems we study, we can refer to this as the error in the DFT energy. We call this Δ -DFT. In fact, we show that the error in the training curve for Δ -DFT drops far faster than those for learning either the DFT or the CC energies themselves, i.e., the error in DFT is much more amenable to interpolation than the DFT energy itself. This, combined with the use of molecular symmetry, conveniently reduces the amount of training needed to achieve quantum chemical accuracy, allowing us to extract CC energies from standard DFT calculations, with essentially no additional cost (beyond the initial generation of training data). We even find that we can construct a sufficiently accurate ML-HK map from potentials to DFT densities to be used as input into our Δ -DFT, so as to produce no significant increase in error. Thus, we have created a DFT scheme that costs no more than standard DFT calculations, but yields CC accuracy, for cases where many energy evaluations are required.

Our new theoretical tools are illustrated by application to the water molecule, as shown in Fig. 1a. We use the same PBE density as a functional of the potential, $n[\mathbf{v}](\mathbf{r})$, as in Ref. [57], but now with various ML maps of the energy as a functional of the density, $E[n]$. We refer to these combinations using generic notation when describing the models, with the specific electronic structure methods (e.g. using the PBE functional [58]) detailed in the Methods section. In our nomenclature, $E^{\text{DFT}}[n^{\text{DFT}}]$ denotes a standard DFT calculation, and a subscript ML indicates that the map is found via machine learning. Thus Ref. [57] showed that, with sufficient training, $E_{\text{ML}}^{\text{DFT}}[n_{\text{ML}}^{\text{DFT}}]$ could be made sufficiently accurate to run MD simulations, bypassing the need to solve the KS equations.

However, the E^{DFT} energies of conformers sampled during finite-temperature MD simulations do not reflect the potential energy surface determined by more accurate CC calculations, as shown for water in Fig. 1b. The DFT energy errors are not a simple function of the energy relative to the minimum energy geometry (see Supplementary Fig. 2), as short O-H bond lengths tend to be too high in energy and stretched bonds are overstabilized. While MD trajectories based on CC energies are prohibitively expensive without ML methods, the following sections demonstrate that self-consistent sampling of the more accurate CC potential energy surface can be achieved by using a machine-learned energy functional, even for molecules such as resorcinol, where the DFT energies near conformational energy barriers are quantitatively incorrect.

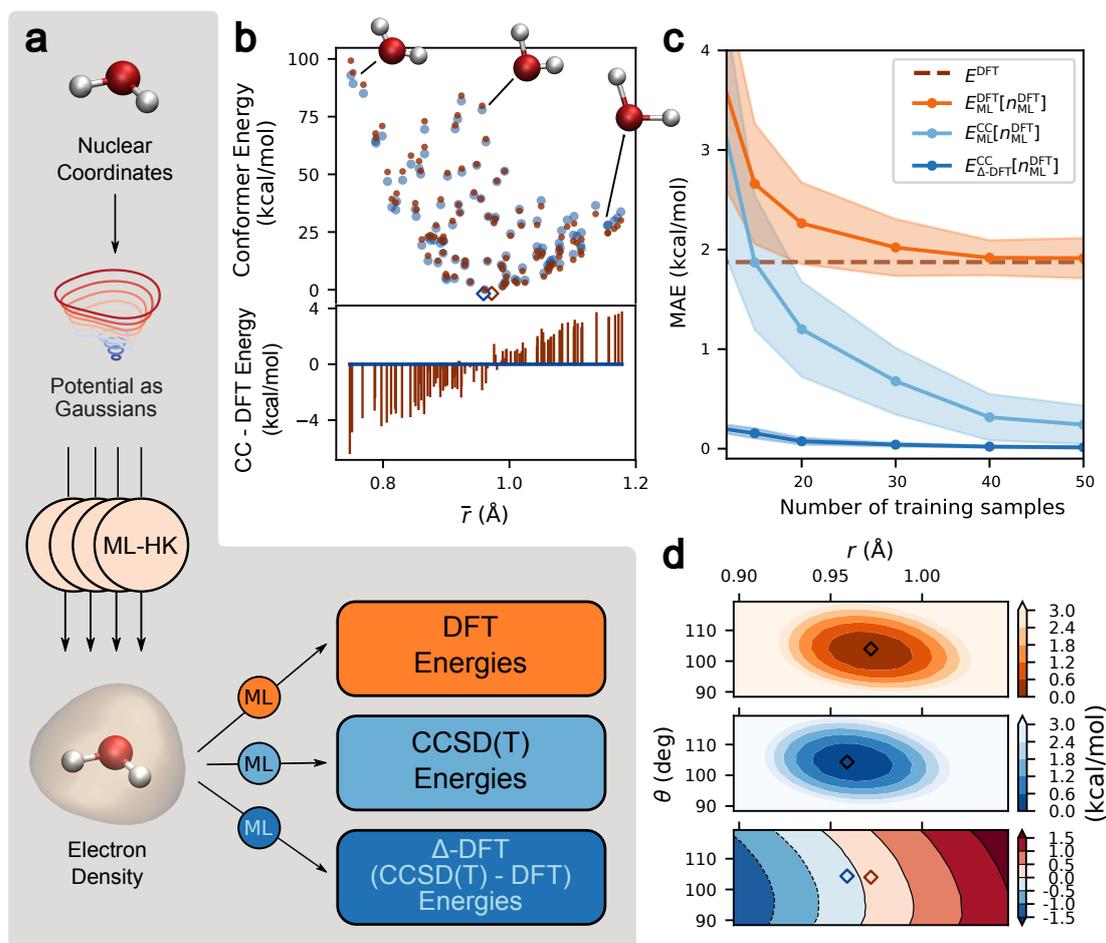


Figure 1: Illustration of density-based machine learning for water conformer energies. For all panels, DFT energies (orange) are shown alongside CC energies (blue) for the same molecular conformers, with optimized geometries indicated by open diamonds. a) The nuclear potential, represented by an approximate Gaussians potential, is the input to a set of ML models that return the electron density [57]. This learned density is the input for independent ML predictions of molecular energies based on DFT or CC electronic structure calculations, b) Calculated energies for CC (dark blue) and DFT (dark orange) for 102 sample geometries relative to the lowest training energy (top), along with the relative energy errors for DFT compared to CC for each conformer (bottom), c) Average out-of-sample prediction errors for the different ML functionals compared to the reference E^{CC} energies. The MAE of the E^{DFT} energies w.r.t. E^{CC} is also shown as a dashed line, d) The potential energy surface of symmetric water geometries for $E_{\text{ML}}^{\text{DFT}}$ (orange) and $E_{\Delta\text{-DFT}}^{\text{CC}}$ (blue) after applying the $\Delta\text{-DFT}$ correction (bottom). For this figure, DFT calculations use the PBE functional and CC calculations use CCSD(T) (see Methods for more details).

Results

Theory

Routine DFT calculations use some approximate XC functional and solve the Kohn-Sham equations self-consistently. But an alternative approach has long been considered (e.g., Ref. [13]), in which the

exact energy, E , is found by correcting an *approximate* self-consistent DFT calculation:

$$E = E^{\text{DFT}}[n^{\text{DFT}}] + \Delta E[n^{\text{DFT}}], \quad (1)$$

where DFT denotes the approximate DFT calculation, and ΔE , evaluated on the approximate density, is *defined* to yield the exact energy. This is *not* the exact functional of standard KS-DFT, but can be a more practical alternative. Thus one solves the KS equations within that approximation, but corrects the final energy by ΔE . Moreover, using the results from the Supplementary Information, one can in principle even construct the exact density from a sequence of such calculations.

CC accuracy from ML DFT

To demonstrate the methodology behind the map in Eq. (1), we begin with learning the CC energy directly for 102 random water geometries (Fig. 1b and Supplementary Fig. 1). The mean absolute error (MAE) of DFT energies relative to the CC energies (relative to the lowest energy conformer in the training set) is 1.86 kcal/mol, with maximum errors of more than 6 kcal/mol. Details of our approach are found in the Methods section. Briefly, after obtaining the predicted DFT density coefficients $\mathbf{u}^{\text{ML}}[\mathbf{v}]$ using the ML-HK map [57], we now use a second KRR model to predict energies from a higher level of theory, in this case CC energies:

$$E_{\text{ML}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}] = \sum_{i=1}^M \alpha_i k(\mathbf{u}_{\text{ML}}[\mathbf{v}], \mathbf{u}_{\text{ML}}[\mathbf{v}_i]). \quad (2)$$

Thus we create $E_{\text{ML}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}]$, the chemically accurate coupled-cluster energy, as a functional of the learned DFT density. (This corresponds to learning $E^{\text{DFT}} + \Delta E$ in Eq. (1).)

The performance of the $E_{\text{ML}}^{\text{DFT}}[n_{\text{ML}}^{\text{DFT}}]$ and $E_{\text{ML}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}]$ models was evaluated for training sets with 10, 15, 20, 30, 40 or 50 geometries, while the test set consisted of 52 geometries (Fig. 1c). Due to the small size of the dataset, we used cross-validation in order to obtain more stable estimates for the prediction accuracy of the models. Details of the evaluation procedure are found in the Methods section. As expected, the accuracy of each model improves with increasing training set size, but the benefit of predicting CC energies compared to DFT energies is immediately obvious. For this data set, the MAE of E^{DFT} relative to E^{CC} (used here as the ground truth) is reached by $E_{\text{ML}}^{\text{DFT}}[n_{\text{ML}}^{\text{DFT}}]$ with 40 training geometries. Quantum chemical accuracy of 1 kcal/mol is obtained using slightly fewer (30) samples for the energy functional $E_{\text{ML}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}]$, and an improved MAE of 0.24 kcal/mol with 50 training samples. Once constructed, the time to evaluate $E_{\text{ML}}[n]$ is the same regardless of which energy its trained on (for a fixed amount of training data). There is a clear benefit of training the model on the more accurate CC energies as long as a good performance can be achieved with a small number of samples from the more computationally expensive method.

Note that although all molecules used in this work are normal at equilibrium i.e., the energy change due to replacing a PBE density with a more accurate density should be negligible, they might not be when far from equilibrium. But this is irrelevant to the ML-CC energy map, as it learns the accurate energy even as a functional of an inaccurate density, as in Eq. (1). In an abnormal case, use of the CC density inside our ML-CC energy functional would make it (paradoxically) less accurate.

Δ -DFT: reducing the CC cost

Inspired by the delta learning approach [72], we also propose a machine learning framework that is able to leverage densities and energies from lower level theories (e.g. DFT) to predict CC level energies. Specifically this is achieved by correcting DFT energies by using delta learning, which we denote as Δ -DFT. Instead of predicting the CC energies directly using our machine learning model, we can

instead train a new map $\Delta E_{\text{ML}}^{\text{CC-DFT}}[n_{\text{ML}}^{\text{DFT}}]$ that yields the error in a DFT calculation (relative to CC) for each geometry (i.e., the second term in Eq. (1)). We define the corresponding total energy as

$$E_{\Delta\text{-DFT}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}] = E^{\text{DFT}}[n^{\text{DFT}}] + \Delta E_{\text{ML}}^{\text{CC-DFT}}[n_{\text{ML}}^{\text{DFT}}]. \quad (3)$$

Correcting the DFT energies in this way shows a dramatic improvement the model performance, as seen in Fig. 1c. Remarkably, with only 10 training samples, the MAE of this $E_{\Delta\text{-DFT}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}]$ model is already lower than the error of $E_{\text{ML}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}]$ trained with 50 samples; using 50 training samples reduces the MAE of the Δ -DFT model to only 0.013 kcal/mol. The Δ -DFT correction is easier to learn than the energies themselves, as illustrated in Fig. 1d for symmetric water geometries that were not included in the previous data set. Although the optimized geometry differs slightly between DFT and CC, the Δ -DFT approach provides a smooth map between the two types of electronic structure calculations as a functional of the density. The comparison between the Δ -DFT and total energy ML models is further explored with larger molecules in the following sections.

Δ -DFT with molecular symmetries

The next molecule chosen to evaluate our ML model is ethanol, with geometries and energies from the MD17 dataset [41, 42]. This molecule has two types of geometric minima, with the alcohol OH in either an *anti* or doubly-degenerate *gauche* position, as well as a freely rotating CH₃ group that introduces additional variability into the possible geometries. Supplementary Fig. 4 shows the atomic distributions of the ethanol dataset after alignment based on heavy atom positions. Adding nuclear degrees of freedom makes the learning problem more difficult, so we expect that a greater number of training samples is needed to achieve chemical accuracy for the range of thermally-accessible geometries. The dataset contains 1000 training and 1000 test samples with both DFT and CC energies. The ML-HK map incorporates equivalence for each element, but we can also exploit the mirror symmetry of the molecule by reflecting H atoms through the plane defined by the three heavy atoms, effectively doubling the size of the training set, as outlined in the Methods section below. To differentiate the models trained on datasets augmented by these symmetries we add an *s* in front of the machine learning model (e.g. *sML*).

Table 1 shows the prediction accuracies of the various *sML* models for ethanol compared to some other state-of-the-art ML methods for the same dataset. The prediction error for DFT and CC energies is roughly equal to that of ML models trained only on energies, although ML models incorporating additional force data are superior. However, using the $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{s\text{ML}}^{\text{DFT}}]$ functional to correct low-cost DFT energies achieves a MAE for CC energies comparable to those of most accurate force-based models. We also note that using only the original 1000 training geometries roughly doubles the MAE of all ML models, e.g., the $E_{\Delta\text{-DFT}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}]$ functional has a MAE of only 0.15 kcal/mol (see Supplementary Table 1).

Optimization using ML functionals

Neither the training nor test configurations from the MD17 dataset [41, 42] include the minimum energy conformers of ethanol. Using the ML models, we predicted the energy of the *anti* and *gauche* conformers optimized using MP2/6-31G* or the electronic structure methods used to generate the model energies. Although all training geometries have energies more than 4.5 kcal/mol higher than the global minimum, the ML models are able to predict the energies of the minima with errors below chemical accuracy (see Table 2).

In addition, the machine-learned potential energy surfaces are sufficiently smooth to optimize ethanol using the ML models themselves. Calculations for each conformer start from geometries optimized using MP2/6-31G*, which are slightly different from both DFT- and CC-optimized geometries.

Figure 2b shows that despite the sparsity of training data near the minimum energy configurations, the ML models trained with different energies can differentiate between the DFT and CC minima with remarkable fidelity.

Molecular dynamics using CC energies

The final molecular example of 1,3-benzenediol (resorcinol) illustrates the utility of learning both the $E_{sML}^{CC}[n_{sML}^{DFT}]$ and $E_{s\Delta-DFT}^{CC}[n_{sML}^{DFT}]$ functionals for the same system. Combining the $E_{sML}^{CC}[n_{sML}^{DFT}]$ with the more expensive and accurate $E_{s\Delta-DFT}^{CC}[n_{sML}^{DFT}]$ method, we demonstrate how to run self-consistent MD simulations that can be used to explore the configurational phase space based on CC energies.

Resorcinol model performance

Resorcinol has two rotatable OH groups, two molecular symmetry operations, and many more degrees of freedom than water or ethanol, making this a more stringent test of the ML functionals. The initial data sets are generated from 1 ns classical MD simulations at 500 K and 300 K for the training and test sets, respectively (details are found in the Methods section). This training set is augmented with the two symmetries, resulting in an effective training set size of 4000 geometries. The molecular geometries in the MD-generated training set have energies between 7 and 50 kcal/mol above the equilibrium conformer, so the four local minima are also included using geometries from MP2/6-31G* optimizations. These conformers, which differ in the orientation of the two alcohol groups, are separated by a rotational barrier of ≈ 4 kcal/mol (See Supplementary Fig. 9). The maximum relative energy error of the DFT energies compared to those from CC is 6.1 kcal/mol and 6.7 kcal/mol for geometries included in the training and test sets, respectively.

As for water and ethanol, the ML model performance improves with expanding training set size (see Supplementary Fig. 10). Table 3 shows the resulting MAE for the resorcinol dataset when training on all 1004 training geometries (4004 training points). While the MAE of predicted energies is around 1.3 kcal/mol for both $E_{sML}^{DFT}[n_{sML}^{DFT}]$ and $E_{sML}^{CC}[n_{sML}^{DFT}]$, the error when using $E_{s\Delta-DFT}^{CC}[n_{sML}^{DFT}]$ is only 0.11 kcal/mol. There is some error attributable to using learned densities; The *sML* models have lower prediction MAEs when using the true DFT densities, although there is essentially no difference in model performance for the Δ -DFT approach.

CC-based MD trajectories

Although DFT energies may be sufficient for some molecules, the ability to use CC energies to determine the equilibrium geometries and thermal fluctuations is a promising advance. For resorcinol, the DFT energies differ significantly from the CC energies, particularly near the OH rotational barrier that separates conformers. Conformational changes are also a rare event in the MD trajectories, making it crucial to describe the transitions accurately. For example, the exploration of the OH dihedral angles during 10 ps MD trajectories from DFT-based NVT simulations at 350 K are shown in Supplementary Fig. 11. In this simulation, only one conformational change is observed, despite several excursions away from the local minima.

Using the $E_{s\Delta-DFT}^{CC}$ approach, we could easily correct energies after running a conventional DFT-MD simulation. However, as shown in Supplementary Fig. 12, for snapshots along a 1.5 ps NVE simulation starting from a point near a conformer change, the MAE of DFT energies compared to CC energies for each snapshot is 1.0 kcal/mol, with a maximum of just under 4.5 kcal/mol. Therefore, a more promising use of the ML functionals is to run MD simulations using CC energies directly. An example $E_{s\Delta-DFT}^{CC}[n_{sML}^{DFT}]$ trajectory starting from a random training point is shown in Supplementary Fig. 13, with an MAE of 0.2 kcal/mol.

Starting from a different point in the DFT-generated trajectory illustrates the importance of generating MD trajectories consistent with coupled-cluster energies. As seen in Fig. 3 for NVE simulations from the same initial condition, a DFT-based trajectory does not have sufficient kinetic energy to traverse the rotational barrier, while the conformer switch does occur for trajectories based on coupled-cluster energies. Astonishingly, the $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ trajectory has a MAE of only 0.18 kcal/mol relative to the true CC energies over a range of more than 15 kcal/mol.

As the Δ -DFT method requires a running DFT calculations at each step of the trajectory, we can overcome this computational cost by combining the ML models. The middle panel of Fig. 3b shows the CC trajectory using a reversible multi-timestep integrator [73] to evaluate energies and forces primarily with the $E_{\text{sML}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ model, with periodic force corrections based on the more accurate $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ every 3 steps (see Supplementary Information for more details). The resulting trajectory has a MAE of 3.8 kcal/mol relative to the true CC energies, with the largest errors in regions that are sparsely represented in the training set. This self-consistent exploration of the configurational space with the combined ML models provides an opportunity to improve the sampling in a cost-effective manner.

Discussion

DFT is used in at least 30,000 scientific papers each year [74], and because of its low cost relative to *ab initio* methods, energies of large molecules can be computed. For most such calculations, CCSD(T) would be more accurate and reliable, but is unaffordable. In this work, we have established one possible bridge between the two methods.

There are two distinct modes in which our results can be applied. With Δ -DFT, the cost of a gas-phase MD simulation is essentially that of the DFT-based MD with a given approximate functional, plus the cost of evaluating a few dozen CCSD(T) energies. No forces are needed for training, making training set generation cheaper than other methods with similar performance. Compared to other machine learning models, Δ -DFT is well-behaved and stable far outside of the training set, since the zero-mean prior allows it to fall back on DFT results when far from the training set. Combining Δ -DFT with the ML models for DFT energies of Ref. [57], yields both the speed from bypassing the KS equations and the accuracy of CCSD(T). While this yields perfectly accurate potential energy surfaces within the training manifold, it occasionally yields inaccurate forces in an MD simulation, which can be corrected with the Δ -DFT forces using the appropriate integrators, as shown above.

Clearly, our methodology can be applied to any gas-phase MD simulation for which CCSD(T) is affordable for several dozen carefully chosen configurations. For larger molecules, recent schemes that embed an *ab initio* heart within a larger DFT calculation could be treatable by this method, especially if Δ -DFT need only be applied to the *ab initio* portion of the calculation. Even gas-phase MD has many applications. Earlier studies focused on comparing equilibrium properties from simulations excluding or including (via the Feynman path integral) nuclear quantum effects [3–6]. More recent studies have focused on accurate spectroscopy and exploration of reactivity in small complexes and clusters [7–10].

Another potential application might be in the optimization of molecular geometries at the CCSD(T) level. Very often, when DFT energetics are tested against CCSD(T) energies, DFT geometries are used, due to the prohibitive cost of finding an optimum geometry at the CCSD(T) level. For a large soft molecules, finding the geometry can require hundreds of evaluations of energies and forces. But we have shown how relatively few energies are needed in Δ -DFT to produce an accurate PES. Perhaps Δ -DFT can speed up such searches, producing CC geometries for molecules that were previously prohibitive.

Machine-learning represents an entirely new approach to density functional construction, produc-

ing density functionals that automatically avoid the biases built in to human-designed functionals. As shown here, ML provides a natural framework for incorporating results from more accurate electronic structure methods, thus bridging the gap between the CC and the DFT worlds while maintaining the versatility of DFT to describe electronic properties beyond energy and forces such as the dipole moment, molecular polarizability, NMR chemical shifts, etc. Along with these insights, the long and successful history of KS-DFT suggests that using the density as a descriptor may thus prove to be an excellent strategy for improved simulations in the future.

Methods

Machine learning model

In order to predict the total energy given only the atomic positions of a molecule, we can use the ML-HK map introduced in Ref. [57], with the entire procedure being illustrated in Fig. 1a. Initially, we characterize the Hamiltonian by the external nuclear potential, which we approximate using a Gaussians potential v [75] as

$$v(\mathbf{r}) = \sum_{\alpha=1}^{N^a} Z_{\alpha} \exp\left(\frac{-\|\mathbf{r} - \mathbf{R}_{\alpha}\|^2}{2\gamma^2}\right), \quad (4)$$

where \mathbf{r} are the current coordinates of the spatial grid, \mathbf{R}_{α} is a matrix containing the atom coordinates and Z_{α} are the nuclear charges of the atoms. This Gaussians potential then is evaluated on a 3D grid around the molecule and used as a descriptor for the ML model. We will use \mathbf{v} to refer to the vector representation of the potential evaluated on a 3D grid. For each molecule, cross-validation is used to determine the width parameter, γ , and the grid spacing used to discretize the resulting potential.

After obtaining the Gaussians potential, we use a kernel ridge regression (KRR) model to learn the approximate DFT valence electron density. In order to simplify the learning problem and avoid representing the density with a 3D grid, we use a Fourier basis representation of the densities, and consequently learn the basis coefficients instead of the density grid points:

$$n_{\text{ML}}[\mathbf{v}](\mathbf{r}) = \sum_{l=1}^L u^{(l)}[\mathbf{v}]\phi_l(\mathbf{r}). \quad (5)$$

In our case, using 12500 basis functions (25 per dimension) proved sufficient for good performance. Using KRR to learn these basis coefficients makes the problem more tractable for 3D densities and more importantly, the orthogonality of the basis functions allows us to learn the individual coefficients independently:

$$u_{\text{ML}}^{(l)}[\mathbf{v}] = \sum_{i=1}^M \beta_i^{(l)} k(\mathbf{v}, \mathbf{v}_i), \quad (6)$$

where $\beta^{(l)}$ are the KRR coefficients and k is the kernel function being used.

The independent and direct predicting the basis coefficients makes the ML-HK map more efficient and easier to scale to larger molecules, since the complexity only depends on the number of basis functions. Additionally, we can use the predicted basis coefficients to reconstruct the continuous density at any point in space, making the predicted density independent of a fixed grid and enabling computations such as numerical integrals to be performed at an arbitrary accuracy.

As a final step, another KRR model is used to learn the total energy from the density basis coefficients:

$$E_{\text{ML}}[n_{\text{ML}}] = \sum_{i=1}^M \alpha_i k(\mathbf{u}_{\text{ML}}[\mathbf{v}], \mathbf{u}_{\text{ML}}[\mathbf{v}_i]), \quad (7)$$

where k is the Gaussian kernel.

Exploiting point group symmetries

Our machine learning model can be easily enriched using the point group symmetries. To extract the point group symmetries and the corresponding transformation matrices we used the SYVA software package [76]. Consequently, we can multiply the size of the training set by the number of point group symmetries without needing any additional quantum chemical calculations simply by applying the point group transformations on our existing data. We denote models trained on datasets augmented by such symmetries by adding an s in front of the ML notation (e.g. sML).

Cross-validation and hyperparameter optimization

Due to the small number of training and test samples, when evaluating the models on the water dataset, the data was shuffled 40 times, and for each shuffle a subset of 50 geometries was selected as the training set, with the remaining 52 being used as the out-of-sample test set. For the smaller training sets, a subset of the 50 training geometries was selected using k -means sampling.

The hyperparameters for all models were tuned using 5-fold cross-validation on the training set. For the ML-HK map from potentials to densities, the following 3 hyperparameters were optimized individually for each dataset: the width parameter of the Gaussians potential γ , the spacing of the Gaussians potential grid, and the width parameter σ of the Gaussian kernel. For each subsequent density to energy map $E_{\text{ML}}^*[n]$, only the width parameter of the Gaussian kernel needs to be chosen using cross-validation. Specific values are reported in the Supplementary Information.

Classical molecular dynamics

Training and test set geometries for resorcinol (1,3-benzenediol) were selected from classical MD simulations. The four local minima were optimized using MP2/6-31g* in Gaussian09 [77]. Symmetric atomic charge assignments are from a RESP fit [78] to the HF/6-31g* calculations for the three distinct geometries, with Boltzmann weights determined by the relative MP2 energies. All other standard GAFF parameters [79] for the MD simulations were assigned using the AmberTools package [80]. To generate resorcinol conformers, classical isothermal MD simulations were run at 300 K and 500 K using the PINY_MD package [81] with massive Nosé-Hoover chain (NHC) thermostats [82] for atomic degrees of freedom (length = 4, τ = 20 fs, Suzuki-Yoshida order = 7, multiple time step = 4) and a time step of 1 fs.

For the resorcinol training set, we selected 1000 conformers closest to k -means centers from the 1 ns classical MD trajectory run at 500 K. The test set comprises 1000 randomly selected snapshots from the 1 ns 300 K classical MD simulation. Data sets are aligned by minimizing the root mean square deviation (RMSD) of carbon atoms to the global minimum energy conformer.

DFT molecular dynamics

Resorcinol Born-Oppenheimer MD simulations were run using DFT in the QUICKSTEP package [83] of CP2K v. 2.6.2 [84]. The PBE exchange-correlation functional [58] was used in the Gaussian

and plane wave (GPW) scheme [85] with DZVP-MOLOPT-GTH (m-DZVP) basis sets [86] paired with appropriate dual-space GTH pseudopotentials [87, 88]. Wave functions were converged to 1E-7 Hartree using the orbital transformation method [89] on a multiple grid ($n = 5$) with a cutoff of 900 Ry for the system in a cubic box ($L = 20$ bohr). For the NVT simulation, a temperature of 350 K was maintained using massive NHC thermostats [82] (length = 4, $\tau = 10$ fs, Suzuki-Yoshida order = 7, multiple time step = 4) and a time step of 0.5 fs.

ML molecular dynamics

We used the Atomistic Simulation Environment (ASE) [90] with a 0.5 fs timestep to run MD with ML energies. For the NVT simulation, a temperature of 350 K maintained via a Langevin thermostat with a friction value of 0.01 atomic units (0.413 fs^{-1}). Atomic forces were calculated using the finite difference method with $\epsilon = 0.001 \text{ \AA}$.

Electronic structure calculations

Optimizations for ethanol conformers were run using MP2/6-31g* in Gaussian09 [77]. DFT calculations for the ML models were run using Quantum ESPRESSO code [91] with the PBE exchange-correlation functional [58] and projector augmented waves (PAWs) [92, 93] with Troullier-Martin pseudopotentials describing the ionic cores [94]. Molecules were simulated in a cubic box ($L = 20$ bohr) with a wave function cutoff of 90 Ry. All coupled-cluster calculations were run using Orca [95] with CCSD(T)/aug-cc-pVTZ [96] for water or CCSD(T)/cc-pVDZ [96] for resorcinol.

Data availability

Data sets will be made available upon publication.

Acknowledgments

The authors thank Dr. Felix Brockherde and Joseph Cendagorta for helpful discussions, Dr. Li Li for the initial water data set, and Dr. Huziel Saucedo and Stefan Chmiela for the optimized geometries of ethanol. Calculations were run on NYU IT High Performance Computing resources and at TUB. Work at NYU was supported by the U.S. Army Research Office under contract/grant number W911NF-13-1-0387 (L.V.-M. and M.E.T.). This work was supported by the German Ministry for Education and Research as Berlin Big Data Centre (01IS14013A) and Berlin Center for Machine Learning (01IS18037I). Partial funding by DFG is acknowledged (EXC 2046/1, project-ID: 390685689). This work was also supported by the Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2017-0-01779). K.B. was supported by NSF grant CHE 1856165. This publication only reflects the authors views. Funding agencies are not liable for any use that may be made of the information contained herein.

Author contributions

L.V.-M. initiated the project and M.B. and L.V.-M. ran all simulations. M.E.T, K.-R.M., and K.B. conceived the theory and co-supervised the project. All authors guided the project design, contributed to data analysis, and co-wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

- [1] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55(22):2471–2474, Nov 1985.
- [2] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, Nov 1965.
- [3] M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello. On the quantum nature of the shared proton in hydrogen bonds. *Science*, 275:817, 1997.
- [4] S. Miura, M. E. Tuckerman, and M. L. Klein. An ab initio path integral molecular dynamics study of double proton transfer in the formic acid dimer. *Journal of Chemical Physics*, 109:5920, 1998.
- [5] M. E. Tuckerman and D. Marx. Heavy-atom skeleton quantization and proton tunneling in "intermediate-barrier" hydrogen bonds. *Physical Review Letters*, 86:4946, 2001.
- [6] X. Z. Li, B. Walker, and A. Michaelides. Quantum nature of the hydrogen bond. *Proceedings of the National Academy of Sciences, USA*, 108:6369, 2011.
- [7] A. Kaczmarek, M. Shiga, and D. Marx. Quantum effects on vibrational and electronic spectra of hydrazine studied by "on-the-fly". *Journal of Physical Chemistry A*, 113:1985, 2009.
- [8] H. Wang and N. Agmon. Complete assignment of the infrared spectrum of the gas-phase protonated ammonia dimer. *Journal of Physical Chemistry A*, 120:3117, 2016.
- [9] N. R. Samala and N. Agmon. Structure, spectroscopy, and dynamics of the phenol-(water)(2) cluster at low and high temperatures. *Journal of Chemical Physics*, 147:234307, 2017.
- [10] T. Jarvinen, J. Lundell, and P. Dopieralski. Ab initio molecular dynamics study of overtone excitations in formic acid and its water complex. *Theoretical Chemistry Accounts*, 137:100, 2018.
- [11] R. G. Parr and W. Yang. *Density Functional Theory of Atoms and Molecules*. Oxford University Press, 1989.
- [12] Carlos Fiolhais, F. Nogueira, and M. Marques. *A Primer in Density Functional Theory*. Springer-Verlag, New York, 2003.
- [13] M. Levy and A. Görling. Correlation energy density-functional formulas from correlating first-order density matrices. *Phys. Rev. A*, 52:R1808, 1995.
- [14] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, 108(5):058301, 2012.
- [15] Grégoire Montavon, Katja Hansen, Siamac Fazli, Matthias Rupp, Franziska Biegler, Andreas Ziehe, Alexandre Tkatchenko, O. Anatole von Lilienfeld, and Klaus-Robert Müller. Learning invariant representations of molecules for atomization energy prediction. In *Adv. Neural. Inf. Process. Syst.* 25, pages 440–448, 2012.

- [16] Grégoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.*, 15(9):095003, 2013.
- [17] V. Botu and R. Ramprasad. Learning scheme to predict atomic forces and accelerate materials simulations. *Phys. Rev. B*, 92:094306, Sep 2015.
- [18] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O. Anatole von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.*, 6(12):2326–2331, 2015.
- [19] Albert P. Bartók and Gábor Csányi. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.*, 115(16):1051–1057, 2015.
- [20] Matthias Rupp, Raghunathan Ramakrishnan, and O. Anatole von Lilienfeld. Machine learning for quantum mechanical properties of atoms in molecules. *J. Phys. Chem. Lett.*, 6(16):3309–3313, 2015.
- [21] Tristan Bereau, Denis Andrienko, and O. Anatole von Lilienfeld. Transferable atomic multipole machine learning models for small organic molecules. *J. Chem. Theory Comput.*, 11(7):3225–3233, 2015.
- [22] Sandip De, Albert P Bartók, Gábor Csányi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics*, 18(20):13754–13769, 2016.
- [23] Evgeny V. Podryabinkin and Alexander V. Shapeev. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.*, 140:171 – 180, 2017.
- [24] Kristof T Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus-Robert Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature Communications*, 8:13890, 2017.
- [25] Kristof T. Schütt, Huziel E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. SchNet—a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.
- [26] Felix A. Faber, Luke Hutchison, Bing Huang, Justin Gilmer, Samuel S Schoenholz, George E Dahl, Oriol Vinyals, Steven Kearnes, Patrick F. Riley, and O. Anatole von Lilienfeld. Prediction errors of molecular machine learning models lower than hybrid DFT error. *J. Chem. Theory Comput.*, 13(11):5255–5264, 2017.
- [27] Kun Yao, John E. Herr, and John Parkhill. The many-body expansion combined with neural networks. *J. Chem. Phys.*, 146(1):014106, 2017.
- [28] Michael Eickenberg, Georgios Exarchakis, Matthew Hirn, Stéphane Mallat, and Louis Thiry. Solid harmonic wavelet scattering for predictions of molecule properties. *J. Chem. Phys.*, 148(24):241732, 2018.
- [29] Kevin Ryczko, Kyle Mills, Iryna Luchak, Christa Homenick, and Isaac Tamblyn. Convolutional neural networks for atomistic systems. *Comput. Mater. Sci.*, 149:134–142, 2018.

- [30] Justin S. Smith, Ben Nebgen, Nicholas Lubbers, Olexandr Isayev, and Adrian E. Roitberg. Less is more: Sampling chemical space with active learning. *The Journal of chemical physics*, 148(24):241733, 2018.
- [31] Andrea Grisafi, David M. Wilkins, Gábor Csányi, and Michele Ceriotti. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.*, 120:036002, Jan 2018.
- [32] Wiktor Pronobis, Alexandre Tkatchenko, and Klaus-Robert Müller. Many-body descriptors for predicting molecular properties with machine learning: Analysis of pairwise and three-body interactions in molecules. *J. Chem. Theory Comput.*, 14(6):2991–3003, 2018.
- [33] Felix A. Faber, Anders S. Christensen, Bing Huang, and O. Anatole von Lilienfeld. Alchemical and structural distribution based representation for universal quantum machine learning. *The Journal of Chemical Physics*, 148(24):241717, 2018.
- [34] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018.
- [35] Truong Son Hy, Shubhendu Trivedi, Horace Pan, Brandon M. Anderson, and Risi Kondor. Predicting molecular properties with covariant compositional networks. *J. Chem. Phys.*, 148(24):241745, 2018.
- [36] Zhenwei Li, James R. Kermode, and Alessandro De Vita. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.*, 114:096405, 2015.
- [37] Michael Gastegger, Jorg Behler, and Philipp Marquetand. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.*, 8:6924–6935, 2017.
- [38] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Saucedo 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.
- [39] S.T. John and Gábor Csányi. Many-body coarse-grained interactions using gaussian approximation potentials. *J. Phys. Chem. B*, 121(48):10934–10949, 2017.
- [40] Tran Doan Huan, Rohit Batra, James Chapman, Sridevi Krishnan, Lihua Chen, and Rampi Ramprasad. A universal strategy for the creation of machine learning-based atomistic force fields. *NPJ Comput. Mater.*, 3(1):37, 2017.
- [41] Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Saucedo, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017.
- [42] Stefan Chmiela, Huziel E. Saucedo, Klaus-Robert Müller, and Alexandre Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, 9(1):3887, 2018.
- [43] Stefan Chmiela, Huziel E. Saucedo, Igor Poltavsky, Klaus-Robert Müller, and Alexandre Tkatchenko. sgdml: Constructing accurate and data efficient molecular force fields using machine learning, 2018.

- [44] Kenta Kanamori, Kazuaki Toyoura, Junya Honda, Kazuki Hattori, Atsuto Seko, Masayuki Karasuyama, Kazuki Shitara, Motoki Shiga, Akihide Kuwabara, and Ichiro Takeuchi. Exploring a potential energy surface by machine learning for characterizing atomic transport. *Phys. Rev. B*, 97(12):125124, 2018.
- [45] Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and Weinan E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.*, 120(14):143001, 2018.
- [46] Aldo Glielmo, Claudio Zeni, and Alessandro De Vita. Efficient nonparametric n-body force fields from machine learning. *Physical Review B*, 97(18):184307, 2018.
- [47] Anders S. Christensen, Felix A. Faber, and O. Anatole von Lilienfeld. Operators in quantum machine learning: Response properties in chemical space. *J. Phys. Chem.*, 150(6):064105, 2019.
- [48] Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noé. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.*, 9(1):5, 2018.
- [49] Frank Noé and Hao Wu. Boltzmann generators - sampling equilibrium states of many-body systems with deep learning. *arXiv preprint arXiv:1812.01729*, 2018.
- [50] Huziel E. Saucedo, Stefan Chmiela, Igor Poltavsky, Klaus-Robert Müller, and Alexandre Tkatchenko. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.*, 150(11):114102, 2019.
- [51] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. Finding density functionals with machine learning. *Phys. Rev. Lett.*, 108:253002, Jun 2012.
- [52] John C. Snyder, Matthias Rupp, Katja Hansen, Leo Blooston, Klaus-Robert Müller, and Kieron Burke. Orbital-free bond breaking via machine learning. *J. Chem. Phys.*, 139:224104, 2013.
- [53] John C. Snyder, Matthias Rupp, Klaus-Robert Müller, and Kieron Burke. Nonlinear gradient denoising: Finding accurate extrema from inaccurate functional derivatives. *International Journal of Quantum Chemistry*, 115(16):1102–1114, 2015.
- [54] Li Li, John C. Snyder, Isabelle M. Pelaschier, Jessica Huang, Uma-Naresh Niranjan, Paul Duncan, Matthias Rupp, Klaus-Robert Müller, and Kieron Burke. Understanding machine-learned density functionals. *International Journal of Quantum Chemistry*, 116(11):819–833, 2016.
- [55] Li Li, Thomas E. Baker, Steven R. White, and Kieron Burke. Pure density functional for strong correlation and the thermodynamic limit from machine learning. *Phys. Rev. B*, 94:245129, Dec 2016.
- [56] Jacob Hollingsworth, Li Li, Thomas E. Baker, and Kieron Burke. Can exact conditions improve machine-learned density functionals? *The Journal of Chemical Physics*, 148(24):241743, June 2018.
- [57] 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.
- [58] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, Oct 1996.

- [59] Matthew Welborn, Lixue Cheng, and Thomas F. Miller. Transferability in machine learning for electronic structure via the molecular orbital basis. *Journal of Chemical Theory and Computation*, 14(9):4772–4779, 2018.
- [60] Junji Seino, Ryo Kageyama, Mikito Fujinami, Yasuhiro Iwabata, 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.
- [61] Kevin Ryczko, David Strubbe, and Isaac Tamblyn. Deep learning and density functional theory. *arXiv preprint arXiv:1811.08928*, 2018.
- [62] Anton V. Sinitskiy and Vijay S. Pande. Deep neural network computes electron densities and energies of a large set of organic molecules faster than density functional theory (DFT). *arXiv preprint arXiv:1809.02723*, 2018.
- [63] Andrea Grisafi, David M. Wilkins, Benjamin A.R. Meyer, Alberto Fabrizio, Clemence Corminboeuf, and Michele Ceriotti. A transferable machine-learning model of the electron density. *arXiv preprint arXiv:1809.05349*, 2018.
- [64] Anand Chandrasekaran, Deepak Kamal, Rohit Batra, Chiho Kim, Lihua Chen, and Rampi Ramprasad. Solving the electronic structure problem with machine learning. *npj Computational Materials*, 5(1):22, 2019.
- [65] Lixue Cheng, Matthew Welborn, Anders S. Christensen, and Thomas F Miller III. A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules. *The Journal of Chemical Physics*, 150(13):131103, 2019.
- [66] Sebastian Dick and Marivi Fernandez-Serra. Learning from the density to correct total energy and forces in first principle simulations. *arXiv preprint arXiv:1812.06572*, 2018.
- [67] Kieron Burke and Lucas O. Wagner. DFT in a nutshell. *Int. J. Quant. Chem.*, 113:96–101, 01/2013 2013.
- [68] Min-Cheol Kim, Eunji Sim, and Kieron Burke. Understanding and reducing errors in density functional calculations. *Phys. Rev. Lett.*, 111:073003, Aug 2013.
- [69] Adam Wasserman, Jonathan Nafziger, Kaili Jiang, Min-Cheol Kim, Eunji Sim, and Kieron Burke. The importance of being inconsistent. *Annual Review of Physical Chemistry*, 68(1):555–581, 2017.
- [70] Eunji Sim, Suhwan Song, and Kieron Burke. Quantifying density errors in DFT. *J. Phys. Chem. Lett.*, 9:6385–6392, October 2018.
- [71] Iris Ginzburg and David Horn. Combined neural networks for time series analysis. In *Advances in Neural Information Processing Systems*, pages 224–231, 1994.
- [72] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Big data meets quantum chemistry approximations: the δ -machine learning approach. *Journal of Chemical Theory and Computation*, 11(5):2087–2096, 2015.
- [73] Mark E. Tuckerman, Bruce J. Berne, and Glenn J. Martyna. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3):1990–2001, 1992.
- [74] Aurora Pribram-Jones, David A. Gross, and Kieron Burke. Dft: A theory full of holes? *Annual Review of Physical Chemistry*, 66(1):283–304, 2015.

- [75] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical Review Letters*, 104(13):136403, 2010.
- [76] László Gyevi-Nagy and Gyula Tasi. Syva: A program to analyze symmetry of molecules based on vector algebra. *Computer Physics Communications*, 215:156–164, 2017.
- [77] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, et al. Gaussian 09, 2009.
- [78] Christopher I. Bayly, Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.
- [79] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.
- [80] Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Antechamber: an accessory software package for molecular mechanical calculations. *Journal of the American Chemical Society*, 222:U403, 2001.
- [81] Mark E. Tuckerman, D.A. Yarne, Shane O. Samuelson, Adam L. Hughes, and Glenn J. Martyna. Exploiting multiple levels of parallelism in molecular dynamics based calculations via modern techniques and software paradigms on distributed memory computers. *Computer Physics Communications*, 128(1?2):333 – 376, 2000.
- [82] 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.
- [83] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Computer Physics Communications*, 167(2):103–128, April 2005.
- [84] Jürg Hutter, Marcella Iannuzzi, Florian Schiffmann, and Joost VandeVondele. CP2K: Atomistic simulations of condensed matter systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(1):15–25, June 2014.
- [85] Gerald Lippert, Jürg Hutter, and Michele Parrinello. A hybrid Gaussian and plane wave density functional scheme. *Molecular Physics*, 92(3):477–488, December 2010.
- [86] Joost VandeVondele and Jürg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics*, 127(11):114105, 2007.
- [87] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Phys. Rev. B*, 54:1703–1710, Jul 1996.
- [88] M Krack. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theoretica Chimica Acta*, 114(1-3):145–152, May 2005.
- [89] Joost VandeVondele and Jürg Hutter. An efficient orbital transformation method for electronic structure calculations. *The Journal of Chemical Physics*, 118(10):4365, 2003.
- [90] S. R. Bahn and K. W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.*, 4(3):56–66, MAY-JUN 2002.

- [91] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L. Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclausero, Ari P. Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M. Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(39):395502, September 2009.
- [92] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59(3):1758–1775, January 1999.
- [93] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50(24):17953–17979, December 1994.
- [94] N. Troullier and José Luriaas Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43:1993–2006, Jan 1991.
- [95] Frank Neese. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):73–78, 2012.
- [96] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989.

Molecule/Model	$E_{\text{SchNet}}^{\text{DFT}}$ [25]	$E_{\text{SchNet}}^{\text{DFT}}$ (with forces) [25]	$E_{\text{sGDML}}^{\text{CC}}$ [42]	$E_{\text{sML}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$	$E_{\text{s}\Delta\text{-DFT}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$
Ethanol	0.93	0.08	0.05	1.10	0.09

Table 1: Mean average errors (kcal/mol) of the *sML* maps for ethanol trained on 1000 samples, compared to other published state-of-the-art ML models. Previously published results with DFT energies are shown where CC energies were not available.

	MP2 minima		DFT minima		CC minima	
	<i>anti</i>	<i>gauche</i>	<i>anti</i>	<i>gauche</i>	<i>anti</i>	<i>gauche</i>
$E_{\text{sML}}^{\text{DFT}}[n_{\text{sML}}^{\text{DFT}}]$	0.22	0.44	0.30	0.55	0.04	0.58
$E_{\text{sML}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$	0.12	0.49	0.19	0.62	0.13	0.66
$E_{\text{s}\Delta\text{-DFT}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$	0.06	0.01	0.06	0.02	0.01	0.01

Table 2: Energy errors (kcal/mol) of the *sML*-HK maps for ethanol at conventionally-optimized geometries. MP2 and PBE have *gauche* as the global minimum, but CCSD(T) global minimum is *anti*. However, the energy differences between the minima are smaller than model errors, except for the case of Δ -DFT.

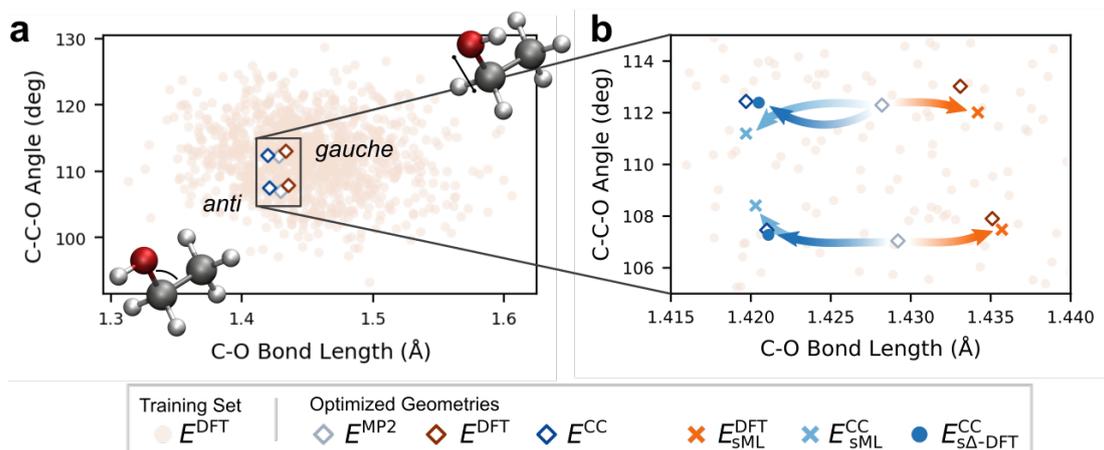


Figure 2: Molecular geometries of ethanol from the ML training set and optimizations. a) 1000 unique configurations used for training, along with the *anti* and *gauche* minima optimized using conventional electronic structure methods b) the configurational space near the minima. Starting from MP2 geometries, the E_{ML} -based optimizations reproduce the subtle differences in DFT- and CC-optimized geometries. For this figure, DFT calculations use the PBE+TS functional and CC calculations use CCSD(T) (see Refs. [41,42] for more details).

Model	$E_{\text{sML}}^{\text{DFT}}$	$E_{\text{sML}}^{\text{CC}}$	$E_{\text{s}\Delta\text{-DFT}}^{\text{CC}}$
$[n_{\text{sML}}^{\text{DFT}}]$	1.26	1.37	0.11
$[n_{\text{sML}}^{\text{DFT}}]$	0.94	0.99	0.11

Table 3: Mean absolute errors (kcal/mol) of the ML models trained on different electronic structure energies for densities learned by the ML-HK map and true densities as inputs.

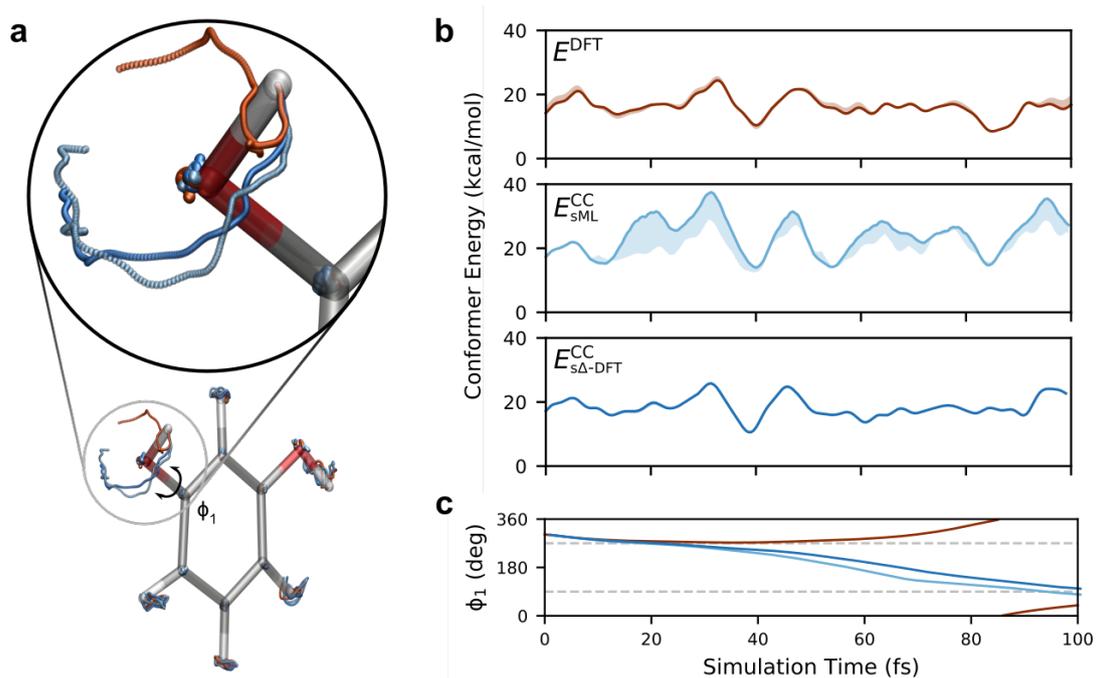


Figure 3: Resorcinol dynamics from an initial condition near a conformational change showing a) the atomic positions explored during 100 fs NVE MD trajectories run with standard DFT (dark orange), $E^{\text{CC}}_{\text{sML}}[n^{\text{DFT}}]$ with RESPA-corrected forces (light blue), and $E^{\text{CC}}_{\text{s}\Delta\text{-DFT}}[n^{\text{DFT}}]$ (blue), b) the conformer energy along each trajectory (solid lines), with the error relative to CC shown as a shaded line width, and c) the evolution of the C-C-O-H dihedral angle for each trajectory with dashed grey lines indicating the barrier between conformers. For this figure, all DFT calculations use PBE and all CC energies are from CCSD(T).

BogjeskiVogt_mldft.pdf (1.52 MiB)

[view on ChemRxiv](#) • [download file](#)

Density functionals with quantum chemical accuracy: From machine learning to molecular dynamics Supplementary Information

Mihail Bogojeski^{*1}, Leslie Vogt-Maranto^{*2}, Mark E. Tuckerman^{†2,3,4}, Klaus-Robert Müller^{†1,5,6}, and Kieron Burke^{†7,8}

¹Machine Learning Group, Technische Universität Berlin, Marchstr. 23, 10587 Berlin, Germany

²Department of Chemistry, New York University, New York, NY 10003, USA

³Courant Institute of Mathematical Science, New York University, New York, NY 10012, USA

⁴NYU-ECNU Center for Computational Chemistry at NYU Shanghai, 3663 Zhongshan Road North, Shanghai 200062, China

⁵Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 02841, Korea

⁶Max-Planck-Institut für Informatik, Stuhlsatzenhausweg, 66123 Saarbrücken, Germany

⁷Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA

⁸Department of Chemistry, University of California, Irvine, CA 92697, USA

1 Molecular Information

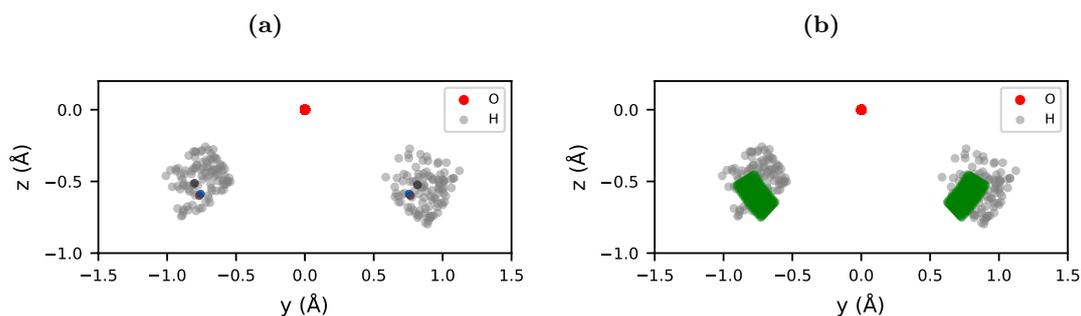
Water

The water molecule has three internal degrees of freedom, two O-H bonds and the H-O-H angle, with bond lengths of 0.97 Å and a bond angle of 104.0° (1.82 radians) for the geometry optimized with the PBE functional. A preliminary dataset composed of 1000 molecular geometries was generated by uniformly sampling O-H bond lengths and H-O-H angles. The 102 geometries for the training set were selected based on the minimum energy conformer of the 1000 structures (1.6 kcal/mol above the optimized geometry), with bond lengths in the range 0.97 ± 0.25 Å and the bond angle in the range of $115 \pm 26^\circ$ (2.00 ± 0.45 radians). In order to simplify the learning problem, the molecules were aligned in the xy-plane with the bisector of the H-O-H angle along the y-axis and the longer O-H bond in the same quadrant. The resulting geometries are shown in Supplementary Fig. 1.

*Contributed equally.

†To whom correspondence should be addressed.

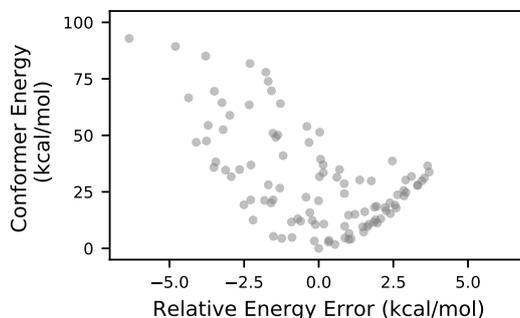
M.E.T.(email:mark.tuckerman@nyu.edu), K.B.(email:kieron@uci.edu), K.-R.M.(email:klaus-robert.mueller@tu-berlin.de)



Supplementary Figure 1: Distribution of 102 water molecule geometries included in the dataset, with bond lengths in the range of 0.72 to 1.22 Å and bond angles in the range of 93.4 to 140.7° (1.63 to 2.46 radians). a) Darker circles indicate the minimum energy conformer of the training set (black) and the geometries after optimization with PBE (orange) or CCSD(T) (blue), b) The extent of the symmetric water test set is shown in green with bond lengths in the range of 0.90 to 1.05 Å and bond angles in the range of 88.5 to 119.5° (1.54 to 2.08 radians).

The conventional coupled cluster calculations were run in Molpro Quantum Chemistry Software [1] or Orca v.3.0.3 [2] using CCSD(T)/aug-cc-pVTZ [3]. For all water geometries in this work, using a single reference wave function is appropriate (T1 diagnostic < 0.02).

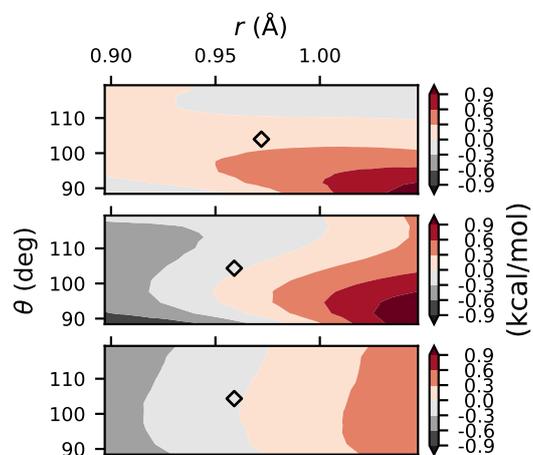
The model is trained on total calculated energies, but there is a shift in the energy range due to the differing bonding energy between electronic structure methods. Therefore, we report relative energies in Fig. 1b based on the lowest energy in the grand training set. These relative energy errors are also shown in Supplementary Fig. 2 and are used to calculate the MAE reported in Fig. 1c. The potential energy surfaces in Fig. 1d are shown relative to the lowest energy conformer for each energy method.



Supplementary Figure 2: Relative energy errors (CC-DFT) for the 102 water geometries in the grand training set plotted against the relative energy as calculated using CC.

The symmetric water data set was constructed for the range of bond lengths and angles sampled during a DFT-MD simulation at 300 K using deviations of 3σ from the average value. The resulting bond range is 0.97 ± 0.075 Å and the angle range is $104.0 \pm 15.5^\circ$ (1.81 ± 0.27 radians). For the symmetric water conformers, the errors for predicted relative energies are qualitatively and quantitatively different for the direct ML energy functionals and the Δ -DFT approach, as shown in Supplementary Fig. 3. The errors in the direct methods reflect the overlap between the training set geometries and the out-of-sample test set, with the largest errors for small angles on the edge of the training set (coordinate range shown in Supplementary Fig. 1). The Δ -DFT errors are significantly smaller and

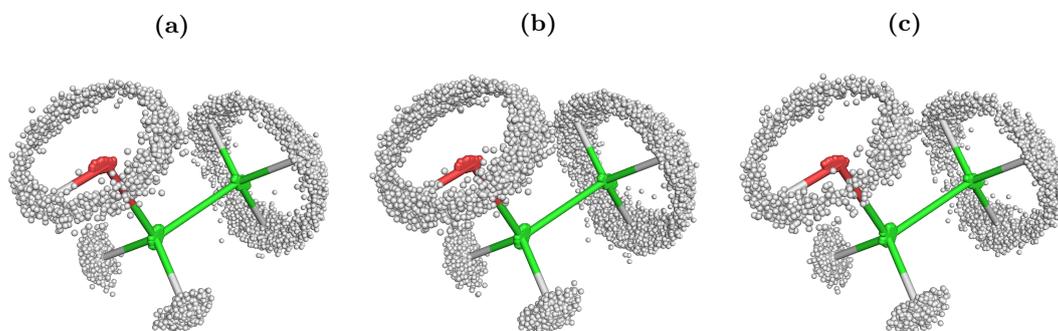
qualitatively different, as they depend much more strongly on the bond length than the bond angle.



Supplementary Figure 3: The relative energy errors for symmetric water geometries compared to the values calculated using traditional electronic structure methods using $E_{ML}^{DFT}[n_{ML}^{DFT}]$ (top, PBE), $E_{ML}^{CC}[n_{ML}^{DFT}]$ (middle, CCSD(T)), and $E_{\Delta-DFT}^{CC}[n_{ML}^{DFT}]$ (bottom, CCSD(T)). The optimized geometries are indicated by open diamonds for each method.

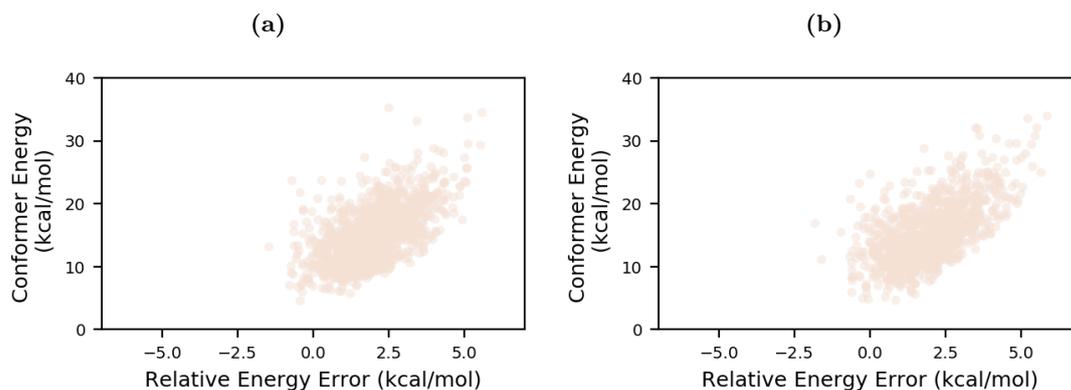
Ethanol

Training and test set data for ethanol is from the MD17 dataset [4, 5]. The 1000 unique geometries in the training set are aligned with the three heavy atoms defining a reflection plane for the symmetry-augmented data set. Supplementary Fig. 4 shows the atomic distributions of the ethanol datasets after alignment and inclusion of symmetry-generated training points.



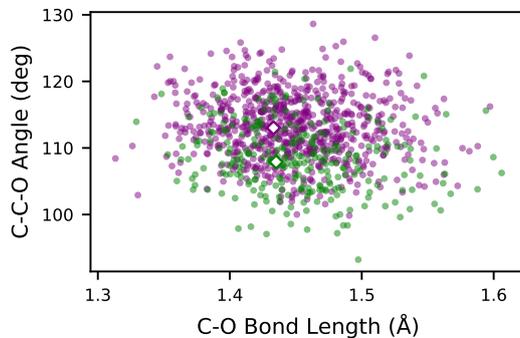
Supplementary Figure 4: Distribution of ethanol geometries included in the dataset, including geometries of a) the original 1000 training set, b) the 2000 point training set generated using mirror symmetry, and c) the 1000 point test set.

The specific data generation methods are reported in Refs. [4, 5]; Briefly, the conformers were generated using DFT-MD at 500 K (PBE+TS) and the coupled cluster energies use CCSD(T)/cc-pVTZ. For all ethanol coupled cluster calculations, the T1 diagnostic is below 0.02. In addition to DFT- and CC-optimized structures, we also optimize each local minima using MP2/6-31g* in Gaussian09 [6]. All geometries are similar, with the largest difference seen for the C-O bond length (DFT > MP2 > CC), as shown in Fig. 2. We report relative energies for ethanol based on the lowest energy minima for each electronic structure method (*anti* for CC by 0.08 kcal/mol and *gauche* for DFT by 0.10 kcal/mol). These relative energy errors are shown in Supplementary Fig. 5, with maximum relative energy errors of 5.6 kcal/mol and 5.9 kcal/mol for geometries included in the training and test sets, respectively.



Supplementary Figure 5: Relative energy errors (CC-DFT) plotted against the relative conformer energy calculated using CC for a) the 1000 unique ethanol geometries in the training set and b) the 1000 ethanol geometries in the test set.

The local minima of ethanol differ most clearly in the dihedral angle of the alcohol OH, but the optimized geometries also differ in the C-C-O angle. The training set shown in Fig. 2a is also shown in Supplementary Fig. 6 sorted by OH dihedral angle.



Supplementary Figure 6: Ethanol training geometries sorted by conformer so that all structures with the alcohol OH in the *anti* (*gauche*) basin are shown in green (purple). The open diamonds show the location of the corresponding DFT-optimized geometries.

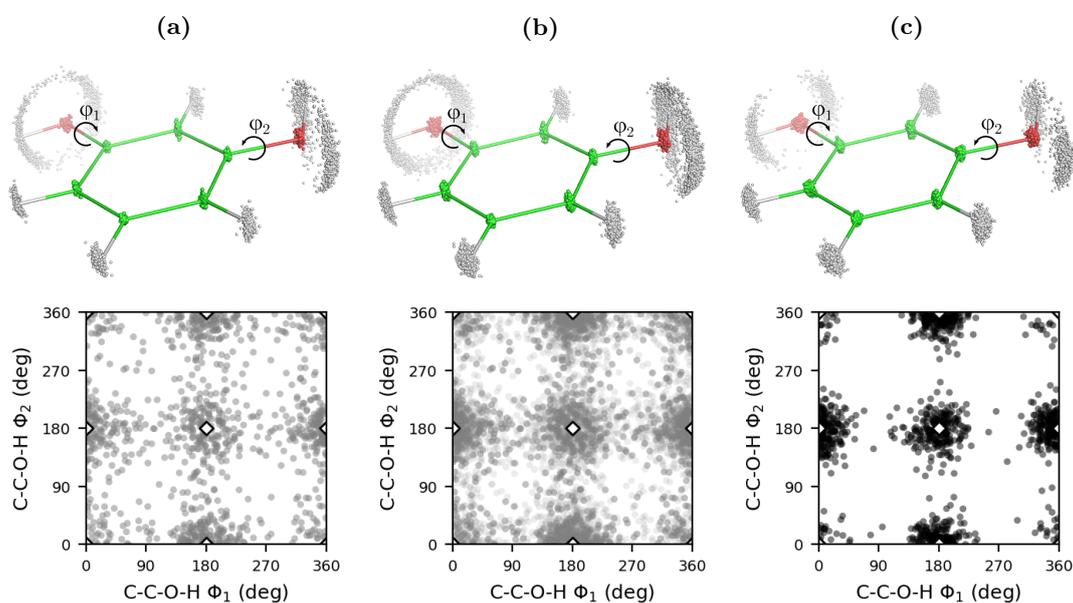
Dataset/Model	$E_{\text{ML}}^{\text{DFT}}[n_{\text{ML}}^{\text{DFT}}]$	$E_{\text{ML}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}]$	$E_{\Delta\text{-DFT}}^{\text{CC}}[n_{\text{ML}}^{\text{DFT}}]$
No symmetries	1.73	1.92	0.15
With symmetries	0.99	1.10	0.09

Supplementary Table 1: Mean average errors (kcal/mol) of the ML maps for ethanol trained on 1000 samples, comparing errors with and without augmenting the dataset using symmetry operations.

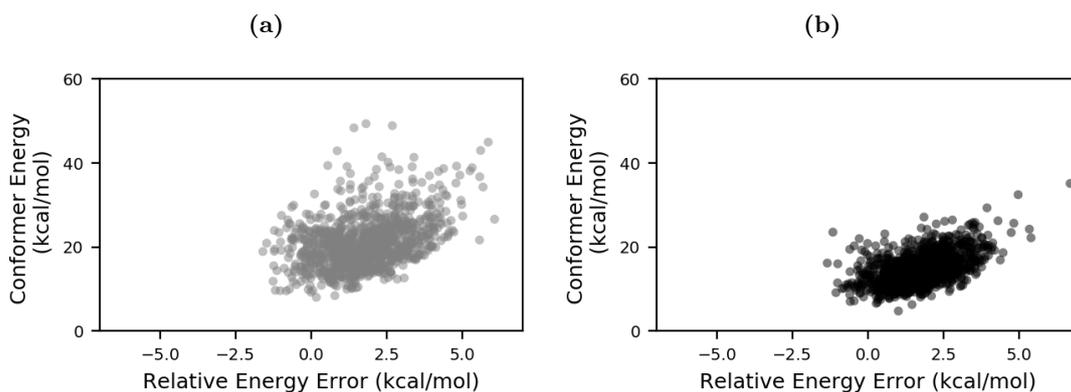
Resorcinol

Resorcinol data set

In order to construct the ML functionals, the resorcinol data sets were generated using standard classical MD simulations. To sample more extreme geometries, training set points were selected from a 500 K trajectory, while the test set was taken from a 300 K simulation. After aligning the molecules to have all C atoms in one plane, the root mean squared deviation (RMSD_C) is 0.045 Å for the 300 K data set and 0.053 Å for the 500 K data set. Supplementary Fig. 7 shows the geometries and C-C-O-H dihedral angles in the original training set, the training set with symmetry operations applied, and the test set. The finite temperature classical MD simulations primarily sample geometries around the minimum energies, but all conformers are more than 7 kcal/mol higher in energy than the global minimum. The all-electron CCSD(T)/cc-pVDZ [3] calculations were run using Orca v.3.0.3 [2], and the single reference wave functions have a maximum T1 diagnostic of 0.012 for all conformers.



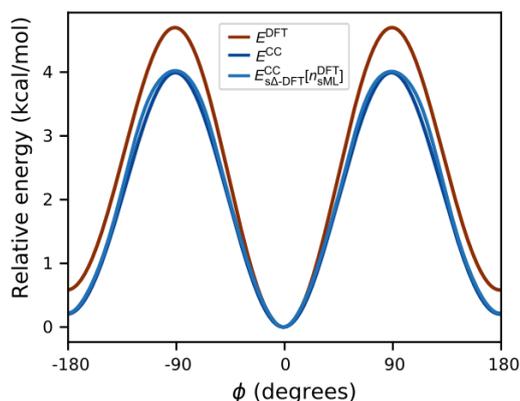
Supplementary Figure 7: Sampled geometries for resorcinol showing atom distribution (top) and distribution of the dihedral angle of the two -OH groups (bottom) for the 1004 point training set (a), 4004 point training set generated with molecular symmetry operations (b), and 1000 point test set (c). The minimum energy conformers are shown as sticks and open black diamonds.



Supplementary Figure 8: Relative energy errors (CC-DFT) plotted against the relative conformer energy calculated using CC for a) the 1000 unique resorcinol geometries in the training set (additional four minima not shown) and b) the 1000 resorcinol geometries in the test set.

Resorcinol rotational barrier

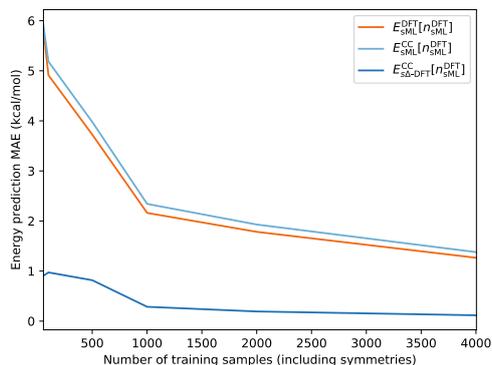
The sparse sampling of geometries away from the minima is due to an energy barrier for each OH group rotation. For an optimized geometry, rotation of a single OH requires crossing an energy barrier of almost 4 kcal/mol based on CC calculations, as shown in Supplementary Fig. 9. The barrier height based on PBE energies is even larger, but can successfully be corrected using the Δ -DFT approach.



Supplementary Figure 9: Relative energies for DFT and CC showing the difference in the OH rotational barrier between the two methods, along with the energy predicted by the $E_{s\Delta-DFT}^{CC}[n^{DFT}_{sML}]$ energy map.

Resorcinol model performance

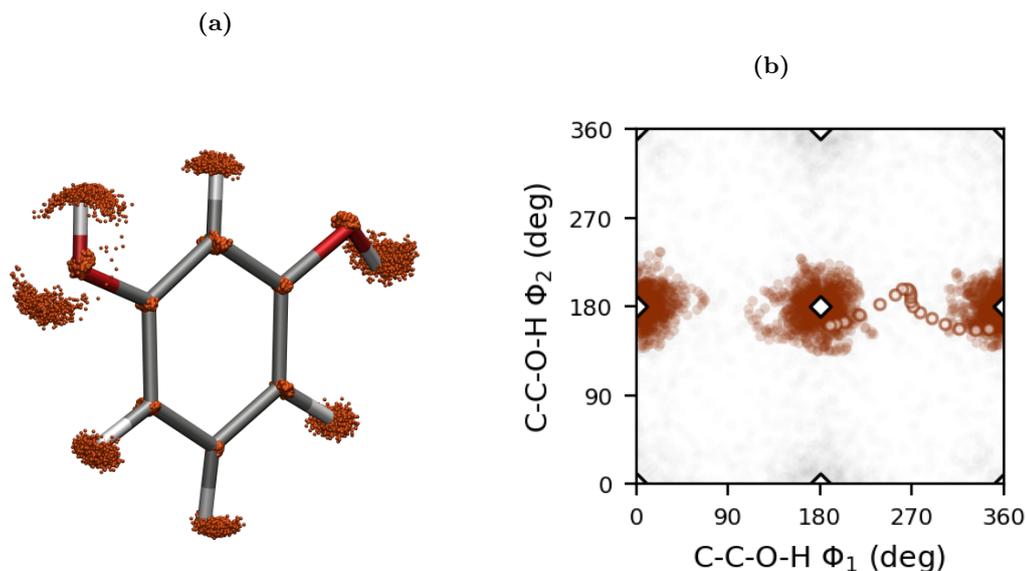
Supplementary Fig. 10 shows the MAE of the different models for different training set sizes of the resorcinol dataset. All training set sizes are counted with the symmetric points obtained by applying the symmetry operations, making the number of unique geometries 1/4 of the total number of training samples.



Supplementary Figure 10: Comparison of the out-of-sample prediction performance for the DFT, CC, and Δ -DFT energy prediction models for different training set sizes (sample counts include symmetries).

Resorcinol DFT-based MD

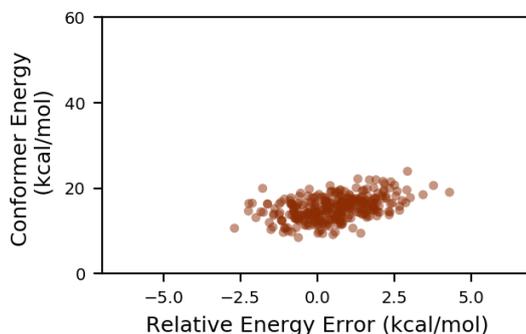
For a DFT-based MD simulation, resorcinol samples two different conformational basins during a 10 ps NVT simulation at 350 K. The aligned atomic positions and C-C-O-H dihedral angles are shown in Supplementary Fig. 11.



Supplementary Figure 11: DFT-based MD for resorcinol at 350 K samples two conformational basins with a) atomic coordinates after alignment for the 10 ps trajectory, and b) the C-C-O-H dihedral angles during the trajectory, with the portion of the trajectory that crosses the rotational energy barrier highlighted with open circles and local minima shown as open diamonds.

Starting from an initial condition close to the barrier crossing event, a DFT-based MD simulation was run using NVE for 1.5 ps. The relative CC energy for snapshots along this trajectory are shown in Supplementary Fig. 12 along with the relative energy errors of the DFT calculations themselves

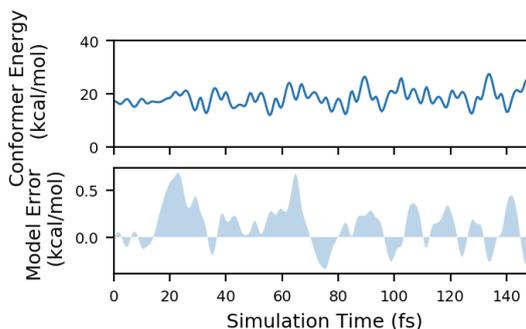
(relative to CC).



Supplementary Figure 12: Relative energy errors (CC-DFT) plotted against the relative conformer energy calculated using CC for an NVE DFT-based MD trajectory for resorcinol starting from an initial condition near the barrier crossing.

Resorcinol ML MD

The $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{s\text{ML}}^{\text{DFT}}]$ model can be used to generate a self-consistent MD trajectory, as seen in Supplementary Fig. 13. Starting from a random training point, the 150 fs trajectory has a MAE of only 0.2 kcal/mol relative to the true CC energies. There is no drift in energy error as the trajectory proceeds, indicating that the Δ -DFT approach is stable for a range of conformers.

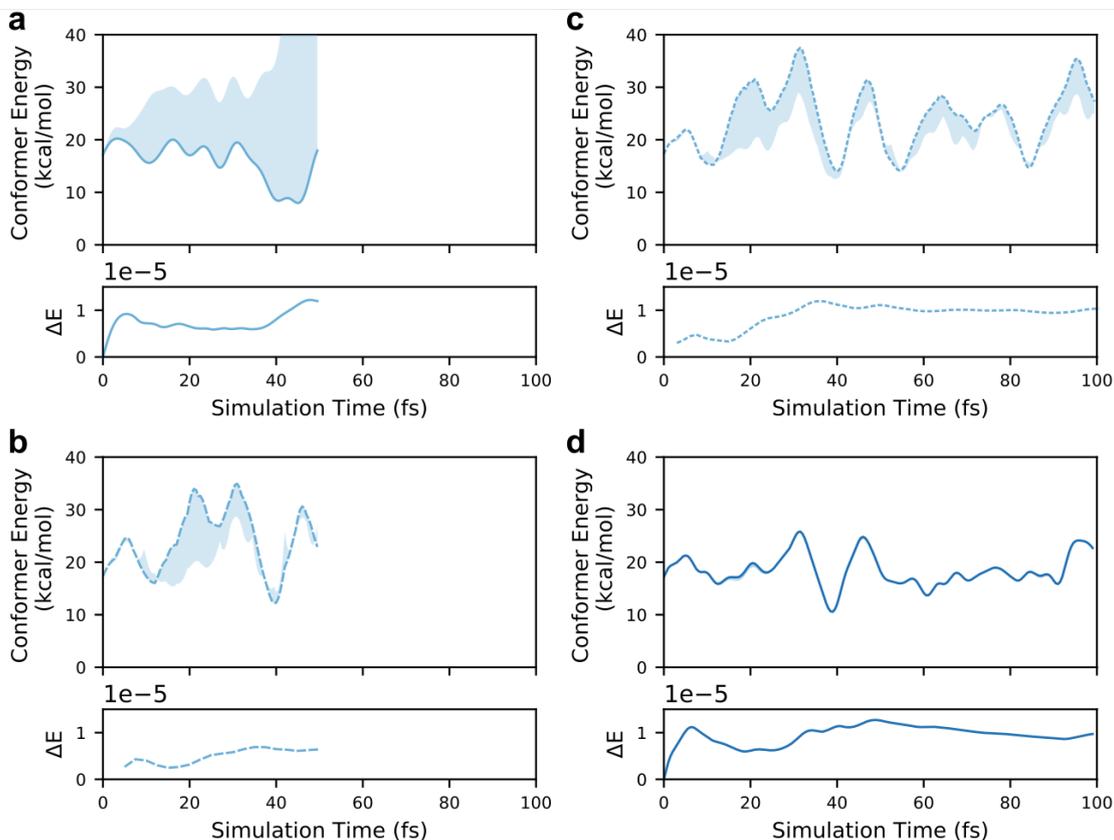


Supplementary Figure 13: $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{s\text{ML}}^{\text{DFT}}]$ energy along the self-generated trajectory is shown in the top panel. The MAE relative to CC is 0.2 kcal/mol, which is smaller than the linewidth, so the energy error is shown separately in the bottom panel.

Resorcinol Multiple time-step MD

The $E_{s\text{ML}}^{\text{CC}}[n_{s\text{ML}}^{\text{DFT}}]$ model can also be used to generate a self-consistent MD trajectory. Starting from an initial condition (atomic positions and velocities) from the DFT-based NVT simulation, the $E_{s\text{ML}}^{\text{CC}}[n_{s\text{ML}}^{\text{DFT}}]$ trajectory explores a region of phase space outside of the training set, as seen in the error relative to CC shown in Supplementary Fig. 14a. The sparsity of training data near this initial position leads to uncertainty in the model. This issue can be mitigated by correcting the forces every few steps using the $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{s\text{ML}}^{\text{DFT}}]$ model (see Supplementary Information Section 6) to prevent excursions into high energy regions. Correcting the forces every fifth time step ($m = 5$) using the

$E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ model already improves the conformer sampling (Supplementary Fig. 14b), with a similar performance seen for a $m = 3$ trajectory (Supplementary Fig. 14c). The trajectories shown in Supplementary Fig. 14c and d are the same as in Fig. 3b.



Supplementary Figure 14: For NVE trajectories starting from the same initial condition, $E_{\text{sML}}^{\text{CC}}$ trajectory energies are shown in the top panels (with errors relative to the true CC energies shown as shaded regions) and conserved quantities in the bottom panels for a) a simulation run with $E_{\text{sML}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ energies and forces, b) a $E_{\text{sML}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ simulation where the $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ is used to correct the forces every fifth time step, c) a $E_{\text{sML}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ simulation where the $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ is used to correct the forces every third time step, and d) a $E_{s\Delta\text{-DFT}}^{\text{CC}}[n_{\text{sML}}^{\text{DFT}}]$ simulation.

2 Using non-self-consistent densities

Our model appears to violate one of the basic results of density functional theory, as it produces CCSD(T) energies from PBE densities. Hohenberg and Kohn [7] showed that, for any given approximate energy functional, one can minimize the energy functional to find a self-consistent formula for the density. The Kohn-Sham scheme is defined [8] to find that density when only the exchange-correlation contribution to the energy is approximated. The purpose of this section is show how the self-consistent density could be found, at least in principle, and also to argue that the energetic consequences of using the PBE density would be negligible here.

There is an exact formula for extracting the energy from any approximation for the ground-state energy for a given external potential [9]:

$$n(\mathbf{r}) = \frac{\delta E[v]}{\delta v(\mathbf{r})} \quad (1)$$

where $E[v]$ is the ground-state energy associated with one-body potential $v(\mathbf{r})$. This could be used to extract a density pointwise from any such approximation: Add a small narrow Gaussian centered at \mathbf{r}_0 to $v(\mathbf{r})$, and note the corresponding change in energy. In the limit of infinitely narrow, infinitely weak perturbations, this yields the density at \mathbf{r}_0 . Of course, such a procedure is highly impractical in a standard basis set of atom-centered Gaussians, but could be easily employed to find specific moments of the density. If the perturbation is a weak static electric field, the prescription yields the dipole moment, as can be seen by multiplying both sides by \mathbf{r} and integrating over all space.

Almost all electronic structure calculations in chemistry and materials science are aimed at finding accurate ground-state energies and the many properties that can be derived from them, such as geometries and barriers. The error in any DFT calculation can be split into two contributions, the functional error and a density-driven error (the energy error due to an incorrect density) [10]. In most DFT calculations (including all those given here), the self-consistent density is so accurate that the energy error is dominated by the functional error [11]: using the exact density in the approximate functional has negligible effect on the energy error. Recent arguments that attempt to distinguish the quality of functionals by constructing metrics of density errors [12] have not held up when analyzed in terms of energies [11].

We can use Supplementary Eq. (1) to analyze the present situation. We know it must be satisfied by the PBE density and energy functional. Thus the difference between the PBE and CCSD(T) densities is simply

$$\Delta n(\mathbf{r}) = \frac{\delta \Delta E^{\text{PBE}}[v]}{\delta v(\mathbf{r})}. \quad (2)$$

This will be a very small energy for normal systems. The fact that the energy difference is easier to learn than the PBE energy itself suggests a smoothness of energy difference with respect to the potential, making density differences tiny.

We also note an additional twist on this question in the context of machine learning. Long ago, Görling and Levy [13] and others pointed out that one could define an exact energy functional on an approximate density, such as the HF density. In fact, as was noted by Li *et al.* [14], to learn accurate *energies*, a very crude representation of the density suffices, so long as it forms a sufficiently useful feature for the energy. In a prototype problem (particle in a box with potential well), with even a very small grid (far too coarse to find accurate solutions to the Schrödinger equation) and essentially exact energies, one could still use kernel ridge regression to find a highly accurate ML functional.

Thus use of PBE densities to find CCSD(T) energies is both practical and theoretically allowed and well understood. On the practical side, it completely avoids the need to extract CCSD(T) densities to train upon. Because the density is not needed to perform a CCSD(T) calculation, it is not available

from many CCSD(T) codes. On the theoretical side, we know (a) the errors in energies will be very small, (b) how to correct them if need be, and (c) that the kernel ridge regression has no difficulty learning on the PBE density (and might also be fine on a much cruder density).

On the other hand, use of non-self-consistent densities breaks the standard relation between energies and forces from the Hellmann-Feynman theorem, and small errors in energies do not automatically imply small errors in forces. Since we use numerical derivatives of energies throughout this paper, we extract the correct forces, but there will be corrections if analytical derivatives are attempted. Whether or not these are significant is beyond the scope of the present work.

3 Kernel ridge regression

Kernel ridge regression (KRR) [15] is a powerful machine learning method for non-linear regression. Non-linearity is achieved by incorporating the kernel trick into Kernel ridge regression, extending linear ridge regression, which finds the optimal linear mapping from the inputs to the labels under ℓ_2 regularization, by exploiting the kernel trick to map the inputs to a high-dimensional non-linear feature space. Let $\mathbf{x}_1, \dots, \mathbf{x}_M \in \mathbb{R}^d$ be the training data points and let $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M]^T$ be their respective labels. The KRR model for a new input sample \mathbf{x}^* is then given by:

$$\mathbf{y}^* = \sum_{i=1}^M \alpha_j k(\mathbf{x}^*, \mathbf{x}_i), \quad (3)$$

where k is a kernel function and $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_M]^T$ are the model weights. The model weights are obtained by solving the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \left\{ \sum_{i=1}^m \left| \mathbf{y}_i - \sum_{j=1}^m \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \right|^2 + \lambda \boldsymbol{\alpha} \mathbf{K} \boldsymbol{\alpha} \right\} \quad (4)$$

where λ is a regularization parameter and \mathbf{K} is the kernel matrix with $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. The analytical solution to the minimization problem is then given by

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{Y}. \quad (5)$$

In this paper we use the Gaussian (radial basis function) kernel

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right), \quad (6)$$

where the kernel width σ is a model parameter that needs to be tuned using cross-validation.

4 Decomposability of the ML-HK map

When using the ML-HK map to predict electron density, the contributions of the prediction error to the cost function are given by

$$\begin{aligned} err(\boldsymbol{\beta}) &= \sum_{i=1}^M \|n_i - n_{\text{ML}}[\mathbf{v}_i]\|_{\mathcal{L}_2}^2 \\ &= \sum_{i=1}^M \left\| n_i - \sum_{l=1}^L u_{\text{ML}}^{(l)}[\mathbf{v}_i] \phi_l \right\|_{\mathcal{L}_2}^2. \end{aligned} \quad (7)$$

By writing the density in terms of its basis representation and assuming orthogonality of the basis functions we obtain

$$\begin{aligned}
err(\boldsymbol{\beta}) &= \sum_{i=1}^M \left\| \sum_{l=1}^L u_i^{(l)} \phi_l - \sum_{l=1}^L u_{\text{ML}}^{(l)}[\mathbf{v}_i] \phi_l \right\|_{\mathcal{L}_2}^2 \\
&= \sum_{i=1}^M \left\| \sum_{l=1}^L \left(u_i^{(l)} - u_{\text{ML}}^{(l)}[\mathbf{v}_i] \right) \phi_l \right\|_{\mathcal{L}_2}^2 \\
&= \sum_{i=1}^M \int \sum_{l=1}^L \left(u_i^{(l)} - u_{\text{ML}}^{(l)}[\mathbf{v}_i] \right) \phi_l(\mathbf{r}) \sum_{l'=1}^L \left(u_i^{(l')} - u_{\text{ML}}^{(l')}[\mathbf{v}_i] \right) \phi_{l'}^*(\mathbf{r}) d\mathbf{r} \\
&= \sum_{i=1}^M \sum_{l,l'=1}^L \left(u_i^{(l)} - u_{\text{ML}}^{(l)}[\mathbf{v}_i] \right) \left(u_i^{(l')} - u_{\text{ML}}^{(l')}[\mathbf{v}_i] \right) \int \phi_l(\mathbf{r}) \phi_{l'}^*(\mathbf{r}) d\mathbf{r} \\
&= \sum_{i=1}^M \sum_{l=1}^L \left(u_i^{(l)} - u_{\text{ML}}^{(l)}[\mathbf{v}_i] \right)^2 \\
&= \sum_{i=1}^M \sum_{l=1}^L \left(u_i^{(l)} - \sum_{j=1}^M \beta_j^{(l)} k(\mathbf{v}_i, \mathbf{v}_j) \right)^2.
\end{aligned} \tag{8}$$

The resulting equation shows that the error can be decomposed into the independent error contributions for each of the basis coefficients. By viewing the errors independently we obtain L separate KRR minimization problems, and analogously to equations 4 and 5 we obtain the analytical solutions

$$\boldsymbol{\beta}^{(l)} = \left(\mathbf{K}_{\sigma^{(l)}} + \lambda^{(l)} \mathbf{I} \right)^{-1} \mathbf{u}^{(l)}, \quad l = 1, \dots, L, \tag{9}$$

where for each basis function ϕ_l , $\lambda^{(l)}$ is a regularization parameter, $\mathbf{u}^{(l)}$ is a vector containing the training set coefficients for the l -th basis function and $\mathbf{K}_{\sigma^{(l)}}$ is a Gaussian kernel matrix with width $\sigma^{(l)}$.

5 Cross-validation

All hyperparameters used in the model are estimated solely on the training set. The width γ and spacing Δ hyperparameters for the artificial Gaussians potential as well as the kernel width σ and the regularization parameter λ were optimized individually for each molecule. In both cases the hyperparameter optimization was performed using cross-validation [16] on the training set. After training and cross-validation, the model is fixed and is applied unchanged on the out-of-sample test set. The optimal hyperparameters for the artificial potentials grid selected using the cross-validation procedure are given in Supplementary Table 2. In order to speed up the search for hyperparameters for the KRR models, we use a fixed regularization hyperparameter value of $\lambda = 2.22 * 10^{-16}$, while only optimizing the kernel width σ . While this may not yield optimal results, in our experience it does not affect the performance significantly, while greatly speeding up the computation time required for cross-validation. Supplementary Tables 3, 4, and 5 show the resulting optimized values for the kernel width σ across the different models and molecules.

Parameters/Molecule	Water	Ethanol	Resorcinol
Grid spacing	0.33	0.19	0.20
Gaussian width γ	0.60	0.36	0.42

Supplementary Table 2: Hyperparameter values for the artificial Gaussians potential selected by cross-validation for every molecule.

Parameter/Model	Water			
	$n_{\text{ML}}^{\text{DFT}}$	$E_{\text{ML}}^{\text{DFT}}$	$E_{\text{ML}}^{\text{CC}}$	$E_{\Delta\text{-DFT}}^{\text{CC}}$
Kernel width σ	7.98	20982	30738	49397629

Supplementary Table 3: Kernel width parameters selected by cross-validation for the different KRR models for the water dataset.

Parameter/Model	Ethanol			
	$n_{\text{ML}}^{\text{DFT}}$	$E_{\text{ML}}^{\text{DFT}}$	$E_{\text{ML}}^{\text{CC}}$	$E_{\Delta\text{-DFT}}^{\text{CC}}$
Kernel width σ	423.9	126976.19	126976.19	409446.22

Supplementary Table 4: Kernel width parameters selected by cross-validation for the different KRR models for the ethanol dataset.

Parameter/Model	Resorcinol			
	$n_{\text{ML}}^{\text{DFT}}$	$E_{\text{ML}}^{\text{DFT}}$	$E_{\text{ML}}^{\text{CC}}$	$E_{\Delta\text{-DFT}}^{\text{CC}}$
Kernel width σ	302.74	160793.66	345085.9	955296.68

Supplementary Table 5: Kernel width parameters selected by cross-validation for the different KRR models for the resorcinol dataset.

6 Multiple time-step molecular dynamics in the ML framework

The generation of a molecular dynamics (MD) trajectory via numerical solution of the equations of motion $\dot{\mathbf{R}}_\alpha = \mathbf{P}_\alpha/M_\alpha$, $\dot{\mathbf{P}}_\alpha = \mathbf{F}_\alpha$, possibly also coupled to a thermostat, where $\alpha = 1, \dots, N$ indexes the N atoms, M_α is the mass of the α th atom, \mathbf{P}_α is its momentum, \mathbf{R}_α is its position, and \mathbf{F}_α is the force on this atom. In DFT, if the true external potential is $v_{\text{ext}}(\mathbf{r}, \mathbf{R})$, where \mathbf{R} denotes the full set of atomic coordinates, then the force is given by

$$\begin{aligned}\mathbf{F}_\alpha &= - \int d\mathbf{r} n(\mathbf{r}) \nabla_\alpha v_{\text{ext}}(\mathbf{r}, \mathbf{R}) + \mathbf{F}_{\alpha, \text{NN}} \\ &\equiv \mathbf{F}_\alpha^{(\text{elec})} + \mathbf{F}_{\alpha, \text{NN}}\end{aligned}\quad (10)$$

where $\mathbf{F}_{\alpha, \text{NN}}$ is the force due to the nuclear-nuclear electronic repulsion. The first term refers to the force originating from the electron-nuclear interaction. In the machine-learning framework, the nesting of functional dependencies, leading to the progression $\mathbf{R} \rightarrow v \rightarrow n \rightarrow E$, such that the electronic force is given by

$$\mathbf{F}_\alpha^{(\text{elec})} = - \int d\mathbf{r} d\mathbf{r}' \frac{\delta E_{\text{ML}}}{\delta n_{\text{ML}}(\mathbf{r})} \frac{\delta n_{\text{ML}}(\mathbf{r})}{\delta v(\mathbf{r}', \mathbf{R})} \nabla_\alpha v(\mathbf{r}', \mathbf{R}) \quad (11)$$

In a standard MD calculation, the numerical integration algorithm for generating P steps of MD using a time step Δt , is structured according to the pseudocode shown below:

$$\begin{aligned}&\text{for } i \text{ in range}(P) \\ &\quad \text{for } \alpha \text{ in range}(N) \\ &\quad\quad \mathbf{P}_\alpha \leftarrow \mathbf{P}_\alpha + \Delta t * \mathbf{F}_\alpha / 2 \\ &\quad\quad \mathbf{R}_\alpha \leftarrow \mathbf{R}_\alpha + \Delta t * \mathbf{P}_\alpha / M_\alpha \\ &\quad\quad \text{Update Forces} \\ &\quad\quad \mathbf{P}_\alpha \leftarrow \mathbf{P}_\alpha + \Delta t * \mathbf{F}_\alpha / 2\end{aligned}\quad (12)$$

The obvious bottleneck in an MD calculation, which often restricts the value of P , *i.e.*, the time scale that can be accessed, is the computational overhead associated with the force calculation, as each step requires a full calculation of \mathbf{F}_α . The computational time required to generate an MD trajectory can be reduced if the force can be subdivided into a component that has a low computational overhead and a correction that varies on a slower time scale and carries most of the computational overhead of the full force calculation. Denoting the former of these as a reference for $\mathbf{F}_\alpha^{(\text{ref})}$ and the correction as $\delta\mathbf{F}_\alpha$, the full force is then $\mathbf{F}_\alpha = \mathbf{F}_\alpha^{(\text{ref})} + \delta\mathbf{F}_\alpha$. With this force decomposition, a reversible, symplectic multiple time-step integration algorithm can be constructed [17] based on the assumption that the correction $\delta\mathbf{F}_\alpha$ only needs to be updated every m steps, where typically $m \sim 5$, which, if the computational overhead of the reference force calculation is negligible compared to the correction, will reduce the computational cost of the calculation by a factor of m . The algorithm, called the reversible reference system propagator algorithm (reversible RESPA), has the same structure as that shown in Supplementary Eq. (12) with the following provision: In each step, the default is to use $\mathbf{F}_\alpha^{(\text{ref})}$ in place of \mathbf{F}_α in each step, and the force update is only an update of the reference force; every m steps, however, one uses the force $\mathbf{F}_\alpha^{(\text{ref})} + m\delta\mathbf{F}_\alpha$, and the force update requires calculation of both the reference force and the correction.

In our Δ -machine learning approach, a natural force decomposition arises from the expression for

the energy

$$\begin{aligned} E^{\text{CC}}[n] &= E^{\text{DFT}}[n] + \Delta E_{\text{ML}}[n] \\ &= E_{\text{ML}}^{\text{CC}}[n] + (E^{\text{DFT}}[n] + \Delta E_{\text{ML}}[n] - E_{\text{ML}}^{\text{CC}}[n]) \\ &\equiv E_{\text{ML}}^{\text{CC}}[n] + \delta E_{\text{ML}}[n] \end{aligned} \tag{13}$$

where the density $n(\mathbf{r})$ is either the explicit PBE density or the density from the Hohenberg-Kohn map. Note that, in the second line, we have added and subtracted the direct CCSD(T) ML model. We, thus, associated the force obtained from $E_{\text{ML}}[n]$ with the reference force $\mathbf{F}_{\alpha}^{(\text{ref})}$, the computational overhead of which is quite low. We then associate the correction $\delta \mathbf{F}_{\alpha}$ with the energy correction $\delta E_{\text{ML}}[n]$. As this term requires a full DFT calculation, its computational overhead is significantly higher, and the overall reduction in computational time is very nearly equal to the value of m .

References

- [1] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, et al. Molpro, version 2015.1, a package of ab initio programs, 2015.
- [2] Frank Neese. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):73–78, 2012.
- [3] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989.
- [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] Stefan Chmiela, Huziel E. Sauceda, Klaus-Robert Müller, and Alexandre Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, 9(1):3887, 2018.
- [6] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, et al. Gaussian 09, 2009.
- [7] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, Nov 1964.
- [8] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, Nov 1965.
- [9] R. G. Parr and W. Yang. *Density Functional Theory of Atoms and Molecules*. Oxford University Press, 1989.
- [10] Adam Wasserman, Jonathan Nafziger, Kaili Jiang, Min-Cheol Kim, Eunji Sim, and Kieron Burke. The importance of being inconsistent. *Annual Review of Physical Chemistry*, 68(1):555–581, 2017.
- [11] Eunji Sim, Suhwan Song, and Kieron Burke. Quantifying density errors in DFT. *J. Phys. Chem. Lett.*, 9:6385–6392, October 2018.
- [12] Michael G. Medvedev, Ivan S. Bushmarinov, Jianwei Sun, John P. Perdew, and Konstantin A. Lyssenko. Density functional theory is straying from the path toward the exact functional. *Science*, 355(6320):49–52, 2017.
- [13] M. Levy and A. Görling. Correlation energy density-functional formulas from correlating first-order density matrices. *Phys. Rev. A*, 52:R1808, 1995.
- [14] Li Li, John C. Snyder, Isabelle M. Pelaschier, Jessica Huang, Uma-Naresh Niranjana, Paul Duncan, Matthias Rupp, Klaus-Robert Müller, and Kieron Burke. Understanding machine-learned density functionals. *International Journal of Quantum Chemistry*, 116(11):819–833, 2016.
- [15] Trevor Hastie, Robert Tibshirani, and J.H. Friedman. *The elements of statistical learning: Data mining, inference, and prediction*, 2009.
- [16] Katja Hansen, Grégoire Montavon, Franziska Biegler, Siamac Fazli, Matthias Rupp, Matthias Scheffler, O. Anatole von Lilienfeld, Alexandre Tkatchenko, and Klaus-Robert Müller. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.*, 9(8):3404–3419, 2013.

- [17] Mark E. Tuckerman, Bruce J. Berne, and Glenn J. Martyna. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3):1990–2001, 1992.

BogjeskiVogt_mldft_esi.pdf (3.61 MiB)

[view on ChemRxiv](#) • [download file](#)
