The Journal of Chemical Physics

RESEARCH ARTICLE | APRIL 19 2023

A semilocal machine-learning correction to density functional approximations

JingChun Wang; Yao Wang; Rui-Xue Xu; ... et. al

J. Chem. Phys. 158, 154107 (2023)
https://doi.org/10.1063/5.0148438

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

A semilocal machine-learning correction to density functional approximations

Cite as: J. Chem. Phys. 158, 154107 (2023); doi: 10.1063/5.0148438
Submitted: 1 March 2023 • Accepted: 5 April 2023 •
Published Online: 19 April 2023

JingChun Wang,[1,2] Yao Wang,[3] Rui-Xue Xu,[2,3] GuanHua Chen,[4] and Xiao Zheng[1,3,a)]

AFFILIATIONS

[1] Department of Chemistry, Fudan University, Shanghai 200433, China
[2] CAS Key Laboratory of Precision and Intelligent Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China
[3] Hefei National Research Center for Physical Sciences at the Microscale and Synergetic Innovation Center of Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, Anhui 230026, China
[4] Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong, China

[a)] **Author to whom correspondence should be addressed:** xzheng@fudan.edu.cn

ABSTRACT

Machine learning (ML) has demonstrated its potential usefulness for the development of density functional theory methods. In this work, we construct an ML model to correct the density functional approximations, which adopts semilocal descriptors of electron density and density derivative and is trained by accurate reference data of relative and absolute energies. The resulting ML-corrected functional is tested on a comprehensive dataset including various types of energetic properties. Particularly, the ML-corrected Becke's three parameters and the Lee–Yang–Parr correlation (B3LYP) functional achieves a substantial improvement over the original B3LYP on the prediction of total energies of atoms and molecules and atomization energies, and a marginal improvement on the prediction of ionization potentials, electron affinities, and bond dissociation energies; whereas, it preserves the same level of accuracy for isomerization energies and reaction barrier heights. The ML-corrected functional allows for fully self-consistent-field calculation with similar efficiency to the parent functional. This study highlights the progress of building an ML correction toward achieving a functional that performs uniformly better than B3LYP.

I. INTRODUCTION

Density functional theory (DFT) has become one of the most promising branches of quantum chemistry and has facilitated the development of various disciplines in natural science and engineering.[1–3] With a wide range of applications in molecular and bulk systems, DFT methods achieve an excellent balance between accuracy and efficiency for general-purpose calculations. The Kohn–Sham (KS) scheme[4] enables the calculation of electron density and energy in a self-consistent-field (SCF) manner. The success of KS-DFT lies in the conversion of solving complex many-body wavefunctions of real electronic systems to the solution of effective noninteracting systems, given the exchange-correlation (XC) functional. Over the years, there have been many efforts of various density functional approximations (DFAs) seeking better formulations of the XC functional with increasing accuracy. Despite the success of DFAs designed by analytic derivation subject to physical constraints[5,6] or fitting with semi-empirical parameters,[7–9] a universally accurate and computationally efficient DFA remains to be found. One way to achieve this, as is described by "Jacob's ladder,"[10] is to incorporate more sophisticated density descriptors into the design of DFA to systematically enhance its accuracy.

Machine learning (ML) techniques[11–13] have been employed to improve existing density functional methods and to develop new DFAs.[14,15] Since the early attempt by Tozer *et al.*[16] in which an artificial neural network (ANN)[17] was employed to fit the Zhao–Morrison–Parr XC potential, numerous works have been devoted to combining the DFT methods and ML techniques to enhance the performance of DFAs toward chemical accuracy.[18–64] For instance, Nagai *et al.* have proposed a systematic approach to construct ANN-based DFAs by using density descriptors on each rung of "Jacob's ladder."[37] In a recent study, an XC functional constructed with a deep neural network was trained on fractional charge and fractional spin systems to satisfy rigorous physical constraints.[55]

The resulting DM21 functional has demonstrated promising performance for the treatment of charge delocalization and strong correlation. Pokharel *et al.* have designed a deep neural network to de-orbitalize the strongly constrained and appropriately normed (SCAN) functional,[6] and the resulting ML functional replicates the performance of SCAN by utilizing the information of electron density and density derivative while avoiding the use of the orbital-dependent kinetic energy density.[56] Recently, Nagai *et al.* have constructed an ML-corrected SCAN functional.[38] By introducing a number of asymptotic constraints as regularization for the ML parameters, they have improved the accuracy of the ML-based functional and extended its applicability from molecules to bulk solids.

It is practically more appealing to design a functional correction than to build a whole new DFA from scratch, provided that the correction does not break appreciably the physical constraints intended by the original DFA. In their early work,[16] Tozer *et al.* proposed the construction of a local correction to the XC potential with ANN. They have also discussed the challenges of building an XC potential with ANN directly from electron density and density gradient. In our previous work, we have constructed an ML correction to the Perdew–Burke–Ernzerhof (PBE) functional[65] by establishing a semilocal mapping from the electron density and reducing the gradient to a correction to the XC energy density.[33] The resulting ML-corrected PBE is directly applicable to any molecular system and yields significantly improved heats of formation (HOF) while maintaining the accuracy of PBE for other thermochemical and kinetic properties. However, our previous attempt has several limitations. The first limitation lies in the preparation of reference data. For a molecular species, the error of calculated thermochemical energy is determined by comparing it to the highly accurate experimental data and then assigned to each point in the **r**-space. Ideally, every **r**-point contributes an independent data entry to the construction of the semilocal mapping. However, the pointwise errors originating from the same global error are not mutually independent. Moreover, to facilitate the training of the ML model, we attributed the errors of the calculated HOF to the molecules alone but considered the atomic energies to be constants. It is certainly desirable to treat the molecular and atomic species on the same footing. The second limitation is that the resulting ML correction works only in a post-SCF manner.

In this work, we aim to develop an accurate and efficient semilocal ML correction to DFAs, which is implemented in an SCF manner, and thus remove the above two limitations. The hybrid functional with Becke's three parameters and the Lee–Yang–Parr correlation (B3LYP)[66–68] achieves an outstanding trade-off between accuracy and efficiency and sets a milestone for DFAs. While there are DFAs that belong to higher rungs of the "Jacob's ladder" and hence are much more accurate than the B3LYP functional, e.g., the doubly hybrid functionals,[69–73] we choose to focus on computationally less demanding DFAs, with the hope that the ML-corrected DFAs are immediately applicable to general chemical systems. Thus, the goal of this work is to address the apparent challenge "to develop a functional that performs uniformly better than B3LYP."[74]

The remainder of this paper is organized as follows: Section II describes the construction of the ML model for correcting the DFAs. Section III exhibits the numerical performance of the ML-DFA across a variety of test sets along with an in-depth analysis of the numerical outcomes. Concluding remarks are finally given in Sec. IV.

II. METHODOLOGY

A. General design of the ML-corrected DFA

Instead of building a superior DFA entirely from scratch, we propose to design a semilocal correction term, $\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}(\mathbf{r})$, represented by an ML model, to the XC energy functional, $E_{\mathrm{XC}}^{\mathrm{DFA}}$. The resulting ML-corrected DFA is abbreviated as ML-DFA and expressed as

$$E_{\mathrm{XC}}^{\mathrm{ML-DFA}} = E_{\mathrm{XC}}^{\mathrm{DFA}} + \int \mathrm{d}\mathbf{r}\rho(\mathbf{r})\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}(\mathbf{r}). \tag{1}$$

For simplicity, the density descriptors of common generalized gradient approximation (GGA) functionals are adopted as the features of the ML model,[33] resulting in a semilocal mapping that eventually completes the KS mapping of the parent functional,

$$\{r_s, \zeta, s\} \mapsto \Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}(\mathbf{r}). \tag{2}$$

Here, the Wigner–Seitz radius $r_s = (4\pi/3)^{-1/3}\rho^{-1/3}$, the relative spin polarization $\zeta = (\rho_\uparrow - \rho_\downarrow)/\rho$, and the reduced density gradient $s = |\nabla\rho|/[2(3\pi^2)^{1/3}\rho^{4/3}]$ represent the electron density and density gradient.[65] In principle, the semilocal ML correction can be applied to any DFA.

To enable an SCF implementation of the ML-DFA, the XC potential of the ML-corrected functional is calculated as follows:

$$v_{\mathrm{XC};\sigma}^{\mathrm{ML-DFA}} \equiv \frac{\delta E_{\mathrm{XC}}^{\mathrm{ML-DFA}}[\rho]}{\delta\rho_\sigma} = v_{\mathrm{XC};\sigma}^{\mathrm{DFA}} + \Delta v_{\mathrm{XC};\sigma}^{\mathrm{ML}}, \tag{3}$$

with $\sigma = \uparrow, \downarrow$ and

$$\begin{aligned}\Delta v_{\mathrm{XC};\sigma}^{\mathrm{ML}}(\mathbf{r}) &= \frac{\partial f^{\mathrm{ML}}}{\partial\rho_\sigma} - \nabla\cdot\frac{\partial f^{\mathrm{ML}}}{\partial\nabla\rho_\sigma} \\ &= \frac{\partial f^{\mathrm{ML}}}{\partial r_s}\left(-\frac{4\pi}{9}\right)r_s^4 + \frac{\partial f^{\mathrm{ML}}}{\partial s}\left(-\frac{s}{\pi r_s^3}\right) \\ &\quad + \frac{\partial f^{\mathrm{ML}}}{\partial\zeta}\left(\frac{4\pi}{3}\right)r_s^3\left[\mathrm{sgn}(\sigma) - \zeta\right] \\ &\quad + \nabla\cdot\left(\frac{\partial f^{\mathrm{ML}}}{\partial s}\frac{3^{14/3}s^3}{16\pi^{5/3}r_s^{12}}\nabla r_s\right),\end{aligned} \tag{4}$$

where $f^{\mathrm{ML}} = \rho\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}$; see the Appendix for the detailed derivation of Eq. (4). The challenge then is to find a proper form of ML-DFA so that the corrected XC functional yields the most accurate prediction for thermochemical properties.

B. The ANN correction to XC energy density

According to Eq. (4), the ML correction to the XC potential requires the calculation of the derivatives $\partial\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}/\partial r_s$, $\partial\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}/\partial s$, and $\partial\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}/\partial\zeta$, which means that the ML model has to be differentiable. Albeit highly accurate and flexible, the XGBoost platform[75] employed in our previous work is non-differentiable and thus is not applicable for modeling $\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}$. Instead, ANN is completely differentiable with continuous activation functions and is chosen as the ML model in this work.

The ML correction to the XC energy density, $\Delta\epsilon_{\rm XC}^{\rm ML}(\mathbf{r})$, is modeled by an ANN that is integrated with the parent DFA to perform SCF calculations. The ANN model is defined locally at each $\mathbf{r}$-point of the numerical integration grids set up for the density functional calculations. The weights of the ANN are determined via a non-gradient optimization algorithm instead of the gradient descent algorithm in conventional ML.

Throughout this work, the ANN model is implemented by using PyTorch,[76] an open source machine learning framework. The automatic differentiation mechanism of PyTorch is convenient for computing the functional derivatives required for $\Delta v_{{\rm XC},\sigma}^{\rm ML}$. The activation function is chosen as the sigmoid function,[17] which is differentiable on its domain $\mathbb{R}$. The network structure is chosen to be $3\times20\times20\times20\times1$ with three input neurons holding the GGA-level density descriptors $\{r_s,\zeta,s\}$, one output neuron giving the $\Delta\epsilon_{\rm XC}^{\rm ML}(\mathbf{r})$, and three fully connected hidden layers with 20 neurons in each layer encoding the complex semilocal mapping. The number of hidden layers of ANN in this work was not pre-determined but discovered gradually by trial and error. It is found that a one- or two-layer network often makes the SCF calculation hard to converge due to extreme values of the functional derivatives, while a network deeper than five layers is difficult to be optimized to give an effective correction to the original DFA. The number of hidden layers is finally set to three for the ANN model.

C. Training of the ANN model

To determine the optimal weights of the ANN, the gradient-based algorithms for conventional supervised ML require a direct label of each sample defined at the $\mathbf{r}$-points for the optimization. Usually, highly accurate thermochemical energies obtained from experimental measurements or high-level quantum chemistry calculations are adopted as reference data. However, reference energy data do not correspond to a single $\mathbf{r}$-point but involve a large number of $\mathbf{r}$-points associated with a chemical species.

As mentioned in Sec. I, it is rather difficult to decompose the error associated with the energetic data of a species into pointwise contributions. Consequently, it is difficult to apply the commonly used back propagation strategy and gradient-based algorithms for optimizing the weights of ANN straightforwardly. Another factor that hinders the back propagation of errors is that the ML-corrected KS mapping still involves an SCF procedure so the loss function (see below) has an implicit dependence on the density descriptors. With the current architecture of the ML-corrected functional, the error can only propagate forwardly through the ANN model. Thus, a gradient-free optimization algorithm, the particle swarm optimization (PSO) algorithm,[77,78] is employed to determine the weights $\mathbf{w}$ of the ML model.

The PSO algorithm focuses on a single quantity, the loss function Ω, to optimize the weights of the ANN. The training samples are defined locally at the integration $\mathbf{r}$-points of the training species as the data entry with input features of $\{r_s,\zeta,s\}$ and output of $\Delta\epsilon_{\rm XC}^{\rm ML}(\mathbf{r})$, but the target values for $\Delta\epsilon_{\rm XC}^{\rm ML}(\mathbf{r})$ are not explicitly given. Instead, the goal of learning is to adjust the weights so that the integration of the ANN outputs for all the $\mathbf{r}$-points associated with a species is as close as possible to the reference energy of that species. The training proceeds as follows. Given a specific set of weights $\mathbf{w}$, SCF calculations are carried out with the ML-corrected DFA to obtain the energy data of the chemical species in the training set. These energy data are then evaluated by comparing them with the reference energies to provide a global loss $\Omega(\mathbf{w})$. The PSO algorithm takes the global loss as the feedback to update the $\mathbf{w}$ directly, without calculating the gradients of $\Omega(\mathbf{w})$.

The reference energies in the training set consist of two parts, relative energies involving several species and absolute energies of individual species. The relative energies are taken as the atomization energies (AEs) of a number of molecules chosen from the G2/97 test set,[79] which are calculated by the explicitly correlated coupled-cluster theory with iterative single, double and perturbative triple excitations, CCSD(T)(F12), using the frozen-core approach and augmented with higher-order corrections for core/core-valence correlation and higher-excitation effects.[80] We also include the total energies (TEs) of these molecules at their equilibrium geometries and the TEs of their constituent atoms calculated at the same level as the reference energies.

While the performance of some DFAs relies heavily on the cancellation of errors between species,[74,81] it is more favorable if the superior accuracy originates from the correct KS mapping. For the latter, it has been demonstrated that the explicit consideration of absolute energies in the development of DFAs can improve the prediction of both TEs and relative energies, leading to enhanced applicability to more species and properties.[68,74,82] In this work, the including of absolute energies during training is expected to promote the accuracy of ML-DFA by serving as a regularization for the optimization of the weights.

Given the training set, the loss function Ω is defined as

$$\Omega(\mathbf{w}) = \frac{1}{M}\sum_{m=1}^{M}\left|\frac{(\mathrm{AE})_m^{\mathrm{ML-DFA}}-(\mathrm{AE})_m^{\mathrm{ref}}}{(\mathrm{AE})^{\mathrm{DFA}}(\mathrm{H_2O})}\right| + \frac{\alpha}{N}\sum_{n=1}^{N}\left|\frac{E_{\mathrm{tot},n}^{\mathrm{ML-DFA}}-E_{\mathrm{tot},n}^{\mathrm{ref}}}{E_{\mathrm{tot}}^{\mathrm{DFA}}(\mathrm{H_2O})}\right|, \tag{5}$$

with

$$(\mathrm{AE})_m^{\mathrm{ML-DFA}} = \sum_{a\in m}E_{\mathrm{tot},a}^{\mathrm{ML-DFA}} - E_{\mathrm{tot},m}^{\mathrm{ML-DFA}} \tag{6}$$

and

$$E_{\rm XC}^{\mathrm{ML-DFA}} = E_{\rm XC}^{\rm DFA} + \int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r})\,\Delta\epsilon_{\rm XC}^{\rm ML}(\mathbf{r};\mathbf{w}). \tag{7}$$

Here, M is the number of AEs in the training set, N is the number of TEs of related molecules and constituent atoms, $\sum_{a\in m}$ denotes the summation over all the constituent atoms of a molecule, and $(\mathrm{AE})_m^{\mathrm{ref}}$ and $E_{\mathrm{tot},n}^{\mathrm{ref}}$ are the reference values of AEs and TEs. The errors of the AEs and TEs are scaled to be dimensionless by the corresponding values of the H_2O molecule calculated with the original DFA. To assess training errors, the XC energy density of ML-DFA is first integrated over $\mathbf{r}$-space to obtain TEs of the training species, and then, the related AEs are calculated and compared to the corresponding reference energies to generate the global loss.

The ratio of the absolute energy to the relative energy is adjusted by the hyperparameter α in the loss function. The ratio plays a critical role in the optimization as it balances the correction of the model to the energy difference and to the absolute energy prediction itself, eventually extending the applicability of ML-DFA to the

prediction of a wider range of chemical species. Despite its importance, it requires extra effort to find the optimal value of the ratio, since there is no prior knowledge available. The initial value of α is determined by performing sample statistics for the G2/97 set, i.e., the ratio is set to $\Sigma_m^M \Delta(\mathrm{AE})_m^{\mathrm{DFA}} / \Sigma_m^M \Delta E_{\mathrm{tot},m}^{\mathrm{DFA}}$, which is 0.16. $\Delta(\mathrm{AE})_m^{\mathrm{DFA}}$ and $\Delta E_{\mathrm{tot},m}^{\mathrm{DFA}}$ are the errors of AEs and TEs associated with the original DFA, respectively, for species in the G2/97 set. The value of α is then adjusted to minimize the validation error.

Figure 1 illustrates the workflow for constructing the ML-corrected DFA. The training protocol of ML-DFA can be generally described as a two-stage nested global optimization of weights $\mathbf{w}$. The initial set of weights, denoted as $\mathbf{w}_0$, is chosen to be $\mathbf{0}$ since the starting point of ML-DFA is the original DFA. At the ith step of the optimization, the following two stages are accessed iteratively.

1. Given the set of weights $\mathbf{w}_i$ from the last step, do KS SCF calculations with a maximum of 50 cycles over all the species in the training set. The calculations yield the XC energy density $\{\epsilon_{\mathrm{XC}}^{(i+1)}(\mathbf{r};\mathbf{w}_i)\}$ by ML-DFA for the ith step.
2. Integrate $\{\epsilon_{\mathrm{XC}}^{(i+1)}(\mathbf{r};\mathbf{w}_i)\}$ over $\mathbf{r}$-space for all the species and calculate the global loss $\Omega(\mathbf{w}_i)$ by Eq. (5). If the global loss is within the pre-defined threshold e_1 and all the residual deviations of KS SCF calculations have reached the convergence threshold e_2, i.e., $\Omega(\mathbf{w}_i) \le e_1$ and $e_{\mathrm{SCF}}^{(i+1)} \le e_2$, the weights $\mathbf{w}_i$ are deemed optimal, and the training process is completed. If the former e_1 criteria are not satisfied, the optimization algorithm updates the weights of ANN to $\mathbf{w}_{i+1}$ by minimizing the current loss $\Omega(\mathbf{w}_i)$ and then returns to stage 1 with the updated weights $\mathbf{w}_{i+1}$ and repeats the process until the two convergence criteria are met. Specifically, if the latter e_2 criteria are not satisfied, which indicates a failure in converging the SCF calculations, the current weight $\mathbf{w}_i$ is discarded, and the optimization continues to the next step.

D. Dataset and computational details

The dataset adopted for training, validation, and prediction involves absolute energies and relative energies. Specifically, the latter include AEs, HOF, ionization potentials (IPs), electronic affinities (EAs), isomerization energies (IEs), bond dissociation energies (BDEs), and reaction barriers.

The composition of the training set is an essential factor that affects the optimization of the ANN model. Although incorporating more species in the training set is likely to yield a more accurate model in terms of test performance, the size and number of species in the training set are limited by the computational resources available under the non-gradient optimization protocol, since the optimization involves hundreds of iterations of full SCF calculations on the entire training set. For a larger training set, it also requires a more sophisticated design and tuning of the loss function in Eq. (5) to achieve better training performance. This is because, otherwise, the contribution of the species with small errors in AE or TE will be flooded by those with significant errors during the optimization. We choose three molecules $\{H_2O, C_2H_2, SO_2\}$ from the G2/97 set and adopt the AEs and TEs of these molecules along with the TEs of their constituent atoms $\{H, C, O, S\}$ that occur most frequently in the original set to form the training set. We assess the training samples at $\mathbf{r}$-points with the NCI-plot[83] in comparison to the data points of the G2/97 set; see the supplementary material for details. According to the distributions of the $\mathbf{r}$-points, the training set has covered a large fraction of the space (ρ, s) for species in the G2/97 set.

The three training species are taken from the 148 AEs of the G2/97 test set,[79] and the remaining 145 AEs form the validation set, designated as G2-AE. The reference energies in the original G2/97 test set were given as HOF from experimental records. To avoid introducing any unnecessary systematic errors from the thermal correction for the HOF calculation, the reference energies for the AEs in both the training and validation sets along with the reference TEs for the corresponding 148 molecules and the constituent atoms are calculated by a high-level wavefunction method.[80] According to the original research,[80,84] the average error associated with this computational protocol is expected to be 0.1 kJ/mol per valence electron, and 95% of the AEs in the G2-AE set were predicted with an error smaller than 2.1 kJ/mol (within the chemical accuracy of 1 kcal/mol) with respect to the active thermochemical tables

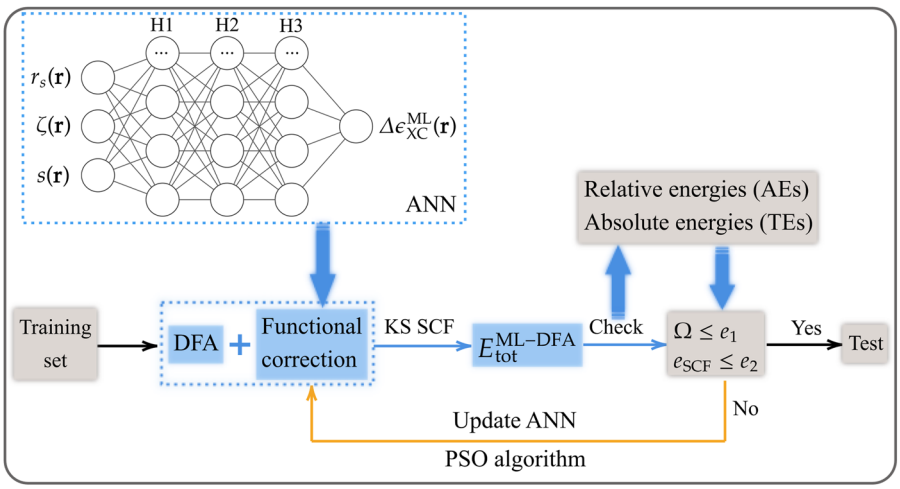

FIG. 1. Schematic diagram for the construction of the ML-corrected DFA. Details about the training process are elaborated in Sec. II C. The total error is evaluated via the loss function Ω, and e_{SCF} is the residual error of the SCF calculations, where e_1 and e_2 are the pre-defined thresholds for the corresponding errors.

(ATcT).[85] The TEs of the species in the G2-AE set are designated as the G2-TE test set, and the remaining 14 atoms from the first three rows of the main group elements form the TAE (total atomic energy) set.

To evaluate the performance of the ML-corrected DFA, we choose to examine a number of established datasets in the literature. These include a set of six AEs of platonic hydrocarbon cages C_nH_n ($n = 4, 6, 8, 10, 12, 20$) taken from the PlatonicTAE6 subset of the MGCDB84 set[86] (denoted as P6-AE) and a set of 19 AEs of alkanes ($n = 1 \ldots 8$) taken from the AlkAtom19 subset of the MGCDB84 set (denoted as Alk-AE). Due to the absence of accurate reference data for AEs in the G3-3 test set,[75,87] HOF of the G3-3 set are directly taken to form the G3-HOF test set. 84 IPs and 58 EAs are taken from the IP and EA subsets of the G2/97 set[88] to form the G2-IP and G2-EA test sets, respectively. 42 randomly selected BDEs form the BDE42 test set with the reference energies taken from experiments[89] and high precision calculations.[90] 99 BDEs are taken from the BDE99nonMR subset of the MGCDB84 set (denoted as BDE99) to further evaluate the performance on bond energy calculations. For the test of IE, 20 IEs are taken from the ISOMERIZATION20 subset of the MGCDB84 set to form the ISO20 test set, and another 8 IEs from the C20C24 subset of the MGCDB84 set form the ISO-C test set. In addition to thermochemical properties, we also evaluate the predictive power of ML-DFA on kinetic energies related to chemical reactions. 38 hydrogen transfer reaction barrier heights and 38 non-hydrogen transfer reaction barrier heights are taken from the HTBH38/08 and NHTBH38/08 sets[82,91,92] to form the HTBH and NHTBH test sets, respectively.

The density functional calculations in this work are performed, unless otherwise noted, by using the Python-based Simulations of Chemistry Framework (PySCF),[93–95] an open-source library for electronic structure calculations. The XC functional of ML-DFA is handled by the convenient customized interface of the XC functional in PySCF. During each step of the KS SCF calculation, the functional derivative of the ANN-presented functional correction can be computed on-the-fly, given the input of density descriptors using the automatic differentiation mechanism of PyTorch, and the resulting ML-corrected XC potential and energy density are fed to PySCF to continue the SCF calculation. Unless otherwise noted, the density functional calculations are carried out with the Def2-TZVPD basis set,[96–99] which is an extension of the standard Def2-TZVP basis set with the addition of diffuse functions.

The non-gradient optimization algorithm is implemented via an open source optimization library called Nevergrad.[100] Among all available optimization algorithms, the population-based PSO method is finally chosen because of its outstanding balance between simplicity and productivity. During the optimization, a stopping condition is added to implement the SCF convergence criterion mentioned in Sec. II C. The condition is invoked to exclude the weights of the ANN with which the SCF calculations performed by ML-DFA do not converge on any training species within a given number of iterations. While the condition might slow down the optimization of the ANN model, it can effectively guarantee an SCF-convergent model after training. The same strategy can be extended to include exact physical constraints in the optimization of ML models during the training of ML-DFA.

In this work, we investigate the ML correction to three DFAs: the PBE functional[65] and two types of B3LYP functional. The difference between the two types of widely used B3YLP functional lies in the use of different versions of the Vosko-Wilk-Nusair (VWN) correlation functional,[101] as reported in previous studies.[102,103] The default B3LYP functional in PySCF, GAMESS,[104] and some other quantum chemical programs adopts the VWN5[102] correlation functional parameterized from the values of calculations on the correlation energy of the uniform electron gas (UEG) as a basis for the fit and is hereafter denoted B3LYP; whereas, the default B3LYP functional in Gaussian[105] adopts the VWN3[102] correlation functional fitted with respect to the random phase approximation (RPA) and is hereafter denoted B3LYPG. The accuracy of the predictions by the two functionals has a small difference in relative energies but a relatively large difference in absolute energies. For the sake of clarity, the ML-corrected PBE functional will be referred to as ML-PBE, and the ML-corrected versions of the B3LYP and B3LYPG functional will be referred to as ML-B3LYP and ML-B3LYPG, respectively.

III. RESULTS AND DISCUSSION

A. Numerical performance of the ML-corrected functional

We first implement the ML correction on the PBE functional. Several trained models are selected as potential candidates, and their predictive power is evaluated on the test sets. It is found that these ML corrections provide only a marginal improvement in terms of accuracy over the original PBE functional for the prediction of various thermochemical and kinetic properties. In particular, given the hyperparameters experimented with, the ML-corrected PBE can only predict accurate AEs or TEs but cannot improve both simultaneously. This suggests that a more sophisticated parent DFA is probably needed to capture the main ingredients of the KS mapping so that the remaining contribution can be well described by ML-based correction.

We then impose the ML correction on the B3LYP and B3LYPG functionals, which belong to a higher rung of the "Jacob's ladder." The overall mean absolute errors (MAEs) of ML-B3LYP and ML-B3LYPG functionals on the test sets are exhibited in Fig. 2 in comparison with the original B3LYP and B3LYPG functionals. The displayed numbers are associated with the best ML model that yields the smallest training and validation MAEs. Since ML-B3LYP and ML-B3LYPG have similar numerical performance over the various test sets, the following discussion will focus mainly on the former for brevity.

As shown in Fig. 2, the original B3LYP functional gives a relatively large MAE on the G2-AE set. The MAE on the G3-HOF test set containing some larger molecules is around 10 kcal/mol, about twice as large compared to the MAE on the G2-AE set. With the ML-corrected B3LYP, the MAE reduces to about 3 kcal/mol on the G2-AE set, and an even more substantial reduction of about 70% in MAE is observed on the G3-HOF test set. Since conventional DFAs often produce appreciable errors for the prediction of thermochemical energies, especially the AE and HOF,[86] the ML correction effectively and consistently improves the accuracy and transferability of the original B3LYP functional on these properties for small- and medium-sized molecules. With the TEs of both molecules and atoms calculated with the same ML-corrected functional, the inconsistency issue in our previous work[33] regarding the calculation of AE or HOF is solved.

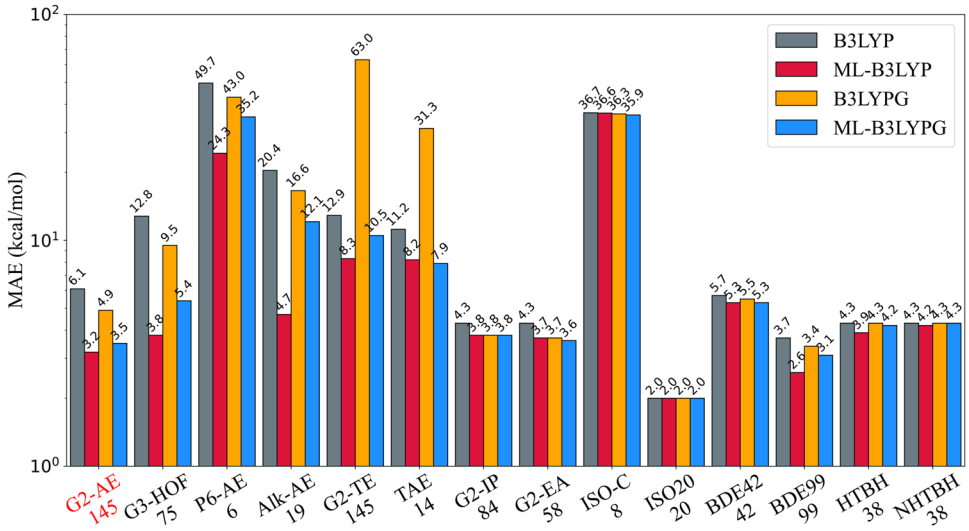

FIG. 2. Performance of ML-corrected B3LYP and B3LYPG functionals in comparison with their uncorrected counterparts. The number over each bar is the MAE of the corresponding dataset yielded by a particular DFA; see the main text for details. G2-AE set is taken as the validation set during the training process. The reference values for the TEs associated with the G2-TE and TAE sets are consistent with the AEs of the G2-AE set reported in Ref. 80. For computing HOF of the G3-HOF set, a scaling factor of 0.98 is applied to the thermal correction according to Refs. 79 and 87. The calculation of BDEs in the BDE42 set is based on geometries optimized at the level of B3LYP/6-31G(d) extracted from computational chemistry databases,[106–109] and the corresponding reference values for comparison are adopted from Refs. 89 and 90. The geometries and reference energies for the P6-AE, Alk-AE, BDE99, ISO20, and ISO-C sets are adopted from Ref. 86.

Regarding the absolute energies, both the ML-B3LYP and ML-B3LYPG functionals achieve a substantial improvement over their uncorrected counterparts. With the original B3LYP and B3LYPG, the MAEs on the G2-TE and TAE sets are larger than 10 kcal/mol. Particularly, the B3LYPG yields exceptionally large MAEs on the G2-TE and TAE sets. The prediction errors by ML-B3LYP on the molecular and atomic TEs in the G2-TE and TAE set have seen a 30% decrease from the errors with the original B3LYP, while the MAEs of B3LYPG are substantially reduced to 1/6 and 1/4 for the molecular and atomic TEs, respectively. The improvement in the prediction of TEs by the ML correction is consistent with the improved accuracy for the prediction of AEs. In principle, this implies that the improved performance is not only due to the cancellation of errors but also caused by a more accurate description of the KS mapping; whereas, in practice, this indicates that the enhanced transferability of the ML correction should allow for a more universal improvement on a wider range of species and properties.

We further evaluate the performance of ML-B3LYP on the P6-AE set, a set of organic molecules containing up to 20 non-hydrogen atoms, and the Alk-AE set, a set of 1–8 alkane AEs. For both sets, the ML-B3LYP functional effectively reduces the MAEs of 49.7 and 20.4 kcal/mol by the original B3LYP to 24.3 and 4.7 kcal/mol, respectively. The performance of the ML-B3LYP functional on all the AE-related test sets suggests that the ML correction can significantly enhance the accuracy of predicted AEs, and the inclusion of absolute energies in the training set is beneficial for the learning task.

Regarding the prediction of IP and EA, ML-B3LYP improves slightly over B3LYP; see Fig. 2. The results can be reasoned from two aspects. First, the MAEs produced by the original B3LYP functional on the G2-IP and G2-EA sets are around 4 kcal/mol, much smaller than those of the AE-related test sets. To minimize the loss function that collects all the errors at equivalent levels, the training process focuses primarily on suppressing the largest errors, i.e., those on the AEs. A more sophisticated ML model is probably needed to further reduce the prediction errors for IP and EA. Second, the prediction of IP and EA involves the calculation of charged species. However, the present training set only consists of neutral molecules and atoms. The inclusion of charged species into the training set and the loss function of the ML model will likely improve further the prediction on IP and EA.

Regarding the prediction of IEs, we assess the performance of ML-B3LYP on the ISO20 and ISO-C test sets, which have been explored in a previous benchmark study.[86] The ISO-C set is composed of the isomerization energies of C_{20} and C_{24} in their ground-state geometries, and the ISO20 test set involves molecules consisting of 1–4 non-hydrogen atoms. The ML-B3LYP basically preserves the same level of accuracy for the prediction of IEs on these two tests.

Regarding the prediction of chemical bonding energies, we examine two test sets, BDE42 and BDE99. The ML-B3LYP improves slightly over the original B3LYP functional. The BDE42 set is composed of species with C–C, C–O, and C–N bonds ranging from small molecules of a few atoms to relatively large organic compounds of 7–9 carbon atoms. The ML-B3LYP yields almost the same MAE as the original B3LYP. Particularly, the MAE on the smaller molecules in the BDE42 set is about 2 kcal/mol, while the MAE on the larger molecules with branched C–C bonds is about 9 kcal/mol. The somewhat more extensive BDE99 set contains 83 BDEs of small molecules with H–C, H–O, C–C, and several other types of bonds

between elements of the first three rows of the Periodic Table, for which the ML-B3LYP exhibits a minor improvement over the original B3LYP functional. Besides the accumulation of errors with the increasing size of the molecule, the minor improvement of BDE prediction is also ascribed to the lack of chemical information on radical species in the training set.

Regarding the prediction of reaction energy barriers, the ML-B3LYP has a marginal improvement over the original B3LYP functional. As shown in Fig. 2, the MAEs on the hydrogen and non-hydrogen reaction barrier heights by the ML-B3LYP functional are nearly the same as those by the original B3LYP functional. Similar to the case of IP, EA, and IE, the rather minor improvement in the accuracy of barrier height prediction may be ascribed to the lack of transition-state species in the training set.

In brief, the numerical results demonstrated in Fig. 2 verify the effectiveness of the ML correction to the two types of B3LYP functionals that are frequently employed in density functional calculations. In this work, the ML model only adopts the AEs and TEs of neutral chemical species to train the ANN model. The resulting ML-B3LYP and ML-B3LYPG functionals improve substantially the prediction of AEs (HOF) and TEs over the uncorrected counterparts for species beyond the training set but rather marginally on IPs, EAs, IEs, BDEs, and barrier heights. Among all the four functionals examined in Fig. 2, the B3LYP functional yields the largest errors on most of the test sets, while the ML-B3LYP functional has the overall best performance for the prediction of various energetic properties. Since the ML-B3LYP functional generally outperforms the original B3LYP, it is potentially applicable in density functional calculations as a safe and superior replacement for the latter. Therefore, the above findings indicate that a semilocal ML correction is practically useful in complementing an approximate KS mapping.

B. Further analysis of accuracy and efficiency

1. Error accumulation

Since many DFAs including the B3LYP suffer from the problem that the error increases with the size of the system,[27,87,110] we would like to find out to what extent the semilocal ML-correction addresses this general problem. As is shown in Fig. 3, the error in molecular AE or HOF calculated by the original B3LYP functional accumulates with the number of non-hydrogen atoms. The ML correction substantially mitigates the effect of error accumulation of B3LYP, as the regression coefficient of ML-B3LYP is 0.5 kcal/mol per atom, less than 1/4 of that of B3LYP. The reduction in error accumulation results from the systematic correction model trained in a SCF manner, which enables the ML-B3LYP to provide more accurate prediction for larger chemical systems.

2. Computational efficiency

It is important to test the computational efficiency of the ML-corrected DFA. In Fig. 4, two medium-sized organic molecules selected from the test sets are calculated by B3LYP and ML-B3LYP, and the convergence of total energy during the SCF process is depicted to compare the efficiency of the two functionals. Despite a slight delay in reaching the SCF convergence between ML-B3LYP and B3LYP, it usually requires only one or two more iterations for the ML-B3LYP to converge for various types of molecules because of the relatively small magnitudes of the ML correction. The extra computational effort for the evaluation of the ML correction is usually negligible compared to the other parts of the calculation and thus negligible in most cases.

We further evaluate the computational efficiency of the ML-B3LYP on 173 molecules with more than one non-hydrogen atom. Denote $\tilde{r} = t^{\mathrm{ML-B3LYP}}/t^{\mathrm{B3LYP}}$ as the relative computation time between the SCF calculations with the ML-B3LYP and B3LYP functionals. The histogram for the 173 molecules is depicted in Fig. 5. For most of the molecules (about 84%), the computational time with ML-B3LYP is similar to that with B3LYP ($1 \le \tilde{r} \le 2$). There are a few molecules for which $\tilde{r} > 2$. These large $\tilde{r}$ values are simply because the molecules are so small that the time for loading the ML model becomes discernible. There are some molecules (about 10%) that the calculation with ML-B3LYP is even faster than that with B3LYP ($\tilde{r} < 1$). For large molecules, calculations with ML-B3LYP are usually slightly slower than with B3LYP. For instance, ML-B3LYP takes 15% more time than B3LYP for the SCF calculation of the C_{60} fullerene molecule. Therefore, the ML-B3LYP functional can be conveniently implemented and efficiently applied in the same

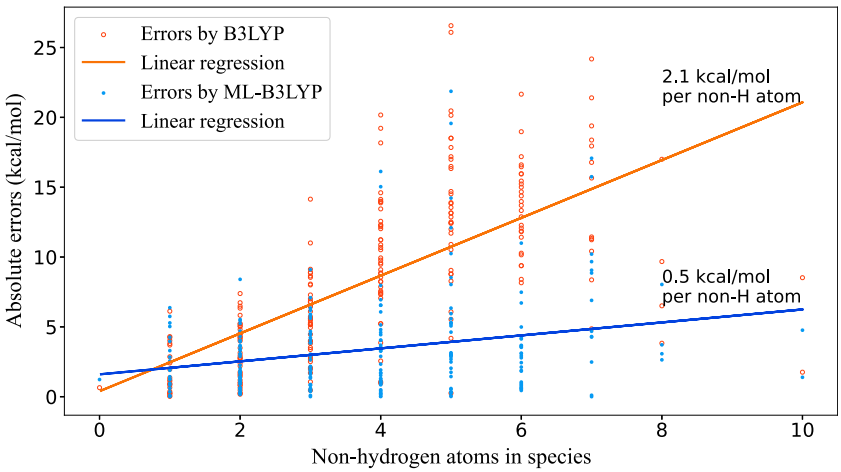

FIG. 3. Error in molecular AE or HOF vs the number of non-hydrogen atoms for molecular species in the G2-AE and G3-HOF set. The dots are the absolute errors and the lines are linear regressions for the errors of the two test sets by the ML-corrected and the original B3LYP functional. The regression coefficients are 2.1 and 0.5 kcal/mol per non-H atom for B3LYP and ML-B3LYP, respectively.

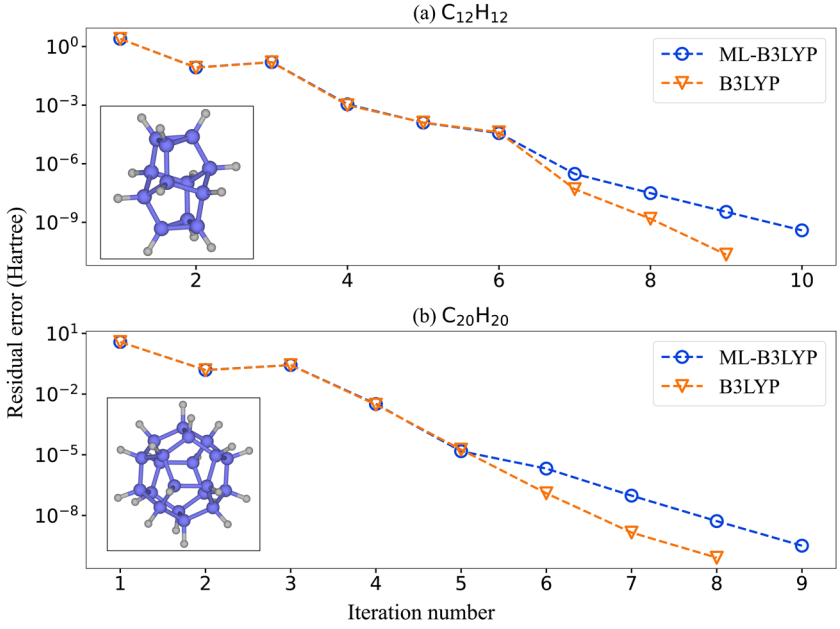

FIG. 4. Comparison of SCF convergence between the B3LYP and ML-B3LYP functionals for (a) octahedrane ($C_{12}H_{12}$) and (b) dodecahedrane ($C_{20}H_{20}$). The residual errors of total energies are plotted with respect to the SCF iteration numbers. The convergence criterion is set to 1×10^{-9} Hartree.

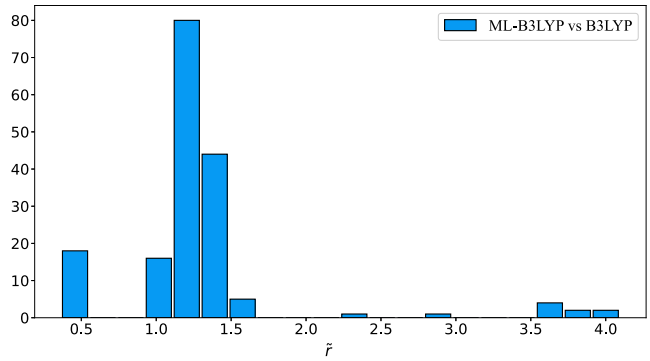

FIG. 5. A histogram of the relative computational time for SCF calculations by ML-B3LYP vs B3LYP over 173 molecules taken from the test set.

way as the original B3LYP functional, with no significant increase in computational efforts.

3. Accuracy of electron density

The electron density calculated by B3LYP and ML-B3LYP is also been examined. We take 60 small molecules from the test sets and calculate their electron densities on the same **r**-points by the B3LYP, ML-B3LYP, and CCSD(T) methods with the same basis set. By taking the numerical outcomes of CCSD(T) as the reference data, the MAE of the electron density calculated by B3LYP or ML-B3LYP is evaluated via

$$D^{\mathrm{DFA}} = \frac{1}{N}\sum_{i=1}^{N}\left|\rho^{\mathrm{CCSD(T)}}(\mathbf{r}_i) - \rho^{\mathrm{DFA}}(\mathbf{r}_i)\right|, \tag{8}$$

where $\rho^{\mathrm{DFA}}(\mathbf{r}_i)$ is the electron density of a species at $\mathbf{r}_i$ calculated by DFA, and N is the total number of **r**-points. Define the relative error as

$$\Delta\tilde{D} = \frac{D^{\mathrm{ML-B3LYP}} - D^{\mathrm{B3LYP}}}{D^{\mathrm{B3LYP}}}. \tag{9}$$

A negative $\Delta\tilde{D}$ means that the electron density calculated by ML-B3LYP is closer to that of CCSD(T) than by B3LYP and thus more accurate.

Figure 6 depicts the histogram on $\Delta\tilde{D}$ of 60 molecules taken from the test sets. Apparently, most of the molecules fall within the range of $-0.2 \le \Delta\tilde{D} \le 0.2$, which means that the electron density calculated by ML-B3LYP does not deviate significantly from that by the original B3LYP functional. More importantly, the majority of molecules have a negative $\Delta\tilde{D}$, indicating a minor improvement of electron density by ML-B3LYP over the original B3LYP. Such an improvement is a direct consequence of the small yet effective change in $\epsilon_{\mathrm{XC}}^{\mathrm{DFA}}(\mathbf{r})$ by the ML correction, which is eventually consistent with the correction scheme proposed in Eq. (1). Since the ML correction is implemented in an SCF manner, a small correction to the original XC energy density is expected to yield an effective improvement to the total energy of a chemical system, with the electron density corrected self-consistently.

4. Magnitude of correction

To assess the magnitude of the ML correction, a comparison is made between $\epsilon_{\mathrm{XC}}^{\mathrm{B3LYP}}(\mathbf{r})$ and $\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}(\mathbf{r})$ resulted from the SCF calculations over the G2-AE set. It is found that $|\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}(\mathbf{r})/\epsilon_{\mathrm{XC}}^{\mathrm{B3LYP}}(\mathbf{r})| < 0.1$ for 99.5% of the **r**-points and $|\Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}(\mathbf{r})/\epsilon_{\mathrm{XC}}^{\mathrm{B3LYP}}(\mathbf{r})| < 0.01$ for 98.2% of the **r**-points; see the supplementary material for details. Such a comparison suggests that, although the constructed semilocal correction has a relatively small magnitude, it is sufficient to improve the accuracy of the parent DFA.

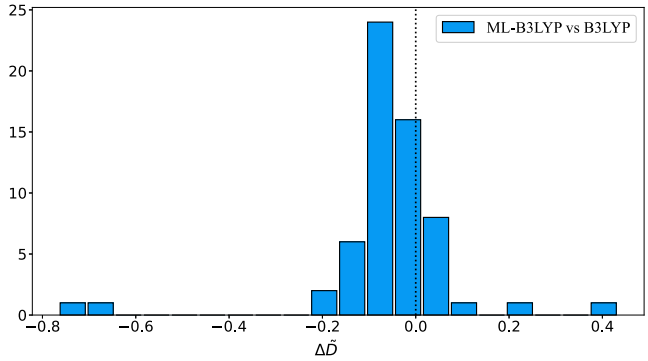

FIG. 6. A histogram on the relative error in electron density by ML-B3LYP vs B3YLP for 60 molecules with the result of CCSD(T) taken as the reference. The dotted line is $\Delta\tilde{D} = 0$. A negative $\Delta\tilde{D}$ means that the electron density calculated by ML-B3LYP is closer to the result of CCSD(T) and hence deemed as more accurate, whereas a positive value indicates that the density by ML-B3LYP deviates further away from the prediction of CCSD(T) and thus less accurate.

5. Influence of integration grids

As the integration grids sometimes affect the SCF energies by DFAs, ML-B3LYP is assessed with different levels of grids on the G2-AE set. The grids are chosen to be at level 4, (99, 590) and level 5 in PySCF, where (99, 590) is referred to as the UltraFine grid level in Gaussian, and the grids at levels 4 and 5 are sparser and denser than (99, 590), respectively. The output energies and the MAEs on the G2-AE set by ML-B3LYP are the same for all the grids tested, suggesting that the default integration grid is sufficient for accurate calculations by ML-B3LYP.

6. Basis set dependence

The ML-B3LYP and B3LYP functionals are tested in this work on the entire dataset with all levels of basis sets in the Def2 family. The consistent results indicate a weak dependence of the results on the choice of basis set, provided that the basis set has at least a medium size.

IV. CONCLUDING REMARKS

This work has developed a correction scheme for improving the performance of DFAs with an ML model that is built from semilocal density descriptors. The correction is constructed with an ANN model whose parameters are determined by the non-gradient global optimization of the loss function consisting of AEs and TEs for several chemical species. The resulting ML-corrected B3LYP functional yields more accurate predictions for both the energy difference properties and the absolute energies of general chemical species than the original B3LYP. In general, ML-B3LYP provides a more accurate and more balanced performance of calculations on thermochemical and reaction kinetic properties than the original B3LYP, especially for TE and AE, and slightly improves the accuracy of prediction for IP, EA, and BDE. Implemented on the PySCF platform as a customized density functional, ML-B3LYP can be immediately applied to general chemical species the same way as the original B3LYP functional, without a significant increase in computational effort.

The introduction of absolute energy for training the ML correction model contributes to a more stable optimization and leads to robust models with consistently more accurate predictions of both chemical energies and electron density compared to the parent functional. With the modest semilocal correction of the XC energy density, improved and more balanced performance of DFAs for chemical predictions with SCF calculations can be safely achieved without significantly degrading compliance with strict physical conditions. The ML-B3LYP built in this work is an enhanced B3LYP functional and can be a favorable replacement for both types of B3LYP functionals commonly used in electronic structure calculations. The computational efficiency of the SCF convergence for the original DFA is effectively preserved by the ML-DFA.

In summary, this work explores the possibility of developing "a functional that performs uniformly better than B3LYP,"[74] and presents the ML-corrected functional that generally improves the predictive accuracy of B3LYP on various properties. The results demonstrate the effectiveness of combining ML corrections with DFAs for better performance. For now, the predictions by the ML-corrected functional are not within the chemical accuracy and will be promoted in further studies.

Several aspects of this work need to be further explored. First, the current ML correction can be modified and developed to extend its applicability to more types of DFAs and various chemical systems by considering exact physical constraints. The physical constraints can be integrated into the training process, similar to the control of the SCF convergence for the ML-DFA during the optimization, to guide the correction model of the ML-DFA toward chemical accuracy. Meanwhile, a more complete and balanced training set can be built to include the chemical environment of ionic and radical species as well as species in non-equilibrium geometries. A systematic protocol is to be established to determine the optimal composition of the training set so that the set is adequate for training a correction model with high accuracy and broad transferability. Moreover, Yaron and co-workers[111,112] have successfully integrated quantum chemistry calculation into an ANN for the prediction of molecular properties. This was achieved by incorporating the density functional tight binding (DFTB) theory as a layer in the network and combining the back propagation of the DFTB Hamiltonian and network parameters with the SCF calculation for the charge distribution to train the ANN. Their findings suggest that it is possible to design a practical approach to optimize the ML-corrected KS mapping by back propagation and gradient descent algorithms through an SCF procedure, which will be further explored. Furthermore, the applicability of ML-DFA can be extended to periodic solids. In future work, we will explore an appropriate way to integrate the information on non-equilibrium geometries into the training process of ML-DFA to improve its predictive power on barrier height and other kinetic properties. It will also be attempted to further improve the predictions by the ML-DFA of the binding energies through a proper introduction of the radical reference data in the training process of the functional. Finally, similar to the ML correction term defined in this work, an ML correction term to capture non-covalent interactions can be added to the XC energy density of DFAs to correct the predictions of the original DFAs on non-covalent interactions, possibly aided by a proper design of non-local density descriptors.

SUPPLEMENTAL MATERIAL

See the supplementary material is available free of charge at:

- SM.pdf: Details about the datasets and numerical performance of the ML-corrected functionals.
- The code for performing density functional calculations with the proposed ML-B3LYP functional with PySCF can be obtained from https://github.com/beckhamwjc/ML-DFA.

ACKNOWLEDGMENTS

Support from the National Natural Science Foundation of China (Grant Nos. 21973086, 22103073, and 22173088) and the Hong Kong Quantum AI Lab, AIR@InnoHK of the Hong Kong Government, is gratefully acknowledged. The authors thank Professor Igor Ying Zhang, Professor Sai Duan, and Mr. ShiRong Wang at Fudan University for the fruitful discussion. J. Wang is grateful for the guidance on numerical aspects from Dr. Jie Liu at USTC and Dr. Tian Lu and colleagues at Beijing Kein Research Center for Natural Sciences.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

JingChun Wang: Investigation (lead); Methodology (equal); Writing – original draft (lead); Writing – review & editing (supporting). **Yao Wang**: Investigation (supporting); Methodology (supporting); Supervision (supporting); Writing – review & editing (supporting). **Rui-Xue Xu**: Funding acquisition (supporting); Resources (lead); Supervision (supporting); Writing – review & editing (supporting). **GuanHua Chen**: Conceptualization (lead); Methodology (supporting); Resources (supporting); Writing – review & editing (supporting). **Xiao Zheng**: Funding acquisition (lead); Methodology (equal); Supervision (lead); Writing – review & editing (lead).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: DERIVATION OF ML-CORRECTED XC POTENTIAL

According to the definitions in Eqs. (1) and (3), the functional derivative of $\Delta v^{\mathrm{ML}}_{\mathrm{XC};\sigma}$ is calculated in the same manner as the XC potential of GGA functionals,[1,2]

$$\Delta v^{\mathrm{ML}}_{\mathrm{XC};\sigma}(\mathbf{r}) = \frac{\delta \int \mathrm{d}\mathbf{r} f^{\mathrm{ML}}(r_s,s,\zeta)}{\delta\rho_\sigma} = \frac{\partial f^{\mathrm{ML}}}{\partial\rho_\sigma} - \nabla\cdot\frac{\partial f^{\mathrm{ML}}}{\partial\nabla\rho_\sigma}, \tag{A1}$$

where $f^{\mathrm{ML}} = \rho\Delta\epsilon^{\mathrm{ML}}_{\mathrm{XC}}$. According to the chain rule for functional differentiation, the first part of the derivative in Eq. (A1) is directly calculated, given the fact that the arguments of f^{ML}, $\{r_s,s,\zeta\}$, are all functions of the electron density ρ,

$$\begin{aligned}\frac{\partial f^{\mathrm{ML}}}{\partial\rho_\sigma} &= \frac{\partial f^{\mathrm{ML}}}{\partial r_s}\frac{\partial r_s}{\partial\rho_\sigma} + \frac{\partial f^{\mathrm{ML}}}{\partial s}\frac{\partial s}{\partial\rho_\sigma} + \frac{\partial f^{\mathrm{ML}}}{\partial\zeta}\frac{\partial\zeta}{\partial\rho_\sigma}\\ &= \frac{\partial f^{\mathrm{ML}}}{\partial r_s}\left(-\frac{4\pi}{9}\right)r_s^4 + \frac{\partial f^{\mathrm{ML}}}{\partial s}\left(-\frac{s}{\pi r_s^3}\right)\\ &\quad + \frac{\partial f^{\mathrm{ML}}}{\partial\zeta}\left(\frac{4\pi}{3}\right)r_s^3\left[\mathrm{sgn}\,(\sigma)-\zeta\right].\end{aligned} \tag{A2}$$

According to the chain rule for function differentiation, the second part of the derivative in Eq. (A1) is directly calculated, given the fact that only s is an explicit function of the density derivative $\nabla\rho_\sigma$ while r_s and ζ are not,

$$\begin{aligned}\nabla\cdot\frac{\partial f^{\mathrm{ML}}}{\partial\nabla\rho_\sigma} &= \nabla\cdot\left(\frac{\partial f^{\mathrm{ML}}}{\partial s}\frac{\partial s}{\partial\nabla\rho_\sigma}\right)\\ &= \nabla\cdot\left(\frac{\partial f^{\mathrm{ML}}}{\partial s}\frac{3^{14/3}s^3}{16\pi^{5/3}r_s^{12}}\nabla r_s\right).\end{aligned} \tag{A3}$$

Therefore, we have

$$\begin{aligned}\Delta v^{\mathrm{ML}}_{\mathrm{XC};\sigma}(\mathbf{r}) &= \frac{\partial f^{\mathrm{ML}}}{\partial r_s}\left(-\frac{4\pi}{9}\right)r_s^4 + \frac{\partial f^{\mathrm{ML}}}{\partial s}\left(-\frac{s}{\pi r_s^3}\right)\\ &\quad + \frac{\partial f^{\mathrm{ML}}}{\partial\zeta}\left(\frac{4\pi}{3}\right)r_s^3\left[\mathrm{sgn}\,(\sigma)-\zeta\right]\\ &\quad + \nabla\cdot\left(\frac{\partial f^{\mathrm{ML}}}{\partial s}\frac{3^{14/3}s^3}{16\pi^{5/3}r_s^{12}}\nabla r_s\right),\end{aligned} \tag{A4}$$

as in Eq. (4).

REFERENCES

[1] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, International Series of Monographs on Chemistry (Oxford University Press, New York, 1995).
[2] E. Engel and R. M. Dreizler, *Density Functional Theory: An Advanced Course* (Springer Berlin Heidelberg, Berlin, Germany, 2013).
[3] R. O. Jones, "Density functional theory: Its origins, rise to prominence, and future," Rev. Mod. Phys. **87**, 897 (2015).
[4] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. **140**, A1133 (1965).
[5] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, "Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids," Phys. Rev. Lett. **91**, 146401 (2003).
[6] J. Sun, A. Ruzsinszky, and J. P. Perdew, "Strongly constrained and appropriately normed semilocal density functional," Phys. Rev. Lett. **115**, 036402 (2015).
[7] A. D. Becke, "Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals," J. Chem. Phys. **107**, 8554 (1997).
[8] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, "Development and assessment of new exchange-correlation functionals," J. Chem. Phys. **109**, 6264 (1998).
[9] Y. Zhao and D. G. Truhlar, "A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions," J. Chem. Phys. **125**, 194101 (2006).

[10]J. P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy," AIP Conf. Proc. **577**, 1 (2001).
[11]C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
[12]S. Marsland, *Machine Learning: An Algorithmic Perspective*, 2nd ed. (Chapman & Hall/CRC, 2014).
[13]M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science **349**, 255 (2015).
[14]B. Kalita, L. Li, R. J. McCarty, and K. Burke, "Learning to approximate density functionals," Acc. Chem. Res. **54**, 818 (2021).
[15]J. Wu, G. Chen, J. Wang, and X. Zheng, "Chapter 23—Redesigning density functional theory with machine learning," in *Quantum Chemistry in the Age of Machine Learning*, edited by P. O. Dral (Elsevier, 2023), pp. 531–558.
[16]D. J. Tozer, V. E. Ingamells, and N. C. Handy, "Exchange-correlation potentials," J. Chem. Phys. **105**, 9200 (1996).
[17]S. S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. (Pearson Education, Upper Saddle River, NJ, 2009).
[18]X. Zheng, L. Hu, X. Wang, and G. Chen, "A generalized exchange-correlation functional: The neural-networks approach," Chem. Phys. Lett. **390**, 186 (2004).
[19]Q. Liu, J. Wang, P. Du, L. Hu, X. Zheng, and G. Chen, "Improving the performance of long-range-corrected exchange-correlation functional with an embedded neural network," J. Phys. Chem. A **121**, 7273 (2017).
[20]J. Wu and X. Xu, "The X1 method for accurate and efficient prediction of heats of formation," J. Chem. Phys. **127**, 214105 (2007).
[21]J. Wu and X. Xu, "Improving the B3LYP bond energies by using the X1 method," J. Chem. Phys. **129**, 164103 (2008).
[22]J. Wu, I. Ying Zhang, and X. Xu, "The X1s method for accurate bond dissociation energies," ChemPhysChem **11**, 2561 (2010).
[23]J. Wu and X. Xu, "Accurate prediction of heats of formation by a combined method of B3LYP and neural network correction," J. Comput. Chem. **30**, 1424 (2009).
[24]Y. Zhou, J. Wu, and X. Xu, "Improving B3LYP heats of formation with three-dimensional molecular descriptors," J. Comput. Chem. **37**, 1175 (2016).
[25]X. Wang, L. Hu, L. Wong, and G. Chen, "A combined first-principles calculation and neural networks correction approach for evaluating Gibbs energy of formation," Mol. Simul. **30**, 9 (2004).
[26]X.-M. Duan, G.-L. Song, Z.-H. Li, X.-J. Wang, G.-H. Chen, and K.-N. Fan, "Accurate prediction of heat of formation by combining Hartree–Fock/density functional theory calculation with linear regression correction approach," J. Chem. Phys. **121**, 7086 (2004).
[27]L. Hu, X. Wang, L. Wong, and G. Chen, "Combined first-principles calculation and neural-network correction approach for heat of formation," J. Chem. Phys. **119**, 11501 (2003).
[28]W. Ting-Ting, L. Wen-Long, C. Zhang-Hui, and M. Ling, "Correcting the systematic error of the density functional theory calculation: The alternate combination approach of genetic algorithm and neural network," Chin. Phys. B **19**, 076401 (2010).
[29]H. Li, L. Shi, M. Zhang, Z. Su, X. Wang, L. Hu, and G. Chen, "Improving the accuracy of density-functional theory calculation: The genetic algorithm and neural network approach," J. Chem. Phys. **126**, 144101 (2007).
[30]X. Wang, L. Wong, L. Hu, C. Chan, Z. Su, and G. Chen, "Improving the accuracy of density-functional theory calculation: The statistical correction approach," J. Phys. Chem. A **108**, 8514 (2004).
[31]J. Sun, J. Wu, T. Song, L. Hu, K. Shan, and G. Chen, "Alternative approach to chemical accuracy: A neural networks-based first-principles method for heat of formation of molecules made of H, C, N, O, F, S, and Cl," J. Phys. Chem. A **118**, 9120 (2014).
[32]A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, "Machine-learning approach for one- and two-body corrections to density functional theory: Applications to molecular and condensed water," Phys. Rev. B **88**, 054104 (2013).
[33]J. Wang, D. Zhang, R.-X. Xu, C. Yam, G. Chen, and X. Zheng, "Improving density functional prediction of molecular thermochemical properties with a machine-learning-corrected generalized gradient approximation," J. Phys. Chem. A **126**, 970 (2022).
[34]S.-C. Lin and M. Oettel, "A classical density functional from machine learning and a convolutional neural network," SciPost Phys. **6**, 25 (2019).
[35]F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, "Bypassing the Kohn-Sham equations with machine learning," Nat. Commun. **8**, 872 (2017).
[36]R. Nagai, R. Akashi, S. Sasaki, and S. Tsuneyuki, "Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability," J. Chem. Phys. **148**, 241737 (2018).
[37]R. Nagai, R. Akashi, and O. Sugino, "Completing density functional theory by machine learning hidden messages from molecules," npj Comput. Mater. **6**, 43 (2020).
[38]R. Nagai, R. Akashi, and O. Sugino, "Machine-learning-based exchange correlation functional with physical asymptotic constraints," Phys. Rev. Res. **4**, 013106 (2022).
[39]Y. Chen, L. Zhang, H. Wang, and W. E, "DeePKS: A comprehensive data-driven approach toward chemically accurate density functional theory," J. Chem. Theory Comput. **17**, 170 (2021).
[40]S. Dick and M. Fernandez-Serra, "Machine learning accurate exchange and correlation functionals of the electronic density," Nat. Commun. **11**, 3509 (2020).
[41]J. Schmidt, C. L. Benavides-Riveros, and M. A. L. Marques, "Machine learning the physical nonlocal exchange–correlation functional of density-functional theory," J. Phys. Chem. Lett. **10**, 6425 (2019).
[42]J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K.-R. Müller, and K. Burke, "Orbital-free bond breaking via machine learning," J. Chem. Phys. **139**, 224104 (2013).
[43]M. Fujinami, R. Kageyama, J. Seino, Y. Ikabata, and H. Nakai, "Orbital-free density functional theory calculation applying semi-local machine-learned kinetic energy density functional and kinetic potential," Chem. Phys. Lett. **748**, 137358 (2020).
[44]R. A. Vargas-Hernández, "Bayesian optimization for calibrating and selecting hybrid-density functional models," J. Phys. Chem. A **124**, 4053 (2020).
[45]J. T. Margraf and K. Reuter, "Pure non-local machine-learned density functional theory for electron correlation," Nat. Commun. **12**, 344 (2021).
[46]J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, "Semi-local machine-learned kinetic energy density functional with third-order gradients of electron density," J. Chem. Phys. **148**, 241705 (2018).
[47]Y. Zhou, J. Wu, S. Chen, and G. Chen, "Toward the exact exchange–correlation potential: A three-dimensional convolutional neural network construct," J. Phys. Chem. Lett. **10**, 7264 (2019).
[48]L. Li, J. C. Snyder, I. M. Pelaschier, J. Huang, U.-N. Niranjan, P. Duncan, M. Rupp, K.-R. Müller, and K. Burke, "Understanding machine-learned density functionals," Int. J. Quantum Chem. **116**, 819 (2016).
[49]K. Ryczko, D. A. Strubbe, and I. Tamblyn, "Deep learning and density-functional theory," Phys. Rev. A **100**, 022512 (2019).
[50]J. R. Moreno, G. Carleo, and A. Georges, "Deep learning the Hohenberg-Kohn maps of density functional theory," Phys. Rev. Lett. **125**, 076402 (2020).
[51]X. Lei and A. J. Medford, "Design and analysis of machine learning exchange-correlation functionals via rotationally invariant convolutional descriptors," Phys. Rev. Mater. **3**, 063801 (2019).
[52]M. Michael Denner, M. H. Fischer, and T. Neupert, "Efficient learning of a one-dimensional density functional theory," Phys. Rev. Res. **2**, 033388 (2020).
[53]J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, "Finding density functionals with machine learning," Phys. Rev. Lett. **108**, 253002 (2012).
[54]L. Li, T. E. Baker, S. R. White, and K. Burke, "Pure density functional for strong correlation and the thermodynamic limit from machine learning," Phys. Rev. B **94**, 245129 (2016).
[55]J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, and A. J. Cohen, "Pushing the frontiers of density functionals by solving the fractional electron problem," Science **374**, 1385 (2021).
[56]K. Pokharel, J. W. Furness, Y. Yao, V. Blum, T. J. P. Irons, A. M. Teale, and J. Sun, "Exact constraints and appropriate norms in machine-learned exchange-correlation functionals," J. Chem. Phys. **157**, 174106 (2022).

[57]H. Bhattacharjee, N. Anesiadis, and D. G. Vlachos, "Regularized machine learning on molecular graph model explains systematic error in DFT enthalpies," Sci. Rep. **11**, 14372 (2021).
[58]J. Gedeon, J. Schmidt, M. J. P. Hodgson, J. Wetherell, C. L. Benavides-Riveros, and M. A. L. Marques, "Machine learning the derivative discontinuity of density-functional theory," Mach. Learn.: Sci. Technol. **3**, 015011 (2021).
[59]R. Meyer, M. Weichselbaum, and A. W. Hauser, "Machine learning approaches toward orbital-free density functional theory: Simultaneous training on the kinetic energy density functional and its functional derivative," J. Chem. Theory Comput. **16**, 5685 (2020).
[60]M. Alghadeer, A. Al-Aswad, and F. H. Alharbi, "Highly accurate machine learning model for kinetic energy density functional," Phys. Lett. A **414**, 127621 (2021).
[61]E. Cuierrier, P.-O. Roy, and M. Ernzerhof, "Constructing and representing exchange–correlation holes through artificial neural networks," J. Chem. Phys. **155**, 174121 (2021).
[62]L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, "Kohn-Sham equations as regularizer: Building prior knowledge into machine-learned physics," Phys. Rev. Lett. **126**, 036401 (2021).
[63]P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, and P. O. Dral, "Artificial intelligence-enhanced quantum chemical method with broad applicability," Nat. Commun. **12**, 7022 (2021).
[64]Y. Liu, C. Zhang, Z. Liu, D. G. Truhlar, Y. Wang, and X. He, "Supervised learning of a chemistry functional with damped dispersion," Nat. Comput. Sci. **3**, 48 (2023).
[65]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865 (1996).
[66]A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," Phys. Rev. A **38**, 3098 (1988).
[67]C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," Phys. Rev. B **37**, 785 (1988).
[68]A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," J. Chem. Phys. **98**, 5648 (1993).
[69]S. Grimme, "Semiempirical hybrid density functional with perturbative second-order correlation," J. Chem. Phys. **124**, 034108 (2006).
[70]N. Mardirossian and M. Head-Gordon, "Survival of the most transferable at the top of Jacob's ladder: Defining and testing the ωB97M(2) double hybrid density functional," J. Chem. Phys. **148**, 241736 (2018).
[71]N. Q. Su, Z. Zhu, and X. Xu, "Doubly hybrid density functionals that correctly describe both density and energy for atoms," Proc. Natl. Acad. Sci. U. S. A. **115**, 2287 (2018).
[72]I. Y. Zhang and X. Xu, "Exploring the limits of the XYG3-type doubly hybrid approximations for the main-group chemistry: The xDH@B3LYP model," J. Phys. Chem. Lett. **12**, 2638 (2021).
[73]Y. Wang, Y. Li, J. Chen, I. Y. Zhang, and X. Xu, "Doubly hybrid functionals close to chemical accuracy for both finite and extended systems: Implementation and test of XYG3 and XYGJ-OS," JACS Au **1**, 543 (2021).
[74]A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Challenges for density functional theory," Chem. Rev. **112**, 289 (2012).
[75]T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'16* (Association for Computing Machinery, New York, 2016), pp. 785–794.
[76]A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), pp. 8024–8035.
[77]R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science* (IEEE, 1995), pp. 39–43.
[78]R. Eberhart and Y. Shi, "Comparing inertia weights and constriction factors in particle swarm optimization," in *Proceedings of the 2000 Congress on Evolutionary Computation CEC00 (Cat. No. 00TH8512)* (IEEE, 2000), Vol. 1, pp. 84–88.
[79]L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, "Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation," J. Chem. Phys. **106**, 1063 (1997).
[80]R. Haunschild and W. Klopper, "New accurate reference energies for the G2/97 test set," J. Chem. Phys. **136**, 164102 (2012).
[81]P. Verma and D. G. Truhlar, "Status and challenges of density functional theory," Trends Chem. **2**, 302 (2020).
[82]R. Peverati and D. G. Truhlar, "Quest for a universal density functional: The accuracy of density functionals across a broad spectrum of databases in chemistry and physics," Philos. Trans. R. Soc. London, Ser. A **372**, 20120476 (2014).
[83]E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang, "Revealing noncovalent interactions," J. Am. Chem. Soc. **132**, 6498 (2010).
[84]W. Klopper, R. A. Bachorz, C. Hättig, and D. P. Tew, "Accurate computational thermochemistry from explicitly correlated coupled-cluster theory," Theor. Chem. Acc. **126**, 289 (2010).
[85]W. Klopper, B. Ruscic, D. P. Tew, F. A. Bischoff, and S. Wolfsegger, "Atomization energies from coupled-cluster calculations augmented with explicitly-correlated perturbation theory," Chem. Phys. **356**, 14 (2009), part of the Special Issue: Moving Frontiers in Quantum Chemistry.
[86]N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals," Mol. Phys. **115**, 2315 (2017).
[87]L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, "Assessment of Gaussian-3 and density functional theories for a larger experimental test set," J. Chem. Phys. **112**, 7374 (2000).
[88]L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, "Assessment of Gaussian-2 and density functional theories for the computation of ionization potentials and electron affinities," J. Chem. Phys. **109**, 42 (1998).
[89]Y.-R. Luo, *Handbook of Bond Dissociation Energies in Organic Compounds* (CRC Press, Boca Raton, 2003).
[90]V. K. Prasad, M. H. Khalilian, A. Otero-de-la-Roza, and G. A. DiLabio, "BSE49, a diverse, high-quality benchmark dataset of separation energies of chemical bonds," Sci. Data **8**, 300 (2021).
[91]J. Zheng, Y. Zhao, and D. G. Truhlar, "The DBH24/08 database and its use to assess electronic structure model chemistries for chemical reaction barrier heights," J. Chem. Theory Comput. **5**, 808 (2009).
[92]Y. Zhao, N. E. Schultz, and D. G. Truhlar, "Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions," J. Chem. Theory Comput. **2**, 364 (2006).
[93]Q. Sun, "Libcint: An efficient general integral library for Gaussian basis functions," J. Comput. Chem. **36**, 1664 (2015).
[94]Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, "PySCF: The Python-based simulations of chemistry framework," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1340 (2018).
[95]Q. Sun *et al.*, "Recent developments in the PySCF program package," J. Chem. Phys. **153**, 024109 (2020).
[96]F. Weigend, "Accurate Coulomb-fitting basis sets for H to Rn," Phys. Chem. Chem. Phys. **8**, 1057 (2006).
[97]F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[98]D. Rappoport and F. Furche, "Property-optimized Gaussian basis sets for molecular response calculations," J. Chem. Phys. **133**, 134105 (2010).
[99]L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, "A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions," Phys. Chem. Chem. Phys. **19**, 32184 (2017).
[100]J. Rapin and O. Teytaud, Nevergrad—A gradient-free optimization platform, https://GitHub.com/FacebookResearch/Nevergrad, 2018.

[101]S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: A critical analysis," Can. J. Phys. **58**, 1200 (1980).
[102]R. H. Hertwig and W. Koch, "On the parameterization of the local correlation functional. What is Becke-3-LYP?," Chem. Phys. Lett. **268**, 345 (1997).
[103]B. Chan, P. M. W. Gill, and L. Radom, "Performance of gradient-corrected and hybrid density functional theory: Role of the underlying local density approximation and the gradient correction," J. Chem. Theory Comput. **8**, 4899 (2012).
[104]G. M. J. Barca *et al.*, "Recent developments in the general atomic and molecular electronic structure system," J. Chem. Phys. **152**, 154102 (2020).
[105]M. J. Frisch *et al.*, Gaussian 16, Revision C.01, Gaussian, Inc., Wallingford, CT, 2016.
[106]R. D. Johnson III, NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, https://cccbdb.nist.gov/, 2020; accessed 01 April 2021.
[107]N. Maho, "The PubChemQC project: A large chemical database from the first principle calculations," AIP Conf. Proc. **1702**, 090058 (2015).
[108]M. Nakata and T. Shimazaki, "PubChemQC project: A large-scale first-principles electronic structure database for data-driven chemistry," J. Chem. Inf. Model. **57**, 1300 (2017).
[109]M. Nakata, T. Shimazaki, M. Hashimoto, and T. Maeda, "PubChemQC PM6: Data sets of 221 million molecules with optimized molecular geometries and electronic properties," J. Chem. Inf. Model. **60**, 5891 (2020).
[110]M. D. Wodrich, C. Corminboeuf, and P. v. R. Schleyer, "Systematic errors in computed alkane energies using B3LYP and other popular DFT functionals," Org. Lett. **8**, 3631 (2006).
[111]H. Li, C. Collins, M. Tanha, G. J. Gordon, and D. J. Yaron, "A density functional tight binding layer for deep learning of chemical Hamiltonians," J. Chem. Theory Comput. **14**, 5764 (2018).
[112]F. Hu, F. He, and D. J. Yaron, "Semiempirical Hamiltonians learned from data can have accuracy comparable to density functional theory," arXiv.2210.11682v2.