Improving Density Functional Prediction of Molecular Thermochemical Properties with a Machine-Learning-Corrected Generalized Gradient Approximation

JingChun Wang,[1] DaDi Zhang,[2] Rui-Xue Xu,[1] ChiYung Yam,[3] GuanHua Chen,[4] and Xiao Zheng[1,*]

[1] *Hefei National Laboratory for Physical Sciences at the Microscale & Synergetic Innovation Center of Quantum Information and Quantum Physics & CAS Center for Excellence in Nanoscience, University of Science and Technology of China, Hefei, Anhui 230026, China*

[2] *National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei, China*

[3] *Beijing Computational Science Research Center, 100084 Beijing, P. R. China*

[4] *Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong, China*

(Dated: May 14, 2021)

The past decade has seen an increasing interest in designing sophisticated density functional approximations (DFAs) by integrating the power of machine learning (ML) techniques. However, application of the ML-based DFAs is often confined to simple model systems. In this work, we construct an ML correction to the widely used Perdew-Burke-Ernzerhof (PBE) functional by establishing a semilocal mapping from the electron density and reduced gradient to the exchange-correlation energy density. The resulting ML-corrected PBE is immediately applicable to any real molecule, and yields significantly improved heats of formation while preserving the accuracy for other thermochemical and kinetic properties. This work highlights the prospect of combining the power of data-driven ML methods with physics-inspired derivations for reaching the heaven of chemical accuracy.

I. INTRODUCTION

Density functional theory (DFT) has achieved enormous success in physics, chemistry, biology and material science as an efficient tool to study electronic structures and reactions in molecules, condensed phases and extended many-body systems since 1990s [1]. The success goes to the modern paradigm of DFT which consists of the Hohenberg-Kohn (HK) theorem [2] and the Kohn-Sham (KS) formalism [3], with the former offering a one-to-one mapping between ground-state electron density and external potential and the later providing a practical approach to finding the ground-state energy by adopting certain approximation for the exchange-correlation (XC) functional.

A great variety of density functional approximations (DFAs) have been proposed. Diversified strategies have been employed, such as the analytic analysis on simple physical models, the explicit imposition of exact physical constraints or conditions, the adiabatic connection [4] between the reference noninteracting system and real interacting system, and the use of semiempirical or empirical parameters whose values are determined by optimizing the numerical accuracy of density functional calculations [5–12]. Despite the progress made, the chemical accuracy has not been achieved universally within the framework of KS-DFT. This is partly because the mathematical representation of the KS mapping, $\rho(\mathbf{r}) \mapsto v_{\mathrm{XC}}(\mathbf{r})$, with $\rho(\mathbf{r})$ and $v_{\mathrm{XC}}(\mathbf{r})$ being the electron density and XC potential, is not sophisticated enough.

One way to attain more expressive representation of the KS mapping is to introduce more intricate density descriptors. Following this idea, DFAs invoking more complex density related quantities have been constructed, which belong to higher rungs of the "Jocab's ladder" [13]. It has been demonstrated that the growing complexity of functional form indeed leads to enhanced accuracy [14]. However, practical application of DFT still faces challenges even with the most sophisticated manually designed DFAs.

An alternative approach to enhance the representation of the KS mapping has become increasingly popular, which is to exploit the machine learning (ML) techniques [15–31]. As early as in 1996, Tozer *et al.* [32] have used an artificial neural network (NN) to fit the Zhao–Morrison–Parr XC potential [33]. This pioneering work demonstrated that it is entirely possible to represent the KS mapping by means of ML models. Recently, Zhou *et al.* have employed a deep learning technique – the convolutional NN – to yield in principle the exact KS mapping [34]. Nagai *et al.* have constructed an NN-based mapping from a series of density descriptors to the XC energy density $\epsilon_{\mathrm{XC}}(\mathbf{r})$, which results in a family of NN-based DFAs [35].

Aside from the KS mapping, other ML-based mapping schemes have also been proposed to enhance the predictive power of DFT [36–38]. For instance, NN models have been used to optimize the values of semiempirical parameters [39, 40] in DFAs such as B3LYP and LC-BLYP. Snyder *et al.* have adopted an ML approach to construct the mapping $\rho(\mathbf{r}) \mapsto T$, with T being the kinetic energy functional, for one-dimensional (1D) noninteracting models [41]. Brockherde *et al.* have attempted to bypass the KS equations by directly learning the density-potential and energy-density maps with ML [42]. Dick and Fernandez-Serra [43, 44] and Chen *et al.* [45] have independently constructed an NN-based mapping from the generalized KS single-electron reduced density matrix to an energy correction term.

Despite the above exciting progress, the application

* xz58@ustc.edu.cn

of ML-based KS mapping has been confined to simple model systems or specific types of molecules. Moreover, to achieve a satisfactory accuracy, in some ML-based mapping schemes, sophisticated density descriptors corresponding to higher rungs of the "Jacob's ladder" had to be invoked [35]. The goal of this work is to construct an accurate and practical ML-corrected KS mapping which is potentially useful for generic systems. In particular, we restrict the choice of density descriptors to electron density and its first derivatives, and thus the resulting KS mapping is a generalized gradient approximation (GGA) for the XC functional. A GGA-type functional could offer a simple interpretation of the KS mapping [46], and meanwhile allows for an efficient treatment of extensive systems such as solids. Furthermore, a semilocal mapping also has the advantages that it is intrinsically universal and transferable, automatically preserves the system's spatial symmetry, and does not require the use of any auxiliary atomic basis functions.

Regarding the predictive power of the ML-corrected mapping, we shall focus on thermochemical properties for which conventional GGA functionals yield large errors. For such properties, the functional-driven error usually dominates over density-driven error [47]. Therefore, for numerical convenience, in this work an ML model which presents a correction to a parent GGA functional is constructed in a post-self-consistent-field (post-SCF) manner.

The remainder of this paper is organized as follows. Section II provides a detailed account of our construct of the ML-corrected GGA functional as well as the datasets adopted to determine and assess the parameters involved in the ML model. In Sec. III we demonstrate the numerical performance of the constructed ML-corrected functional and present extensive discussion. Concluding remarks are finally given in Sec. IV.

II. METHODOLOGY

A. ML-PBE: An ML-corrected GGA functional

We choose the widely used Perdew–Burke–Ernzerhof (PBE) functional [48] as the parent DFA. Although the PBE functional yields much improved thermochemical properties than the local density approximation (LDA), its overall accuracy is still far from satisfactory. This is partly because, while the analytic form of the PBE functional satisfies a number of exact physical constraints, it is not sophisticated enough to account for all the subtle features of electron density as well as their influence on the XC energy.

Instead of constructing a more superior GGA functional completely from scratch, we consider a correction to the PBE energy functional, $E_{\rm XC}^{\rm PBE}$, and an ML model is adopted to represent the energy correction, $\Delta E_{\rm XC}^{\rm ML}$. The ML-corrected PBE functional, abbreviated as ML-PBE hereafter, assumes the following form:

$$\begin{aligned} E_{\rm XC}^{\rm ML-PBE} &= E_{\rm XC}^{\rm PBE} + \Delta E_{\rm XC}^{\rm ML} \\ &= \int d\mathbf{r}\, \rho \left[\epsilon_{\rm XC}^{\rm PBE}(r_s,\zeta,s) + \Delta\epsilon_{\rm XC}^{\rm ML}\right] \\ &= \int d\mathbf{r}\, \rho\, \epsilon_{\rm X}^{\rm unif}(\rho)\, F_{\rm XC}^{\rm PBE}(r_s,\zeta,s)\, F^{\rm ML}. \end{aligned} \quad (1)$$

Here, $\epsilon_{\rm X}^{\rm unif}$ is the exchange energy density of a uniform electron gas, and $\epsilon_{\rm XC}^{\rm PBE}$ and $\Delta\epsilon_{\rm XC}^{\rm ML}$ are the XC energy densities corresponding to the PBE functional and the ML correction, respectively. Regarding the involving density-related quantities, $\zeta = (\rho_\uparrow - \rho_\downarrow)/\rho$ is the relative spin polarization, and $s = |\nabla\rho|/(2k_F\rho)$ is the reduced density gradient, with $k_F = (3\pi^2\rho)^{1/3}$ and $r_s = (4\pi\rho/3)^{-1/3}$ being the Fermi wavevector and Wigner–Seitz radius, respectively [48]. Upon the last equality of Eq. (1), the ML-PBE functional is recast into a compact form by referring to the LDA exchange and defining two enhancement factors, $F_{\rm XC}^{\rm PBE}$ and $F^{\rm ML}$, where $F_{\rm XC}^{\rm PBE}$ enhances the LDA exchange to the PBE functional, and $F^{\rm ML}$ enhances the PBE to the ML-corrected PBE.

The challenge then is to construct a universal and accurate semilocal mapping, $\{\rho_\uparrow(\mathbf{r}), \rho_\downarrow(\mathbf{r}), s_\uparrow(\mathbf{r}), s_\downarrow(\mathbf{r})\} \mapsto \Delta\epsilon_{\rm XC}^{\rm ML}(\mathbf{r})$, with $s_\sigma = |\nabla\rho_\sigma|/[2(3\pi^2)^{1/3}\rho_\sigma^{4/3}]$ ($\sigma = \uparrow, \downarrow$) being the spin-specific reduced density gradient. With the inclusion of an ML correction, the exact physical constraints satisfied by the parent DFA are likely to be compromised. This presents another challenge for the development of ML-corrected DFA, which is however beyond the scope of the present paper.

Figure 1 illustrates the workflow for establishing the ML-based KS mapping. To construct a semilocal mapping, an individual point in the $\mathbf{r}$-space of any chemical species is taken as an independent sample. Each sample is associated with a data entry, with the density descriptors $\{\rho_\sigma(\mathbf{r}), s_\sigma(\mathbf{r})\}$ representing the features of the sample and $\Delta\epsilon_{\rm XC}^{\rm ML}(\mathbf{r})$ as the label. The key is to acquire a sufficient amount of pointwise data to enable supervised learning of the ML model. To this end, highly accurate reference data $\Delta\epsilon_{\rm XC}^{\rm ref}(\mathbf{r})$ are needed to guide the learning process, which aims to minimizing the loss function,

$$\Omega = \frac{1}{M}\sum_{m=1}^{M}\int d\mathbf{r}\left|\Delta\epsilon_{{\rm XC},m}^{\rm ML}(\mathbf{r}) - \Delta\epsilon_{{\rm XC},m}^{\rm ref}(\mathbf{r})\right|, \quad (2)$$

where M is the number of species in the training dataset.

Usually, it is the energy difference between two species that can be obtained accurately from experiment or from high-level quantum chemistry calculation. However, to enable the training of a semilocal ML model, the reference value for the energy of each species, $E_m^{\rm ref}$, is required. Moreover, the discrepancy between the energy calculated by the PBE functional and the corresponding reference value, $\Delta E_{{\rm XC},m} = E_{{\rm tot},m}^{\rm ref} - E_{{\rm tot},m}^{\rm PBE}$, is to be decomposed into pointwise errors, which are then assigned to each individual $\mathbf{r}$-point. In the present work, the decomposition of error is done by presuming that the correction to

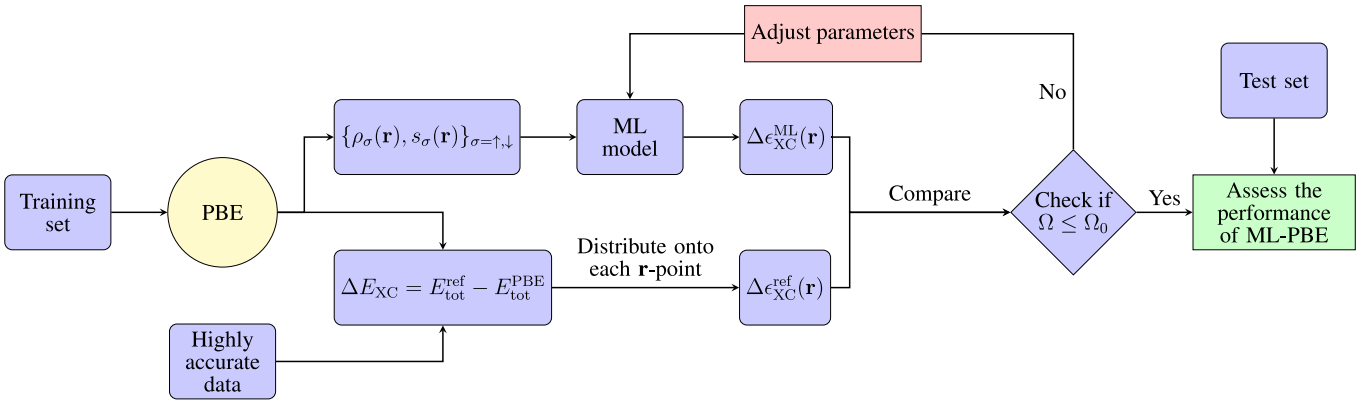

FIG. 1. Schematic diagram illustrating the workflow for constructing the ML-corrected PBE functional. Details are elaborated in the main text. Here, Ω_0 is a preset threshold for the minimization of loss function Ω.

energy density is proportional to the XC energy density itself, i.e.,

$$\Delta\epsilon_{{\rm XC},m}^{\rm ref}(\mathbf{r}) = \frac{\epsilon_{{\rm XC},m}^{\rm DFA}(\mathbf{r})}{E_{{\rm XC},m}^{\rm DFA}} \Delta E_{{\rm XC},m}. \quad (3)$$

In practice, the $\mathbf{r}$-space integral of a function $f(\mathbf{r})$ is often evaluated via a summation over discretized grids, i.e., $\int d\mathbf{r}\, f(\mathbf{r}) \mapsto \sum_i f(\mathbf{r}_i) W(\mathbf{r}_i)$, with $W(\mathbf{r}_i)$ being the weight of the $\mathbf{r}_i$ grid. Equation (3) is thus rewritten as

$$\Delta\epsilon_{{\rm XC},m}^{\rm ref}(\mathbf{r}_i) = \frac{\epsilon_{{\rm XC},m}^{\rm DFA}(\mathbf{r}_i) W(\mathbf{r}_i)}{\sum_i \rho_m(\mathbf{r}_i) \epsilon_{{\rm XC},m}^{\rm DFA}(\mathbf{r}_i) W(\mathbf{r}_i)} \Delta E_{{\rm XC},m}. \quad (4)$$

In principle, the DFA in Eqs. (3) and (4) can be the ML-PBE functional so as to form a self-closed training process; while in practice, the DFA is chosen as the original PBE, which is found to yield a more accurate ML model; see the Supporting Information (SI) for more details. Since each $\mathbf{r}_i$ grid point presents an independent sample, a few chemical species are already capable of providing an ample amount of samples, which suffice to train a sophisticated ML model.

B. Datasets and ML model

The dataset used in this work consists of two parts: a training set for optimizing the ML model, and a test set for evaluating the performance of the ML-corrected functional. The training set contains 166 energetic data, including 148 standard heats of formation (HOF) taken from the G2/97 set [49] (denoted as the G2-HOF set hereafter) and 18 total energies of neutral atoms of the first three periods (from H to Ar). Note that the HOF (or the atomization energy) of a molecular species involves the calculation of the molecular energy as well as the energies of all the constituent atoms. For numerical convenience, the error associated with an HOF data is assigned completely to the molecular species, i.e., $\Delta E_{{\rm XC},m}$ in Eqs. (3) and (4), while the PBE energies of neutral atoms are taken as constant parameters in the training stage.

The test set contains 75 HOF from the G3-3 subset of the G3/99 set [50] (denoted as the G3-HOF set hereafter), and 88 ionization potentials (IPs) and 58 electron affinities (EAs) from the G2/97 set [51] (denoted as the G2-IP and G2-EA sets). The G3-HOF set consists mainly of organic molecules containing 2–10 carbon atoms as well as covalent compounds made of elements from the first three rows of periodic table.

We also assess the predictive power of ML-PBE beyond thermochemical properties by examining energies for reaction kinetics. Specifically, the test set also includes 38 barrier heights for hydrogen-transfer reactions from the HTBH38/08 set [52–54], and 38 barrier heights for non-hydrogen-transfer reactions from the NHTBH38/08 set [52–54] (denoted as the HTBH38 and NHTBH38 sets). There are around 2.4 million $\mathbf{r}$-points in the training set, and around 6 million in the test set.

In the evaluation stage, the total energies of the species in the test set (including atoms, molecules, ions and transition state complexes) are calculated with the ML-PBE functional. The only exception is the total atomic energies for the assessment of the G3-HOF set. To be consistent with the training process, the PBE energies of neutral atoms are taken as constant parameters for the computation of HOF.

The density functional calculations are performed by the Gaussian 16 suite of program [55]. For each species, the atom-centered $\mathbf{r}$-points [56, 57] are generated by an in-house built quantum chemistry software package QM^{4}D [58], with which the density descriptors $\{\rho_\sigma(\mathbf{r}), s_\sigma(\mathbf{r})\}$ on these $\mathbf{r}$-points are then evaluated. The Gaussian basis set 6-311++G(3df, 3pd) is employed throughout this work unless specified. Computational details are provided in the SI.

A number of ML models were tested for constructing the semilocal KS mapping of our interest. These include some well-established deep learning techniques such as the convolutional NN. Among all the candidates, the XG-

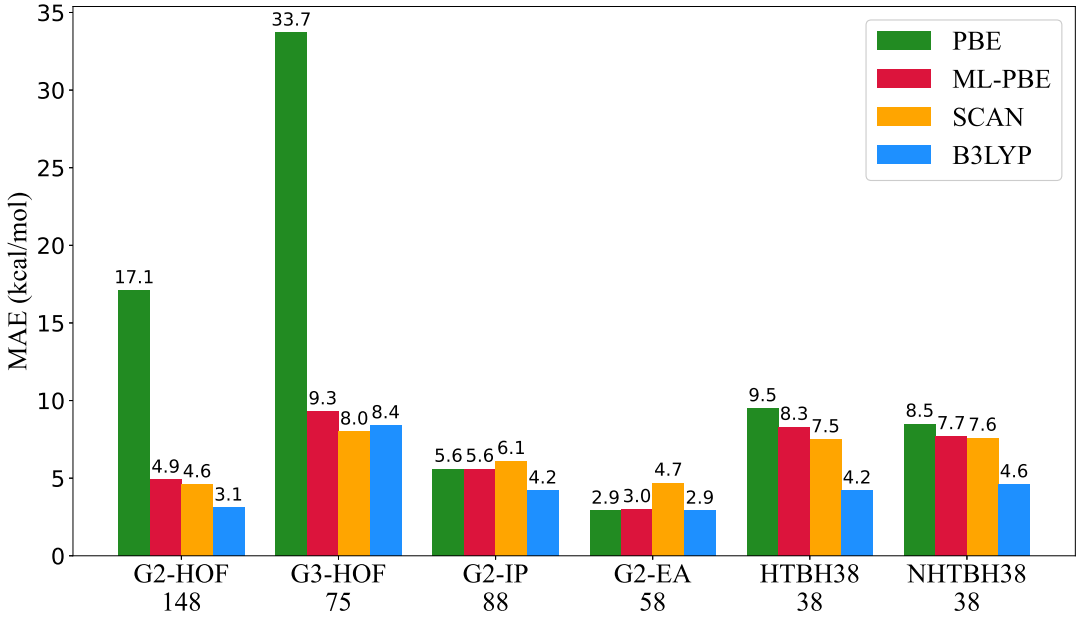

FIG. 2. Performance of ML-PBE in comparison with the PBE, SCAN and B3LYP functionals. The number over each bar is the MAE of the corresponding dataset yielded by a particular DFA; see Sec. II B for the details about the datasets. The MAEs of the G2-IP and G2-EA sets by SCAN are evaluated via calculations done with the 6-311++G(3df, 3pd) basis set, while the other MAEs associated with SCAN are taken from Refs. 59 and 60. The MAEs of the HTBH38/08 and NHTBH38/08 sets by B3LYP are evaluated via calculations done with the MG3S basis set [61], while the other MAEs associated with B3LYP are extracted from Refs. 53 and 62.

Boost [63], which implements the gradient boosting decision tree (GBDT) algorithm [64, 65], excels in the overall performance, and is finally adopted to construct the ML mapping. A generic function $F(x)$ learned by the GBDT has the form of

$$F(x) = \sum_{t=1}^{T} f_t(x), \tag{5}$$

where the basis function $f_t(x)$ is a simple decision tree. The model is trained additively, i.e., at each epoch a new basis function $f_t(x)$ is added to the existing model, followed by optimization of the involving parameters.

Owing to its prevailing advantages in numerical efficiency and robustness, XGBoost has gained increasing interests from practitioners in various fields of chemistry and materials science, and is particularly favorable for our task. In contrast, a sophisticated NN model with a large number of hidden neurons, according to our practical experience, is not compatible with a feature vector of only four dimensions in our problem.

III. RESULTS AND DISCUSSION

To assess the performance of the ML-PBE functional, the mean absolute errors (MAEs) for the various datasets yielded by ML-PBE are exhibited in Fig. 2. For the purpose of comparison, the MAEs of the same datasets yielded by the original PBE functional, the strongly constrained and appropriately normed (SCAN) functional [66] which is a meta-GGA on the third rung of the Jacob's ladder, and the Becke-3-Lee-Yang-Parr (B3LYP) [67–69] hybrid functional which belongs to the fourth rung of the Jacob's ladder, are also displayed.

As shown clearly in Fig. 2, the original PBE yields appreciable errors for the HOF, with MAEs as large as 17.1 and 33.7 kcal/mol for the G2-HOF and G3-HOF sets, respectively. By invoking the ML correction to XC energy density, $\Delta\epsilon_{\rm XC}^{\rm ML}(\mathbf{r})$, the resulting ML-PBE achieves a much improved accuracy for the calculation of HOF. Specifically, the MAEs are substantially reduced to 4.9 and 9.3 kcal/mol for the G2-HOF and G3-HOF sets, respectively. Such errors are already comparable to those yielded by the meta-GGA SCAN. With both PBE and ML-PBE, the MAE for the G3-HOF set is about twice that for the G2-HOF set, because of the larger sizes of molecules in the former dataset. Such error accumulation over the $\mathbf{r}$-space will be scrutinized below.

The ML correction does not improve the prediction of IP and EA, as the MAE yielded by ML-PBE is almost the same as that by the original PBE; see Fig. 2. There are two possible reasons. First, the original PBE yields much larger errors for HOF than for IP and EA, and thus the ML model will focus mainly on the former errors to achieve an overall balanced performance across various energetic properties. Second, the training dataset only involves neutral species, and thus the ML model might

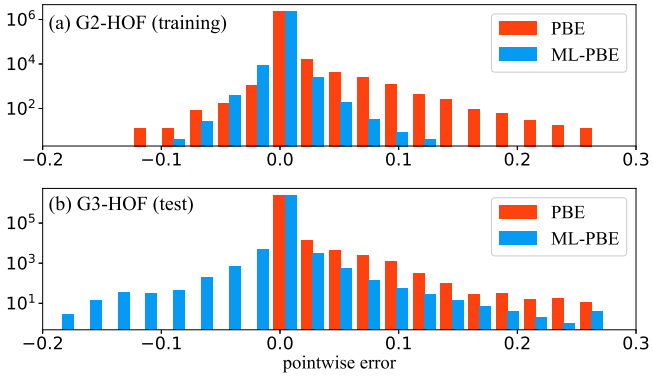

FIG. 3. Histogram of pointwise errors in the XC energy density for PBE and ML-PBE, $\Delta\tilde{\epsilon}_{\rm XC}^{\rm PBE}(\mathbf{r})$ and $\Delta\tilde{\epsilon}_{\rm XC}^{\rm ML-PBE}(\mathbf{r})$, assessed for the (a) G2-HOF and (b) G3-HOF sets. The horizontal position of a bar represents the magnitude of pointwise error in unit of kcal/mol, while the height of a bar stands for the number of $\mathbf{r}$-points possessing that particular amount of error. The vertical axis is in a logarithmic scale.

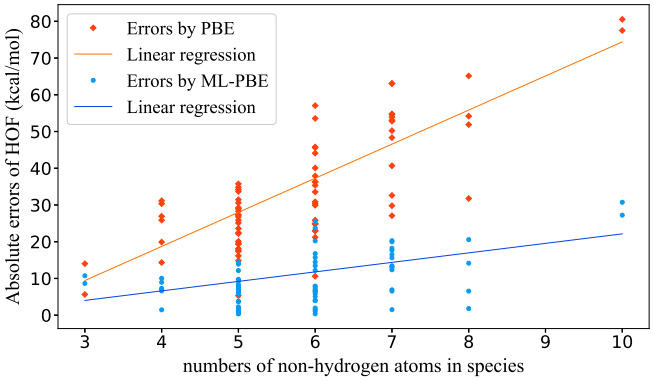

FIG. 4. The error in the XC energy ($\Delta E_{\rm XC}$) versus the number of non-hydrogen atoms ($N_{\rm A}$) for the molecular species in the G3-HOF set. The lines are linear regressions of the errors, with the slopes being 9.3 and 2.6 kcal/mol per atom for PBE and ML-PBE, respectively.

not gain enough knowledge about the charged species. We have attempted to add several ionic species to the training set, but there was not much change in the MAEs. This confirms that the training of the ML model is predominantly driven by the errors of HOF. Nevertheless, both PBE and ML-PBE outperform SCAN in the prediction of IP and EA.

We also extend the assessment of ML-PBE to kinetic properties. As demonstrated in Fig. 2, the ML correction leads to a marginal improvement in the prediction of reaction energy barriers. Just like the situation of IP and EA, since the training dataset consists mainly of molecular species at their equilibrium geometries, the ML model does not learn much about the transition-state species. Moreover, the discrepancy of calculated reaction barriers from the reference values is closely related to the delocalization error of a semilocal DFA, which has been analyzed and understood from the perspective of fractional charges [70–74]. To correct the delocalization error, exact physical constraints such as the Perdew-Parr-Levy-Balduz condition [75] needs to be imposed explicitly [72, 74, 76, 77], which is however beyond the scope of this work. The ML-PBE produces slightly larger MAEs than SCAN for the HTBH38 and NHTBH38 sets.

From the above analysis, it is reasonable to conclude that the ML-PBE generally outperforms the original PBE in the prediction of thermochemical and kinetic properties of molecules, and achieves almost the same level of accuracy as the meta-GGA SCAN. This justifies our proposed strategy of constructing a sophisticated KS mapping by using relatively simple density descriptors. On the other hand, the ML-PBE underperforms the hybrid B3LYP functional for all the energetic properties examined in Fig. 2. This accentuates the limitation of a semilocal functional form, as well as the necessity of building nonlocal ingredients into the ML correction to XC energy for improving further its predictive power.

We proceed to examine whether the distribution of pointwise errors in the XC energy density is improved by the ML correction. Figure 3(a) and (b) depict the histogram of pointwise errors, $\Delta\tilde{\epsilon}_{\rm XC}^{\rm PBE}(\mathbf{r}) = \epsilon_{\rm XC}^{\rm ref}(\mathbf{r}) - \epsilon_{\rm XC}^{\rm PBE}(\mathbf{r})$ and $\Delta\tilde{\epsilon}_{\rm XC}^{\rm ML-PBE}(\mathbf{r}) = \epsilon_{\rm XC}^{\rm ref}(\mathbf{r}) - \epsilon_{\rm XC}^{\rm ML-PBE}(\mathbf{r})$, assessed for the G2-HOF training set and the G3-HOF test set, respectively. Clearly, the original PBE gives rise to a rather biased error distribution with a long tail of positive values, while with the ML correction the pointwise errors become more centralized and balanced. It is noticed that the calculation result on the G3-HOF set by PBE is so biased that the deviation from the reference values are all positive. Nevertheless, ML-PBE counterbalances the biased positive distribution of errors by a relatively larger amount of negative deviations, leading to a drastic drop of MAE. The improvement is overall consistent for both the training and test sets, though the error distribution is somewhat broader for the latter.

As mentioned earlier, a semilocal KS mapping inevitably leads to the accumulation of pointwise errors, i.e., the error in the total XC energy $\Delta E_{\rm XC}$ increases linearly with the size of the system. As shown in Fig. 4, with the original PBE, $\Delta E_{\rm XC}$ of a molecular species grows rapidly with the number of constituent atoms $N_{\rm A}$, and a linear regression indicates a slope of 9.3 kcal/mol per atom. Although the linear increase of $\Delta E_{\rm XC}$ still persists after adopting the ML correction, the rate of growth is greatly suppressed, as verified by a significantly reduced slope of 2.6 kcal/mol per atom. Such a reduction in the slope is consistent with the drop of MAE associated with the G3-HOF set; cf. Fig. 2.

An intriguing question is: how large is the ML correction as compared to the XC energy density of PBE? To answer this question, we examine the magnitude of the enhancement factor corresponding to the ML correction, $F^{\rm ML} = 1 + \Delta\epsilon_{\rm XC}^{\rm ML}/\epsilon_{\rm XC}^{\rm PBE}$. Figure 5 exhibits the distribution of $F^{\rm ML}$ for all the $\mathbf{r}$-points associated with the

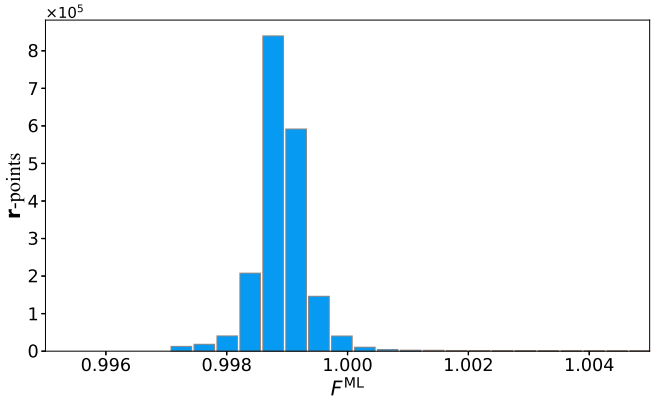

FIG. 5. A histogram showing the distribution of the pointwise enhancement factor F^{ML} for all the species in the training set. The height of a bar represents the number of $\mathbf{r}$-points assuming a particular value of F^{ML}.

species in the training set. It is found that a predominant majority of F^{ML} falls within a tiny interval between 0.997 and 1.0, indicating that the ML correction amounts to only about 0.3% of the magnitude of $\epsilon_{\mathrm{XC}}^{\mathrm{PBE}}$. Therefore, the ML correction scheme proposed in this work is in fact rather conservative, and thus the physical considerations behind the original design of the PBE functional are largely preserved in the ML-PBE.

Nagai and co-workers have also constructed ML-based GGA and meta-GGA functionals [35]. Different from our proposed scheme of finding a correction to an existing DFA, they attempted to establish NN-based DFAs completely from scratch, and thus the enhancement factor of a resulting NN-DFA deviates appreciably from those of the existing DFAs in certain regions of density descriptors; cf. Fig. 5 in Ref. 35. In contrast, we always have $F_{\mathrm{XC}}^{\mathrm{ML-PBE}} \simeq F_{\mathrm{XC}}^{\mathrm{PBE}}$ because of the small magnitudes of the ML correction.

In terms of numerical performance, the NN-GGA obtained by Nagai and co-workers yields an MAE of 11.0 kcal/mol for the G2-HOF set and an MAE of 9.6 kcal/mol for the HTBH38 and NHTBH38 sets combined [35]; while the corresponding MAEs are 4.9 kcal/mol and 8.0 kcal/mol respectively by our proposed ML-PBE. In fact, the ML-PBE achieves an MAE of only 9.3 kcal/mol for the G3-HOF set, which consists of molecules considerably larger than those in the G2-HOF set. Therefore, the ML-PBE is expected to outperform the NN-GGA obtained in Ref. 35 for the prediction of thermochemical energies. Of course, it is worth pointing out that much improved accuracy has been achieved by Nagai and co-workers, with the use of more sophisticated density descriptors [35].

IV. CONCLUDING REMARKS

To summarize, in this work we construct an ML correction to the PBE functional by establishing a semilocal mapping $\{\rho_\sigma(\mathbf{r}), s_\sigma(\mathbf{r})\} \mapsto \Delta\epsilon_{\mathrm{XC}}^{\mathrm{ML}}(\mathbf{r})$. The resulting GGA functional, ML-PBE, yields substantially improved HOF of molecular species than the original PBE, and its overall performance on the prediction of molecular thermochemical and kinetic properties is comparable to the widely used meta-GGA SCAN. The encouraging performance of ML-PBE confirms that integrating the power of data-driven machine learning with physics-inspired derivation is a promising approach towards the heaven of chemical accuracy.

In principle, the ML-PBE functional constructed in this work is immediately applicable to general electronic systems including molecules, clusters, and bulk solids. However, limited by the computational resources and codes at our disposal, numerical calculations were restricted to simple molecules. Further optimization and more extensive assessment of the ML-corrected DFA are certainly desired.

There is still much room for improvement regarding the design of the ML-based KS mapping. First, the construct of a semilocal mapping requires the error in the total energy of a molecule be decomposed into pointwise contributions. The present scheme relies on the presumption of Eq. (3). A more rational way is perhaps to take the XC energy density of a highly accurate hyper-GGA as $\epsilon_{\mathrm{XC}}^{\mathrm{ref}}(\mathbf{r})$. Second, the ML model obtained by this work merely provides a post-SCF correction to the energy, while the electron density is left uncorrected. In principle, the XC potential $v_{\mathrm{XC}}(\mathbf{r})$ can be evaluated by exploiting the analytic gradients of the ML model, which should enable a self-consistent correction to both the energy and the electron density. This is to be pursued in our future work.

Finally, to preserve the computational efficiency of a GGA functional, the present work focuses only on a semilocal KS mapping. However, it has been demonstrated that nonlocal density descriptors may be crucial for addressing long-range electron correlation effects such as the dispersive interaction [34, 35]. This thus raises another challenge: incorporating the nonlocal dependence effectively and efficiently into an ML model. Work along this direction is underway.

ACKNOWLEDGMENTS

The support from the Ministry of Science and Technology of China (Grant Nos. 2016YFA0400900, 2016YFA0200600, and 2017YFA0204904), the National Natural Science Foundation of China (Grant Nos. 21973086 and 21633006), and the Fundamental Research Funds for the Central Universities (Grant No. WK2060000018) is gratefully acknowledged. Computational resources are provided by the Supercomputing Center of University of Science and Technology of China. J.C.W. is grateful for the guidance on numerical aspects from Dr. Tian Lu and colleagues at Beijing Kein Research Center for Natural Sciences.

[1] R. O. Jones, "Density functional theory: Its origins, rise to prominence, and future," Rev. Mod. Phys. **87**, 897 (2015).

[2] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev. **136**, B864 (1964).

[3] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. **140**, A1133 (1965).

[4] A. D. Becke, "A new mixing of Hartree–Fock and local density-functional theories," J. Chem. Phys. **98**, 1372 (1993).

[5] S. Grimme, "Semiempirical hybrid density functional with perturbative second-order correlation," J. Chem. Phys. **124**, 034108 (2006).

[6] Y. Zhang, X. Xu, and W. A. Goddard, "Doubly hybrid density functional for accurate descriptions of nonbond interactions, thermochemistry, and thermochemical kinetics," Proc. Natl. Acad. Sci. U.S.A. **106**, 4963 (2009).

[7] Y. Zhao, B. J. Lynch, and D. G. Truhlar, "Doubly hybrid meta DFT: New multi-coefficient correlation and density functional methods for thermochemistry and thermochemical kinetics," J. Phys. Chem. A **108**, 4786 (2004).

[8] T. Leininger, H. Stoll, H.-J. Werner, and A. Savin, "Combining long-range configuration interaction with short-range density functionals," Chem. Phys. Lett. **275**, 151 (1997).

[9] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened Coulomb potential," J. Chem. Phys. **118**, 8207 (2003).

[10] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, "A long-range correction scheme for generalized-gradient-approximation exchange functionals," J. Chem. Phys. **115**, 3540 (2001).

[11] T. Yanai, D. P. Tew, and N. C. Handy, "A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP)," Chem. Phys. Lett. **393**, 51 (2004).

[12] P. Verma, Y. Wang, S. Ghosh, X. He, and D. G. Truhlar, "Revised M11 exchange-correlation functional for electronic excitation energies and ground-state properties," J. Phys. Chem. A **123**, 2966 (2019).

[13] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, "Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits," J. Chem. Phys. **123**, 062201 (2005).

[14] D. Hait and M. Head-Gordon, "How accurate is density functional theory at predicting dipole moments? An assessment using a new database of 200 benchmark values," J. Chem. Theory Comput. **14**, 1969 (2018).

[15] K. Ryczko, D. A. Strubbe, and I. Tamblyn, "Deep learning and density-functional theory," Phys. Rev. A **100**, 022512 (2019).

[16] M. M. Denner, M. H. Fischer, and T. Neupert, "Efficient learning of a one-dimensional density functional theory," Phys. Rev. Res. **2**, 033388 (2020).

[17] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, "Prediction errors of molecular machine learning models lower than hybrid DFT error," J. Chem. Theory Comput. **13**, 5255 (2017).

[18] J. T. Margraf and K. Reuter, "Pure non-local machine-learned density functional theory for electron correlation," Nat. Commun. **12**, 344 (2021).

[19] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, "Solving the electronic structure problem with machine learning," npj Comput. Mater. **5**, 22 (2019).

[20] A. T. Fowler, C. J. Pickard, and J. A. Elliott, "Managing uncertainty in data-derived densities to accelerate density functional theory," J. Phys.: Mater. **2**, 034001 (2019).

[21] J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, and K. W. Jacobsen, "Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation," Phys. Rev. B **85**, 235149 (2012).

[22] 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).

[23] M. Aldegunde, J. R. Kermode, and N. Zabaras, "Development of an exchange–correlation functional with uncertainty quantification capabilities for density functional theory," J. Comput. Phys. **311**, 173 (2016).

[24] J. Nelson, R. Tiwari, and S. Sanvito, "Machine learning density functional theory for the Hubbard model," Phys. Rev. B **99**, 075132 (2019).

[25] 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).

[26] J. Wellendorff, K. T. Lundgaard, K. W. Jacobsen, and T. Bligaard, "mBEEF: An accurate semi-local Bayesian error estimation density functional," J. Chem. Phys. **140**, 144107 (2014).

[27] 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).

[28] M. Fritz, M. Fernández-Serra, and J. M. Soler, "Optimization of an exchange-correlation density functional for water," J. Chem. Phys. **144**, 224101 (2016).

[29] M. Tsubaki and T. Mizoguchi, "Quantum deep field: Data-driven wave function, electron density generation, and atomization energy prediction and extrapolation with machine learning," Phys. Rev. Lett. **125**, 206401 (2020).

[30] 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).

[31] Y. Suzuki, R. Nagai, and J. Haruyama, "Machine learning exchange-correlation potential in time-dependent density-functional theory," Phys. Rev. A **101**, 050501 (2020).

[32] D. J. Tozer, V. E. Ingamells, and N. C. Handy, "Exchange-correlation potentials," J. Chem. Phys. **105**, 9200 (1996).

[33] Q. Zhao, R. C. Morrison, and R. G. Parr, "From electron densities to Kohn-Sham kinetic energies, orbital energies, exchange-correlation potentials, and exchange-

correlation energies," Phys. Rev. A **50**, 2138 (1994).
[34] 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).
[35] R. Nagai, R. Akashi, and O. Sugino, "Completing density functional theory by machine learning hidden messages from molecules," npj Comput. Mater. **6**, 43 (2020).
[36] R. A. Vargas-Hernández, "Bayesian optimization for calibrating and selecting hybrid-density functional models," J. Phys. Chem. A **124**, 4053 (2020).
[37] A. Fabrizio, B. Meyer, and C. Corminboeuf, "Machine learning models of the energy curvature vs particle number for optimal tuning of long-range corrected functionals," J. Chem. Phys. **152**, 154103 (2020).
[38] M. Yu, S. Yang, C. Wu, and N. Marom, "Machine learning the Hubbard U parameter in DFT+U using Bayesian optimization," npj Comput. Mater. **6**, 180 (2020).
[39] X. Zheng, L. Hu, X. Wang, and G. Chen, "A generalized exchange-correlation functional: the Neural-Networks approach," Chem. Phys. Lett. **390**, 186 (2004).
[40] 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).
[41] 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).
[42] 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).
[43] S. Dick and M. Fernandez-Serra, "Learning from the density to correct total energy and forces in first principle simulations," J. Chem. Phys. **151**, 144102 (2019).
[44] S. Dick and M. Fernandez-Serra, "Machine learning accurate exchange and correlation functionals of the electronic density," Nat. Commun. **11**, 3509 (2020).
[45] 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).
[46] N. C. Handy, "The importance of Colle–Salvetti for computational density functional theory," Theor. Chem. Acc. **123**, 165 (2009).
[47] M.-C. Kim, E. Sim, and K. Burke, "Understanding and reducing errors in density functional calculations," Phys. Rev. Lett. **111**, 073003 (2013).
[48] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865 (1996).
[49] 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).
[50] 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).
[51] 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).
[52] 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).
[53] 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).
[54] 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).
[55] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian˜16 Revision C.01, 2016, Gaussian Inc. Wallingford CT.
[56] M. E. Mura and P. J. Knowles, "Improved radial grids for quadrature in molecular density-functional calculations," J. Chem. Phys. **104**, 9848 (1996).
[57] V. I. Lebedev and D. N. Laikov, "A quadrature formula for the sphere of the 131st algebraic order of accuracy," Dok Math **59**, 477 (1999).
[58] X. Hu, H. Hu, X. Zheng, X. Zeng, P. Wu, D. Peng, Y. Jin, L. Yu, and W. Yang, (2021), An in-house program for QM/MM simulations. Available online at: `http://www.qm4d.info`.
[59] K. Hui and J.-D. Chai, "SCAN-based hybrid and double-hybrid density functionals from models without fitted parameters," J. Chem. Phys. **144**, 044114 (2016).
[60] Y. Mo, G. Tian, and J. Tao, "Performance of a nonempirical exchange functional from density matrix expansion: comparative study with different correlations," Phys. Chem. Chem. Phys. **19**, 21707 (2017).
[61] B. J. Lynch, Y. Zhao, and D. G. Truhlar, "Effectiveness of diffuse basis functions for calculating relative energies by density functional theory," J. Phys. Chem. A **107**, 1384 (2003).
[62] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes," J. Chem. Phys. **119**, 12129 (2003).
[63] T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System, in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, page 785–794, 2016.
[64] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," Ann. Statist. **29**, 1189 (2001).

[65] J. H. Friedman, "Stochastic gradient boosting," Comput. Stat. Data Anal. **38**, 367 (2002).

[66] J. Sun, A. Ruzsinszky, and J. P. Perdew, "Strongly constrained and appropriately normed semilocal density functional," Phys. Rev. Lett. **115**, 036402 (2015).

[67] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," Phys. Rev. A **38**, 3098 (1988).

[68] 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).

[69] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," J. Chem. Phys. **98**, 5648 (1993).

[70] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Fractional charge perspective on the band gap in density-functional theory," Phys. Rev. B **77**, 115123 (2008).

[71] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Insights into current limitations of density functional theory," Science **321**, 792 (2008).

[72] C. Li, X. Zheng, A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Local scaling correction for reducing delocalization error in density functional approximations," Phys. Rev. Lett. **114**, 053001 (2015).

[73] X. Zheng, M. Liu, E. R. Johnson, J. Contreras-García, and W. Yang, "Delocalization error of density-functional approximations: A distinct manifestation in hydrogen molecular chains," J. Chem. Phys. **137**, 214106 (2012).

[74] C. Li, X. Zheng, N. Q. Su, and W. Yang, "Localized orbital scaling correction for systematic elimination of delocalization error in density functional approximations," Natl. Sci. Rev. **5**, 203 (2017).

[75] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, "Density-Functional Theory for fractional particle number: Derivative discontinuities of the energy," Phys. Rev. Lett. **49**, 1691 (1982).

[76] X. Zheng, A. J. Cohen, P. Mori-Sánchez, X. Hu, and W. Yang, "Improving band gap prediction in density functional theory from molecules to solids," Phys. Rev. Lett. **107**, 026403 (2011).

[77] X. Zheng, C. Li, D. Zhang, and W. Yang, "Scaling correction approaches for reducing delocalization error in density functional approximations," Sci. China Chem. **58**, 1825 (2015).