Learning Generalized Residual Exchange-Correlation-Uncertain Functional for Density Functional Theory

Sizhuo Jin[1], Shuo Chen[2], Jianjun Qian[1], Ying Tai[2], Jun Li[1]*

[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China.
[2]School of Intelligence Science and Technology, Nanjing University, Suzhou, China.
{jinsizhuo, csjqian, junli}@njust.edu.cn, shuo.chen.ya@foxmail.com, yingtai@nju.edu.cn

Abstract

Density Functional Theory (DFT) stands as a widely used and efficient approach for addressing the many-electron Schrödinger equation across various domains such as physics, chemistry, and biology. However, a core challenge that persists over the long term pertains to refining the exchange-correlation (XC) approximation. This approximation significantly influences the triumphs and shortcomings observed in DFT applications. Nonetheless, a prevalent issue among XC approximations is the presence of systematic errors, stemming from deviations from the mathematical properties of the exact XC functional. For example, although both B3LYP and DM21 (DeepMind 21) exhibit improvements over previous benchmarks, there is still potential for further refinement. In this paper, we propose a strategy for enhancing XC approximations by estimating the neural uncertainty of the XC functional, named Residual XC-Uncertain Functional. Specifically, our approach involves training a neural network to predict both the mean and variance of the XC functional, treating it as a Gaussian distribution. To ensure stability in each sampling point, we construct the mean by combining traditional XC approximations with our neural predictions, mitigating the risk of divergence or vanishing values. It is crucial to highlight that our methodology excels particularly in cases where systematic errors are pronounced. Empirical outcomes from three benchmark tests substantiate the superiority of our approach over existing state-of-the-art methods. Our approach not only surpasses related techniques but also significantly outperforms both the popular B3LYP and the recent DM21 methods, achieving average RMSE improvements of 62% and 37%, respectively, across the three benchmarks: W4-17, G21EA, and G21IP.

Introduction

Density Functional Theory (DFT) has gained widespread popularity as a technique for calculating electronic structures and energies, unveiling physical properties in various fields including chemistry, material science, biology, and medicine. This popularity arises from a well-accepted compromise between computational cost and accuracy, as highlighted by the consensus (Burke 2012). While an exact functional method from electron density to energy theoretically exists and has been demonstrated (Hohenberg and Kohn 1964), the exponential time complexity in relation to the number of electrons renders practical implementation of DFT unfeasible.To address this, Kohn-Sham (K-S) formalism was introduced (Kohn and Sham 1965), significantly reducing the computational complexity of DFT to cubic order. This achievement was realized by leveraging the concept of non-interacting electrons within a many-electron system, leading to the formulation of a set of self-consistent equations. Despite this advancement, a persistent challenge lies in the development of suitable XC approximations for the K-S equations in practical calculations (Medvedev et al. 2017).

In the preceding decades, three classical XC approximations have emerged. The Local Density Approximation (LDA) (Kohn and Sham 1965), stands as the simplest and most prevalent functional for accurately describing molecular geometries. However, LDA exhibits a significant drawback by strongly overbinding molecules, leading to an overestimation of about 1 eV per bond. To address this limitation, the Generalized Gradient Approximation (GGA) (Perdew 1986), takes a stride by incorporating the first derivative of the density, thus improving the accuracy of chemical calculations and rectifying the overbinding issue present in LDA. Furthermore, B3LYP (Becke 1993) ingeniously integrates the Hartree-Fock (HF) exchange functional, amalgamating it with various local exchange and correlation functionals, such as LDA (Kohn and Sham 1965), Becke's 1988 (B88) (Becke 1988), Lee-Yang-Parr (LYP) (Lee, Yang, and Parr 1988), and Vosko-Wilk-Nusair (VWN) (Vosko, Wilk, and Nusair 1980). B3LYP has become a cornerstone in various domains of chemistry due to its improved accuracy. However, the drawback lies in the fact that these hybrid functional coefficients are largely manually calibrated, yielding hand-designed functionals that often fail to surpass the B3LYP (Cohen, Mori-Sanchez, and Yang 2012).

Recently, DeepMind 21 (DM21) (Kirkpatrick et al. 2021) has introduced a novel approach that combines local LDA correlation and HF exchange functionals through a Multilayer Perceptron (MLP). This MLP predicts the local enhancement factors, essentially the combined coefficients, enabling the integration of local energies throughout space. Once trained, DM21 proves effective in Self-Consistent Field (SCF) calculations, offering precise results for intricate systems like chemical reactions. Although DM21 surpasses

*corresponding author

leading hybrid XC functionals such as B3LYP (Becke 1993), M06-2X (Zhao and Truhlar 2008), and ωB97X (Chai and Head-Gordon 2008), all of them still have significant systematic errors compared to exact (or highly accurate) functionals, see the DM21 results in benchmark datasets like GMTKN55 (Goerigk and Grimme 2010) and QM9 (Kim, Park, and Choi 2019). In addition, to estimate the above systematic errors, (Mortensen et al. 2005; Aldegunde, Kermode, and Zabaras 2016) use Bayesian statistics to construct an ideal functional that combines an ensemble of existing XC functionals and a Gaussian error for uncertainty quantification.

To better capture systematic errors, we introduce a residual uncertainty framework to characterize the eXchange-Correlation (XC) functional. Drawing inspiration from uncertainty modeling in machine learning and computer vision (Kendall and Gal 2017) and the Bayesian error estimation of DFT (Mortensen et al. 2005), we conceive of the XC functional as a Gaussian distribution characterized by its mean and variance. Subsequently, we develop a neural network to predict both the mean and variance, facilitating an accurate approximation of the XC functional. This neural network takes into account both local and nonlocal features of the occupied Kohn-Sham (KS) orbitals (Kirkpatrick et al. 2021). Challenges arise in achieving convergence in Self-Consistent Field (SCF) calculations for the KS equations when directly predicting the mean and variance. This sensitivity is attributed to the integration of local energies across all space. To mitigate this issue, we are inspired by the residual learning (He et al. 2016), and construct the mean by amalgamating traditional XC approximations with the network predictive mean. Moreover, we regulate the predictive variance using a scaling factor to prevent excessive disruptions. Consequently, based on the aforementioned analysis, we propose a Residual Bayesian network (RBNet) designed to forecast both the mean and variance of the XC functional. It is noteworthy that our RBNet can seamlessly integrate with established traditional XC approximations (*e.g.*, B3LYP and DM21). To confirm this, our approach exhibits substantial enhancements over both B3LYP and DM21 methodologies. Overall, our contributions can be summarized as follow:

- We propose a generalized residual XC-uncertain functional for density functional theory by introducing a Residual Bayesian Network (RBNet), a novel Residual uncertainty method designed to enhance the modeling of the XC functional by forecasting both mean and variance. To the best of our knowledge, our contribution lies in pioneering the integration of neural uncertainty into the development of XC functionals.
- Furthermore, our approach exhibits wide applicability across various XC approximation methods. This is particularly advantageous when these methods exhibit substantial systematic errors, as our approach excels in rectifying and enhancing their performance.
- We comprehensively evaluate the efficacy of our methodology using four prominent public benchmarks: W4-17, G21EA, and G21IP. The experimental findings unequivocally demonstrate the remarkable superiority of our approach compared to existing exchange-correlation approximation algorithms. Crucially, our approach consistently outperforms not only the renowned B3LYP method but also the recent DM21 model, underscoring its exceptional generalization capability.

Related Work

DFT Using Deep Learning

Density Functional Theory (DFT) has revolutionized the application of quantum mechanics to intricate challenges in chemistry, materials science, biology, and medicine. Reviews such as (Cohen, Mori-Sanchez, and Yang 2012; Burke 2012; Medvedev et al. 2017) encompass most conventional DFT methods. In this context, we focus on a comprehensive examination of eXchange-Correlation (XC) approximation techniques utilizing deep learning (Pederson, Kalita, and Burke 2022). The concept involves training a neural network to serve as a universal XC functional, capable of accurately reproducing both the exact XC energy and potential (Schmidt, Benavides-Riveros, and Marques 2019). Additionally, we explore the utilization of a three-dimensional convolutional neural network designed to approach an exact XC potential (Zhou et al. 2019). Another avenue involves NeuralXC, which employs a Behler-Parrinello-style neural network architecture to generate non-local XC functionals (Dick and Fernandez-Serra 2020).

The implementation of neural networks for the exchange-correlation functional holds the potential to enhance simulation precision within the framework of differentiable Kohn-Sham density functional theory (Kasim and Vinko 2021). Here, we adopt machine learning algorithms to semi-automate the creation of X and XC holes via artificial neural networks (Cuierrier, Roy, and Ernzerhof 2021). The act of training neural networks for the exchange-correlation functional introduces implicit regularization, significantly bolstering generalization capabilities (Li et al. 2021). Furthermore, by applying a machine learning model adhering to physical asymptotic constraints, the XC functional is constructed by separating it into exchange (X) and correlation (C) components (Nagai, Akashi, and Sugino 2022). To enhance the strongly constrained and appropriately normed SCAN functional, a deep neural network is ingeniously designed, leveraging precise constraint adherence and appropriate norms (Pokharel et al. 2022).

In recent developments, DM21 employs a multilayer perceptron (MLP) to predict local enhancement factors of local LDA correlation and HF exchange functionals, particularly for the integration of local energies throughout space (Kirkpatrick et al. 2021). A derivative of this approach, known as many-body MB-DM21 potential, finds application in simulating liquid water across varying temperatures and ambient pressures (Palos et al. 2022). Despite the considerable advancements of these machine learning methodologies compared to traditional XC methods, they still exhibit notable systematic errors in benchmarking. This paper introduces a novel strategy to mitigate these systematic errors, thereby enhancing the precision of XC functional calculations.

Backgrounds

B3LYP (Becke 1993) stands out as one of the most extensively employed exchange-correlation (XC) functionals in density functional theory (DFT) calculations. This hybrid functional amalgamates the HF exchange functional, denoted as e_x^{HF}, with the VWN correlation functional (Vosko, Wilk, and Nusair 1980), along with the local density approximation e_x^{LDA} (Kohn and Sham 1965), the Becke 88 three-parameter exchange functional e_x^{GGA} (Becke 1988), and the LYP correlation functional e_c^{GGA} (Lee, Yang, and Parr 1988). The formula for B3LYP is defined as follows:

$$\begin{aligned} e_{xc}^{\mathrm{B3LYP}} &= e_x^{\mathrm{LDA}} + e_c^{\mathrm{LDA}} + a(e_x^{\mathrm{HF}} - e_x^{\mathrm{LDA}}) \\ &\quad + b(e_x^{\mathrm{GGA}} - e_x^{\mathrm{LDA}}) + c(e_c^{\mathrm{GGA}} - e_c^{\mathrm{LDA}}), \end{aligned} \tag{1}$$

where the parameters are commonly set as $a = 0.20$, $b = 0.72$, and $c = 0.81$. Due to its affordability, versatility, and potential for enhancement, the B3LYP functional has spurred the development of composite methods like B3LYP-3c (Pracht, Grant, and Grimme 2020) and B3LYP-D3-DCP (van Santen and DiLabio 2015). Typically, B3LYP is employed alongside dispersion corrections such as DFT-D, compensating for its limited consideration of noncovalent interactions.

DeepMind 21 (DM21) (Kirkpatrick et al. 2021), is firstly developed by DeepMind in 2021, which involves training a multi-layer perceptron network to address systematic errors resulting from the violation of mathematical properties of the exact exchange-correlation (XC) functional. This approach effectively characterizes instances of artificial charge delocalization and strong correlation, outperforming conventional XC functionals like B3LYP (Becke 1993) and ωB97X (Chai and Head-Gordon 2008). The DM21 model can be described by the following formalization:

$$e_{xc}^{\mathrm{DM21}} = a_1 e^{\mathrm{LDA}} + a_2 e^{\mathrm{HF}} + a_3 e^{\omega \mathrm{HF}}, \tag{2}$$

where the local enhancement factors a_1, a_2 and a_3 are obtained as the three-dimensional outputs of a multi-layer perceptron network trained on molecular data as well as on hypothetical systems with fractional charge and spin. Furthermore, an extension of DM21, known as Many-body MB-DM21 (Palos et al. 2022), applies the principles of DM21 within the many-body expansion of the energy. This extension is utilized in simulations of liquid water, considering temperature variations under ambient pressure conditions.

Methodology

In this section, we introduce a pioneering approach for crafting a generalized XC functional within KS-DFT calculations. Initially, we harness the concept of uncertainty to formulate a versatile structure for the XC functional while scrutinizing its inherent constraints. Subsequently, we incorporate an additional residual uncertainty to enhance the modeling of the XC functional. Finally, we elucidate the training loss associated with our proposed approach.

Generalized XC-Uncertain Functional

Due to the inherent difficulty in obtaining the precise XC functional, we draw inspirations from Bayesian deep learning techniques (Kendall and Gal 2017) and Bayesian error estimation of DFT (Mortensen et al. 2005). In this context, we create a neural network capable of directly forecasting both the mean and variance of the XC energy density. Mathematically, we establish a generalized XC-uncertain functional represented by a Gaussian distribution characterized by mean and variance parameters:

$$e_{xc} = \widehat{e}_{xc} + \widehat{\sigma}\epsilon, \;\; [\widehat{e}_{xc}, \widehat{\sigma}^2] = f(\mathbf{x}(\mathbf{r}); \theta), \;\; \epsilon \sim \mathcal{N}(0, I), \tag{3}$$

where $f(\cdot;\cdot)$ denotes a Bayesian neural network parameterized by model weights θ, while $\mathbf{x}(\mathbf{r})$ represents the input features at the spatial location $\mathbf{r}$. A single network can be employed to process the input $\mathbf{x}(\mathbf{r})$, with its architecture designed to predict both $\widehat{e}_{xc}$ and $\widehat{\sigma}^2$ concurrently. Following the DM21 feature (Kirkpatrick et al. 2021), we adopt an 11-dimensional vector representation for the feature $\mathbf{x}(\mathbf{r})$:

$$\begin{aligned} \mathbf{x}(\mathbf{r}) = &[\rho_\uparrow, \rho_\downarrow, |\nabla\rho_\uparrow|^2, |\nabla\rho_\downarrow|^2, |\nabla\rho|^2, \\ &\tau_\uparrow, \tau_\downarrow, e_\uparrow^{\omega_1 \mathrm{HF}}, e_\downarrow^{\omega_1 \mathrm{HF}}, e_\uparrow^{\omega_2 \mathrm{HF}}, e_\downarrow^{\omega_2 \mathrm{HF}}], \end{aligned} \tag{4}$$

where these features are the functions of the location $\mathbf{r}$, and denoted by

- The densities with spins: $\rho_\uparrow(\mathbf{r})$, $\rho_\downarrow(\mathbf{r})$
- The square norm of the gradient of the densities: $|\nabla\rho_\uparrow(\mathbf{r})|^2$, $|\nabla\rho_\downarrow(\mathbf{r})|^2$, $|\nabla\rho(\mathbf{r})|^2 = |\nabla(\rho_\uparrow(\mathbf{r}) + \rho_\downarrow(\mathbf{r}))|^2$
- The kinetic energy densities with spins: $\tau_\uparrow(\mathbf{r})$, $\tau_\downarrow(\mathbf{r})$
- The local HF features with $\omega_1 = 0.4$ and $\omega_2 \to \infty$: $e_\uparrow^{\omega_1 \mathrm{HF}}(\mathbf{r})$, $e_\downarrow^{\omega_1 \mathrm{HF}}(\mathbf{r})$, $e_\uparrow^{\omega_2 \mathrm{HF}}(\mathbf{r})$ and $e_\downarrow^{\omega_2 \mathrm{HF}}(\mathbf{r})$.

Using the XC energy density e_{xc} in Eq.(3), the XC energy $E_{xc}[n]$ is constructed by integrating the density e_{xc}:

$$E_{xc} = E_{xc}^{\mathrm{U}} + \sigma\epsilon, \epsilon \sim \mathcal{N}(0, 1), \tag{5}$$

where $\rho(\mathbf{r})$ is the electron density, $E_{xc}^{\mathrm{U}}[\rho] = \int \rho(\mathbf{r})\widehat{e}_{xc}(\mathbf{r})d\mathbf{r} = \sum_{\mathbf{r}} \rho(\mathbf{r})\overline{e}_{xc}(\mathbf{r})$ and $\sigma[\rho] = \int \rho(\mathbf{r})\widehat{\sigma}(\mathbf{r})\epsilon d\mathbf{r} = \sum_{\mathbf{r}} \rho(\mathbf{r})\widehat{\sigma}(\mathbf{r})\epsilon = \sigma\epsilon$. Here, the integral represents the sum of the function on all grand points. To learn this general Uncertainty of the XC functional, we employ a minimization objective:

$$\mathcal{L}_{\mathrm{U}} = \frac{1}{N}\sum_i \frac{1}{2}\sigma_i^2 \|E_{xc,i}^{\mathrm{U}} - E_{xc,i}^*\|^2 + \frac{1}{2}\log\sigma_i^2, \tag{6}$$

where N represents the total count of reactions, where i signifies the ith individual reaction. $E_{xc,i}^{\mathrm{U}}$ corresponds to the predicted total exchange-correlation energy change for the reaction (difference between product and reactant exchange-correlation energies), while $E_{xc,i}^*$ stands for a precise exchange-correlation energy for the reaction obtained from literature total energy calculations.

However, learning both $\widehat{e}_{xc}$ and $\widehat{\sigma}^2$ directly poses challenges due to integration, resulting in two distinct issues. Firstly, the sensitivity of $\widehat{e}_{xc}$ to loss convergence is heightened by the stochastically initialized network parameters. Secondly, the accumulation of variance frequently triggers loss explosion. To address these dilemmas, we introduce a concept of residual uncertainty in the subsequent subsection.

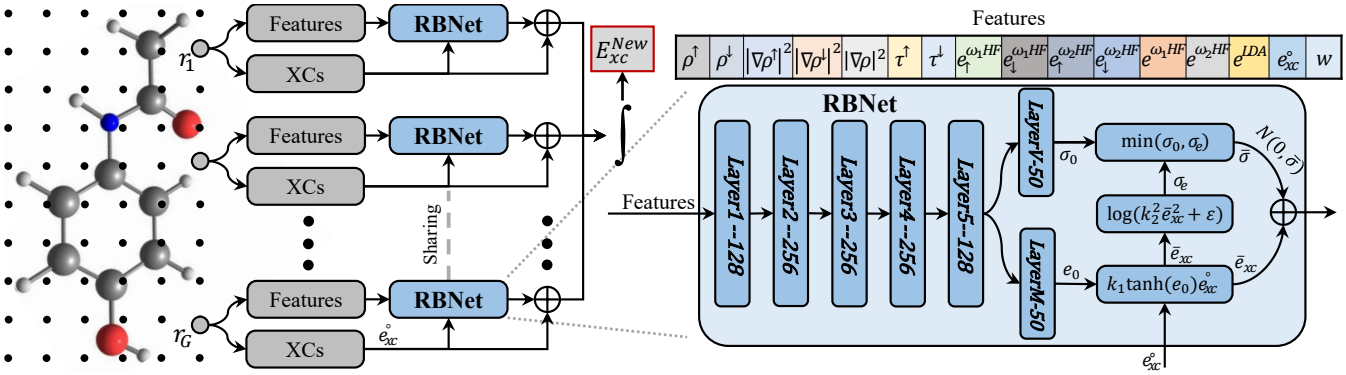

Figure 1: Overview of our residual XC-uncertain functional architecture. The features of the electron density are computed through sampling KS orbitals on grid points. In contrast to conventional XC functionals that produce a local XC energy density e_{xc}°, our RBNet model predicts a residual energy density situated between the ideal energy density and the e_{xc}°.

Residual XC-Uncertain Functional

To address the aforementioned issues, we are inspired by the residual learning (He et al. 2016) and introduce a concept of residual uncertainty to enhance the modeling of XC-Uncertain energy density, as depicted in Fig.1. Our core concept involves learning a residual density situated between the theoretical energy density $\widehat{e}_{xc}$ and the conventional energy density e_{xc}°. Additionally, we employ the established energy density e_{xc}° to impose constraints on the residual mean and variance, effectively preventing them from causing instability. Consequently, we define the residual XC-uncertain energy density as follows:

$$\begin{aligned} e_{xc} &= e_{xc}^{\circ} + \underbrace{\overline{e}_{xc} + \overline{\sigma}\epsilon}_{\text{RBNet}}, \epsilon \sim \mathcal{N}(0, I), \\ [\overline{e}_{xc}, \overline{\sigma}] &= g(\mathbf{y}(\mathbf{r}), e_{xc}^{\circ}; \phi), \end{aligned} \tag{7}$$

where e_{xc}° denotes any conventional XC energy density method (in this paper, we primarily consider B3LYP ($e_{xc}^{\circ} = e_{xc}^{\text{B3LYP}}$) (Becke 1993) and DM21 ($e_{xc}^{\circ} = e_{xc}^{\text{DM21}}$) (Kirkpatrick et al. 2021), although numerous other methods are also applied). The function $g(\cdot;\cdot)$ represents a residual Bayesian neural network characterized by model weights ϕ along with two scaling factors k_1 and k_2, while $\mathbf{x}(\mathbf{r})$ stands for the input features at location $\mathbf{r}$.

Features. Here, we extend the DM21 feature from 11-dimensional vector to 16-dimensianl vector:

$$\mathbf{y}(\mathbf{r}) = [\mathbf{x}(\mathbf{r}), e^{\omega_1 \text{HF}}, e^{\omega_2 \text{HF}}, e^{\text{LDA}}, e_{xc}^{\circ}, w], \tag{8}$$

where $\mathbf{x}(\mathbf{r})$ is defined in Eq.(4), and others are denoted by

- The local HF energy with $\omega_1 = 0.4$ and $\omega_2 \to \infty$, $e^{\omega_1 \text{HF}} = e_{\uparrow}^{\omega_1 \text{HF}}(\mathbf{r}) + e_{\downarrow}^{\omega_1 \text{HF}}(\mathbf{r})$, $e^{\omega_2 \text{HF}} = e_{\uparrow}^{\omega_2 \text{HF}}(\mathbf{r}) + e_{\downarrow}^{\omega_2 \text{HF}}(\mathbf{r})$,
- The LDA energy, $e^{\text{LDA}}(\mathbf{r})$, any known energy $e_{xc}^{\circ}(\mathbf{r})$, and w is the weight of the grand point.

Residual Bayesian Network (RBNet). To overcome the difficulty of learning the ideal $\widehat{e}_{xc}$ in the Eq.(3), we adopt the conventional XC energy density e_{xc}° as a reference point. This enables us to model a *residual XC energy density* $\overline{e}_{xc}$:

$$\overline{e}_{xc} = \widehat{e}_{xc} - e_{xc}^{\circ}. \tag{9}$$

Subsequently, we introduce a residual Bayesian network denoted as $g(\mathbf{y}(\mathbf{r}), e_{xc}^{\circ}; \phi)$ to simultaneously learn both $\overline{e}_{xc}$ and $\overline{\sigma}^2$ using our novel feature $\mathbf{y}(\mathbf{r})$. As illustrated in Figure1, the RBNet comprises a unified network structure featuring two heads for the prediction of temporary values e_0 and σ_0. To ensure the realism of XC energy densities, we formulate $\overline{e}_{xc}$ and $\overline{\sigma}^2$ as follows:

$$\begin{aligned} \overline{e}_{xc} &= k_1 \tanh(e_0) e_{xc}^{\circ}, \\ \overline{\sigma} &= \min\{\sigma_0, \sigma_e\}, \quad \sigma_e = \log(k_2^2 \overline{e}_{xc}^2 + \varepsilon), \end{aligned} \tag{10}$$

where ε is introduced as a small constant to prevent $\overline{e}_{xc}$ from becoming excessively small. The scaling factors $0 < k_1 < 2$ and $0 < k_2 < 2$ play a role in shaping the model's behavior. Specifically, we set $\varepsilon = 10^{-4}$ and $k_1 = k_2 = 1$.

Training Loss

Before presenting the training loss, we first show the energy through the XC energy density in Eq.(7), referred to as residual uncertainty, which is subsequently determined via the following integral:

$$E_{xc} = E_{xc}^{\circ} + E_{xc}^{\text{RU}} + \sigma\epsilon, \epsilon \sim \mathcal{N}(0, I), \tag{11}$$

where the terms are defined as follows:

$$\begin{cases} E_{xc}^{\circ}[\rho] = \int \rho(\mathbf{r}) e_{xc}^{\circ}(\mathbf{r})\, d\mathbf{r} = \sum_{\mathbf{r}} \rho(\mathbf{r}) e_{xc}^{\circ}(\mathbf{r}), \\ E_{xc}^{\text{RU}}[\rho] = \int \rho(\mathbf{r})\, \overline{e}_{xc}(\mathbf{r})\, d\mathbf{r} = \sum_{\mathbf{r}} \rho(\mathbf{r})\, \overline{e}_{xc}(\mathbf{r}), \\ \sigma[\rho] = \int \rho(\mathbf{r})\, \overline{\sigma}(\mathbf{r})\, \epsilon\, d\mathbf{r} = \sum_{\mathbf{r}} \rho(\mathbf{r})\, \overline{\sigma}(\mathbf{r})\, \epsilon = \sigma\, \epsilon. \end{cases}$$

We adopt a Gaussian likelihood to model the residual uncertainty, which leads to a minimization objective, $\mathcal{L}_{\text{RBNet}} = \frac{1}{N} \sum_i \frac{1}{2}\sigma_i^2 |E_{xc,i}^{\circ} - E_{xc,i}^{*} + E_{xc,i}^{\text{RU}}|^2 + \frac{1}{2} \log \sigma_i^2$, where $E_{xc,i}^{\circ}$ signifies the exchange-correlation energy output from the traditional method for the total reaction (product minus reactant exchange-correlation energy), $E_{xc,i}^{\text{RU}}$ represents our model's prediction, and $E_{xc,i}^{*}$ is defined in Eq.(6). Following the approach (Kendall and Gal 2017), we steer the network to predict the logarithm of the variance, $s_i := \log \sigma_i^2$:

$$\mathcal{L}_{\text{RBNet}} = \frac{1}{N} \sum_i \frac{1}{2} \exp(-s_i) \| E_{xc,i}^{\circ} - E_{xc,i}^{*} + E_{xc,i}^{\text{RU}} \|^2 + \frac{1}{2} s_i. \tag{12}$$

Methods	W4-17			G21EA			G21IP		
	RMSE (↓)	MAE (↓)	MAD (↓)	RMSE (↓)	MAE (↓)	MAD (↓)	RMSE (↓)	MAE (↓)	MAD (↓)
SCAN-D3(BJ)	4.36	6.32	4.63	4.02	3.48	3.50	5.57	4.53	4.53
M06-2X-D3(0)	4.20	3.66	2.28	5.75	1.94	1.84	3.97	2.86	2.80
ωB97X-V	4.97	2.77	3.65	2.31	1.92	1.85	3.83	3.16	3.21
MN15-D3(BJ)	6.19	4.76	4.78	1.89	1.47	**1.21**	4.12	3.16	3.15
B3LYP-D3(BJ)	6.05±2.9E-8	4.68±4.4E-8	2.94±4.4E-8	6.31±4.0E-11	5.04±9.7E-10	3.45±5.4E-10	5.83±1.1E-3	4.09±1.2E-13	3.67±2.8E-13
B3LYP+RBNet	**1.95±4.9E-8**✓	1.29±1.2E-7✓	**1.30±1.4E-7**✓	**1.67±9.5E-4** ✓	**1.38±3.1E-4** ✓	1.35±2.4E-4 ✓	**2.60±3.4E-3**✓	**1.97±1.2E-3** ✓	**1.79±9.6E-4** ✓
	+67.77%	+72.44%	+55.78%	+73.53%	+72.62%	+60.87%	+55.40%	+51.83%	+51.23%
DM21	3.41±1.3E-2	2.36±1.3E-2	2.13±4.9E-3	2.66±7.1E-7	2.09±1.9E-6	1.37±8.4E-3	9.97±1.9E-2	5.05±2.9E-2	6.54±7.4E-2
DM21+RBNet	2.45±1.2E-5✓	1.74±5.5E-4✓	1.32±1.3E-5✓	2.34±2.6E-3 ✓	1.74±5.5E-4 ✓	1.33±9.3E-4 ★	2.86±3.0E-3✓	2.01±7.6E-4 ✓	2.10±1.3E-3 ✓
	+28.15%	+16.75%	+38.03%	+12.03%	+16.75%	+2.92%	+71.31%	+60.20%	+67.89%

Table 1: State-of-the-art performance (kcal/mol) by our RBNet (residual XC-uncertain functional) on W4-17, G21EA and G21IP datasets. We can see that our RBNet outperforms both B3LYP and DM21 significantly, validating its robustness and generalizability. The **bold** and underline font indicate the best and second-best outcomes. The standard deviation is presented in scientific notation. The symbol ✓ denotes successful passage of Welch's t-test at a significance level of 0.05, whereas ★ indicates successful passage at a significance level of 0.10.

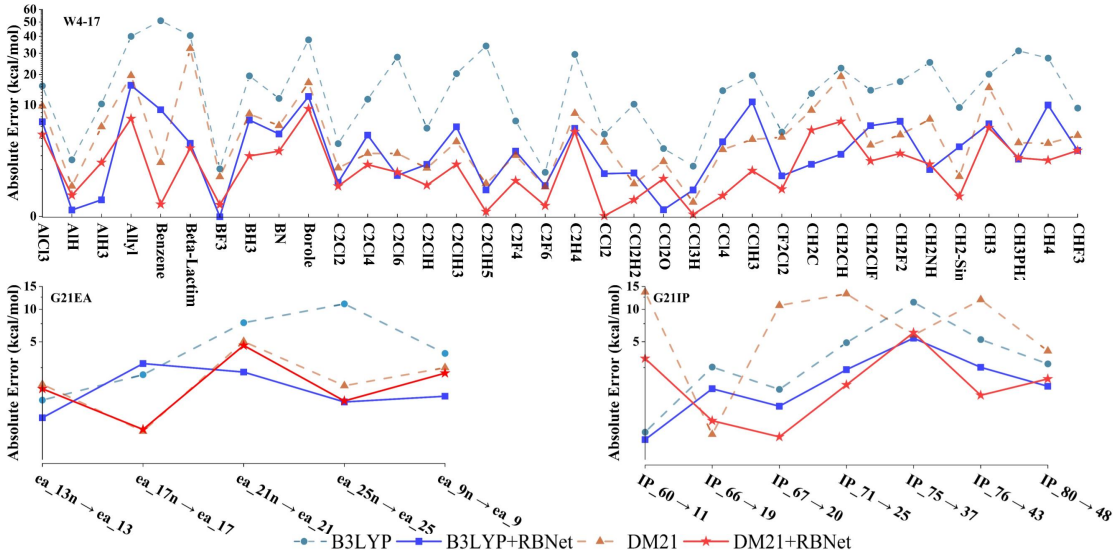

Figure 2: Absolute errors (in kcal/mol) relative to the accurate energy computed from literature.

Experiments

In this section, we assess the efficacy of our methods using chemical reactions across three well-known datasets: W4-17, G21EA, and G21IP. We present the details of the experimental setup, highlight the key findings, and conduct a comprehensive parameter analysis, all of which collectively underscore the effectiveness and superiority of our RBNet.

Experimental Setup

Datasets. We consider three subsets, W4-17, G21EA and G21IP within the GMTKN55 dataset (Goerigk and Grimme 2010). The remaining portion of the analysis was performed using our proprietary single-atom energy dataset known as Atom.**W4-17** (Karton, Sylvetsky, and Martin 2017) is a comprehensive collection of total atomisation energies for 200 species. These energies describe the transformational energy that occurs when one mole of a species is completely dissociated into its individual constituent atoms in the gas phase. Put simply, the total atomisation energy signifies the energy change associated with breaking all bonds within a species and converting all its atoms into isolated entities. Furthermore, **G21EA** (Curtiss et al. 1991) encompasses 25 reactions involving adiabatic electron affinities, while **G21IP** (Curtiss et al. 1991) incorporates 36 reactions associated with adiabatic ionisation potentials. These two datasets capture the energy changes that occur as species either gain or lose a single electron. Additionally, we computed data for isolated atoms that are implicated within these datasets, contributing to the control of training outcomes.

Partition Settings. We partition the datasets into training, validation, and test sets using a 6:2:2 ratio. All related atoms are included in the training set, encompassing both individual atoms and molecules consisting solely of single elements found within the dataset. Subsequently, we address each atom type based on descending atomic numbers. Should a given atom not yet be represented in the training set, a molecule

containing that atom is randomly selected and added to the training set. This process ensures the inclusion of molecules featuring atom counts of 2, 3, 4, 5, 6, and those exceeding 6. Lastly, the remaining molecules are distributed randomly among the training, validation, and test sets until the desired dataset sizes are achieved.

Implemented details. Our RBNet is a single network with two heads and its structure is $16 - 128 - 256 - 256 - 256 - 128 - \begin{cases} 50 - 1 \\ 50 - 1 \end{cases}$. We train the RBNet using the loss in the Eq.(12). The parameters are set to $k_1 = 1$, $k_2 = 1$. For model optimization, we use the SGD optimizer and set an initial learning rate of $1e^{-3}$ with a cosine annealed schedule. We train all methods — epochs. We implement our model using Python 3.7.16 with Pytorch 1.12.1 on a server equipped with Intel(R) Xeon(R) Gold 6230R CPU and GeForce RTX 3090.

Evaluation metrics. For all datasets, we utilize the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Deviation (MAD) as evaluation criteria to assess the performance of all methods. RMSE quantifies the square root of the average squared disparity between predicted and actual values within a dataset. MAE gauges the mean absolute disparity between predicted and actual values in a dataset. A lower MAE indicates a better fitting model for the dataset. MAD holds particular value as a metric due to its resilience to outliers compared to alternative measures of dispersion like standard deviation and variance.

Compared methods. To validate the viability of our approach, we conduct a comparative analysis against six state-of-the-art techniques as outlined below:

- SCAN-D3(BJ) (Sun, Ruzsinszky, and Perdew 2015) stands as a meta-generalized-gradient approximation (meta-GGA), compliant with all 17 exact constraints.
- M06-2X-D3(0) (Zhao and Truhlar 2008) introduces a highly nonlocal functional featuring double the nonlocal exchange content (2X).
- MN15-D3(BJ) (Yu et al. 2016) presents a hybrid XC functional that amalgamates nonlocal Hartree–Fock exchange energy, nonseparable local exchange–correlation energy, and additional correlation energy.
- ωB97X-V (Mardirossian and Head-Gordon 2014) boasts a configuration with 7 linear parameters (2 for local exchange, 4 for local correlation, and 1 for short-range exact exchange) and 3 nonlinear parameters (1 for range-separation and 2 for non-local correlation), totaling 10 optimized parameters.
- B3LYP-D3(BJ) (Becke 1993), one of the most prevalent hybrid XC functionals employed in DFT calculations, is defined by Eq.(1).
- DM21 (Kirkpatrick et al. 2021) predicts three local enhancement factors, blending the local density approximation e_{xc}^{LDA}, local Hartree Fock exchange energy densities $e^{\omega_1 \mathrm{HF}}$, and $e^{\omega_2 \mathrm{HF}}$ as expressed in Eq.(2).

In addition, our method interfaces with two established techniques. B3LYP+RBNet combines our residual uncertain XC functional with B3LYP by adopting it as the conventional XC functional. DM21+RBNet further merges our residual uncertain XC functional with DM21 as the conventional XC functional. Note that D3(BJ) and D3(0) are the basis of the wave function in the DFT calculation.

Features	W4-17			G21EA			G21IP		
	RMSE	MAE	MAD	RMSE	MAE	MAD	RMSE	MAE	MAD
$\mathbf{x}(\mathbf{r})$	2.35	3.27	2.16	3.51	3.23	2.75	3.47	2.82	2.92
$\mathbf{y}(\mathbf{r})$	**1.95**	**1.29**	**1.30**	**2.82**	**1.38**	**1.35**	**2.54**	**1.92**	**1.75**

Table 2: Ablation study of the features $\mathbf{y}(\mathbf{r})$ and $\mathbf{x}(\mathbf{r})$.

Methods	W4-17			G21EA			G21IP		
	RMSE	MAE	MAD	RMSE	MAE	MAD	RMSE	MAE	MAD
B3LYP (Baseline)	6.05	4.68	2.94	6.32	5.04	3.45	5.31	4.02	3.66
+ResNet	3.94	2.71	2.27	3.27	2.62	1.71	3.74	2.97	2.74
+RBNet	**1.95**	**1.29**	**1.30**	**2.82**	**1.38**	**1.35**	**2.54**	**1.92**	**1.75**

Table 3: Ablation study of the ResNet and our RBNet.

Main Results

Error Results. Table 1 presents the error results. Our RBNet, a residual uncertain XC functional, exhibits remarkable improvements across all datasets compared to the six XC functionals. Further insights are as follows.

Firstly, it is evident that both our B3LYP+RBNet and DM21+RBNet functionals outperform all state-of-the-art XC functionals in terms of RMSE, MAE, and MAD on the W4-17 and G21IP datasets. For instance, B3LYP+RBNet surpasses DM21 by over 1.27 kcal/mol on W4-17 and ωB97X-V by more than 1.29 kcal/mol on G21IP. Moreover, our approach demonstrates comparability to MN15-D3(BJ) on the G21EA dataset. Secondly, B3LYP+RBNet exhibits superior capabilities in comparison to our baseline B3LYP, showcasing substantial RMSE improvements of 4.10, 3.50, and 2.77 kcal/mol on the three datasets, respectively. Similar observations apply to DM21+RBNet. This emphasizes RBNet's potential for generalization to other related methods. Fig.2 visually depicts the errors (in kcal/mol) relative to the accurate energies computed from literature for each test reaction.

Visualization. We utilize visualization techniques to analyze the spatial distribution of residuals acquired through our RBNet on both the XC energy and the XC potential. To gain a comprehensive understanding of our RBNet's performance, we focus on the XC energy and potential residuals. The results are illustrated in Figure3, which shows the local residual distribution between B3LYP and B3LYP+RBNet in the sulfanyl (HS) molecule for $\log(\overline{e}_{xc})$ and $\log(\overline{v}_{xc})$. Likewise, Figure4 shows the distribution of local residuals between DM21 and DM21+RBNet in the ethyne (C_2H_2) molecule for $\log(\overline{e}_{xc})$ and $\log(\overline{v}_{xc})$. The spatial distribution unmistakably reveals elevated residual values around the nuclei and along the molecular boundaries. Additionally, a discernible spatial symmetry is observed in the distribution, centered around the molecular structure. **Ablation Study.** Since we design a new feature $\mathbf{y}(\mathbf{r})$ and propose the RBNet, we conduct two ablation studies to validate their effectiveness, utilizing B3LYP as the baseline. First, comparing our feature $\mathbf{y}(\mathbf{r})$ to

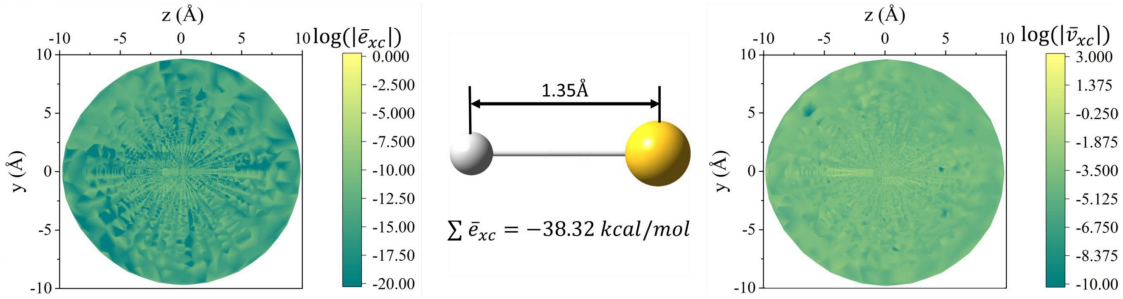

Figure 3: Spatial distribution of the residuals learned by our RBNet using B3LYP for sulfanyl (HS) in the G21EA dataset. The left and right panels show the spatial distribution of the residuals for $\log(\bar{e}_{xc})$ and for $\log(\bar{v}_{xc})$, while the middle panel displays the molecular geometry of HS. The highest local residuals are in the region around the nuclei and at the boundaries.

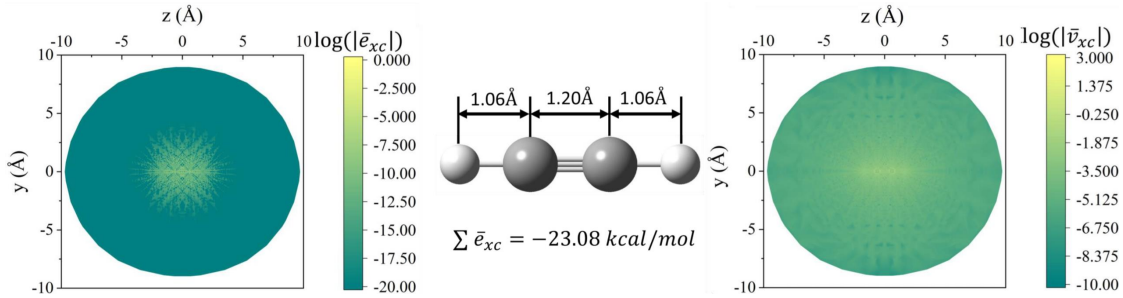

Figure 4: Spatial distribution of the residuals learned by our RBNet using DM21 for ethyne (C_2H_2) in the W4-17 dataset. The left and right panels show the spatial distribution of the residuals for $\log(\bar{e}_{xc})$ and for $\log(\bar{v}_{xc})$, while the middle panel displays the molecular geometry of C_2H_2. The highest local residuals are in the region around the nuclei and at the boundaries.

the feature $\mathbf{x}(\mathbf{r})$ from DM21, Table 2 illustrates that, with B3LYP+RBNet, our $\mathbf{y}(\mathbf{r})$ outperforms $\mathbf{x}(\mathbf{r})$ across all three datasets. Second, by incorporating a ResNet without variance and our proposed RBNet to learn XC functionals, denoted as B3LYP+ResNet and B3LYP+RBNet, respectively, Table 3 reports the results to show that our RBNet exhibits the capability to capture superior XC functionals, thus enhancing the accuracy of DFT.

Parameter Analysis. Our network incorporates two essential hyperparameters: k_1 and k_2. The role of k_1 is to define the upper limit for the mean of the correction, while k_2 serves to set the upper limit for the variance of the correction. When they increase or decrease, the mean and variance will get bigger or smaller, resulting in an explosion or no-change of the sum energy. To identify optimal values, we employ a control variable approach. In Fig.5, during practical implementation, we ascertain that $k_1 = 1$ and $k_2 = 1$.

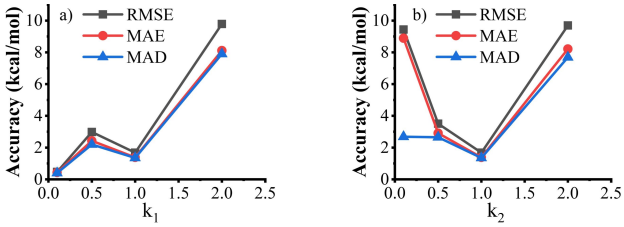

Figure 5: Parameter analysis of two parameters k_1 and k_2. Left shows the results of k_1 varied from 0.1 to 2.0 when $k_2 = 1$, while right showcase the results of k_2 varied from 0.1 to 2.0 when $k_1 = 1$.

Conclusion

In this paper, we introduced a novel uncertain format for the XC Functional, referred to as the General XC-Uncertain Functional, which involves training a single neural network to determine both the mean and variance. Nevertheless, directly learning these parameters is challenging due to issues such as non-convergence in integration results or the potential for training loss explosion. To overcome this, we proposed a solution involving a residual XC-uncertain functional. It was achieved by introducing a residual Bayesian network, designed to learn the residual approximation between the ideal (ground-truth) functional and the conventional XC functional. Our experimental findings demonstrated the effectiveness of our approach, showcasing its superiority over existing state-of-the-art XC functionals, like B3LYP and DM21.

Moreover, there are two directions for investigation: 1) *Enhancing Accuracy:* Despite significant strides in minimizing the disparity between our predictions and the ground-truth values, additional efforts are imperative to refine the precision of our predictions further. 2) *Generalization:* Our method is inherently adaptable for extension to accommodate other XC functionals. This underscores the versatility and potential broader applicability of our approach.

Acknowledgements

Jun Li was partially supported by the National Science Fund of China under Grant Nos. 62072242.

References

Aldegunde, M.; Kermode, J. R.; and Zabaras, N. 2016. Development of an exchange–correlation functional with uncertainty quantification capabilities for density functional theory. *J. Chem. Phys.*, 311: 173–195.

Becke, A. D. 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38: 3098.

Becke, A. D. 1993. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98: 5648.

Burke, K. 2012. Perspective on density functional theory. *J. Chem. Phys.*, 136: 150901.

Chai, J.-D.; and Head-Gordon, M. 2008. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.*, 128: 084106.

Cohen, A. J.; Mori-Sanchez, P.; and Yang, W. 2012. Challenges for Density Functional Theory. *Chem. Rev.*, 112: 289–320.

Cuierrier, E.; Roy, P.-O.; and Ernzerhof, M. 2021. Constructing and representing exchange–correlation holes through artificial neural networks. *J. Chem. Phys.*, 155: 174121.

Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; and Pople, J. A. 1991. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.*, 94: 7221–7230.

Dick, S.; and Fernandez-Serra, M. 2020. Machine learning accurate exchange and correlation functionals of the electronic density. *Nature Commun.*, 11: 3059.

Goerigk, L.; and Grimme, S. 2010. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions, Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *J. Chem. Theory Comput.*, 6: 107–126.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. *CVPR*, 770–778.

Hohenberg, P.; and Kohn, W. 1964. Inhomogeneous electron gas. *Phys. Rev.*, 136: B864.

Karton, A.; Sylvetsky, N.; and Martin, J. M. L. 2017. W4-17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. *J. Chem. Theory Comput.*, 38: 2063–2075.

Kasim, M.; and Vinko, S. 2021. Learning the Exchange-Correlation Functional from Nature with Fully Differentiable Density Functional Theory. *Phys. Rev. Lett.*, 127: 126403.

Kendall, A.; and Gal, Y. 2017. What uncertainties do we need in bayesian deep learning for computer vision? *NeurIPS*, 30.

Kim, H.; Park, J. Y.; and Choi, S. 2019. Energy refinement and analysis of structures in the QM9 database via a highly accurate quantum chemical method. *Sci. Data*, 6: 109.

Kirkpatrick, H.; McMorrow, B.; Turban, D. H. P.; Gaunt1, A. L.; Spencer, J. S.; Matthews, A. G. D. G.; Obika, A.; Thiry, L.; Fortunato, M.; Pfau, D.; Castellanos, L. R.; Petersen, S.; Nelson, A. W. R.; Kohli, P.; Mori-Sánchez, P.; Hassabis, D.; and Cohen, A. J. 2021. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science*, 374: 1385–1389.

Kohn, W.; and Sham, L. J. 1965. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140: A1133.

Lee, C.; Yang, W.; and Parr, R. G. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37: 785.

Li, L.; Hoyer, S.; Pederson, R.; Sun, R.; Cubuk, E. D.; Riley, P.; and Burke, K. 2021. Kohn-Sham Equations as Regularizer: Building Prior Knowledge into Machine-Learned Physics. *Phys. Rev. Lett.*, 126: 036401.

Mardirossian, N.; and Head-Gordon, M. 2014. xB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.*, 16: 9904.

Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; and Lyssenko, K. A. 2017. Density functional theory is straying from the path toward the exact functional. *Science*, 355: 49–52.

Mortensen, J. J.; Kaasbjerg, K.; Frederiksen, S. L.; Nørskov, J. K.; Sethna, J. P.; and Jacobsen, K. W. 2005. Bayesian Error Estimation in Density-Functional Theory. *Phys. Rev. Lett.*, 95: 216401.

Nagai, R.; Akashi, R.; and Sugino, O. 2022. Machine-learning-based exchange correlation functional with physical asymptotic constraints. *Phys. Rev. Res.*, 4: 013106.

Palos, E.; Lambros, E.; Dasgupta, S.; and Paesani, F. 2022. Density functional theory of water with the machine-learned DM21 functional. *J. Chem. Phys.*, 156: 161103.

Pederson, R.; Kalita, B.; and Burke, K. 2022. Machine learning and density functional theory. *Nature Rev. Phys.*, 4: 357–358.

Perdew, J. P. 1986. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B*, 33: 8822(R).

Pokharel, K.; Furness, J. W.; Yao, Y.; Blum, V.; Irons, T. J.; Teale, A. M.; and Sun, J. 2022. Exact constraints and appropriate norms in machine learned exchange-correlation functionals. *J. Chem. Phys.*, 157(17): 174106.

Pracht, P.; Grant, D. F.; and Grimme, S. 2020. Comprehensive Assessment of GFN Tight-Binding and Composite Density Functional Theory Methods for Calculating Gas-Phase Infrared Spectra. *J. Chem. Theory Comput.*, 16(11): 7044–7060.

Schmidt, J.; Benavides-Riveros, C. L.; and Marques, M. A. L. 2019. Machine Learning the Physical Nonlocal Exchange Correlation Functional of Density-Functional Theory. *J. Phys. Chem. Lett.*, 10: 6425–6431.

Sun, J.; Ruzsinszky, A.; and Perdew, J. P. 2015. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.*, 115: 036402.

van Santen, J. A.; and DiLabio, G. A. 2015. Dispersion Corrections Improve the Accuracy of Both Noncovalent and Covalent Interactions Energies Predicted by a Density-Functional Theory Approximation. *J. Phys. Chem. A*, 119(25): 6703–6713.

Vosko, S. H.; Wilk, L.; and Nusair, M. 1980. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian J. of Phys.*, 58(8): 1200–1211.

Yu, H. S.; He, X.; Li, S. L.; and Truhlar, D. G. 2016. MN15: a Kohn–Sham global-hybrid exchange–correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chem. Sci.*, 7: 5032–5051.

Zhao, Y.; and Truhlar, D. G. 2008. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, 120: 215–241.

Zhou, Y.; Wu, J.; Chen, S.; and Chen, G. 2019. Toward the Exact Exchange Correlation Potential: A ThreeDimensional Convolutional Neural Network Construct. *J. Phys. Chem. Lett.*, 10: 7264–7269.