Density Functional Theory as a Data Science

Tsuneda, Takao

(Citation)
Chemical Record,20(7):618-639

(Issue Date)
2020-07

(Resource Type)
journal article

(Version)
Accepted Manuscript

Density Functional Theory as a Data Science

Takao Tsuneda*

Graduate School of Science, Technology, and Innovation, Kobe University, Kobe 657-8501, Japan

Abstract

The development of density functional theory (DFT) functionals and physical corrections are reviewed focusing on the physical meanings and the semiempirical parameters from the viewpoint of data science. This review shows that DFT exchange-correlation functionals have been developed under many strict physical conditions with minimizing the number of the semiempirical parameters, except for some recent functionals. Major physical corrections for exchange-correlation functionals are also shown to have clear physical meanings independent of the functionals, though they inevitably require minimum semiempirical parameters dependent on the functionals combined. We, therefore, interpret that DFT functionals with physical corrections are the most sophisticated target functions that are physically legitimated, even from the viewpoint of data science.

*Electronic address: tsuneda@phoenix.kobe-u.ac.jp

I. INTRODUCTION

Density functional theory (DFT) is the central theory in current quantum chemistry (QC) [1]. Though DFT was developed in the field of materials science [1, 2], it has been applied in QC calculations since mid 1990s, and it is used in more than 90 % of QC calculations as of 2019. The cause for the high-frequency use comes from the high cost performance of DFT calculations, i.e., the high accuracy close to chemical accuracy, which is within the error of several kcal/mol in energy, and the low computational time compared to those of *ab initio* wavefunction theories [1]. In the Kohn-Sham equation, which is the basic function of DFT, the difference from the *ab initio* Hartree-Fock (HF) equation is only exchange-correlation potential part [1, 2]. The performance of DFT calculations, therefore, depend on the quality of DFT exchange-correlation functionals used.

Development of DFT exchange and correlation functionals has close relations to data science [3]. DFT functionals are usually developed in the following steps:

1. Collecting physical effects contributing to the electronic states of systems,
2. Determining DFT functional by models combining the physical effects with semiempirical parameters, and
3. Sophisticating DFT functionals by increasing the number of physical effects for improving their applicability and then by minimizing the number of semiempirical parameters for removing artificial problems such as the local minima in calculated potential energy surfaces.

These steps are obviously very close to the steps of determining target functions in data science:

1. Collecting features (independent variables) contributing to predictions,
2. Determining target functions by various data-scientific models using many features with parameters, and
3. Sophisticating the models by increasing the number of the features for improving the generalization performance and then by minimizing them and accompanying parameters for avoiding the overfitting.

Therefore, from the viewpoint of data science, DFT functionals are the target functions, whose features are physical effects, for reproducing e.g., chemical properties, reaction processes in chemistry. Considering that DFT now dominates QC calculations, it is not too much to say that DFT functional development is the most successful data science in chemistry.

This analogy between DFT functional development and data science suggests various methods for sophisticating DFT functionals. In the case of low generalization performance due to the lack of learning called "underfitting", new features are added by collecting the new datasets of these features in data science. However, since it is usually difficult to prepare enough datasets, there are several methods for virtually increasing the number of data [3]: e.g., "holdout", which splits the datasets into the training and test datasets, and its "cross validation", which crosses the roles of the training and test datasets. These methods are applicable to the DFT functional development. As a sophisticated data-scientific method for improving the generalization performance, there is "ensemble learning", which combines a number of models to enhance the prediction [3]: e.g., "bootstrap aggregating (bagging)" and "boosting". For improving the generalization performance, the bagging first performs the bootstrap, in which the dataset of a certain number of rows is repeatedly sampled with replacement allowing overlaps from the original learning dataset of the same number of rows to generate new datasets, and then constructs the model of each new dataset called "weak learner", and finally aggregates the prediction results. The representative bagging is "random forest", which uses decision trees as weak learners. On the other hand, the boosting firstly constructs a model for the original learning dataset, secondly compares the predictions and correct answers to figure out incorrectly-predicted samples, thirdly makes new learning datasets by weighting the incorrectly-predicted samples, and finally returns to the construction of models repeatedly. The representative boosting is "gradient boosting", which uses the gradient of the errors to determine the weight of sample datasets. There are many DFT functionals combining different types of functionals and corrections [3]: e.g., semiempirical functionals. Note that the combination of the different corrections should not use the weighting for accuracy, because the physical corrections have physical meanings as a whole, and therefore, the weighting of the corrections provide artificial effects on the electronic states. On the other hand, DFT functionals actually contain physically-undefined parts such as the large density gradient parts of generalized-gradient-approximation (GGA)

functionals, as mentioned in Sec. IV A. The ensemble learning could be an efficient tool for improving these parts of DFT functionals. Therefore, the data-scientific methods give new tools for improving the applicability of DFT functionals.

In the case of the "overfitting" of models due to the excessive number of features compared to the number of datasets, dimension reduction is performed to decrease the number of features for avoiding the increase of the upper limit of generalization errors in data science. The dimension reduction is classified into two types [3]: one is "feature selection" choosing the subsets of features and another is "feature extraction" transforming a feature space axis to another one. The most frequently-used feature selection is "stepwise regression", which adds important features from the set of explanatory variables and removes unnecessary features stepwise. In the DFT functional development, the stepwise regression corresponds to the selection of significant physical effects for reproducing calculation systems such as reaction systems under a threshold. For example, relativistic effects are usually insignificant to reproduce light atom systems, though they are necessary to reproduce heavy atom systems. The most famous feature extraction is "principal component analysis", in which significant features are figured out using the orthogonal transformation of the set of possibly correlated variables to the set of linearly uncorrelated variables (principal components). In DFT functional development, the double counting of the same physical effects should be removed to enhance the applicability. The long-range correction (LC), for instance, avoids the double counting of exchange effects by separating exchange interactions into short- and long-range parts and assigning them into a DFT functional and the HF exchange integral (see Sec. V A). Therefore, the data-scientific methods also have a potential for evaluating the double counting rates of physical effects in DFT functionals and corrections.

As mentioned above, the analogy of DFT functional development and data science is beneficial to consider the strategy for enhancing DFT functional development or data science. After considering requirements for the development of machine-learned DFT, we explore several major DFT functionals and their physical corrections from the viewpoint of data science and consider their advantages and disadvantages as the features of a target integrated functional. We anticipate that this review will be the first step to integrate DFT and data science in near future.

II. REQUIREMENTS FOR DEVELOPING MACHINE-LEARNED DFT

Data science has, so far, been applied to various predictions in theoretical chemistry calculations [4]: e.g., energies and chemical properties, reaction models, potential energy surfaces, and interatomic potentials and analyses in molecular dynamics simulations. However, this section focuses only on the requirements for making the best use of machine learning in the development of DFT functionals, because DFT functionals account for all these predictions and, therefore, have a wide range of applications. For making use of machine learning in the functional developments, it is required to answer the following questions:

1. What should be chosen as the descriptors of DFT functionals?

2. What points are required for implemented parameters?

3. How significant are the fundamental conditions for the accuracy of predictions?

4. What should be targeted by DFT functionals as prediction models?

Let us consider the answers for these questions.

Since the descriptors of machine learnings should be independent as much as possible to enhance the generalization performance of predictions. By analogy, DFT functionals are also required to consist of independent elements to the maximum extent possible. DFT functionals are usually composed of exchange and correlation functionals and their corrections. Therefore, these functionals and corrections should be independent of each other to develop DFT functionals of high generalization performance. This indicates that exchange-correlation functionals should be derived in physically well-defined manners and the physically meaningful corrections should be incorporated without double counting.

Parameters of machine-learned functions should be minimized in number and have physical meanings. As mentioned in Sec. I, the parameters are minimized in the final phase of machine learnings to avoid the overfitting of prediction models It has actually been reported that the increase in the number of parameters in DFT functionals causes spurious local minima in calculated potential energy surfaces [5]. In machine learnings, principal features of prediction models are screened by regression coefficients. If the parameters have no physical meanings, it is difficult to interpret the regression coefficient and to make clear the cause

for the significance of the principal features, which is needed to sophisticate the prediction models, i.e., DFT functionals.

The significance of the fundamental conditions in the functionals has also been emphasized. Medvedev et al. [6, 7] showed that semiempirical functionals worsen the reproducibility of electron density and indicated that these functionals do not go up the Jacob's ladder toward universal functional due to the violation of the fundamental conditions. Hollingsworth et al. [8] investigated the effect of the fundamental conditions on machine-learned functionals and consequently found the improvement of the approximation to a kinetic energy functional. From the viewpoint of machine learning, it is natural to use the fundamental conditions as a constraint for the functionals, because it is an effective strategy for enhancing the accuracy of prediction models to decrease the number of descriptors by imposing reasonable conditions to the prediction models.

For obtaining high generalization performance of functionals, evaluation criteria should be reviewed for the development of conventional functionals, though the errors of the calculated chemical properties in various benchmark sets have been uncritically adopted. Except for the DFT functional developments, the most frequently-used data science in the data science for chemistry is chemoinformatics [9], in which the reaction prediction models are constructed by the learning of bond species and chemical property values. Following the philosophy of chemoinformatics, many reaction prediction models have been developed by the machine learnings using the calculated properties of DFT to estimate reactivities without DFT calculations [10, 11]. In recent years, several machine-learned functionals have also been developed [12–15] especially for kinetic energy functionals [8, 16, 17]. Most functionals including machine-learned ones have been obtained by minimizing the errors for many molecular chemical properties of various benchmark sets. However, Pernot and Savin [18] recently questioned this strategy by showing that "summary statics", e.g., root mean square errors and mean absolute errors, are inappropriate as the criteria of prediction models. They also pointed out that atomization energies should not be included in the training datasets to improve the generalization performance of functionals due to their serious dependence on molecular sizes. This indicates that the detail review is necessary for choosing the evaluation criteria and benchmark sets in the datasets of machine learning.

III. CLASSIFICATION OF DFT FUNCTIONALS AND THEIR PHYSICAL CORRECTIONS

The basic equation of DFT is the Kohn-Sham equation [2], in which the Fock operator $\hat{F}$ contains an exchange-correlation functional,

$$\hat{F}\phi_i = \epsilon_i \phi_i, \tag{1}$$

$$\hat{F} = \hat{h} + 2\sum_j^n \hat{J}_j + v_{\rm xc}, \tag{2}$$

where ϕ_i and ϵ_i are the i-th orbital and orbital energy, respectively, n is the number of occupied orbitals, $\hat{h}$ is the one-electron Hamiltonian operator, $\hat{J}_j$ is the Coulomb operator for the j-th orbital, and $v_{\rm xc}$ is the exchange-correlation potential functional. For this Fock operator, the total electronic energy is given as

$$E = \sum_i^n \left(h_i + 2\sum_j^n J_{ij} \right) + E_{\rm xc}, \tag{3}$$

where $E_{\rm xc}$ is the exchange-correlation energy functional, which defines the exchange-correlation potential functional,

$$v_{\rm xc} = \frac{\delta E_{\rm xc}}{\delta \rho}, \tag{4}$$

and the one-electron and Coulomb integrals, h_i and J_{ij}, are provided as

$$h_i = \int d^3\mathbf{r} \phi_i^*(\mathbf{r}) \left\{ -\frac{1}{2}\boldsymbol{\nabla}^2 + V_{\rm ne}(\mathbf{r}) \right\} \phi_i(\mathbf{r}), \tag{5}$$

$$J_{ij} = \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 \phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\frac{1}{r_{12}}\phi_i(\mathbf{r}_1)\phi_j(\mathbf{r}_2), \tag{6}$$

where $\mathbf{r}i$ is the coordinate of i-th electron and $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$. In the generalized Kohn-Sham method [19], the Kohn-Sham equation is also established for the inclusion of orbital-dependent potentials such as the HF exchange integral,

$$\hat{O}^S[\{\phi_i\}]\phi_j + v_{\rm eff}\phi_j = \epsilon_j \phi_j, \tag{7}$$

where $\hat{O}^S$ is the orbital-dependent operator in the Fock operator and $v_{\rm eff}$ is the effective potential functional for the remaining orbital-independent part, which is usually the density functional part of the exchange-correlation potential functionals. Note that the Fock operator of the (generalized) Kohn-Sham equation is the same as that of the HF equation except

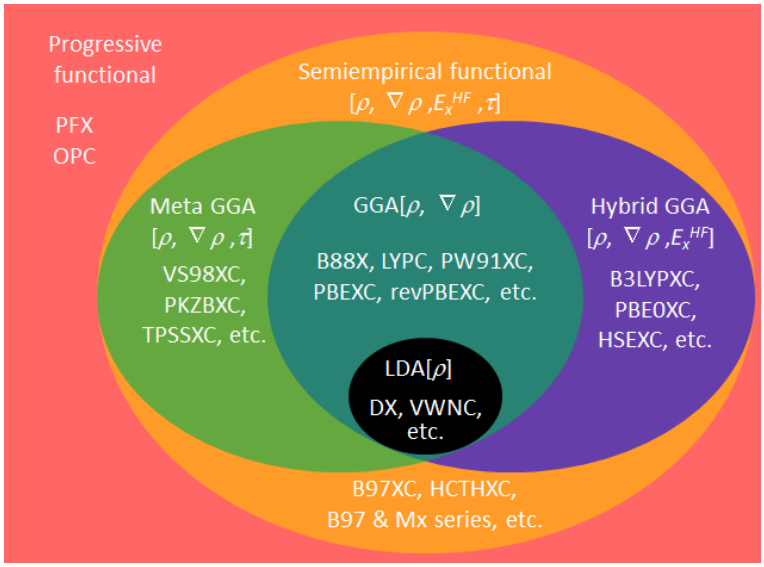

FIG. 1: Classification of major exchange-correlation functionals. The suffix "X" and "C" indicate exchange and correlation functionals, respectively.

for the exchange-correlation functional part. Therefore, the accuracy of the (generalized) Kohn-Sham method depends only on the quality of the exchange-correlation functionals used.

Exchange-correlation functionals can be classified on the basis of their features. Figure 1 shows a usual classification of major exchange-correlation functionals. As shown in the figure, the major exchange-correlation functionals are roughly categorized into six types:

- Local density approximation (LDA) functionals of only electron density ρ: $E_{\mathrm{LDA}}[\rho]$,
- Generalized gradient approximation (GGA) functionals correcting LDA functionals with the density gradient $\boldsymbol{\nabla}\rho$: $E_{\mathrm{GGA}}[\rho, \nabla\rho]$,
- Meta-GGA functionals correcting GGA functionals with the kinetic energy density τ and/or the density Laplacian $\nabla^2\rho$: $E_{\mathrm{mGGA}}[\rho, \nabla\rho, \tau, \nabla^2\rho]$,
- Hybrid functionals mixing the HF exchange integral ($E_{\mathrm{x}}^{\mathrm{HF}}$) at a constant ratio: $E_{\mathrm{hGGA}}[\rho, \nabla\rho, E_{\mathrm{x}}^{\mathrm{HF}}]$,
- Semiempirical functionals that are developed to reproduce accurate properties with many semiempirical parameters: $E_{\mathrm{semiemp}}[\rho, \nabla\rho, \tau, E_{\mathrm{x}}^{\mathrm{HF}}]$, and

- Progressive functionals transforming in accordance with τ functional and exchange integral kernel K_σ for exchange and correlation functionals, respectively: $E_{\mathrm{x}}[\tau]$, $E_{\mathrm{c}}[K_\sigma]$.

There are various other types of functionals: e.g., functionals explicitly containing the set of orbitals $\{\phi_i\}$. Note, however, that as mentioned in Sec. V, most of them are assigned to the physical corrections. We, therefore, consider that exchange-correlation functionals are basically classified by these features.

Exchange-correlation functionals that do not fall into this classification mostly correspond to the physical corrections for these functionals. The physical corrections are similar to the complement of new features in data science, though the formers are essentially required when they have significant effects. Major physical corrections are given as follows:

- Long-range correction complementing long-range electron-electron exchange interactions for exchange functionals,
- Self-interaction corrections removing the self-interaction error of exchange functionals,
- Van der Waals corrections supplementing van der Waals interactions for correlation functionals, and
- Relativistic corrections complementing relativity theory-based effects on electronic motions especially in kinetic energies.

Note that the corrections for kinetic energies are often interpreted as the corrections for correlation functionals due to the definition of electron correlation [20]. The characteristics of these corrections are summarized as follows:

- Correcting for clearly-defined physical effects that should be essentially incorporated,
- Being independent of the formulations of corrected functionals, and
- Requiring no or only a few semiempirical parameters to be incorporated.

In Secs. IV and V, the classified exchange-correlation functionals and the physical corrections are sampled, respectively, focusing on the physical meaning of semiempirical parameters. Note that this review intends to establish the connection between the DFT functional development and data science. In the field of DFT, huge number of functionals have been

developed [1]: Major functionals are semiempirical functionals such as Minnesota-series (Mx) functionals [21, 22], which intend to reproduce accurate properties using as many semi-empirical parameters as needed, and double-hybrid functionals [23], which combines GGA functionals with both HF exchange integrals and perturbation correlation energies. For related methods, optimized effective potential (OEP) method [24, 25], which intends to provide the orbital-dependent effective potentials corresponding to various methods such as perturbation theory, and constrained search method [26, 27], which determines potentials directly from the highly accurate electron densities, have been frequently used. However, in this review, only popular functionals and related methods, most of which are implemented in the current packages of Gaussian [28] and GAMESS [29], are introduced to focus on the requisite for functionals and physical corrections in data science, i.e., the minimum and clear physical meaning of parameters.

IV. EXCHANGE-CORRELATION FUNCTIONALS

In this section, several major exchange-correlation functionals are explored from the viewpoint of data science as the models of exchange and correlation energies focusing on incorporated semiempirical parameters. Following the classification shown in Sec. III, six types of exchange-correlation functionals are discussed for the physical meanings of the functional formulations and the semiempirical parameters.

A. LDA and GGA exchange functionals

LDA exchange functional essentially has no semiempirical parameter, because it has the exact form called "Dirac exchange functional",

$$E_{\mathrm{x}}^{\mathrm{LDA}} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int d^3\mathbf{r}\rho^{4/3}(\mathbf{r}). \tag{8}$$

This functional is used as the local density limit of GGA exchange functionals, as mentioned later. However, this LDA functional often has one semiempirical parameter to alternate to the HF exchange integral to be mainly used in solid state calculations [30]. In this parametrized functional called "Xα functional", a semiempirical parameter α is multiplied to the exchange potential in order to improve the underestimation of exchange effects. This semiempirical parameter is, therefore, a practical tool and has no physical meanings.

GGA exchange functionals are generally modeled by the functions of dimensionless parameter,

$$x_\sigma = \frac{|\boldsymbol{\nabla}\rho_\sigma|}{\rho_\sigma^{4/3}}, \tag{9}$$

in the K_σ of the general functional form,

$$E_{\mathrm{x}}^{\mathrm{GGA}} = -\frac{1}{2}\sum_\sigma \int \rho_\sigma^{4/3} K_\sigma d^3\mathbf{r}. \tag{10}$$

That is, GGA functionals are different only for the dependence on x_σ. Different to models in data science, GGA exchange functional models are physically regulated by the fundamental conditions for exchange energy. Figure 2 displays the K_σs of four major GGA exchange functionals: the Becke 1988 (B88) [31], the Perdew-Wang 1991 (PW91) [32, 33], the Perdew-Burke-Ernzerhof (PBE) [5], and the revised PBE (revPBE) [34] functionals. As shown in

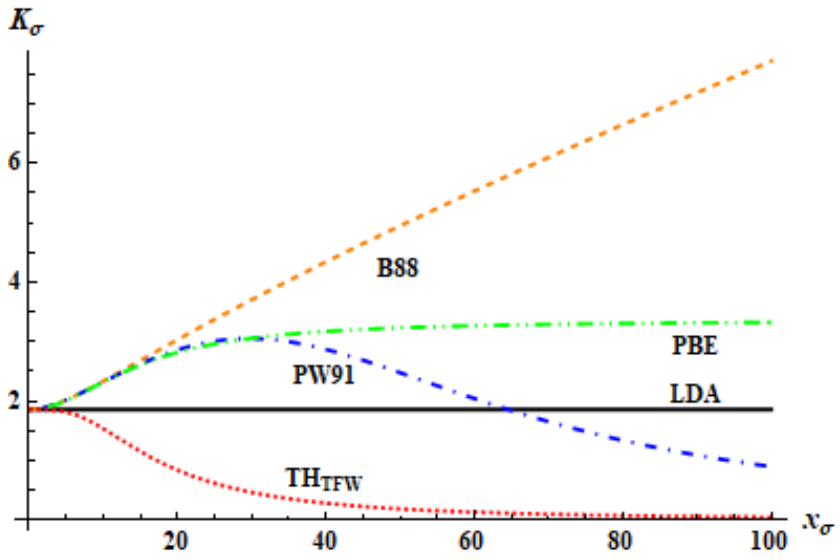

FIG. 2: The plot of K_σ, defined in Eq. (10), in terms of x_σ in Eq. (9) for major GGA exchange functionals.

the figure, GGA exchange functionals are clearly different for large x_σ, i.e., in the region of low electron density and/or high density gradient, because there is no fundamental condition regulating the exchange energy in this region, as mentioned later.

There are several fundamental conditions for exchange energy that GGA exchange functionals should satisfy. Followings are the fundamental conditions, which have been adopted to develop major GGA exchange functionals:

1. Exchange energy is negative for non-zero electron density, $\rho \neq 0$,

$$E_{\rm x}[\rho] \; < \; 0. \tag{11}$$

2. In slowly varying electron density regions ($\nabla\rho << \rho$), the exchange energy is expanded using x_σ in Eq. (9) as [35]

$$\lim_{x_\sigma \to 0} E_{\rm x} \; = \; -3 \left(\frac{3}{4\pi}\right)^{1/3} \sum_\sigma \int d^3\mathbf{r} \rho_\sigma^{4/3} \left[1 + \frac{5x_\sigma^2}{162(6\pi^2)^{2/3}} + O(x_\sigma^4)\right]. \tag{12}$$

Note that the coefficient of the x_σ term in Eq. (12) is controversial, because it is derived to be doubled, i.e., $5x_\sigma^2/81(6\pi^2)^{2/3}$, in another study [36], and many GGA exchange functionals have used nearly double values for practical advantage.

3. Coordinate-scaling conditions strongly regulate GGA exchange functionals. For the uniform coordinate-scaling for cubic coordinates (x, y, z) ($\rho(x, y, z) \to \rho_\lambda = \lambda^3\rho(\lambda x, \lambda y, \lambda z)$), the second-order nonuniform coordinate scaling for planar coordinates (x, y) ($\rho(x, y, z) \to \rho^{xy}_{\lambda\lambda} = \lambda^2\rho(\lambda x, \lambda y, z)$) and the first-order nonuniform coordinate scaling for linear coordinates x ($\rho(x, y, z) \to \rho^x_\lambda = \lambda\rho(\lambda x, y, z)$), exchange energy is scaled as

$$E_{\rm x}[\rho_\lambda] = \lambda E_{\rm x}[\rho], \tag{13}$$

$$\lim_{\lambda\to\infty} \frac{1}{\lambda} E_{\rm x}[\rho^{xy}_{\lambda\lambda}] = \text{const.} \neq 0, \tag{14}$$

$$\lim_{\lambda\to 0} \frac{1}{\lambda} E_{\rm x}[\rho^{xy}_{\lambda\lambda}] = \text{const.} \neq 0, \tag{15}$$

$$\lim_{\lambda\to\infty} E_{\rm x}[\rho^x_\lambda] = \text{const.} \neq 0, \tag{16}$$

$$\lim_{\lambda\to 0} E_{\rm x}[\rho^x_\lambda] = \text{const.} \neq 0, \tag{17}$$

respectively.

4. The Lieb-Oxford bound condition [37],

$$E_{\rm x}[\rho] \geq -1.679 \int d^3\mathbf{r}\rho^{4/3}, \tag{18}$$

has been often used to set the upper limit of exchange energy. Note that the HF exchange integral violates this condition. This condition is, therefore, applicable only to GGA exchange functionals.

5. GGA exchange functionals should be self-interaction free. That is, they should not contain the self-interaction error, which is the Coulomb self-interaction that is not cancelled out with the exchange self-interaction and remains due to the use of the exchange functional. To determine whether or not the exchange energy is self-interaction-free, the self-interaction-free conditions for one-electron systems is often used [34],

$$E_{\rm x}[q\rho_1] = q^2 E_{\rm x}[\rho_1], \tag{19}$$

where ρ_1 is the electron density of one-electron systems. Note that this equation is based on the density matrix of self-interacting electrons. The far-from-nucleus asymptotic behavior condition for exchange energy ($\bar{E}_{\rm x}$ is the integral kernel of $E_{\rm x}$),

$$\lim_{r\to\infty} \rho^{-1}\bar{E}_{\rm x} = -\frac{1}{2r}, \tag{20}$$

is also derived from the self-interaction density matrix [38]. Furthermore, there is the self-interaction-free condition for N electrons [39], which is called "energy linearity theorem for fractional occupations". Note that the long-range correction makes GGA exchange functionals approximately meet this condition [40]. This indicates that the lack of long-range exchange interactions causes the self-interaction errors from multiple electron systems in GGA exchange functionals.

Compared to models in data science, GGA exchange functionals generally gives very high generalization performance due to the canonical prerequisite that they should be derived from physically legitimate models obeying these fundamental conditions. Under the favor of this prerequisite, the remaining subject is how to construct GGA exchange functionals from the physically legitimate models. As far as we know, there is no fundamental conditions regulating GGA exchange functionals for large x_σ. This indicates that GGA exchange functionals in this region are adjustable to reproduce accurate results. That is why GGA exchange functionals are very different for large x_σ as shown in Fig. 2, as mentioned above.

The B88 exchange functional [31] succeeded to reproduce accurate electronic properties in the framework of the Kohn-Sham method for the first time [1]. Therefore, many GGA exchange functionals have so far been developed to modify this functional. For many years, the B88 exchange functional [31] has been the most popular GGA exchange functional in QC calculations. This functional was derived for the form in the low density/high density gradient region to satisfy the far-from-nucleus asymptotic interaction condition in Eq. (20),

$$K_\sigma^{\mathrm{B88}} \;=\; K_\sigma^{\mathrm{LDA}} + \frac{2\zeta x_\sigma^2}{1+6\zeta x_\sigma \sinh^{-1} x_\sigma}, \tag{21}$$

where

$$K_\sigma^{\mathrm{LDA}} \;=\; 3\left(\frac{3}{4\pi}\right)^{1/3}, \tag{22}$$

and the only semiempirical parameter is $\zeta = 0.0042$. The second term in the right-hand side of Eq. (21), which is the GGA term of this functional, is implemented to satisfy Eq. (20),

$$-\frac{1}{2}\rho_\sigma^{1/3}\left(\frac{2\zeta x_\sigma^2}{1+6\zeta x_\sigma \sinh^{-1} x_\sigma}\right) \xrightarrow{r\to\infty} -\frac{1}{2r}, \tag{23}$$

for the Slater-type 1s orbital, e.g., density $\rho_\sigma^{\mathrm{1s}} = \pi^{-1}\exp(-2r)$, by using the relation,

$$\sinh^{-1}\left[\mathrm{const.}r^b \exp(ar)\right] \xrightarrow{r\to\infty} ar, \tag{24}$$

where a and b are arbitrary real numbers. The only semiempirical parameter is used to adjust the magnitude of the GGA term for relatively small x_σ. Therefore, the B88 exchange functional is interpreted as a correction for the LDA exchange functional to satisfy the far-from-nucleus asymptotic interaction condition.

The PW91 exchange functional [32, 33] modifies the B88 exchange functional to satisfy as many fundamental physical conditions as possible,

$$K_\sigma^{\rm PW91} = K_\sigma^{\rm LDA}\left[\frac{1+6\zeta x_\sigma \sinh^{-1} x_\sigma}{1+6\zeta x_\sigma \sinh^{-1} x_\sigma + dx_\sigma^4/(48\pi^2)^{4/3}} + \frac{(a-b\exp[-cx_\sigma^2/(48\pi^2)^{2/3}])x_\sigma^2/(48\pi^2)^{2/3}}{1+6\zeta x_\sigma \sinh^{-1} x_\sigma + dx_\sigma^4/(48\pi^2)^{4/3}}\right], \tag{25}$$

where $a = 0.2473$, $b = 0.1508$, $c = 100$ and $d = 0.004$ are semiempirical parameters that are introduced to satisfy many fundamental physical conditions or to give accurate electronic properties, and ζ is the semiempirical parameter of the B88 functional. Note that this functional gives very different dependence on large x_σ from the B88 one. This inevitably indicates that this functional violates the far-from-nucleus asymptotic interaction condition.

The PBE exchange functional [5] chooses significant fundamental conditions to be applied with only two semiempirical parameters in order to simplify the PW91 exchange functional,

$$K_\sigma^{\rm PBE} = K_\sigma^{\rm LDA}\left[1+\kappa-\frac{\kappa}{1+\mu x_\sigma^2/(48\pi^2)^{2/3}\kappa}\right], \tag{26}$$

where semiempirical parameters $\mu = 0.21951$ and $\kappa = 0.804$ are the values for satisfying the fundamental conditions and for providing accurate electronic properties. The K_σ approaches a constant for large x_σ as shown in Fig. 2. This indicates that this functional is insensitive in the low density/high density gradient region. The revPBE exchange functional [34] is different from the PBE one only for the semiempirical parameters: $\mu = 0.967$ and $\kappa = 0.235$.

Similarly to the above GGA exchange functionals, many other GGA exchange functionals have also been developed to satisfy significant fundamental conditions for exchange energy by fitting several semiempirical parameters. In contrast, a parameter-free exchange functional (TH) [36] is developed from a new standpoint: it is directly derived from a density matrix expansion [41] at the momentum, which is calculated by the kinetic energy density τ_σ,

$$K_\sigma^{\rm TH} = \frac{27\pi}{10\tau_\sigma}\rho_\sigma^{5/3}\left(1+\frac{7x_\sigma^2\rho_\sigma^{5/3}}{216\tau_\sigma}\right). \tag{27}$$

Note that this functional contains the σ-spin kinetic energy density τ_σ with no semiempirical parameters. However, this kinetic energy density requires a GGA kinetic energy functional,

TABLE I: Fundamental conditions for exchange energy, semiempirical parameters, and the comparison of GGA exchange functionals for their validities [1]. For the slowly-varying density limit, two probable behaviors are suggested.

Condition/Parameters	LDA	PW91	PBE	B88	TH_{TFW}
Fundamental conditions					
Negative exchange energy	Yes	Yes	Yes	Yes	Yes
LDA limit	Yes	Yes	Yes	Yes	Yes
Slowly-varying density limit	-	Yes?	Yes?	No	Yes?
Uniform coordinate scaling	Yes	Yes	Yes	Yes	Yes
Nonuniform coordinate scaling	No	No	No	No	No
Lieb-Oxford bound	Yes	Yes	Yes	No	Yes
Self-interaction-free for 1 electron	No	No	No	No	No
Far-from-nucleus asymptotic behavior	No	No	No	Y/N	No
Semiempirical parameters					
Number of parameters	0	4	2	1	0
Physical meaning of parameters	-	Yes	Yes	Yes	-

because it usually gives a divergent energy for the exact kinetic energy density due to divide-by-zero. Nevertheless, this is obviously a promising functional, because applying the Thomas-Fermi-von Weizsäcker (TFW) GGA kinetic energy functional [42–44] to the kinetic energy density term, this functional accurately reproduces atomic exchange energies (average 97.8 % of the exact values with mean absolute error of 0.154 hartree) in spite of its parameterless form [36]. Due to this kinetic energy density term, this functional is classified into not GGA functionals but "progressive functional", which is defined to transform in accordance with the combined term.

Table I summarizes the physical legitimacy of GGA exchange functionals by checking the fulfillment of the fundamental conditions for exchange energy and the physical meaning of semiempirical parameters. This table shows that the TH functional has very high physical legitimacy equivalent to those of PW91 and PBE, though the former is not developed to satisfy these fundamental conditions different from the latters. This TH functional also

has no semiempirical parameter, though it contains the kinetic energy density term, which should be derived from a kinetic energy functional. Therefore, this functional can provide the best performance by using a physically legitimate kinetic energy density with minimal parameters (preferably one parameter).

B. LDA and GGA correlation functionals

Different from the LDA exchange functional, there is no exact LDA correlation functional. Conventional LDA correlation functionals, therefore, have been developed to satisfy several fundamental conditions for the correlation energy of uniform electron gas: e.g., the high density limit [45], low density limit [46], and random phase approximation (RPA) expression [47]. Note, however, that the middle-range density regions of correlation functionals are not regulated by these limits and these limits are inconsistent with the RPA expression [48]. Conventional LDA correlation functionals have complicated formulations to solve these problems: The Vosko-Wilk-Nusair (VWN) [49] and Perdew-Wang (PW) [33] LDA correlation functionals are frequently used in QC and solid state calculations, respectively.

The VWN LDA correlation functional [49] is constructed to get the RPA expression converge to the high and low density limits. Using a Padè interpolation model, this functional is developed by fitting many semiempirical parameters to the correlation energy of the quantum Monte Carlo for the uniform electron gas,

$$\begin{aligned} E_{\rm c}^{\rm VWN} \;=\; & \int d^3\mathbf{r}\,\Big\{\epsilon_0^{\rm VWN}(x_{01},a_1,b_1,c_1) \\ & +\epsilon_0^{\rm VWN}(x_{02},a_2,b_2,c_2)\left(\frac{9(f_1[\zeta]-2)}{2}\right)(1-\zeta^4) \\ & +\left[\epsilon_0^{\rm VWN}(x_{03},a_3,b_3,c_3)-\epsilon_0^{\rm VWN}(x_{01},a_1,b_1,c_1)\right]\frac{f_1(\zeta)-2}{2^{1/3}-1}\zeta^4\Big\}, \end{aligned} \tag{28}$$

where $\epsilon_0^{\rm VWN}$ is expressed as

$$\begin{aligned} \epsilon_0^{\rm VWN}(x_0,a,b,c) \;=\; & \frac{a}{2}\left\{\ln\!\left(\frac{r_{\rm s}}{r_{\rm s}+b\sqrt{r_{\rm s}}+c}\right)+\frac{2b}{\sqrt{4c-b^2}}\tan^{-1}\!\left(\frac{\sqrt{4c-b^2}}{2\sqrt{r_{\rm s}}+b}\right)\right. \\ & -\frac{bx_0}{x_0^2+bx_0+c}\left[\ln\!\left(\frac{(\sqrt{r_{\rm s}}-x_0)^2}{r_{\rm s}+b\sqrt{r_{\rm s}}+c}\right)\right. \\ & \left.\left.+\frac{2(b+2x_0)}{\sqrt{4c-b^2}}\tan^{-1}\!\left(\frac{\sqrt{4c-b^2}}{2\sqrt{r_{\rm s}}+b}\right)\right]\right\}, \end{aligned} \tag{29}$$

$f_1[\zeta] = \left\{(1+\zeta)^{4/3} + (1-\zeta)^{4/3}\right\}/2$, $\zeta = (\rho_\alpha - \rho_\beta)/(\rho_\alpha + \rho_\beta)$, and r_s is Wigner-Seitz radius defined by $r_\mathrm{s} = (3/4\pi\rho)^{1/3}$. In this functional, twelve semiempirical parameters, x_{0i}, a_i, b_i, c_i for i=1 - 3, are used to be fitted to the correlation energy of the uniform electron gas. Note that many hybrid functionals such as B3LYP [50] contains this VWN functional.

The PW LDA correlation functional [33] simplifies this VWN LDA correlation functional to decrease the semiempirical parameters and to satisfy more fundamental physical conditions,

$$E_\mathrm{c}^{\text{PW-LDA}}[\rho] \;=\; -2a \int d^3\mathbf{r}\rho\,(1-\alpha r_\mathrm{s}) \ln\left[1 + \frac{1}{2a\left(\beta_1 r_\mathrm{s}^{1/2} + \beta_2 r_\mathrm{s} + \beta_3 r_\mathrm{s}^{3/2} + \beta_4 r_\mathrm{s}^2\right)}\right]. \quad (30)$$

Even though, this functional includes six semiempirical parameters: $a = 0.031097$, $\alpha = 0.21370$, $\beta_1 = 7.5957$, $\beta_2 = 3.5876$, $\beta_3 = 1.6382$, and $\beta_4 = 0.49294$ for adjusting to the uniform electron gas correlation energy. This functional is used in many GGA correlation functionals such as the PBE functional.

Conventional GGA correlation functionals are roughly classified into two types: the density gradient approximation (DGA)-type and the Colle-Salvetti (CS)-type functionals. Similarly to the GGA exchange functionals, DGA-type correlation functionals are derived from models satisfying some fundamental conditions for correlation energy by fitting semiempirical parameters to reproduce accurate correlation energies of molecules. The CS-type correlation functional [51] is derived from a correlation wavefunction, in which an uncorrelated wavefunction is multiplied by a correlation function satisfying the correlation cusp condition for short-range electron-electron interactions.

Followings are several major fundamental conditions for correlation energy:

1. Correlation energy is always negative for non-zero electron density, $\rho \neq 0$,

$$E_\mathrm{c}[\rho] \;\leq\; 0. \quad (31)$$

   Though correlation energy is zero in one-electron systems, it is not necessarily zero for unoccupied orbitals, even in the hydrogen atom [52].

2. In slowly varying electron density regions, the correlation energy is expanded using a dimensionless parameter $x = |\nabla\rho|/\rho^{4/3}$ as [35]

$$\lim_{x\to 0} E_\mathrm{c} \;=\; \int d^3\mathbf{r}\left\{c_1[\rho] + c_2[\rho]x^2 + O(x^4)\right\}. \quad (32)$$

3. For rapidly varying (low-density-high-gradient) electron density regions, where the density gradient is much larger than the electron density, the correlation energy behaves as [53, 54]

$$\lim_{x\to\infty} \rho^{-1}\bar{E}_{\rm c} \;=\; 0, \tag{33}$$

where $\bar{E}_{\rm c}$ is the integral kernel of the correlation energy. Under the favor of this condition, correlation functionals are regulated for the low-density-high-gradient region in contrast to exchange functionals. This limit appears prominently in small electron density regions, where dispersion forces usually determine the character of the interatomic bonds. Note that this condition gives a density dependence of the exchange energy due to the use of the LDA exchange functional in the derivation.

4. In coordinate-scaling conditions, the order of correlation energy is also regulated for the scaling of the electron density, which corresponds to the scaling of coordinates: for the uniform coordinate-scaling,

$$E_{\rm c}[\rho_\lambda] \;<\; \lambda E_{\rm c}[\rho] \quad (\lambda < 1), \tag{34}$$

$$E_{\rm c}[\rho_\lambda] \;>\; \lambda E_{\rm c}[\rho] \quad (\lambda > 1), \tag{35}$$

$$\lim_{\lambda\to\infty} E_{\rm c}[\rho_\lambda] \;=\; \text{const.} \neq 0, \tag{36}$$

$$\lim_{\lambda\to 0} \frac{1}{\lambda} E_{\rm c}[\rho_\lambda] \;=\; \text{const.} \neq 0. \tag{37}$$

For the nonuniform coordinate-scaling, the following conditions are also established: in the second-order scaling,

$$\lim_{\lambda\to\infty} E_{\rm c}[\rho^{xy}_{\lambda\lambda}] \;=\; 0, \tag{38}$$

$$\lim_{\lambda\to 0} \frac{1}{\lambda^2} E_{\rm c}[\rho^{xy}_{\lambda\lambda}] \;=\; \text{const.} \neq 0, \tag{39}$$

and in the first-order scaling,

$$\lim_{\lambda\to\infty} \lambda E_{\rm c}[\rho^{x}_{\lambda}] \;=\; \text{const.} \neq 0, \tag{40}$$

$$\lim_{\lambda\to 0} \frac{1}{\lambda} E_{\rm c}[\rho^{x}_{\lambda}] \;=\; 0. \tag{41}$$

5. Since one-electron systems contain no electron correlation self-interaction in two-electron interactions, the self-interaction-free conditions for one-electron systems is

given as

$$E_{\rm c}^{\alpha\beta}[q\rho_1] \;=\; 0, \tag{42}$$

where ρ_1 is the electron density of one-electron systems. Interestingly, Eq. (42) is derived from the density matrix of self-interacting electrons.

The most frequently-used DGA-type correlation functionals are the PW91 and PBE correlation functionals. The advantage of these functionals is satisfying many fundamental conditions. Instead, the disadvantage is the complicated forms and physically-meaningless semiempirical parameters.

The PW91 correlation functional [33] expands the PW LDA correlation functional with the GGA term, which is determined using many semiempirical parameters to satisfy the fundamental conditions for correlation energy,

$$E_{\rm c}^{\rm PW91}[\rho,s,t] \;=\; E_{\rm c}^{\text{PW-LDA}}[\rho] + \int d^3\mathbf{r}\rho H[\rho,s,t] \tag{43}$$

$$\begin{aligned} H[\rho,s,t] \;=\; & \frac{\beta^2}{2\alpha}\ln\left[1+\frac{2\alpha}{\beta}\frac{t^2+At^4}{1+At^2+A^2t^4}\right] \\ & +C_{c_0}\left[C_1+\frac{C_2+C_3r_{\rm s}+C_4r_{\rm s}^2}{1+C_5r_{\rm s}+C_6r_{\rm s}^2+C_7r_{\rm s}^3}-C_{c_1}\right]t^2\exp(-100s^2), \end{aligned} \tag{44}$$

where $\bar{E}_{\rm c}^{\text{PW-LDA}}$ is the integral kernel of the PW LDA correlation functional, the coefficient,

$$A \;=\; \frac{2\alpha}{\beta}\left[\exp\left(\frac{-2\alpha\bar{E}_{\rm c}^{\text{PW-LDA}}[\rho]}{\beta^2\rho}\right)-1\right]^{-1}, \tag{45}$$

and the dimensionless parameters,

$$s \;=\; \frac{|\boldsymbol{\nabla}\rho|}{2k_{\rm F}\rho} \;\;\text{and}\;\; t=\frac{|\boldsymbol{\nabla}\rho|}{2k_{\rm s}\rho}, \tag{46}$$

in which $k_{\rm F}=(3\pi^2\rho)^{1/3}$ and $k_{\rm s}=(4k_{\rm F}/\pi)^{1/2}$. This GGA term contains 11 semiempirical parameters, α, β, C_{c_0}, C_{c_1} and C_i for i=1 - 7, which are adjusted to give accurate electronic properties.

The PBE correlation functional [5] solves a few problems in the PW91 correlation functional, e.g., the large number of parameters inducing artificial local minima in potential energy surfaces and the implicit inclusion of exchange effects [1], by focusing on three significant fundamental conditions to be satisfied,

$$E_{\rm c}^{\rm PBE}[\rho,\zeta,t] \;=\; E_{\rm c}^{\text{PW-LDA}}[\rho] + \int d^3\mathbf{r}\rho H[\rho,\zeta,t] \tag{47}$$

$$H[\rho,\zeta,t] \;=\; \gamma\phi^3\ln\left[1+\frac{\beta}{\gamma}t'^2\left(\frac{1+At'^2}{1+At'^2+A^2t'^4}\right)\right], \tag{48}$$

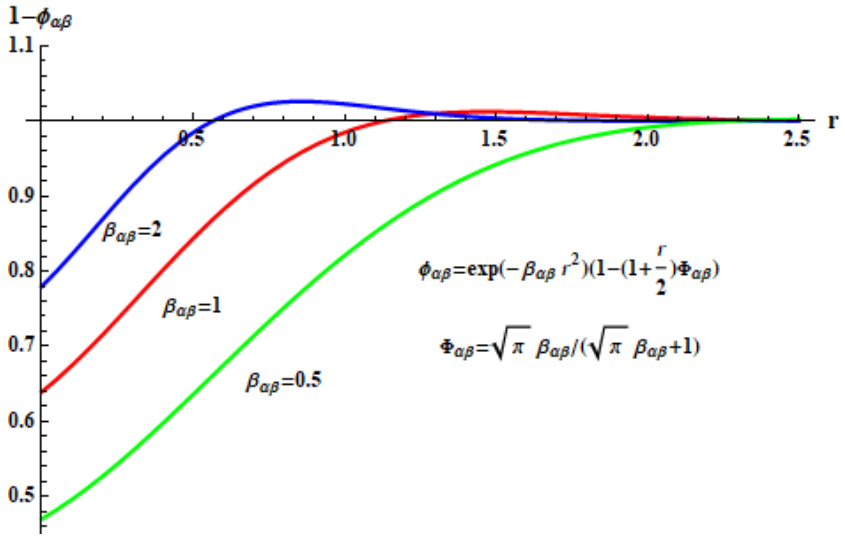

FIG. 3: Form of the correlation function for the Colle-Salvetti-type correlation functionals. $\Psi_{\rm uncorr}$ is a wavefunction containing no electron correlation effect.

where

$$A = \frac{\beta}{\gamma}\left[\exp\left(-\frac{\bar{E}_{\rm c}^{\text{PW-LDA}}[\rho]}{\gamma\phi^2\rho}\right)-1\right]^{-1}, \tag{49}$$

$$\phi = \frac{1}{2}\left[(1+\zeta)^{2/3}+(1-\zeta)^{2/3}\right], \tag{50}$$

$\zeta = (\rho_\alpha - \rho_\beta)/(\rho_\alpha + \rho_\beta)$. As a result, the number of semiempirical parameters in the GGA term is dramatically reduced from 11 to 2, $\gamma = (1-\ln 2)/\pi^2 = 0.031091$ and $\beta = 0.066725$, that are fitted to provide accurate electronic properties.

Representative CS-type correlation functionals are the original CS [51], Lee-Yang-Parr (LYP) [55] and the one-parameter progressive (OP) [56] functionals. In these functionals, correlation holes are determined to satisfy the correlation cusp condition. The general form of the correlation function is drawn in Fig. 3. In this correlation function, the size of correlation hole is determined by $\beta_{\alpha\beta}$, which usually depends on the electron density.

The original CS correlation functional is an electron correlation correction for the HF method,

$$E_{\rm c}^{\rm CS}[\rho, \boldsymbol{\nabla}_{\mathbf{r}}^2 P_{\rm 2HF}] = -\int d^3\mathbf{r}\frac{a}{1+d\beta}\rho\left[1+b\rho^{-8/3}\left[\boldsymbol{\nabla}_{\mathbf{r}}^2 P_{\rm 2HF}\left(\mathbf{r}-\frac{\mathbf{s}}{2},\mathbf{r}+\frac{\mathbf{s}}{2}\right)\right]_{\mathbf{s}=\mathbf{0}}\exp(-c/\beta)\right], \tag{51}$$

where $\beta = q\rho^{1/3}$. This functional contains five semiempirical parameters, $a = 0.04918$,

$b = 0.06598$, $c = 0.58$, $d = 0.8$, and $q = 2.29$. The second derivative term of the diagonal term of the second-order density matrix, which is very costly in practical calculations, is also included in this functional. Therefore, this functional is not a GGA functional.

The LYP correlation functional transforms the costly second derivative term of the CS functional to a density gradient term to dramatically reduce the computational time, reduced the computational time. This functional is in the following form:

$$E_{\rm c}^{\rm LYP}[\rho, \boldsymbol{\nabla}\rho, \boldsymbol{\nabla}^2\rho] = -\int d^3\mathbf{r} \frac{a}{1+d\rho^{-1/3}} \left\{\rho + b\rho^{-2/3}\left[C_{\rm F}\rho^{5/3} - 2t_{\rm W} + \frac{1}{9}\left(t_{\rm W} + \frac{1}{2}\boldsymbol{\nabla}^2\rho\right)\right]\exp\left(-c\rho^{-1/3}\right)\right\}, \quad (52)$$

where

$$t_{\rm W} = \frac{1}{8}\left(\frac{|\boldsymbol{\nabla}\rho|^2}{\rho} - \boldsymbol{\nabla}^2\rho\right) \quad (53)$$

and $\beta = q\rho^{1/3}$ and $C_{\rm F} = 3(3\pi^2)^{2/3}/10$, which are adjusted to give accurate electronic properties. This functional also contains five semiempirical parameters, $a = 0.04918$, $b = 0.7628$, $c = 0.58$, $d = 0.8$, and $q = 2.29$. Though density Laplacian $\boldsymbol{\nabla}^2\rho$ is included in Eq. (52), this is usually approximated as a density gradient term by partial integral transformation. That is why this functional is assigned to a GGA functional. The most serious problem of this functional is that the sign of the density Laplacian term is set to be opposite to that of the CS functional probably for practical reason [56]. This problem partly causes that this functional hardly satisfies the fundamental conditions for correlation energy.

The OP correlation functional [56] is derived assuming that the exchange hole coming from the combined exchange functional is proportional to the exclusion volume of the correlation hole, and is derived for opposite-spin electron pairs with only one semiempirical parameter,

$$E_c^{\rm OP} = -\int d^3\mathbf{r}\rho_\alpha\rho_\beta \frac{1.5214\beta_{\alpha\beta} + 0.5764}{\beta_{\alpha\beta}^4 + 1.1284\beta_{\alpha\beta}^3 + 0.3183\beta_{\alpha\beta}^2}, \quad (54)$$

where

$$\beta_{\alpha\beta} = q_{\alpha\beta}\left(\rho_\alpha^{-1/3}K_\alpha^{-1} + \rho_\beta^{-1/3}K_\beta^{-1}\right)^{-1}, \quad (55)$$

and K_σ is the exchange functional term defined in Eq. (10). Note that the values in Eq. (54), i.e., 1.5214, 0.5763, 1.1284 and 0.3183, are not semiempirical fitted parameters but

TABLE II: Fundamental conditions for correlation energy, semiempiirical parameters, and the comparison of GGA correlation functionals for their validities [1].

Condition/Parameters	LDA	PW91	PBE	LYP	OP_{B88}
Fundamental conditions					
Negative correlation energy	Yes	No	Yes	No	Yes
LDA limit	No	No	No	No	Yes
Slowly-varying density limit	-	Yes	Yes	No	Yes
Rapidly-varying density limit	No	Yes	Yes	No	Yes
Uniform coordinate scaling	No	No	Yes	No	Yes
Nonuniform coordinate scaling	No	Yes	No	No	No
Self-interaction-free for 1 electron	No	No	No	Yes	Yes
Semiempirical parameters					
Number of parameters	12 or 6	11	2	4	1
Physical meaning of parameters	No	No	No	No	Yes

physically-driven constants. The semiempirical parameter is only $q_{\alpha\beta}$, which is determined for each exchange functional combined: e.g., $q_{\alpha\beta} = 2.367$ if this functional is combined with the B88 and PBE exchange functionals. Therefore, this parameter is related to the proportionality constant between the sizes of exchange and correlation holes. Since the OP functional is a progressive functional transforming in accordance with a combined exchange functional, it is an LDA correlation functional for the K_σ of the LDA exchange functional and is a GGA correlation functional for the K_σ of a GGA exchange functional. The OP functional gives correlation energies similar to those of the LYP correlation functional, despite of the only one semiempirical parameter included. This functional also satisfies the most fundamental conditions for correlation energy in all correlation functionals, though no fundamental conditions are taken into account in the derivation.

Table II displays the fundamental conditions for correlation energy, semiempirical parameters, and the physical legitimacy of GGA correlation functionals [1]. The table indicates that OP correlation is the most physically-legitimate functional satisfying the most fundamental conditions, despite it is not derived neglecting these conditions and contains only

one semiempirical parameter, which has physical meaning.

C. Meta-GGA functionals

The meta-GGA functional upgrades GGA functional by adding the term(s) of kinetic energy density τ. This upgrade is justified by the fact that the kinetic energy density τ and the density Laplacian $\boldsymbol{\nabla}^2\rho$ are included in the density matrix expansion around the Fermi momentum [41],

$$k_{\mathrm{F}\sigma} = (6\pi^2\rho_\sigma)^{1/3}, \tag{56}$$

while the exchange energy is represented by the density matrix. Actually, the kinetic energy density is approximated using the density gradient and density Laplacian,

$$\tau = \sum_\sigma \tau_\sigma = \frac{1}{2}\sum_\sigma\sum_i |\boldsymbol{\nabla}\phi_{i\sigma}|^2 \tag{57}$$

$$= \sum_\sigma \left[C_{\mathrm{F}}\rho_\sigma^{5/3} + \frac{1}{72}\frac{|\boldsymbol{\nabla}\rho_\sigma|^2}{\rho_\sigma} + \frac{1}{6}\boldsymbol{\nabla}^2\rho_\sigma + O(\boldsymbol{\nabla}^4) \right]. \tag{58}$$

Major meta-GGA exchange-correlation functionals include the van Voorhis-Scuseria 1998 (VS98) [57], the Perdew-Kurt-Zupan-Blaha (PKZB) [58], the Tao-Perdew-Staroverov-Scuseria (TPSS) [59], and the strongly constrained and appropriately normed (SCAN) [60, 61] functionals.

The VS98 meta-GGA exchange-correlation functional [57] is derived from the analytical expansion of the density matrix [41], similarly to the TH exchange functional in Eq. (27),

$$E_{\mathrm{x}}^{\mathrm{VS98}} = \sum_\sigma \int d^3\mathbf{r}\rho_\sigma^{4/3} h_\sigma^{\mathrm{VS98}}(x_\sigma, z_\sigma), \tag{59}$$

$$E_{\mathrm{c}\sigma\sigma'}^{\mathrm{VS98}} = \int d^3\mathbf{r}\hat{E}_{\mathrm{c}\sigma\sigma'}^{\mathrm{LDA}} h_\sigma(\sqrt{x_\alpha^2 + x_\beta^2}, z_\alpha + z_\beta), \tag{60}$$

and

$$E_{\mathrm{c}\sigma\sigma}^{\mathrm{VS98}} = \int d^3\mathbf{r}\hat{E}_{\mathrm{c}\sigma\sigma}^{\mathrm{LDA}} h^{\mathrm{VS98}}\sigma(x_\sigma, z_\sigma), \tag{61}$$

where

$$h_\sigma^{\mathrm{VS98}}(x_\sigma, z_\sigma) = \frac{a}{\gamma_\sigma(x_\sigma, z_\sigma)} + \frac{bx_\sigma^2 + cz_\sigma}{\gamma_\sigma^2(x_\sigma, z_\sigma)} + \frac{dx_\sigma^4 + ex_\sigma^2 z_\sigma + fz_\sigma^2}{\gamma_\sigma^3(x_\sigma, z_\sigma)}, \tag{62}$$

and $\gamma_\sigma(x_\sigma, z_\sigma) = 1+\alpha(x_\sigma^2+z_\sigma)$. $\hat{E}^{\rm LDA}_{{\rm c}\sigma_1\sigma_2}$ is the integral kernel of the $\sigma_1\sigma_2$-pair LDA correlation functional. Both x_σ and z_σ are dimensionless parameters: x_σ is given in (9) and z_σ is defined using the kinetic energy density as

$$z_\sigma = \frac{\tau_\sigma - \tau_\sigma^{\rm TF}}{\rho_\sigma^{5/3}} = \frac{\tau_\sigma}{\rho_\sigma^{5/3}} - C_{\rm F}. \tag{63}$$

This functional contains twenty one semiempirical parameters: seven semiempirical parameters (a through f and α) for each exchange, parallel-spin or opposite-spin pair correlation functional. Due to a considerable number of semiempirical parameters and its formulation of dimensionless x_σ and z_σ, this functional is also taken as the first "semiempirical functionals", mentioned in Sec. IV E. These semiempirical parameters are fitted to give accurate electronic properties of molecules and therefore have no physical meanings.

The PKZB meta-GGA exchange-correlation functional [58] extends the PBE-GGA functional to the inclusion of the kinetic energy density to satisfy the fundamental conditions containing the density Laplacian,

$$E_{\rm x}^{\rm PKZB} = \sum_\sigma \int d^3\mathbf{r} \bar{E}_{\rm x}^{\rm LDA}[2\rho_\sigma] \left\{ 1 + \kappa - \frac{\kappa}{1 + x/\kappa} \right\}, \tag{64}$$

where

$$x = \frac{10}{81}p + \frac{146}{2025}\tilde{q}^2 - \frac{73}{405}\tilde{q}p + \left[D + \frac{1}{\kappa}\left(\frac{10}{81}\right)^2 \right] p^2, \tag{65}$$

$$\tilde{q} = \frac{3\tau_\sigma}{2(3\pi^2)^{2/3}\rho_\sigma^{5/3}} - \frac{9}{20} - \frac{p}{12}, \tag{66}$$

and

$$p = \frac{|\boldsymbol{\nabla}\rho_\sigma|^2}{4(6\pi^2)^{2/3}\rho_\sigma^{8/3}}. \tag{67}$$

This functional contains two semiempirical parameters, $D = 0.113$ and $\kappa = 0.804$, which are fitted to give accurate exchange energies. The PKZB correlation functional removes the self-interaction error for one electron from the PBE-GGA correlation functional,

$$\begin{aligned} E_{\rm c}^{\rm PKZB} = \int d^3\mathbf{r} \Bigg\{ & \bar{E}_{\rm c}^{\rm PBE}(\rho_\alpha, \rho_\beta, \boldsymbol{\nabla}\rho_\alpha, \boldsymbol{\nabla}\rho_\beta) \left[1 + C \left(\frac{\sum_\sigma \tau_\sigma^{\rm W}}{\sum_\sigma \tau_\sigma} \right)^2 \right] \\ & - (1+C) \sum_\sigma \left(\frac{\tau_\sigma^{\rm W}}{\tau_\sigma} \right)^2 \bar{E}_{\rm c}^{\rm PBE}(\rho_\sigma, 0, \boldsymbol{\nabla}\rho_\sigma, 0) \Bigg\}, \end{aligned} \tag{68}$$

where $\tau_\sigma^{\rm W}$ is the von Weizsäcker kinetic energy density, which is defined by the von Weizsäcker kinetic energy,

$$T_{\rm W} = \int d^3\mathbf{r}\tau^{\rm W} = \sum_\sigma \int d^3\mathbf{r}\tau_\sigma^{\rm W} = \frac{1}{8}\sum_\sigma \int d^3\mathbf{r}\frac{|\boldsymbol{\nabla}\rho_\sigma(\mathbf{r})|^2}{\rho_\sigma(\mathbf{r})}. \tag{69}$$

This extension adds one semiempirical parameter $C = 0.53$ for reproducing accurate correlation energies.

The TPSS meta-GGA exchange-correlation functional [59] intends to remove the semiempirical parameters in the PKZB functional to construct a nonempirical meta-GGA functional,

$$E_{\rm c}^{\rm TPSS} = \int d^3\mathbf{r}\bar{E}_{\rm c}^{\rm revPKZB}\left[1 + d\left(\frac{\tau^{\rm W}}{\tau}\right)^3\right], \tag{70}$$

and

$$\begin{aligned}\bar{E}_{\rm c}^{\rm revPKZB} &= \bar{E}_{\rm c}^{\rm PBE}[\rho_\alpha, \rho_\beta, \boldsymbol{\nabla}\rho_\alpha, \boldsymbol{\nabla}\rho_\beta]\left[1 + C(\zeta,\xi)\left(\frac{\tau^{\rm W}}{\tau}\right)^2\right] \\ &\quad - [1 + C(\zeta,\xi)]\left(\frac{\tau^{\rm W}}{\tau}\right)^2 \sum_\sigma \frac{\rho_\sigma}{\rho}\bar{E}_{\rm c}^{\rm max},\end{aligned} \tag{71}$$

where

$$C(\zeta,\xi) = \frac{0.53 + 0.87\zeta^2 + 0.50\zeta^4 + 2.26\zeta^6}{\left\{1 + \xi^2\left[(1+\zeta)^{-4/3} + (1-\zeta)^{-4/3}\right]/2\right\}^4}, \tag{72}$$

$$\bar{E}_{\rm c}^{\rm max} = \max\left\{\bar{E}_{\rm c}^{\rm PBE}[\rho_\sigma, 0, \boldsymbol{\nabla}\rho_\sigma, 0], \bar{E}_{\rm c}^{\rm PBE}[\rho_\alpha, \rho_\beta, \boldsymbol{\nabla}\rho_\alpha, \boldsymbol{\nabla}\rho_\beta]\right\}, \tag{73}$$

$\xi = |\boldsymbol{\nabla}\zeta|/2(3\pi^2\rho)^{1/3}$, $\zeta = (\rho_\alpha - \rho_\beta)/\rho$, and $d = 2.8$. However, this functional contains six semiempirical parameters with four constants in $C(\zeta,\xi)$ for removing the self-interaction error, though these parameters are interpreted as "fundamental constants". These semiempirical parameters are used to satisfy the fundamental conditions for correlation energy and to provide accurate electronic properties of molecules.

The SCAN meta-GGA exchange-correlation functional [60] improves the accuracy of the TPSS functional by imposing the exact constraints on meta-GGA and adding some norms for rare-gas atoms and nonbonded interactions. There is a hybrid version of this functional [61]. However, this functional has a very complicated formulation and contains many fitted semiempirical parameters (6 and 11 parameters for exchange and correlation functionals, respectively, though these are interpreted as fundamental constants) for satisfying all fundamental conditions of meta-GGA. For the use in machine learning, the formulation should be simplified and the number of the parameters should be dramatically reduced.

D. Hybrid functionals

Based on the concept of the adiabatic connection, which makes the Kohn-Sham energies of the independent electron model link to those of the fully-interacting electron one, hybrid functionals mix the HF exchange integral with GGA exchange functionals at a constant ratio,

$$E_{\mathrm{x}} = \int_0^1 d\lambda E_{\mathrm{x}}^{\lambda} \approx E_{\mathrm{x}}^{\mathrm{GGA}} + \lambda \left(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{GGA}} \right), \tag{74}$$

where λ is termed the coupling strength parameter. The adiabatic connection used in hybrid functionals is not between non-interacting and interacting systems in the usual connection but between non-correlated and correlated systems. Note that hybrid functionals are not the generic term of functionals combining the HF exchange integral with exchange functionals. Instead, hybrid functionals are defined to obey the ansatz that the exact exchange energy is situated between the GGA exchange energy functional and the HF exchange integral. Representative hybrid functionals are B3LYP [50], PBE0 [62], and Heyd-Scuseria-Ernzerhof (HSE) [63] functionals.

The B3LYP hybrid functional [50] is the most frequently used DFT functional in QC calculations,

$$E_{\mathrm{xc}}^{\mathrm{B3LYP}} = E_{\mathrm{xc}}^{\mathrm{LDA}} + a_1 \left(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{LDA}} \right) + a_2 \Delta E_{\mathrm{x}}^{\mathrm{B88}} + a_3 \left(E_{\mathrm{c}}^{\mathrm{LYP}} - E_{\mathrm{c}}^{\mathrm{VWN\text{-}LDA}} \right) \tag{75}$$

This functional introduces three semiempirical parameters, which are fitted to give accurate chemical properties and therefore have no physical meanings, for adiabatically connecting the HF exchange integral with the B88 exchange functional and the LYP-GGA correlation functional with the LDA correlation functional: $a_1 = 0.2$, $a_2 = 0.72$, and $a_3 = 0.81$. This functional often provides surprisingly accurate chemical properties for small molecules, though it often has serious problems for large systems [64–68].

The PBE0 hybrid functional [62] is a perturbation correction for the PBE-GGA exchange-correlation functional using the energy difference between the exchange functional and the HF exchange integral,

$$E_{\mathrm{xc}}^{\mathrm{PBE0}} = E_{\mathrm{xc}}^{\mathrm{PBE}} + \frac{1}{4} \left(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{PBE}} \right). \tag{76}$$

This functional contains no additional semiempirical parameter, though it includes them in the PBE exchange-correlation functional.

The HSE hybrid functional [63] corrects for the short-range part of the PBE exchange-correlation functional by mixing the HF exchange integral at a constant rate,

$$E_{\rm xc}^{\rm HSE} = aE_{\rm x}^{\text{SR-HF}} + (1-a)E_{\rm x}^{\rm PBE} + E_{\rm c}^{\rm PBE}, \tag{77}$$

where $E_{\rm x}^{\text{SR-HF}}$ is the short-range part of the HF exchange integral, which is separated by the standard error function. This functional is also called as a "range-separated functional" due to the division of the standard error function in a manner similar to the long-range correction (Sec. V A), though the effect of this correction is opposite to that of the long-range correction. Similarly to that of the PBE0 functional, this functional also contains no additional semiempirical parameter with $a = 1/4$ as the perturbation coefficient, though it includes the semiempirical parameters in the PBE functional.

E. Semiempirical functionals

Semiempirical functionals focus on reproducing accurate chemical properties incorporating a considerable number of semiempirical parameters and above-mentioned dimensionless parameters, i.e., x_σ and z_σ, Representative semiempirical functionals are B97 [69] and Hamprecht-Cohen-Tozer-Handy (HCTH) [70] functionals, and their derivatives: the B97-series and the Mx-series semiempirical functionals.

The B97 semiempirical functional [69] is also assigned to a hybrid meta-GGA functional,

$$E_{\rm xc}^{\rm B97} = E_{\rm x}^{\rm B97} + E_{\rm c}^{\rm B97} + c_{\rm x}E_{\rm x}^{\rm HF}, \tag{78}$$

$$E_{\rm x}^{\rm B97} = \sum_\sigma \int d^3\mathbf{r}\bar{E}_{{\rm x}\sigma}^{\rm LDA}[\rho_\sigma]g_{{\rm x}\sigma}(x_\sigma^2), \tag{79}$$

$$E_{\rm c}^{\rm B97} = \sum_\sigma \int d^3\mathbf{r}\bar{E}_{{\rm c}\sigma\sigma}^{\rm LDA}[\rho_\sigma]g_{{\rm c}\sigma\sigma}(x_\sigma^2) + \int d^3\mathbf{r}\bar{E}_{{\rm c}\alpha\beta}^{\rm LDA}[\rho_\alpha,\rho_\beta]g_{{\rm c}\alpha\beta}\left(\frac{x_\alpha^2+x_\beta^2}{2}\right), \tag{80}$$

where

$$g_{{\rm x}\sigma} = \sum_i^m C_{{\rm x}\sigma i}\left[\gamma_{{\rm x}\sigma}x_\sigma^2\left(1+\gamma_{{\rm x}\sigma}x_\sigma^2\right)^{-1}\right]^i, \tag{81}$$

$$g_{{\rm c}\sigma\sigma} = \sum_i^m C_{{\rm c}\sigma\sigma i}\left[\gamma_{{\rm c}\sigma\sigma}x_\sigma^2\left(1+\gamma_{{\rm c}\sigma\sigma}x_\sigma^2\right)^{-1}\right]^i, \tag{82}$$

$$g_{{\rm c}\alpha\beta} = \sum_i^m C_{{\rm c}\alpha\beta i}\left[\frac{1}{2}\gamma_{{\rm c}\alpha\beta}(x_\alpha^2+x_\beta^2)\left(1+\frac{1}{2}\gamma_{{\rm c}\alpha\beta}(x_\alpha^2+x_\beta^2)\right)^{-1}\right]^i, \tag{83}$$

where $\gamma_{\mathrm{x}\sigma} = 0.004$, $\gamma_{\mathrm{c}\sigma\sigma} = 0.2$, and $\gamma_{\mathrm{c}\alpha\beta} = 0.006$. Besides these three parameters, this functional contains ten semiempirical parameters: c_{x}, $c_{\mathrm{x}\sigma i}$, $c_{\mathrm{c}\sigma\sigma i}$, and $c_{\mathrm{x}\alpha\beta i}$ for $i = 0, 1, 2$. These semiempirical parameters have no physical meanings.

The HCTH semiempirical functional [70] transforms this B97 functional to a GGA functional by removing c_{x} and increasing the numbers of $c_{\mathrm{x}\sigma i}$, $c_{\mathrm{c}\sigma\sigma i}$, and $c_{\mathrm{x}\alpha\beta i}$ to $i = 0$ to 4. This functional contains fifteen semiempirical parameters in total.

The B97-series semiempirical functionals modifies these functionals: e.g., B97-1 functional changes the values of the semiempirical parameter, B97-D functional [71] is a dispersion correction for the B97 functional, and ωB97 functional [72] uses the long-range correction (Sec. V A) for the B97 functional.

The Mx-series semiempirical functionals [21, 22] are hybrid meta-GGA functionals, which is the combination of the PBE exchange and the B97 correlation functionals with the correction for the kinetic energy density terms,

$$E_{\mathrm{x}}^{\mathrm{M06}} = \frac{X}{100} E_{\mathrm{x}}^{\mathrm{HF}} + \left(1 - \frac{X}{100}\right) E_{\mathrm{x}}^{\mathrm{M06\text{-}DFT}}, \tag{84}$$

$$E_{\mathrm{x}}^{\mathrm{M06\text{-}DFT}} = E_{\mathrm{x}}^{\mathrm{revM05}} + E_{\mathrm{x}}^{\mathrm{revVS98}}, \tag{85}$$

$$E_{\mathrm{x}}^{\mathrm{revM05}} = \sum_{\sigma} \int d^3\mathbf{r} \bar{E}_{\mathrm{x}\sigma}^{\mathrm{PBE}}[\rho_\sigma] f_{\mathrm{x}\sigma}(w_\sigma), \tag{86}$$

where

$$f_{\mathrm{x}\sigma}(w_\sigma) = \sum_{i}^{m} C_{\mathrm{x}\sigma i} \left(\frac{\tau_\sigma^{\mathrm{LDA}} - \tau_\sigma}{\tau_\sigma^{\mathrm{LDA}} + \tau_\sigma}\right)^i, \tag{87}$$

and

$$E_{\mathrm{c}}^{\mathrm{M06}} = E_{\mathrm{c}\sigma\sigma}^{\mathrm{revB97\text{-}SIC}} + E_{\mathrm{c}\sigma\sigma}^{\mathrm{revVS98\text{-}SIC}} + E_{\mathrm{c}\alpha\beta}^{\mathrm{revB97}} + E_{\mathrm{c}\alpha\beta}^{\mathrm{revVS98}}, \tag{88}$$

$$E_{\mathrm{c}\sigma\sigma}^{\mathrm{M06}} = \int d^3\mathbf{r} D_\sigma \left(\bar{E}_{\mathrm{c}\sigma\sigma}^{\mathrm{revB97\text{-}SIC}} + \bar{E}_{\mathrm{c}\sigma\sigma}^{\mathrm{revVS98\text{-}SIC}}\right), \tag{89}$$

where the self-interaction correction term

$$D_\sigma = 1 - \frac{x_\sigma^2}{4\left(z_\sigma + C_{\mathrm{F}}\right)}, \tag{90}$$

and $\tau_\sigma^{\mathrm{LDA}} = C_{\mathrm{F}} \rho_\sigma^{5/3}$. In these equations, "rev" indicates the revision of the semiempirical parameters. Representative Mx-series functionals are M06 functional combining the M05 functional, which is the correction of the PBE exchange functional for the kinetic energy density, and the above B97 correlation functional with the VS98 exchange-correlation functional

and the HF exchange integral. This functional contains thirty six semiempirical parameters in total: X, $\gamma_{c\sigma\sigma}$, $\gamma_{c\alpha\beta}$, $C_{x\sigma i}$ (i=0 - 11), $c_{c\sigma\sigma i}$ (i=0 - 4), $c_{x\alpha\beta i}$ (i=0 - 4), d_i (i=0 - 2), $d_{c\sigma\sigma i}$ (i=0 - 4), and $d_{c\alpha\beta i}$ (i=0 - 4). The M06-2x functional removes the VS98 exchange functional part from the M06 exchange functional and consequently uses thirty five semiempirical parameters, in which d_i (i=0 - 2) are decreased. Other Mx-series functionals are the M06-L functional, which excludes the HF exchange integral (38 parameters) and the M06-HF functional (38 parameters), which uses the HF exchange integral in the exchange part. The Bose-Martin kinetic (BMK) functional [73] is also a semiempirical functional. The semiempirical parameters of these functionals have no physical meanings. That may be the reason why these semiempirical functionals are reported to worsen the calculated electron density from those of other conventional functionals [6, 7].

V. PHYSICAL CORRECTIONS FOR DFT FUNCTIONALS

Physical corrections are defined to be the physical effects that are neglected in exchange-correlation functionals and are different from terms that are introduced to give accurate electronic properties. Though all physical corrections should essentially be incorporated in any exchange-correlation functional, they are selected in accordance with calculated systems to save computational time. However, since the size of physical corrections depends on exchange-correlation functionals used, they often require parameters adjusted to these functionals. In this section, major physical corrections are introduced focusing on the physical meanings of the corrections and semiempirical parameters.

A. Long-range correction

Long-range correction [74, 75] complements long-range electron-electron exchange interactions into exchange functionals. In conventional exchange functionals, exchange interactions are usually represented by the models of several features: electron density ρ, density gradient $\nabla\rho$, density Laplacian $\nabla^2\rho$, kinetic energy density τ, and the constant mixing of the HF exchange integral $E_{\mathrm{x}}^{\mathrm{HF}}$. Since all these features but $E_{\mathrm{x}}^{\mathrm{HF}}$ are one-electron properties, these exchange functionals inevitably neglect the long-range exchange interactions. Note that the HF exchange integral covers all the long-range exchange interactions. Therefore, the long-range exchange interactions are incorporated by combining the short-range part of exchange functionals with the long-range part of the HF exchange integral. In the long-range correction, the two-electron operator, $1/r_{12}$ is first divided by, e.g., the standard error function [74],

$$\frac{1}{r_{12}} = \frac{1-\mathrm{erf}(\mu r_{12})}{r_{12}} + \frac{\mathrm{erf}(\mu r_{12})}{r_{12}}, \tag{91}$$

where μ is a semiempirical parameter for determining the division ratio. However, conventional exchange functionals are not easily divided into the short- and long-range parts, because they are mostly not derived from the divisible density matrix form. To solve this problem, exchange functionals are assumed to correspond to the momentum k_σ [74]. In this assumption, the momentum is represented by K_σ in Eq. (10) as [74]

$$k_\sigma = \left(\frac{9\pi}{K_\sigma}\right)^{1/2} \rho_\sigma^{1/3}. \tag{92}$$

For the K_σ of LDA exchange functional [76] in Eq. (8), $K_\sigma^{\mathrm{LDA}} = 3\,(3/4\pi)^{1/3}$ [75], the momentum k_σ becomes the Fermi momentum, $k_{\mathrm{F}\sigma}$ in Eq. (56). Based on this assumption, the short-range part of exchange functionals in Eq. (10) is derived as

$$E_{\mathrm{x}}^{\mathrm{LC(sr)}} = -\frac{1}{2}\sum_\sigma \int d^3\mathbf{r}\rho_\sigma^{4/3} K_\sigma \left\{1 - \frac{8}{3}a_\sigma\left[\sqrt{\pi}\mathrm{erf}\left(\frac{1}{2a_\sigma}\right) + 2a_\sigma\left(b_\sigma - c_\sigma\right)\right]\right\}, \tag{93}$$

where a_σ, b_σ, and c_σ are given as

$$a_\sigma = \frac{\mu}{2k_\sigma} = \frac{\mu K_\sigma^{1/2}}{6\sqrt{\pi}\rho_\sigma^{1/3}}, \tag{94}$$

$$b_\sigma = \exp\left(-\frac{1}{4a_\sigma^2}\right) - 1, \tag{95}$$

$$c_\sigma = 2a_\sigma^2 b_\sigma + \frac{1}{2}, \tag{96}$$

while the long-range part of the HF exchange integral is derived by multiplying the standard error function for the two-electron operator as

$$E_{\mathrm{x}}^{\mathrm{LC(lr)}} = -\frac{1}{2}\sum_\sigma\sum_i^n\sum_j^n \iint d^3\mathbf{r}_1 d^3\mathbf{r}_2 \phi_{i\sigma}^*(\mathbf{r}_1)\phi_{j\sigma}^*(\mathbf{r}_1)\frac{\mathrm{erf}\left(\mu r_{12}\right)}{r_{12}}\phi_{i\sigma}(\mathbf{r}_2)\phi_{j\sigma}(\mathbf{r}_2). \tag{97}$$

This correction contains one semiempirical parameter μ, which naturally depends on the exchange functional corrected: e.g., $\mu = 0.33$ for the B88 and PBE exchange functionals [77]. This parameter determines the division rate of the short- and long-range parts in Eq. (91). Besides the original functionals such as LC-BOP and LC-BLYP [74], representative functionals using the long-range correction are CAM-B3LYP [78], LC-ωPBE [79], and ωB97X [72], tuned range-separated hybrid [80] and short- and long-range corrected (SLC) [81] functionals.

B. Self-interaction correction

Self-interaction correction (SIC) removes the self-interaction error of exchange functionals, which erroneously remains after the cancellation of the Coulomb and exchange self-interactions due to the use of exchange functionals [82–84], from the Kohn-Sham energy,

$$E = E^{\mathrm{KS}} - \sum_i^n \left(J_{ii} + E_{\mathrm{xc}}[\rho_i]\right). \tag{98}$$

Though this equation has been used especially in many solid state calculations, it is known to have a serious problem: exchange functionals are usually not invariant for the unitary transformation differently from the HF exchange integral. To solve this problem, conventional SIC calculations have localized the Kohn-Sham orbitals before the SIC operation [85]. However, the SIC energy consequently depends on the orbital localization methods. This SIC is also reported to worsen Kohn-Sham orbital energies [86] and to cause the overestimated band gaps of semiconductors [87].

Pseudospectral regional (PR) SIC [52, 88] solves these problems by replacing exchange functionals with the HF exchange integral via the pseudospectral (PS) technique [89] only in self-interaction regions on the basis of regional SIC [90],

$$E_{\mathrm{x}}^{\mathrm{PR}} \;=\; [1 - f_{\mathrm{RS}}(t)]\, E_{\mathrm{x}}^{\mathrm{DF}} + f_{\mathrm{RS}}(t) E_{\mathrm{x}}^{\mathrm{SI}}, \tag{99}$$

where f_{RS} is a step function for clipping the self-interaction regions by way of kinetic energy density [52]. The pseudospectral HF exchange integral is the transformation of the HF exchange integral for enabling regional numerical quadrature calculations such as

$$E_{\mathrm{x}}^{\mathrm{SI}} \;=\; -\frac{1}{4} \sum_{\mu\nu\lambda\kappa} \int d^3\mathbf{r} P_{\mu\nu} P_{\lambda\kappa} \chi_\nu^*(\mathbf{r}) \chi_\lambda(\mathbf{r}) \int d^3\mathbf{r}' \frac{\chi_\kappa^*(\mathbf{r}') \chi_\mu(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|}. \tag{100}$$

where $\mathbf{r}$ and $\mathbf{r}'$ are the position vectors of electrons, $P_{\mu\nu}$ is the $\mu\nu$-component of a density matrix, and χ_λ is the λ-th atomic orbital. This PRSIC solves not only the problems of the usual SIC but also core orbital energies and core excitation energies that are seriously underestimated in usual DFT calculations [52]. This correction contains two semiempirical parameters in f_{RS}, in which the threshold for determining the regions depends on the atomic nuclear charges.

Self-interaction error is also included in many correlation functionals, as shown in the calculations of one-electron systems such as hydrogen atom. Since the LDA correlation functionals contain the self-interaction error in the parallel-spin pair ($\alpha\alpha$ and $\beta\beta$) terms, many conventional GGA correlation functionals using these LDA functionals also include this error without removal. The most efficient way to solve this problem is to use only opposite-spin pair ($\alpha\beta$) terms as done in the OP correlation functional [56].

C. Van der Waals dispersion correction

Van der Waals (vdW) correction complements van der Waals electron correlations, which are essentially neglected in most correlation functionals. VdW interactions include dipole-dipole, dipole-induced dipole, and dispersion interactions [91]. Note that these vdW interactions are included in the Coulomb-exchange interactions, which are already included in usual Kohn-Sham equations, except for the dispersion interactions.

The dispersion interaction is an electron correlation, which universally acts between bodies. As a classical expression, the London potential function between two heterogeneous bodies, which is developed on the basis of perturbation theory, is used [92],

$$V_{\rm disp}^{\rm London}(\mathbf{r}) = -\frac{3}{2}\frac{\alpha_{\rm A}\alpha_{\rm B}}{{R_{\rm AB}}^6}\frac{I_{\rm A}I_{\rm B}}{I_{\rm A}+I_{\rm B}}, \tag{101}$$

where $R_{\rm AB}$ is the distance between systems A and B, $\alpha_{\rm X}$ is the linear response for an electric field producing an induced dipole moment of nonpolar system X, and $I_{\rm X}$ is the ionization potential of partial system X. Since this dispersion interaction is a long-range electron-electron correlation, it is not usually incorporated in correlation functionals. The dispersion interaction, therefore, should be explicitly supplemented in correlation functionals.

Various dispersion corrections have so far been suggested: classical dispersion corrections [92], semiempirical dispersion-corrected functionals, combinations with perturbation theories, linear-response theories, and van der Waals (dispersion) functionals. The simplest dispersion correction is based on the London potential function,

$$E_{\rm disp}^{\rm London} = -\sum_{\rm A>B}\frac{C_6^{\rm AB}}{{R_{\rm AB}}^6}f_{\rm damp}(R_{\rm AB}), \tag{102}$$

where A and B are usually the labels of atoms, $C_6^{\rm AB}$ is a semiempirical parameter dependent on the atomic species A and B, and $f_{\rm damp}$ is a damping function for cutting off unnecessary short-range interactions. Since this correction well reproduces dispersion interactions between distant systems, it is widely used in classical molecular dynamics (MD) simulations. However, this correction is not applicable to intramolecular dispersion interactions, because it is difficult to determine a system-independent damping factor for small $R_{\rm AB}$.

The semiempirical dispersion-corrected functionals modify correlation functionals. Representative semiempirical dispersion-corrected functionals are DFT-D functionals [71]. The

DFT-D functionals have three levels: DFT-D1, DFT-D2, and DFT-D3 [93],

$$E^{\text{DFT-D2}} = E^{\text{KS}} + E^{(2)}_{\text{disp}} \tag{103}$$

$$E^{\text{DFT-D3}} = E^{\text{KS}} + E^{(2)}_{\text{disp}} + E^{(3)}_{\text{disp}} \tag{104}$$

$$E^{(2)}_{\text{disp}} = -\sum_{\text{A>B}} \sum_{n=6,8,10,\cdots} s_n \frac{C^{\text{AB}}_n}{{R_{\text{AB}}}^n} f^n_{\text{damp}}(R_{\text{AB}}), \tag{105}$$

$$E^{(3)}_{\text{disp}} = -\sum_{\text{A>B>C}} \frac{C^{\text{ABC}}_9 (3\cos\theta_a \cos\theta_b \cos\theta_c + 1)}{(R_{\text{AB}} R_{\text{BC}} R_{\text{CA}})^3} f^{(3)}_{\text{damp}}(\bar{R}_{\text{ABC}}), \tag{106}$$

where A, B and C are atomic labels, θ_a, θ_b, and θ_c are the internal angles of the ABC triangles, and $\bar{R}_{\text{ABC}}$ are geometrically averaged radii. The damping functions are given by

$$f^n_{\text{damp}}(R_{\text{AB}}) = \left\{1 + 6\left(R_{\text{AB}}/(s_{r,n} R^{\text{AB}}_0)\right)^{-\alpha_n}\right\}^{-1}, \tag{107}$$

$$f^{(3)}_{\text{damp}}(\bar{R}_{\text{ABC}}) = \left\{1 + 6\left(\bar{R}_{\text{ABC}}/(4R^{\text{ABC}}_0/3)\right)^{-16}\right\}^{-1}. \tag{108}$$

For dispersion coefficients, C^{AB}_n and C^{ABC}_9, time-dependent Kohn-Sham (TDKS) equation [94] and recursion relations are used to determine the values for each atomic pair and trio [95, 96]. All of the remaining parameters, R^{AB}_0, R^{ABC}_0, coefficients s_n ($n = 8, 10, \cdots$) and s_6, are semiempirical parameters adjusted to weakly-interacting systems [93].

Perturbation theories such as the MP2 method [97] requires no damping factors to be applied to either intermolecular or intramolecular dispersion interactions,

$$E_{\text{MP2}} = -\sum_{i<j}^{n_{\text{occ}}} \sum_{a<b}^{n_{\text{vir}}} \frac{|\langle ij|ab\rangle - \langle ij|ba\rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}, \tag{109}$$

where ϵ_i is the i-th orbital energy and

$$\langle ij|ab\rangle = \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 \phi^*_i(\mathbf{r}_1)\phi^*_j(\mathbf{r}_2)\frac{1}{r_{12}}\phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_2). \tag{110}$$

Representative perturbation-based dispersion-corrected DFT are DFT symmetry-adapted perturbation theory (DFT-SAPT) [98] and double-hybrid functionals [99]. Although DFT-SAPT is a promising dispersion calculation method for clearly separated systems, it cannot reproduce intramolecular dispersion interactions and needs much computational time. The double-hybrid functionals such as B2PLYP functional [23] mix perturbation energies in correlation functionals at a constant ratio with two semiempirical parameters a_{x} and a_{c},

$$E_{\text{xc}} = (1 - a_{\text{x}})E_{\text{x}} + a_x E_{\text{HF}} + (1 - a_{\text{c}})E_{\text{c}} + a_{\text{c}} E_{\text{MP2}}. \tag{111}$$

Although this method gives middle-range electron correlations, it does not cover the whole long-range dispersion interactions.

Linear-response theories give dispersion interactions in the framework of the Kohn-Sham method. Representative linear-response theories in DFT is adiabatic connection / fluctuation-dissipation theorem (AC/FDT) method [100]. The AC/FDT method calculates electron correlation as the energy response quantity for the spontaneous fluctuations of electronic motions coming from the perturbation of the interelectronic interactions,

$$E_{\mathrm{c}} \;=\; -\int_0^1 d\lambda \iint d^3\mathbf{r} d^3\mathbf{r}' \frac{1}{|\mathbf{r}-\mathbf{r}'|}\left[\frac{1}{2\pi}\int_0^\infty du\left\{\chi_\lambda\left(\mathbf{r},\mathbf{r}',\mathrm{i}u\right)-\chi_0\left(\mathbf{r},\mathbf{r}',\mathrm{i}u\right)\right\}\right], \quad (112)$$

where χ_λ and χ_0 are density response functions for interacting and independent electrons, respectively, which are calculated by solving the Dyson equation,

$$\begin{aligned}\chi_\lambda\left(\mathbf{r},\mathbf{r}',\omega\right) \;=\; & \chi_0\left(\mathbf{r},\mathbf{r}',\omega\right)+\iint d^3\mathbf{r}_1 d^3\mathbf{r}_2 \chi_0\left(\mathbf{r},\mathbf{r}_1,\omega\right) \\ & \times\left\{\frac{\lambda}{|\mathbf{r}_1-\mathbf{r}_2|}+f_{\mathrm{xc}}^\lambda\left(\mathbf{r}_1,\mathbf{r}_2,\omega\right)\right\}\chi_\lambda\left(\mathbf{r}_2,\mathbf{r}',\omega\right), \end{aligned} \quad (113)$$

where f_{xc}^λ is the exchange-correlation integral kernel for interacting systems. The correlation energy in Eq. (112) is calculated by solving the TDKS equation [94]. The long-range part of this correlation energy is used as the dispersion correction called "RPAx" dispersion correction [101]. The AC/FDT correlation energy also includes somewhat nondynamical correlation effects. This AC/FDT method is physically the most superior dispersion correction but requires much more computational time than the Kohn-Sham calculations.

VdW (dispersion) functionals reduce the computational time of the AC/FDT method while maintaining accuracy. Representative vdW functionals are Andersson-Langreth-Lundqvist (ALL96) [102], Dobson-Dinte (DD96) [103], Dion-Rydberg-Schröder-Langreth-Lundqvist (DRSLL) [104], Vydrov-van Voorhis 2009 (VV09) [105], and local response dispersion (LRD) [106] functionals. The ALL functional uses LDA for the electron density response function of the AC/FDT method [102],

$$E_{\mathrm{disp}}^{\mathrm{ALL}}\left[\rho\right] = -\frac{6}{4\pi^{3/2}}\int_{\mathrm{V}_1} d^3\mathbf{r}_1 \int_{\mathrm{V}_2} d^3\mathbf{r}_2 \frac{\sqrt{\rho(\mathbf{r}_1)}\sqrt{\rho(\mathbf{r}_1)}}{\sqrt{\rho(\mathbf{r}_1)}+\sqrt{\rho(\mathbf{r}_1)}}\frac{1}{r_{12}^6}. \quad (114)$$

This is occasionally the same functional as the DD96 functional, which was developed in the same year [103]. The computational time of this functional is usually less than that of exchange-correlation functionals despite of the numerical two-electron integrals included, because the integral calculations usually neglect spatial regions of small momentum variations

and core regions. However, this functional needs damping functions $f_{\rm damp}$ for short-range electron-electron distances, similar to Eq. (102) and is, therefore, applicable to the interactions of well-separated electron distributions. The DRSLL functional is a dispersion functional available for the short-range electron-electron distances [104],

$$E_{\rm disp}^{\rm DRSLL}[\rho] = \int d^3\mathbf{r}_1 \int d^3\mathbf{r}_2 \rho(\mathbf{r}_1)\phi(\mathbf{r}_1,\mathbf{r}_2)\rho(\mathbf{r}_2), \tag{115}$$

$$\phi(\mathbf{r}_1,\mathbf{r}_2) = \frac{2}{\pi^2}\int_0^\infty daa^2 \int dbb^2 W(a,b) T\left(\nu_1(a),\nu_1(b),\nu_2(a),\nu_2(b)\right), \tag{116}$$

where

$$T(w,x,y,z) = \frac{1}{2}\left[\frac{1}{w+x}+\frac{1}{y+x}\right]\left[\frac{1}{(w+y)(x+z)}+\frac{1}{(w+z)(y+x)}\right], \tag{117}$$

$$W(a,b) = \frac{2}{a^3b^2}\left[(3-a^2)b\cos b\sin a + (3-b^2)a\cos a\sin b \right. \\ \left. +(a^2+b^2-3)\sin a\sin b - 3ab\cos a\cos b\right], \tag{118}$$

$\nu_i(y) = y^2/2\left(1-\exp(-4\pi y^2/9d_i^2)\right)$ and $d_i = r_{12}q_0(\mathbf{r}_i)$. The q_0 is given using the Fermi momentum, $k_{\rm F} = (3\pi^2\rho)^{1/3}$, as

$$q_0(\mathbf{r}) = k_{\rm F}(\mathbf{r})\left[1+0.09434\left(\frac{\boldsymbol{\nabla}\rho(\mathbf{r})}{2k_{\rm F}(\mathbf{r})\rho(\mathbf{r})}\right)^2\right]. \tag{119}$$

This functional requires no damping function, because it naturally approaches zero for short electron-electron distances. Therefore, this functional can reproduce intramolecular dispersion interactions. VdW-DF method [104] combines these vdW functionals with GGA exchange-correlation functionals. For other vdW functionals that can calculate intramolecular dispersion interactions, the VV09 functional [105] uses the dielectric model smoothing the real-space cutoff, the LRD functional [106] combines the DD96 local response approximation functional [103] with the real-space cutoff [105].

VdW corrections inevitably depend on correlation functionals used. Actually, these vdW corrections contain semiempirical parameters in, e.g., the damping functions. For sophisticating the vdW corrections, it is, therefore, required to reduce the number of the semiempirical parameters in the damping functions.

Note, finally, that repulsions balanced with dispersion attractions are equivalently significant in dispersion calculations. That is, it is actually known that the repulsions from the long-range correction for exchange functionals in Sec. (V A) are well balanced with dispersion attractions to give the most accurate dispersion binding energies [107]. This is called "LC+vdW" method [107].

D. Relativistic corrections

Relativistic corrections supplement the relativity theory-based effect on electronic motions [108, 109]. In particular, the relativistic corrections in the field of QC are based on the special relativity theory [108], in which the principle of relativity establishing the invariance of physical laws for the Lorentz transformation and the principle of the constancy of the speed of light for coordinate systems moving at a constant speed in Minkovski's four-dimensional space-time coordinates are established.

The Schrödinger equation is relativistically invariant for the Lorentz transformation. The Dirac equation is developed to make the kinetic energy part of the Schrödinger equation invariant for the Lorentz transformation [110],

$$\hat{H}^{\mathrm{D}}\Psi = \left[c\boldsymbol{\alpha}\cdot\hat{\mathbf{p}} + \boldsymbol{\beta}mc^2 + V\right]\Psi = i\frac{\partial\Psi}{\partial t}, \tag{120}$$

where

$$\beta = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{pmatrix},\ \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{121}$$

$$\boldsymbol{\alpha}_w = \begin{pmatrix} \mathbf{0} & \boldsymbol{\sigma}_w \\ \boldsymbol{\sigma}_w & \mathbf{0} \end{pmatrix} \quad (w = x, y, z), \tag{122}$$

$$\boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{123}$$

The electron mass m and the speed of light c, which are both unity in atomic units, are explicitly written in this section. The $\boldsymbol{\sigma}$ is called the Pauli spin matrix [111]. This equation is invariant for the Lorentz transformation, because the momentum $\mathbf{p} = -i\boldsymbol{\nabla}$ is the first derivative in terms of space.

The Dirac equation leads to two significant findings [110]: One finding is that the wavefunction becomes four-component,

$$\Psi = \begin{pmatrix} \Psi^{\mathrm{L}}_{\alpha} \\ \Psi^{\mathrm{L}}_{\beta} \\ \Psi^{\mathrm{S}}_{\alpha} \\ \Psi^{\mathrm{S}}_{\beta} \end{pmatrix}. \tag{124}$$

Another finding is that there must be continuum states below $-2mc^2$ in energy and these continuum states are interpreted to be occupied by an infinite number of positrons. Since the β term in Eq. (120) provides the inconveniently large rest energy, 5.11×10^5 eV, the β is replaced with

$$\beta' = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -2\mathbf{I} \end{pmatrix}. \tag{125}$$

This β' causes this finding, though this is disproven by quantum electrodynamics except for the presence of positrons. This finding leads to the interpretation that in Eq. (124), large-component wavefunction Ψ^{L} indicates the wavefunction of electronic motions, while small-component wavefunction Ψ^{S} implies an electronic wavefunction affected by positronic motions.

The time-independent Dirac equation is given as

$$\left[c\boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + \boldsymbol{\beta}' mc^2 + V\right] \Psi = E\Psi \tag{126}$$

Since this equation is written as

$$c(\boldsymbol{\sigma} \cdot \hat{\mathbf{p}})\Psi^{\mathrm{S}} + V\Psi^{\mathrm{L}} = E\Psi^{\mathrm{L}} \tag{127}$$

$$c(\boldsymbol{\sigma} \cdot \hat{\mathbf{p}})\Psi^{\mathrm{L}} + (-2mc^2 + V)\Psi^{\mathrm{S}} = E\Psi^{\mathrm{S}}, \tag{128}$$

the small-component wavefunction is represented by the large-component one,

$$\Psi^{\mathrm{S}} = (E + 2mc^2 - V)^{-1} c(\boldsymbol{\sigma} \cdot \hat{\mathbf{p}})\Psi^{\mathrm{L}} = \left(1 + \frac{E - V}{2mc^2}\right)^{-1} \frac{\boldsymbol{\sigma} \cdot \hat{\mathbf{p}}}{2mc} \Psi^{\mathrm{L}}, \tag{129}$$

Based on Eq. (129), the time-independent Dirac equation is obtained as

$$\left[\frac{1}{2m}(\boldsymbol{\sigma} \cdot \hat{\mathbf{p}}) \left(1 + \frac{E - V}{2mc^2}\right)^{-1} (\boldsymbol{\sigma} \cdot \hat{\mathbf{p}}) + V\right] \Psi^{\mathrm{L}} = E\Psi^{\mathrm{L}}. \tag{130}$$

Assuming that the speed of light would be infinity, this equation gives the nonrelativistic Schrödinger equation,

$$\left[\frac{\hat{\mathbf{p}}^2}{2m} + V\right] \Psi^{\mathrm{L}} = E\Psi^{\mathrm{L}}. \tag{131}$$

Relativistic corrections are also significant in the nuclear-electron and electron-electron potentials, i.e., nuclear-electron potential $1/r$ and electron-electron potential $1/r_{12}$, because

these potentials are variant for the Lorentz transformation due to the dependence on space. That is, the retardation is present for long-range interactions compared to short-range interactions. Though this retardation is not easy to be incorporated, a simpler formulation is derived using the Taylor expansion of the electron-electron interaction potential in terms of the fine structure constant $1/c$,

$$V_{\mathrm{ee}}(\mathbf{r}_{12}) = \frac{1}{r_{12}} - \frac{1}{r_{12}}\left[\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 + \frac{(\boldsymbol{\alpha}_1 \times \mathbf{r}_{12})(\boldsymbol{\alpha}_1 \times \mathbf{r}_{12})}{r_{12}^2}\right] \tag{132}$$

$$= \frac{1}{r_{12}} - \frac{1}{2r_{12}}\left[\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 + \frac{(\boldsymbol{\alpha}_1 \cdot \mathbf{r}_{12})(\boldsymbol{\alpha}_1 \cdot \mathbf{r}_{12})}{r_{12}^2}\right], \tag{133}$$

where $\boldsymbol{\alpha}_i$ is the α matrix in Eq. (122) for the i-th particle. For the nuclear-electron potential, the relativistic correction is usually neglected, because it is on the order of $1/c^3$. The relativistic correction for the electron-electron interaction potential is called the Breit interaction [112], in which $-\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2/r_{12}$ is separately called the Gaunt interaction [113]. This Breit interaction has the order of $1/c^2$, because it contains no $(1/c)$-order term. Even for this Breit interaction, the high-order perturbation terms are considerably complicated. Although these terms also require long computational times, they have a minimal effect on chemical reactions and properties. Therefore, these terms are usually neglected in QC calculations, even for heavy atoms.

Though the relativistic correction for the kinetic energy in the Dirac equation is applicable to the Kohn-Sham equation as done in the Dirac-Kohn-Sham method [114, 115], the Dirac-Kohn-Sham calculations require a huge computational time that is approximately 25 times longer than that of the Kohn-Sham calculations due to the use of the small-component wavefunction, producing the coupling of the large-component and small-component wavefunctions. Two-component relativistic approximations are suggested to solve this problem. In the two-component approximations, $\hat{K}$ is expanded for slow electronic motions compared to the speed of light,

$$\hat{K} = \left(1 + \frac{E - V}{2mc^2}\right)^{-1} \approx 1 - \frac{E - V}{2mc^2} + \cdots \tag{134}$$

up to the order of $1/c^2$ to the Dirac equation, such as

$$\left[\frac{\hat{\mathbf{p}}^2}{2m} + V - \frac{\hat{\mathbf{p}}^4}{8m^3c^2} + \frac{Z\hat{\mathbf{s}} \cdot \hat{\mathbf{l}}}{2m^2c^2r^3} + \frac{Z\pi\delta(\mathbf{r})}{2m^2c^2}\right]\Psi^{\mathrm{L}} = E\Psi^{\mathrm{L}}, \tag{135}$$

where $\hat{\mathbf{l}}$ is the orbital angular momentum operator and δ is the delta function. This equation is called the Breit-Pauli equation [112]. Compared to Eq. (131), the additional terms

are the mass velocity correction, which is the effect of velocity on mass, the spin-orbit interaction, which is the magnetic interaction between electronic spin and the orbital, and the Darwin correction, which comes from the high-frequency vibrational electronic motion around the equilibrium geometry, in sequence. The mass velocity and Darwin corrections are collectively called the scalar relativistic correction. However, the expansion of this $\hat{K}$ is often inappropriate due to the divergence of the potential ($V \to \infty$) near nuclei. To avoid this divergence, $\hat{K}$ is frequently replaced with $\hat{K}'$ defined by

$$(E + 2mc^2 - V)^{-1} = (2mc^2 - V)^{-1} \left(1 + \frac{E}{2mc^2 - V}\right)^{-1} = (2mc^2 - V)^{-1}\hat{K}'. \quad (136)$$

By this approximation, the divergence of the expansion due to $E/(2mc^2 - V) \ll 1$ is avoided. The $\hat{K}'$ is also approximated as unity in the zeroth-order regular approximation (ZORA),

The Foldy-Wouthuysen transformation [116] is another major two-component approximation. This approximation makes the large-component and small-component submatrices of the Dirac Hamiltonian matrix $\hat{\mathbf{H}}_{\mathrm{D}}$ linear independent by a unitary transformation,

$$\hat{\mathbf{H}}^{\mathrm{FW}} = U\hat{\mathbf{H}}^{\mathrm{D}}U^{\dagger} = \begin{pmatrix} \hat{H}^{\mathrm{L}} & 0 \\ 0 & \hat{H}^{\mathrm{S}} \end{pmatrix}. \quad (137)$$

For free electrons ($V = 0$), this transformation is written as

$$U_0 = A_p(1 + \beta R_p), \quad (138)$$

where

$$A_p = \left(\frac{E_p + mc^2}{2E_p}\right)^{1/2}, \quad (139)$$

$$E_p = \left(\hat{\mathbf{p}}^2c^2 + m^2c^4\right)^{1/2}, \quad (140)$$

$$R_p = \frac{c\boldsymbol{\alpha} \cdot \hat{\mathbf{p}}}{E_p + mc^2}. \quad (141)$$

This transformation leads to the Hamiltonian operator for using a potential as

$$\hat{\mathbf{H}}^{\mathrm{FW}} = U_0\hat{\mathbf{H}}^{\mathrm{D}}U_0^{\dagger} = \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{O}_1, \quad (142)$$

where

$$\mathcal{E}_0 = \beta E_p - mc^2, \quad (143)$$

$$\mathcal{E}_1 = A_p(V + R_pVR_p)A_p, \quad (144)$$

$$\mathcal{O}_1 = \beta A_p \left[R_p, V\right] A_p. \quad (145)$$

This formulation includes the lowest-order off-diagonal term of the Hamiltonian matrix, $\mathcal{O}_1$, which leads to the production of singular points. However, these singular points can be eliminated by further appropriate unitary transformations. In this strategy, the Douglas-Kroll transformation solves the singular point problem [117],

$$\hat{\mathbf{H}}^{\mathrm{DK}} = U\hat{\mathbf{H}}^{\mathrm{D}}U^{\dagger} = \cdots U_4U_3U_2U_1\hat{\mathbf{H}}^{\mathrm{FW}}U_1^{\dagger}U_2^{\dagger}U_3^{\dagger}U_4^{\dagger}\cdots = \sum_{i=0}^{\infty}\mathcal{E}_i. \tag{146}$$

This summation is usually terminated at particular numbers of unitary transformations: 2 in DK2, 3 in DK3 and so forth. This method is also called the Douglas-Kroll-Hess transformation [118].

Since these relativistic corrections are fundamentally carried out only for the kinetic energy, which is not approximated in the (generalized) Kohn-Sham method, these relativistic corrections need no semiempirical parameters to be implemented. However, if the kinetic energy will be approximated, the relativistic effects inevitably need semiempirical parameters and give huge dependence on the approximated form and the parameters due to the enormous size of the kinetic energy.

VI. CONCLUSIONS

In this review, the development of DFT functionals and their physical corrections is explored focusing on their physical meanings and the semiempirical parameters included from the viewpoint of data science. DFT exchange-correlation functionals have been developed under many strict physical conditions with minimizing the number of the semiempirical parameters, except for some recent functionals like semiempirical functionals. Physical corrections for exchange-correlation functionals have clear physical meanings independent of the DFT functionals, though they inevitably require minimum semiempirical parameters dependent on the functionals combined. Considering their remarkable accuracy in electronic property calculations, DFT functionals with physical corrections are, therefore, interpreted to be the most sophisticated target functions that are physically legitimated, even from the viewpoint of data science. However, a minimum of semiempirical parameters are included dependent on the combination of the functionals and the corrections. Conventional data-scientific methods will propose the optimum semiempirical parameters through many channels.

ACKNOWLEDGMENTS

We would like to pay honors to Prof. Leo Radom of the University of Sydney for his great achievements. This research was supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) (Grants: 17H01188 and 16KT0047).

Figure Captions

Fig. 1.

Classification of major exchange-correlation functionals. The suffix "X" and "C" indicate exchange and correlation functionals, respectively.

Fig. 2.

The plot of K_σ, defined in Eq. (10), in terms of x_σ in Eq. (9) for major GGA exchange functionals.

Fig. 3.

Form of the correlation function for the Colle-Salvetti-type correlation functionals. $\Psi_{\rm uncorr}$ is a wavefunction containing no electron correlation effect.

[1] T. Tsuneda, *Density Functional Theory in Quantum Chemistry* (Springer, Tokyo, 2014).

[2] W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).

[3] J. D. Kelleher, B. Mac Namee, and A. D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics Algorithms, Worked examples, and Case Studies* (The MIT Press, Cambridge, UK, 2014).

[4] M. Rupp, O. A. von Lilienfeld, and K. Burke, J. Chem. Phys. **148**, 241401(1 (2018).

[5] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[6] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, Science **355**, 49 (2017).

[7] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, Science **356**, 496c (2017).

[8] J. Hollingsworth, L. Li, T. E. Baker, and K. Burke, J. Chem. Phys. **148**, 241743(1 (2018).

[9] J. Gasteiger and T. Engel, eds., *Chemoinformatics* (Wiley-VCH, Weinheim, 2003).

[10] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Nature Commun. **8**, 872(1 (2017).

[11] 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, J. Chem. Theory Comput. **13**, 5255 (2017).

[12] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, Phys. Rev. Lett. **108**, 253002(1 (2012).

[13] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, Phys. Rev. B **88**, 054104(1 (2013).

[14] N. Mardirossian and M. Head-Gordon, J. Chem. Phys. **148**, 241736(1 (2018).

[15] R. Nagai, R. Akashi, S. Sasaki, and S. Tsuneyuki, J. Chem. Phys. **148**, 241737(1 (2018).

[16] K. Yao and J. Parkhill, J. Chem. Theory Comput. **12**, 1139 (2016).

[17] J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, J. Chem. Phys. **148**, 241705(1 (2018).

[18] P. Pernot and A. Savin, J. Chem. Phys. **148**, 241707(1 (2018).

[19] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, Phys. Rev. B **53**, 3764 (1986).

[20] P.-O. Löwdin, Phys. Rev. **97**, 1509 (1955).

[21] Y. Zhao and D. G. Truhlar, J. Chem. Phys. **125**, 194101(1 (2006).

[22] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008).

[23] S. Grimme, J. Chem. Phys. **124**, 034108 (2006).

[24] J. D. Talman and W. F. Shadwick, Phys. Rev. A **14**, 36 (1976).

[25] J. B. Krieger, Y. Li, and G. J. Iafrate, Phys. Rev. A **45**, 101 (1992).

[26] Q. Zhao, R. C. Morrison, and R. G. Parr, Phys. Rev. A **50**, 2138 (1984).

[27] Q. Wu and W. Yang, J. Chem. Phys. **116**, 515 (2001).

[28] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, et al., *Gaussian09 Revision D.01* (2009), gaussian Inc. Wallingford CT.

[29] M. S. Gordon and M. W. Schmidt, in *Theory and Applications of Computational Chemistry: the first forty years*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Amsterdam, 2005), pp. 1167–1189.

[30] J. C. Slater, Phys. Rev. **81**, 385 (1951).

[31] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[32] J. P. Perdew, in *Electronic structure of solids '91*, edited by P. Ziesche and H. Eschrigh (Akademie Verlag, Berlin, 1991).

[33] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[34] Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).

[35] L. Kleinman and S. Lee, Phys. Rev. B **37**, 4634 (1988).

[36] T. Tsuneda and K. Hirao, Phys. Rev. B **62**, 15527 (2000).

[37] E. H. Lieb and S. Oxford, Int. J. Quantum Chem. **19**, 427 (1981).

[38] M. Levy, J. P. Perdew, and V. Sahni, Phys. Rev. A **30**, 2745 (1984).

[39] P. Mori-Sanchez, A. Cohen, and W. Yang, J. Chem. Phys. **125**, 201102(1 (2006).

[40] T. Tsuneda, J.-W. Song, S. Suzuki, and K. Hirao, J. Chem. Phys. **133**, 174101(1 (2010).

[41] J. W. Negele and D. Vautherin, Phys. Rev. C **5**, 1472 (1972).

[42] L. H. Thomas, Proc. Cam. Phyl. Soc. **23**, 542 (1927).

[43] E. Fermi, Z. Phys. **48**, 73 (1928).

[44] C. F. Weizsäcker, Z. Phys. **96**, 431 (1935).

[45] M. Gell-Mann and K. A. Bruecker, Phys. Rev. **106**, 364 (1957).

[46] W. J. Carr, Phys. Rev. **122**, 1437 (1961).

[47] L. Hedin and B. I. Lundqvist, J. Phys. C: Solid State Phys. **4**, 2064 (1971).

[48] T. Tsuneda, M. Kamiya, N. Morinaga, and K. Hirao, J. Chem. Phys. **114**, 6505 (2001).

[49] S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).

[50] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[51] R. Colle and O. Salvetti, Theor. Chim. Acta **37**, 329 (1975).

[52] A. Nakata and T. Tsuneda, J. Chem. Phys. **139**, 064102(1 (2013).

[53] S. K. Ma and K. A. Brueckner, Phys. Rev. **165**, 18 (1968).

[54] R. M. Dreizler and E. K. U. Gross, *Density-Functional Theory An Approach to the Quantum Many-Body Problem* (Springer, Berlin, 1990).

[55] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[56] T. Tsuneda, T. Suzumura, and K. Hirao, J. Chem. Phys. **110**, 10664 (1999).

[57] T. van Voorhis and G. E. Scuseria, J. Chem. Phys. **109**, 400 (1998).

[58] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha, Phys. Rev. Lett. **82**, 2544 (1999).

[59] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401(1 (2008).

[60] J. Sun, A. Ruzsinszky, and J. P. Perdew, Phys. Rev. Lett. **115**, 036402(1 (2015).

[61] K. Hui and J.-D. Chai, J. Chem. Phys. **144**, 044114(1 (2016).

[62] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[63] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).

[64] S. N. Pieniazek, F. R. Clemente, and K. N. Houk, Angew. Chem. Int. Ed. **47**, 7746 (2008).

[65] S. E. Wheeler, A. Moran, S. N. Pieniazek, and K. N. Houk, J. Phys. Chem. A **38**, 10376 (2009).

[66] M. Grimme, S. Steinmetz and M. Korth, J. Org. Chem. **72**, 2118 (2007).

[67] A. Dreuw, J. L. Weisman, and M. Head-Gordon, J. Chem. Phys. **119**, 2943 (2003).

[68] B. Champagne, E. A. Perpéte, and D. Jacquemin, J. Phys. Chem. A **104**, 4755 (2000).

[69] A. D. Becke, J. Chem. Phys. **107**, 8554 (1997).

[70] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998).

[71] J. Antony and S. Grimme, Phys. Chem. Chem. Phys. **8**, 5287 (2006).

[72] J.-D. Chai and M. Head-Gordon, J. Chem. Phys. **128**, 084106(1 (2008).

[73] A. D. Boese and J. M. L. Martin, J. Chem. Phys. **121**, 3405 (2004).

[74] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).

[75] A. Savin, in *Recent Developments and Applications of Modern Density Functional Theory*, edited by J. J. Seminario (Elsevier, Amsterdam, 1996), pp. 327–357.

[76] P. A. M. Dirac, Camb. Phil. Soc. **26**, 376 (1930).

[77] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao, J. Chem. Phys. **120**, 8425 (2004).

[78] T. Yanai, D. P. Tew, and N. C. Handy, Chem. Phys. Lett. **91**, 51 (2004).

[79] O. A. Vydrov, J. Heyd, A. Krukau, and G. E. Scuseria, J. Chem. Phys. **125**, 074106(1 (2006).

[80] R. Baer, E. Livshits, and U. Salzner, Ann. Rev. Phys. Chem. **61**, 85 (2010).

[81] C.-W. Wang, K. Hui, and J.-D. Chai, J. Chem. Phys. **145**, 204101(1 (2016).

[82] D. R. Hartree, Math. Proc. Camb. Phil. Soc. **24**, 89 (1928).

[83] E. Fermi and E. Amaldi, Accad. Ital. Rome **6**, 117 (1934).

[84] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[85] B. G. Johnson, C. A. Gonzales, P. M. W. Gill, and J. A. Pople, Chem. Phys. Lett. **221**, 100 (1994).

[86] O. A. Vydrov, G. E. Scuseria, and J. P. Perdew, J. Chem. Phys. **126**, 154109(1 (2007).

[87] I. G. Gerber, J. G. Angyan, M. Marsman, and G. Kresse, J. Chem. Phys. **125**, 054101(1 (2007).

[88] A. Nakata, T. Tsuneda, and K. Hirao, J. Phys. Chem. A **114**, 8521 (2010).

[89] S. A. Orszag, Stud. Appl. Math. **51**, 253 (1972).

[90] T. Tsuneda, M. Kamiya, and K. Hirao, J. Comput. Chem. **24**, 1592 (2003).

[91] J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1992).

[92] F. W. London, Z. Phys. **63**, 245 (1930).

[93] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[94] T. Tsuneda and K. Hirao, in *Theoretical and Quantum Chemistry at the 21st Century Dawn End*, edited by T. Chakraborty and R. Carbo (Apple Academic Press, Florida, 2018), pp. 153–195.

[95] H. Casimir and D. Polder, Phys. Rev. **73**, 360 (1948).

[96] G. Starkschall and R. Gordon, J. Chem. Phys. **56**, 2801 (1972).

[97] R. McWeeny, *Methods of Molecular Quantum Mechanics 2nd Ed.* (Academic Press, San Diego, 1992).

[98] H. L. Williams and C. F. Chabalowski, J. Phys. Chem. A **105**, 646 (2001).

[99] T. Schwabe and S. Grimme, Phys. Chem. Chem. Phys. **9**, 3397 (2007).

[100] D. C. Langreth and J. P. Perdew, Solid State Commun. **17**, 1425 (1975).

[101] W. Zhu, J. Toulouse, A. Savin, and J. G. Angyan, J. Chem. Phys. **131**, 174105 (2009).

[102] Y. Andersson, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **76**, 102 (1996).

[103] J. F. Dobson and B. P. Dinte, Phys. Rev. Lett. **76**, 1780 (1996).

[104] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004).

[105] O. A. Vydrov and T. van Voorhis, J. Chem. Phys. **130**, 104105 (2009).

[106] T. Sato and H. Nakai, J. Chem. Phys. **131**, 224104 (2009).

[107] M. Kamiya, T. Tsuneda, and K. Hirao, J. Chem. Phys. **117**, 6010 (2002).

[108] A. Einstein, Ann. Phys. **17**, 891 (1905).

[109] A. Einstein, Ann. Phys. **49**, 769 (1916).

[110] P. A. M. Dirac, Proc. Roy. Soc. A **117**, 610 (1928).

[111] W. Pauli, Z. Phys. **31**, 765 (1925).

[112] G. Breit, Phys. Rev. **34**, 553 (1929).

[113] I. P. Gaunt, Proc. R. Soc. Lond. A **124**, 163 (1929).

[114] A. K. Rajagopal, J. Phys. C: Solid State Phys. **11**, L943 (1978).

[115] A. H. MacDonald and S. H. Vosko, J. Phys. C: Solid State Phys. **12**, 2977 (1979).

[116] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).

[117] M. Douglas and N. M. Kroll, Ann. Phys. **82**, 89 (1974).

[118] G. Jansen and B. A. Hess, Phys. Rev. A **39**, 6016 (1989).