# Can machines learn density functionals? Past, present, and future of ML in DFT

Ryosuke Akashi[0009−0003−9399−2335], Mihira Sogal and
Kieron Burke[0000−0002−6159−0054]

**Abstract**
Density functional theory has become the world's favorite electronic structure method, and is routinely applied to both materials and molecules. Here, we review recent attempts to use modern machine-learning to improve density functional approximations. Many different researchers have tried many different approaches, but some common themes and lessons have emerged. We discuss these trends and where they might bring us in the future.

Ryosuke Akashi
National Institutes for Quantum Science and Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan, e-mail: `akashi.ryosuke@qst.go.jp`

Mihira Sogal and Kieron Burke
University of California, Irvine, CA 92697, USA

# Contents

# 1 Introduction

Machine learning (ML) allows us to develop a calculable and reusable surrogate of any task comprising of regression (including classification or pattern recognition). After conquering human pattern recognition [1] and some famous games [2, 3], this technology is rapidly percolating into every branch of quantum chemistry and physics as a provider of versatile tools. In these fields a general aim is to solve the electronic many-body Schrödinger equation. The human process of solving the Schrödinger equation is as follows: first, relevant degrees of freedom must be defined for the system, e.g. spatial and spin coordinates of atoms and electrons; next, the Hamiltonian corresponding to the system can be written explicitly. Finally, the Hamiltonian is diagonalized to find the wavefunctions (eigenvectors) and corresponding energies (eigenvalues) of the system. The exact numerical solution of the Schrödinger equation is extremely costly, scaling exponentially in the number of electrons in the system. As a result, exact diagonalization of the Hamiltonian is seldom computationally feasible (for realistic Hamiltonians), and approximate methods must be used instead.

The most widely used method, and the focus of this Chapter, is density functional theory (DFT), for which Walter Kohn shared the Nobel Prize in Chemistry in 1998 [4–6]. DFT is an alternative formulation of the many-electron problem in which the charge density $n(\mathbf{r})$ is treated as the fundamental object of interest as opposed to the wavefunction $\psi$. While $\psi$ is a function of the spatial and spin coordinates of every electron in the system (for a total of $4M$ variables where $M$ is the number of electrons), $n(\mathbf{r})$ depends only on the three general spatial coordinates $(x, y, z)$. Thus, the framework of DFT dramatically speeds up the process of solving the many-electron problem. While DFT provides a strong balance between accuracy and computational cost, the cost of executing a DFT calculation is still significant, of order $O(M^3)$. Furthermore, even within DFT, approximations are required so that calculations can be executed at reasonable cost. Machine learning is a promising strategy by which to improve both the speed and accuracy of DFT calculations. In fact, the capacity for ML to accelerate the approximate solution of the Schrödinger equation spans the entire process outlined above, and is not restricted to DFT alone [7]. Ongoing research even includes the development of ML methods that can bypass DFT altogether, such as machine-learned potentials trained on DFT data (see other chapters). Nonetheless, the contributions of ML to the advancement of DFT methods is a conceptually rich and a practically significant topic, and the following sections are devoted to a review of the recent progress.

In Section 2 we explain the basic foundations of DFT, and introduce the problem of approximating the exchange-correlation functional. In Section 3 we showcase some historical developments of approximate functionals found by humans in order to contextualize the more recent developments from ML. We summarize in Section 4 the technical and theoretical achievements enabled by ML. Section 5 is devoted to discussions on the generalization accuracy of the machines seen in the literature and potential directions for furthur research. Section 6 concludes the review with some perspectives on the future.

## 2 Background: Almost a century of DFT

Density functional theory begins with the foundational papers of Thomas and Fermi [8] [9]. Here we summarize modern DFT.

### 2.1 Foundations

The electronic $N$-body dynamics is governed by the following Hamiltonian

$$H = \sum_{i=1}^{N} \left[ -\frac{\nabla_i^2}{2} + V_{\text{ion}}(\mathbf{r}_i) \right] + \frac{1}{2} \sum_{i \neq j}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{1}$$

and the Schrödinger equation for the $N-$body wave function $\Psi$ is

$$H\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N). \tag{2}$$

where $\mathbf{r}_i$ are the spatial coordinates of the electrons, and $\Psi$ is antisymmetric with respect to particle exchange.

In this chapter, we restrict ourselves to the non-relativistic, non-magnetic case and adopt the atomic unit ($e = c = \hbar = m_{\text{e}} = 1$). We also suppress spin degrees of freedom in all equations for simplicity–although these are included in all modern DFT calculations.

This equation determines the series of many-body eigenstates once the one-body potential $V_{\text{ion}}(\mathbf{r})$ generated by the ionic configuration is given. In this sense, the eigenstates and their energy ordering are well-defined functionals of $V_{\text{ion}}(\mathbf{r})$, a scalar value on real space.

Hohenberg and Kohn formulated two important theorems [4]. First, the charge density distribution of the ground state $n(\mathbf{r})$ determines the corresponding $V_{\text{ion}}(\mathbf{r})$ unambiguously (if it exists) that has such an $n(\mathbf{r})$ in the ground state (in the non-degenerate case). Second, there exists a variational equation with $n(\mathbf{r})$ as the variable that yields the ground state energy: Formally, there is a functional of the whole distribution of $n(\mathbf{r})$ bounded by the ground state energy for given $V_{\text{ion}}(\mathbf{r})$.

$$E[n, V_{\text{ion}}] \equiv F[n] + \int d\mathbf{r} n(\mathbf{r}) V_{\text{ion}}(\mathbf{r}) \geq E_{\text{GS}}[V_{\text{ion}}], \tag{3}$$

for which the equality holds only when $n(\mathbf{r})$ is equal to that at the true ground state for $V_{\text{ion}}(\mathbf{r})$. The functional $F[n]$, as expressed later by Levy [10] and Lieb [11], is

$$F[n] \equiv \min_{\Psi \to n} \langle \Psi | \left( \hat{T}_{\text{e}} + \hat{U}_{\text{ee}} \right) | \Psi \rangle, \tag{4}$$

$$\hat{T}_{\text{e}} = \sum_{i=1}^{N} \left( -\frac{\nabla_i^2}{2} \right), \quad \hat{U}_{\text{ee}} = \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{5}$$

where the minimization is taken with constraint that the many-body state $\Psi$ yields the given charge density distribution through

$$n(\mathbf{r}) = \int d\mathbf{r}_2 d\mathbf{r}_3 \cdots d\mathbf{r}_N |\Psi(\mathbf{r}, \mathbf{r}_2, \cdots, \mathbf{r}_N)|^2. \tag{6}$$

We express the dependence as functional by squared parentheses "[]". The functional $F[n]$ is universal, which simply means it is a density functional with no dependence on $V_{\text{ion}}(\mathbf{r})$.

   The Euler equation for the ground state charge density is:

$$\frac{\delta F[n]}{\delta n(\mathbf{r})} + V_{\text{ion}}(\mathbf{r}) = 0, \tag{7}$$

if we fix the particle number. This equation clarifies the input-output relation of the ground-state quantum theory: Scalar function of space ($V_{\text{ion}}(\mathbf{r})$) in, scalar function of space ($n(\mathbf{r})$) out. Of course this equation does not simplify the problem since, to obtain the exact formula of $F[n]$, we are required to solve the Schrödinger equation. Thus, accurate and calculable models for $F[n]$ are essential for efficient calculations.

   People have sought two paths for solving the DFT variational equation (7). The first is the Kohn-Sham (KS) framework, and the second is the older orbital-free (OF) DFT. For both, the derivation of useful approximations to $F[n]$ has been the challenge. The latter in principle yields cheaper computational workflows but simultaneously imposes the additional subordinate (and difficult) challenge of how to calculate the kinetic energy of without the concept of orbitals.

## 2.2 The Kohn-Sham equations

Kohn and Sham [5] created a reference system that is formally solved more easily than the many-body Schrödinger equation, but yields the same ground-state charge density, via a transformation of the variational equation (7), as

$$\left[ -\frac{\nabla^2}{2} + V_{\text{eff}}[n](\mathbf{r}) \right] \varphi_a(\mathbf{r}) = \varepsilon_a \varphi_a(\mathbf{r}), \tag{8}$$

$$n(\mathbf{r}) = \sum_a \theta(\mu - \varepsilon_a) |\varphi_a(\mathbf{r})|^2 \equiv \sum_{a:\text{occ.}} |\varphi_a(\mathbf{r})|^2. \tag{9}$$

These are the Kohn-Sham (KS) equations. The effective potential $V_{\text{eff}}[n](\mathbf{r})$ is related to the universal functional $F[n]$ through

$$V_{\text{eff}}[n](\mathbf{r}) = V_{\text{ion}}(\mathbf{r}) + V_{\text{H}}[n](\mathbf{r}) + V_{\text{xc}}[n](\mathbf{r}), \tag{10}$$

where

$$V_H[n](\mathbf{r}) = \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad V_{xc}[n](\mathbf{r}) = \frac{\delta E_{xc}}{\delta n(\mathbf{r})}, \tag{11}$$

and the exchange-correlation (XC) energy functional is defined as:

$$E_{xc}[n] = F[n] - T_s[n] - E_H[n], \tag{12}$$

where

$$T_s[n] = \min_{\Phi \to n} \langle \Phi | \hat{T}_e | \Phi \rangle, \quad E_H[n] = \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \tag{13}$$

The kinetic energy functional $T_s[n]$ is defined by the constrained minimization of Eq. (4). The majority of kinetic energy is thus treated by an auxiliary set of single-particle orbitals $\{\varphi_a\}$ ($a$: spin-orbital) that form $\Phi$, while the remaining quantum effects are condensed into the exchange-correlation energy $E_{xc}[n]$. The ground state total energy $E_{GS}[V_{ion}]$ is derived from the stationary solution of the KS equations by

$$E_{GS}[V_{ion}] = \sum_{a:\text{occ.}} \varepsilon_a - E_H[n] + E_{xc}[n] - \int d\mathbf{r} n(\mathbf{r}) V_{xc}(\mathbf{r}). \tag{14}$$

This framework opens a path to practical first-principles calculations for the total energy of electronic systems such as molecules and solids, if any useful approximations to $E_{xc}$ are available. It should be considered as a formalization of Slater's $X_\alpha$ method [12]. The development of useful models for $E_{xc}[n]$ has thus become a grand challenge in theoretical chemistry, condensed matter physics, and material science.

The reference one-particle system can be taken somewhat arbitrarily, as we can formulate an equation that is different from the original KS but in principle gives an identical ground state charge density [13]. For some arbitrary operator $\hat{O}$ on the KS orbitals, we can construct the following system:

$$\left[\hat{O}[\{\varphi\}] + V_{eff}^{gKS}[n](\mathbf{r})\right] \varphi_a(\mathbf{r}) = \varepsilon_a \varphi_a(\mathbf{r}), \tag{15}$$

$$\frac{\delta S[\Phi]}{\delta \varphi_a(\mathbf{r})} \equiv \hat{O}[\{\varphi\}] \varphi_a(\mathbf{r}) \tag{16}$$

$$V_{eff}^{gKS}[n](\mathbf{r}) \equiv V_{ion}(\mathbf{r}) + \frac{\delta (F[n] - F^s[n])}{\delta n(\mathbf{r})}, \quad F^s[n] \equiv \min_{\Phi \to n} S[\Phi]. \tag{17}$$

Here, $S[\Phi]$ is an arbitrary functional of the Slater state $\Phi$ with subtle mathematical requirements to $F^s[n]$. This *generalized* KS framework justifies including supplemental one-body operators (such as the Fock exchange) in the KS equations . This generalization can make ground-state calculations much more convenient while yielding practically identical ground-state energies. However, the spectrum of eigenvalues can be much improved in a Generalized KS calculation [14].

## 2.3 Orbital-free DFT

Orbital-free DFT (OF-DFT) aims to solve the variational equation (7) directly in terms of the density $n(\mathbf{r})$. As in KS-DFT, a separation of the universal functional $F[n]$ is used; $F[n] = T_s[n] + E_H[n] + E_{xc}[n]$. The OF-DFT approach poses additional challenges: (i) First, $T_s[n]$ has to be modeled without solving for single-particle orbitals. (ii) Second, treatment of $V_{ion}(\mathbf{r})$ also raises an issue. In the KS approach one can omit the core electrons from the equations by using a pseudopotential to describe the effects of the core states on the valence states by the valence-orbital dependent projector (this becomes problematic in OF-DFT). We do not discuss pseudopotentials in depth here, but refer the reader to Ref. [15] for a thorough review.

## 3 Human development of the functional

Various conditions on the exact functional are well known [16]. The early development of approximate functionals was often grounded in interpolations between these well-studied exact conditions. Approximations constructed this way automatically satisfy exact conditions–but are also calculable in the intermediate regimes. Machine learning brings a sophisticated tool to the interpolation step. We review a few approximate functionals so that the historical context of the more recent ML development may be clarified.

Conventionally the XC energy models have been developed based on the separation into the exchange and correlation parts $E_{xc}[n] = E_x[n] + E_c[n]$. The exchange part is defined by the Coulomb exchange integral with an appropriate single-particle basis. Defined with the KS orbitals, it is functional of $n(\mathbf{r})$, as each orbital is an implicit density functional (here spin is required, as there is no exchange contribution from opposite spins):

$$E_x[n] = -\frac{1}{2} \sum_{a,b:\text{occ.}} \int d\mathbf{r} d\mathbf{r}' \frac{\varphi_a^*(\mathbf{r})\varphi_b^*(\mathbf{r}')\varphi_b(\mathbf{r})\varphi_a(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \tag{18}$$

The correlation part $E_c[n]$ contains all other quantum effects. $E_x[n]$ and $E_c[n]$ can each be written as an integral over an XC energy per electron, e.g:

$$E_c[n] = \int d\mathbf{r} n(\mathbf{r}) \varepsilon_c[n](\mathbf{r}). \tag{19}$$

Note that choices of energy per electron are ambiguous since they can have arbitrary components that integrate to zero in total [17]. The definitions above can also be reasonably extended for the two-component spin DFT, where the functionals depend on $n_\uparrow(\mathbf{r})$ and $n_\downarrow(\mathbf{r})$ representing the spin-up and down components of the

electron charge density. The relative spin polarization introduced $\zeta(\mathbf{r}) = (n_\uparrow(\mathbf{r}) - n_\downarrow(\mathbf{r}))/(n_\uparrow(\mathbf{r}) + n_\downarrow(\mathbf{r}))$ is often used to parameterize spin dependence.

## 3.1 Jacob's Ladder of Density Functional Approximations

Approximations to the XC functional can be organized into a hierarchy which Schmidt and Perdew [18] coined as Jacob's ladder. The ladder spans the distance between the (crude) Hartree approximation and the sought-after Heaven of Chemical Accuracy (the ~ 1 kcal/mol accuracy required to predict the energy of a chemical reaction). The lowest "rung" of the ladder is the simplest approximation to the XC functional: the local density approximation (LDA). In the LDA, the XC functional is constructed as a *local* functional of the density–that is, it depends on the density at that point alone. Improvements on the LDA (the higher "rungs") are guided by the concept of a gradient expansion [4, 5]. Going up the rungs, the approximate models of the exchange-correlation energy densities incorporate dependence on the charge density at the site $n(\mathbf{r})$ [local (spin) density approximation, L(S)DA], its gradient $\nabla n(\mathbf{r})$ (generalized gradient approximation, GGA), Laplacian and/or kinetic energy density (meta GGA), occupied KS orbitals, (hybrid), and unoccupied ones (double hybrid, RPA). Below we review some example functionals from the various rungs. An excellent summary of approximate XC functionals can be found in [19].

## 3.2 Local density approximation

In the LDA, the energies per electron are given by

$$\varepsilon_{\text{x,c}}^{\text{LDA}}[n](\mathbf{r}) = \varepsilon_{\text{x,c}}^{\text{unif}}(n(\mathbf{r})). \tag{20}$$

where $\varepsilon_{\text{x,c}}^{\text{unif}}(n(\mathbf{r}))$ is the energy per electron of the uniform gas with density $n$. To develop the model for $\varepsilon_{\text{x,c}}(n)$, people refer to the system's behavior in the uniform electron gas limit $V_{\text{ion}} = 0$. Often the charge density is characterized by the Wigner-Seitz radius $r_{\text{s}} = \sqrt[3]{3/(4\pi n)} \equiv r_{\text{s}}(n)$. The dependence on the spin density is also included, called the local spin density approximation (LSDA):

$$\varepsilon_{\text{x,c}}^{\text{LSDA}}[n_\uparrow, n_\downarrow](\mathbf{r}) = \varepsilon_{\text{x,c}}^{\text{unif}}(n(\mathbf{r}), \zeta(\mathbf{r})). \tag{21}$$

The exchange energy per electron in the uniform electron gas is analytically calculated to be

$$\varepsilon_{\text{x}}^{\text{unif}}(n) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{1/3}. \tag{22}$$

First Bloch [20], then Dirac [21] derived this form for Eq. (20).

### 3.2.1 Correlation of the Uniform Gas

Several parameterizations of $\varepsilon_c^{\mathrm{unif}}(n)$ exist, the most commonly used being VWN [22], Perdew-Zunger [23], and PW92 [24]. All three were constructed with the same basic strategy. First, an analytical form is chosen such that it recovers the correct asymptotic behavior in either the high-density ($r_s \to 0$) or low-density ($r_s \to \infty$) limit, or both. This is followed by numerically fitting a small number of free parameters to the Monte Carlo results of Ceperley and Alder[25, 26]; the interpolation serves to extend the range of $r_s$ for which the model is accurate. All three of these parameterizations agree with one another to within a roughly 2-percent error (which is substantially smaller than the error of LDA for inhomogenous systems). The parameter determination for the spin-polarized case used Misawa's spin scaling for the correlation energy [27].

## 3.3 Generalized gradient approximation (GGA)

The next rung of Jacob's ladder corresponds to inclusion of the density gradient

$$\varepsilon_{\mathrm{xc}}^{\mathrm{GGA}}[n](\mathbf{r}) = \varepsilon_{\mathrm{xc}}^{\mathrm{GGA}}(n(\mathbf{r}), |\nabla n(\mathbf{r})|), \tag{23}$$

called generalized gradient approximation (GGA). The development of useful GGA model functionals has been much advanced through '80s and '90s. Although there have been many proposed models, we here explain the Perdew-Burke-Ernzerhof (PBE) model [28], that well represents how the analytical human functional development proceeds.

The PBE correlation energy per electron is

$$\varepsilon_c^{\mathrm{PBE}}[n_\uparrow, n_\downarrow](\mathbf{r}) = \varepsilon_c^{\mathrm{unif}}(r_s(\mathbf{r}), \zeta(\mathbf{r})) + H(r_s(\mathbf{r}), \zeta(\mathbf{r}), t(\mathbf{r})), \tag{24}$$

where $t$ is a dimensionless measure of the gradient. The form of $H$ is chosen to satisfy three conditions: (a) In the small-$t$ limit it reduces to the second order gradient expansion [29] $H \sim \beta t^2$, from which $\beta$ is known to be $\simeq 0.066725$ . (b) In the rapidly varying limit $t \to \infty$, the correlation energy should vanish: $H \to -\varepsilon_c^{\mathrm{UEG}}$. (c) The total correlation energy must scale to a constant under uniform scaling to the high-density limit (Ref.[30]).

The exchange energy per electron in the unpolarized case $\zeta = 0$ is formulated as

$$\varepsilon_{\mathrm{x}}^{\mathrm{PBE}}[n](\mathbf{r}) \equiv \varepsilon_{\mathrm{x}}^{\mathrm{unif}}(n(\mathbf{r})) F_{\mathrm{x}}(s(\mathbf{r})), \tag{25}$$

with $s = |\nabla n|/(2k_{\mathrm{F}} n)$ and $F_{\mathrm{x}}$ is called the *enhancement factor*. The spin dependent formula is derived by the exact spin-scaling relationship for exchange [31]. The PBE form satisfies the following three conditions: (d) Under the uniform density scaling described above, $E_{\mathrm{x}}$ must scale linearly (Ref. [32]). (e) It cancels the gradient correction to the correlation energy as $s \to 0$ so that the model recovers the accurate

linear response of LDA. (f) It satisfies the Lieb-Oxford inequality [33] for all $n$ and $s$, thus ensuring satisfaction for all possible densities.

In the above construction, the formulas have four parameters determined from the conditions. The artificial aspect is in selecting the forms of $H$ [Eq. (24)] and $F$ [Eq. (25)], which are mathematically simple for humans and minimally parametrized. Many variations on the PBE approximation have since been suggested, including RPBE [34], revPBE [35], and PBEsol [36].

## 3.4 Physically fitted functionals

Becke initiated a complementary approach for functional development. He introduced a model exchange energy within the GGA with a single parameter and determined it so that it optimally reproduces the Hartree-Fock energies in 6 rare gas atoms up to Rn [37], called B88. This parametrization satisfied exact conditions on spin-scaling and recovery of the uniform limit. Moreover, the form of the exchange energy per electron yields the correct asymptotic behavior in the evanescent region of an atom or molecule–a crucial property of chemical bonding [38], [39].

### 3.4.1 Hybrid Functionals

Based on intuitive ideas of the adiabatic connection curves of approximate functionals (see also Ref.[40] for recent advance), Becke further proposed [41] to modify his B88 model by replacing a fraction of LSDA exchange with exact exchange, and adding correlation at the GGA level:

$$E_{xc}^{B3PW91} = E_{xc}^{LSDA} + a_0(E_x - E_x^{LSDA}) + a_x(E_x^{B88} - E_x^{LSDA}) + a_c(E_c^{PW91} - E_c^{LSDA}),$$
(26)

where $E_c^{PW91}$ is from Ref. [42]. He chose the three parameters $(a_0, a_x, a_c)$ so that experimentally observed reaction energies for approximately 100 reactions are well reproduced. The variant of this model with $E_c^{PW91}$ replaced by the Lee-Yang-Parr form [43] is the B3LYP functional [44, 45], the most used hybrid functional (4th rung of the ladder) in quantum chemistry. Although Becke himself emphasized the physics [41], the scheme stimulated researchers, yielding various data-tuned functionals in the next decades (see Ref. [46] for a review). Many of these developments, not originally declaring themselves as ML, used the same fundamental strategies as modern ML-DFT; but often were contrary in spirit to Becke's philosophy.

# 4 Machine-learning

Machine learning technology has introduced a paradigm that is in some ways similar to the variational theory for quantum many-body Hamiltonian. When we calculate the many-body ground-state wavefunction of an interacting quantum system by the variational method, we prepare a model wavefunction with numerous parameters and optimize those so that the energy expectation value is minimized. The derived wavefunction is used for calculating the observable properties such as correlation functions, but of course this specific wavefunction is only used for this one system.

This scheme is not regarded as useful for finding approximate functionals in DFT since the latter aims to utilize the derived functional for systems not referenced in the optimization step. Even if we could achieve an extreme accuracy for a specific system, such an optimized model would never be applicable to others. To this field, ML has provided a framework to optimize a model function for *a group* of systems, keeping its accuracy even when applied out of the group (via some form of regularization). The surge of ML developments in DFT has demonstrated that highly parameterized functionals may have general accuracy, at least for an appropriately defined range of applications.

In this section, we attempt to summarize numerous recent results applying ML to design density functionals, and use these to draw general insights into the possibilities that ML has opened for improved DFT calculations.

## 4.1 General discussion

ML in DFT generally concerns modeling of the functionals appearing in the previous sections, such as exchange-correlation energy $E_{\text{xc}}[n]$. The tricky thing with DFT is that the density functionals and $n(\mathbf{r})$ are, in practice, both outputs of the Hamiltonian that are simultaneously determined after solving the KS equations. Because of this there has been no convenient analytic theory that relates $n(\mathbf{r})$ and functionals as input and output, unlike the wave-function based perturbation theory that relates the single-particle basis to physical quantities. Machine learning provides versatile tools to optimize models relating those using a large number of pairs $(n(\mathbf{r}), \text{functional})$ produced by accurate calculations in specific cases.

The typical procedure starts from modeling of a target functional in DFT, dubbed as $f[n]$, by any regressor that can accept any value of $n(\mathbf{r})$. Practically this is implemented as a vector-to-scalar[1] mapping $n(\mathbf{r}) \rightarrow f[n]$, where $n(\mathbf{r})$, a non-negative distribution in space, is approximately represented as a vector; $n(\mathbf{r}) \approx \mathbf{n}$. How to efficiently represent $\mathbf{n}$ is also of a matter of ongoing development. The regressor model chosen has many parameters. We then collect specific examples of values $(n^{(i)}, f^{(i)}[n])$ for various different potentials to serve as training data. These reference data must be accurate enough for the intended applications of the ML

---

[1] Hereafter we assume $f$ to be a scalar for simplicity but we can also take $f$ as vector or tensor.

model. The optimization of the model parameters is executed so that a loss function

$$L = \frac{1}{N_{\text{data}}} \sum_{i=1}^{N_{\text{data}}} D(f^{(i)}, f_{\text{model}}[n^{(i)}]) + (\text{reg.}).  \tag{27}$$

is minimized. Here $D(f, f')$ represents any non-negative function that measures distance between inputs $f$ and $f'$; e.g., $D(f, f') = (f - f')^2$ and "reg." denotes the regularization term that prevents overfitting to the training data. While the building blocks for this procedure predate ML, ML technology has provided both powerful model functions and efficient methods for optimization and regularization.

### 4.1.1 Universal regressors

In the DFT context, two approaches are often used for the above $f_{\text{model}}$: Kernel ridge regression (KRR) and neural network (NN). We explain the features of those models very briefly.

Kernel ridge regression (KRR) [47] has a form

$$f(\mathbf{v}_*) = \sum_{j=1}^{N_{\text{data}}} \alpha_j k(\mathbf{v}_j, \mathbf{v}),  \tag{28}$$

where $\mathbf{v}$ denotes any descriptor vector for $f$. The kernel function $k(\mathbf{x}, \mathbf{x}')$ can have any form as long as the $N_{\text{data}} \times N_{\text{data}}$ matrix formed by any set of data $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_{N_{\text{data}}}\}$, $\mathcal{K}_{ij} = k(\mathbf{v}_i, \mathbf{v}_j)$, is positive definite. Assuming ridge regularization, the weight $\alpha_i$ is exactly determined by a matrix inversion of size $N_{\text{data}} \times N_{\text{data}}$.

Neural networks (NN) come in many varieties. The typical fully connected NN is formed by repetition of non-linear vector-to-vector transformation $\mathbf{v} \rightarrow \mathbf{v}'$

$$\mathbf{v}' = \phi(W\mathbf{v} + \mathbf{b}),  \tag{29}$$

which is often represented by the celebrated graph (probably seen in other Chapters) composed of layers of nodes and edges connecting those. Here, $W$ and $\mathbf{b}$ are the parameter matrix and vector to be optimized and $\phi$ denotes some non-linear transformation operated on the respective vector components. The intermediate vector dimensions are arbitrary.

Both KRR and NN are both assured to be universal approximators [48–51], in the sense that the model can numerically reproduce *any* reference function–provided that the model has a sufficient number of parameters. This is achieved by increasing the reference data and intermediate dimensions for the KRR and NN, respectively.

The universal character of these models is quite beneficial for developing approximate density functionals. In principle, the functionals $f[n]$ have rigorous (but not directly calculable) forms that are only inferred from specific cases. NN and KRR-based models are capable of representing any dependences between $n(\mathbf{r})$ and

$f$, by fitting the specific data. They can also attain some robustness against overfitting when the optimized models are applied to unseen cases.

Although those two approaches are mathematically equivalent in some limits [52], a crucial difference between the KRR and NN exists in practice. KRR needs to store the training data for later calculations, but once the definition of the loss and training data are specified, the weights $\{\alpha_i\}$ can be rigorously determined. The NN, on the other hand, treats the training data explicitly only at the model optimization. The information of the training systems are then encoded implicitly in the NN model parameters. However, the optimum parameters depend on the optimization methods and, if one adopts stochastic methods such as dropout and batch optimization, their specific values are not reproducible even with fixed NN architecture and training systems. This is quite disturbing from a traditional quantum-chemical viewpoint.

### 4.1.2 Limitation by data

The accuracy of ML-based functionals is limited by the training data, but accurate data available now for training is still little in amount. The most accurate data are those from experimental observations, which are limited in number of the systems as well as in the variety of quantities. Few quantities are measurable with enough accuracy, like the energy-related quantities and structural properties such as bond lengths and angles and lattice parameters. Hence, in most of the ML-based DFT developments, researchers refer to databases augmented by theoretical calculations. In datasets based on existing theoretical methods, the procedure to generate the training data pair $(n(\mathbf{r}), f)$ is free from any uncertainty like noise in experimental data–allowing for more consistent benchmarking and comparison across ML models. Currently, the most accurate and executable solver for molecules is the coupled cluster (CC) method [53, 54]. Fortunately, in many cases coupled-cluster singles and doubles and perturbative triples (CCSD(T) [55]) achieves chemical accuracy (errors less than 1 kcal/mol or about 30 meV for covalent bonds in small molecules) and therefore versatile ML-DFT functionals with CC accuracy are worth aiming for.

### 4.1.3 ML Usage with DFT

People have explored various fitting tasks concerning the density functionals using ML. Theoretically there is a nuanced difference among the papers in the level of the ML usage: They either (i) preserve the rigorous framework of DFT, (ii) supplement or completely replace the variable $n(\mathbf{r})$ with atomic potentials $V_{\text{ion}}(\mathbf{r})$, or (iii) with element labels. Level (ii) is distinct from (iii) as it is based on the potential functional theory, which takes $V_{\text{ion}}(\mathbf{r})$ as the variable for the corresponding ground state. This is a formal dual of DFT [11, 56] and therefore is capable of carrying the same information. Using the element descriptors in addition to $n(\mathbf{r})$ seems useful for efficient fitting in some cases, but at the price of departure from the density-potential functional framework.
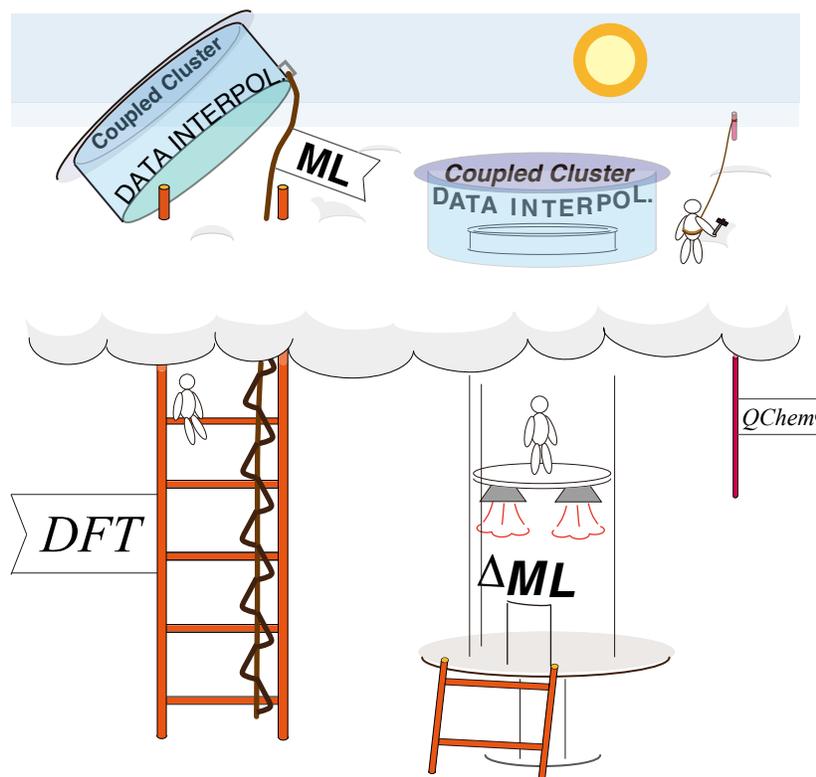
Fig. 1: A cartoon illustration of the relationship between machine learning and DFT. On the left hand side, we show the (Jacobs) ladder between Hartree-accuracy and chemical accuracy "heaven" (the blue sky). Traditional DFT often involves using the Jacob's ladder hierarchy to add successively more complex "ingredients" (density-gradients, etc.) to a functional. Machine learning can supplement this methodology by learning new, unorthodox ingredients–thus the rope of ML provides an alternative pathway towards coupled-cluster, or even chemical accuracy. In the middle, we show a pure ML approach. The "rocket booster" of ML allows us to circumvent the Jacob's ladder entirely; but the accuracy of an ML model is bounded by the training data. Practically, this "glass ceiling" is coupled-cluster theory. On the extreme right, a scientist has scaled the Quantum Chemistry rope to build improved methods and better scaling codes–and thus prepares to break the CC ceiling by providing higher-quality training data.

Across the studies we address in this review, ML methodologies fit the Jacob's Ladder paradigm to varying degrees. Some try better parametrizations of the functionals used in the KS or OF framework, adapting the definitions of variables in the

ladder rungs. Others seek modified implementations of the ladder, by laying "rungs" of different variables other than original, like the non-local weighted density [57]. Still others fit the difference of the functionals calculated with the high-level DF or quantum chemical approximations and low-level DFT by extra variables that may be of level (ii) or (iii). Such studies would be understood as building an ML path to Heaven on the basis of (low-level) DFT, among which we find the current most successful achievements in the cost-accuracy balance as we will highlight below. For all those, we again remind that the readers should be aware of the limitation by data. We sketch the situation by Fig. 1.

### 4.1.4  Interpolation, extrapolation, and generalization

In principle, a trained model should not be applied to systems that are not represented in the training data. For a trained machine the target systems are classified into two groups. Systems for which the model is expected to provide accurate results are called "in-domain" or "in-sample". Systems that are poorly represented by the training data, and therefore are not expected to be accurately captured by the model, are termed "out-of-domain" or "out-of-sample". In this review, we use "interpolation" to refer to the application of a model to an in-domain system, and "extrapolation" to refer to the application of a model to an out-of-domain system.

Note that whether a task constitutes interpolation or extrapolation is judged only after application; the system for which the machine yields accurate predictions is *found to be* in the "domain of interpolation". The interesting issue for ML applications to DFT would be the possibility that any training scheme can result in an interpolation domain which is broader than one would expect from the training data. Because within the DFT framework the variable is $n(\mathbf{r})$, the original atoms are anonymized, and therefore the interpolation domain can become broader than expected from the literal atomistic composition of the systems. We call the success of the models to unseen systems "generalization", rather than extrapolation.

## 4.2  Fitting of functionals

Here we show some of the successful cases for making well-behaved density functionals $f[n] \approx f(\mathbf{n})$ with machine learning. The functional $f[n]$ could be the total energy, the exchange-correlation energy, the exchange-correlation potential, forces, or any other quantities related to the ground state determined by the specified $n(\mathbf{r})$.

### 4.2.1  Kinetic energy

In 2012, Snyder et al. [58, 59] calculated the exact ground-state charge density and non-interacting kinetic energy in a one-dimensional model, where the model

ionic potential was replaced by a sum of Gaussians controlled by a few parameters. They demonstrated that the kinetic energy $T_s[n]$ is well fitted by the KRR model as a functional of $n(\mathbf{r})$ being a vector $\{n(x_1), n(x_2), \cdots\}$ on a discretized spatial grid. This work is the first demonstration that a fully nonlocal functional can be implemented for accurate reuse, as long as the range of application is well controlled: In this study the training and test potentials were both within the specific range of the Gaussian model parameter space. This has been also shown to hold for the exchange-correlation energy in one dimension [60].

The faithful fitting of fully nonlocal functionals in three dimensions is far more difficult since the representation of $[n]$ requires a large amount of data. Yao and Parkhill [61] pursued $T_s[n]$ by a convolutional NN form, whereas Brockherde et al. [62] demonstrated a bypass of the functional relation $V_{ion} \rightarrow n$ in the KS framework by KRR for molecules in 3D, avoiding even the explicit construction of $T_s[n]$.

Semilocal approaches to $T_s[n]$ with ML have been widely explored. Seino et al. [63] targeted the kinetic energy density $\tau(\mathbf{r})$ defined with a local decomposition $T_s[n] = \int d\mathbf{r}\tau([n]; \mathbf{r})$. They collected the grid data for the converged $n(\mathbf{r})$ and $\tau(\mathbf{r})$ calculated from KS-DFT for (i) atoms from H to Ne and (ii) 19 small molecules composed of C, H, O, and N, and separated those into the training and test grid points in dataset (i) and (ii), respectively. They implemented the semilocal form $\tau[n](\mathbf{r}) = \tau(\mathbf{n}(\mathbf{r}), |\nabla \mathbf{n}(\mathbf{r})|, \nabla^2\mathbf{n}(\mathbf{r}), |\nabla\nabla^2\mathbf{n}(\mathbf{r})|)$ using the NN and examined the fitting accuracy. Golub and Manzhos dug further into the fitting problem with focus on the instability near the atom [64] and ML model dependency [65]. Semilocal modeling tends to suffer from error in molecular reaction energies of order 0.1 Hartree compared with the KS-DFT, which still remains a challenge even with ML. Still, it may be useful for estimating equilibrium structural properties as shown by Imoto et al. [66]. Inclusion of the atomic kind and positions as descriptors can be useful for improvement [67, 68], though is not a conclusive solution. All such attempts must account for the ambiguity in the energy density discussed earlier.

Nonlocal density descriptors are generally expressed in the following convolution

$$\varrho(\mathbf{r}) = \int d\mathbf{r}' w(\mathbf{r} - \mathbf{r}')n(\mathbf{r}'), \tag{30}$$

and seem useful for capturing the nonlocal nature of $T_s[n]$. Note that the convolution weight $w$ can also be a target of ML modeling. Ryczko et al. [69] demonstrated a fitting of a voxel-to-voxel NN model for $n(\mathbf{r})$ to $\delta T_s/\delta n(\mathbf{r})$ for graphene sheet. The decomposition of $n(\mathbf{r})$ by atom-centered basis functions has also been used as an efficient method for incorporating the nonlocality [70, 71]. The nonlocal descriptors [72] are generally found to reduce the fitting error by an order of magnitude compared to semilocal approximations.

Using machine-learned energies for the variational solution Eq.(7) imposes another challenge since an ML model fitted to energies suffers from numerical noise when we take derivatives [58]. Training sets can never be dense enough to infer the derivatives in all directions in the high dimensional descriptor space. The numerical instability induced by this may be mitigated by stopping the derivatives going out of the sampling space [58, 73, 74]. The finite difference method is also a stable

choice [65], but with increased computational cost relative to the direct derivatives. Defining the loss function to include the functional derivatives was also shown to be effective [66, 75]. For NN models, the noisy behavior may be mitigated by carefully designing the NN architecture [76].

The kinetic "potential" used for the variational approach, $\delta T_s[n]/\delta n(\mathbf{r})$, has significant nonlocality. To fully take advantage of the OF framework than KS, the nonlocal calculation must be kept as cheap as possible. Ghasemi and Kühne attempted to learn the "source" of the potential rather than the potential itself [77], which could alleviate the nonlocality.

A remarkable mention goes to a recent result by Zhang et al. [71], in which they claim that the total energy difference of ML OF-DFT from KS-DFT was well under the chemical accuracy for broad unlearned molecules. They used the atom-centered density descriptor and developed a complicated NN architecture using the transformer [78], which is rapidly gaining popularity in various ML usages for its performance. This achievement may be the cornerstone on which the OF-DFT can next aim at true chemical accuracy.

### 4.2.2 Exchange-correlation energy

Fitting of the exchange-correlation energy by ML dates back to 1996 by Tozer, Ingamells and Handy [79]. For small molecules composed of the first and second row elements, they calculated accurate charge densities using coupled clusters theory, solved the inversion problem of the KS equation [80] to obtain the KS potential, fitted a NN model function, and applied the optimized functional to unseen molecules (still restricted to the second-row) to see its performance. They found that the optimized LDA functional deviates from an analytical LDA model to better fit the data. Thanks to the correct asymptotic behavior of the training KS potential, their NN LDA functional yields significant improvement for the ionization energy for the restricted group of test systems. Surprisingly, this study has already established the template design of the modern ML-DFT developments, as described later.

Other researchers exploited Becke's three parameter hybrid form (Sec. 3.4) by making the mixing parameters $(a_0, a_x, a_c)$ [Eq. (26)] system-dependent. Zheng et al. [81] implemented in 2004 a small NN which relates $n(\mathbf{r})$ calculated with the B3LYP to $(a_0, a_x, a_c)$ for correcting $E_{xc}$ so that the reaction energies of the reference molecules are well reproduced, showing fair improvement in the test accuracy relative to B3LYP. Similar methods, where a small number of parameters entering the hybrid exchange-correlation functionals are derived from lower-level theories, have been further developed more recently [82, 83]. This is not strictly DFT since the parameters are explicitly dependent on atomic numbers. These are modest usages of ML, whereas we showcase more drastic use below.

Exact fitting of the fully nonlocal $E_{xc}[n]$ is interesting as it may help to infer the analytic properties of the exact exchange and correlation. For one-dimensional hydrogen chains, Li and coauthors fitted the exact exchange-correlation energy $E_{xc}[n]$ calculated with the density matrix renormalization group [60] with KRR, showing

that the test error rapidly converges to zero by increasing the training data. Nagai et al. [84] performed a fitting of the fully nonlocal exchange-correlation potential $V_{xc}(\mathbf{r})$ in an extremely simple one-dimensional model derived from the exact diagonalization and KS inversion, as a vector-to-vector mapping $\mathbf{n} \rightarrow \mathbf{V}_{xc} = \{V_{xc}(x_1), V_{xc}(x_2), \cdots\}$ by the NN. Such exact fitting becomes demanding when going to three dimensions due to the difficulty of representing the density distribution in an efficient fashion. Zhou et al. [85] trained a nonlocal convolutional NN model that maps $V_{xc}(\mathbf{r})$ from $n$ in the region nearby $\mathbf{r}$ for H-He diatomic molecules in three dimensions. On using the KS-inverted $V_{xc}(\mathbf{r})$ for training $\varepsilon_{xc}[n](\mathbf{r})$, we also feature Schmidt et al. [86] for 1D and Gavini et al. [87] for 3D systems.

ML has also been helpful for development starting from the semilocal approximation, toward the chemical accuracy world but by going up paths forked from Jacob's Ladder (Fig. 1). The semilocal approximation is represented by the vector-to-scalar relation $\varepsilon_{x,c}[n](\mathbf{r}) \approx \varepsilon_{x,c}(\mathbf{g}(\mathbf{r}))$ with e.g. $\mathbf{g}(\mathbf{r}) = (n(\mathbf{r}), |\nabla n(\mathbf{r})|)$ for the GGA. Extension of the descriptor vector $\mathbf{g}$ has been widely explored, exploiting the fact that the procedure for optimizing ML models is executable for any descriptor.

A variety of nonlocal descriptors in the form Eq. (30) have been pursued, which are generalizations of the weighted density approximation for the exchange-correlation hole [57]. Schmidt et al. [86] applied the convolutional NN form to the one dimensional model. Lei and Medford [88] introduced the decomposition of the three dimensional charge density by the Maxwell-Cartesian spherical harmonics and showed that the B3LYP exchange-correlation functional may be accurately fitted by a NN. Nagai et al. [89] implemented a NN model including a weighted density descriptor and applied to the KS equations, achieving accuracy comparable to a hybrid functional, without resorting to the generalized KS framework including the exact exchange operator (4th rung). Bystrom and Kozinsky [90, 91] constructed a convolution-type density descriptor that is invariant under the uniform scaling transformation $n(\mathbf{r}) \rightarrow \lambda^3 n(\lambda \mathbf{r})$, that is necessary for modeling exchange functionals that guarantee exact scale invariance (see later).

The atomic decomposition of the charge density $n(\mathbf{r}) \approx \sum_{I:\text{atom}} n_I(\mathbf{r})$ has also been pursued for efficient descriptor design for the 3D charge density. The theory by Bartók et al. [92] on how to represent the atomic configuration surrounding a certain atom and its extension to high-dimensional quantities [93] have been utilized by several researchers [94–96]. They generally decompose the calculated $n(\mathbf{r})$ in an atom-centered basis and formulate rotationally covariant descriptors from the coefficients. A scheme of this kind allows for a representation of the entire distribution which is very sparse compared to those obtained from system-independent basis like plane-waves. The disadvantage is that one has to store the labels of atom types and positions: the functionals thus developed are in a strict sense not density functionals and may lose DFT's capability of generalizing across different types of atoms; yet they seem useful if the range of application is appropriately limited. [96, 97]. Generalization to different atom types can be recovered by forcing the model $E_{xc}$ to depend on $n(\mathbf{r})$ but not explicitly on atom type [96, 98], as shown in the case of OF-DFT [71].

Not only the charge density, but *any* quantities that are calculated using the converged KS Hamiltonian can also be taken as descriptors since they are all density functionals. The DeepMind21 (DM21) functional [99] incorporates the local HF exchange energy density and LDA exchange energy density into the descriptors, the former of which is a functional of the KS orbitals [100]. Riemelmoser et al. [101] formulated an extremely nonlocal density descriptor calculable with the fast Fourier transform. Polak et al. [102] adopt artificial densities calculated from unoccupied KS states as descriptors for the correlation energy density. A more data-based approach to the descriptor design would be to use the dimensionality reduction techniques as preprocessers of the functional input. Gong and coworkers [103] trained an NN encoder of density distribution in three dimensions so that it can relate the distributions matched by the translation and scale transformation and utilized the encoded features as descriptors of the exchange energy.

### 4.2.3 Baselining and Δ-learning

One key to succesful performances of ML-DFT is to leverage the wealth of intuition around human-made functional forms in concert with ML. One methodology for doing so is Δ-learning; in which the "baseline" of the model is an existing analytical functional form. Then, the *difference* between the analytical form and the reference data (i.e the Δ) is learned. Very often, these methods are presented in modest fashion (sometimes in the supplementary materials rather than the main text) despite their pivotal importance. We do not delve into the details, but discuss some general strategies.

In most of the ML applications based on the semilocal-to-nonlocal strategy, the model definition relies on the traditional decomposition or factorization of the exchange and correlation energies such as Eqs. (25) and (24). These definitions are useful for applying constraints on the machines as detailed later.

Ramakrishnan et al. [105] put forward a Δ-ML concept, represented by the following formula

$$P_{\mathrm{t}}(\mathbf{d}_{\mathrm{t}}) \approx P_{\mathrm{b}}(\mathbf{d}_{\mathrm{b}}) + \Delta P(\mathbf{d}_{\mathrm{b}}). \tag{31}$$

Here, one wants to learn a target ("t") property $P_{\mathrm{t}}$ as a function of descriptors of target value $\mathbf{d}_{\mathrm{t}}$, assuming that the descriptor value is also determined self-consistently referring to $P_{\mathrm{t}}$ (for example $P$ and $\mathbf{d}$ may be the total energy and optimum atomic positions of molecules, respectively). In the DFT context $\mathbf{d}$ can be $n(\mathbf{r})$ [104, 106]. For the learning task, one chooses a baseline theory "b" that is calculated with lower computational costs and lets the machine learn the residual $\Delta P$ as a function of approximate value of descriptor $\mathbf{d}_{\mathrm{b}}$ derived from the baseline theory. This approach is straightforwardly applied to Jacob's ladder, to learn the higher level energies as a functional of $n(\mathbf{r})$ derived from lower-level theories. Thanks to this, the calculation of the total energy with CCSD accuracy has been achieved with DFT cost [104].
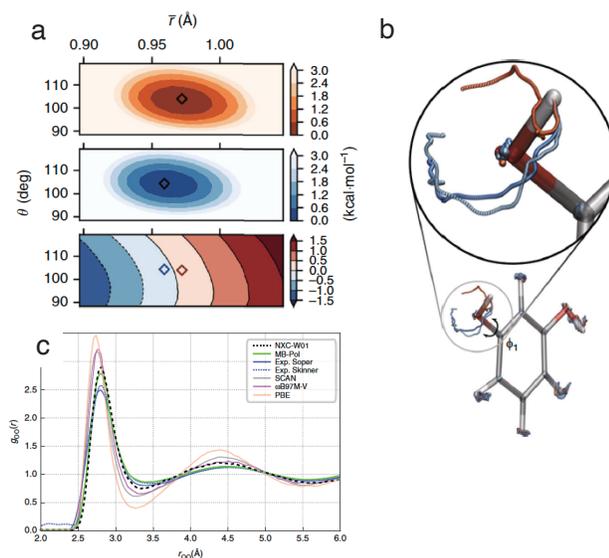
Fig. 2: Dynamics on accurate potential surfaces by $\Delta$-DFT. **a** The potential for water molecule calculated by the baseline theory (GGA, top) is corrected to the surface of CC accuracy (middle). The correction ML term (bottom) shows smooth and modest behavior, which corrects the minimum from red to blue diamond. Taken from Ref. [104] under the terms of the CC BY 4.0 license. **b** The dynamics of resorcinol on the accurate surface generates rotational trajectories of a branch (light and dark blue), different from the GGA force field (red). Taken from Ref. [104] under the terms of the CC BY 4.0 license. **c** The radial distribution function for O-O in water from molecular dynamics, calculated by a ML nonlocal $E_{xc}$ trained by the water trimer ("NXC-W01" in legend). Taken from Ref. [97] under the terms of the CC BY 4.0 license.

The remaining error due to the incorrect $\mathbf{d}_b$ is related to the density-driven error concept [107], which we revisit later in this text.

### 4.2.4 Derivatives of the functionals

Similarly to the case of $T_s[n]$, using the trained $F_{model}[n]$ for solving the KS equation requires some care. Although in $F_{model}[n]$ we are able to take arbitrary descriptors, to use it for calculating $V_{Hxc}(\mathbf{r}) = \delta F_{model}[n]/\delta n(\mathbf{r})$ the descriptors have to be formally differentiable, though a large class of descriptors written in the convolutional forms can be differentiated [108]. Compared with the $T_s[n]$ case, the effect of the numerical noise in $\delta F_{model}[n]/\delta n(\mathbf{r})$ may be mitigated by the explicit use of the kinetic energy

operator $\hat{T} = -\nabla^2/2$ [84]. Still, stabilizing the KS cycles would be essential for reliable applications. Regularizations based on perturbation theory [99] and taking unconverged KS steps into the loss [109] have been proposed.

Another important derivative is the force on atoms $\mathbf{F}_I = -\partial E_{GS}[n]/\partial \mathbf{R}_I$, which is essential for performing a molecular dynamics simulation. If one can learn an accurate charge density, for the Coulombic system, the Hellmann-Feynman force formula [110]

$$\mathbf{F}_I = - \int d\mathbf{r} n(\mathbf{r}) \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}|^3} (\mathbf{R}_I - \mathbf{r}) + \sum_{I' \neq I} \frac{Z_I Z_{I'} e^2}{|\mathbf{R}_I - \mathbf{R}_{I'}|^3} (\mathbf{R}_I - \mathbf{R}_{I'}) \qquad (32)$$

enables us to calculate the force with the same accuracy. Dick and Fernandez-Serra [97] thus executed molecular dynamics with an accurate force field for liquid phases of small molecules, obtaining accurate radial distribution functions. For frameworks that do not target accurate charge density, some additional precalculation is required such as the atomic-position derivative of the atom-centered basis set [98] similar to the Pulay term [111]. Bogojeski et al. demonstrated accurate MD simulation of a molecule, by intermittently correcting the baseline force by the correction term calculated with the finite-difference method to the ML higher-rung energy [104].

### 4.2.5 Other quantities

The quantity of central interest in DFT calculations is the energy, while other quantities at the ground state, including responses functions, are also functionals of $n(\mathbf{r})$. Moreno, Carleo and Georges [112] attempted the fitting of the many body wavefunction and two-body correlation function in a 1D extended Hubbard model as function(al)s of occupations on discrete sites [113], implying that those high-dimensional quantities are on reduced latent dimensions of site occupations in line with the Hohenberg-Kohn theorems [113]. See Denner *et al.* for a related study [114]. Applications to single-particle excitations from the ground states [115] and real-time dynamics [116] have also been demonstrated.

### 4.3 Learning through the KS cycle

For the training of the universal functionals, one may desire to refer to quantities other than the functional itself, such as $n(\mathbf{r})$, from the expectation that the condition number in the model optimization should be increased and agreement with the observable quantities are improved. This is also a matter of curiosity that whether people can improve accuracy of $n(\mathbf{r})$, the fundamental variable of DFT, even if accurate observation of it is difficult. The task appears tricky when the quantities are not dependent explicitly on the model parameters but dependent only through the
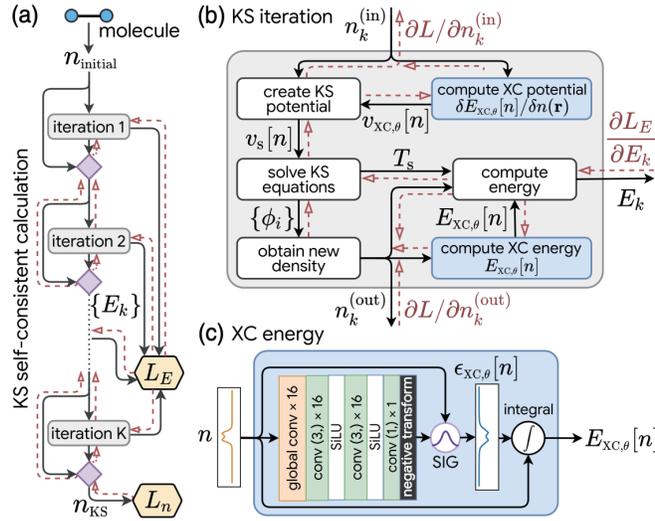
Fig. 3: **(a)–(c)** Learning through the differentiable KS equation, taken from Ref. [109] under the terms of the CC BY 4.0 license. The exchange-correlation energy [**(c)**] is modeled. The loss function is defined with the energies at unconverged steps and converged charge density [**(a)**]. The parameter derivatives backpropagate through the iterative calculations by the chain rule [**(b)**]. This helps produce a functional which converges.

KS solution using parametrized functionals. Nagai et al. [89] trained the exchange-correlation energy density with the loss function defined by $n(\mathbf{r})$. They optimized the parameters by derivative-free Monte Carlo sampling and indeed showed that the training on only three molecules can generalize to 147 molecules, although the optimization was extremely time-consuming. Similar generalization from a few to hundreds of molecules has been demonstrated by Wang et al. [117], showing the efficiency of the training on $n(\mathbf{r})$. Thankfully, since then, the usual gradient-based optimization has been already made applicable as we sketch below.

Chen et al. [98] took another approach which may be applied to learning of other quantities calculated through the self-consistent KS solution. Let $Q$ be any quantity to learn which is in theory formulated with the outputs of the converged KS equation; e.g., $Q = n$. Their learning scheme is then expressed as the following minimization[2]

$$\min_{\omega} \sum_{i=1}^{N_{\text{data}}} \left[ E_{\text{data}}^{(i)} - E_{\text{KS}}^{(i)}(\omega) \right], \tag{33}$$

---

[2] We omit the force term in the original work as it is not necessary for the present discussion.

where $E_{\mathrm{KS}}^{(i)}$ denotes the ground state energy from *a modified* KS equation modeled with a set of parameters $\omega$:

$$E_{\mathrm{KS}}^{(i)} = \min_{\{\varphi_j\}, \Sigma_{j:\mathrm{occ.}} |\varphi_j|^2 = n(\mathbf{r})} \left[ F_\omega[n] + E_{\mathrm{ext}}^{(i)}[n] + \lambda_Q D[Q[n], Q_{\mathrm{data}}^{(i)}] \right]. \quad (34)$$

The parametrized universal function $F_\omega$ is assumed to be the target to model. Here $D[Q, Q']$ is a given nonnegative function which becomes zero only when $Q$ and $Q'$ agree., whereas $\lambda_Q$ is the regularization parameter. As a key trick, they switched the order of the minimizations from $\min_\omega \left( \min_{\{\varphi_j\}, \Sigma_{j:\mathrm{occ.}}} \cdots \right)$ to $\min_{\{\varphi_j\}, \Sigma_{j:\mathrm{occ.}}} (\min_\omega \cdots)$. The gradient optimization of the parameter $\omega$ is then executed as usual with a fixed set of orbitals $\{\varphi_j^{(i)}\}$ where $i$ and $j$ run over the training systems and occupied states, respectively. The orbitals are infrequently updated by the parametrized KS equation including a penalty potential $V_Q^{(i)} = \lambda_Q \delta D[Q[n], Q_{\mathrm{data}}^{(i)}] / \delta n$. Those optimizations are repeated until convergence, after which the optimized $F_\omega$ is in hand that is tuned to best yield $Q$ as well as $E$ for given training systems.

The idea that the KS self-consistent calculation can be made differentiable is remarkable [109]. Any mathematical algorithm implemented in a programming language is executed by sequences of fundamental operations; addition, subtraction, multiplication, and division. From this, even complicated algorithms can be described in terms of composite functions formed by those fundamental operations and therefore its derivatives can be formulated by the chain rule. The differentiable implementation of the KS self-consistent cycle has then emerged and enabled us to differentiate the physical quantities calculated via the self-consistent KS solution. For a one dimensional model, Li et al. [109] optimized the NN-model exchange-correlation energy with respect to the loss function referring to the converged charge density. Soon after Kasim and Vinko [118] and Dick and Fernandez-Serra [119] demonstrated a three-dimensional implementation for the optimization of the NN-model meta GGA. This technology is widely applicable to ML-DFT as it extends the variety of the learning framework, though the differentiable programming has its own challenges as it requires large memory to store the numerical derivatives.

## 4.4 Machine-learning and exact conditions

Even with the recent flourish of ML-DFT, the traditional analytical approaches remain important. Rigorous equalities and inequalities from the latter are useful for regularizing the behavior of the trained ML models when applied to less familiar systems. Usage of those as constraints have been explored as summarized below.

Uniform scaling equalities and inequalities are properties satisfied by the exact functional [121]. Under uniform scale transformation, the charge density transforms as:

Table 1: Performance of the ML meta GGA model ("pcNN-based") with exact constraints; data taken from Ref. [120]. The model was trained on three molecules. "NN-based" represents the model from Ref. [89], where the constraints were not imposed. (a) Atomization energy benckmark on 144 untrained molecules. (b) Optimum lattice constants of 48 solids. Parentheses in column "NN-based" indicates that 6 systems were excluded because of ill convergence of the KS cycle. (MAE, mean absolute error; ME, mean error; SD, standard deviation of signed error)

(a). Atomization energies of molecules

| XC | MAE (kcal/mol) | ME (kcal/mol) | SD (kcal/mol) |
|---|---|---|---|
| PBE | 17.3 | 16.2 | 13.1 |
| SCAN | 6.2 | -4.5 | 5.7 |
| NN-based | 4.8 | 1.8 | 6.3 |
| pcNN-based | 3.6 | 0.3 | 4.5 |

(b). Optimized lattice constants of solids

| XC | MAE (mÅ) | ME (mÅ) | SD (mÅ) |
|---|---|---|---|
| PBE | 38.1 | 33.9 | 44.2 |
| SCAN | 22.3 | -7.5 | 28.5 |
| NN-based | (22.9) | (0.8) | (32.0) |
| pcNN-based | 19.1 | -2.5 | 26.5 |

$$n(\mathbf{r}) \rightarrow n_\gamma(\mathbf{r}) = \gamma^D n(\gamma \mathbf{r}), \tag{35}$$

where $D$ denotes the spatial dimension. With this transformation, the exact kinetic energy transforms as $T_s[n] \rightarrow T_s[n_\gamma] = \gamma^2 T_s[n]$ and the exact exchange obeys $E_x[n] \rightarrow E_x[n_\gamma] = \gamma E_x[n]$. Hollingsworth et al. [122] proposed a preprocess to transform the input $n(\mathbf{r})$ to a definite scale $n(\mathbf{r}) \rightarrow \tilde{n}(\mathbf{r})$ so that the density distributions related by the scaling transformation are identified. The machine then learns only the processed $\tilde{n}(\mathbf{r})$ and the actual energy is calculated via the exact scaling transformation at the postprocess. A loosened version of the preprocess has also been proposed by Gong and coworkers [103]. They trained an NN encoder of density distribution in three dimensions so that it can relate the distributions matched by the translation and scale transformation and utilized the encoded features as descriptors of the exchange energy. Bystrom and Kozinsky formulated a weighted-density type descriptor that is explicitly invariant with respect to the scale transformation [90, 91]. Using this, they implemented an $E_x$ that is forced to satisfy the scaling invariance, and is succiciently accurate to replace exact exchange in global hybrids [102].

Various asymptotic formulas and inequalities are available [16] as constraints on an ML model. Building models which strictly satisfy such constraints–and therefore avoid overfitting to training data–is now possible [28]. Cuierrier *et al.* [123] and Sparrow *et al.* [124] included the variance from the exact constraints to the loss function, with which the model is regularized to comply with those approximately. Lagrange interpolation to the end point of an ML model [120] can force the model more strictly to converge to the desired asymptotic limits. Inequality constraints can also be enforced by using any bounded activation functions at the output [119]. Nagai et al. [120] have shown that, by designing an ML model which recovers the uniform electron limit and other constraints used in the strictly constrained and appropriately

normed (SCAN) meta GGA [16], the resulting functional is reliably applicable to solid systems, even when it was trained only on molecules (Table 1).

Linear dependence of the total energy on total number/spin of electrons is a significant exact constraint. DFT applied to non-integer $N = \int d\mathbf{r} n(\mathbf{r})$ by defining the non-integer ground state as a linear superposition of the ground state density matrices with integer electrons [125, 126]. Ultimately, this is simply the $T \to 0$ limit of Mermin's finite-temperature generalization of the Hohenberg-Kohn theorems to the grand canonical ensemble [127]. From this definition, the ground state energy as a function of $N$, $E_{\mathrm{GS}}(N)$, is linear between integer values of $N$ [126, 128]. This linearity condition, which is usually violated by approximate functionals, is found to be tied to a proper description of charge transfer within the system with definite $N$, affecting the accuracy of molecular dissociation and the charge gap [129, 130]. Researchers implemented the charge-spin linearity conditions to the model by construction [131] or training of the model with data enforcing the linearity [99]. In particular, the DM21 functional by Kirkpatrick and colleagues [99] succeeds in reproducing the dissociation curves of various molecules, passing the stringent test for DFT. Unfortunately, that approximation is costly to evaluate, and has significant convergence issues.

An important insight on using the exact constraints can be found in Hollingsworth et al. [122]. In their test of learning $T_{\mathrm{s}}[n]$ compliant to the scaling relationship in one dimensional models, significant improvement was seen in the case of the Hooke's atom, whereas for the stretched 1D $H_2$ the performance improvement was marginal. This is due to the failure of learning the size consistency because of the scaling preprocess $n(\mathbf{r}) \to \tilde{n}(\mathbf{r})$. The effective use of the constraints thus requires knowledge of which constraints regularize which physics [132]. It would be of great interest to apply such conditions with more modern ML methods to quantities such as the exact exchange.

## 4.5 Summary of Machine-Learning assisted DFT development

In Tables 2–5 we feature representative pioneering contributions mentioned in this section with their brief descriptions in the right columns. Table 2 summarizes DF quantitites as targets of ML fitting, whereas Tables 3 and 4 append significant conceptual and technical developments. For limiting the scope, we have omitted papers whose theoretical framework departs from the strict DFT. Table 5 is for some influential papers from out of ML-DFT.

Table 2: Representative contributions to the ML-DFT.

| Learning object | |
| --- | --- |
| *Kinetic energy* | |
| Snyder 2012-13 [58, 59] | Exact $T_s$ in nonlocal form in a 1D model with KRR |
| Yao & Parkill 2016 [61] | $\tau(\mathbf{r})$ in C-H molecules by weighted density approximation with NN |
| Seino 2018 [63] | $\tau(\mathbf{r})$ in small molecules by semilocal form with NN |
| *XC energy and potential* | |
| Ingamells & Handy 1996 [133] | XC potential from the KS inversion for light atoms from semilocal descriptors with NN |
| Tozer 1996 [79] | |
| Tozer 1997 [134] | Same but with flexible polynomial |
| Zheng 2004 [81] | Optimal B3LYP mixing parameters $a_0, a_x, a_c$ for small molecules as density functionals with NN |
| Li 2016 [74] | Accurate DMRG $E_{xc}[n]$ for 1D systems with KRR |
| Nagai 2018 [84] | XC potential from the KS inversion for a 1D model with a fully nonlocal NN model |
| Zhou 2019 [85] | XC potential from the KS inversion for H-He diatomic systems with a nonlocal convolutional NN model |
| Nagai 2020 [89] | $\varepsilon_{xc}(\mathbf{r})$ in small molecules in the weighted density approximation with NN, optimized with reference to CC energy and $n(\mathbf{r})$ via KS equation |
| Margraf & Reuter 2021 [96] | MP2 and CC $E_c$ in small molecules in a nonlocal expression with atomic decomposition of $n(\mathbf{r})$ implemented with KRR |
| *Density as potential functional* | |
| Brockherde 2017 [62] | GGA $n(\mathbf{r})$ as nonlocal potential functional in 1D and molecular systems via basis expansion with KRR |
| *Δ-ML (density functional from inaccurate density)* | |
| Bogojeski 2020 [104] | Training on CCSD(T) energies in small molecules by learning correction to PBE from PBE $n(\mathbf{r})$ with KRR |
| *Other quantities* | |
| Moreno 2020 [112] | Many particle wavefunction and correlation function with NN |
| Suzuki 2020 [116] | Time-dependent XC potential with NN |
| Bystrom 2024 [115] | $E_x$ and its derivative discontinuity by KRR with density matrix based descriptors |

Table 3: (Cont'd) Representative contributions in ML-DFT.

| *Descriptor design* | |
| --- | --- |
| Tozer 1996 [79] | Semilocal descriptors for $V_{xc}(\mathbf{r})$ along with traditional generalized gradient approach |
| Zheng 2004 [81] | Macroscopic quantities calculated from $n(\mathbf{r})$ for the B3LYP mixing parameters |
| Snyder 2012-13 [58, 59] | Fully nonlocal $n(\mathbf{r})$ for $T_s$ |
| Yao & Parkhill 2016 [61] | Weighted density for $\tau(\mathbf{r})$: Linear filters directing nearby atoms |
| Lei & Medford 2019 [88] | Weighted density for $\varepsilon_x(\mathbf{r})$: Atom-centered spherical harmonics |
| Margraf & Reuter 2021 [96] | Nonlocal $n(\mathbf{r})$ dependence with its density fitting decomposition and per-atom KRR kernel for $E_c$ |
| Kirkpatrick 2021 [99] | LDA and Hartree-Fock KS $\varepsilon_x(\mathbf{r})$ for $E_{xc}$ |
| Gong 2023 [103] | $n(\mathbf{r})$ in NN-encoded representation for $E_x$ |
| Riemelmoser 2023 [101] | Weighted density for $\varepsilon_{xc}(\mathbf{r})$: Atom-centered spherical harmonics efficiently calculated with fast Fourier transform |

| *Using derivatives for self consistent $n(\mathbf{r})$ calculation* | |
| --- | --- |
| Snyder 2012-13  [58, 59] | Regularizing the derivative by projection onto training data space |
| Schmidt 2019 [86] | Training $E_{xc}$ with loss referring to $\delta E_{xc}/\delta n(\mathbf{r}) = V_{xc}(\mathbf{r})$ derived from the KS inversion |
| Nagai 2020  [89] | Training $\varepsilon_{xc}$ with loss referring to $n(\mathbf{r})$ and energies: derivative-free optimization |
| Meyer 2020 [75] | Training $T_s$ with loss referring to $\delta T_s/\delta n$ |
| Chen 2021  [98] | Training $\varepsilon_{xc}$ with loss referring to $n(\mathbf{r})$: Penalty potential |
| Li 2021  [109], Kasim & Vinko 2021 [118], Dick & Fernandez-Serra 2021 [119] | Training $\varepsilon_{xc}$ with loss referring to $n(\mathbf{r})$: Differentiable programming |
| Kirkpatrick 2021 [99] | Training $E_{xc}$ with perturbation theory based regularization term in loss |
| Kanungo 2024  [87] | Training $\varepsilon_{xc}$ with loss referring to $\delta E_{xc}/\delta n(\mathbf{r}) = V_{xc}(\mathbf{r})$ derived from the KS inversion in 3D |

Table 4: (Cont'd) Representative contributions in ML-DFT.

| *Making machine respect exact conditions* | |
| --- | --- |
| *Uniform scaling condition* | |
| Hollingsworth 2018 [122] | Normalizing $n$ to a unified scale |
| Bystrom & Kozinsky 2022 [90] | Formulating scale invariant weighted density |
| Gong 2023 [103] | Using an encoder that identify two $n(\mathbf{r})$ related by scale transformation |
| *Asymptotic conditions* | |
| Cuierrier 2021 [123], Dick & Fernandez-Serra 2021 [119], Sparrow 2022 [124], | Defining loss with departure from exact conditions |
| Nagai 2022 [120] | Model explicitly constrained by Lagrange interpolation form |
| *Linearity with respect to electron number* | |
| Kirkpatrick 2021 [99] | Training the model with fractional electron-number cases |
| Gedeon 2022 [131] | Defining the model with derivative discontinuity by design |
| *Global rotational invariance* | |
| Margraf & Reuter 2021 [96] | Rotationally invariant descriptors from atom-centered expansion coefficients of $n(\mathbf{r})$ |
| Chen 2021 [98] | Rotationally invariant descriptors with eigenvalues of density matrix |

Table 5: Signficant papers that have helped ML-DFT.

| | |
| --- | --- |
| Bartok 2013 [92], Grisafi 2018 [93] | Rotationally covariant descriptors for atomic environment |
| Kim 2013 [107] | Understanding XC approximation by density-driven and functional-driven errors |
| Ramakrishnan 2015 [105] | $\Delta$-learning concept |

# 5 Where do we go from here?

There has been much progress in the proofs of new concepts in the DFT thanks to ML. In this section we attempt to summarize the achievements and raise an issue of the importance for the DFT framework deduced from the ML applications.

## 5.1 Generalization: Summary

We classify the types of generalization presented in published papers: *Configurational generalization* [58–60, 62, 64, 67–70, 73–75, 77, 84, 86, 96, 97, 104, 109, 112, 114, 122, 131, 135–138] means generalization within a fixed system, where the training is performed for various atomic configurations and test is done for untrained configurations of the same system. *Compositional generalization* [63, 66, 71, 72, 79, 87, 89, 91, 98, 99, 101, 115, 118–120, 139–148] denotes the case where the training is performed for a certain group of compounds and the test is done for other compounds not seen during training. The latter generalization is the very goal of DFT development, whereas the former generalization is also useful for thermodynamic methods like molecular dynamics. Some studies indicate that ML can achieve configurational generalization by skipping rungs in Jacob's ladder while retaining coupled cluster accuracy, based on the $\Delta$-learning framework. Bogojeski et al. [104] implemented a KRR model that relates $\mathbf{n}$ to the energy difference between CC and GGA, $\Delta E_{\mathrm{model}}^{\mathrm{CC-GGA}}[n_{\mathrm{model}}^{\mathrm{DFT}}]$[3], and built models optimized for molecular $H_2O$, ethanol and resorcinol; all of which yielded accurate potential surfaces with errors < 1 kcal/mol from the CC results. Dick and Fernandez-Serra [97] also built a model for the energy difference from the baseline functional (GGA) that is also which yields more accurate $n$ than the baseline approximation. Their model trained on the $H_2O$ trimer was used for molecular dynamics with supercell containing 96 $H_2O$ molecules, yielding very accurate radial distribution functions. Both studies attain chemical accuracy at GGA cost, skipping three steps of the ladder Fig. 1.

We can find some unexpected generalizations from the configurational studies, where the machines retain accuracy even in the regimes to which we would not anticipate extrapolation. In Brockherde et al. [62], a model was trained on malonaldehyde, where configurations on the verge of the proton transfer were not present. Nevertheless, it accurately described the proton transfer process when applied to the molecular dynamics (Fig. 4b), though less accurately than for other configurations. Generalization of this kind is what physics-oriented researchers hope to achieve in ML applications, i.e. that the machine finds some apparent extrapolation is actually an interpolation, according to the latent space embedded in the training model. The configurational studies also give us clear insights where the models do not general-

---

[3] The model accepts approximate $n_{\mathrm{model}}^{\mathrm{DFT}}$, which is calculated by another KRR model that relates the ionic potential and $n$ calculated from the semilocal DFT; the calculation of this term only can therefore be done by further reduced cost.

ize and why. Nagai et al. [84] examined a simple one-dimensional Hamiltonian with variable number of electrons bound in a potential. They showed that the machine trained solely on potentials which bind only one electron never generalizes to potentials which bind two (Fig. 4a), and vice versa. Closer analysis of such successful and failed generalizations in the literature may give us insights that are brought back to the analytical DFT development.

The compositional kind of generalizations are usually shown by the study design where researchers take training data for compound databases and divide them into training, validation and test groups. In such studies the DM21 functional [99] should be worth highlighting (Fig. 4c). They prepared exchange-correlation energies with CCSD(T) accuracy for a training set consisting of 1161 molecular reactions and 1074 cases of fractionally charged or spin-polarized atoms from H to Ar. The exchange-correlation energy density was modeled as a functional of the semilocal density descriptor calculated with the B3LYP functional (fourth rung). Training was done with an additional regularization term in the loss that is intended for stabilizing the KS self-consistent loop. The trained NN model for the exchange-correlation energy density was tested by the KS self-consistent calculations for over $10^5$ systems from the GMTKN55 and QM9 datasets, showing accuracy of double hybrid (fifth rung); one level skipped. The similarity between the training and test data seems low, which may indicate that the machine learned some latent dimensions in which the test systems are regarded interpolation.

The most interesting goal for the compositional models would be to use it for exploration of new compounds. For this, the models have to attain some confidence when they are applied to undiscovered materials in *in silico* , but this goal still seems distant. For instance, some "out-of-paper" tests for DM21 reported untested systems where the DM21's accuracy declines, such as water clusters [149] and transition metal compounds [150]. Analyses of the latent space in the trained compositional functionals in some way would help human to judge if the functionals is applicable to any given undiscovered systems. As an related effort let us recall Pokharel et al. [141] which examine how broadly a functional which learned a few selected "norm" systems, to which the model is assured to agree, may generalize. A more data-based approach to formulate the similarity measure between the databases is also interesting [151].
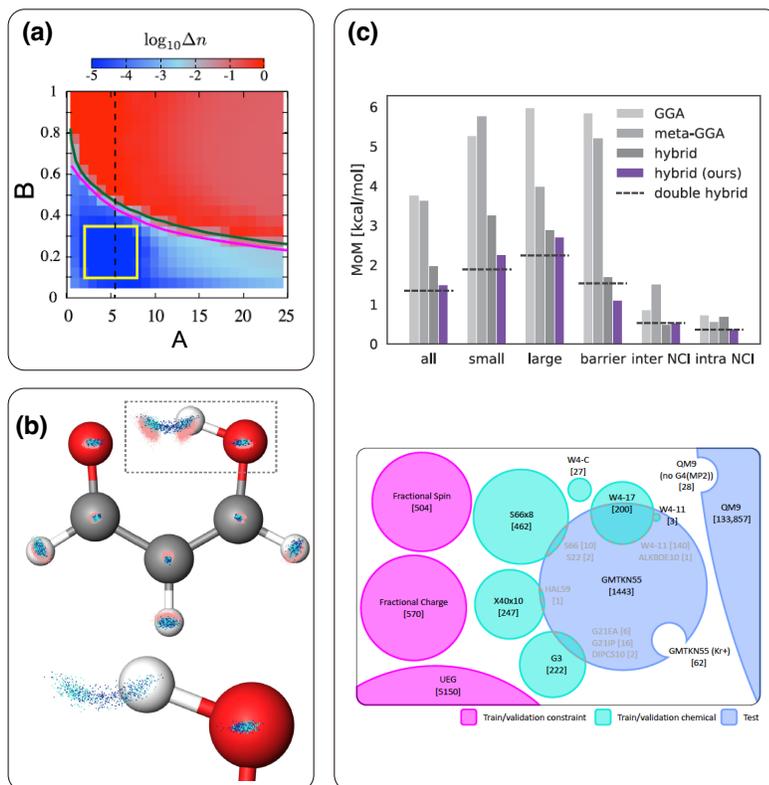
Fig. 4: Suggestive results concerning generalization. **a** Clear example of failed generalization. In a 1D model the ionic potential is controlled by two parameters *A* and *B*. The machine exclusively trained on data in a single phase (yellow frame) does not generalize beyond the phase boundary. Reprinted from R. Nagai, R. Akashi, S. Sasaki and S. Tsuneyuki "Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability", J. Chem. Phys. **148**, 241737 (2018) with the permission of AIP Publishing [84]. **b** Configurations of malonaldehyde. The model trained with red points generate geometries including the proton transfer when applied to MD (blue), that agree well with those generated by the first-principles MD. Taken from Ref. [62] under the terms of the Creative Commons CC BY. **c** (top) Performance of the DM21 functional reaching the accuracy of double hybrid, at the cost of a single hybrid. (bottom) The training and test data, whose overlap seems small. From J. Kirkpatrick *et al.*, "Pushing the frontiers of density functionals by solving the fractional electron problem", Science **374**, 1385 (2021) [99]. Reprinted with permission from AAAS.
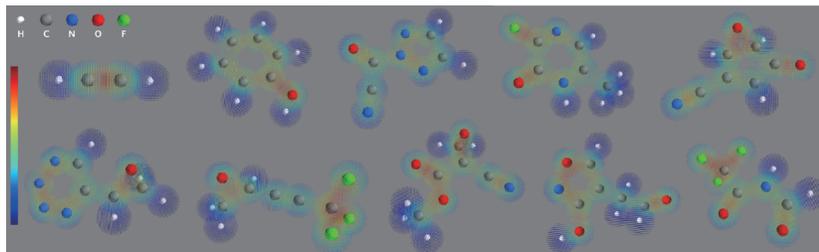
Fig. 5: "Charge density" as an intermediate field value in a trained machine, which was not directly referenced in the training, taken from Ref. [139] under the terms of the CC BY 4.0 license.

## 5.2 How to measure the accuracy of the density

Ground-state DFT is a theory based on the ground-state charge density. However, it really only uses the density as an auxillary variable. Its chief use is to produce approximate energies for various different configurations of nuclear potentials. That is what the thousands of applications that are published each year mostly do.

However, whether we need accurate densities for accurate machine learning can also be in question [107, 152–154]. As a matter of fact, there is much fascination about development of models that bypass DFT altogether, by formulating the machines with descriptors derived from atomic positions and compositions of the system concerned, from which it is implied that for well-controlled interpolation tasks we do not have to refer to the charge density. This is used to make both machine-learned potentials and also models for other properties, such as dipole moments, that we would usually extract from integrals over the density.

Table 6: Relevant dimension estimation of approximate densities for describing energies; data taken from Ref. [104].

| Density/Energy | $E_{\mathrm{SML}}^{\mathrm{DFT}}$ | $E_{\mathrm{SML}}^{\mathrm{CC}}$ | $E_{\mathrm{SA\text{-}DFT}}^{\mathrm{CC}}$ | $E_{\mathrm{SA}}^{\mathrm{SAD}}$ | $E_{\mathrm{SA\text{-}SAD}}^{\mathrm{CC}}$ |
|---|---|---|---|---|---|
| $n^{\mathrm{DFT}}$ | 1196.56 | 1208.05 | 4335.93 | n/d | n/d |
| $n_s^{\mathrm{DFT}}$ | 1162.68 | 1161.95 | 4010.18 | n/d | n/d |
| $n^{\mathrm{SAD}}$ | 661.39 | 623.38 | 3639.97 | 677.66 | 374.98 |

n/d = not determined

If we do not need too accurate $n(\mathbf{r})$ for practical accuracy, it should further accelerate our development since the accurate calculation of $n(\mathbf{r})$ beyond Hartree-Fock or semilocal DFT for training is itself a demanding task, and also the storage of accurate $n(\mathbf{r})$ requires much more disk resources than energy-related observables.

On this issue let us refer to Brockherde et al. [62]. Two machines were tested there: one is a machine that maps the one-body potential $V_{ext}$ to the total energy (termed machine-learning Kohn-Sham map), and the other is a composite machine in which two machines respectively map $V_{ext}(\mathbf{r})$ to $n(\mathbf{r})$ (termed machine-learning Hohenberg-Kohn map) and $n(\mathbf{r})$ to $E$. The latter $n(\mathbf{r}) \rightarrow E$ machine then learns the functional relating the approximate $n(\mathbf{r})$ to the total energy, say, $E[n_{approx.}]$, that is of course different from the exact density functional. Interestingly, the latter machine $V_{ext}(\mathbf{r}) \rightarrow n(\mathbf{r}) \rightarrow E$ achieved reduction of the test errors more easily. This scheme has been further sophisticated as $\Delta$-DFT for learning the exchange-correlation energy with the coupled-cluster accuracy [104].

Tsubaki and Mizoguchi [139] also showed an interesting demonstration. They developed a machine composed of two functionals that commonly use the LCAO coefficients of the orbitals as input. One relates the coefficients to the atomization energy as the NN output. The other first calculate the charge density $n(\mathbf{r})$ by the exact KS formula (Eq. 9) and next calculate the external potential from the calculated $n(\mathbf{r})$ through a NN. The two NNs are optimized so that the end data, atomization energy and the potential, agree well with the training data. On training it does not appear to matter whether the intermediate $n(\mathbf{r})$ is accurate; however, the trained machine gave plausible (but of course inaccurate) $n(\mathbf{r})$ (Fig. 5). Incorporating the orbital-to-$n(\mathbf{r})$ relation into the machine thus seems effective for efficient learning of the energy, even if the machine does not learn $n(\mathbf{r})$ directly.

Bogojeski *et al.* [104] examined the accuracy of $E_{xc}[n]$ with different levels of approximation to $n(\mathbf{r})$. They applied the relevant dimension estimation method in the data science [155] and calculated the signal-noise ratio, finding that, as $n(\mathbf{r})$ becomes inaccurate, the noise in the input-output pair becomes larger (Table 6), which indicates that an accurate $n(\mathbf{r})$ carries more information relevant to $E_{xc}$.

For understanding the issue of learning accurate densities, the framework of density-corrected DFT (DC-DFT) [107] should prove useful. A general analysis (formalized in a broader context than ML alone) allows for the decomposition of the error of a DFT calculation into two pieces: (i) the error due to the approximation used for the form of $E_{XC}$ and (ii) the error due to the approximate density which is found by minimizing the approximate functional. Letting $\tilde{E}$ and $\tilde{n}(\mathbf{r})$ denote the approximate energy functional and density respectively; with $E$ and $n(\mathbf{r})$ being the exact functional and density, the total error can be written as:

$$\Delta E = \tilde{E}[\tilde{n}] - E[n] = (\tilde{E}[n] - E[n]) + (\tilde{E}[\tilde{n}] - \tilde{E}[n]) \tag{36}$$

where $\tilde{E}[n] - E[n]$ is called the "functional driven error" and $\tilde{E}[\tilde{n}] - \tilde{E}[n]$ is called the "density driven error". In the majority of cases, the functional-driven error dominates, but in some important systems, such as those containing radicals, the density-driven error is dominant. A DC-DFT calculation specifically seeks to reduce the density-driven error by evaluating an approximate functional on a *different* density than its variational minimum. Such a calculation is refered to as "inconsistent" in the literature, as opposed to the *self*-consistent solution of KS equations [156]. Often,

Hartree Fock densities are used for this purpose (such a calculation is called HF-DFT), and for many molecules, this greatly reduces the density-driven error [153]. While HF-DFT has so far been the main practical manifestation of density correction, the principle of DC-DFT is much broader, and is not restricted to Hartree Fock densities alone. This opens a path for constructing highly accurate ML functionals by learning the energy as a functional of a non-self-consistent density.

## 6 The future?

The application of machine learning in the physical sciences is producing a paradigm shift in these fields, creating many new capabilities that are driven by data and compute power. This was recognized by the Nobel prize in both physics and chemistry in 2024. A generation from now, ML will be as much a part of science as calculus is today.

In the specific area of electronic structure, the greatest impact so far has been the development of meachine-learned potentials (MLPs), which are already being used to generate new science (see other chapters). A great example is the MD simulation of a shock wave passing through a slab of carbon, with 18 billion carbon atoms [157]. Such a calculation is both impossible with DFT and not desirable, if (essentially) the exact same results are produced by the MLP. More recently, foundational models, trained on the entire periodic table, have appeared [158]. While less accurate and robust than is desirable, the ability to perform much larger simulations on almost any phase of matter will totally alter all sciences it is applied to, at both the computational and experimental levels. Such codes will be used by theorists and experimentalists alike.

Where does this leave DFT itself? Will it be no more than the starting point for such models, by generating their training data? The answer is yes and no. It will be used to generate training data, but also used to validate the models in places where they fail. It also seems likely that it will also be used to fine-tune to improve accuracy in specific situations. This is very analogous to how high-level quantum chemical and sometimes Monte Carlo data are used to benchmark and check DFT calculations today.

Which brings us to the topic of this chapter. The use of ML to directly improve approximate density functionals. We have discussed two specific functionals, the KS kinetic energy functional and the XC functional. While important pioneering steps have been made so far, no all purpose approximation has yet appeared that has been widely adopted by the user community. The hopes of DM21 have been dashed by its high computational cost and the difficulties with convergence.

We also want to make another distinction. Some work is aimed at taking a human-designed approximate form, and trying to improve over human attempts with that approximate form. Examples are GGA's and meta-GGA's and hybrids. An early example is the $\omega$B97 series of functionals [46, 159], which were data driven, and essentially regularized fits to avoid overfitting. Such attempts are unlikely to work on

problems where the human-designed counterparts fail, e.g. for strongly correlated systems. What they aim to do is improve accuracy on the systems that already work. A clever aspect of DM21 was to take a (relatively) little-explored human-designed form, so both improved accuracy and new capabilities could both be possible.

On the other hand, the rest of the work creates functionals that look nothing like anything suggested by people, i.e., which use density inputs that are not the usual semilocal suspects. Some even use the entire density itself, allowing the possibility of creating the exact functional defined by HK long ago. With enough data in the right form, it seems one can drive down errors far below the current forms, and create functionals that work even for strongly correlated systems.

So why has this approach not been more broadly explored? Ultimately, it creates a new and unexpected problem: it tends to generalize poorly. The amount of training data needed to allow generalization across chemical species appears daunting and impractical. This difficulty can be traced back to use of the *entire* density. A tiny change in a density anywhere can, in principle, substantially change the ML energy. In practice, this almost never happens, but occasionally it is important. It is very difficult for a general scheme to reproduce this behavior. It is useful to muse on the generality of the KS scheme with the simple LDA approximation, using only the density at a single point to generate the XC energy– it has zero dependence on what the chemical or material species is.

This suggests that for such methods to become more useful, and go beyond proofs-of-principle, one needs to identify a limited set of specific features in the density that can be calculated relatively easily and allow generalization across systems. These obviously include the usual semilocal features, but also must include some that capture effects relevant to electron localization, such as the electron number in the vicinity of a nucleus. Much experimentation is likely to be needed to discover such features.

The next decade should provide widespread use of foundational MLPs, with (we hope) some spectacular scientific progress based on these capabilities. We can also hope to see XC or orbital-free approximations developed with ML becoming widely available and used, improving in accuracy over human-designed approximations in at least some areas, and maybe even working in new areas where traditional DFT approximations fail.

We shall see.

## 7 Funding

# References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in Advances in neural information processing systems, Vol. 25, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger (2012), `https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf`.

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning", arXiv:1312.5602, `https://arxiv.org/abs/1312.5602`.

[3] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search", Nature **529**, 484–489 (2016), `https://doi.org/10.1038/nature16961`.

[4] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", Phys. Rev. **136**, B864–B871 (1964), `https://link.aps.org/doi/10.1103/PhysRev.136.B864`.

[5] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects", Phys. Rev. **140**, A1133–A1138 (1965), `https://link.aps.org/doi/10.1103/PhysRev.140.A1133`.

[6] *The 1998 nobel prize in chemistry*, https://www.nobelprize.org/prizes/chemistry/1998/summary/, Accessed: 27th January 2025.

[7] B. Huang, G. F. von Rudorff, and O. A. von Lilienfeld, "The central role of density functional theory in the ai age", Science **381**, 170–175 (2023), `https://www.science.org/doi/abs/10.1126/science.abn3445`.

[8] L. H. Thomas, "The calculation of atomic fields", Mathematical Proceedings of the Cambridge Philosophical Society **23**, 542–548 (1927).

[9] E. Fermi, "Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente", Zeitschrift für Physik **48**, 73–79 (1928), `https://doi.org/10.1007/BF01351576`.

[10] M. Levy, "Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem", Proc. Natl. Acad. Sci. USA **76**, 6062–6065 (1979).

[11] E. H. Lieb, "Density functionals for coloumb systems", International Journal of Quantum Chemistry **XXIV**, 243–277 (1983).

[12] K. J. J.C. Slater, "Self-consistent field x-alpha cluster method for polyatomic molecules and solids", Phys. Rev. B (1972).

[13] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, "Generalized kohn-sham schemes and the band-gap problem", Phys. Rev. B **53**, 3764–3774 (1996), `https://link.aps.org/doi/10.1103/PhysRevB.53.3764`.

[14]   R. Garrick, A. Natan, T. Gould, and L. Kronik, "Exact generalized kohn-sham theory for hybrid functionals", Phys. Rev. X **10**, 021040 (2020), `https://link.aps.org/doi/10.1103/PhysRevX.10.021040`.

[15]   W. Mi, K. Luo, S. B. Trickey, and M. Pavanello, "Orbital-free density functional theory: an attractive electronic structure method for large-scale first-principles simulations", Chemical Reviews **123**, PMID: 37870767, 12039–12104 (2023), `https://doi.org/10.1021/acs.chemrev.2c00758`.

[16]   J. Sun, A. Ruzsinszky, and J. P. Perdew, "Strongly constrained and appropriately normed semilocal density functional", Phys. Rev. Lett. **115**, 036402 (2015), `https://link.aps.org/doi/10.1103/PhysRevLett.115.036402`.

[17]   K. Burke, F. G. Cruz, and K.-C. Lam, "Unambiguous exchange-correlation energy density", The Journal of Chemical Physics **109**, 8161–8167 (1998), `https://doi.org/10.1063/1.477479`.

[18]   J. P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy", AIP Conference Proceedings **577**, 1–20 (2001), `https://doi.org/10.1063/1.1390175`.

[19]   J. Toulouse, *Review of approximations for the exchange-correlation energy in density-functional theory, in "density functional theory"* (2023), pp. 1–90.

[20]   F. Bloch, "Bemerkung zur elektronentheorie des ferromagnetismus und der elektrischen leitfähigkeit", Zeitschrift für Physik **57**, 545–555 (1929), `https://doi.org/10.1007/BF01340281`.

[21]   P. A. M. Dirac, "Note on exchange phenomena in the thomas atom", Mathematical Proceedings of the Cambridge Philosophical Society **26**, 376–385 (1930).

[22]   M. N. S.H. Vosko L. Wilk, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis", Canadian Journal of Physics **58**, 1200 (1980).

[23]   J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems", Phys. Rev. B **23**, 5048–5079 (1981), `https://link.aps.org/doi/10.1103/PhysRevB.23.5048`.

[24]   Y. W. John Perdew, "Accurate and simple analytic representation of the electron-gas correlation energy", Phys. Rev. B **45**, 12344 (1992).

[25]   D. Ceperley, "Ground state of the fermion one-component plasma: a monte carlo study in two and three dimensions", Phys. Rev. B **18**, 3126–3138 (1978), `https://link.aps.org/doi/10.1103/PhysRevB.18.3126`.

[26]   D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method", Phys. Rev. Lett. **45**, 566–569 (1980), `https://link.aps.org/doi/10.1103/PhysRevLett.45.566`.

[27]   S. Misawa, "Ferromagnetism of an electron gas", Phys. Rev. **140**, A1645–A1648 (1965), `https://link.aps.org/doi/10.1103/PhysRev.140.A1645`.

[28]   J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple", Phys. Rev. Lett. **77**, 3865–3868 (1996), `https://link.aps.org/doi/10.1103/PhysRevLett.77.3865`.

[29]  Y. Wang and J. P. Perdew, "Spin scaling of the electron-gas correlation energy in the high-density limit", Phys. Rev. B **43**, 8911–8916 (1991), `https://link.aps.org/doi/10.1103/PhysRevB.43.8911`.

[30]  M. Levy, "Asymptotic coordinate scaling bound for exchange-correlation energy in density-functional theory", International Journal of Quantum Chemistry **36**, 617–619 (1989), `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560360864`.

[31]  G. L. Oliver and J. P. Perdew, "Spin-density gradient expansion for the kinetic energy", Phys. Rev. A **20**, 397–403 (1979), `https://link.aps.org/doi/10.1103/PhysRevA.20.397`.

[32]  M. Levy and J. P. Perdew, "Hellmann-feynman, virial, and scaling requisites for the exact universal density functionals. shape of the correlation potential and diamagnetic susceptibility for atoms", Phys. Rev. A **32**, 2010–2021 (1985), `https://link.aps.org/doi/10.1103/PhysRevA.32.2010`.

[33]  E. H. Lieb and S. Oxford, "Improved lower bound on the indirect coulomb energy", International Journal of Quantum Chemistry **19**, 427–439 (1981), `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560190306`.

[34]  J. K. N. Bjørk Hammer Lars Bruno Hansen, "Improved adsorption energetics within density-functional theory using revised perdew-burke-erznerhof functionals", Phys. Rev. B **59**, 7143 (199).

[35]  W. Y. Yingkai Zhang, "Comment on 'generalized gradient approximation made simple'", PRL **80**, 890 (1998).

[36]  J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, "Restoring the density-gradient expansion for exchange in solids and surfaces", Phys. Rev. Lett. **100**, 136406 (2007).

[37]  A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior", Phys. Rev. A **38**, 3098–3100 (1988), `https://link.aps.org/doi/10.1103/PhysRevA.38.3098`.

[38]  P. Elliot and K. Burke, "Non-empirical dedication of the parameter in the b88 exchange functional", Canadian Journal of Chemistry, `10.1139/V09-095` (2009), `https://doi.org/10.1139/V09-095`.

[39]  A. Cancio, G. P. Chen, B. T. Krull, and K. Burke, "Fitting a round peg into a round hole: asymptotically correcting the generalized gradient approximation for correlation", Journal of Chemical Physics **149**, 084116 (2018), `https://doi.org/10.1063/1.5021597`.

[40]  A. M. Wibowo-Teale, B. C. Huynh, T. Helgaker, and D. J. Tozer, "Classical reaction barriers in dft: an adiabatic-connection perspective", Journal of Chemical Theory and Computation **21**, 124–137 (2025), `https://doi.org/10.1021/acs.jctc.4c01038`.

[41]  A. D. Becke, "Density-functional thermochemistry. iii. the role of exact exchange", The Journal of Chemical Physics **98**, 5648–5652 (1993), `https://doi.org/10.1063/1.464913`.

[42]    J. P. Perdew, "Electronic structure of solids '91", in (Akademie Verlag, Berlin, 1991), p. 11.

[43]    C. Lee, W. Yang, and R. G. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density", Phys. Rev. B **37**, 785–789 (1988), `https://link.aps.org/doi/10.1103/PhysRevB.37.785`.

[44]    K. Kim and K. D. Jordan, "Comparison of density functional and mp2 calculations on the water monomer and dimer", The Journal of Physical Chemistry **98**, 10089–10094 (1994), `https://doi.org/10.1021/j100091a024`.

[45]    P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields", The Journal of Physical Chemistry **98**, 11623–11627 (1994), `https://doi.org/10.1021/j100096a001`.

[46]    N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals", Molecular Physics **115**, 2315–2372 (2017), `https://doi.org/10.1080/00268976.2017.1333644`.

[47]    C. M. Bishop, *Pattern recognition and machine learning* (Springer, New York, 2006).

[48]    C. A. Micchelli, Y. Xu, and H. Zhang, "Universal kernels", Journal of Machine Learning Research **7**, 2651–2667 (2006), `http://jmlr.org/papers/v7/micchelli06a.html`.

[49]    K. Hornik, "Approximation capabilities of multilayer feedforward networks", Neural Networks **4**, 251–257 (1991), `https://www.sciencedirect.com/science/article/pii/089360809190009T`.

[50]    G. Cybenko, "Approximation by superpositions of a sigmoidal function", Mathematics of Control, Signals and Systems **2**, 303–314 (1989), `https://doi.org/10.1007/BF02551274`.

[51]    A. Kolmogorov, "On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables", Proceedings of the USSR Academy of Sciences **108**, 179–182 (1956).

[52]    J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, and J. Sohl-Dickstein, "Deep neural networks as gaussian processes", arXiv:1711.00165, `https://arxiv.org/abs/1711.00165`.

[53]    R. J. Bartlett, "Many-body perturbation theory and coupled cluster theory for electron correlation in molecules", Annual Review of Physical Chemistry **32**, 359–401 (1981), `https://www.annualreviews.org/content/journals/10.1146/annurev.pc.32.100181.002043`.

[54]    G. D. Purvis III and R. J. Bartlett, "A full coupled-cluster singles and doubles model: the inclusion of disconnected triples", The Journal of Chemical Physics **76**, 1910–1918 (1982), `https://doi.org/10.1063/1.443164`.

[55]    K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories", Chemical

Physics Letters **157**, 479–483 (1989), `https://www.sciencedirect.com/science/article/pii/S0009261489873956`.

[56]    W. Yang, P. W. Ayers, and Q. Wu, "Potential functionals: dual to density functionals and solution to the $v$-representability problem", Phys. Rev. Lett. **92**, 146404 (2004), `https://link.aps.org/doi/10.1103/PhysRevLett.92.146404`.

[57]    O. Gunnarsson and B. I. Lundqvist, "Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism", Phys. Rev. B **13**, 4274–4298 (1976), `https://link.aps.org/doi/10.1103/PhysRevB.13.4274`.

[58]    J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, "Finding density functionals with machine learning", Physical review letters **108**, 253002 (2012).

[59]    J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K.-R. Müller, and K. Burke, "Orbital-free bond breaking via machine learning", The Journal of Chemical Physics **139**, 224104 (2013), `https://doi.org/10.1063/1.4834075`.

[60]    L. Li, T. E. Baker, S. R. White, and K. Burke, "Pure density functional for strong correlation and the thermodynamic limit from machine learning", Phys. Rev. B **94**, 245129 (2016), `https://link.aps.org/doi/10.1103/PhysRevB.94.245129`.

[61]    K. Yao and J. Parkhill, "Kinetic energy of hydrocarbons as a function of electron density and convolutional neural networks", Journal of Chemical Theory and Computation **12**, PMID: 26812530, 1139–1147 (2016), `https://doi.org/10.1021/acs.jctc.5b01011`.

[62]    F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, "Bypassing the kohn-sham equations with machine learning", Nat. Commun. **8**, 872 (2017).

[63]    J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, "Semi-local machine-learned kinetic energy density functional with third-order gradients of electron density", The Journal of Chemical Physics **148**, 241705 (2018), `https://doi.org/10.1063/1.5007230`.

[64]    P. Golub and S. Manzhos, "Kinetic energy densities based on the fourth order gradient expansion: performance in different classes of materials and improvement via machine learning", Phys. Chem. Chem. Phys. **21**, 378–395 (2019), `http://dx.doi.org/10.1039/C8CP06433D`.

[65]    P. Golub and S. Manzhos, "Conundrum: a program for orbital-free density functional theory calculations", Computer Physics Communications **256**, 107365 (2020), `https://www.sciencedirect.com/science/article/pii/S0010465520301570`.

[66]    F. Imoto, M. Imada, and A. Oshiyama, "Order-$N$ orbital-free density-functional calculations with machine learning of functional derivatives for semiconductors and metals", Phys. Rev. Res. **3**, 033198 (2021), `https://link.aps.org/doi/10.1103/PhysRevResearch.3.033198`.

[67]    J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, "Semi-local machine-learned kinetic energy density functional demonstrating smooth

potential energy curves", Chemical Physics Letters **734**, 136732 (2019), `https : / / www . sciencedirect . com / science / article / pii / S0009261419307134`.

[68]  S. Manzhos and P. Golub, "Data-driven kinetic energy density fitting for orbital-free dft: linear vs gaussian process regression", The Journal of Chemical Physics **153**, 074104 (2020), `https : / / doi . org / 10 . 1063 / 5 . 0015042`.

[69]  K. Ryczko, S. J. Wetzel, R. G. Melko, and I. Tamblyn, "Toward orbital-free density functional theory with small data sets and deep learning", Journal of Chemical Theory and Computation **18**, PMID: 34995061, 1122–1128 (2022), `https://doi.org/10.1021/acs.jctc.1c00812`.

[70]  R. Remme, T. Kaczun, M. Scheurer, A. Dreuw, and F. A. Hamprecht, "Kinet-icnet: deep learning a transferable kinetic energy functional for orbital-free density functional theory", The Journal of Chemical Physics **159**, 144113 (2023), `https://doi.org/10.1063/5.0158275`.

[71]  H. Zhang, S. Liu, J. You, C. Liu, S. Zheng, Z. Lu, T. Wang, N. Zheng, and B. Shao, "Overcoming the barrier of orbital-free density functional theory for molecular systems using deep learning", Nature Computational Science **4**, 210–223 (2024), `https://doi.org/10.1038/s43588-024-00605-8`.

[72]  L. Sun and M. Chen, "Machine learning based nonlocal kinetic energy density functional for simple metals and alloys", Phys. Rev. B **109**, 115135 (2024), `https : / / link . aps . org / doi / 10 . 1103 / PhysRevB . 109 . 115135`.

[73]  J. C. Snyder, M. Rupp, K.-R. Müller, and K. Burke, "Nonlinear gradient de-noising: finding accurate extrema from inaccurate functional derivatives", International Journal of Quantum Chemistry **115**, 1102–1114 (2015), `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.24937`.

[74]  L. Li, J. C. Snyder, I. M. Pelaschier, J. Huang, U.-N. Niranjan, P. Duncan, M. Rupp, K.-R. Müller, and K. Burke, "Understanding machine-learned density functionals", International Journal of Quantum Chemistry **116**, 819–833 (2016), `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.25040`.

[75]  R. Meyer, M. Weichselbaum, and A. W. Hauser, "Machine learning ap-proaches toward orbital-free density functional theory: simultaneous train-ing on the kinetic energy density functional and its functional derivative", Journal of Chemical Theory and Computation **16**, PMID: 32786898, 5685–5694 (2020), `https://doi.org/10.1021/acs.jctc.0c00580`.

[76]  E. Costa, G. Scriva, R. Fazio, and S. Pilati, "Deep-learning density func-tionals for gradient descent optimization", Phys. Rev. E **106**, 045309 (2022), `https://link.aps.org/doi/10.1103/PhysRevE.106.045309`.

[77]  S. A. Ghasemi and T. D. Kühne, "Artificial neural networks for the kinetic energy functional of non-interacting fermions", The Journal of Chemical Physics **154**, 074107 (2021), `https://doi.org/10.1063/5.0037319`.

[78]  R. E. Turner, "An introduction to transformers", arXiv:2304.10557, `https://arxiv.org/abs/2304.10557`.

[79] D. J. Tozer, V. E. Ingamells, and N. C. Handy, "Exchange-correlation potentials", The Journal of Chemical Physics **105**, 9200–9213 (1996), `https://doi.org/10.1063/1.472753`.

[80] Q. Zhao, R. C. Morrison, and R. G. Parr, "From electron densities to kohn-sham kinetic energies, orbital energies, exchange-correlation potentials, and exchange-correlation energies", Phys. Rev. A **50**, 2138–2142 (1994), `https://link.aps.org/doi/10.1103/PhysRevA.50.2138`.

[81] X. Zheng, L. Hu, X. Wang, and G. Chen, "A generalized exchange-correlation functional: the neural-networks approach", Chemical Physics Letters **390**, 186–192 (2004), `https://www.sciencedirect.com/science/article/pii/S0009261404005603`.

[82] Q. Liu, J. Wang, P. Du, L. Hu, X. Zheng, and G. Chen, "Improving the performance of long-range-corrected exchange-correlation functional with an embedded neural network", The Journal of Physical Chemistry A **121**, PMID: 28876064, 7273–7281 (2017), `https://doi.org/10.1021/acs.jpca.7b07045`.

[83] D. Khan, A. James Arthur Price, M. L. Ach, and O. Anatole von Lilienfeld, "Adaptive hybrid density functionals", arXiv:2402.14793, `https://arxiv.org/abs/2402.14793`.

[84] R. Nagai, R. Akashi, S. Sasaki, and S. Tsuneyuki, "Neural-network kohn-sham exchange-correlation potential and its out-of-training transferability", J. Chem. Phys. **148**, 241737 (2018).

[85] Y. Zhou, J. Wu, S. Chen, and G. Chen, "Toward the exact exchange–correlation potential: a three-dimensional convolutional neural network construct", The Journal of Physical Chemistry Letters **10**, PMID: 31690079, 7264–7269 (2019), `https://doi.org/10.1021/acs.jpclett.9b02838`.

[86] J. Schmidt, C. L. Benavides-Riveros, and M. A. L. Marques, "Machine learning the physical nonlocal exchange–correlation functional of density-functional theory", The Journal of Physical Chemistry Letters **10**, 6425–6431 (2019), `https://doi.org/10.1021/acs.jpclett.9b02422`.

[87] B. Kanungo, J. Hatch, P. M. Zimmerman, and V. Gavini, "Learning local and semi-local density functionals from exact exchange-correlation potentials and energies", arXiv:2409.06498, `https://arxiv.org/abs/2409.06498`.

[88] X. Lei and A. J. Medford, "Design and analysis of machine learning exchange-correlation functionals via rotationally invariant convolutional descriptors", Phys. Rev. Mater. **3**, 063801 (2019), `https://link.aps.org/doi/10.1103/PhysRevMaterials.3.063801`.

[89] R. Nagai, R. Akashi, and O. Sugino, "Completing density functional theory by machine learning hidden messages from molecules", Npj Comput. Mater. **6**, 1–8 (2020).

[90] K. Bystrom and B. Kozinsky, "Cider: an expressive, nonlocal feature set for machine learning density functionals with exact constraints", Journal of Chemical Theory and Computation **18**, PMID: 35235322, 2180–2192 (2022), `https://doi.org/10.1021/acs.jctc.1c00904`.

[91]   K. Bystrom and B. Kozinsky, "Nonlocal machine-learned exchange func-
       tional for molecules and solids", Phys. Rev. B **110**, 075130 (2024), `https://link.aps.org/doi/10.1103/PhysRevB.110.075130`.

[92]   A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environ-
       ments", Phys. Rev. B **87**, 184115 (2013), `https://link.aps.org/doi/10.1103/PhysRevB.87.184115`.

[93]   A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, "Symmetry-adapted
       machine learning for tensorial properties of atomistic systems", Phys. Rev.
       Lett. **120**, 036002 (2018), `https://link.aps.org/doi/10.1103/PhysRevLett.120.036002`.

[94]   S. Dick and M. Fernandez-Serra, "Learning from the density to correct total
       energy and forces in first principle simulations", The Journal of Chemical
       Physics **151**, 144102 (2019), `https://doi.org/10.1063/1.5114618`.

[95]   A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and
       M. Ceriotti, "Transferable machine-learning model of the electron density",
       ACS Central Science **5**, PMID: 30693325, 57–64 (2019), `https://doi.org/10.1021/acscentsci.8b00551`.

[96]   J. T. Margraf and K. Reuter, "Pure non-local machine-learned density func-
       tional theory for electron correlation", Nature Communications **12**, 344
       (2021), `https://doi.org/10.1038/s41467-020-20471-y`.

[97]   S. Dick and M. Fernandez-Serra, "Machine learning accurate exchange and
       correlation functionals of the electronic density", Nature Communications
       **11**, 3509 (2020), `https://doi.org/10.1038/s41467-020-17265-7`.

[98]   Y. Chen, L. Zhang, H. Wang, and W. E, "Deepks: a comprehensive data-
       driven approach toward chemically accurate density functional theory", Jour-
       nal of Chemical Theory and Computation **17**, PMID: 33296197, 170–181
       (2021), `https://doi.org/10.1021/acs.jctc.0c00872`.

[99]   J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer,
       A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R.
       Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D.
       Hassabis, and A. J. Cohen, "Pushing the frontiers of density functionals by
       solving the fractional electron problem", Science **374**, 1385–1389 (2021),
       `https://www.science.org/doi/abs/10.1126/science.abj6511`.

[100]  J. P. Perdew, V. N. Staroverov, J. Tao, and G. E. Scuseria, "Density functional
       with full exact exchange, balanced nonlocality of correlation, and constraint
       satisfaction", Phys. Rev. A **78**, 052513 (2008), `https://link.aps.org/doi/10.1103/PhysRevA.78.052513`.

[101]  S. Riemelmoser, C. Verdi, M. Kaltak, and G. Kresse, "Machine learning den-
       sity functionals from the random-phase approximation", Journal of Chemical
       Theory and Computation **19**, PMID: 37800677, 7287–7299 (2023), `https://doi.org/10.1021/acs.jctc.3c00848`.

[102]  E. Polak, H. Zhao, and S. Vuckovic, "Real-space machine learning of corre-
       lation density functionals", ChemRxiv preprint, `https://chemrxiv.org/engage/chemrxiv/article-details/66e75a1b12ff75c3a16e9d00`.

[103] W. Gong, T. Sun, H. Bai, S. T. u. R. Chowdhury, P. Chu, A. Aryal, J. Yu, H. Ling, J. P. Perdew, and Q. Yan, "Incorporation of density scaling constraint in density functional design via contrastive representation learning", Digital Discovery **2**, 1404–1413 (2023), `http://dx.doi.org/10.1039/D3DD00114H`.

[104] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, "Quantum chemical accuracy from density functional approximations via machine learning", Nature Communications **11**, 5223 (2020), `https://doi.org/10.1038/s41467-020-19093-1`.

[105] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Big data meets quantum chemistry approximations: the $\delta$-machine learning approach", Journal of Chemical Theory and Computation **11**, PMID: 26574412, 2087–2096 (2015), `https://doi.org/10.1021/acs.jctc.5b00099`.

[106] A. V. Sinitskiy and V. S. Pande, "Deep neural network computes electron densities and energies of a large set of organic molecules faster than density functional theory (dft)", arXiv:1809.02723, `https://arxiv.org/abs/1809.02723`.

[107] M.-C. Kim, E. Sim, and K. Burke, "Understanding and reducing errors in density functional calculations", Phys. Rev. Lett. **111**, 073003 (2013), `https://link.aps.org/doi/10.1103/PhysRevLett.111.073003`.

[108] S. J. Sahoo, Q. Xu, X. Lei, D. Staros, G. R. Iyer, B. Rubenstein, P. Suryanarayana, and A. J. Medford, "Self-consistent convolutional density functional approximations: application to adsorption at metal surfaces", ChemPhysChem **25**, e202300688 (2024), `https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cphc.202300688`.

[109] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, "Kohn-sham equations as regularizer: building prior knowledge into machine-learned physics", Phys. Rev. Lett. **126**, 036401 (2021).

[110] R. P. Feynman, "Forces in molecules", Phys. Rev. **56**, 340–343 (1939), `https://link.aps.org/doi/10.1103/PhysRev.56.340`.

[111] P. Pulay, "Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules", Molecular Physics **17**, 197–204 (1969), `https://doi.org/10.1080/00268976900100941`.

[112] J. R. Moreno, G. Carleo, and A. Georges, "Deep learning the hohenberg-kohn maps of density functional theory", Phys. Rev. Lett. **125**, 076402 (2020), `https://link.aps.org/doi/10.1103/PhysRevLett.125.076402`.

[113] K. Schönhammer, O. Gunnarsson, and R. M. Noack, "Density-functional theory on a lattice: comparison with exact numerical results for a model with strongly correlated electrons", Phys. Rev. B **52**, 2504–2510 (1995), `https://link.aps.org/doi/10.1103/PhysRevB.52.2504`.

[114] M. M. Denner, M. H. Fischer, and T. Neupert, "Efficient learning of a one-dimensional density functional theory", Phys. Rev. Res. **2**, 033388 (2020), `https://link.aps.org/doi/10.1103/PhysRevResearch.2.033388`.

[115] K. Bystrom, S. Falletta, and B. Kozinsky, "Training machine-learned density functionals on band gaps", Journal of Chemical Theory and Computation **20**, PMID: 39178337, 7516–7532 (2024), `https://doi.org/10.1021/acs.jctc.4c00999`.

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

[117] J. Wang, Y. Wang, R.-X. Xu, G. Chen, and X. Zheng, "A semilocal machine-learning correction to density functional approximations", The Journal of Chemical Physics **158**, 154107 (2023), `https://doi.org/10.1063/5.0148438`.

[118] M. F. Kasim and S. M. Vinko, "Learning the exchange-correlation functional from nature with fully differentiable density functional theory", Phys. Rev. Lett. **127**, 126403 (2021), `https://link.aps.org/doi/10.1103/PhysRevLett.127.126403`.

[119] S. Dick and M. Fernandez-Serra, "Highly accurate and constrained density functional obtained with differentiable programming", Physical Review B **104**, L161109 (2021).

[120] R. Nagai, R. Akashi, and O. Sugino, "Machine-learning-based exchange correlation functional with physical asymptotic constraints", Physical Review Research **4**, 013106 (2022).

[121] M. Levy and J. P. Perdew, "Hellmann-feynman, virial, and scaling requisites for the exact universal density functionals. shape of the correlation potential and diamagnetic susceptibility for atoms", Phys. Rev. A **32**, 2010–2021 (1985), `https://link.aps.org/doi/10.1103/PhysRevA.32.2010`.

[122] J. Hollingsworth, L. Li, T. E. Baker, and K. Burke, "Can exact conditions improve machine-learned density functionals?", J. Chem. Phys. **148**, 241743 (2018).

[123] E. Cuierrier, P.-O. Roy, and M. Ernzerhof, "Constructing and representing exchange–correlation holes through artificial neural networks", The Journal of Chemical Physics **155**, 174121 (2021), `https://doi.org/10.1063/5.0062940`.

[124] Z. M. Sparrow, B. G. Ernst, T. K. Quady, and R. A. DiStasio Jr., "Uniting nonempirical and empirical density functional approximation strategies using constraint-based regularization", The Journal of Physical Chemistry Letters **13**, 6896–6904 (2022), `https://doi.org/10.1021/acs.jpclett.2c00643`.

[125] J. F. Janak, "Proof that $\frac{\partial E}{\partial n_i} = \epsilon$ in density-functional theory", Phys. Rev. B **18**, 7165–7168 (1978), `https://link.aps.org/doi/10.1103/PhysRevB.18.7165`.

[126] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, "Density-functional theory for fractional particle number: derivative discontinuities of the energy", Phys. Rev. Lett. **49**, 1691–1694 (1982), `https://link.aps.org/doi/10.1103/PhysRevLett.49.1691`.

[127]  N. D. Mermin, "Thermal properties of the ingomogeneous electron gas", Phys. Rev. **137**, A1441 (1965).

[128]  J. P. Perdew and M. Levy, "Physical content of the exact kohn-sham or-bital energies: band gaps and derivative discontinuities", Phys. Rev. Lett. **51**, 1884–1887 (1983), `https://link.aps.org/doi/10.1103/PhysRevLett.51.1884`.

[129]  A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Insights into current limitations of density functional theory", Science **321**, 792–794 (2008), `https://www.science.org/doi/abs/10.1126/science.1158722`.

[130]  A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Fractional spins and static correlation error in density functional theory", The Journal of Chemical Physics **129**, 121104 (2008), `https://doi.org/10.1063/1.2987202`.

[131]  J. Gedeon, J. Schmidt, M. J. Hodgson, J. Wetherell, C. L. Benavides-Riveros, and M. A. Marques, "Machine learning the derivative discontinuity of density-functional theory", Machine Learning: Science and Technology **3**, 015011 (2021).

[132]  R. Pederson and K. Burke, "The difference between molecules and materials: reassessing the role of exact conditions in density functional theory", The Journal of Chemical Physics **159**, 214113 (2023), `https://doi.org/10.1063/5.0172058`.

[133]  V. E. Ingamells and N. C. Handy, "Towards accurate exchange-correlation potentials for molecules", Chemical Physics Letters **248**, 373–378 (1996), `https://www.sciencedirect.com/science/article/pii/0009261495013415`.

[134]  D. J. Tozer, N. C. Handy, and W. H. Green, "Exchange-correlation functionals from ab initio electron densities", Chemical Physics Letters **273**, 183–194 (1997), `https://www.sciencedirect.com/science/article/pii/S0009261497005861`.

[135]  A. Ryabov, I. Akhatov, and P. Zhilyaev, "Neural network interpolation of exchange-correlation functional", Scientific Reports **10**, 8000 (2020), `https://doi.org/10.1038/s41598-020-64619-8`.

[136]  M. Fujinami, R. Kageyama, J. Seino, Y. Ikabata, and H. Nakai, "Orbital-free density functional theory calculation applying semi-local machine-learned kinetic energy density functional and kinetic potential", Chemical Physics Letters **748**, 137358 (2020), `https://www.sciencedirect.com/science/article/pii/S0009261420302736`.

[137]  S. Manzhos, J. Lüder, and M. Ihara, "Machine learning of kinetic energy densities with target and feature smoothing: better results with fewer training data", The Journal of Chemical Physics **159**, 234115 (2023), `https://doi.org/10.1063/5.0175689`.

[138]  P. d. Mazo-Sevillano and J. Hermann, "Variational principle to regularize machine-learned density functionals: the non-interacting kinetic-energy functional", The Journal of Chemical Physics **159**, 194107 (2023), `https://doi.org/10.1063/5.0166432`.

[139]  M. Tsubaki and T. Mizoguchi, "Quantum deep field: data-driven wave function, electron density generation, and atomization energy prediction and extrapolation with machine learning", Phys. Rev. Lett. **125**, 206401 (2020), `https://link.aps.org/doi/10.1103/PhysRevLett.125.206401`.

[140]  E. Cuierrier, P.-O. Roy, R. Wang, and M. Ernzerhof, "The fourth-order expansion of the exchange hole and neural networks to construct exchange–correlation functionals", The Journal of Chemical Physics **157**, 171103 (2022), `https://doi.org/10.1063/5.0122761`.

[141]  K. Pokharel, J. W. Furness, Y. Yao, V. Blum, T. J. P. Irons, A. M. Teale, and J. Sun, "Exact constraints and appropriate norms in machine-learned exchange-correlation functionals", The Journal of Chemical Physics **157**, 174106 (2022), `https://doi.org/10.1063/5.0111183`.

[142]  P. A. M. Casares, J. S. Baker, M. Medvidović, R. d. Reis, and J. M. Arrazola, "Graddft. a software library for machine learning enhanced density functional theory", The Journal of Chemical Physics **160**, 062501 (2024), `https://doi.org/10.1063/5.0181037`.

[143]  J. P. A. de Mendonça, L. A. Mariano, E. Devijver, N. Jakse, and R. Poloni, "Artificial neural network-based density functional approach for adiabatic energy differences in transition metal complexes", Journal of Chemical Theory and Computation **19**, 7555–7566 (2023), `https://doi.org/10.1021/acs.jctc.3c00600`.

[144]  A. Ryabov, I. Akhatov, and P. Zhilyaev, "Application of two-component neural network for exchange-correlation functional interpolation", Scientific Reports **12**, 14133 (2022), `https://doi.org/10.1038/s41598-022-18083-1`.

[145]  Z. Chen and W. Yang, "Development of a machine learning finite-range nonlocal density functional", The Journal of Chemical Physics **160**, 014105 (2024), `https://doi.org/10.1063/5.0179149`.

[146]  B. Kalita, R. Pederson, J. Chen, L. Li, and K. Burke, "How well does kohn–sham regularizer work for weakly correlated systems?", The Journal of Physical Chemistry Letters **13**, 2540–2547 (2022), `https://doi.org/10.1021/acs.jpclett.2c00371`.

[147]  J. Lüder, M. Ihara, and S. Manzhos, "A machine-learned kinetic energy model for light weight metals and compounds of group iii-v elements", Electronic Structure **6**, 045002 (2024), `https://dx.doi.org/10.1088/2516-1075/ad7e8d`.

[148]  G. Luise, C.-W. Huang, T. Vogels, D. P. Kooi, S. Ehlert, S. Lanius, K. J. H. Giesbertz, A. Karton, D. Gunceler, M. Stanley, W. P. Bruinsma, L. Huang, X. Wei, J. G. Torres, A. Katbashev, R. C. Zavaleta, B. Máté, S.-O. Kaba, R. Sordillo, Y. Chen, D. B. Williams-Young, C. M. Bishop, J. Hermann, R. van den Berg, and P. Gori-Giorgi, "Accurate and scalable exchange-correlation with deep learning", arXiv:2506.14665.

[149]  E. Palos, E. Lambros, S. Dasgupta, and F. Paesani, "Density functional theory of water with the machine-learned dm21 functional", The Journal of

Chemical Physics **156**, 161103 (2022), `https://doi.org/10.1063/5.0090862`.

[150] H. Zhao, T. Gould, and S. Vuckovic, "Deep mind 21 functional does not extrapolate to transition metal chemistry", Phys. Chem. Chem. Phys. **26**, 12289–12298 (2024), `http://dx.doi.org/10.1039/D4CP00878B`.

[151] T. Gould, B. Chan, S. G. Dale, and S. Vuckovic, "Identifying and embedding transferability in data-driven representations of chemical space", Chem. Sci. **15**, 11122–11133 (2024), `http://dx.doi.org/10.1039/D4SC02358G`.

[152] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, "Density functional theory is straying from the path toward the exact functional", Science **355**, 49–52 (2017), `https://www.science.org/doi/abs/10.1126/science.aah5975`.

[153] E. Sim, S. Song, and K. Burke, "Quantifying density errors in dft", The Journal of Physical Chemistry Letters **9**, PMID: 30335392, 6385–6392 (2018), `https://doi.org/10.1021/acs.jpclett.8b02855`.

[154] L. Cheng, P. B. Szabó, Z. Schätzle, D. Kooi, J. Köhler, K. J. H. Giesbertz, F. Noé, J. Hermann, P. Gori-Giorgi, and A. Foster, "Highly accurate real-space electron densities with neural networks", arXiv (2024), `https://arxiv.org/abs/2409.01306v1`.

[155] M. L. Braun, J. M. Buhmann, and K.-R. Müller, "On relevant dimensions in kernel feature spaces", Journal of Machine Learning Research **9**, 1875–1908 (2008), `http://jmlr.org/papers/v9/braun08a.html`.

[156] A. Wasserman, J. Nafziger, K. Jiang, M.-C. Kim, E. Sim, and K. Burke, "The importance of being inconsistent", Annual Review of Physical Chemistry (2017).

[157] K. Nguyen-Cong, J. T. Willman, S. G. Moore, A. B. Belonoshko, R. Gayatri, E. Weinberg, M. A. Wood, A. P. Thompson, and I. I. Oleynik, "Billion atom molecular dynamics simulations of carbon at extreme conditions and experimental time and length scales", in Proceedings of the international conference for high performance computing, networking, storage and analysis, SC '21 (2021), `https://doi.org/10.1145/3458817.3487400`.

[158] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, S. M. Blau, V. Cărare, J. P. Darby, S. De, F. Della Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, F. Falcioni, E. Fako, A. C. Ferrari, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. Holm, J. Jaafar, S. Hofmann, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. Stenczel, K. V. Svahn, C. Sutton, T. D. Swinburne, J. Tilly, C. van der Oord, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills, and G. Csányi,

"A foundation model for atomistic materials chemistry", arXiv:2401.00096, `https://arxiv.org/abs/2401.00096`.

[159] J.-D. Chai and M. Head-Gordon, "Systematic optimization of long-range corrected hybrid density functionals", The Journal of Chemical Physics **128**, 084106 (2008), `https://doi.org/10.1063/1.2834918`.