Exact constraints and appropriate norms in machine-learned exchange-correlation functionals

Cite as: J. Chem. Phys. **157**, 174106 (2022); https://doi.org/10.1063/5.0111183 Submitted: 18 July 2022 • Accepted: 10 October 2022 • Accepted Manuscript Online: 10 October 2022 • Published Online: 03 November 2022

Kanun Pokharel, 🔟 James W. Furness, Yi Yao, et al.



ARTICLES YOU MAY BE INTERESTED IN

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

Many recent density functionals are numerically ill-behaved The Journal of Chemical Physics **157**, 174114 (2022); https://doi.org/10.1063/5.0121187

Koopmans' theorem and selection rules for one-electron ionization processes in orbitally degenerate systems

The Journal of Chemical Physics 157, 174103 (2022); https://doi.org/10.1063/5.0110124





J. Chem. Phys. **157**, 174106 (2022); https://doi.org/10.1063/5.0111183 © 2022 Author(s).

ſŢ

Export Citation

View Online

Exact constraints and appropriate norms in machine-learned exchange-correlation functionals

Cite as: J. Chem. Phys. 157, 174106 (2022); doi: 10.1063/5.0111183 Submitted: 18 July 2022 • Accepted: 10 October 2022 • Published Online: 3 November 2022

Kanun Pokharel,^{1,a)} James W. Furness,^{1,b)} D Yi Yao,^{2,3} Volker Blum,² D Tom J. P. Irons,⁴ Andrew M. Teale,^{4,5} and Jianwei Sun^{1,c)}

AFFILIATIONS

¹ Department of Physics and Engineering Physics, Tulane University, New Orleans, Louisiana 70118, USA

- ²Thomas Lord Department of Mechanical Engineering and Material Science, Duke University, Durham,
- North Carolina 27708, USA
- ³University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA
- ⁴School of Chemistry, University of Nottingham, Nottingham NG7 2RD, United Kingdom
- ⁵ Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, P.O. Box 1033, Blindern, N-0315 Oslo, Norway

^{a)}kpokhare@tulane.edu

^b]furness@tulane.edu ^{c]}Author to whom correspondence should be addressed: jsun@tulane.edu

ABSTRACT

Machine learning techniques have received growing attention as an alternative strategy for developing general-purpose density functional approximations, augmenting the historically successful approach of human-designed functionals derived to obey mathematical constraints known for the exact exchange-correlation functional. More recently, efforts have been made to reconcile the two techniques, integrating machine learning and exact-constraint satisfaction. We continue this integrated approach, designing a deep neural network that exploits the exact constraint and appropriate norm philosophy to de-orbitalize the strongly constrained and appropriately normed (SCAN) functional. The deep neural network is trained to replicate the SCAN functional from only electron density and local derivative information, avoiding the use of the orbital-dependent kinetic energy density. The performance and transferability of the machine-learned functional are demonstrated for molecular and periodic systems.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0111183

I. INTRODUCTION

The density functional theory (DFT) of Hohenberg, Kohn, and Sham^{1,2} allows for efficient computation of material properties by avoiding the complicated many-electron wave function in favor of the computationally convenient electron density when solving the electronic structure problem. Due to its useful accuracy and efficiency, DFT has become the most widely used computational approach for solving electronic structure problems in chemistry and condensed matter physics.

In the Kohn-Sham (KS) formulation of DFT, the majority of the energy is calculated exactly, leaving only a small portion of the energy, known as the exchange-correlation (XC) energy, to be approximated. There has been extensive research on improving approximations to the XC energy, and the resulting functionals are roughly categorized into a hierarchy of increasing complexity and expected accuracy.³ The meta-generalized gradient approximations (mGGAs), the highest category that depends only on semi-local ingredients, are becoming increasingly popular for allowing high accuracy at a favorable computational cost. The mGGA functionals commonly consist of a per-particle XC energy density, $\epsilon_{\rm xc}$, built from three ingredients: the electron density, $n(\mathbf{r})$, its gradient, $|\nabla n(\mathbf{r})|$, and the kinetic energy density $\tau(\mathbf{r}) = 1/2\sum_{i}^{occ.} |\nabla \varphi_i(\mathbf{r})|^2$, where φ_i are the occupied KS orbitals. Although less common, density functionals depending on the density Laplacian $\nabla^2 n(\mathbf{r})$ instead of (or in addition to) τ are also included at the mGGA level. The total XC energy for the system is calculated by multiplying this perparticle XC energy density by the local electron density $n(\mathbf{r})$ and integrating over all space,

$$E_{\rm xc} = \int d\mathbf{r} \epsilon_{\rm xc}^{\rm mGGA}(n(\mathbf{r}), |\nabla n(\mathbf{r})|, \tau(\mathbf{r}), \nabla^2 n(\mathbf{r}))n(\mathbf{r}), \qquad (1)$$

$$= \int d\mathbf{r} F_{\rm xc}^{\rm mGGA}(n(\mathbf{r}), |\nabla n(\mathbf{r})|, \tau(\mathbf{r}), \nabla^2 n(\mathbf{r}))n(\mathbf{r}) \epsilon_{\rm x}^{\rm LDA}(\mathbf{r}), \quad (2)$$

where F_{xc}^{mGGA} is the XC enhancement factor, and $\epsilon_x^{LDA}(\mathbf{r}) = -(3/4\pi)(3\pi^2 n)^{1/3}$ is the exchange energy per-particle of the uniform electron gas of electron density *n*. Note the XC energy density is $e_{xc} = n\epsilon_{xc}$.

The kinetic energy density, τ , is commonly used in mGGAs to recognize different chemical environments through iso-orbital indicator variables,⁴⁻⁶ and as a component of the spherically averaged exchange hole expansion.7 Although theoretically convenient, τ introduces an implicit dependence on the KS orbitals, which brings some complications. (1) It reduces computational efficiency in molecular codes, as the contraction of basis function derivatives required for τ is more costly than that required for GGA quantities. For Fourier transform-based periodic codes, the performance cost is worse still, as τ cannot be constructed for derivatives of the density alone (as would be preferable) and, instead, derivatives of the basis functions are required. (2) It prevents the functional being used in orbital-free DFT calculations. (3) Evaluation of the XC potential for τ -dependent functionals requires either optimized effective potential (OEP) techniques^{8,9} or a generalized KS scheme.¹⁰⁻¹² Although a generalized KS treatment can be computationally convenient, the effective XC potential operator of a τ -dependent mGGA is no longer a multiplicative function, $v_{xc}(\mathbf{r})$, and is, instead, a non-local operator, \hat{v}_{xc} .

Despite the potential advantages offered avoiding the use of orbital-dependent ingredients such as τ , $\nabla^2 n(\mathbf{r})$ remains a less explored ingredient and its physical significance for the XC energy is unclear. Recently, Mejia-Rodriguez and Trickey^{13,14} replaced $\tau(\mathbf{r})$ with functions of $\nabla^2 n(\mathbf{r})$ in many mGGA XC functionals to recover similar (but not identical) performance to the parent functionals. This suggests an intriguing but unclear relationship between $\tau(\mathbf{r})$ and $\nabla^2 n(\mathbf{r})$, although an explicit relationship remains elusive despite significant effort.¹⁵

Machine learning (ML) has proven to be a powerful tool for building complicated non-linear mappings for which little theoretical guidance exists. It has proved successful in building complex models across a wide variety of fields including robotics,¹ pattern recognition,^{18,19} drug design,²⁰⁻²² and gaming.²³ Within DFT research, there has been a recent practice of applying ML to construct density functionals. Synder et al. used an ML approximation to construct an orbital-free non-interacting kinetic energy functional $T_s[n]$ for spinless fermion systems.² Brockherde et al. used ML to learn the Hohenberg-Kohn (HK) map between electron density and external potential to give a mechanism that bypasses solving the KS equations.²⁶ Several other works have focused on the XC potential problem.²⁷⁻³⁰ Specifically, interesting for this paper is the recent work of Dick and Fernandez-Serra, in which they use machine-learning techniques to self-consistently determine an mGGA similar to the Strongly Constrained and Appropriately Normed (SCAN) functional⁵ that satisfies some of the exact constraints.³¹ A perspective surveying the current state of ML in computational chemistry and materials science was recently published by Westermayr *et al.* in Ref. 32.

The SCAN functional⁵ has proven to be effective for describing a wide variety of systems,³³ such as liquid water and ice,³⁴ semiconductor materials,³⁵ and metal oxides,³⁶ as well as for key properties of correlated materials like cuprates.^{37–39} SCAN's success is credited to its adherence to all of the known exact-constraints applicable to a meta-GGA functional, along with the philosophy of using "appropriate norms" to set free parameters with minimal empiricism. These appropriate norms are provided by the systems for which a semi-local density functional approximation can be expected to be highly accurate, that is, the total energies of systems with highly localized exchange-correlation holes.⁵

In light of this dual success of both constraint-driven design and ML techniques, a question arises. Is the philosophy of exact constraints and appropriate norms compatible with ML for functional design? Growing evidence that exact constraint adherence can improve ML transferability suggests it is. Indeed, earlier works from Hollingsworth *et al.* show that enforcing coordinate scaling constraints can improve machine-learned functionals.⁴⁰ Nagai *et al.*⁴¹ recently introduced a method to analytically impose constraints on the asymptotic behavior of an ML XC functional, finding generally improved accuracy. More recently, Kirkpatrick and co-workers⁴² developed a functional DM21 (DeepMind21) that includes specific training to approximately obey two classes of constraints on systems with fractional electrons, which are fractional charge systems and fractional spin systems.⁴²

In our work, we explore this idea of exact constraints and appropriate norms satisfaction by training a deep artificial neural network (ANN) to reproduce the XC enhancement factor of the SCAN functional using $\nabla^2 n(\mathbf{r})$ instead of $\tau(\mathbf{r})$, a similar goal to the SCAN-L functional.¹⁴ The de-orbitalization of SCAN stands as a convenient application for exploring the idea of constraint satisfaction in ML functionals, with SCAN-L providing an analytical benchmark for the task.

The ML models will be trained to perform the transformation,

$$F_{xc}^{SCAN}(n(\mathbf{r}), |\nabla n(\mathbf{r})|, \tau(\mathbf{r})),$$

$$\downarrow \qquad (3)$$

$$F_{xc}^{SCAN-ML}(n(\mathbf{r}), |\nabla n(\mathbf{r})|, \nabla^2 n(\mathbf{r})).$$

We approach this mapping of orbital-free ingredients onto the SCAN XC enhancement factor by using two different ML models adhering to different numbers of exact conditions. A similar transformation could be made considering energy densities directly, although we do not do so here. While setting the energy density as the objective directly would naturally bias training toward energetically important regions, it is simpler to engineer exact constraint adherence into the enhancement factor. One model is a single totally connected ANN trained for Eq. (3) directly, termed the "combined model." The other model is built as two complementary exchangeand correlation-like ANNs designed to obey exact spin-scaling constraints, termed the "spin-scaled model." We also impose the general Lieb–Oxford bound⁴³ on these models to create models that also satisfy this constraint.

II. EXACT CONSTRAINTS

Although the exact XC functional remains unknown, it is known to obey many mathematical conditions, commonly called the "exact constraints" of XC functionals. Currently, 17 exact constraints⁵ are known to apply at the semi-local functional level. These can be broken down as conditions for the exchange energy: (1) negativity, (2) spin-scaling,⁴⁴ (3) uniform density scaling,⁴⁵ (4) the slowly varying density gradient expansion (to fourth order),⁴⁶ (5) nonuniform density scaling,^{47,48} and (6) a tight bound for two-electron densities.^{43,49} For correlation: (7) non-positivity, (8) the slowly varying density gradient expansion (to second order),⁵⁰ (9) uniform density scaling to the high-density limit,⁴⁵ (10) uniform density scaling to the low-density limit,⁴⁵ (11) zero correlation energy for any one-electron spin-polarized density, and (12) nonuniform density scaling.^{47,48} Finally, there are constraints known for the exchange and correlation together: (13) size extensivity, (14) the general Lieb–Oxford bound, 43,51,52 (15) weak dependence upon relative spin polarization in the low-density limit,^{53,54} (16) static linear response of the uniform electron gas,⁵⁵ and (17) the Lieb-Oxford bound for two-electron densities.4

A subset of the 17 constraints is amenable to an ML model. The first of these is the behavior of the exchange energy under constraint (3): uniform density scaling,

$$n_{\gamma}(\mathbf{r}) = \gamma^3 n(\gamma \mathbf{r}), \qquad (4)$$

where γ is a positive real number. The exact exchange energy is known to scale as

$$E_{\mathbf{x}}[n_{\gamma}] = \gamma E_{\mathbf{x}}[n], \qquad (5)$$

under this transformation. The effect of this condition on kernel ridge regression models was investigated by Hollingsworth *et al.* in Ref. 40 for Hooke's atom model systems, concluding that its inclusion improved the ML functional performance. This constraint can be imposed by restricting the model to only take specific dimensionless quantities as input, although we do not do so here.

Constraint (2), the spin-scaling relation for exchange energy,

$$E_{\mathrm{x}}[n_{\uparrow},n_{\downarrow}] = \frac{E_{\mathrm{x}}[2n_{\uparrow}] + E_{\mathrm{x}}[2n_{\downarrow}]}{2},\tag{6}$$

is simple to enforce for ML exchange models by requiring separate exchange and correlation models, and that the same exchange model handles each spin channel independently.

Constraint (14), the general Lieb–Oxford bound on the XC enhancement factor, states that

$$0 \le F_{\rm xc}(\mathbf{r}) \le 2.215. \tag{7}$$

These bounds can be enforced on ML models by including a post-processing step that maps the ML model output, denoted as $ANN(\mathbf{r})$, to the desired domain, e.g.,

$$F_{\rm xc}^{\rm ML}(\mathbf{r}) = \frac{2.215}{1 + \rm ANN}(\mathbf{r})^2.$$
(8)

Note that in Eq. (8), ANN (\mathbf{r}) is an auxiliary quantity within the model, and the neural network output alone is, therefore, not an appropriate XC enhancement factor by itself. A similar approach can be applied to impose constraint (6), the tight bound for the exchange enhancement factor $F_x(\mathbf{r}) \in [0, 1.174]$, if the exchange and correlation models are separated. Conveniently, such post-processing also enforces constraints (1) and (7), non-positivity, by constraining $F_{\rm xc} \ge 0$. It appears more challenging to enforce the exact constraints outside this subset in ML models. For example, while enforcing the second (and fourth)-order gradient expansions for correlation (and exchange) is relatively straightforward in analytical functionals, the ML design contains thousands of parameters, which cannot be fully controlled. Thus, it is non-trivial to enforce such gradient expansion constraints on the model a priori. Despite this, the nature of supervised training against methods that obey such constraints (such as SCAN) will result in the trained model effectively learning the constraints to some degree. However, without the rigorous enforcement described above, it is unclear how well such adherence will transfer out of the training domain.

III. INPUT DOMAIN

Identifying the input domain is a critical part of the ML model design as a model's performance can be strongly dependent upon the nature of its inputs. Since our central interest is to replace the kinetic energy density $\tau(\mathbf{r})$ dependence, we will only consider orbital-free ingredients. Four density inputs were initially identified,

$$s = \left(\frac{3}{4\pi n}\right)^{1/3} - \text{Wigner} - \text{Seitz radius},$$
 (9)

$$s = \frac{|\nabla n|}{2(3\pi^2)^{1/3}n^{4/3}} - \text{Reduced density gradient}, \tag{10}$$

$$\zeta = \frac{n_{\uparrow} - n_{\downarrow}}{n_{\uparrow} + n_{\downarrow}} - \text{Spin polarization,}$$
(11)

$$q = \frac{\nabla^2 n}{4(3\pi^2)^{2/3} n^{5/3}} - \text{Reduced density laplacian.}$$
(12)

The Weigner–Seitz radius is the radius of a sphere, which, on average, contains one electron for a uniform density *n*. The reduced density gradient introduces inhomogeneity, which measures how fast and how much the density varies on the scale of the local Fermi wavelength $2\pi/k_F$, where $k_F = (3\pi^2 n)^{1/3}$ is the Fermi wavevector. The reduced density Laplacian also measures the density inhomogeneity and can distinguish bonds, in contrast to the reduced gradient, which vanishes at the middle of the bond.¹⁵ The above dimensionless ingredients are preferred for XC functionals rather than using the density variables directly as the correct uniform coordinate density-scaling behavior can be satisfied with them in conventional XC functionals.⁴⁵ We note that the ML exchange models, here, break the formal scaling limits, in practice, as we supply density information to the exchange model through r_s , thereby violating the exchange coordinate-scaling condition.

Further exploration revealed that including additional ingredients directly from the SCAN exchange and correlation functionals (see supplemental material of Ref. 5) could improve the model performance,

J. Chem. Phys. **157**, 174106 (2022); doi: 10.1063/5.0111183 Published under an exclusive license by AIP Publishing

$$\epsilon_{\rm c}^0(r_s, s, \zeta)$$
 – Single orbital correlation, (13)

 $\epsilon_{\rm c}^{\rm l}(r_s, s, \zeta)$ – Slowly varying correlation, (14)

$$g_x(s)$$
 – Exchange inhomogeneity, (15)

$$h_0 = 1.174 - \text{Single orbital exchange},$$
 (16)

$$h_1 = 1.065 - \frac{0.065}{\left(1 + \frac{10s^2/81}{0.065}\right)} - X$$
 second order gradient expansion.

(17)

These additional inputs are combinations of the original density ingredients and do not provide any new information directly. However, their inclusion makes learning more efficient, as it reduces the manipulations that the network must learn. This limits dependence on the network to only learning τ -dependent aspects, rather than requiring it to learn every detail of the SCAN functional.

The possible range of the input parameters is very different from the desired range of the model outputs: $0 \le F_{xc}^{ML} \le 2.215$. For example, the domain of r_s and s is $[0, +\infty)$, while q is $(-\infty, \infty)$. Such a mismatch in the magnitude of input and output is known to be challenging for ML models. To correct for this, we pre-processed the unbounded inputs using the hyperbolic tangent function, tanh(x),⁵⁶ to smoothly map the unbounded quantities to (-1, +1). With pre-processing, the inputs are defined as

$$\tilde{r}_s = \tanh(r_s),\tag{18}$$

$$\tilde{s} = \tanh(s),$$
 (19)

$$\tilde{q} = \tanh(q). \tag{20}$$

Finally, we pre-process ζ as

$$\tilde{\zeta} = \frac{1}{2} \Big[(1+\zeta)^{4/3} + (1-\zeta)^{4/3} \Big], \tag{21}$$

to ensure the model is a symmetric function of spin polarization, following the spin-scaling of the exchange energy^{44,57} of the uniform electron gas.² The inputs of Eqs. (13)–(17) are unprocessed as their ranges are already correctly bounded. Note that the pre-processed variables [Eqs. (18)–(21)] are only supplied to the network and are not used to generate the additional inputs of Eqs. (13)–(17).

Having identified the input domain, a training dataset consisting of 20 atoms was generated by using accurate spherical Hartree–Fock Slater-type orbitals^{58–60} containing open- and closed-shell atoms (He, Li, Be, B, C, N, O, F, Ne, Na, P, Cl, Ar, K, Cr, Cu, Cu⁺, As, Kr, Xe) with *s*, *p*, and *d* valence shells. Spherical Hartree–Fock Slater-type orbitals were chosen to match the data used to fit parameters in SCAN, although we expand the set of atoms significantly beyond the rare gas atoms used to fit SCAN. The energetically important region of the atomic density is typically between 0 and 4 *a*₀ or bohr, and the density of each atom was, therefore, sampled in shells of decreasing sample density for models, with 1300 radial points uniformly sampling in *r* < 1 bohr (core), 800 in $1 \le r < 4$ bohr and 500 in the tail region $4 \le r < 10$ bohr.

These atomic training data were augmented with densities from the "appropriate norm" systems used in SCAN's construction. The first norm is the one electron hydrogen atom, which is used to ensure that SCAN is one-electron self-correlation-free (constraint 11). The second and third norms are the jellium surface densities for $r_s = 2, 3, 4, and 6,^{61,62}$ and the converged SCAN orbitals of the compressed argon dimer with nuclear separations of 1.6, 1.8, and 2.0 Å, which were used to fix SCAN's interpolation function parameters.⁵ We restrict the training dataset to only these appropriate norms and avoid including chemically bonded systems, such that any resulting chemical bonding shown by the trained model is a prediction arising from underlying transferable physics rather than simple fitting. While increasing the domain of the training data is expected to give the resulting ML models higher accuracy for a wider range of problems, our intent, here, is to observe how exact constraint satisfaction can transfer knowledge from minimal training data onto diverse problems.

IV. NEURAL NETWORK ARCHITECTURE AND TRAINING

In this work, all the networks were built on the basis of the ML framework Tensorflow.⁶³ For the models, the non-linearity in the mapping is acquired by using sigmoid,⁶⁴ tanh, and exponential linear unit (elu)⁶⁵ activation functions, chosen as commonly used continuously differentiable activation functions. Ensuring smooth activation functions was found to be essential for obtaining reasonable XC potentials, as discussed below. The dataset was randomly divided into a training (80%) set and a validation (20%) set, using the train-test split feature of sci-kit learn.⁶⁶ Hyper-parameter searching identified a three-layered model with sigmoid to sigmoid to tanh activation functions as preferable for the combined model, while a two-layered model with tanh to elu activation functions was preferred for the spin-scaled model. The network weights and biases were optimized by stochastic gradient descent with the Adam optimizer⁶⁷ using a learning rate of 0.005. The training dataset is comprised of samples of $F_{\rm xc}(r)$ from the DFT quadrature grid. A gradient step is applied after each sample for a stochastic gradient descent. The optimized model was chosen as that which minimizes the error for the validation set, generally found after one complete pass of the training data.

A. Combined model

Figure 1(a) presents a schematic for the simple neural network (NN) architecture, termed the "combined model." As the name suggests, the combined model receives inputs constructed from total density $(n_{\uparrow} + n_{\downarrow})$ as the features and targets SCAN's exchange-correlation enhancement factor. Different numbers of hidden layers and neuron counts were tested, with a three-layered model with 100, 50, and 20 neurons in the respective layers found to perform best.

For the combined model, the loss function to be optimized in the learning process is defined as

$$\mathcal{L}_{\text{combined}} = \frac{1}{N} \sum_{i}^{N} \left(F_{\text{xc}}^{\text{ML}} - F_{\text{xc}}^{\text{SCAN}} \right)^{2}, \qquad (22)$$

where N is the number of training data points, thus minimizing the mean square difference between the SCAN XC enhancement factor and the learned XC enhancement factor. The Lieb–Oxford bound for the combined model is introduced as a post-processing mechanism following the explanation in Eq. (8).



FIG. 1. ML model architecture and workflow for (a) combined and (b) spin-scaled models. For the combined model, total density $(n_{\uparrow} + n_{\downarrow})$ is given as the input. In the spin-scaled model, the upper architecture is for the exchange model learning, while the lower is the correlation-like model. Spin-scaling is satisfied in the exchange model represented by $(2n_{\sigma})$, where $\sigma = \uparrow \downarrow$ spin channels.

B. Spin-scaled model

The spin-scaled model follows a more complex architecture that allows it to obey the spin-scaling exact constraint by treating exchange and correlation separately, as discussed above. The overall architecture for the spin-scaled model is presented in Fig. 1(b). The spin-scaled model is comprised of two separate networks, one for exchange and the other for correlation. These networks are trained separately and later combined to form the complete model. Training is, therefore, carried out as a two-step process.

This separation by spin channel reduces the input domain for the exchange network to six features for exchange, $\{\tilde{r}_{s\sigma}, \tilde{s}_{\sigma}, \tilde{q}_{\sigma}, g_{x\sigma}, h_{0\sigma}, h_{1\sigma}\}$, separately generated for each spin σ . The exchange network is trained first to minimize the mean square difference with the SCAN exchange enhancement factor defined as

$$\mathcal{L}_{\text{exchange}} = \frac{1}{N} \sum_{i}^{N} \left(F_{x}^{\text{ML}} - F_{x}^{\text{SCAN}} \right)^{2}, \qquad (23)$$

with the spin-scaled exchange enhancement factor,

$$F_{\rm x}^{\rm ML} = \frac{F_{\rm x}(2n_{\uparrow})e_{\rm x}^{\rm LDA}(2n_{\uparrow}) + F_{\rm x}(2n_{\downarrow})e_{\rm x}^{\rm LDA}(2n_{\downarrow})}{2e_{\rm x}^{\rm LDA}(n_{\uparrow}+n_{\downarrow})}.$$
 (24)

As the exchange energy must be invariant to the permutation of spin labels, the same exchange network is used for both spin channels and should be trained on both spin channels of the training data. The exchange network has two layers with 80 and 40 neurons at the first and second layers, respectively, and the activation functions are tanh and exponential linear unit (elu).⁶⁵

The correlation energy is not subject to the same spin-scaling constraint and is handled by a separate model. This second model takes a reduced set of the total density $(n_{\uparrow} + n_{\downarrow})$ input variables suitable for correlation: { $\tilde{r}_{s}, \tilde{s}, \tilde{q}, \zeta^{2}, \epsilon_{o}^{0}, \epsilon_{o}^{1}$ }. This second network has the same hyperparameters as the exchange network.

TABLE I. Deviation of the spin-scaled exchange network total exchange energy (hartrees) from the SCAN exchange functional against which it was trained. Calculated for the set of spherical atom densities and the molecules of the G3 test set. Mean error (ME) and Mean absolute error (MAE) are presented, and full data are given in supplementary material. Here, "LO" abbreviates "Lieb–Oxford."

		No LO bound	LO bound
Atoms	ME MAE	-0.147 0.153	-0.541 0.543
G3	ME MAE	-0.209 0.209	$\begin{array}{c} -0.448\\ 0.448\end{array}$

The loss function for the second network is

$$\mathcal{L}_{\text{correlation}} = \frac{1}{N} \sum_{i}^{N} \left[F_{\text{c}}^{\text{ML}} - \left(F_{\text{xc}}^{\text{SCAN}} - F_{\text{x}}^{\text{ML}} \right) \right]^{2}, \quad (25)$$

where F_x^{ML} is the output of the previously trained exchange network. This second network is, therefore, not a true model of SCAN correlation as the loss function of Eq. (25) drives it to compensate for deficiencies in the exchange network, although correlation effects will dominate if the exchange network is accurate. Table I shows the deviation of the LO bounded and un-bounded exchange networks (to be discussed below) in total exchange energy from SCAN for the spherical atoms, which were part of the training and test sets, and the molecules of the G3 set, which were not in the training set. From this, we see that deviation is typically <1% for both networks, although the bounded model deviates more significantly than the unbounded, indicating that good exchange models have been learned.

Finally, the total enhancement factor is obtained by summing the exchange network and second network enhancement factors as

$$F_{\rm xc}^{\rm spin-\, scaled} = F_{\rm x}^{\rm ML} + F_{\rm c}^{\rm ML}.$$
 (26)

The Lieb–Oxford bounds for the spin-scaled model are introduced separately for exchange and the correlation parts as they are trained separately. For the exchange part, we follow similar mechanism as explained in Eq. (8), where the bound is introduced to the exchange enhancement factor as

$$F_{\rm x}^{\rm ML-bound}(\mathbf{r}) = \frac{1.174}{1 + \rm ANN(\mathbf{r})^2}.$$
 (27)

Here, we choose a tight bound of 1.174 for the exchange.

For correlation, we introduce the bound by modifying the loss function to include an additional penalty term,

$$\mathcal{L}_{\text{correlation}}^{\text{bound}} = \frac{1}{N} \sum_{i}^{N} \left\{ \left[F_{\text{c}}^{\text{ML}} - \left(F_{\text{xc}}^{\text{SCAN}} - F_{\text{x}}^{\text{ML}} \right) \right]^{2} + \mu \times \text{relu} \left[F_{\text{xc}}^{\text{ML}} - 2.215 \right] \right\},$$
(28)

where μ value is chosen to be 20, and relu is the rectified linear unit.⁶⁸ Here, we see that if the total XC enhancement factor is smaller than the tight bound 2.215, the penalty term is zero, whereas any value of the total XC enhancement factor greater than 2.215 will incur a penalty. This total loss function is minimized to satisfy the Lieb–Oxford bound introduced in the ML model. It is possible that the loss functions used here may be prone to over-fitting. Such error could be mitigated by including additional penalty terms on the smoothness of the model, or including dropout layers into the network during training.

V. RESULTS AND DISCUSSION

A. Atomic performance

Figure 2 shows the SCAN and ML-model XC enhancement for the silicon atom, which was not part of the training set. Figures 2(a) and 2(b) show combined and spin-scaled models, respectively,



FIG. 2. XC enhancement factor plots for the test silicon atom, which was not included in the training set for (a) combined model and (b) spin-scaled model without Lieb–Oxford bound, while (c) and (d) represent the same as (a) and (b) with Lieb–Oxford bound. All models are compared against SCAN functional. Density was obtained from accurate spherical HF orbitals.^{58–60}

without the Lieb–Oxford bound, while Figs. 2(c) and 2(d) include the Lieb–Oxford bound constraint. All the ML models show good agreement with the SCAN's XC enhancement factor for this system. The combined model without the Lieb–Oxford bound shows slightly less variation in the energetically important region between r = 0 and r = 4 bohr.

During training, we did not target the total XC energy directly, in favor of learning the XC enhancement factor of an existing mGGA instead. This switch toward learning the XC enhancement factor, a local property, has three benefits. Most importantly, it reduces the complexity of the training by avoiding summing the derivatives of many training points in a numerical integration batch. Second, far more training, data are available for a given functional's local XC enhancement factor than for total XC energies, as every point in the all-space integration of any system can now be considered a training point. We note that all training points were weighted equally. As some points contribute more to the total energy than others learning the XC energy density directly would implicitly bias training toward energetically important regions, such as core densities, and away from less important regions, such as asymptotic densities. Third, this avoids the introduction of a gauge freedom in which many different energy density functions can integrate to the same total energy, which could result in learning a model that gives reasonable total energy, but poor local accuracy to the SCAN energy density. Despite this training against the local XC enhancement factor, it will be shown that the models are successful in recovering the total XC energy for the training atom sets, predicting the global property from local training.

B. Molecular test sets

For the models trained against data from atomic systems, a real challenge is to generalize to problems outside the training domain. We examine this by looking at model accuracy for the open- and closed-shell molecules of the G3 test set.⁶⁹ The input ingredients (density, and its gradient and Laplacian) and the SCAN $F_{\rm xc}^{\rm SCAN}$ target for all molecular calculations were generated from self-consistent SCAN orbitals in the 6-311++G(3df,3pd) basis set.^{70,71} All molecular calculations were carried out using the QUantum

TABLE II. Mean absolute difference (MAD) in kcal mol⁻¹ for G3 set of 226 molecular atomization energies.⁶⁹ All SCAN G3 calculations were performed fully self-consistently with the 6-311++G(3df,3pd) basis set^{70,71} in the QUEST program.⁷² The ML calculations were performed non-self-consistently from SCAN orbitals. The ML models with Lieb–Oxford bound are denoted by "LO" within the table. Data are reported as the mean absolute difference between the total energy calculated using the ML model or SCAN-L and that calculated using the parent SCAN functional.

Difference to SCAN						
ML models	Closed shell (MAD)	Open shell (MAD)	(G3) (MAD)			
Combined	7.49	7.89	7.69			
Combined-LO	11.88	9.26	10.58			
Spin-scaled	8.03	4.85	6.44			
Spin-scaled-LO	7.66	5.66	6.66			
ŚĊAN-L	10.74	4.77	7.75			

TABLE III. Mean absolute error (MAE) in kcal mol⁻¹ for G3 set of 226 molecular atomization energies.⁶⁹ Errors are given relative to standard reference values for the G3 set.

Reference comparison					
ML models	Closed shell (MAE)	Open shell (MAE)	(G3) (MAE)		
Combined	7.80	5.17	6.48		
Combined-LO	13.24	7.26	10.25		
Spin-scaled	8.55	3.73	6.14		
Spin-scaled-LO	11.14	6.35	8.75		
ŚĊAN	6.53	6.23	6.49		
SCAN-L	7.26	4.55	5.91		

Electronic Structure Techniques (QUEST) program.⁷² A dense numerical integration grid was used to ensure SCAN calculations were properly converged, defined in QUEST⁷² as a Lebedev angular grid of degree 41^{73} and an LMG radial grid with an accuracy threshold of 10^{-15} .⁷⁴

The mean absolute difference or error in atomization energy for the G3 set is summarized in Tables II and III for open-shell, closed-shell, and total collections. Table II shows errors relative to atomization energies calculated by using the parent SCAN functional to illustrate model fidelity. Table III shows errors relative to standard reference total energies for the set to illustrate the models' chemical accuracy. Overall, all models showed comparable performance giving accuracy close to the SCAN functional (MAE 6.53 kcal mol⁻¹). In particular, enforcing the correct spin-scaling relation in the ML models improved accuracy for the open-shell systems, although this was accompanied by a small reduction in accuracy for the closed-shell systems. Introduction of the Lieb–Oxford bound to the combined model deteriorated performance for both open- and closed-shell systems. This unexpected poor performance suggests that enforcing the Lieb–Oxford bound through Eq. (8) does not seem to be a successful strategy.

It is possible that the reduced performance found is a result of locally imposing the Lieb–Oxford bound on the model. As the Lieb–Oxford bound is a constraint defined for the total XC energy, imposing it as a constraint on the integrated energy may be more successful; however, this would require a significantly different training scheme. We found that, in practice, the Lieb–Oxford bound was obeyed by the unbounded models for all training and test examples as a result of having been trained on Lieb–Oxford bound obeying examples. It is unclear how general this adherence is across general densities, however, in contrast to the bounded models for which the Lieb–Oxford bound is obeyed by construction.

Although such post-processing is theoretically convenient, the results from the G3 set and lattice constant below show that it significantly limited model learning during training. We understand this as an effect of the non-linear normalization of Eqs. (8) and (27). These two equations present challenging task of requiring the ANN to learn an unbounded function and requiring the ANN to be very large when the target $F_{\rm xc}(F_{\rm x})$ is small. The spin-scaled model performs better than the combined model because the number to return (1.174) is relatively smaller compared to 2.215 for the combined model and, hence, the error is smaller. An alternative solution for enforcing this constraint could be to simply truncate the range of the network output,

$$F_{\rm xc}^{\rm ML} = \max(0, \min(2.215, \text{ANN}(r))),$$
 (29)

and

$$F_{\rm x}^{\rm ML} = \max(0, \min(1.174, \text{ANN}(r))),$$
 (30)



FIG. 3. Exchange-correlation enhancement factor plotted along the bond axis for O₂ (open-shell) and CO (closed-shell) from (a) combined model and (b) spin-scaled model without Lieb–Oxford bound. (c) and (d) are the same as (a) and (b) but for models including Lieb–Oxford bound. The gray vertical lines mark atomic positions.

however, this may introduce undesirable discontinuities in the partial derivatives of the model, transferring into a non-physical XC potential.

To better understand how faithfully the ML models are reproducing the SCAN functional, Fig. 3 compares the XC enhancement factor, F_{xc}, for the open-shell O₂ and closed-shell CO molecules. We see that all the models are accurate in the areas immediately around the nuclei. This is expected from the good atomic performance as these core regions are relatively unchanged by the covalent bonding. The models deviate more severely around the bond center, with all models underestimating F_{xc}^{SCAN} . In these regions, $|\nabla n| \to 0$ and hence $s \rightarrow 0$, while the density *n* remains significant. Such regions are under-represented in the training set, appearing only in small regions at the center of the compressed Ar₂ diatomic. This suggests that the spin-scaling relation enforced in the spin-scaled model as well as the Lieb-Oxford bound are insufficient to transfer learning from a training set that does not include chemical bonding, onto systems that are chemically bound. Further constraint satisfaction, or inclusion of bonding data into the training set, is likely necessary to improve model accuracy at these important points. We also note that the combined model with no Lieb-Oxford bound exhibits sharp spikes in $F_{\rm xc}$ in the bonding regions that are not seen for other ML models or the SCAN functional.

C. Lattice constants of solids

The transferability of the ML functionals was further tested by calculating the lattice constants of 20 solids from the LC20 test set.⁷⁵ This tests the model's ability to generalize into further unseen chemical environments in periodic systems, as well as requiring description of the energy as a function of nuclear displacement. The LC20 set is, therefore, a sensitive test of model transferability. The equilibrium lattice constants were determined by a nine-point fit of total energy per unit cell to the Vinet equation of state around the SCAN equilibrium unit cell volume (V_0) in a range of $V_0 \pm 10\%$.^{77,78}

Figure 4 compares LC20 results from ML models with the SCAN and SCAN-L functionals. The SCAN results for lattice constants are taken from Ref. 5 and SCAN-L results from Ref. 14. The ML calculations were obtained non-self-consistently using Perdew–Burke–Ernzerhof (PBE) densities, not SCAN densities as was used for molecular calculations. This is because SCAN suffered numerical instabilities in calculations using FHI-aims,⁷⁶ with some systems failing to converge. FHI-aims was used for periodic calculations rather than VASP (which was used for Refs. 5 and 14) as all-electron calculations. We should, therefore, be aware of a possible density-driven error for the ML models resulting from the PBE densities on which they were evaluated, which could be significant.

The results of Fig. 4 show that the ML models do not perform as well for solids as they do for gas phase atoms and molecules, suggesting difficulty in generalizing knowledge into this untrained domain. The spin-scaled model performed significantly better than the combined model both with and without enforcing the Lieb–Oxford bound. Although the performance of ML models is poor compared to SCAN and SCAN-L, they improve slightly on the PBE GGA⁵¹ (MAE 0.060 Å). Having said that, it suggests a similar conclusion to the molecular tests; that the spin-scaling and the Lieb–Oxford



FIG. 4. Mean Absolute Error in predicted lattice constants (MAE, Å) for SCAN, whose results are obtained from Ref. 5, SCAN-L, whose results are obtained from Ref. 14, combined, combined with Lieb–Oxford bound (denoted "LO" within figure), spin-scaled and spin-scaled with Lieb–Oxford bound for the LC20 set of 20 solids with lattice constants ranging from 3.415 to 6.042 Å.⁷⁵ The ML calculations were performed using fixed self-consistent PBE densities generated using FHI-aims.⁷⁶

bound are insufficient to transfer learning into the periodic systems. The choice to model the full XC enhancement factor may also contribute to this poor transferability, rather than only training the orbital dependent parts as in Ref. 13. Although some extended character is present in the jellium surfaces, the training data are predominantly isolated systems and, hence, extended systems are under-represented. Inclusion of such information in the training set, or adjusting the model to target τ directly, is likely necessary to improve the model accuracy.

D. Exchange-correlation potential

The XC potential is defined as

ı

$$\Phi_{\rm xc}(\mathbf{r}) = \frac{\delta E_{\rm xc}}{\delta n(\mathbf{r})}.$$
(31)

This is constructed for the ML models from the partial derivatives of the model with respect to its ingredients,

$$\left\{\frac{\partial \epsilon_{\rm xc}^{\rm ML}}{\partial n(\mathbf{r})}, \frac{\partial \epsilon_{\rm xc}^{\rm ML}}{\partial |\nabla n(\mathbf{r})|}, \frac{\partial \epsilon_{\rm xc}^{\rm ML}}{\partial \nabla^2 n(\mathbf{r})}\right\},\tag{32}$$

through repeated application of the chain rule. In practice, this is achieved by using the back-propagation mechanism of the machine learning framework.⁷⁹ Here, the choice of the activation functions for the neuron layers is critically important for obtaining a smooth XC potential appropriate for well-converged self-consistent field (SCF) calculations. When the activation functions are differentiated during back-propagation, the use of activation functions with discontinuous derivatives, such as the popular rectified linear unit (relu),⁸⁰ may introduce discontinuities into the XC potential that can harm SCF convergence and computational efficiency. We, therefore, used only smooth sigmoid, tanh, and elu activation functions within the present models.

Figure 5 shows partial derivatives of the combined and spinscaled models with respect to input ingredients, compared against the equivalent for the SCAN functional where they exist. Note



FIG. 5. Derivatives of SCAN and ML model XC energy density (combined and spin-scaled) with respect to input ingredients for the silicon atom (not in the training set). ML derivatives are represented by solid lines, while SCAN is dotted. Derivative components for (a)–(d) combined model without LO bound, (e)–(h) spin-scaled model without LO bound, (i)–(l) combined model with LO bound, and (m)–(p) Spin-scaled model with LO bound. Columns (a), (e), (i), and (m) show derivative with respect to density, columns (b), (f), (j), and (n) show derivative with respect to same-spin square gradient, columns (c), (g), (k), and (o) show derivative with respect to opposite-spin square gradient, and columns (d), (h), (l), and (p) show derivative with respect to density Laplacian.

that $\epsilon_{\rm xc}^{\rm SCAN}(n, |\nabla n|, \tau)$ and $\epsilon_{\rm xc}^{\rm ML}(n, |\nabla n|, \nabla^2 n)$ are necessarily different functions, even though $\epsilon_{\rm xc}^{\rm ML}$ has been trained to reproduce $\epsilon_{\rm xc}^{\rm SCAN}$. Hence, we should not expect their partial derivatives in *n* and $|\nabla n|$ to match. Figures 5(a)–5(d) show the combined model partial derivatives for the test silicon atom, while Figs. 5(e)–5(h) show the same for the spin-scaled model.

Figures 5(a), 5(e), 5(i), and 5(m) show that the density partial derivative is comparable to SCAN for all models. All models exhibit oscillations in this derivative, although these are less severe for the spin-scaled models than the combined models. Figures 5(b) and 5(c) and Figs. 5(j) and 5(k) for the combined model without and with Lieb–Oxford bound constraint and Figs. 5(f) and 5(g) and Figs. 5(n) and 5(o) for the spin-scaled model, again with and without Lieb–Oxford bound, show the partial derivatives with respect to the same spin and cross-spin gradient components. Finally, Figs. 5(d), 5(h), 5(i), and 5(k) show the Laplacian partial derivatives.

It is not clear to what degree the oscillations of the present partial derivatives may affect the SCF performance of the ML models without performing such calculations, which is beyond the scope of this paper. We can reasonably expect the smoother spin-scaled model to outperform the combined model in this regard, however, as a result of the reduced oscillations seen in Fig. 5. The harmful impact of such oscillations may be significantly damped if the regions they occur in are not energetically important and the curvature minimization method of Cancio, Wagner, and Wood⁸¹ could be explored for controlling the potentials of Laplacian-dependent ML functionals. We note that the SCAN functional is known to have problematic oscillations in its XC potential, which reduce its computational efficiency but generally do not prevent SCF convergence.^{82–86} The similarity between the ML and SCAN partial derivatives of Fig. 5, therefore, suggests that self-consistency can likely be reached, although this may be sensitive to the choice of starting guess orbitals if they are different from the converged orbitals used in training.

VI. DISCUSSION

A broad goal of this work has been to explore how exact constraint adherence can enhance the transferability of ML density functionals beyond a limited set of training data for which their accuracy is theoretically well justified. For the semi-local meta-GGA functionals, these are systems for which the XC hole remains well localized (atoms and compressed non-bonding diatoms), or for which the long-range parts of the X and C holes cancel (jellium).^{87,88} It is unclear to what extent our models' overall fidelity to SCAN is a limitation of this restricted training data, and to what extent it reflects a fundamental difficulty in representing SCAN from orbital-free ingredients. The success of the SCAN-L and related functionals¹³ suggests that the former, although repeating the present study with a significantly expanded training set that covers more of the chemical space, could provide compelling evidence one way or the other; however, it is beyond the present scope.

VII. CONCLUSION

We have explored how a philosophy of exact constraints and appropriate norms can be combined with ML techniques in functional design. We have shown a simple test of this idea as a de-orbitalization of the SCAN functional, using the density Laplacian, $\nabla^2 n(\mathbf{r})$, in place of kinetic energy density $\tau(\mathbf{r})$. Four ML functionals were developed, enforcing a spin-scaling constraint, the Lieb–Oxford bound, both, and neither. These models were trained using a dataset with no chemical bonding, preferring the norms appropriate for semi-local functionals.⁵ The model satisfying both the spin-scaling constraint and the Lieb–Oxford bound generally achieved a more balanced performance across the properties tested, although performance was worse than that achieved by the analytical SCAN-L de-orbitalization for solids. Given that the model performance was generally improved when both constraints were imposed, it is reasonable to believe that engineering in further constraints can enhance the robustness of the ML models out of the training domain.

SUPPLEMENTARY MATERIAL

Supplementary material contains: (1) Individual atomization energy data for the AE6 test set, (2) Individual reaction barrier height data for the BH6 test set, (3) Complementary plots for Fig. 3 showing angle-averaged XC energy density, (4) Individual data for the error in atomic and G3 exchange energies from the exchange component of the spin-scaled model, and (5) Individual data for the error in the G3 test set.

ACKNOWLEDGMENTS

J.W.F. thanks Michael Withnall for his useful discussion and guidance on ML techniques. J.W.F., K.P., and J.S. acknowledge discussions with John Perdew on semi-local XC functionals. J.S. thanks Kieron Burke for his useful discussion on ML techniques. K.P., J.W.F., and J.S. acknowledge the support by the U.S. DOE, Office of Science, Basic Energy Sciences (Grant No. DE-SC0019350). This work was partially supported by the Research Council of Norway through its Centres of Excellence scheme, Project No. 262695. We acknowledge financial support from the European Research Council under H2020/ERC Consolidator Grant top DFT (Grant No. 772259).

AUTHOR DECLARATIONS

Conflict of Interest

The authors declare the following competing financial interest(s): V.B. is a member of the executive board of MS1P e.V., the non-profit which licenses the FHI-aims electronic structure code used to create some of the data reported in this work. V.B. does not receive any financial gains from this position. All other authors have no conflicts to disclose.

Author Contributions

J.S. and J.W.F. designed the project. K.P., J.W.F., Y.Y., V.B., T.J.P.I., and A.M.T. developed the computational methods, and K.P. performed all calculations. K.P., J.W.F., and J.S. analyzed the data and wrote the manuscript.

Kanun Pokharel: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). James W. Furness: Conceptualization (equal); Formal analysis

ARTICLE

(equal); Methodology (equal); Project administration (equal); Software (equal); Supervision (equal); Validation (equal); Writing original draft (equal); Writing - review & editing (equal). Yi Yao: Data curation (supporting); Investigation (supporting); Software (supporting); Writing - review & editing (equal). Volker Blum: Data curation (supporting); Investigation (supporting); Software (supporting); Writing - review & editing (equal). Tom J. P. Irons: Investigation (supporting); Resources (supporting); Software (supporting); Writing - review & editing (equal). Andrew M. Teale: Investigation (supporting); Resources (supporting); Software (supporting); Writing - review & editing (equal). Jianwei Sun: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Writing - review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. Parameters for the ML models presented are provided as supplementary material Tensorflow metagraph parameter files.

REFERENCES

¹P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev. **136**(3B), B864 (1964).

²W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. **140**(4A), A1133 (1965).

³J. P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy," AIP Conf. Proc. 577, 1–20 (2001).

⁴J. Sun, B. Xiao, Y. Fang, R. Haunschild, P. Hao, A. Ruzsinszky, G. I. Csonka, G. E. Scuseria, and J. P. Perdew, "Density functionals that recognize covalent, metallic, and weak bonds," Phys. Rev. Lett. **111**(10), 106401 (2013).

⁵J. Sun, A. Ruzsinszky, and J. P. Perdew, "Strongly constrained and appropriately normed semilocal density functional," Phys. Rev. Lett. **115**(3), 036402 (2015).

⁶J. W. Furness and J. Sun, "Enhancing the efficiency of density functionals with an improved iso-orbital indicator," Phys. Rev. B **99**(4), 041119(R) (2019).

⁷A. D. Becke, "A new inhomogeneity parameter in density-functional theory," J. Chem. Phys. **109**(6), 2092–2098 (1998).

⁸A. Görling and M. Levy, "Exact Kohn-Sham scheme based on perturbation theory," Phys. Rev. A **50**(1), 196 (1994).

⁹Q. Wu and W. Yang, "Algebraic equation and iterative optimization for the optimized effective potential in density functional theory," J. Theor. Comput. Chem. 2(4), 627–638 (2003).

¹⁰ 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(7), 3764–3774 (1996).
 ¹¹ R. Neumann, R. H. Nobes, and N. C. Handy, "Exchange functionals and potentials," Mol. Phys. 87(1), 1–36 (1996).

¹²Q. Wu and W. Yang, "A direct optimization method for calculating density functionals and exchange-correlation potentials from electron densities," J. Chem. Phys. **118**(6), 2498–2509 (2003).

¹³D. Mejia-Rodriguez and S. Trickey, "Deorbitalization strategies for metageneralized-gradient-approximation exchange-correlation functionals," Phys. Rev. A 96(5), 052512 (2017).

¹⁴D. Mejia-Rodriguez and S. B. Trickey, "Deorbitalized meta-GGA exchangecorrelation functionals in solids," Phys. Rev. B 98(11), 115161 (2018). ¹⁵F. Della Sala, E. Fabiano, and L. A. Constantin, "Kinetic-energy-density dependent semilocal exchange-correlation functionals," Int. J. Quantum Chem. 116(22), 1641–1694 (2016).

¹⁶M. Dorigo and U. Schnepf, "Genetics-based machine learning and behaviorbased robotics: A new synthesis," IEEE Trans. Syst., Man, Cybern. 23(1), 141–154 (1993).

¹⁷ M. A. Salichs, R. Barber, A. M. Khamis, M. Malfaz, J. F. Gorostiza, R. Pacheco, R. Rivas, A. Corrales, E. Delgado, and D. Garcia, "Maggie: A robotic platform for human-robot social interaction," in 2006 IEEE Conference on Robotics, Automation and Mechatronics (IEEE, 2006), pp. 1–7.

¹⁸N. MacLeod, M. Benfield, and P. Culverhouse, "Time to automate identification," Nature 467(7312), 154–155 (2010).

¹⁹L. Chittka and A. Dyer, "Your face looks familiar," Nature **481**(7380), 154–155 (2012).

²⁰E. A. Ashley, "The precision medicine initiative: A new national effort," JAMA, J. Am. Med. Assoc. **313**(21), 2119–2120 (2015).

²¹ M. C. Schatz and B. Langmead, "The DNA data deluge," IEEE Spectrum 50(7), 28–33 (2013).

²²H. Ding, I. Takigawa, H. Mamitsuka, and S. Zhu, "Similarity-based machine learning methods for predicting drug-target interactions: A brief review," Briefings Bioinf. 15(5), 734–747 (2014).

²³D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of Go with deep neural networks and tree search," Nature **529**(7587), 484–489 (2016).

²⁴ J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, "Finding density functionals with machine learning," Phys. Rev. Lett. **108**(25), 253002 (2012).

²⁵J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K.-R. Müller, and K. Burke, "Orbital-free bond breaking via machine learning," J. Chem. Phys. 139(22), 224104 (2013).

²⁶F. 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(1), 872 (2017).

²⁷Q. Liu, J. Wang, P. Du, L. Hu, X. Zheng, and G. Chen, "Improving the performance of long-range-corrected exchange-correlation functional with an embedded neural network," J. Phys. Chem. A **121**(38), 7273–7281 (2017).

²⁸ 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**(24), 241737 (2018).

²⁹J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, and K. W. Jacobsen, "Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation," Phys. Rev. B **85**(23), 235149 (2012).

³⁰L. Li, T. E. Baker, S. R. White, K. Burke *et al.*, "Pure density functional for strong correlation and the thermodynamic limit from machine learning," Phys. Rev. B 94(24), 245129 (2016).

³¹S. Dick and M. Fernandez-Serra, "Highly accurate and constrained density functional obtained with differentiable programming," Phys. Rev. B **104**(16), L161109 (2021).

³²J. Westermayr, M. Gastegger, K. T. Schütt, and R. J. Maurer, "Perspective on integrating machine learning into computational chemistry and materials science," J. Chem. Phys. **154**, 230903 (2021).

³³J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu *et al.*, "Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional," Nat. Chem. 8(9), 831 (2016).

³⁴M. Chen, H.-Y. Ko, R. C. Remsing, M. F. Calegari Andrade, B. Santra, Z. Sun, A. Selloni, R. Car, M. L. Klein, J. P. Perdew, and X. Wu, "Ab initio theory and modeling of water," Proc. Natl. Acad. Sci. U. S. A. **114**(41), 10846–10851 (2017).

³⁵R. C. Remsing, M. L. Klein, and J. Sun, "Dependence of the structure and dynamics of liquid silicon on the choice of density functional approximation," Phys. Rev. B **96**(2), 024203 (2017).

³⁶G. S. Gautam and E. A. Carter, "Evaluating transition metal oxides within DFT-SCAN and SCAN+U frameworks for solar thermochemical applications," Phys. Rev. Mater. 2(9), 095401 (2018).

³⁷J. W. Furness, Y. Zhang, C. Lane, I. G. Buda, B. Barbiellini, R. S. Markiewicz, A. Bansil, and J. Sun, "An accurate first-principles treatment of doping-dependent electronic structure of high-temperature cuprate superconductors," Commun. Phys. 1(1), 11 (2018).

³⁸C. Lane, J. W. Furness, I. G. Buda, Y. Zhang, R. S. Markiewicz, B. Barbiellini, J. Sun, and A. Bansil, "Antiferromagnetic ground state of La₂CuO₄: A parameterfree *ab initio* description," Phys. Rev. B **98**(12), 125140 (2018).

³⁹K. Pokharel, C. Lane, J. W. Furness, R. Zhang, J. Ning, B. Barbiellini, R. S. Markiewicz, Y. Zhang, A. Bansil, and J. Sun, "Sensitivity of the electronic and magnetic structures of high-temperature cuprate superconductors to exchange-correlation density functionals," npj Comput. Mater. **8**, 1 (2020).

⁴⁰ J. Hollingsworth, L. Li, T. E. Baker, and K. Burke, "Can exact conditions improve machine-learned density functionals?," J. Chem. Phys. **148**(24), 241743 (2018).

⁴¹ R. Nagai, R. Akashi, and O. Sugino, "Machine-learning-based exchange correlation functional with physical asymptotic constraints," Phys. Rev. Res. 4(1), 013106 (2022).

⁴²J. 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 *et al.*, "Pushing the frontiers of density functionals by solving the fractional electron problem," Science **374**(6573), 1385–1389 (2021).

⁴³E. H. Lieb and S. Oxford, "Improved lower bound on the indirect Coulomb energy," Int. J. Quantum Chem. 19(3), 427–439 (1981).

⁴⁴G. L. Oliver and J. P. Perdew, "Spin-density gradient expansion for the kinetic energy," Phys. Rev. A **20**(2), 397–403 (1979).

⁴⁵M. 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**(4), 2010 (1985).

⁴⁶P. S. Svendsen and U. von Barth, "Gradient expansion of the exchange energy from second-order density response theory," Phys. Rev. B 54(24), 17402 (1996).

⁴⁷L. Pollack and J. P. Perdew, "Evaluating density functional performance for the quasi-two-dimensional electron gas," J. Phys.: Condens. Matter 12(7), 1239 (2000).

⁴⁸J. P. Perdew and Y. Wang, "Accurate and simple analytic representation of the electron-gas correlation energy," Phys. Rev. B **45**(23), 13244 (1992).

⁴⁹J. P. Perdew, A. Ruzsinszky, J. Sun, and K. Burke, "Gedanken densities and exact constraints in density functional theory," J. Chem. Phys. **140**(18), 18A533 (2014).

⁵⁰J. P. Perdew and Y. Wang, "Erratum: Accurate and simple analytic representation of the electron-gas correlation energy [Phys. Rev. B 45, 13244 (1992)]," Phys. Rev. B 98(7), 079904 (2018).

⁵¹J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**(18), 3865–3868 (1996).

⁵²J. P. Perdew, P. Ziesche, and H. Eschrig, in Electronic Structure of Solids' 91, 1991.

⁵³ M. Seidl, J. P. Perdew, and S. Kurth, "Simulation of all-order density-functional perturbation theory, using the second order and the strong-correlation limit," Phys. Rev. Lett. 84(22), 5070 (2000).

⁵⁴J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, "Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids," Phys. Rev. Lett. **91**(14), 146401 (2003).

⁵⁵J. Tao, J. P. Perdew, L. M. Almeida, C. Fiolhais, and S. Kümmel, "Nonempirical density functionals investigated for jellium: Spin-polarized surfaces, spherical clusters, and bulk linear response," Phys. Rev. B 77(24), 245107 (2008).

⁵⁶B. L. Kalman and S. C. Kwasny, "Why tanh: Choosing a sigmoidal function," in Proceedings of the 1992 IJCNN International Joint Conference on Neural Networks (IEEE, 1992), Vol. 4, pp. 578–581.

⁵⁷A. D. Kaplan and M. Levy, J. P. Perdew Predictive Power of the Exact Constraints and Appropriate Norms in Density Functional Theory, arXiv:2207.03855.
⁵⁸E. Clementi and C. Roetti, "Roothaan-Hartree-Fock atomic wavefunctions: Basis functions and their coefficients for ground and certain excited states of neutral and ionized atoms, Z≤54," At. Data Nucl. Data Tables 14(3-4), 177-478 (1974).

⁵⁹T. Koga, K. Kanayama, S. Watanabe, and A. J. Thakkar, "Analytical Hartree–Fock wave functions subject to cusp and asymptotic constraints: He to Xe, Li⁺ to Cs⁺, H⁻ to I⁻," Int. J. Quantum Chem. **71**(6), 491–497 (1999).

⁶⁰J. W. Furness and S. Lehtola, Hartree-Fock orbitals for spherical atoms: A python toolbox, https://github.com/JFurness1/AtomicOrbitals, 2021.

⁶¹L. M. Almeida, J. P. Perdew, and C. Fiolhais, "Surface and curvature energies from jellium spheres: Density functional hierarchy and quantum Monte Carlo," Phys. Rev. B **66**(7), 075115 (2002).

⁶² B. Wood, N. D. M. Hine, W. M. C. Foulkes, and P. García-González, "Quantum Monte Carlo calculations of the surface energy of an electron gas," Phys. Rev. B 76(3), 035403 (2007).

⁶³ M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation (OSDI'16)* (USENIX Association2560 Ninth St. Suite 215 Berkeley, CA, 2016), pp. 265–283.

⁶⁴J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *International Workshop on Artificial Neural Networks* (Springer, 1995), pp. 195–201.

⁶⁵D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," arXiv:1511.07289 (2015).

⁶⁶F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res. **12**, 2825–2830 (2011).

⁶⁷D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2014).

⁶⁸V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in ICML'10: Proceedings of the 27th International Conference on Machine Learning, 2010.

⁶⁹L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, "Assessment of Gaussian-3 and density functional theories for a larger experimental test set," J. Chem. Phys. **112**(17), 7374–7383 (2000).

⁷⁰T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. V. R. Schleyer, "Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F," J. Comput. Chem. 4(3), 294–301 (1983).

⁷¹ M. J. Frisch, J. A. Pople, and J. S. Binkley, "Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets," J. Chem. Phys. **80**(7), 3265–3269 (1984).

⁷²A. M. Teale, J. W. Furness, T. J. P. Irons, M. S. Ryley, J. Zemen, M. Withnal, and J. Elsdon, QUantum Electronic Structure Techniques (QUEST), 2017.

⁷³ V. I. Lebedev and D. N. Laikov, "A quadrature formula for the sphere of the 131st algebraic order of accuracy," Dokl. Math. **59**(3), 477–481 (1999).

⁷⁴ R. Lindh, P. A. Malmqvist, and L. Gagliardi, "Molecular integrals by numerical quadrature. I. Radial integration," Theor. Chem. Acc. **106**(3), 178–187 (2001).

⁷⁵ J. Sun, M. Marsman, G. I. Csonka, A. Ruzsinszky, P. Hao, Y. S. Kim, G. Kresse, and J. P. Perdew, "Self-consistent meta-generalized gradient approximation within the projector-augmented-wave method," Phys. Rev. B 84(3), 035117 (2011).

⁷⁶V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, "*Ab initio* molecular simulations with numeric atom-centered orbitals," Comput. Phys. Commun. **180**(11), 2175–2196 (2009).

⁷⁷P. Vinet, J. Ferrante, J. R. Smith, and J. H. Rose, "A universal equation of state for solids," J. Phys. C: Solid State Phys. **19**(20), L467 (1986).

⁷⁸P. Vinet, J. H. Rose, J. Ferrante, and J. R. Smith, "Universal features of the equation of state of solids," J. Phys.: Condens. Matter 1(11), 1941 (1989).

⁷⁹ H. J. Liao, J. G. Liu, L. Wang, and T. Xiang, "Differentiable programming tensor networks," Phys. Rev. X 9(3), 031041 (2019).

⁸⁰J. Brownlee, "A gentle introduction to the rectified linear unit (ReLU)," Mach. Learn. Mastery 6, (2019).

⁸¹ A. C. Cancio, C. E. Wagner, and S. A. Wood, "Laplacian-based models for the exchange energy," Int. J. Quantum Chem. **112**(24), 3796–3806 (2012). ⁸²Z.-h. Yang, H. Peng, J. Sun, and J. P. Perdew, "More realistic band gaps from meta-generalized gradient approximations: Only in a generalized Kohn-Sham scheme," Phys. Rev. B 93(20), 205205 (2016).

⁸³Y. Yao and Y. Kanai, "Plane-wave pseudopotential implementation and performance of SCAN meta-GGA exchange-correlation functional for extended systems," J. Chem. Phys. **146**(22), 224105 (2017).

⁸⁴Y. Yamamoto, C. M. Diaz, L. Basurto, K. A. Jackson, T. Baruah, and R. R. Zope, "Fermi-Löwdin orbital self-interaction correction using the strongly constrained and appropriately normed meta-GGA functional," J. Chem. Phys. **151**(15), 154105 (2019). ⁸⁵A. P. Bartók and J. R. Yates, "Regularized SCAN functional," J. Chem. Phys. 150(16), 161101 (2019).

⁸⁶ J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, "Accurate and numerically efficient r²SCAN meta-generalized gradient approximation," J. Phys. Chem. Lett. **11**(19), 8208–8215 (2020).

⁸⁷J. M. Pitarke and J. P. Perdew, "Metal surface energy: Persistent cancellation of short-range correlation effects beyond the random phase approximation," Phys. Rev. B 67, 045101 (2003).

⁸⁸D. C. Langreth and J. P. Perdew, "Exchange-correlation energy of a metallic surface: Wave-vector analysis," Phys. Rev. B **15**, 2884–2901 (1977).