THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Cite This: J. Phys. Chem. Lett. 2019, 10, 6425−6431

Letter

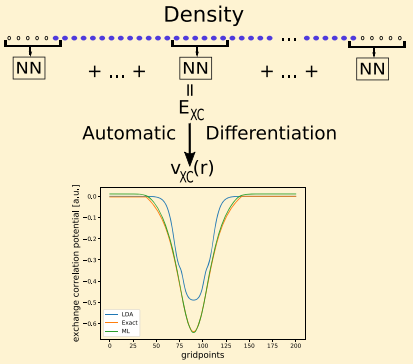
pubs.acs.org/JPCL

Machine Learning the Physical Nonlocal Exchange−Correlation Functional of Density-Functional Theory

Jonathan Schmidt, Carlos L. Benavides-Riveros,* and Miguel A. L. Marques*

Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany

ABSTRACT: We train a neural network as the universal exchange−correlation functional of density-functional theory that simultaneously reproduces both the exact exchange−correlation energy and the potential. This functional is extremely nonlocal but retains the computational scaling of traditional local or semilocal approximations. It therefore holds the promise of solving some of the delocalization problems that plague density-functional theory, while maintaining the computational efficiency that characterizes the Kohn−Sham equations. Furthermore, by using automatic differentiation, a capability present in modern machine-learning frameworks, we impose the exact mathematical relation between the exchange−correlation energy and the potential, leading to a fully consistent method. We demonstrate the feasibility of our approach by looking at one-dimensional systems with two strongly correlated electrons, where density-functional methods are known to fail, and investigate the behavior and performance of our functional by varying the degree of nonlocality.



Nowadays density-functional theory (DFT) is the cornerstone of computational theoretical physics and quantum chemistry, as it provides the prevalent method for the calculation of the electronic structure of both solids and molecules. Based on the Hohenberg−Kohn theorems,[1] DFT re-formulates the quantum many-electron problem as a theory of the ground-state electronic density $n(\mathbf{r})$. The success of DFT is to a large extent due to the existence of a system of noninteracting electrons (the Kohn−Sham system) that has the same ground-state density as the interacting electrons. This leads to the Kohn−Sham equations, a set of self-consistent equations for one-particle orbitals.[2] In such a formalism the ground-state (GS) energy can be expressed as

$$E_{\mathrm{GS}} = \sum_i \epsilon_i + E_{\mathrm{xc}}[n] - \int \mathrm{d}^3 r\, v_{\mathrm{xc}}(\mathbf{r})\, n(\mathbf{r}) - E_{\mathrm{H}}[n] \tag{1}$$

where ϵ_i are the eigenvalues of the Kohn−Sham Hamiltonian, $v_{\mathrm{xc}}(\mathbf{r})$ is the exchange−correlation potential, $E_{\mathrm{H}}[n]$ is the Hartree energy, and $E_{\mathrm{xc}}[n]$ is the exchange−correlation energy. The exchange−correlation potential is defined as the functional derivative of the universal exchange−correlation energy functional:

$$v_{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}[n]}{\delta n(\mathbf{r})} \tag{2}$$

Due to the Hohenberg−Kohn existence theorems, if the exact exchange−correlation energy functional $E_{\mathrm{xc}}[n]$ is known, Kohn−Sham DFT then yields the exact ground-state energy and the exact ground-state electronic density.

Traditionally, "educated" formal expressions of the exchange−correlation energy functional have been proposed by a combination of theoretical insight, highly accurate Monte Carlo,[3] or quantum chemical simulations or by fitting general expressions to experimental data. In general, functionals can be sorted according to Jacob's ladder:[4] the lowest rung of the ladder is occupied by local-density approximations (LDA) that use solely single density points as inputs.[5−7] The second rung is occupied by generalized-gradient approximations (GGA) that include the gradient of the density.[8,9] This is followed by the meta-GGAs[10] (that use the kinetic-energy density) and hybrid functionals[11−13] (that mix a fraction of nonlocal Fock exchange) on the subsequent rungs. Note that more than 500 of these functionals have been proposed in the past decades,[14] although most of them with rather limited impact.

In spite of the success of DFT in dealing efficiently with electronic systems, it still suffers from stubborn quantitative and qualitative failures. For instance, barriers of chemical reactions, band gaps of materials, or molecular dissociation energies are usually underestimated.[15] Degenerate or near-degenerate states are also poorly described by DFT. While hybrid functionals can alleviate some of the problems of traditional semilocal functionals, they come at a greatly increased computational cost that limits severely the number and size of systems that can be researched. It is believed that many of these problems originate in the delocalization and static correlation errors which plague approximate functionals.[16−18] Roughly speaking, the delocalization error refers to the tendency of DFT functionals to spread out the electron density, while the static correlation arises from the difficulty of describing degenerated states with a single Slater determinant.[19]

Received: August 19, 2019
Accepted: October 4, 2019
Published: October 9, 2019

DOI: 10.1021/acs.jpclett.9b02422
J. Phys. Chem. Lett. 2019, 10, 6425−6431

More recently, machine learning (ML) has revolutionized many fields of computational sciences, such as image or speech recognition,[20,21] and has found countless applications in material science.[22−24] Within DFT, the application of ML techniques to the formulation of density functionals has already a long history.[25] In 2012, an ML approximation for the kinetic energy functional $T_s[n]$ was constructed for a system of noninteracting spinless fermions.[26,27] Yao et al.[28] developed a convolutional neural network to reproduce the kinetic energy functional for molecules. They already mentioned the possibility of using the functional derivative of the neural network for minimization purposes. In order to exploit the Hohenberg−Kohn density-potential map, an ML model was later trained to learn the fundamental relation of DFT between external potentials and electronic densities.[29] These works focused mainly on developing functionals for the total energy or the noninteracting kinetic energy to facilitate orbital-free DFT calculations. More recently, some works have addressed the problem of training the exchange−correlation potential.[30−34] However, this line of research has been limited by the fact that the exchange−correlation potential was not obtained from the exchange−correlation energy through the functional derivative of eq 2.

It is true that one can find in the literature a series of approximations to the exchange−correlation functionals that do not fulfill eq 2. For example, the Krieger−Lee−Iafrate approximation[35] breaks this connection in order to simplify the implementations of orbital functionals using the optimized effective potential method.[36,37] Sometimes, it is also convenient to approximate directly the potential (e.g., in the van Leeuwen-Baerends GGA from 1994[38] or the modified Becke−Johnson potential[39,40]), leading again to expressions that do not obey eq 2. These so-called "stray" functionals[41] have found some important applications. For example, the modified Becke−Johnson is one of the most successful functionals to calculate electronic band-gaps.[42] Unfortunately, they are also found to break a series of exact theorems and conditions,[41] severely limiting their universality and range of applicability. By and large, it is highly advantageous to develop consistent functionals that obey the important eq 2.

Modern ML frameworks, like pytorch[43] and tensorflow,[44] allow for automatic differentiation with respect to any parameter. Recently, Nagai et al. used this functionality to train exchange−correlation potentials for molecules.[45] They trained neural networks through a Monte Carlo updating scheme to reproduce accurate energies and densities of molecules. The functionals by Nagai and coauthors follow the traditional approaches of an LDA, GGA, meta-GGA and add a related near-region approximation. Although a clear step forward, using traditional forms for the exchange−correlation functional is unlikely to lead to fundamentally better, disruptive approximations to the exchange−correlation functionals. New paradigms have to be sought in order to unleash the power of ML techniques to its full extent.

In this paper we use the autodifferentiation functionality to train neural-network exchange−correlation functionals through back-propagation. The networks are trained to reproduce not only the correct exchange−correlation energy E_{xc} but also the exchange−correlation potential $v_{xc}(r)$ consistently as its functional derivative with respect to the density. Consequently, the resulting functional allows for self-consistent calculations and can easily be integrated into existing Kohn−Sham DFT frameworks. Furthermore, these functionals can be made highly nonlocal by using the information on the density in a *finite* neighborhood as input to the neural network, allowing for far more nonlocality than traditional LDA or GGA functionals, despite having the same computational scaling with system size. Therefore, this approach promises to alleviate the delocalization problems of DFT and to improve its accuracy without the computational expense of hybrid functionals. To demonstrate the feasibility of this approach, we developed an ML functional for the exchange−correlation energy and exchange−correlation potential based on exact results for two electrons in one-dimension (1D).

The letter presents the details of the data set, training process, and neural networks. The exact dependence of the functional on the degree of locality and its behavior is also discussed, as well as our results for the 1D homogeneous electron gas and the H_2 molecule along the dissociation path. Finally, we finish the letter discussing our conclusions and future research directions.

Data. The training data was produced by solving exactly the one-dimensional two-electron problem in the external potential generated by up to three different nuclei. Softening the Coulomb interaction,

$$\frac{1}{r} \rightarrow \frac{1}{\sqrt{1 + x^2}} \tag{3}$$

we obtain the 1D Hamiltonian driven by the interaction of the two electrons, namely,

$$H(x_1,x_2) = -\sum_{i=1}^{2}\left[\frac{1}{2}\partial_i^2 + v(x_i)\right] + \frac{1}{\sqrt{1 + (x_1 - x_2)^2}}$$

where the external potential is given by the superposition of three potentials,

$$v(x) = \sum_{k=1}^{3}\frac{Z_k}{\sqrt{1 + (x - a_k)^2}} \tag{4}$$

The total charge of the nuclei $Z = \sum_k Z_k$ is equal to 2 or 3. Qualitatively close to real 3D systems, this 1D model is known as a theoretical laboratory for studying strong correlation and developing exchange−correlation density functionals for DFT.[46] Since the ground-state problem of the Hamiltonian $H(x_1,x_2)$ can be treated as a one-particle problem in two dimensions, the problem can be solved exactly.

We sampled 20 000 systems and calculated their exact ground-state energy and ground-state electronic density. We used a grid spacing of 0.1 au, and a box size of 20 au, leading to a grid with 201 points. The nuclei positions a_i in eq 4 were normally distributed with zero mean and variance of 4 au We then solved the corresponding inverse Kohn−Sham problem in OCTOPUS[47] to find the exact exchange−correlation energy and potential. Since the inversion is known to be numerically unstable,[48] we removed outliers that result from these instabilities. We used up to 12 800 of these systems for training, 6400 for validation during the training, and 2000 systems for the test set. Furthermore, training was considerably improved when removing outliers with $E_{xc} > -0.55$ au from the training set. No outliers were removed from the test set to allow for a completely unbiased evaluation of the functionals.

In general, one would have to double the data by mirroring the systems to learn the correct symmetry. However, in this specific case one can simply build the symmetry directly into

DOI: 10.1021/acs.jpclett.9b02422
J. Phys. Chem. Lett. 2019, 10, 6425−6431

the neural network functional, as explained in the next subsection.

Topology of the Neural Network. Our ML functional scans the density of the total system, as illustrated in Figure 1. The

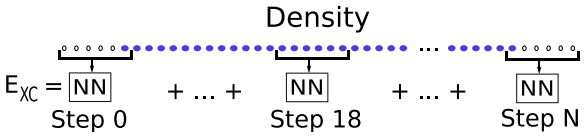


Figure 1. Structure of the ML functional in 1D with the degree of locality equal to $\kappa = 6$ (see text). At the borders, the density is padded with $\kappa - 1 = 5$ zeros. Starting at one of the borders, the network calculates the local exchange−correlation energy for κ points. In the next step, the input of the network is moved by one grid point and it is evaluated again. The network itself is a simple fully connected neural network.

density in a neighborhood of the test point is used as the input for a 4 or 5 layer fully connected neural network, that then outputs a *local* exchange−correlation energy. Specifically, the network takes a certain number of density points as input, which we call κ, the *kernel size*. This is the degree of locality of the ML functional. At the borders of the system the density is padded with $\kappa - 1$ zeros. Starting at one border the network calculates the local exchange−correlation energy. In the next step the input of the network is moved by one grid point, and it is evaluated again. As described in Figure 1, this process continues until the other border is reached. We arrive at the total exchange−correlation energy of the system by summing over all network outputs. The padding and the scanning with a certain kernel size are inspired by standard convolutional neural networks and can also be implemented as such by concatenating the data along the channel-dimensions in between standard convolutional layers. Due to the homogeneity of space, the functional has to be symmetric with respect to its input densities. The symmetry is ensured by initializing the weights of the first layer symmetrically along the spacial dimension. To arrive at the final scalar output we used four- or five-layer fully connected neural networks.

The possible selection of activation functions (i.e., the nonlinearities that follow each multiplication with a weight matrix of a neural network) was rather limited, because typical functions (e.g., rectified linear units) were not usable due to their lacking differentiability at zero (using relus actually resulted in piece-wise linear potentials). To avoid this problem, we chose exponential linear functions.[49] Different numbers and sizes of hidden layers were also tested, and we settled on the minimum number of parameters that could be used without underfitting. In this work all networks were built on the basis of the ML framework pytorch.[43] The library Ignite[50] was used to simplify the training process and tensorboardX to integrate tensorboard[44] into pytorch. The network weights were optimized with Adam[51] using default parameters from pytorch.

For the loss function (i.e., the cost function which is going to be optimized in the learning process), we should keep in mind that the objective is not only to obtain small errors for the exchange−correlation energy. To arrive at the correct density through the solution of the self-consistent Kohn−Sham equations, also the exchange−correlation potential should be as close as possible to the exact one. Furthermore, we want not only to ensure a small error for the potential but also its smoothness. In addition, the error of the exchange−correlation energy, as well as the error of the integral $\int \mathrm{d}x\, v_{xc}(x)\, n(x)$ (that appears in the expression for the total energy (1)), should be minimized.

In order to achieve all these goals concurrently, we used the following loss function, where θ are the parameters of the neural network that have to be optimized:

$$L(\theta, n_i) = \alpha \mathrm{MSE}(E_{xc}) + \beta \mathrm{MSE}(v_{xc}) + \gamma \mathrm{MSE}\left(\frac{\mathrm{d}v_{xc}(x)}{\mathrm{d}x}\right) + \delta \mathrm{MSE}\left(E_{xc} - \int \mathrm{d}x\, v_{xc}(x)\, n(x)\right) \quad (5)$$

This function is a weighted combination of the mean-squared errors (MSE) of the exchange−correlation energy, the exchange−correlation potential, its numerical spatial derivative, and the difference between the exchange−correlation energy and the integral over the potential. This latter term is part of the formula for the total energy (1) and theoretically allows for some error cancellation. We also attempted to use the integral as a separate term in eq 5. Depending on the network, one or the other term produced better results. Finally, the weights α, β, γ, and δ in eq 5 are also optimized as part of the hyperparameter optimization. Usual values for $\alpha, \beta, \gamma, \delta$ are 1.0, 100.0, 10.0, 1.0.

The training for the exchange−correlation energy converges quite fast after a few hundred epochs (i.e., one complete pass of the training data). The convergence of the potential can take thousands of epochs depending on the training set and batch size. At each training step the model was saved if it improved the validation error for the potential. The model with the lowest validation error was later used for the self-consistent Kohn−Sham calculations. As the amount of memory that is needed per sample is quite limited, very large batch sizes (e.g., 4096) are possible, allowing for a far more efficient parallelization of the training. Training with larger batchsizes seems to produce better convergence. However, it leads to a strong increase in the error during validation with self-consistent Kohn−Sham calculations. Smaller batch sizes (32, 64, 128) improve the error by up to 50% and provide the best generalization ability of the functionals, in consistency with the literature.[52,53]

Evaluation. We trained neural networks with various kernel sizes and used them within a self-consistent Kohn−Sham calculations for a test set of 2000 systems created with the method described above. The self-consistent Kohn−Sham calculations were run using a self-written code. For all kernel sizes, models with different hyper-parameters were evaluated on a validation set of 250 systems. The training was not completely converged at this stage. Only for the best models of each kernel size did we continue the training and evaluate the models on the test set. To compare various models with respect to the LDA functional of DFT, we chose energy differences, as these are physically more meaningful.

In Figure 2 we plot the exchange−correlation potential resulting from self-consistent calculations with a ML functional trained according to the loss function (5) for two test-systems and compared it with the exact and LDA predicted exchange−correlation potentials. In Figure 3 we plot the same information for one test system; the ML functional is this time trained only on the exchange−correlation energy (i.e., not on the potential). Whenever the machine-learned functional is trained on both the energy and the potential, the exchange−correlation potential presents a great improvement in

DOI: 10.1021/acs.jpclett.9b02422
J. Phys. Chem. Lett. 2019, 10, 6425−6431

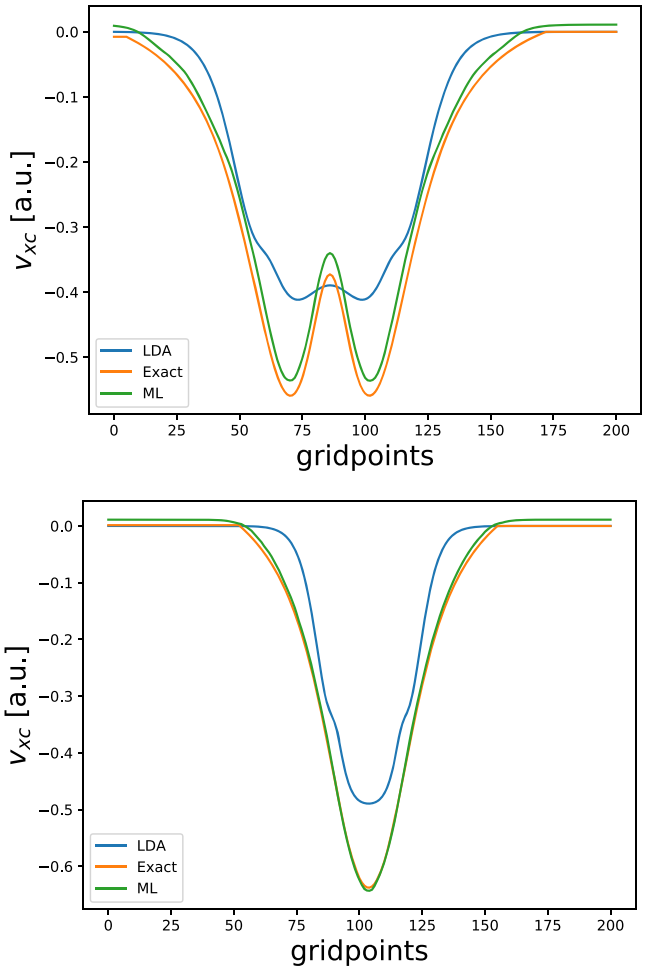


Figure 2. Comparison of exchange−correlation potentials of an 1D-LDA,[54] the exact potential, and our ML-functional with kernel size 30 for two different systems.

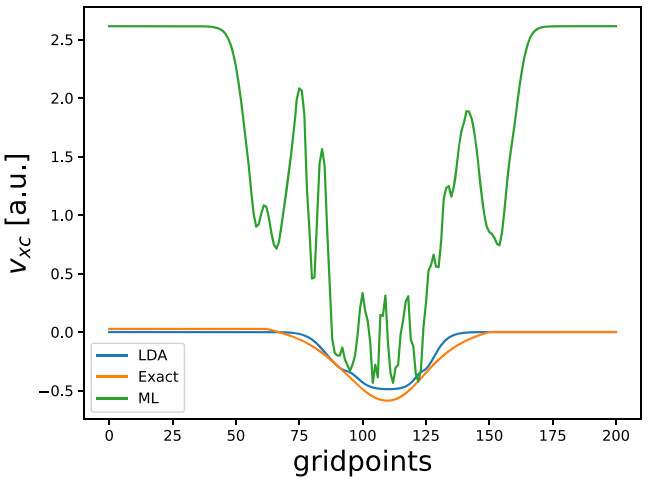


Figure 3. Comparison of exchange−correlation potentials of an 1D-LDA,[54] the exact potential, and our ML-functional trained only with the exchange−correlation energy.

comparison to the traditional LDA functional, while the functional trained only on the energy fails completely. Remarkably, the functionals trained with the loss-function (5) also show a qualitatively closer behavior to the exact exchange−correlation potentials. The results for the predicted total energies of the test set (relative to the energy of H_2 at its equilibrium distance) are presented in Table 1. Our ML-LDA

Table 1. Mean Absolute Errors (MAE) for the Total Energy in Self-Consistent Calculations for Various Kernel Sizes of Our ML-DFT Functional, Relative to the Error of the One-Dimensional LDA of Ref 54[a]

kernel size	MAE(ML)/MAE(LDA) [%]
LDA	100
1	38.1
15	21.8
30	8.2
60	8.2
120	7.1
180	6.5

[a]For reference, the mean absolute error of the LDA of ref 54 is 1.4×10^{-2} au.

(i.e., the functional with kernel size $\kappa = 1$) already performs better than the traditional LDA. As the ML-LDA was trained to reproduce the exchange−correlation energy of heterogeneous systems while traditional LDAs are "trained" for constant densities, the difference in performance is not surprising. Increasing the kernel size leads to a monotonical decrease of the error, and improves the results by more than a factor of 6 for the larger sizes. The optimal kernel size is, in our opinion, around 30 (i.e., 3 au), as larger kernels do not provide a significant advantage. Furthermore, some of the functionals with larger kernel sizes also demonstrate unphysical behavior (see below).

Ultimately, the more nonlocal the functional is, the higher the complexity and the larger the number of parameters. This reason, together with the need to represent more long-range interactions that are based on different physical principles (such as van der Waals interaction), makes the training considerably more difficult. One approach to circumvent this problem is to keep the nonlocality limited to ranges on the scale of molecular bonds. This allows for simpler training and still includes most of the nonlocality that is required for the exchange−correlation energy. Another possibility would be to enlarge the nonlocality by increasing the architectural complexity of the functional.

Efforts to decrease the number of training systems for a kernel size of 30 lead to a slightly increased error of 11% using 800 samples for training. Although the scaling of the networks to realistic three-dimensional systems is nontrivial, we expect that both the number of trainable parameters and the density points in the training set will grow cubically when transitioning to three-dimensional systems. In this sense, we expect a similar demand for training data as in 1D. Furthermore, realistic systems are usually far larger and therefore provide more "local" training samples per system for the neural network. Recent research by Nagai and coauthors points in the same direction.[45] Indeed, they only required a few sample molecules and used far more parameters to learn a much more local (and in this sense simpler) functional than the ones used here.

Previously, we tested the ML-functional on sample systems belonging to the distribution of the training data. Now, we go a step further and test how our functionals perform self-consistently on systems outside this distribution as well as a couple of paradigmatic cases. The systems in the training set had external potentials arising from 2 and 3 nuclei. In order to

DOI: 10.1021/acs.jpclett.9b02422
J. Phys. Chem. Lett. 2019, 10, 6425−6431

go beyond these systems, we tested the functional with kernel size 30 also on a test set of 150 systems with 4 nuclei. Using the same functional as in Table 1, we arrive at an error for the total energy more than eight times smaller than with the LDA (MAE(ML)/MAE(LDA) = 11.9%). Naturally, the error increased outside the training distribution. However, considering the different nature of the highly charged systems with 4 nuclei, this hints at a good generalization ability of the functional.

In Figure 4 we present our calculations for the H_2 molecule in 1D along the dissociation path with functionals of varying

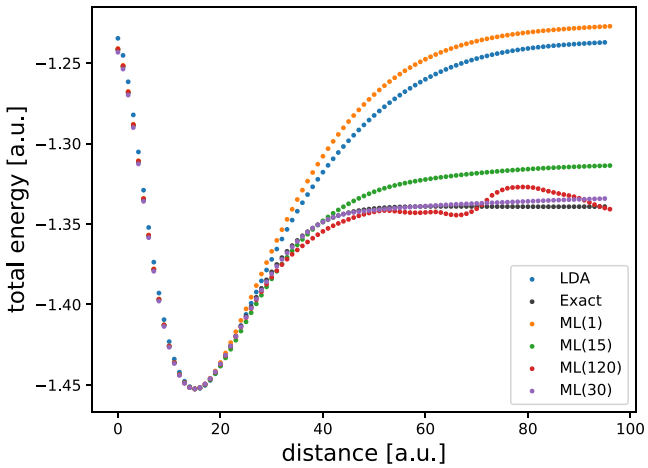


Figure 4. Dissociation curves of the 1D H_2 molecule with ML functionals of varying nonlocality (i.e., kernels of 1, 15, 120, 30) in comparison to the exact and the LDA results.

nonlocality in comparison to the exact result. The curves in Figure 4 are shifted to have the same equilibrium energies. As is well-known, the traditional LDA completely fails to produce the correct dissociation limit.[54] The same behavior is observed for the ML LDA (with a kernel size of 1). Remarkably, using increasingly more nonlocal functionals, we can reduce the relative energy error to 3.2% of the LDA error.

It is obvious that even the functional with a kernel size of 30 will start failing above a certain distance. This is a conceptual problem of local KS-DFT and can only be alleviated and not eliminated in our approach. It can already be considered a success that our functionals are able to reproduce the dissociation curve reasonably well far beyond their own degree of nonlocality.

Yet it has to be noted that not all functionals performed that well. Some of the functionals with larger kernels failed to reproduce a physical behavior with respect to the dissociation distance and produced multiple local minima and maxima. Despite these problems, they still return the correct equilibrium distance and on average far better energies than the LDA. This unphysical behavior of some functionals just stresses the fact that a rigorous validation on a multitude of different systems will be essential to arrive at a working functional. It has to be noted that a larger training set and a longer training time was far more beneficial for this validation than for example the average error. As systems similar to the dissociated molecule are most likely outliers of the training data this is not surprising. Note that the need for more training data can, however, be avoided by active learning and a thoughtful construction of the training set.

Finally, we study the homogeneous-electron gas, a model system that is used in the construction of the majority of exchange−correlation functionals. We can simulate this system with our neural networks by providing them with a constant electronic density as input. The results can then be compared to the numerically exact values for the energy density as obtained, for example, from quantum Monte Carlo simulations.[54] Our results are depicted in Figure 5 as a function of

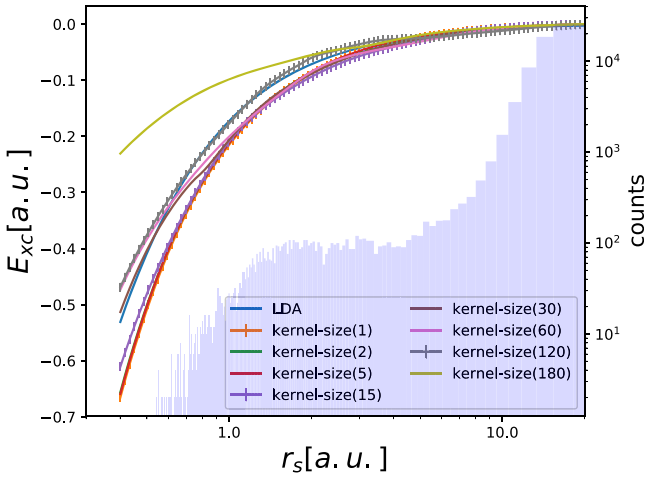


Figure 5. Exchange−correlation energy per unit volume of a 1D homogeneous electron gas from quantum Monte Carlo calculations[54] (curve labeled LDA) is compared with several ML functionals evaluated at constant density. We also plot a histogram of the number of systems of the training set containing more than three grid points with a density within a bin size of 0.01 au. The number of counts can be read on the right axis. Notice that there is basically no system with $r_s < 1$. The machine-learned curves are shifted to be exactly zero at zero density.

the Wigner−Seitz radius $r_s = 1/2n$. Remarkably, our ML functional with kernel size equal to 120 reproduces the exchange−correlation energy of the 1D homogeneous electron gas, especially for $r_s > 2$, where the majority of counts in the training data were located. Furthermore, kernel sizes lower than or equal to 60 are practically indistinguishable for larger r_s and behave qualitatively very close to the homogeneous electron gas. The largest kernel size (say, 180) underestimates the exchange−correlation energy.

Some difficulties for our neural-network functionals are evident. First, the functionals were trained only for systems with a specific size while the homogeneous electron gas is, in fact, an infinite periodic system. Second, as the histogram in Figure 5 illustrates, the training data does not contain almost any samples with high densities ($r_s < 1$). Naturally, the availability of training data similar to the homogeneous electron gas is even more important for the more nonlocal functionals as they take into account larger regions of space. The first challenge results in the fact that the nonzero biases in each layer cause the neural networks to output a nonzero value for zero density. When training for different system sizes, there are several ways to avoid this failure. First, one could solve the problem by adding systems padded with different amounts of zeros at the border to force the neural network to learn the correct relationship. As a second possibility, one could force all biases of all layers to zero, however, this would severely limit the expressibility of the networks. To circumvent this problem, and in order to compare the behavior of the energy with

DOI: 10.1021/acs.jpclett.9b02422
J. Phys. Chem. Lett. 2019, 10, 6425−6431

respect to the Wigner−Seitz radius, we shifted the curves in Figure 5 to yield zero energy for zero density.

Despite the small amount of training data at high density, functionals with larger kernel sizes still generalize on average far better to the homogeneous electron gas. While this is the case for most models, there are some rare cases, similar to the problems with the H2 dissociation, where large kernel sizes produce unphysical behavior. It is not obvious whether this result will remain true in three dimensions, it is nevertheless promising that the extra nonlocal information in the larger kernels might help the functionals to be generalized. Constant densities will be an essential feature of a functional for solid-state physics. Fortunately, exact training data in the form of quantum Monte Carlo calculations already exists for this purpose and can be easily incorporated in our training sets.

In conclusion, in this Letter we have demonstrated the viability of learning an exchange−correlation potential, via the differentiation of the exchange−correlation energy in a physically consistent manner. This procedure allows for standard self-consistent Kohn−Sham calculations. From the presented data, it is evident that neural-network functionals trained on the exchange−correlation potential and energy have the potential to be far more precise than previous local DFT functionals. Increasing the nonlocality of the functional allows for an extremely precise treatment of the electronic interaction on the scale of at least a few atomic units and, to a certain extent, even solve long-standing problems in DFT like, e.g., molecular dissociation.

For simplicity, we trained a neural network to the one-dimensional two-electron problem in the external potential generated by up to to three nuclei. Training three-dimensional systems will have to be accomplished by using data obtained with coupled-cluster, full configuration-interaction, or quantum Monte Carlo. While sufficient data to train a universal functional still has to be created, exchange−correlation energies and potentials for a few small molecules already exists and can provide a good starting point. The density representation on a grid is unfortunately not feasible for more general systems, as grid sizes and forms will vary. However, we think this can easily be circumvented by representing the density locally in some basis sets (e.g., Gaussians).

Finally, there is already a long history within DFT in the development of empirical functionals.[32,33,55−57] The machine learning paradigm allows us to drastically increase the amount of data used for the training and the complexity of these functionals. Including known exact conditions of the exchange−correlation functional in the learning process as constraints in the minimization will still be helpful[58] and provide further conditions for validation. Furthermore, as the functionals will have to work in practically every density environment, the importance of an extremely in-depth validation cannot be overstated and will be essential to arrive at a widely used functional.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: carlos.benavides-riveros@physik.uni-halle.de.
*E-mail: miguel.marques@physik.uni-halle.de.

ORCID

Carlos L. Benavides-Riveros: 0000-0001-6924-727X
Miguel A. L. Marques: 0000-0003-0170-8222

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge partial support from the German DFG through project MA-6786/1.

REFERENCES

(1) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864.
(2) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.
(3) Ceperley, D. M.; Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **1980**, *45*, 566.
(4) Perdew, J. P.; Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **2000**, *577*, 1.
(5) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200.
(6) Cole, L. A.; Perdew, J. P. Calculated electron affinities of the elements. *Phys. Rev. A: At., Mol., Opt. Phys.* **1982**, *25*, 1265.
(7) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13244.
(8) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(9) Perdew, J. P.; Burke, K.; Wang, Y. Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 16533.
(10) Sun, J.; Remsing, R. C.; Zhang, Y.; Sun, Z.; Ruzsinszky, A.; Peng, H.; Yang, Z.; Paul, A.; Waghmare, U.; Wu, X.; Klein, M. L.; Perdew, J. P. Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. *Nat. Chem.* **2016**, *8*, 831.
(11) Becke, A. D. A new mixing of Hartree−Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372.
(12) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **1996**, *105*, 9982.
(13) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158.
(14) Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L. Recent developments in libxc − − A comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1.
(15) Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **2014**, *140*, 18A301.
(16) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792.
(17) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Fractional spins and static correlation error in density functional theory. *J. Chem. Phys.* **2008**, *129*, 121104.
(18) Skone, J. H.; Govoni, M.; Galli, G. Self-consistent hybrid functional for condensed systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 195112.
(19) Benavides-Riveros, C. L.; Lathiotakis, N. N.; Marques, M. A. L. Towards a formal definition of static and dynamic electronic correlations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 12655.
(20) Liu, S.-S.; Tian, Y.-T. *Advances in Neural Networks*; Springer: Berlin, Heidelberg, 2010; pp 144−151.
(21) Waibel, A., Lee, K.-F., Eds. *Readings in Speech Recognition*; Morgan Kaufmann, 1990.
(22) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547.

DOI: 10.1021/acs.jpclett.9b02422
J. Phys. Chem. Lett. 2019, 10, 6425−6431

(23) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **2019**, *5*, 83.
(24) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.
(25) Tozer, D. J.; Ingamells, V. E.; Handy, N. C. Exchange-correlation potentials. *J. Chem. Phys.* **1996**, *105*, 9200.
(26) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.
(27) Snyder, J. C.; Rupp, M.; Hansen, K.; Blooston, L.; Müller, K.-R.; Burke, K. Orbital-free bond breaking via machine learning. *J. Chem. Phys.* **2013**, *139*, 224104.
(28) Yao, K.; Parkhill, J. Kinetic Energy of Hydrocarbons as a Function of Electron Density and Convolutional Neural Networks. *J. Chem. Theory Comput.* **2016**, *12*, 1139.
(29) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **2017**, *8*, 872.
(30) Liu, Q.; Wang, J. C.; Du, P. L.; Hu, L. H.; Zheng, X.; Chen, G. Improving the Performance of Long-Range-Corrected Exchange-Correlation Functional with an Embedded Neural Network. *J. Phys. Chem. A* **2017**, *121*, 7273.
(31) Nagai, R.; Akashi, R.; Sasaki, S.; Tsuneyuki, S. Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability. *J. Chem. Phys.* **2018**, *148*, 241737.
(32) Lundgaard, K. T.; Wellendorff, J.; Voss, J.; Jacobsen, K. W.; Bligaard, T. mBEEF-vdW: Robust fitting of error estimation density functionals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 235162.
(33) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 235149.
(34) Li, L.; Baker, T. E.; White, S. R.; Burke, K. Pure density functional for strong correlation and the thermodynamic limit from machine learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94*, 245129.
(35) Krieger, J. B.; Li, Y.; Iafrate, G. J. Derivation and application of an accurate Kohn-Sham potential with integer discontinuity. *Phys. Lett. A* **1990**, *146*, 256.
(36) Sharp, R. T.; Horton, G. K. A Variational Approach to the Unipotential Many-Electron Problem. *Phys. Rev.* **1953**, *90*, 317.
(37) Talman, J. D.; Shadwick, W. F. Optimized effective atomic central potential. *Phys. Rev. A: At., Mol., Opt. Phys.* **1976**, *14*, 36.
(38) van Leeuwen, R.; Baerends, E. J. Exchange-correlation potential with correct asymptotic behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1994**, *49*, 2421.
(39) Becke, A. D.; Johnson, E. R. A Simple Effective Potential for Exchange. *J. Chem. Phys.* **2006**, *124*, 221101.
(40) Tran, F.; Blaha, P. Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential. *Phys. Rev. Lett.* **2009**, *102*, 226401.
(41) Gaiduk, A. P.; Staroverov, V. N. How to tell when a model Kohn−Sham potential is not a functional derivative. *J. Chem. Phys.* **2009**, *131*, 044107.
(42) Borlido, P.; Aull, T.; Huran, A. W.; Tran, F.; Marques, M. A. L.; Botti, S. Large-Scale Benchmark of Exchange−Correlation Functionals for the Determination of Electronic Band Gaps of Solids. *J. Chem. Theory Comput.* **2019**, *15*, 5069.
(43) Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. *Automatic differentiation in pytorch. NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques*. 2017.
(44) Abadi, M.; et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. https://tensorflow.org/, 2015.
(45) Nagai, R.; Akashi, R.; Sugino, O. Completing density functional theory by machine-learning hidden messages from molecules. *arXiv:1903.00238*, **2019**.
(46) Wagner, L. O.; Stoudenmire, E. M.; Burke, K.; White, S. R. Reference electronic structure calculations in one dimension. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8581.
(47) Andrade, X.; et al. Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31371.
(48) Jensen, D. S.; Wasserman, A. Numerical methods for the inverse problem of density functional theory. *Int. J. Quantum Chem.* **2018**, *118*, e25425.
(49) Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv:1511.07289*, **2015**.
(50) *Ignite*. https://github.com/pytorch/ignite, 2018.
(51) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv:1412.6980*, **2014**.
(52) Masters, D.; Luschi, C. Revisiting small batch training for deep neural networks. *arXiv:1804.07612*, **2018**.
(53) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016; http://www.deeplearningbook.org.
(54) Helbig, N.; Fuks, J. I.; Casula, M.; Verstraete, M. J.; Marques, M. A. L.; Tokatly, I. V.; Rubio, A. Density functional theory beyond the linear regime: Validating an adiabatic local density approximation. *Phys. Rev. A: At., Mol., Opt. Phys.* **2011**, *83*, 032503.
(55) Yu, H. S.; He, X.; Truhlar, D. G. MN15-L: A New Local Exchange-Correlation Functional for Kohn−Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids. *J. Chem. Theory Comput.* **2016**, *12*, 1280.
(56) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn−Sham global-hybrid exchange−correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chem. Sci.* **2016**, *7*, 5032.
(57) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
(58) Hollingsworth, J.; Baker, T. E.; Burke, K. Can exact conditions improve machine-learned density functionals? *J. Chem. Phys.* **2018**, *148*, 241743.

DOI: 10.1021/acs.jpclett.9b02422
J. Phys. Chem. Lett. 2019, 10, 6425−6431