Check for updates

OPEN

Neural network interpolation of exchange-correlation functional

Alexander Ryabov[1,2], Iskander Akhatov[1] & Petr Zhilyaev[1] ✉

Density functional theory (DFT) is one of the most widely used tools to solve the many-body Schrodinger equation. The core uncertainty inside DFT theory is the exchange-correlation (XC) functional, the exact form of which is still unknown. Therefore, the essential part of DFT success is based on the progress in the development of XC approximations. Traditionally, they are built upon analytic solutions in low- and high-density limits and result from quantum Monte Carlo numerical calculations. However, there is no consistent and general scheme of XC interpolation and functional representation. Many different developed parametrizations mainly utilize a number of phenomenological rules to construct a specific XC functional. In contrast, the neural network (NN) approach can provide a general way to parametrize an XC functional without any *a priori* knowledge of its functional form. In this work, we develop NN XC functionals and prove their applicability to 3-dimensional physical systems. We show that both the local density approximation (LDA) and generalized gradient approximation (GGA) are well reproduced by the NN approach. It is demonstrated that the local environment can be easily considered by changing only the number of neurons in the first layer of the NN. The developed NN XC functionals show good results when applied to systems that are not presented in the training/test data. The generalizability of the formulated NN XC framework leads us to believe that it could give superior results in comparison with traditional XC schemes provided training data from high-level theories such as the quantum Monte Carlo and post-Hartree-Fock methods.

Since its initial formulation, density functional theory (DFT)[1,2] has become one of the main methods to solve the many-body Schrodinger equation. The source of the major ambiguity in DFT theory is the lack of a universal form of the exchange-correlation (XC) functional. Therefore, the success of DFT is mostly determined by the improvement of XC representations. The seminal Monte Carlo (MC) simulations of Ceperley and Adler[3] have resulted in a number of practical local density approximations (LDAs)[4–6]. The next major progress was produced by the introduction of the generalized gradient approximation (GGA), which takes into account local gradients of electron density[7–9]. This greatly enhanced the ability of DFT to describe systems with inhomogeneous electron densities, such as molecules and surfaces. A number of other functionals have been suggested with various degrees of complexity, extending DFT to a broad class of material properties[10]. The seeking for better functionals is still a very active area of research[11–14].

Despite the obvious success of the LDA and GGA, their construction is a highly complicated process that includes many heuristics steps. Even the XC functional form itself is based on some assumptions about the local nature of XC interactions. A large number of analytical conditions to fulfill[9] and a commitment to the local functional form make it difficult to incorporate new numerical results from high-level quantum mechanics theories. For example, only numerical data from MC simulations of homogeneous electron gas are used to produce LDA[6] and GGA[9] functionals. Therefore, much of the progress in quantum MC[15,16] and post-Hartree-Fock[17] methods is not adopted by XC approximations. One possible solution to overcome this problem is to use neural networks (NNs) for the interpolation of XC functionals. The NN provides a general way to represent any functional relationship[18] and a highly flexible framework to embody both analytical and numerical training data.

There have been a number of studies in which NNs have been used to describe the potential energy surface to speed up electronic structure calculations[19–26], while fewer studies have been devoted to developing NN XC functionals[27–30]. NN was firstly utilized to approximate XC potential by Tozer *et al.*[31]. This study showed that NN approach is flexible and principally applicable to generate XC potentials to be used in the real physical systems DFT calculations.

[1]Center for Design, Manufacturing and Materials, Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, Moscow, 143026, Russia. [2]Moscow Institute of Physics and Technology (State University), Institutskiy per. 9, Dolgoprudny, Moscow Region, 141700, Russia. ✉e-mail: p.zhilyaev@skoltech.ru
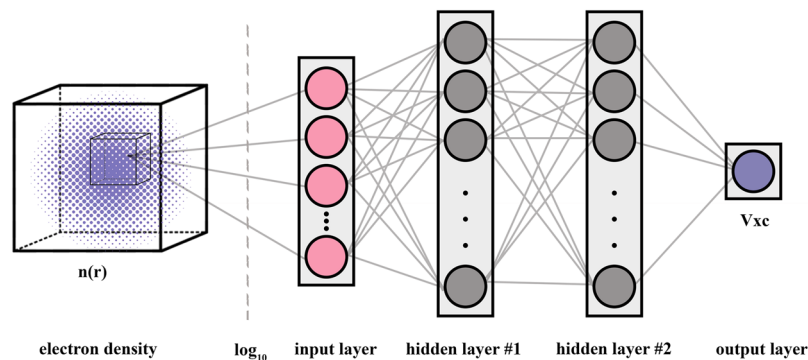
Figure 1. Topology of the developed neural network. Two hidden layers are used, with 1000 and 500 neurons, respectively. A base 10 logarithm is applied to the input data (electron density). This helps to scale the input data to the range that the neural network is most sensitive to.

In this work, we focus on constructing NN XC functionals for 3D systems without any feature engineering. We show that even rather simple NN architectures reproduce LDA and GGA XC potentials with satisfactory accuracy. To construct the LDA, potential point-to-point mapping is used due to the local idea behind the LDA itself. GGA potential region-to-point mapping is achieved by increasing the number of input neurons to take into account the local environment.

This paper is outlined as follows: In methods section, we describe the data generation process, NN topologies and hyperparameters. In results section, we show that the constructed NN XC potentials have the capability to reproduce LDA and GGA results for real physical systems. In conclusion section, we summarize our results and consider future research that could be done with the proposed framework, with a particular emphasis on incorporating data from high-level methods such as quantum Monte Carlo and post-Hartree-Fock.

Methods

To obtain the training data needed to build the NN XC potential, we carry out calculations of three-dimensional (3D) electron gases in a framework of the Kohn-Sham formalism[1,2] for potentials of a simple harmonic oscillator (SHO). This potential takes the form:

$$V_{ext}(\mathbf{r}) = \frac{1}{2}\sum_{i}^{3} k_i x_i^2$$

(1)

where k_i is the spring constant and $\mathbf{r} = (x_1, x_2, x_3)$ is the radius vector. Parameter k_i pertains to a simple harmonic oscillator randomly sampled from intervals $0.01 \leq k_i \leq 0.9$ Ha/a_0^2, where a_0 is 1 Bohr radius.

The XC potential is evaluated on a $40 \times 40 \times 40$ a_0 parallelepiped with $32 \times 32 \times 32$ mesh, which corresponds to a spacing of 1.25 a_0. Three electrons are placed in this 3D space. For each chosen parameter, DFT calculations in real space were performed using the Octopus code[32–34]. The parametrizations of the LDA exchange-correlation functional used include the Slater exchange[35,36] and Perdew-Zunger (modified) correlation part[5]. The GGA parameters used include the exchange and correlation parts from[9,37]. The total number of DFT calculations performed is 150 for each XC parametrization. The ranges of electron density represented by the Wigner-Seitz radius r_s are $[1; 5 \times 10^6]$ and $[1; 10^8]$ for LDA and GGA, respectively. One calculation then provides a $32 \times 32 \times 32$ dataset size for LDA and a $28 \times 28 \times 28$ size for GGA. These yield overall dataset sizes of approximately 5.0 million and 3.3 million for LDA and GGA, respectively, but only 3.3 million samples are used for both functionals. 15% of all datasets are used for the test set.

The composition of Tensorflow[38] and TFLearn[39] is utilized for training the NN. The Adam algorithm with a learning rate of 0.001 is used for training, with MSE loss. The neural architecture used for the LDA is a fully connected network with one input neuron, one linear output neuron and two ReLU[40] hidden layers with 1000 and 500 neurons (see Fig. 1). The only difference in the GGA case is that it takes into account the contribution of the local environment, which corresponds to an increase in the number of neurons in the first layer to 125 ($5 \times 5 \times 5$ cube). It should be noted that in principle the local environment could be taken into account explicitly by passing gradient and higher order derivatives of electron density. We delegate a computation of electron density derivatives to NN itself, making preprocessing step easy and convenient. Considered $5 \times 5 \times 5$ input cube gives information to calculate gradient and Jacobian matrix with $O(\Delta r^2)$ and $O(\Delta r)$ precision correspondingly, where Δr is mesh spacing.

Results and Discussion

Multiple neural architectures, including different activation functions, numbers of layers and numbers of neurons, are tested. We choose robust and rather simple architectures, namely, the point-to-point and region-to-point architectures described above. We also test NNs with only one hidden layer with 1000 neurons, which have poor generalizability and result in a larger loss in comparison with the topology used. It should be noted that the topology chosen is heuristics, and it is possible that another number of hidden layers and neurons could give better results than those provided in this study. The search for the optimal NN topology is a large task that needs to be carried out in future studies.

2

	NN LDA	NN GGA
Num. of inputs	1	125
Dataset size, 10^6	3.3	3.3
r_s range, Bohr	$[1; 5 \times 10^6]$	$[1; 1 \times 10^8]$
Training/test split (%)	85:15	85:15

Table 1. General information about NN XC functionals training process.
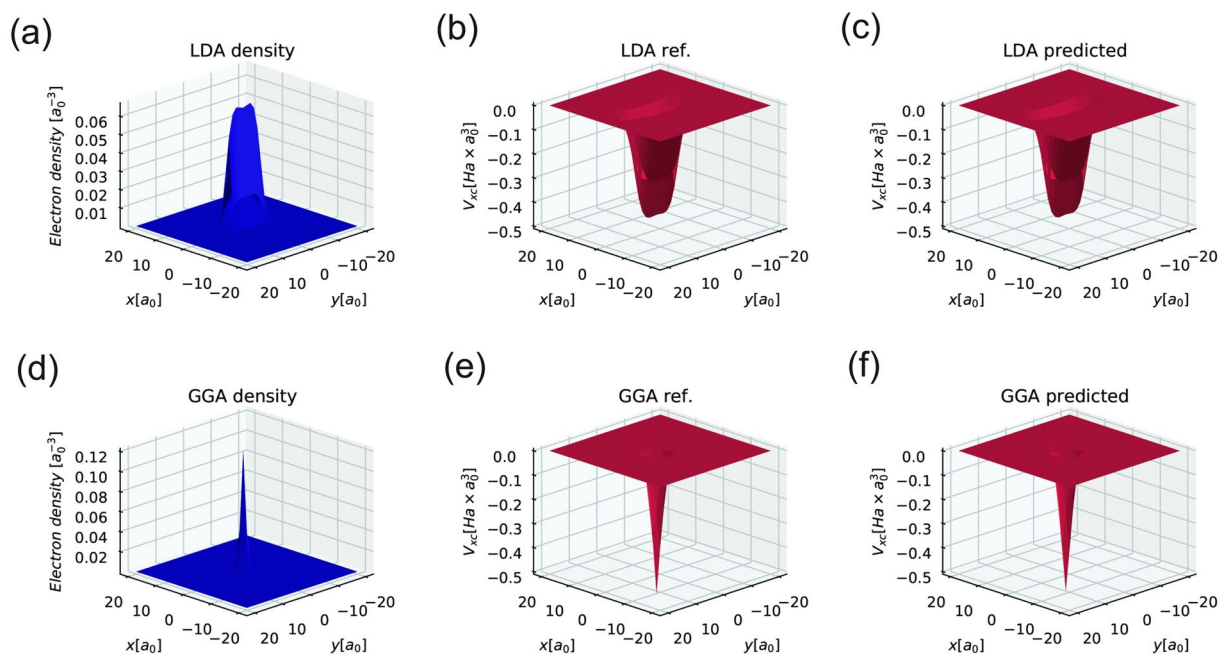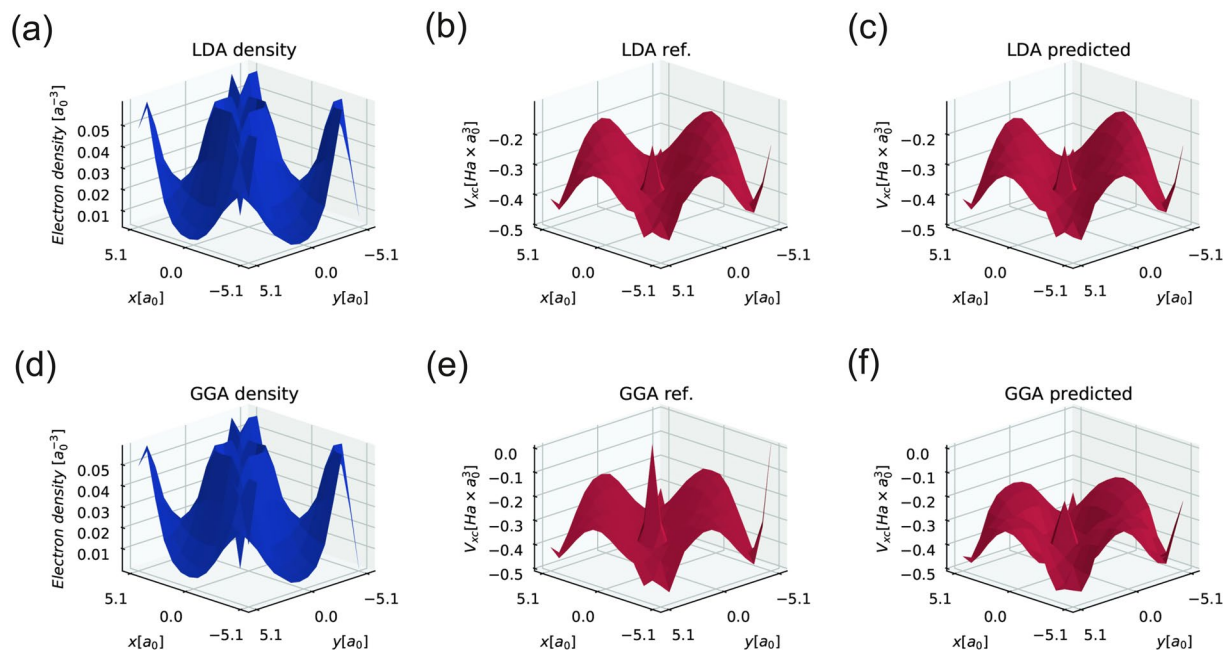

Figure 2. The results of applying NN XC functionals to the self-consistent electron density obtained from solving the Schrodinger equation in the framework of DFT for 3 electrons in a simple harmonic potential. LDA and GGA "true" V_{xc} results are produced by the Octopus package. All presented images are slices of 3D data in the z = 0 plane; (**a**) - the slice of the electron density obtained in the DFT calculation using LDA XC funtional, (**b**) - the slice of the XC potential obtained in the same as (**a**) DFT calculation using LDA XC functional, (**c**) - the slice of the NN XC potential obtained from the (**a**) electron density, (**d**) - the slice of the electron density obtained in the DFT calculation using GGA XC funtional, (**e**) - the slice of the XC potential obtained in the same as (**d**) DFT calculation using GGA XC functional, (**f**) - the slice of the NN XC potential obtained from the (**d**) electron density.

Because the density range contains mostly small values, base-10 logarithmic transformation is chosen for preprocessing the dataset. This scaling greatly improves the convergence speed of the backpropagation algorithm during the NN training cycle. No other preprocessing techniques and feature engineering are applied to the input data.

Training of the LDA NN is stopped after 1300 epochs. The NN LDA mean absolute error (MAE) is 0.03 mHa × Bohr³. The mean minimum and maximum values of V_{xc} averaged over all configurations used for the LDA NN construction are 0.73 and 510 mHa × Bohr³, respectively. Training of the GGA NN is terminated after 320 epochs. The NN GGA mean absolute error (MAE) is 0.156 mHa × Bohr³. The mean minimum and maximum values of Vxc averaged over all configurations used for the GGA NN construction are 0 and 553 mHa × Bohr³, respectively. The results of the training for both the LDA and GGA NNs are summarized in Table 1.

The results of applying an NN to the electron density from the test set are presented in Fig. 2. Both XC potentials provided by the LDA and GGA NNs are in good agreement with the numerical calculations made by Octopus. The special features of the V_{xc} form, for example, the "two humps", are also well reproduced by them. For LDA it is also possible to provide comparison between $V_{xc}^{NN}(r_s)$ and analytical reference function $V_{xc}(r_s)$. Plotted curves are indistinguishable in the range of r_s from 1 to 15, the absolute maximum difference is 1.7 mHa × Bohr³ (see Fig. S1 of supplementary materials).

To test the NN XCs for real physical systems, we consider Si in a diamond structure with 8 basis atoms in a cubic cell (lattice constant: 10.2 a_0). The numerical calculation is performed on a $10.2 \times 10.2 \times 10.2\ a_0$ parallelepiped with a $14 \times 14 \times 14$ mesh, which corresponds to a spacing of 0.729 a_0. The NN LDA and NN GGA MAEs are 0.6 and 18.1 mHa × Bohr³, respectively. The NN LDA results are in excellent agreement with the reference data obtained from Octopus. The NN GGA describes V_{xc} reasonably well, except for regions with high

Figure 3. The results of applying NN XC functionals to the self-consistent electron density obtained from solving the Schrodinger equation in the framework of DFT for a Si diamond structure with 8 basis atoms in a cubic cell. The LDA and GGA "true" V_{xc} results are produced by the Octopus package. All presented images are slices of 3D data in the (001) plane defined by the Miller indices; (**a**) - the slice of the electron density obtained in the DFT calculation using LDA XC funtional, (**b**) - the slice of the XC potential obtained in the same as (**a**) DFT calculation using LDA XC functional, (**c**) - the slice of the NN XC potential obtained from the (**a**) electron density, (**d**) - the slice of the electron density obtained in the DFT calculation using GGA XC funtional, (**e**) - the slice of the XC potential obtained in the same as (**d**) DFT calculation using GGA XC functional, (**f**) - the slice of the NN XC potential obtained from the (**d**) electron density.

density variations (see Fig. 3). We believe that this is because of the different spacings of the training/test data, which lead to the change in electron density gradients perceived by the NN XC.

The NN XC potentials are also tested on slices of 3D data in the plane of the benzene molecule. The numerical calculation is performed on a $14 \times 14 \times 14$ a_0 parallelepiped with a $44 \times 44 \times 44$ mesh, which corresponds to a spacing of 0.318 a_0. The NN LDA and NN GGA MAEs are 1.36 and 18.3 mHa $\times$ Bohr^3, respectively. The overall form of the XC potential obtained by Octopus is similar to the form produced by both NNs. In the case of benzene, the MAE is substantially larger than the MAE obtained during the training/test procedure. This is due to the high electron density values obtained for benzene, which are not presented in the training/test dataset. Moreover, the mesh spacing in the benzene calculation is smaller than that in the dataset, which is important for NN GGA potential evaluation. Despite the lack of appropriate input data, both NN XC potentials show an impressive ability to generalize the XC functional (see Fig. 4).

The results obtained from the developed NN XC potentials show that NN LDA works well, provided that the electron density values are presented in the training/test dataset. Based on Table 2, the NN LDA MAE is rather small for silicon and becomes larger for benzene, in which the electron density contains values that are outside the training/test density range. It is not surprising that the NN GGA (see Table 3) demonstrates a large MAE for both systems considered, as the mesh spacing is not presented in the training procedure. Electron density spatial information not included explicitly in the training set leads to the fact that neural networks can take gradients into account in a way that cannot be properly generalized. Changing the mesh corrupts the learned representations and makes it not quite correct to apply the NN GGA to systems on meshes other than the one learned.

To construct mesh-independent nonlocal functionals such as the GGA, it is necessary to use the coordinates of mesh points at which the electron density is defined. This additional information will help the NN evaluate mesh-independent gradients or other higher-order variations of electron density and include them in the process of the interpolation of the exchange-correlation potential. The type of local coordinate system to use remains the subject of further research.

We also implemented both NN LDA and NN GGA potentials in the OCTOPUS package. The xc-energy was taken from original LDA[5,35,36] and GGA[9,37] functionals. A behaviour during self-consistent cycle and an ability to perform geometry optimization were tested. Lattice constant of Si was scanned in range from 9.6 to 10.95 Bohr with a step 0.15 Bohr. Bond length of a benzene molecule was changed in range from 2.37 to 2.90 Bohr with a step 0.053 Bohr. All performed self-consistent calculations converged with density tolerance 0.5×10^{-5} Bohr^{-3}. In case of LDA dependencies of a total energy on a geometry parameter for both Si and the benzene molecule are in good agreement with results obtained for original xc-potentials (see Fig. 5). There is a notable shift to higher values for NN GGA dependencies of a total energy on a geometry parameter in comparison with one obtained from
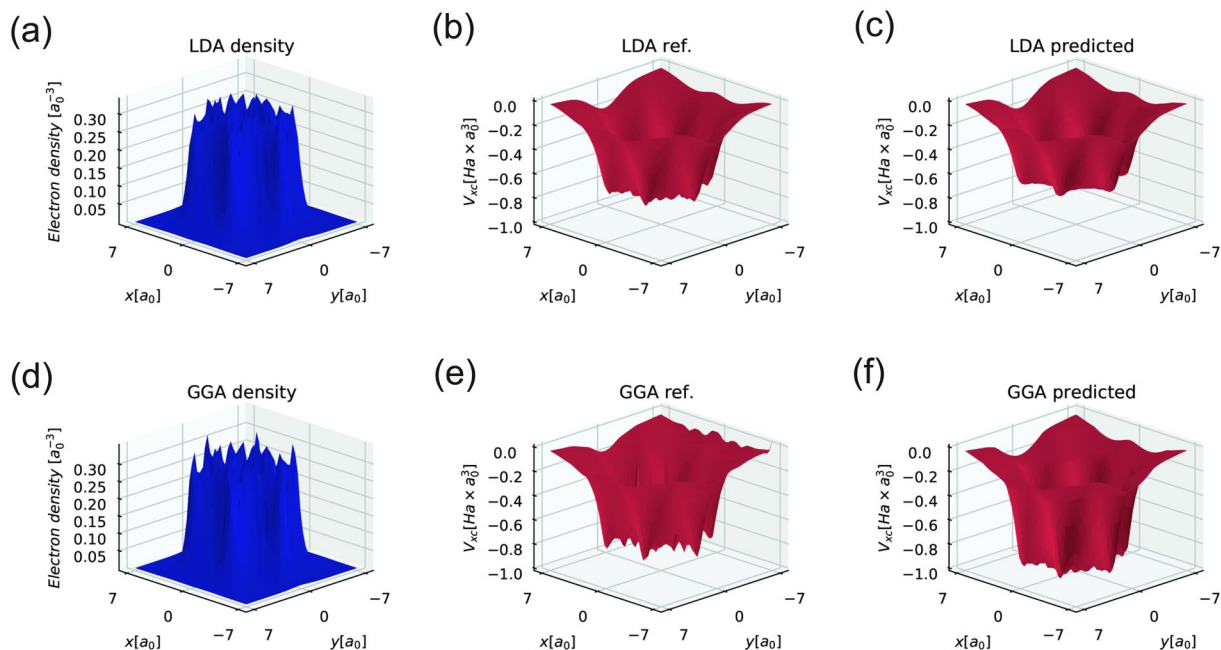
Figure 4. The results of applying NN XC functionals to the self-consistent electron density obtained from solving the Schrodinger equation in the framework of DFT for a benzene molecule. The LDA and GGA "true" V_{xc} results are procuced by the Octopus package. All presented images are slices of 3D data in the plane of the benzene molecule; (**a**) - the slice of the electron density obtained in the DFT calculation using LDA XC funtional, (**b**) - the slice of the XC potential obtained in the same as (**a**) DFT calculation using LDA XC functional, (**c**) - the slice of the NN XC potential obtained from the (**a**) electron density, (**d**) - the slice of the electron density obtained in the DFT calculation using GGA XC funtional, (**e**) - the slice of the XC potential obtained in the same as (**d**) DFT calculation using GGA XC functional, (**f**) - the slice of the NN XC potential obtained from the (**d**) electron density.
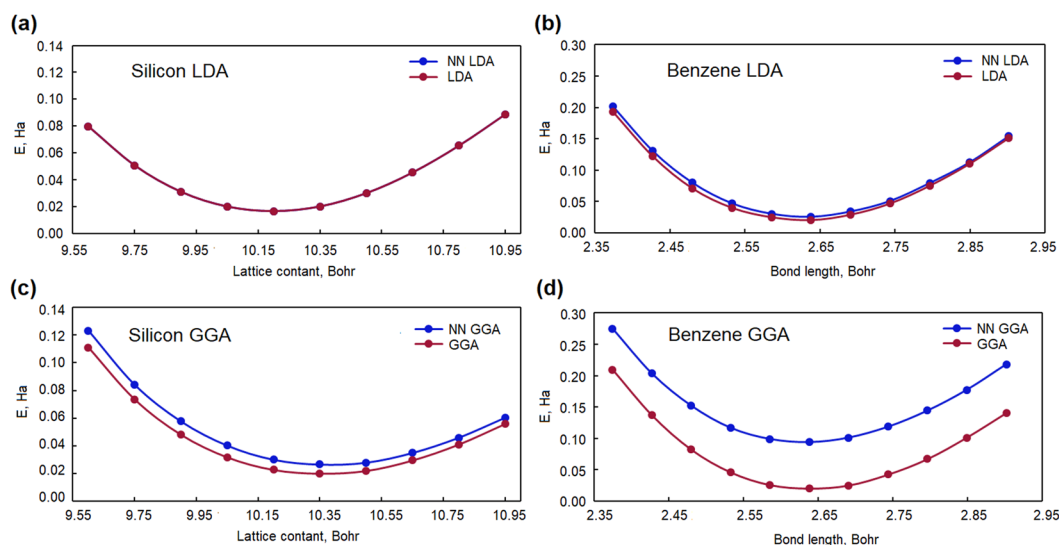
Figure 5. The results of applying NN XC potentials in self-consistent cycle. The xc-energy is calculated from original functional that was used for training NN; (**a**) - Si dependence of total energy on lattice parameter for NN LDA, (**b**) - The benzene molecule dependence of total energy on bond length for NN LDA, (**c**) - Si dependence of total energy on lattice parameter for NN GGA, (**d**) - The benzene molecule dependence of total energy on bond length for NN GGA.

original GGA xc-potential. We attribute this differences to the fact that total xc-energy was not fully consistent with the NN xc-potential. In general, obtained results for self-consistent calculations show that NN xc-potentials are converged and can be used in geometry optimization problems.

	SHO	Si	Benzene
Min Vxc, mHa × Bohr³	0.73	187	4.69
Max Vxc, mHa × Bohr³	514	495	755
MAE, mHa × Bohr³	0.03	0.6	1.36
$n(\mathbf{r})$ is in the training/test set	True	True	False
Mesh is in the training/test set	True	False	False

Table 2. Summary of results for NN LDA.

	SHO	Si	Benzene
Min Vxc, mHa × Bohr³	0	17	0.06
Max Vxc, mHa × Bohr³	553	496	804
MAE, mHa × Bohr³	0.16	18.1	18.3
$n(\mathbf{r})$ is in the training/test set	True	True	False
Mesh is in the training/test set	True	False	False

Table 3. Summary of training results for NN GGA.

Conclusions

In this work, we developed, verified and tested a neural-network approach to interpolating XC functionals. We show that the NN approach gives adequate results for both the LDA and GGA considered. When considering a model system based on three electrons placed in a parallelepiped box with an external potential of a 3D simple harmonic oscillator, the NN is in perfect agreement with numerical calculations from the test dataset. Moreover, the neural network shows a satisfactory error in real 3D physical systems not presented in the training/test data, such as silicon and benzene. Despite the lack of generalizability of the NN GGA for both systems, the analysis of those results led us to the idea of improving the NN approach by searching for a better type of coordinate system for proper representation of the input data. It is important to note that the neural network topology used in this work is very simple and has good potential for further improvements.

The main advantage of the NN approach in comparison with other interpolation techniques for XC functionals is its flexibility to incorporate exchange-correlation data from different sources, such as post-Hartree-Fock and quantum Monte Carlo. It is possible that application of the NN to interpolate high-level XC quantum data could eliminate many heuristics used in the traditional construction of XC functionals.

Data availability

Optimized NN weights and code to use it are available at https://github.com/AlexanderFreeman/MLXC. Implementation of NN potentials into OCTOPUS package is available at https://github.com/AlexanderFreeman/octopus_nn.

References

1. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical review* **136**, B864 (1964).
2. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **140**, A1133 (1965).
3. Ceperley, D. M. & Alder, B. Ground state of the electron gas by a stochastic method. *Physical Review Letters* **45**, 566 (1980).
4. Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of physics* **58**, 1200–1211 (1980).
5. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
6. Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **45**, 13244 (1992).
7. Wang, Y. & Perdew, J. P. Spin scaling of the electron-gas correlation energy in the high-density limit. *Physical Review B* **43**, 8911 (1991).
8. Perdew, J. P. *et al.* Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical review B* **46**, 6671 (1992).
9. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **77**, 3865 (1996).
10. Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics* **115**, 2315–2372 (2017).
11. Lani, G., Di Marino, S., Gerolin, A., van Leeuwen, R. & Gori-Giorgi, P. The adiabatic strictly-correlated-electrons functional: kernel and exact properties. *Physical Chemistry Chemical Physics* **18**, 21092–21101 (2016).
12. Maier, T. M., Haasler, M., Arbuznikov, A. V. & Kaupp, M. New approaches for the calibration of exchange-energy densities in local hybrid functionals. *Physical Chemistry Chemical Physics* **18**, 21133–21144 (2016).
13. Mori-Sánchez, P. & Cohen, A. J. The derivative discontinuity of the exchange–correlation functional. *Physical Chemistry Chemical Physics* **16**, 14378–14387 (2014).
14. Mori-Sánchez, P. & Cohen, A. J. Exact density functional obtained via the levy constrained search. *The journal of physical chemistry letters* **9**, 4910–4914 (2018).

15. Needs, R., Towler, M., Drummond, N. & Ríos, P. L. Continuum variational and diffusion quantum monte carlo calculations. *Journal of Physics: Condensed Matter* **22**, 023201 (2009).
16. Kolorenč, J. & Mitas, L. Applications of quantum monte carlo methods in condensed systems. *Reports on Progress in Physics* **74**, 026502 (2011).
17. Cremer, D. Møller–plesset perturbation theory: from small molecule methods to methods for thousands of atoms. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 509–530 (2011).
18. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems* **2**, 303–314 (1989).
19. Lorenz, S., Groß, A. & Scheffler, M. Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks. *Chemical Physics Letters* **395**, 210–215 (2004).
20. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
21. Balabin, R. M. & Lomakina, E. I. Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies. *The journal of chemical physics* **131**, 074104 (2009).
22. Handley, C. M. & Popelier, P. L. Potential energy surfaces fitted by artificial neural networks. *The Journal of Physical Chemistry A* **114**, 3371–3383 (2010).
23. Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Physical Chemistry Chemical Physics* **13**, 17930–17955 (2011).
24. Schütt, K. *et al*. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, 991–1001 (2017).
25. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nature communications* **8**, 13890 (2017).
26. Xie, C., Zhu, X., Yarkony, D. R. & Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. iv. coupled diabatic potential energy matrices. *The Journal of chemical physics* **149**, 144107 (2018).
27. Nagai, R., Akashi, R., Sasaki, S. & Tsuneyuki, S. Neural-network kohn-sham exchange-correlation potential and its out-of-training transferability. *The Journal of chemical physics* **148**, 241737 (2018).
28. Lei, X. & Medford, A. J. Design and analysis of machine learning exchange-correlation functionals via rotationally invariant convolutional descriptors. Physical Review. *Materials* **3**, 063801 (2019).
29. Nagai, R., Akashi, R. & Sugino, O. Completing density functional theory by machine-learning hidden messages from molecules. *arXiv preprint arXiv:1903.00238* (2019).
30. Ramos, P. & Pavanello, M. Static correlation density functional theory. *arXiv preprint arXiv:1906.06661* (2019).
31. Tozer, D. J., Ingamells, V. E. & Handy, N. C. Exchange-correlation potentials. *The Journal of chemical physics* **105**, 9200–9213 (1996).
32. Andrade, X. *et al*. Real-space grids and the octopus code as tools for the development of new simulation approaches for electronic systems. *Physical Chemistry Chemical Physics* **17**, 31371–31396 (2015).
33. Andrade, X. *et al*. Time-dependent density-functional theory in massively parallel computer architectures: the octopus project. *Journal of Physics: Condensed Matter* **24**, 233202 (2012).
34. Andrade, X. & Aspuru-Guzik, A. Real-space density functional theory on graphical processing units: computational approach and comparison to gaussian basis set methods. *Journal of chemical theory and computation* **9**, 4360–4373 (2013).
35. Dirac, P. A. M. Note on exchange phenomena in the thomas atom. *Mathematical Proceedings of the Cambridge Philosophical Society* **26**, 376–385 (1930).
36. Bloch, M. F. Bemerkung zur elektronentheorie des ferromagnetismus und der elektrischen leitˋahigkeit. *Zeitschrift für Physik* **57**, 545–555 (1929).
37. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple [phys. rev. lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **78**, 1396–1396 (1997).
38. Abadi, M. *et al*. Tensorflow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283 (2016).
39. Tang, Y. Tf. learn: Tensorflow's high-level module for distributed machine learning. *arXiv preprint arXiv:1612.04251* (2016).
40. Nair, V. & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814 (2010).

Author contributions

A.R. and P.Z. performed research. A.R. and P.Z. wrote the paper. P.Z and I.A. supervised the project. All authors reviewed the manuscript

Competing interests

The authors declare no competing interests.

Additional information