Machine-Learned Functionals for Strongly Correlated Systems

Daniel S. King,[†] Donald Truhlar,[*,‡] and Laura Gagliardi[*,¶]

[†]*Department of Chemistry, University of Chicago, Chicago IL*

[‡]*Department of Chemistry, University of Minnesota, Minneapolis MN*

[¶]*Department of Chemistry, Pritzker School of Molecular Engineering, James Franck Institute, Chicago Center for Theoretical Chemistry, University of Chicago, Chicago IL*

E-mail: truhlar@umn.edu; lgagliardi@uchicago.edu

Abstract

We introduce multiconfiguration data-driven functional theory (MC-DDFT) as a new approach to multiconfiguration nonclassical functional theory (MC-NCFT), in which the classical energy of a multiconfigurational wave function is combined with a machine-learned functional for the nonclassical exchange-correlation energy. We also present results obtained by a related approach, multiconfiguration energy-correcting functional theory (MC-ECFT), in which the total energy of a wave function method (e.g. CASSCF or NEVPT2) is corrected with a machine-learned functional. On a dataset of carbene singlet-triplet energy splittings, we demonstrate that these new multiconfiguration data-driven functional methods (MC-DDFMs) are able to achieve near-benchmark performance on systems not used for training while being less active-space dependent than multiconfiguration pair-density functional theory using currently available translated functionals. This data-driven approach appears to hold particular promise for MC-NCFT, as corrections to the CASSCF classical energy appear to be

more generalizable than corrections to total energies yielded by wave function methods such as CASSCF or NEVPT2.

Introduction

Although current Kohn-Sham density functional theory (KS-DFT) is highly accurate for many interesting chemical systems, it is well-known to be less accurate for strongly correlated systems than for systems well-described by a single Slater determinant.[1–7] This has motivated interest in combining density functionals with multiconfigurational wave function methods[8–11] (e.g. CASSCF) that explicitly express the wave function as a superposition of Slater determinants. Because multiconfigurational self-consistent-field (MCSCF) wave function methods are generally limited to an active space of orbitals and electrons that is too small to yield quantitatively accurate correlation energies, one must augment them by a post-MCSCF procedure to obtain quantitative accuracy. The most widely used of these methods include multireference perturbation theory (MRPT)[11–14] (e.g. CASPT2 and NEVPT2) and multireference configuration interaction (MRCI),[15,16] which are both very expensive.

As an alternative to MRPT and MRCI, we have proposed multiconfiguration pair-density functional theory (MC-PDFT)[17] and multiconfiguration density-coherence functional theory (MC-DCFT).[18] These methods share the feature that they compute an energy combining wave function theory for the classical components (kinetic energy, electron-nuclear attraction, and classical electron-electron interactions) with a functional for the nonclassical components of the energy (exchange and correlation), and together they may be grouped as examples of multiconfigurational nonclassical functional theory (MC-NCFT). Here we propose a new way to obtain nonclassical energy functionals, which may be called multiconfigurational data-driven functional theory (MC-DDFT).

The general MC-NCFT energy expression is given by:

$$E_{\mathrm{MC-NCFT}}[\psi^{\mathrm{MC}}] = E^{\mathrm{MC}}_{\mathrm{class}} + E_{\mathrm{nc}}[f[\psi^{\mathrm{MC}}]] \tag{1}$$

where the classical energy $\mathrm{E}^{\mathrm{MC}}_{\mathrm{class}}$ accounts for classical nucleus-nucleus repulsion, nucleus-electron attraction, electron-electron repulsion, and electron kinetic energy, and E^{MC}_{nc} is a "non-classical functional" (NCF) dependent on a featurization f of the reference wave function ψ^{MC}, which may be the density, pair density, density coherence, gradients of these quantities, or even projections of the density onto atom-centered basis functions, which is the approach pursued here.

Despite the encouragingly simple form of eq 1, all NCFs we have developed to date have been based on functional forms originally developed for KS-DFT via a process of translation [19] or conversion.[18] For example, we obtained the tPBE functional by translating the KS-DFT functional PBE;[20] the MC-PDFT method that utilizes this translation is also referred to as tPBE. Recently, neural networks have been used to develop density functionals for KS-DFT with no reference to the standard analytical forms generally utilized in functional development.[21–23] Inspired by this work, we extend this idea to nonclassical functionals in order to develop functionals specifically for use with MC-NCFT without reference to any functionals in KS-DFT.

Training Geometries. We have taken our training geometries from the QMSpin database of Schwilk et. al.[24] which contains carbenes optimized in the singlet state using CASSCF(2,2)/cc-pVDZ-F12 as well as benchmark-quality vertical singlet-triplet splittings obtained using explicitly correlated multireference configuration interaction with single and double excitations and the Davidson quadruples correction (MRCISD-F12+Q).[25–28] In this work we use a subset of these carbenes that contain only carbon and hydrogen atoms.

Network Architecture. We have taken an approach very similar to the recent work of of Dick and Fernandez in their development of NeuralXC.[21] Atomic feature vectors for atoms I are obtained by projecting the density ρ^{MC} or on-top density Π^{MC} onto atom-centered basis

functions ϕ_{nlm} via quadrature:

$$c^{I,\rho}_{nlm} = \int_{\mathbf{r}} \phi^{I}_{nlm}(\mathbf{r})\rho^{\mathrm{MC}}(\mathbf{r}) \qquad c^{I,\Pi}_{nlm} = \int_{\mathbf{r}} \phi^{I}_{nlm}(\mathbf{r})\Pi^{\mathrm{MC}}(\mathbf{r}) \tag{2}$$

and these features are then made rotationally invariant by the transformations:[21,29]

$$d^{I,\rho}_{nl} = \sum_{m} {c^{I,\rho}_{nlm}}^{2} \qquad d^{I,\Pi}_{nl} = \sum_{m} {c^{I,\rho}_{nlm}}^{2} \tag{3}$$

In this work we have opted to use the 108 optimized basis functions developed by Chen et. al. for featurization on each atom;[30] this results in a total of 36 rotationally invariant features for each atom and density ρ or Π: 12 "s" features ($l = 0$, $d^{I,\zeta}_{1,0} \dots d^{I,\zeta}_{12,0}$), 12 "p" features ($l = 1$, $d^{I,\zeta}_{2,1} \dots d^{I,\zeta}_{13,1}$), and 12 "d" features ($l = 2$, $d^{I,\zeta}_{3,2} \dots d^{I,\zeta}_{14,2}$). We then input each atomic feature vector $v_I = \{d^{I,\rho}_{nl}, d^{I,\Pi}_{nl}\}$ into its respective element network, f_{λ_I} to obtain the total energy correction:

$$E = \sum_{I} f_{\lambda_I}(v_I) \tag{4}$$

as in the work of Behler and Parrinello.[31] This approach makes the network both size-extensive and permutationally invariant.

Networks were implemented and developed in PyTorch[32] from the starting point of NeuralXC available on GitHub.[33] Element networks consist of an input layer, n_{layers} fully connected hidden layers each with n_{nodes}, and a one-node output layer, with n_{layers} and n_{nodes} treated as hyperparameters. The GELU activation function[34] was used for all nodes.

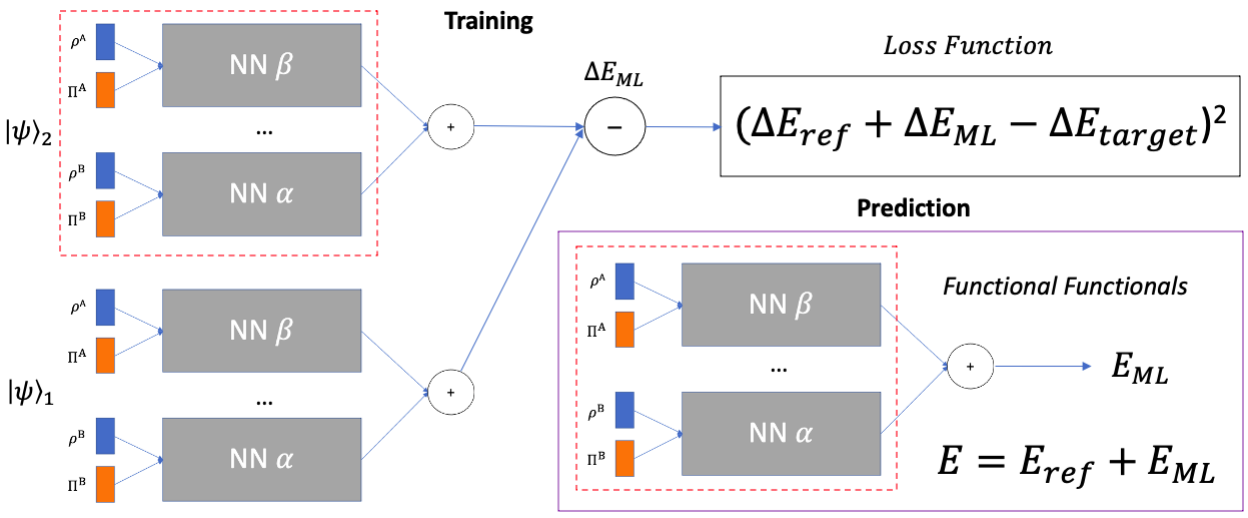

Figure 1: Network training scheme. Given a starting reference energy E_{ref} with output ΔE_{ref}, the element networks $\{\alpha, \beta, \ldots\}$ are regressed to minimize the mean squared deviation between corrected energy differences $\Delta E_{\text{ref}} + \Delta E_{\text{ML}}$ and the target energy difference ΔE_{target}.

Network Training. We focus our non-classical functionals on predicting benchmark-quality energy differences between two states $|\psi\rangle_1$ and $|\psi\rangle_2$. Given a difference in energy between these states from an inexpensive reference method, ΔE_{ref}, we train functionals to minimize the mean squared deviation between the corrected energy difference, $\Delta E_{\text{ref}} + \Delta E_{\text{target}}$, and a target energy difference, ΔE_{target} (in this work, singlet-triplet energy splittings from MRCISD-F12+Q); this training scheme is outlined in Figure 1. Although this centering of the loss function solely on relative energies stands in contrast to previous work in NeuralXC,[21] DeepKS,[22] OrbNet,[23] and KDFA,[29] it has three advantages: (i) it allows benchmark results to be obtained from a variety of different sources (including experiment, which almost always yields relative energies); (ii) relative energies are the quantities of most interest to chemists, since bond energies, energies of reaction, and barrier heights are all relative energies; and (iii) theoretical data used for training is almost always more accurate for relative energies than for absolute energies.

For optimization of parameters and hyperparameters, the 360 carbenes were split into a training set of 289 carbenes, a validation set of 37 carbenes, and a test set of 36 carbenes. All

features were normalized using a StandardScaler fit on the training set,[35] and networks were optimized to reduce the mean squared error loss over the entire training set in Pytorch using the Adam optimizer[36] with a learning rate of 0.01 for a maximum of 20001 steps. A PyTorch scheduler (torch.optim.lr_scheduler.ReduceLROnPlateau) was used to decrease the learning rate over time upon an observed plateau in the loss to a minimum learning rate of 1.1e-7, after which the training was stopped early. The hyperparameters considered were the weight decay of the Adam optimizer and the number of nodes and layers in the element networks; these hyperparameters were optimized using Optuna[37] by minimizing loss on the validation set. The final hyperparameters of all networks and the ranges explored are outlined in the Supporting Information.

Wave Function Generation. State-averaged (2,2)-CASSCF wave functions, along with tPBE and NEVPT2 energies for the singlet and triplet states of each carbene, were obtained using PySCF,[38] as integrated with MC-PDFT capabilities using publicly available development code.[39] Atomic feature vector inputs (eq 2) were obtained via quadrature using the highest grid quality (grid_level=9). During development it was found that these input features converge at significantly lower thresholds than the CASSCF energy, and therefore more stringent CASSCF optimization parameters were used in obtaining the singlet and triplet wave functions to insure consistency (mc.conv_tol = 1e-10, mc.conv_tol_grad = 1e-6, mc.ah_lindep = 1e-14, and mc.ah_conv_tol = 1e-12).

Active Space Selection. With the exception of benzene, the active spaces for all CASSCF calculations were chosen automatically using the ranked-orbital approach.[40] The highest 23 doubly occupied orbitals and the lowest 23 virtual orbitals of an ROHF wave function were individually Boys-localized[41] and the approximate pair coefficient (APC) method[40] was employed on all doubly occupied orbitals and the localized virtual orbitals to approximate orbital entropies (the remaining virtual orbitals were not considered for the active space). These entropies were then used to rank the orbitals in terms of importance, and the final active space was selected by setting a maximum number of allowed CSFs in the wave function

expansion (e.g. max(2,2), max(4,4), and max(6,7)) and dropping orbitals from the active space until the size of the active space satisfied the threshold. In the training data we selected all active spaces at the max(6,7) level.

Active Space Error. Although the ranked-orbital approach above is imperfect at ranking orbitals in importance for the active space, at the max(6,7) level our method failed to select active spaces with qualitatively accurate CASSCF excitation energies (<1eV in absolute error) in only a small number of cases; these cases were rejected from the training, validation, and test sets. However, in addition to the calculations at the max(6,7) level that were used to train the functionals, we performed some tests with minimal active spaces generated at the max(2,2) level, which requires a perfect ranking of the orbitals; in these tests we experienced a much higher failure rate (33%). Therefore, these tests were carried out on a test subset of only 24 carbenes (listed in the Supporting Information).

Types of Data-Driven Functionals. We have developed three types of functionals using the above scheme:

1) Purely data-driven functionals (PDDFs), which correct the classical energy of a wave function:

$$E_{\mathrm{MC-PDDFT}} = E^{\mathrm{MC}}_{\mathrm{class}} + E_{\mathrm{PDD}}[\rho^{\mathrm{MC}}, \Pi^{\mathrm{MC}}] \tag{5}$$

2) Δ-data-driven functionals (ΔDDFs), which correct MC-NCFT calcuations that employ translated or converted functionals from KS-DFT, e.g. tPBE:

$$E_{\mathrm{MC-\Delta DDFT}} = E^{\mathrm{MC}}_{\mathrm{tPBE}} + E_{\mathrm{\Delta DD}}[\rho^{\mathrm{MC}}, \Pi^{\mathrm{MC}}] \tag{6}$$

3) Energy-correcting data-driven functionals (ECDDFs), which correct the total energy of a wave function theory:

$$E_{\mathrm{MC-ECFT}} = E_{\mathrm{tot}}^{\mathrm{MC}} + E_{\Delta\mathrm{DD}}[\rho^{\mathrm{MC}}, \Pi^{\mathrm{MC}}] \tag{7}$$

Below, we present results on the four new functionals we have developed in these categories: DDF21, a purely data-driven functional; ΔtPBE-21, a Δ-data-driven functional that corrects tPBE; ΔCASSCF-21, an energy-correcting functional that corrects CASSCF; and ΔNEVPT2-21, an energy-correcting functional that corrects NEVPT2. All of these functionals can be referred to collectively as "data-driven functionals" (DDFs). When used in combination with their reference energies they form multiconfiguration data-driven functional methods (MC-DDFMs).

Results

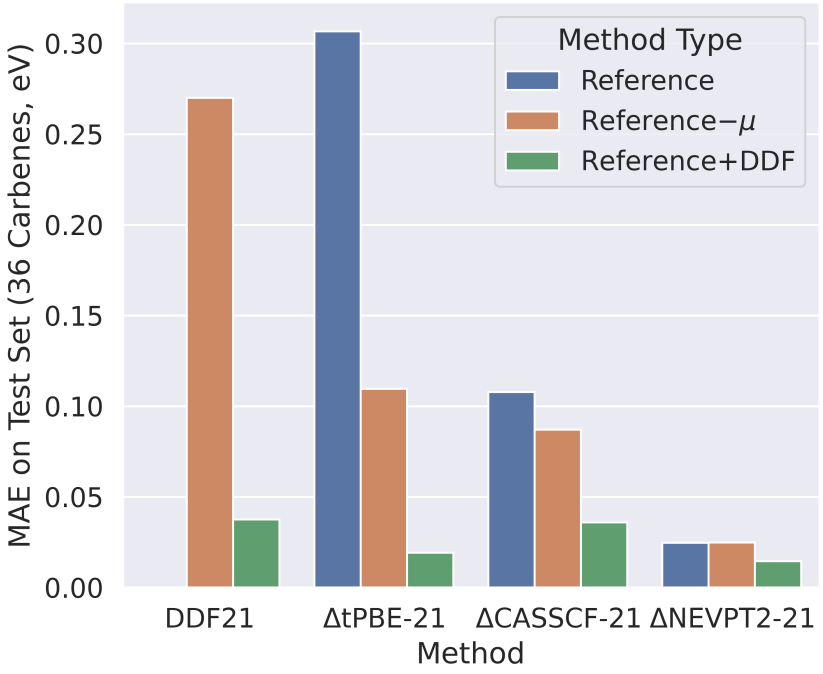

Figure 2: Mean absolute errors (MAEs) on MRCISD-F12+Q benchmark data for a test set of 36 carbenes excluded from the training data. For each MC-DDFM (DDF21, ΔtPBE-21, ΔCASSCF-21, and ΔNEVPT2-21, shown in green), we show the performance of its reference method (tPBE, CASSCF, and NEVPT2, shown in blue) as well as a one-parameter mean-corrected method (Reference-μ) shown in orange. The MAE of the classical CASSCF energy (1.1eV) is not shown due to scale.

Results. Figure 2 shows the performance of the four MC-DDFMs in comparison to their respective reference methods on the test set of 36 carbene singlet-triplet energy splittings. For comparison, we also show the performance of a simple one-parameter mean correction to the singlet-triplet energy splittings, in which ΔE_{ref} is corrected by its mean deviation from MRCISD-F12+Q on the training data. Encouragingly, all four functionals are able to greatly improve upon these one parameter corrections, surpassing the mean absolute errors (MAEs) of their reference methods by factors of 29 (DDF21), 16 (ΔtPBE21), 3 (ΔCASSCF-21), and 2 (ΔNEVPT2-21). Although all the data in the article proper are for functionals that depend on both the density and the pair density, additional results given in the Supporting

Information show that we obtain similarly high accuracy using only density features or only on-top density features.

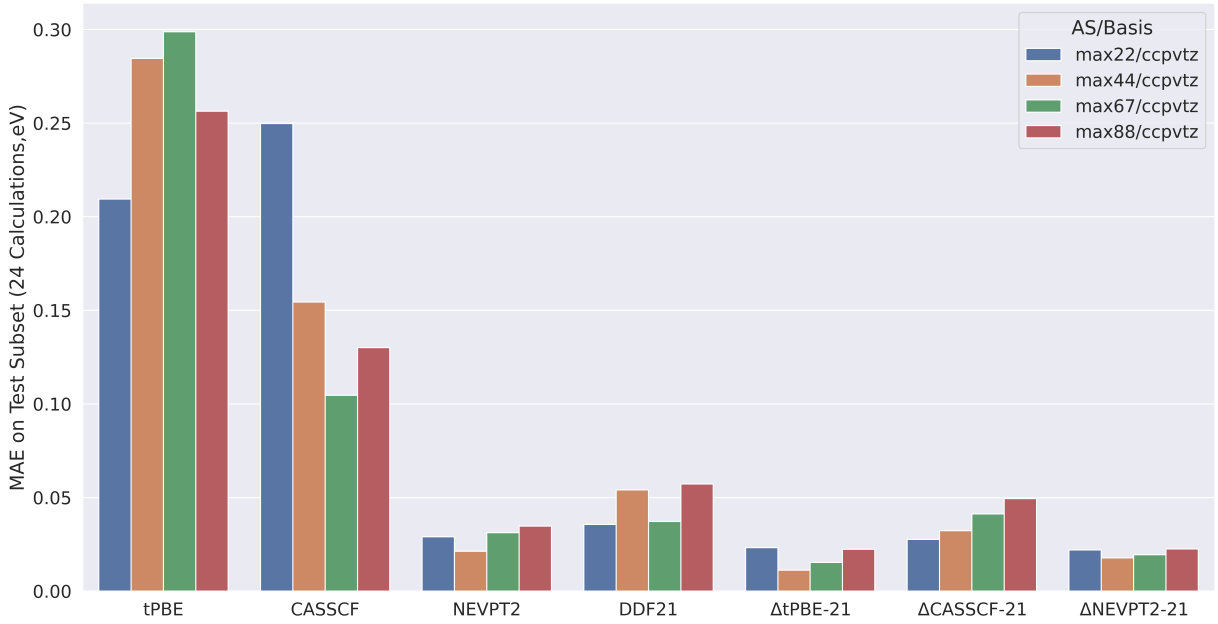

Figure 3: Mean absolute errors on MRCISD-F12+Q benchmark data from a test subset of 24 carbenes for which our automated scheme chose a reasonable (2,2) active space, tested with the cc-PVTZ basis at four different active space sizes: max(2,2), max(4,4), max(6,7), and max(8,8).

We tested the active space dependence of our data-driven functional methods on 24 carbenes with four different active spaces whose number of configurations vary by four orders of magnitude: max(2,2), max(4,4), max(6,7), and max(8,8). Figure 3 shows that all MC-DDFMs maintain their near-benchmark accuracy across this wide range of active spaces, despite being trained on only max(6,7) active spaces. We note that this active space robustness is likely a benefit of the sole dependence of the loss function on relative energies rather than absolute ones. However, we find that one drawback of our approach is that the parameters do not seem to be easily transferable to other basis sets; when switching to either a cc-pVDZ or cc-pVQZ basis the errors of the MC-DDFMs increase dramatically (Supporting Information).

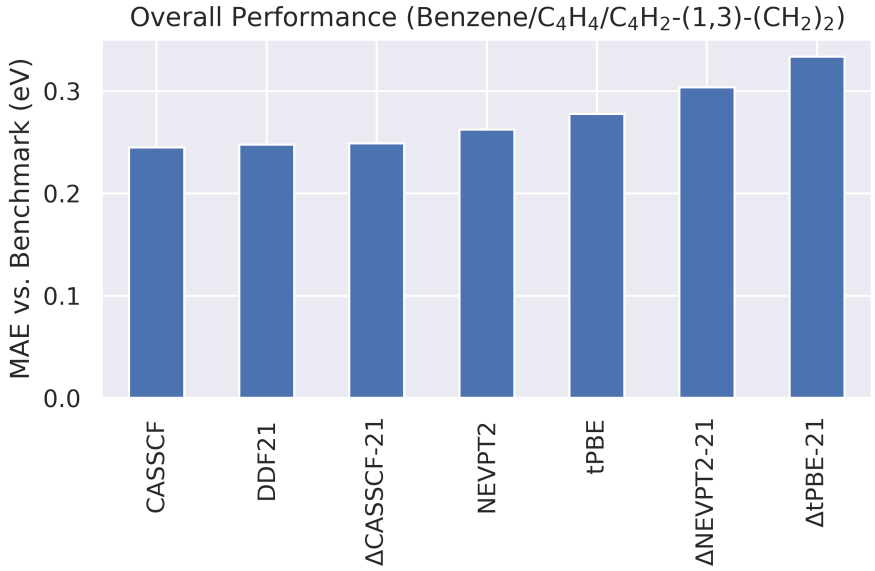

Figure 4: Mean absolute error of reference and data-driven functional methods on three difficult singlet triplet gaps, consisting of one aryl system (C_6H_6, using the standard minimal cc-pVTZ@UNO-(6,6) active space[42]) and two biradical systems (cyclobutadiene, C_4H_4, and 1,3-bis(methylene)-cyclobutadiene (C_4H_2-(1,3)-$(CH_2)_2$), using automatically selected max(10,10) active spaces).

As a final test of generalizability, we tested the MC-DDFMs on three difficult singlet–triplet energy splittings quite different than any data in the training set: benzene and two biradical systems; cyclobutadiene (C_4H_4) and 1,3-bis(methylene)cyclobutadiene (C_4H_2-(1,3)-$(CH_2)_2$) (Figure 4). These systems were taken from previous benchmark studies on translated functionals,[43,44] with benchmarks for benzene taken from experiment[45] and benchmarks for the biradicals from theoretical results.[46] While MC-DDFMs correcting total energies (ΔtPBE-21, ΔCASSCF-21, and ΔCASSCF-21) all did worse on average than their respective reference methods, DDF21 maintains a large improvement upon the CASSCF classical energy, reducing its MAE from 0.77 eV (unshown in Figure 4) to only 0.25 eV. This suggests to us that corrections to the classical energy may be more generalizable than corrections to "complete" methods such as CASSCF or NEVPT2, and similar generalizability in this regard is achieved by MC-DDFMs trained solely on the density or on-top density (Supporting Information).

Conclusions. We have presented a data-driven approach to the development of energy functionals for strongly correlated systems utilizing neural networks parametrized in terms of the density and on-top density. Using a dataset of carbene singlet-triplet energy splittings taken from the QMSpin database,[24] we find that the new multiconfiguraional data-driven functional methods (MC-DDFMs) are able to achieve benchmark-quality accuracy on carbenes not included in the training set and improve markedly on approaches using translated functionals even when extended to different active spaces. Furthermore, this work shows that data-driven functionals hold particularly great promise for multiconfigurational non-classical functional theory, as corrections to the classical energy appear to be more generalizable than corrections to total energies such as yielded by CASSCF or NEVPT2. It will be interesting to see if this good performance can be maintained when the functionals are parameterized using larger and more diverse sets of training data.

Acknowledgment

The authors thank the Inorganometallic Catalyst Design Center (ICDC) under DOE award DE-SC0012702. Additionally, the authors thank the Research Computing Center (RCC) at the University of Chicago for access to computational resources.

Supporting Information Available

Names of the carbenes included in the training, validation, test, and test subsets. Histograms showing systematic errors in reference methods. Basis set dependence of MC-DDFMs. Performance and active space and basis set dependence of MC-DDFMs trained solely on density features or solely on on-top density features. Individual performances on benzene, cyclobutadiene, and 1,3-bis(methylene)-cyclobutadiene.

References

(1) Kohn, W.; Becke, A. D.; Parr, R. G. Density Functional Theory of Electronic Structure. *J. Phys. Chem.* **1996**, *100*, 12974–12980.

(2) Scuseria, G. E.; Staroverov, V. N. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 669–724.

(3) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. Insights Into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792–794.

(4) Cramer, C. J.; Truhlar, D. G. Density Functional Theory for Transition Metals and Transition Metal Chemistry. *Phys. Chem. Chem. Phys* **2009**, *11*, 10757.

(5) Becke, A. D. Perspective: Fifty Years of Density-Functional Theory in Chemical Physics. *J. Chem. Phys.* **2014**, *140*, 18A301.

(6) Yu, H. S.; Li, S. L.; Truhlar, D. G. Perspective: Kohn-Sham Density Functional Theory Descending a Staircase. *J. Chem. Phys.* **2016**, *145*, 130901.

(7) Verma, P.; Truhlar, D. G. Status and Challenges of Density Functional Theory. *Trends Chem.* **2020**, *2*, 302–318.

(8) Nakano, H.; Nakajima, T.; Tsuneda, T.; Hirao, K. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 507–557.

(9) Roos, B. O. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 725–764.

(10) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 1167–1189.

(11) Gaggioli, C. A.; Stoneburner, S. J.; Cramer, C. J.; Gagliardi, L. Beyond Density Functional Theory: The Multiconfigurational Approach To Model Heterogeneous Catalysis. *ACS Catal.* **2019**, *9*, 8481–8502.

(12) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-Order Perturbation Theory With a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(13) Hirao, K. Multireference Møller—Plesset Method. *Chem. Phys. Lett.* **1992**, *190*, 374–380.

(14) Angeli, C.; Cimiraglia, R.; Evangelisti, S.; Leininger, T.; Malrieu, J.-P. Introduction of n -Electron Valence States for Multireference Perturbation Theory. *J. Chem. Phys.* **2001**, *114*, 10252–10264.

(15) Brown, F. B.; Shavitt, I.; Shepard, R. Multireference Configuration Interaction Treatment of Potential Energy Surfaces: Symmetric Dissociation of H2O in a Double-Zeta Basis. *Chem. Phys. Lett.* **1984**, *105*, 363–369.

(16) Werner, H.; Knowles, P. J. An Efficient Internally Contracted Multiconfiguration–reference Configuration Interaction Method. *J. Chem. Phys.* **1988**, *89*, 5803–5814.

(17) Li Manni, G.; Carlson, R. K.; Luo, S.; Ma, D.; Olsen, J.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration Pair-Density Functional Theory. *J. Chem. Theory Comput.* **2014**, *10*, 3669–3680.

(18) Zhang, D.; Hermes, M. R.; Gagliardi, L.; Truhlar, D. G. Multiconfiguration Density-Coherence Functional Theory. *J. Chem. Theory Comput.* **2021**, *17*, 2775–2782.

(19) Li Manni, G.; Smart, S. D.; Alavi, A. Combining the Complete Active Space Self-Consistent Field Method and the Full Configuration Interaction Quantum Monte Carlo

Within a Super-Ci Framework, With Application to Challenging Metal-Porphyrins. *J. Chem. Theory Comput.* **2016**, *12*, 1245–1258.

(20) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(21) Dick, S.; Fernandez-Serra, M. Machine Learning Accurate Exchange and Correlation Functionals of the Electronic Density. *Nat. Commun.* **2020**, *11*.

(22) Chen, Y.; Zhang, L.; Wang, H.; E, W. DeePKS: A Comprehensive Data-Driven Approach Toward Chemically Accurate Density Functional Theory. *J. Chem. Theory Comput.* **2021**, *17*, 170–181.

(23) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. OrbNet: Deep Learning for Quantum Chemistry Using Symmetry-Adapted Atomic-Orbital Features. *J. Chem. Phys.* **2020**, *153*, 124111.

(24) Schwilk, M.; Tahchieva, D. N.; von Lilienfeld, O. A. Large Yet Bounded: Spin Gap Ranges in Carbenes. *arXiv:2004.10600 [physics]* **2020**, arXiv: 2004.10600.

(25) Werner, H.-J.; Knowles, P. J. An Efficient Internally Contracted Multiconfiguration–Reference Configuration Interaction Method. *J. Chem. Phys.* **1988**, *89*, 5803–5814.

(26) Knowles, P. J.; Werner, H.-J. An Efficient Method for the Evaluation of Coupling Coefficients in Configuration Interaction Calculations. *Chem. Phys. Lett.* **1988**, *145*, 514–522.

(27) Knowles, P. J.; Werner, H.-J. Internally Contracted Multiconfiguration-Reference Configuration Interaction Calculations for Excited States. *Theor. Chim. Acta.* **1992**, *84*, 95–103.

(28) Shiozaki, T.; Knizia, G.; Werner, H.-J. Explicitly Correlated Multireference Configuration Interaction: MRCI-F12. *J. Chem. Phys.* **2011**, *134*, 034113.

(29) Margraf, J. T.; Reuter, K. Pure Non-Local Machine-Learned Density Functional Theory for Electron Correlation. *Nat. Commun.* **2021**, *12*.

(30) Chen, Y.; Zhang, L.; Wang, H.; E, W. Ground State Energy Functional With Hartree–Fock Efficiency and Chemical Accuracy. *J. Phys. Chem. A* **2020**, *124*, 7155–7165.

(31) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*.

(32) Paszke, A. et al. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc., 2019; pp 8024–8035.

(33) Dick, S. Semodi/Neuralxc. `https://github.com/semodi/neuralxc`, accessed near 04.15.2021.

(34) Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (Gelus). *arXiv preprint arXiv:1606.08415* **2016**,

(35) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(36) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980* **2014**,

(37) Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA, 2019; pp 2623–2631.

(38) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K. PySCF: The Python-based Simulations of Chemistry Framework. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*.

(39) MatthewRHermes/mrh. `https://github.com/MatthewRHermes/mrh`, accessed near 04.15.2021.

(40) King, D. S.; Gagliardi, L. A Ranked-Orbital Approach to Select Active Spaces for High-Throughput Multireference Computation. *J. Chem. Theory Comput.* **2021**, *17*, 2817–2831.

(41) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.

(42) Tóth, Z.; Pulay, P. Finding Symmetry Breaking Hartree-Fock Solutions: The Case of Triplet Instability. *J. Chem. Phys.* **2016**, *145*, 164102.

(43) Sharma, P.; Bernales, V.; Truhlar, D. G.; Gagliardi, L. Valence ΠΠ* Excitations in Benzene Studied by Multiconfiguration Pair-Density Functional Theory. *J. Phys. Chem. Lett.* **2019**, *10*, 75–81.

(44) Stoneburner, S. J.; Truhlar, D. G.; Gagliardi, L. MC-PDFT Can Calculate Singlet–triplet Splittings of Organic Diradicals. *J. Chem. Phys.* **2018**, *148*, 064108.

(45) Doering, J. P. Low-Energy Electron-Impact Study of the First, Second, and Third Triplet States of Benzene. *J. Chem. Phys.* **1969**, *51*, 2866–2870.

(46) Stoneburner, S. J.; Shen, J.; Ajala, A. O.; Piecuch, P.; Truhlar, D. G.; Gagliardi, L. Systematic Design of Active Spaces for Multi-Reference Calculations of Singlet–triplet Gaps of Organic Diradicals, With Benchmarks Against Doubly Electron-Attached Coupled-Cluster Data. *J. Chem. Phys.* **2017**, *147*, 164120.