TECHNICAL COMMENT

Comment on "Pushing the frontiers of density functionals by solving the fractional electron problem"

Igor S. Gerasimov[1]*, Timofey V. Losev[2,3], Evgeny Yu. Epifanov[2,4], Irina Rudenko[5,6], Ivan S. Bushmarinov[5]*, Alexander A. Ryabov[6,7]*, Petr A. Zhilyaev[7]*, Michael G. Medvedev[2,4]*

Kirkpatrick *et al.* (Reports, 9 December 2021, p. 1385) trained a neural network–based DFT functional, DM21, on fractional-charge (FC) and fractional-spin (FS) systems, and they claim that it has outstanding accuracy for chemical systems exhibiting strong correlation. Here, we show that the ability of DM21 to generalize the behavior of such systems does not follow from the published results and requires revisiting.

The main conclusion of Kirkpatrick *et al.* (*1*) is that the inclusion of FC and FS systems in a training set of a sufficiently flexible density functional allows it to become outstandingly accurate even for systems with FC/FS character, on which density functional theory (DFT) usually fails. To prove this, the authors constructed a local hybrid functional with learnable (neural network–based) enhancement factors and trained it to simultaneously approximate the energies and Kohn-Sham orbitals of the target systems. The training set consists of 1161 reaction/molecular energies, 1074 energies of FC/FS systems, and 931 sets of reasonable Kohn-Sham orbitals. The resulting functional, named DM21 by the authors, was found to be much more accurate than four functionals—SCAN-D3(BJ) (*2*, *3*), ωB97X-V (*4*), M06-2X (*5*), PW6B95-D3(0) (*6*)—commonly used on bond-breaking benchmark (BBB), GMTKN55 (*7*), and QM9 (*8*) test datasets. The authors then showed that if they exclude the FC/FS systems from the training set, the resulting functional (DM21m) becomes much less accurate for the BBB dataset but somewhat more accurate for GMTKN55 and QM9 ones.

Although the conclusion of Kirkpatrick *et al.* about the role of FC/FS systems in the training set may be correct, it is not the only possible explanation for their observations. In our opinion, the improvements in the performance of DM21 on the BBB test dataset relative to DM21m may be caused by a much more prosaic reason: an unintended overlap between the training and test datasets. Indeed, whereas the BBB dataset is composed of the dissociation curves of H_2, Li_2, C_2, N_2, F_2, H_2^+, He_2^+, Li_2^+, B_2^+, Ne_2^+, and Al_2^+ systems, their dissociated counterparts [the most problematic cases for DFT (*9*)] were used to train the DM21 (but not DM21m) functional in forms of fractional-spin and fractional-charge atomics in the FS/FC training sets (Fig. 1A). Notably, the dimers' densities constructed from integer-occupied orbitals already converge into densities of fractional-orbital atomics at intermediate separation (Fig. 1B). Thus, the "flat" region of the dissociation curve has wave function and energy very similar to a pair of individual FS/FC atomics (Fig. 1B), so we can conclude that the nontrivial part of the BBB dataset had a very strong representation in the training set.

This situation highlights the difference between testing a model in physics and in machine learning. In physics, the performance of a model on a test dataset shows how well it represents reality; on the contrary, in machine learning, where a model is trained on some data ("training set"), its performance on a test dataset shows the model's ability to generalize—that is, to understand versus to memorize. Because the machine learning models generally tend to be better at "memorizing" (*10*), it is crucial to test one on systems unlike those used to train it (*11*, *12*); otherwise, the so-called "data leakage" can artificially inflate the performance metrics. Although Kirkpatrick *et al.*'s test dataset has only a small overall overlap with the training one [figure S2 of (*1*)], the BBB subset implicitly overlaps with the training set by ~50% (Fig. 1) and therefore can hardly be considered a fair test. In our opinion, the DM21 performance on it mainly illustrates the DM21 functional's ability to memorize the FS/FC systems seen during training, instead of its ability to generalize the behavior of such systems.

Apart from the BBB dataset, Kirkpatrick *et al.* have also demonstrated good performance of DM21 on other systems where high FS/FC character is expected; however, all these tests do not allow us to make a clear-cut conclusion regarding the "understanding" of the FS/FC systems by DM21:

Adenine-thymine base pair: For this system, the "oracle" used by the authors seems to be incorrect, because it completely neglects hydrogen bonding, which would channel electron density between the fragments, inducing
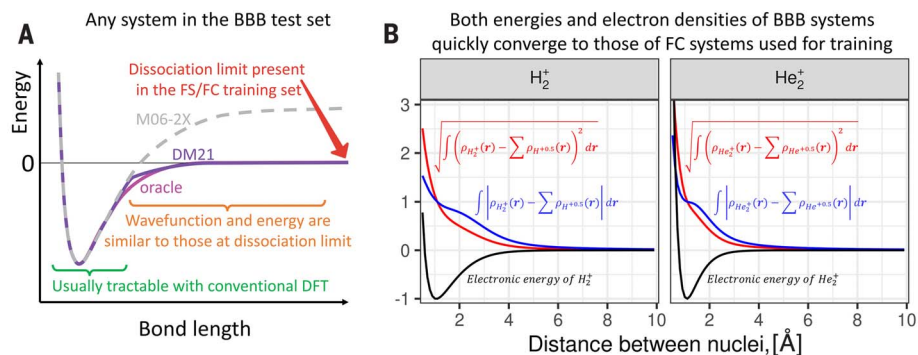
[1]Department of Chemistry, Kyungpook National University, Daegu 41566, South Korea. [2]N. D. Zelinsky Institute of Organic Chemistry of Russian Academy of Sciences, 119991 Moscow, Russian Federation. [3]Department of Chemistry, Lomonosov Moscow State University, 119991 Moscow, Russian Federation. [4]National Research University Higher School of Economics, 101000 Moscow, Russian Federation. [5]Yandex, 119021 Moscow, Russian Federation. [6]Moscow Institute of Physics and Technology (State University), Dolgoprudny, Moscow Region 141700, Russian Federation. [7]Center for Materials Technologies, Skolkovo Institute of Science and Technology, Moscow 143026, Russian Federation.
*Corresponding author. Email: i.s.ger@yandex.ru (I.S.G.); bush-i@yandex.ru (I.S.B.); ryabov.alexandr@phystech.edu (A.A.R.); p.zhilyaev@skoltech.ru (P.A.Z.); medvedev.m.g@gmail.com (M.G.M.)
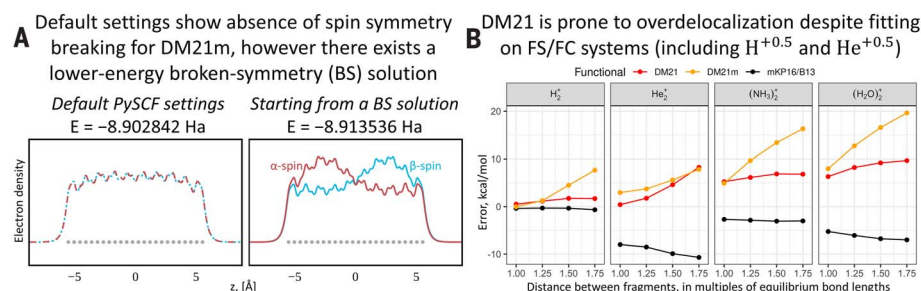

Fig. 1. Leakage of the FS/FC systems into the BBB test dataset. (**A**) A typical dissociation curve for a system in the BBB test dataset. The dissociation limit corresponds to a pair of FC or FS systems, both of which were used to train DM21 (but not DM21m). The "bonding" and "flat" regions are marked with green and orange braces, respectively. (**B**) Resemblance between electron densities of dimeric cations with integer orbital occupations and superpositions of FC atomics; shown electron density difference measures (*15*) are integral of squared errors (red) and integral of absolute errors (blue). For clarity, all functions are normed by absolute values at the equilibrium geometry. Calculations are performed with HF for H, He⁺, and H_2^+, and CCSD for He and He_2^+, using the aug-cc-pV5Z basis set.

A
Default settings show absence of spin symmetry breaking for DM21m, however there exists a lower-energy broken-symmetry (BS) solution

Default PySCF settings
E = −8.902842 Ha

Starting from a BS solution
E = −8.913536 Ha

α-spin   β-spin

Electron density

z, [Å]

B
DM21 is prone to overdelocalization despite fitting on FS/FC systems (including $H^{+0.5}$ and $He^{+0.5}$)

Functional ← DM21 ← DM21m ← mKP16/B13

H_2^+ He_2^+ $(NH_3)_2^+$ $(H_2O)_2^+$

Error, kcal/mol

Distance between fragments, in multiples of equilibrium bond lengths

Fig. 2. Behavior of DM21 family functionals. (**A**) Symmetry-preserved and broken-symmetry solutions of DM21m for the H_{24} system. (**B**) Errors of DM21, DM21m, and mKP16/B13 on the systems from the SIE4x4 dataset. Data are from (*1*) and (*14*).

some charge delocalization even in the exact solution. Thus, we do not know the correct extent of charge delocalization between the two molecules. Furthermore, DM21 was not compared to DM21m on this system.

Compressed chain of hydrogen atoms: This is the most controversial example because Kirkpatrick *et al.* do not provide any details on their claim that "high-level wave function methods did not yield spin polarized solutions." As far as we know, there is no solid evidence that a finite chain of hydrogen atoms cannot have a spin symmetry–broken Kohn-Sham wave function. Thus, the presented behavior of DM21 only shows that it is different from the other functionals, not that it is superior to them.

Moreover, the very fact of spin-symmetric wave function for DM21 needs further verification. We could not converge DM21 for H_{24} using the basis set and code published by Kirkpatrick *et al.*; however, for DM21m we have located both symmetry-preserved and broken-symmetry (BS) solutions, depending on the starting guess (Fig. 2A). Notably, the default settings resulted in the symmetry-preserved solution, which is 44.4 kcal/mol less stable than the BS one. This result suggests that the symmetry-preserved solution located by Kirkpatrick *et al.* for DM21 may be simply a wrong self-consistent field stationary point and, therefore, that a wave function stability check is required to prove that DM21 has no lower-energy BS solutions.

Bicyclobutane reaction barriers: Here, DM21 outperforms DM21m on two conrotatory transition states (TSs); however, the functionals' performances on the disrotatory TSs are very close [table S4 of (*1*)]. This trend is intriguing because it is the disrotatory route, which has high nondynamical correlation (*13*); however, DM21m has no problems with it despite the absence of training on FS/FC systems.

SIE4x4 dataset from the GMTKN55: This was designed to test functional accuracies on FC systems (where self-interaction error is most pronounced) and contains four systems, two of which (H_2^+ and He_2^+) are also present in the BBB dataset (for which dissociation limits were used to train DM21; see above). DM21 outperforms DM21m on all systems from this dataset (Fig. 2B); however, it shows a clear trend of increasing the errors with distance and is outranked by another e^{HF}-dependent mKP16/B13 functional (*14*), which was published by Proynov and Kong in July 2021.

The analysis above shows that the ability of DM21 to generalize the behavior of FS/FC systems does not follow from the available evidence and is yet to be determined. To prove that DM21 has indeed generalized (not memorized) the behavior of FS/FC systems, authors might construct an independent test dataset and compare performances (that is, generalization abilities) of DM21 and DM21m on it.

At the same time, the general idea of introducing exact constraints through a training dataset is supported by the DM21mu functional (differing from DM21m only by the inclusion of the uniform electron gas in the training dataset) strongly outperforming DM21m on all three subsets of the test dataset [figure S4 of (*1*)]. This result is nontrivial by itself, vividly demonstrating the importance of the uniform electron gas constraint for constructing neural network–based DFT functionals.

To conclude, the problem of generalizing the behavior of fractional-electron systems with a neural network–based DFT functional seems to remain open. However, the overall approach of imposing exact constraints on a learned functional via training data is likely to play an important role in the future of density functional theory.

REFERENCES AND NOTES

1. J. Kirkpatrick *et al.*, *Science* **374**, 1385–1389 (2021).
2. J. Sun, A. Ruzsinszky, J. P. Perdew, *Phys. Rev. Lett.* **115**, 036402 (2015).
3. J. Sun *et al.*, *Nat. Chem.* **8**, 831–836 (2016).
4. N. Mardirossian, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **16**, 9904–9924 (2014).
5. Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **120**, 215–241 (2007).
6. Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **109**, 5656–5667 (2005).
7. L. Goerigk *et al.*, *Phys. Chem. Chem. Phys.* **19**, 32184–32215 (2017).
8. R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *Sci. Data* **1**, 140022 (2014).
9. A. J. Cohen, P. Mori-Sánchez, W. Yang, *Science* **321**, 792–794 (2008).
10. V. Feldman, C. Zhang, *Adv. Neur. Inf. Process. Syst.* **33**, 2881–2891 (2020).
11. R. K. Samala, H.-P. Chan, L. Hadjiiski, S. Koneru, *Proc. SPIE* **11314**, 279–284 (2020).
12. S. Kaufman, S. Rosset, C. Perlich, O. Stitelman, *ACM Trans. Knowl. Discov. Data* **6**, 1–21 (2012).
13. R. Berner, A. Lüchow, *J. Phys. Chem. A* **114**, 13222–13227 (2010).
14. E. Proynov, J. Kong, *J. Chem. Theory Comput.* **17**, 4633–4638 (2021).
15. A. A. Marjewski *et al.*, *Mendeleev Commun.* **28**, 225–235 (2018).

Science

Comment on "Pushing the frontiers of density functionals by solving the fractional electron problem"

Igor S. GerasimovTimofey V. LosevEvgeny Yu. EpifanovIrina RudenkoIvan S. BushmarinovAlexander A. RyabovPetr A. ZhilyaevMichael G. Medvedev

View the article online
https://www.science.org/doi/10.1126/science.abq3385
Permissions
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service