

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Developing Density Functional Theory with Physical Prior Knowledge

Permalink

<https://escholarship.org/uc/item/3gh1v4h5>

Author

Pederson, Ryan

Publication Date

2023

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Developing Density Functional Theory with Physical Prior Knowledge

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Physics

by

Ryan Pederson

Dissertation Committee:
Professor Kieron Burke, Chair
Professor Steven R. White
Professor Ruqian Wu

2023

© 2023 Ryan Pederson

DEDICATION

To Lauren, my parents, and my brother.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	vi
LIST OF TABLES	x
ACKNOWLEDGMENTS	xi
VITA	xiii
ABSTRACT OF THE DISSERTATION	xvi
1 Introduction	1
2 Bypassing the energy functional in density functional theory: Direct calculation of electronic energies from conditional probability densities	4
2.1 Abstract	4
2.2 Introduction	5
2.3 Theory	7
2.4 Approximations	10
2.5 Finite temperatures	15
2.6 Conclusion	17
2.7 Acknowledgements	18
3 Conditional probability density functional theory	19
3.1 Abstract	19
3.2 Introduction	20
3.3 Background and Notation	23
3.3.1 DFT fundamentals	23
3.3.2 Adiabatic connection	25
3.3.3 Pair densities and XC holes	26
3.4 Theory	28
3.4.1 CP-DFT	28
3.4.2 Spin-adapted CP-DFT	34
3.4.3 CP-DFT with averaged quantities	37
3.4.4 connection to exact factorization?	38
3.5 Exact conditions	41

3.5.1	Two electron spin-singlet systems	41
3.5.2	Cusp condition	42
3.5.3	Long-range limits	43
3.5.4	Strictly correlated electron limit	46
3.5.5	Dissociation limit	48
3.6	Blue electron approximation	50
3.7	Conclusions	58
3.8	Acknowledgments	59
4	Reassessing the role of exact conditions in density functional theory	60
4.1	Abstract	60
4.2	Introduction	61
4.3	Results	63
4.4	Conclusion	73
4.5	Acknowledgments	73
5	Kohn-Sham equations as regularizer: building prior knowledge into machine-learned physics	74
5.1	Abstract	74
5.2	Introduction	75
5.3	Results	78
5.4	Conclusion	82
5.5	Acknowledgments	83
6	How Well Does Kohn-Sham Regularizer Work for Weakly Correlated Systems?	87
6.1	Abstract	87
6.2	Introduction	88
6.3	Results	91
6.4	Conclusion	102
6.5	Acknowledgments	103
7	Large scale quantum chemistry with tensor processing units	104
7.1	Abstract	104
7.2	Introduction	105
7.3	Results	111
7.3.1	DFT with TPUs	111
7.3.2	Dynamic precision on TPUs	114
7.4	Conclusion	116
7.5	Acknowledgments	119
Bibliography		120
Appendix A	Supplemental Info for Chapter 4	157
Appendix B	Supplemental Info for Chapter 5	179

LIST OF FIGURES

	Page
2.1 CP (blue) and exact (black): (a) XC energy per particle in uniform gas at increasing Wigner–Seitz radii (r_s) and $T = 0$, (b) binding energy curve for H_2 (red is KS-DFT using PBE [313]), and (c) XC free energy per particle at $r_s = 1$ as a function of reduced temperature (T_F is the Fermi temperature). Exact from Ref. [331] in (a), Ref. [155] in (c). Hartree atomic units used throughout.	7
2.2 Pure blue electron approximation (purple), half pure blue approximation (orange), interpolation (Eq. 2.14, blue), and exact (black): (a) normalized XC hole densities for the uniform gas at $r_s = 1$ with exact parameterization from Ref. [332], (b) $U_c(R)$ from H_2 , using exact $n(r)$ and producing errors below 20%, (c) Hooke’s atom $\tilde{v}_s(\mathbf{r}' \mathbf{r})$, with the external potential $r'^2/8$ (black dashed), and (d) Hooke’s atom $\tilde{n}_\mathbf{r}(r')$	12
2.3 Percentage error of uniform gas potential XC free energy per electron for the CP-DFT calculations within KS (blue), TF (purple), and classical (green) approximations relative to the parameterization of Groth et al. [155].	17
3.1 Exact CP densities and potentials in 1D He (left) and 1D Be (right): $\tilde{v}_{s,y}(y')$ is the CP-KS potential with corresponding CP density $\tilde{n}_y(y')$, $v_s(y')$ the KS potential with corresponding ground-state density $n(y')$. Quantities are plotted for reference positions $y = 0$ and $y = 0.8$	28
3.2 Exact CP densities and potentials in 1D Li with $N_\uparrow = 2, N_\downarrow = 1$. The CP-KS potential is $\tilde{v}_{s,y}(y')$ with corresponding total CP density $\tilde{n}_y(y')$ (dashed), the spin-CP-KS potential is $\tilde{v}_{s,x}(y')$ with corresponding spin-CP density $\tilde{n}_x(y')$, $v_{s,\sigma}(y')$ is the KS spin potential with corresponding ground-state spin density $n_\sigma(y')$, and $n(y')$ is the total ground-state density. Quantities are plotted for reference space-spin positions $x = (0, \uparrow)$, $x = (0.8, \uparrow)$ (left) and $x = (0, \downarrow)$, $x = (0.8, \downarrow)$ (right).	36
3.3 CP densities and potentials in 1D He ($\lambda = 1$): $\tilde{v}_{s,y}^{\text{BEA}}(y')$ is the 1D BEA approximation to the CP-KS potential with corresponding CP density $\tilde{n}_y^{\text{BEA}}(y')$, $\tilde{v}_{s,y}(y')$ is the exact CP-KS potential with corresponding exact CP density $\tilde{n}_y(y')$, $v_s(y')$ the exact KS potential with corresponding exact ground-state density $n(y')$, and $v(y')$ the external potential for our 1D He atom. Quantities are plotted for reference position $y = 0$ (left) and for $y = 0.8$ (right).	48

3.4	CP densities and potentials in 1D Be ($\lambda = 1$): $\tilde{v}_{s,y}^{\text{BEA,LDA}}(y')$ is the 1D BEA approximation to the CP-KS potential using the 1D LDA [29] approximation, see Eq. (3.96). The corresponding CP density is $\tilde{n}_{s,y}^{\text{BEA,LDA}}(y')$	50
3.5	Spin-CP densities and potentials in 1D Li with $N_\uparrow = 2, N_\downarrow = 1$, and $\lambda = 1$: $\tilde{v}_{s,x}^{\text{BEA,LDA}}(y')$ is the 1D BEA approximation to the spin-CP-KS potential using the 1D LDA [29] approximation, see Eq. (3.96), with corresponding spin-CP density $\tilde{n}_{s,x}^{\text{BEA,LDA}}(y')$. The exact CP-KS potential is $\tilde{v}_{s,y}(y')$ with corresponding exact CP density $\tilde{n}_y(y')$, $v_s(y')$ the exact KS potential with corresponding exact ground-state density $n(y')$, and $v(y')$ the external potential for our 1D Li atom.	56
3.6	CP densities and potentials for (3D) He atom ($\lambda = 1$): contour plots for the exact CP density $\tilde{n}_r(r')$ (top left) and corresponding exact CP-KS potential $\tilde{v}_r(r')$ (bottom left) are plotted within the $(x', y', z' = 0)$ plane for reference position $r = (0.563, 0, 0)$ (displayed as red cross). The corresponding LBEA results are plotted on the right.	57
4.1	An unpolarized ground-state gedanken density with 2 electrons whose correlation energy is -21 mH in PBE, but +85 mH in LYP. For reference, the He atom density (divided by 7) is plotted.	62
4.2	The distribution $g(s)$ is plotted for various ground-state densities: the gedanken density in Fig. 4.1, the He and N atoms, and the N ₂ molecule. The absolute difference between the N ₂ molecule and N atom distributions is also plotted. Details of the calculations can be found in Supplemental S2.	65
5.1	One-dimensional H ₂ dissociation curves for several ML models trained from two molecules (red diamonds) with optimal models (highlighted in color) selected by the validation molecule at $R = 3$ (black triangles). The top panel shows energy (with E_{nn} , the nucleus-nucleus repulsion energy) with exact values shown by the black dashed line. The bottom panel shows difference from the exact curves with chemical accuracy in grey shadow. (a) directly predicts E from geometries and clearly fails to capture the physics from very limited data. (b-d) shows our method (KSR) with different inputs to the model to align with the first two rungs of Jacob's ladder [326] (LDA and GGA) and then global (a fully non-local functional). Uniform gas LDA [29] is shown in brown. Grey lines denote 15 sampled functionals during training, with darker lines denoting later samples. Atomic units used throughout.	77
5.2	KS-DFT as a differentiable program. Black arrows are the conventional computation flow. The gradients flow along red dashed arrows to minimize the energy loss L_E and density loss L_n . (a) The high-level KS self-consistent calculations with linear density mixing (purple diamonds). (b) A single KS iteration produces $v_{\text{XC},\theta}[n]$ and $E_{\text{XC},\theta}[n]$ by invoking the XC energy calculation twice, once directly and once calculating a derivative using automatic differentiation. (c) The XC energy calculation using the global XC functional.	84

5.3 (a) t-SNE visualization [264] of density trajectories (grey dots) sampled by KSR during training for $R = 3.84$ from initial guess (cross) to exact density (red diamond). Darker trajectories denote later optimization steps t . Note t-SNE projection does not perfectly preserve the distance between densities. The light red ellipse illustrates the vicinity of the exact density within $\log_{10}(\int dx(n_{\text{KS}} - n_{\text{DMRG}})^2/N_e) \leq -4.25$. Densities from each KS iteration in trajectories are plotted in the corresponding highlighted colors for (b) $t = 0$ untrained, (c) $t = 220$ optimal in Figure 5.1, and (d) $t = 560$ overfitting to training with bad generalization on validation.	85
5.4 Test generalization of models as a function of the total number of training examples N_{train} : full KSR (blue), energy only KSR (pink) and direct ML (orange) on (a) holdout H_2 and H_4 , and unseen types of molecules (b) H_2^+ (c) H_2H_2 . Black dashed lines show chemical accuracy. See the supplemental material [1] for training details.	86
5.5 Density and KS potential of H_4 with $R = 4.32$ from neural XC functionals trained with (a) full KSR (blue) and (b) energy only KSR (pink) on training set of size $N_{\text{train}} = 20$. Exact curves are in grey. v_s are shifted by a constant for better comparison.	86
6.1 (a) sKSR-global, sKSR-LDA and sKSR-GGA architectures to calculate ϵ_{xc} from spin-densities. (b) sKSR – differentiable KS-DFT with spin-polarization. Black arrows refer to the conventional computational flow. The gradients flow along red-dashed arrows to minimize the loss during training.	95
6.2 sKSR-LDA trained on 1D LSDA-calculated Li^{++} and He energies and densities. Here $r_s = 1/2n$ and $\epsilon_{\text{xc}}^{\text{unif}}$ corresponds to the XC energy density of the 1D uniform electron gas [30].	96
6.3 (a) The densities obtained using sKSR-global (orange dashes) and the exact ground-state densities (gray), (b) average XC potentials calculated from sKSR-global approximation (red dashes) to ϵ_{xc} and their exact counterparts calculated with DMRG (light blue) for the test molecules in Table. 6.1 at equilibrium separations. The sKSR potentials are shifted by a constant for a better comparison with the exact XC potentials. sKSR-global was trained on H , He , Li , Be , and Be^{++} and validated on Be^+ . Note that, in general, these 1D densities and XC potentials can differ even qualitatively from their 3D analogs.	99
6.4 The binding energy curve of H_2 molecule calculated based on the total energy prediction for H_2 molecule and the energy of the individual H atoms. sKSR-global was evaluated using restricted KS (blue) and unrestricted KS (red dashes) scheme. The DMRG (black) and KSR-global (green) results are also shown. All the neural approximations, with and without spin, are trained on the dataset given in Table. 6.1.	101

6.5	The complete dissociation energy curve of LiH molecule generated with sKSR-LDA (orange), sKSR-GGA (green) and sKSR-global(red). The DMRG (black dashes) and the uniform gas LSDA (blue dashes) results are also shown. The neural XC functional approximations were trained and validated on atoms and ions given in Table. 6.1.	101
6.6	(a) The total density and (b) the average XC potentials of LiH at a bond-distance of 5.92 Bohr calculated with the three neural XC functionals as well as uniform-gas LSDA. The exact (DMRG) average XC potentials are included for comparison.	102
7.1	TPU v3 wall time for $O(N^3)$ density matrix purification, Eqs. (7.5)-(7.7), as a function of the number N of orbitals, for clusters of water molecules, both in single (squares) and double (triangles) precision. A full TPU v3 pod with 2048 cores and 32 TB of memory is expected to handle $N \sim 500\,000$ orbitals in our current implementation (extrapolated results necessitated by temporary resource unavailability).	106
7.2	The two main steps of our implementation of an $O(N^3)$ DFT computation, the Hamiltonian build and computing the ground state density matrix, which we run on CPUs and TPUs, respectively. The DFT code FHI-aims [48, 49] is used to set up the Hamiltonian and the ELSI library [456, 472] is used to facilitate the integration of the TPU-based solver to FHI-aims.	109
7.3	Wall times for a single DFT iteration on water clusters, using a TPU board composed of a CPU host (48 CPU cores) and 8 TPU cores with 128 GB of HBM. Green and purple curves correspond to using single and double precision in the TPU solver, respectively. The dashed blue curve corresponds to the CPU time spent on FHI-aims (always in double precision), and should be subtracted from the other curves in order to obtain the time spent on the TPU solvers. For reference, in red we also plot the time required for a CPU-only computation using the <i>Eigenvalue solvers for Petaflop Applications</i> (ELPA), a highly parallelized eigensolver library [223, 267], run on 48 CPU cores.	114
7.4	Convergence trajectory of an end-to-end dynamic precision DFT calculation on a $(\text{H}_2\text{O})_{10327}$ cluster. The absolute total energy differences between subsequent DFT iterations, i and $i - 1$, are plotted (top). The corresponding difference in real-space densities within the L^1 norm is plotted (bottom).	115

LIST OF TABLES

	Page
2.1 Results for 2-electron Helium-like ions using HF densities. Virial E_c^{CP} is derived from the virial theorem for atoms.	13
2.2 H_2 energies versus R , where E_c^{Blue} is computed from Eq. 2.18 with the exact density.	14
3.1 Electron-electron repulsion energies, V_{ee} , for 1D systems using various methods. All energies are given in Hartree units. Errors with the exact are given in parenthesis. For $N = 3$ (spin-polarized systems), spin-CP-DFT calculations were performed self-consistently for each spin channel. 1D Hartree-Fock (HF) results are given as reference. Exact results are from 1D DMRG calculations [29]. The local BEA of the main text performs better for 3D Coulomb-interacting systems [270].	52
4.1 For each condition, we assess if the local condition is satisfied (or partially satisfied) for an approximation (with more given in Supplemental S11).	67
6.1 Training, validation and test sets for generalizability experiment. The molecules in the test set refer to the relaxed structures.	97
6.2 Total energy errors (in mH), density losses (in 10^{-4} Bohr $^{-1}$), and errors in ionization potentials for atoms and atomization energies in molecules (in mH) calculated using uniform gas LSDA [30], sKSR-LDA, sKSR-GGA, and sKSR-global respectively, for the training, validation, and test sets in Table 6.1. . .	97
7.1 Tabulated results in Fig. 7.1, including also number of atoms and electrons. Wall times for the matrix purification step are shown both for single (FP32) and double (FP64) precision. Energies are relative to the largest calculation, $E[(H_2O)_{N_{\text{mol}}}] / N_{\text{mol}} - E[(H_2O)_{10327}] / 10327$, where N_{mol} is the total number of water molecules. In this sequence, we used a number of TPU cores that grows roughly as N^2 . As a result, walltimes are seen to roughly scale linearly in N , instead of the expected $O(N^3)$ scaling.	107

ACKNOWLEDGMENTS

Reflecting on the factors that have led me to this point, I deeply appreciate the support and direction I have received from my loved ones, friends, and influential mentors throughout my journey. I feel incredibly fortunate to be blessed with both favorable circumstances and positive influences.

Nearly all of my PhD research has been done in close collaboration with my advisor, Kieron Burke. Interacting with such a dedicated, creative, and responsible scientist has been pivotal in my development. I am also grateful to collaborate with Steven White, whose ability to distill physical insight and improve methodologies in our projects was quite inspiring. It was stimulating to work on new and out-of-the-box approaches (even if the ideas seemed crazy at the time) and already see some of them blossom in the field. Overall, I could not have asked for a more positive PhD experience.

I would also like to thank members of the Burke group family for their constant feedback and reassurance throughout my time. Fellow students and postdocs Bhupalee Kalita, John Kozlowski, Thais Scott, Ryan McCarty, Suhwan Song, Pavel Okun, Steven Crisostomo, Brandon Momanyi, Chris Chen, Ke Xu, and Francisca Sagredo all contributed to creating such a warm, friendly, and productive environment in our group. A special thanks goes out to former Burke group members Li Li and Ryan McCarty who were incredible collaborators and mentors. Li directly connected me to colleagues at Google which led to my internship experience there. Ryan McCarty provided career and interview guidance and other countless valuable pieces of advice.

During my PhD, I was also fortunate to pursue an internship research opportunity at Google X. Here I worked closely with my supervisor Guifre Vidal, a fun and brilliant physicist who worked relentlessly to keep our research effort alive and well-regarded in the field. I extend my appreciation to Shrestha Basu Mallick, Volker Blum, Adam Lewis, Markus Hauru, Jackson Beall, John Kozlowski, and Ruyi Song who were all extremely helpful and knowledgeable colleagues during my internship. I am also especially grateful to my advisor Kieron Burke for accommodating this internship opportunity during my PhD despite it being quite inconvenient.

Lastly, I would like to express my heartfelt gratitude to my friends and family who have always been supportive of my scientific interests. Their unwavering encouragement has meant the world to me. I am principally indebted to my wife, Lauren, who traveled across the country with me to pursue my dreams and has provided unfaltering love and support throughout this adventure.

My work was funded by the Department of Energy Grant No. DE-SC0008696 and the Eddleman Quantum Institute at the University of California, Irvine.

This dissertation is a collection of published journal articles. I acknowledge the following copyright holders. Chapter 1 is reproduced with permission from Ryan J. McCarty, Dennis

Perchak, Ryan Pederson, Robert Evans, Yiheng Qiu, Steven R. White, and Kieron Burke, “Bypassing the energy functional in density functional theory: direct calculation of electronic energies from conditional probability densities.” *Physical Review Letters* 125(26), 266401 (2020), <https://doi.org/10.1103/PhysRevLett.125.266401>, Copyright 2023 by the American Physical Society. Chapter 3 is reproduced with permission from Ryan Pederson, Jielun Chen, Steven R. White, and Kieron Burke, “Conditional probability density functional theory.” *Physical Review B* 105(24), 245138 (2022), <https://doi.org/10.1103/PhysRevB.105.245138>, Copyright 2023 by the American Physical Society. Chapter 4 is reproduced with permission from Ryan Pederson and Kieron Burke, “Reassessing the role of exact conditions in density functional theory.” arXiv preprint 2303.01766 (2023). Chapter 5 is reproduced with permission from Li Li, Stephan Hoyer, Ryan Pederson, Ruoxi Sun, Ekin D. Cubuk, Patrick Riley, and Kieron Burke, “Kohn-Sham Equations as Regularizer: Building Prior Knowledge into Machine-Learned Physics.” *Physical Review Letters* 126(3), 036401 (2021), <https://doi.org/10.1103/PhysRevLett.126.036401>, under the Creative Commons Attribution 4.0 International license. Chapter 6 is reproduced with permission from Bhupalee Kalita, Ryan Pederson, Jielun Chen, Li Li, and Kieron Burke, “How well does Kohn-Sham regularizer work for weakly correlated systems?” *The Journal of Physical Chemistry Letters* 13.11, 2540-2547 (2022), Copyright 2023 American Chemical Society. Chapter 7 is reproduced with permission from Ryan Pederson, John Kozlowski, Ruyi Song, Jackson Beall, Martin Ganahl, Markus Hauru, Adam GM Lewis, Yi Yao, Shrestha Basu Mallick, Volker Blum, and Guifre Vidal, “Large Scale Quantum Chemistry with Tensor Processing Units.” *Journal of Chemical Theory and Computation* 19(1), 25-32 (2023), Copyright 2023 American Chemical Society.

VITA

Ryan Pederson

EDUCATION

Doctor of Philosophy in Physics	2018–2023
University name	<i>Irvine, California</i>
Bachelor of Science in Physics	2013–2017
Virginia Tech	<i>Blacksburg, Virginia</i>

RESEARCH EXPERIENCE

Graduate Research Assistant	2018–2023
University of California, Irvine	<i>Irvine, California</i>

TEACHING EXPERIENCE

Graduate Teaching Assistant	2018–2020
University of California, Irvine	<i>Irvine, California</i>

REFEREED JOURNAL PUBLICATIONS

Lipschitz continuity under toric equivalence for asynchronous Boolean networks	2023
R. Chen, J. McNitt, H. Mortveit, R. Pederson, and C. Reidys <i>Chaos: An Interdisciplinary Journal of Nonlinear Science</i> 33(2), 023118	
Seven Useful Questions in Density Functional Theory	2023
S. Crisostomo, R. Pederson, J. Kozlowski, B. Kalita, A. Cancio, K. Datchev, A. Wasserman, S. Song, K. Burke <i>Letters in Mathematical Physics</i> 113(2), 42	
Large scale quantum chemistry with Tensor Processing Units	2022
R. Pederson, J. Kozlowski, R. Song, J. Beall, M. Ganahl, M. Hauru, A.G. Lewis, S.B. Mallick, V. Blum, Y. Yao, and G. Vidal <i>Journal of Chemical Theory and Computation</i> 19(1), 25-32 [Letter]	
Machine learning and density functional theory	2022
R. Pederson, B. Kalita, and K. Burke <i>Nature Reviews Physics</i> 4.6: 357-358	
Conditional probability density functional theory	2022
R. Pederson, J. Chen, S. White, and K. Burke <i>Physical Review B</i> 105.24: 245138 [Editors' Suggestion]	
How Well Does Kohn–Sham Regularizer Work for Weakly Correlated Systems?	2022
B. Kalita, R. Pederson, J. Chen, L. Li, and K. Burke <i>The Journal of Physical Chemistry Letters</i> 13, 2540-2547	
Kohn-Sham equations as regularizer: building prior knowledge into machine learned physics	2021
L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke <i>Physical review letters</i> 126.3: 036401	
Bypassing the energy functional in density functional theory: Direct calculation of electronic energies from conditional probability densities	2020
R. McCarty, D. Perchak, R. Pederson, R. Evans, Y. Qiu, S. White, and K. Burke <i>Physical Review Letters</i> 125.26: 266401	

**Multireference Ab Initio Studies of Magnetic Properties
of Terbium-Based Single-Molecule Magnets**

2019

R. Pederson, A. Wysocki, N. Mayhall, and K. Park *The Journal of Physical Chemistry A* 123(32), 6996-7006

**Attractor Stability in Finite Asynchronous Biological
System Models**

2019

H. Mortveit and R. Pederson *Bulletin of mathematical biology* 81 (5), 1442-1460

ABSTRACT OF THE DISSERTATION

Developing Density Functional Theory with Physical Prior Knowledge

By

Ryan Pederson

Doctor of Philosophy in Physics

University of California, Irvine, 2023

Professor Kieron Burke, Chair

Density functional theory (DFT) has been used extensively over the past several decades and across many branches of science. The success of DFT lies in its relatively low-cost and usefully high accuracy in many practical systems of interest. However, there are still many instances, such as strongly correlated systems or systems at high temperatures, where conventional DFT approaches are no longer reliable. In addition, reliable DFT approaches are often computationally intractable for large system sizes, limiting their scope of application in realistic system settings. This dissertation is a collection of my contributions to address these fundamental challenges in the field. A common theme across all projects is the use of physical prior knowledge to motivate or (in)directly constrain the methods and techniques developed. In Chapter 1, I provide context for the research presented in the following self-contained chapters. Chapter 2 introduces condition probability DFT (CP-DFT) as a new and alternative density functional approach to obtain conditional probability densities and ground-state energies. Chapter 3 expands upon the previous chapter by establishing CP-DFT as a formally exact theory and derives several key physical properties of CP densities and corresponding potentials used in the theory. Chapter 4 analyzes and discusses the role of exact physical conditions (constraints) in developing conventional Kohn-Sham DFT exchange-correlation (XC) approximations. Chapter 5 introduces the Kohn-Sham regularizer method for training neural network-based XC models for strongly correlated systems. Chapter 6 expands on the

previous chapter by developing a spin-adapted Kohn-Sham regularizer and demonstrating impressive generalizability on weakly correlated systems. Finally, Chapter 7 explores the repurposing of Tensor Processing Units – hardware designed for machine-learning tasks – for large-scale DFT calculations by utilizing algorithms that exploit physical properties of the density matrix.

Chapter 1

Introduction

Computational methods for solving electronic structure problems are invaluable in advancing modern scientific research and technological innovation. The applications of such methods are vast, ranging from catalysis [296] and high-temperature superconductor prediction [104] to accelerating drug discovery [69].

Among the various quantum-based computational methods, density functional theory (DFT) particularly stands out due to its capacity to deliver accurate results for a wide range of systems at a relatively low computational cost. DFT provides entirely quantum solutions at a fraction of the cost of directly solving the Schrödinger equation, by mapping the coupled many-body problem to a single-particle problem. In the standard Kohn-Sham (KS) approach [214], the electronic energy is expressed as a functional of the electron density, and only a small portion of the energy, the exchange-correlation (XC) energy, needs to be approximated. KS-DFT is formally exact, that is, given the exact XC energy functional, we recover the exact ground-state density and energy for *any* electronic system [214]. However, there is no systematic way to approximate the exact XC energy functional and researchers have worked on developing XC energy approximations for decades, culminating in hundreds of

different approximations that are in current use [238]. The sheer number of approximations reflects the immense difficulty of finding generally accurate approximations.

Despite a large research effort, several general failures have been identified in modern DFT approximations. For instance, the self-interaction error [333], which associates errors stemming from an electron interacting with itself nonphysically within approximations. Approximations also violate the flat-plane energy condition [284, 79, 306]. In the context of fractional charges (systems with non-integer total charge) and fractional spins (systems with non-integer spin magnetization), the exact energy is a linear interpolation of the energy of the adjacent integer systems. Approximations notoriously miss this, producing embarrassingly large systematic errors in systems as simple as stretched H_2 . Such grievous failures cause approximations to be especially unreliable for strongly correlated electron systems of interest, thus limiting the applicability of DFT calculations. Given an XC approximation, an additional practical challenge in DFT is to perform the calculation as efficiently as possible, and thus further expand the application space of DFT by enabling calculations on systems of larger scale.

This dissertation is a collection of self-contained chapters based on research with my advisor and collaborators to address these challenges in DFT. It is organized as follows. Chapter 1 introduces *conditional probability* (CP) DFT as an alternative DFT approach that determines the ground-state energy of a system by finding the CP density (the conditional probability density of finding an electron at a position \mathbf{r} given an electron at position \mathbf{r}') from a series of independent KS DFT calculations. By directly calculating CP densities, we bypass the need for an approximate XC energy functional. Chapter 3 establishes CP-DFT as a formally exact theory, derives several key properties, and further explores a suitable CP-DFT approximation that was introduced in Chapter 1. Chapter 4 analyzes and discusses the role of exact conditions (proven mathematical properties of the exact functional) in developing XC approximations. Chapter 5 introduces a machine-learning DFT approach, the KS reg-

ularizer, for training neural network-based XC models. This physically-motivated scheme performs well even with limited data and can be used to train sophisticated XC model approximations that generalize well for strongly correlated systems. Chapter 6 expands on this methodology by developing a spin-adapted KS regularizer to treat spin-polarized systems. We showcase the impressive generalizability of this method on weakly correlated molecular systems. Finally, in Chapter 7, we explore the use of Google’s cloud-based Tensor Processing Units (TPUs) to both accelerate and scale-up conventional DFT calculations by utilizing algorithms that exploit physical properties of the one-body density matrix.

Chapter 2

Bypassing the energy functional in density functional theory: Direct calculation of electronic energies from conditional probability densities

This chapter is a reproduction of Ref. [270], which I co-authored with Ryan J. McCarty, Dennis Perchak, Robert Evans, Yiheng Qiu, Steven R. White, and Kieron Burke.

2.1 Abstract

Density functional calculations can fail for want of an accurate exchange-correlation approximation. The energy can instead be extracted from a sequence of density functional calculations of conditional probabilities (CP-DFT). Simple CP approximations yield usefully accurate results for two-electron ions, the hydrogen dimer, and the uniform gas at all

temperatures. CP-DFT has no self-interaction error for one electron, and correctly dissociates H₂, both major challenges. For warm dense matter, classical CP-DFT calculations can overcome the convergence problems of Kohn-Sham DFT.

2.2 Introduction

Modern electronic structure calculations usually focus on finding accurate ground-state energies, as many predicted properties of a molecule or a material depend on this ability [60]. Wavefunction-based methods, such as coupled-cluster theory [32, 75] or quantum Monte Carlo (QMC) [17, 26], directly yield energies. Kohn-Sham (KS) density functional theory (DFT) [214] incorporates the many-electron problem into the exchange-correlation (XC) energy, which must be approximated as a functional of spin densities. Hundreds of XC functionals with distinct approximations are available in standard codes [238], reflecting the tremendous difficulty in finding general, accurate approximations. Recently, KS-DFT at finite temperatures [276] has been tremendously successful in simulations of warm dense matter [150, 55]. However, it inherits all the limitations of ground-state approximations and becomes impossible to converge for very high temperatures [436].

We propose an alternative to KS-DFT, in which we directly calculate conditional probability densities, from which the energy can be calculated. This bypasses all the difficulties of approximating the XC energy. The electronic pair density can always be written as

$$P(\mathbf{r}, \mathbf{r}') = n(\mathbf{r}) \tilde{n}_{\mathbf{r}}(\mathbf{r}'), \quad (2.1)$$

where $n(\mathbf{r})$ is the single particle density, and $\tilde{n}_{\mathbf{r}}(\mathbf{r}')$ is the conditional probability (CP) density of finding an electron at \mathbf{r}' , given an electron at \mathbf{r} . The standard exact KS potential of DFT, $v_s[n](\mathbf{r})$, is defined to yield $n(\mathbf{r})$ in an effective fermionic non-interacting problem [103].

A conditional probability KS potential (CPKS), $v_s[\tilde{n}_\mathbf{r}](\mathbf{r}')$ yields $\tilde{n}_\mathbf{r}(\mathbf{r}')$ from such a KS calculation with $N - 1$ electrons. Because standard KS-DFT calculations usually yield accurate densities [208], an accurate CPKS potential should yield accurate XC energies. Unlike XC approximations built on theories of the XC hole [313], here we *calculate* that hole.

Just as in traditional DFT, we construct a simple, universal approximation for the CPKS potential from exact conditions and the uniform gas. At large separations or high temperatures, the CP potential reduces to adding $1/|\mathbf{r} - \mathbf{r}'|$ to the external potential, as if the missing electron were classical. We call this a blue electron (i.e. distinguishable from all others), recalling the Percus test particle of classical statistical mechanics [310]. At small separations, the electron-electron cusp condition [62] requires adding only $1/2$ this potential (due to the reduced mass). We locally interpolate between these two universal limits with representative results shown in Fig 2.1.

For the uniform gas at zero temperature, our CP potential interpolation is extremely accurate. We added a strong repulsion for $r_s < 1$, to recover the exchange limit. Panel (b) shows the H_2 binding curve, where the inclusion of the electron-electron cusp is vital. Unlike semi-local DFT, CP-DFT dissociates the molecule correctly, remaining spin-unpolarized throughout. Panel (c) shows CPKS calculations for many temperatures, where the error never exceeds 20%. We show later that orbital-free Thomas-Fermi, and even classical, CP calculations agree reasonably with CPKS, are accurate for all $T > T_F$, and have errors that vanish in the high temperature limit, providing an inexpensive alternative when temperatures are beyond the convergence limit of KS-DFT.

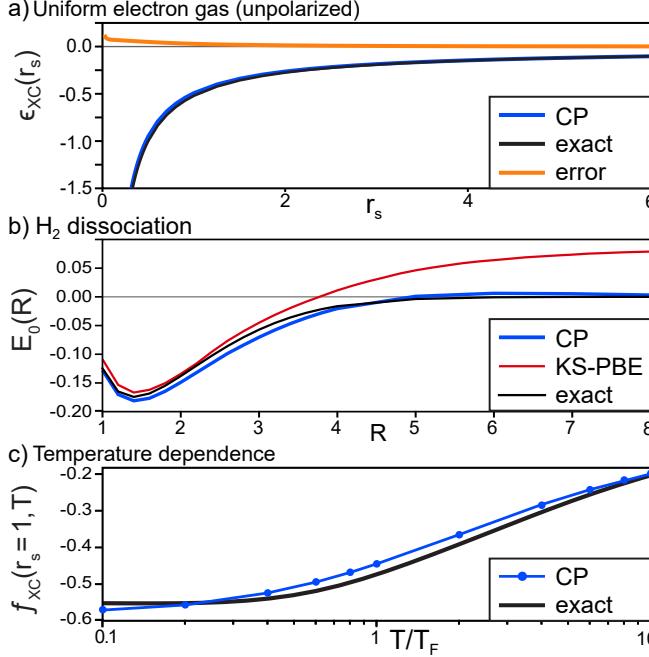


Figure 2.1: CP (blue) and exact (black): (a) XC energy per particle in uniform gas at increasing Wigner–Seitz radii (r_s) and $T = 0$, (b) binding energy curve for H_2 (red is KS-DFT using PBE [313]), and (c) XC free energy per particle at $r_s = 1$ as a function of reduced temperature (T_F is the Fermi temperature). Exact from Ref. [331] in (a), Ref. [155] in (c). Hartree atomic units used throughout.

2.3 Theory

We consider non-relativistic purely electronic problems, and use Hartree atomic units throughout. The pair density of the exact ground-state wavefunction Ψ^λ :

$$P^\lambda(\mathbf{r}_1, \mathbf{r}_2) = N(N-1) \sum_{\sigma_1 \sigma_2} \int d3 \dots dN |\Psi^\lambda(1 \dots N)|^2, \quad (2.2)$$

where N is the number of electrons. Here 1 denotes both \mathbf{r}_1 and σ_1 , the spatial and spin indices. The λ -dependence is the coupling constant in KS DFT, where the repulsion is multiplied by λ but the one-body potential $v^\lambda(\mathbf{r})$ is adjusted to keep the ground-state density $n(\mathbf{r})$ fixed [228]. The XC energy is:

$$E_{XC} = \frac{1}{2} \int_0^1 d\lambda \int d^3r \int d^3r' \frac{n(\mathbf{r}) [\tilde{n}_\mathbf{r}^\lambda(\mathbf{r}') - n(\mathbf{r})]}{|\mathbf{r} - \mathbf{r}'|}, \quad (2.3)$$

with $\tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}') - n(\mathbf{r})$ being the λ -dependent XC hole, defined via the λ -dependent generalization of Eq. 2.1. Setting $\lambda = 1$ in Eq. 2.3 yields U_{xc} , the potential contribution to XC. The integral over λ is called the adiabatic connection.

Denote $v^{\lambda}[n](\mathbf{r})$ as the one body potential that yields the unique ground-state density for electron repulsion $\lambda/|\mathbf{r} - \mathbf{r}'|$. The conditional probability potential is

$$\tilde{v}^{\lambda}(\mathbf{r}'|\mathbf{r}) = v[\tilde{n}_{\mathbf{r}}^{\lambda}](\mathbf{r}') = v^{\lambda}[n](\mathbf{r}') + \Delta \tilde{v}_{\mathbf{r}}^{\lambda}[n](\mathbf{r}'), \quad (2.4)$$

being the unique potential whose ground-state density for Coulomb interacting electrons yields the exact λ -dependent CP density. The CPKS potential is found self-consistently:

$$\tilde{v}_{\text{s}}^{\lambda}(\mathbf{r}'|\mathbf{r}) = v_{\text{s}}[\tilde{n}_{\mathbf{r}}^{\lambda}](\mathbf{r}') = \tilde{v}^{\lambda}(\mathbf{r}'|\mathbf{r}) + v_{\text{HXC}}[\tilde{n}_{\mathbf{r}}^{\lambda}](\mathbf{r}'), \quad (2.5)$$

where v_{HXC} is the Hartree-XC potential [60]. Knowledge of the CP correction potential, $\Delta \tilde{v}_{\mathbf{r}}^{\lambda}[n](\mathbf{r}')$ in Eq. 2.4, allows a self-consistent KS calculation for the exact CP density. Uniqueness of the CP potential is guaranteed by the HK theorem. As $\tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}')$ is non-negative, normalized to $N - 1$, and found from a wavefunction, it is in the standard space of densities, for which we routinely assume KS potentials exist [240, 255].

The above equations are for pure density functionals, and their analogs for spin-density functionals are straightforward (but cumbersome). Decades of research in DFT can be applied to the study of CP densities and potentials, yielding many exact conditions. For example, at $\lambda = 0$ where the exchange hole is never positive,

$$\tilde{n}_{\mathbf{r}}^{\lambda=0}(\mathbf{r}') \leq n(\mathbf{r}'). \quad (2.6)$$

The CP densities satisfy a complementarity principle:

$$\tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}') = \frac{n(\mathbf{r}')}{n(\mathbf{r})} \tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}), \quad (2.7)$$

which is Bayesian, and may be amenable to modern machine-learning methods. The electron coalescence cusp condition requires

$$\frac{\partial \tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}, u)}{\partial u} \Big|_{u=0} = \lambda \tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}), \quad (2.8)$$

where $\mathbf{u} = \mathbf{r}' - \mathbf{r}$ and the left-hand side has been spherically averaged over $\mathbf{r} + \mathbf{u}$ [63]. Using Ref. [245], write

$$\Psi^{\lambda}(1 \dots N) = \sqrt{\frac{n(\mathbf{r}_1)}{N}} \tilde{\Psi}_{\mathbf{r}}^{\lambda}(2 \dots N), \quad (2.9)$$

where $\tilde{\Psi}_{\mathbf{r}}^{\lambda}$ is not antisymmetric under interchange of the electrons, but is uniquely defined by Eq. 2.9, and $\tilde{n}_{\mathbf{r}}^{\lambda}$ is its density. For large r , Ref. [245] shows that $\tilde{\Psi}_{\mathbf{r}}^{\lambda}$ becomes a ground-state of the $N - 1$ particle system and its gradients with respect to \mathbf{r} vanish, yielding

$$\Delta \tilde{v}_{\mathbf{r} \rightarrow \infty}^{\lambda}(\mathbf{r}') \rightarrow \frac{\lambda}{|\mathbf{r} - \mathbf{r}'|}, \quad (2.10)$$

i.e., the blue electron approximation becomes exact in this limit.

For $N = 1$, $\tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}') = 0$, there is no self-interaction error [333]. If $N = 2$, the CP density has just one electron:

$$\tilde{\phi}_{\mathbf{r}}^{\lambda}(\mathbf{r}') = \sqrt{\tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}')} = \sqrt{\frac{2}{n(\mathbf{r})}} \Psi^{\lambda}(\mathbf{r}, \mathbf{r}'), \quad (2.11)$$

yielding

$$\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r}) - \epsilon_r^\lambda = \frac{1}{2} \frac{\nabla'^2 \Psi^\lambda(\mathbf{r}, \mathbf{r}')}{\Psi^\lambda(\mathbf{r}, \mathbf{r}')}, \quad (2.12)$$

where ϵ_r^λ is the eigenvalue of the CPKS potential. Because the wavefunction satisfies the Schrödinger equation, we find

$$\Delta\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r}) + \Delta\tilde{v}_s^\lambda(\mathbf{r}|\mathbf{r}') = \frac{\lambda}{|\mathbf{r} - \mathbf{r}'|} - E^\lambda, \quad (2.13)$$

where $\Delta\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r}) = \tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r}) - v^\lambda[n](\mathbf{r}') - \epsilon_r^\lambda$.

2.4 Approximations

To perform a CP-DFT calculation, we need a general-purpose approximation to the CP potential, $\Delta\tilde{v}_r^\lambda(\mathbf{r}')$. At large separations, the CP potential is simply $\lambda/|\mathbf{r} - \mathbf{r}'|$ for all systems. At small separations, it is $\lambda/(2|\mathbf{r} - \mathbf{r}'|)$, to satisfy the electron-electron cusp condition, for all systems. We interpolate between these two with a simple local density approximation

$$\Delta\tilde{v}_r^\lambda[n](\mathbf{r}') \approx \frac{\lambda}{2|\mathbf{r} - \mathbf{r}'|} \left(1 + \text{Erf}\left(\frac{|\mathbf{r} - \mathbf{r}'|}{r_s(n(\mathbf{r}))}\right)\right), \quad (2.14)$$

where $r_s = (3/(4\pi n))^{1/3}$ is the Wigner-Seitz radius at the reference point. We use this approximation for all ground-state CP calculations in the paper. Fig. 2.1(a) and 2.1(b) use Eq. 2.14 combined with standard DFT approximations for v_{XC} . Fig. 2.1(c) uses simply $\lambda/|\mathbf{r} - \mathbf{r}'|$, as the difference is negligible at significant temperatures.

The N -electron density is trivially a constant, and the one-body potential vanishes. The CP

calculation is for $N - 1$ electrons in a KS potential:

$$v_s(r) = \Delta\tilde{v}(r) + \int d^3r' \frac{\tilde{n}(\mathbf{r}') - n_0}{|\mathbf{r} - \mathbf{r}'|} + v_{\text{XC}}^{\text{LDA}}[\tilde{n}](\mathbf{r}), \quad (2.15)$$

where $n_0 = N/V$ and

$$\Delta\tilde{v}(r) = \Delta\tilde{v}_0^{(\lambda=1)}(n_0, r) + A(r_s)e^{-r^2/2\sigma(r_s)^2}. \quad (2.16)$$

The second term is added to recover the correct high-density limit, i.e., the simple $n^{4/3}$ exchange energy. By calculating many r_s values we can integrate over r_s to perform the adiabatic connection with only $\lambda = 1$. The XC potential is from [426]. The strength and range parameters of the added Gaussian potential are fitted to [426] for $r_s = 0.02$, where exchange dominates. The density is found self-consistently in a sphere using Fermi-surface smearing ($T = 0.05T_F$) and $N = 512$. Imposing zero density flux through the surface of the sphere minimizes boundary effects. Further details will appear in a forthcoming paper.

Fig. 2.2(a) compares the hole density to the parameterization of the uniform gas XC hole [332]. The agreement is very good, with the lowest accuracy from the on-top region, which minimally affects the XC energy.

We applied Eq. 2.14 to highly accurate calculations of 2-electron systems. These calculations were done using a new type of basis function called gausslets [440, 441] which are tailored for density matrix renormalization group calculations [439] and based on wavelets. Gausslets resemble a variable-spaced real-space grid. The two-electron Hamiltonian terms have only two indices, V_{ij} , unlike the four indices needed in a standard basis. The grid-like structure make CP calculations easy to implement. A blue electron sitting at a point in space sits on a gausslet, i , located at its reference, \mathbf{r}_i . The repulsive one-electron potential at i is simply row i of V_{ij} , and integration likewise becomes point-wise sums. Recent innovations add a Gaussian basis to better describe atomic core behavior, further described in a forthcoming

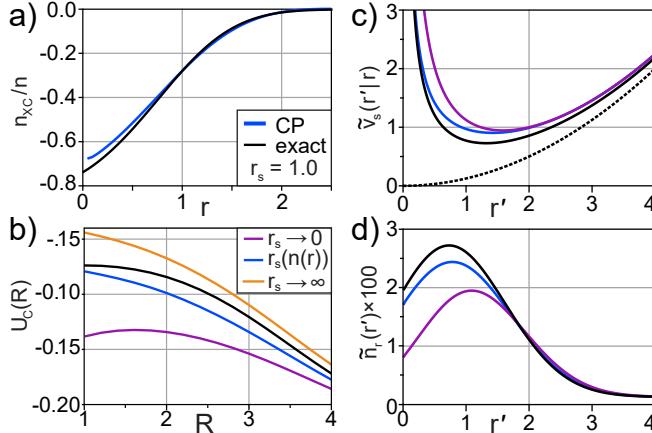


Figure 2.2: Pure blue electron approximation (purple), half pure blue approximation (orange), interpolation (Eq. 2.14, blue), and exact (black): (a) normalized XC hole densities for the uniform gas at $r_s = 1$ with exact parameterization from Ref. [332], (b) $U_c(R)$ from H_2 , using exact $n(r)$ and producing errors below 20%, (c) Hooke's atom $\tilde{v}_s(r'|r)$, with the external potential $r'^2/8$ (black dashed), and (d) Hooke's atom $\tilde{n}_r(r')$.

work. We used 2000 or less gausslets with total energies errors below 0.1 mH for $Z = 1$ and $Z = 2$. To find the conditional probability using Eq. 2.14, we find the ground state of an $N \times N$ matrix with the Lanczos algorithm [227] and repeat N times. Gausslets make an excellent basis for CP calculations, but in any basis, CP calculations are receptive to parallel computing, as each value of \mathbf{r} and λ can be computed independently.

Accurate densities from standard DFT calculations are needed for CP calculations. For 2-electron ions presented in Table 2.1, we choose Hartree-Fock, as it provides a bound density even for H^- [88]. We performed the double integral over \mathbf{r} and \mathbf{r}' to find the potential contribution to correlation, U_c . The virial theorem for atoms (relating the total energy to total kinetic energy, $E = -T$) then allows us to deduce E_c . For He, the ground-state energy error is -6 mH, while that of PBE is +10mH. As $Z \rightarrow \infty$, the CP calculation correctly yields a finite value. At $Z = 1$, the error has increased to 10mH, but H^- is not even bound in a KS-DFT calculation with standard approximations [208].

The virial trick only works for Coulomb-interacting atoms and molecules at equilibrium. Otherwise, we need to perform the adiabatic connection integral. For $N = 2$, we know

Z	E_x^{HF}	$V_{\text{ee}}^{\text{CP}}$	U_c^{CP}	U_c^{Exact}	virial E_c^{CP}	E_c^{Exact}
1.0	-0.3959	0.2918	-0.1041	-0.0698	-0.0523	-0.0420
2.0	-1.0257	0.9301	-0.0956	-0.0786	-0.0479	-0.0421
3.0	-1.6516	1.5521	-0.0995	-0.0832	-0.0504	-0.0435
4.0	-2.2770	2.1750	-0.1020	-0.0857	-0.0525	-0.0443
6.0	-3.5273	3.4226	-0.1047	-0.0881	-0.0563	-0.0452

Table 2.1: Results for 2-electron Helium-like ions using HF densities. Virial E_c^{CP} is derived from the virial theorem for atoms.

the exact result as $\lambda \rightarrow 0$ (exchange limit), where $\tilde{n}_{\mathbf{r}}^{\lambda=0}(\mathbf{r}') = n(\mathbf{r}')/2$. By definition, for 2-electrons we have

$$\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r}) = v_s[n](\mathbf{r}') - \lambda v_{\text{HX}}[n](\mathbf{r}') - v_c^\lambda[n](\mathbf{r}') + \Delta\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r}). \quad (2.17)$$

In practice, obtaining $v_c^\lambda[n](\mathbf{r}')$ is difficult, and we approximate

$$\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r}) \approx \begin{cases} v_s[n](\mathbf{r}'), \lambda = 0 \\ v[n](\mathbf{r}') + (1 - \lambda)v_{\text{HX}}[n](\mathbf{r}') + \Delta\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r}) \end{cases} \quad (2.18)$$

to recover the exchange limit exactly. In the following calculation for H_2 , we utilize the interpolated blue approximation, Eq. 2.14, for $\Delta\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r})$ and the exact density $n(\mathbf{r}')$ throughout. We run for $\lambda \in \{0.0, 0.1, 0.3, 0.5, 0.7, 1.0\}$, and fit to a first-order Padé approximant, which is integrated analytically.

The binding curve for H_2 as a function of bond length is shown in Fig 1(b), with components given in Table 2.2. Fig. 2.2(b), shows $U_c(R)$ for 3 distinct choices of CP potential. As $R \rightarrow \infty$, any version of the blue electron approximation becomes accurate. Consider what

R	E_x	V_{ee}^{Blue}	U_c^{Blue}	U_c^{Exact}	E_c^{Blue}	E_c^{Exact}
1.0	-0.7472	0.6688	-0.0785	-0.0732	-0.0433	-0.0400
2.0	-0.5698	0.4720	-0.0978	-0.0835	-0.0587	-0.0478
4.0	-0.4323	0.2576	-0.1747	-0.1692	-0.1359	-0.1318
8.0	-0.3749	0.1241	-0.2497	-0.2499	-0.2445	-0.2477

Table 2.2: H_2 energies versus R , where E_c^{Blue} is computed from Eq. 2.18 with the exact density.

happens as the bond is stretched. The exact wavefunction has Heitler-London [169] form:

$$\Psi^\lambda(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} (\phi_A(\mathbf{r}_1) \phi_B(\mathbf{r}_2) + \phi_B(\mathbf{r}_1) \phi_A(\mathbf{r}_2)) \quad (2.19)$$

where ϕ_A and ϕ_B are atomic orbitals localized on each of the two protons. This yields a conditional density:

$$n_{\mathbf{r}}^\lambda(\mathbf{r}') = n_B(\mathbf{r}'), \quad \mathbf{r} \text{ near } A \quad (2.20)$$

and vice versa, for all $\lambda \neq 0$. Thus the Coulomb energy of the pair density vanishes due to the lack of overlap, and each atomic region correctly yields a one-electron energy of a separate hydrogen atom. Standard semilocal DFT must choose between retaining the correct spin symmetry, as in the PBE curve of Fig 1(b), or sacrificing accurate spin densities[327]. At the formal level, CP-DFT is an exact theory for bond dissociation, unlike the on-top hole theory of Ref. [327].

Hooke's atom consists of two Coulomb repelling electrons in a harmonic potential of force constant k [193]. At $k = 1/4$, the density is known analytically, and at $r = 0$, the exact $\tilde{v}_s^\lambda(\mathbf{r}'|\mathbf{r})$ is radial. In Fig. 2.2(c) and 2.2(d) we compare the blue electron approximation, our interpolation formula Eq. 2.14, and the exact CP potential and the resulting densities $\tilde{n}_{\mathbf{r}}^\lambda(r')$. Note the accuracy of the blue approximation for large r' , and the cusps as $r' \rightarrow r$ in

the exact and approximate CP densities.

In practical calculations, one does not have access to exact densities, but usually KS-DFT densities from standard approximations are accurate, and in many cases where they are not, Hartree-Fock densities are better[209]. In principle, if neither suffices, densities could be found self-consistently by minimizing the energy from CP calculations with respect to the N -electron density.

2.5 Finite temperatures

Possibly, the most important application of CP-DFT is for thermal equilibrium in warm dense matter [150]. While thermal KS-DFT calculations have been very successful, finding consistent temperature-dependent approximations is more difficult than at zero temperature [102]. Moreover, calculations using KS solvers eventually fail at extremely high temperatures, due to convergence difficulties with orbital sums.

For finite temperatures, Eq 2.3 translates to F_{XC} , the XC contribution to the Helmholtz free energy, which includes entropic contributions [276, 344]. To find accurate CP densities, we solve the KS equations with finite temperature occupations. (Thermal corrections to v_{xc} are argued to have little effect on the orbitals [377]). Fig. 1(c) shows results for the potential XC free energy at $r_s = 1.0$ for a wide range of temperatures. The black curve displays the analytical parameterization (Ref. [155]). The CPKS approximation mildly overestimates f_{xc} for $t = T/T_F$ between about 0.2 and 9. This accuracy has been achieved from our trivial CPKS calculation, without any quantum Monte Carlo or other many-body solver.

But for high temperatures, KS-DFT calculations fail to converge due to the exponential growth in orbitals that contribute, and our calculation is no exception. We therefore performed a much simpler CP calculation using the Thomas-Fermi (TF) approxima-

tion [401, 119], often employed in plasma physics [120, 230], and implementing the simple blue approximation. We first solved the TF equation at $T = 0$ to initiate iterations for a full numerical solution. We make a simple interpolation of Perrot's [335] accurate parameterization of the Helmholtz free energy density $f_0(n)$ of the uniform non-interacting electron gas constructed to yield the correct $T = 0$ and (classical) $T \rightarrow \infty$ limits:

$$f_0(n) = k_B T n \left(\ln(y) - c + a y^{\frac{2}{3}} \right), \quad (2.21)$$

where $y = \pi^2 n / \sqrt{2} (k_B T)^{3/2}$, $c = 1 - \ln(2/\sqrt{\pi})$, and $a = 9(2/3)^{1/3}/10$. The Fermi temperature is given by $k_B T_F = (3\pi^2 n)^{2/3}/2$. As $T \rightarrow 0$, $f_0(n) = 3n k_B T_F / 5$ as required. TF theory corresponds to minimizing the Mermin [276] grand potential functional ignoring XC and making the local density approximation $F[n] = \int d^3 r f_0(n(r))$ for the non-interacting Helmholtz free energy.

Classical connection: In the classical limit TF theory reduces to the Poisson-Boltzmann (PB) theory used to treat electrical double layers and many other properties of electrolyte solutions and ionic liquids [162]. In the high temperature limit we can ignore the third term in Eq 2.21 yielding

$$F[n] = k_B T \int d^3 r n(\mathbf{r}) \left(\ln \left(\frac{n(\mathbf{r}) \lambda^3}{2} \right) - 1 \right), \quad (2.22)$$

where $\lambda = (2\pi/k_B T)^{1/2}$ is the thermal de Broglie wavelength. Eq. 2.22 is identical to the Helmholtz free energy functional of the ideal classical gas, apart from the residual spin degeneracy factor $(2s+1)$. Employing Eq. 2.22 from the outset corresponds to implementing the classical DFT [162, 112] that generates PB theory for the one-component plasma. In the classical limit the TF screening length, λ_{TF} [23], reduces to the Debye length λ_D of the OCP, given by $(\lambda_D)^{-2} = 4\pi e^2 n / k_B T$.

Fig. 2.3 shows relative XC free energy errors as a function of $t = T/T_F$ over a larger temper-

ature range than Fig. 2.1(c). The blue KS approximation (blue curve) performs well across its range. CP-TF (purple) overestimates up to $t \approx 10$; for larger values, all results merge. The classical approximation (green) becomes exact at sufficiently high t .

In the classical limit (Boltzmann statistics) the CP approach is equivalent to the Percus test particle procedure [73, 310]. Fixing a (classical) particle at the origin constitutes an external potential for the others. The resulting one-body density is proportional to the pair correlation function of the liquid [21]. The Percus procedure for quantum systems was pioneered by Chihara [73] and the most successful applications relate to liquid metals and electron-ion correlations [19].

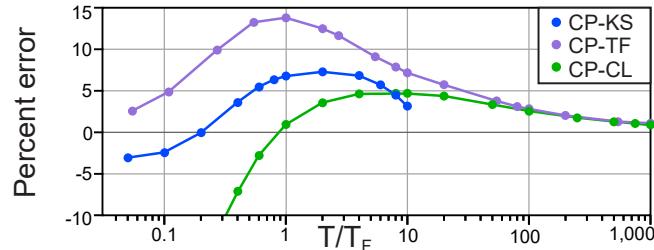


Figure 2.3: Percentage error of uniform gas potential XC free energy per electron for the CP-DFT calculations within KS (blue), TF (purple), and classical (green) approximations relative to the parameterization of Groth et al. [155].

2.6 Conclusion

Lastly, we mention a connection with factorization schemes in the ground state. Eq. 2.9 can be used to find a differential equation for $\tilde{\Psi}_{\mathbf{r}}^{\lambda}(2 \dots N)$. But this is *not* an eigenvalue equation that you solve with given boundary conditions. Such conditional wavefunctions are not always the lowest eigenstate if one treats this as an eigenvalue problem [280]. Moreover, the potential experienced by $\tilde{\Psi}_{\mathbf{r}}^{\lambda}(2 \dots N)$ depends on all $N - 1$ coordinates, so it is not amenable to the standard KS treatment. Thus this seems an unlikely route for deriving other exact properties.

In conclusion, CP-DFT calculations provide a useful alternative to standard KS-DFT. While more expensive, they are highly parallelizable and in important cases, can succeed where KS-DFT often fails. Most importantly, such calculations bypass the need to approximate the XC functional and its potential in difficult cases, such as bond breaking. Our CP potential approximation becomes exact in many limits. It may be exact even for strictly correlated electrons, where

$$\tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}') \rightarrow \sum_{j=1}^{N-1} \delta^{(3)}(\mathbf{r}' - \mathbf{f}_j(\mathbf{r})), \quad (2.23)$$

and $\mathbf{f}_j(\mathbf{r})$ is a co-motion function [149, 136]. Several longer works will follow.

2.7 Acknowledgements

R.J.M. supported by University of California President's Postdoctoral Fellowship, K.B. and D.P. by DOE DE-FG02-08ER46496, R.P. and S.R.W by DOE DE-SC0008696, and R.E. by Leverhulme Trust EM 2020-029/4. K.B. thanks John Perdew for suggesting a variation on this in 1993.

Chapter 3

Conditional probability density functional theory

This chapter is a reproduction of Ref. [305], which I co-authored with Jielun Chen, Steven R. White, and Kieron Burke.

3.1 Abstract

We present conditional probability (CP) density functional theory (DFT) as a formally exact theory. In essence, CP-DFT determines the ground-state energy of a system by finding the CP density from a series of independent Kohn-Sham (KS) DFT calculations. By directly calculating CP densities, we bypass the need for an approximate XC energy functional. In this work we discuss and derive several key properties of the CP density and corresponding CP-KS potential. Illustrative examples are used throughout to help guide the reader through the various concepts and theory presented. We explore a suitable CP-DFT approximation and discuss exact conditions, limitations, and results for selected examples.

3.2 Introduction

Over the past 30 years, density functional theory (DFT) has emerged as a widespread tool in many branches of physics and chemistry [60]. In particular, Kohn-Sham (KS) DFT is an especially popular method to find ground-state energies and electronic densities of molecules and materials. KS-DFT generally scales better with system size than wavefunction-based methods, such as coupled-cluster theory [32, 75] or quantum Monte Carlo (QMC) [26], and is therefore more suitable for modeling realistic systems. However, KS-DFT relies on finding accurate exchange-correlation (XC) energies, which must be approximated as a functional of spin densities. Hundreds of distinct XC approximations currently exist in standard codes [238], reflecting the tremendous difficulty in finding general and accurate approximations. As these standard KS-DFT approximations usually yield accurate self-consistent densities [208], the resulting errors in energies are primarily due to the flaws in the XC energy approximation itself, i.e. the energy error would not change significantly if evaluated on the exact density. Moreover, major deficiencies still remain such as the self-interaction error and the inability to derive accurate energies in strongly-correlated systems [79].

In the exact theory, one can think of XC energies being determined by the pair density, $P^\lambda(\mathbf{r}, \mathbf{r}')$, which is the joint probability density for finding electrons at two points given electron-electron interactions with strength $\lambda/|\mathbf{r}' - \mathbf{r}|$ and fixed ground-state density $n(\mathbf{r})$ [228]. In standard DFT approximations, the on-top pair densities, $P^\lambda(\mathbf{r}, \mathbf{r})$, are quite accurate [319] and several attempts to use pair densities or density matrices in DFT have been made [328, 131, 286]. This pair density can always be written as

$$P^\lambda(\mathbf{r}, \mathbf{r}') = n(\mathbf{r}) \tilde{n}_{\mathbf{r}}^\lambda(\mathbf{r}'), \quad (3.1)$$

where $\tilde{n}_{\mathbf{r}}^\lambda(\mathbf{r}')$ is the conditional probability (CP) density of finding an electron at \mathbf{r}' , given an electron at \mathbf{r} .

In standard KS-DFT, the exact KS potential, $v_s[n](\mathbf{r})$, is defined to yield $n(\mathbf{r})$ in an effective non-interacting electron problem [103]. Analogously, a conditional probability KS potential (CP-KS), $v_s[\tilde{n}_\mathbf{r}^\lambda](\mathbf{r}')$, can be defined at each point \mathbf{r} in the system, so that it yields $\tilde{n}_\mathbf{r}^\lambda(\mathbf{r}')$ from a KS calculation with $N - 1$ electrons. By the Hohenberg-Kohn (HK) theorem [174], the CP-KS potential, if it exists, is unique. The above equations are for pure density functionals, and their analogs for spin-decomposed systems will be presented in the next sections. This approach, recently presented as conditional probability DFT (CP-DFT) in Ref. [270], is formally exact. The exact system density and CP-KS potential yield exact pair densities and hence XC energies. However, just like KS-DFT, CP-DFT is only useful if good approximations can be found for a wide array of problems. A clear advantage of CP-DFT is that it naturally has no self-interaction error for one electron. Since standard KS-DFT calculations usually yield highly accurate total system densities [208], an accurate CP-KS potential approximation should yield highly accurate XC energies.

The CP-DFT approach is analogous to the Overhauser model [298, 96], which was first proposed to approximate the pair correlation function of the uniform electron gas, but has also been extended to non-uniform systems through the *average pair-density functional theory* (APDFT) approach [147, 146]. However, there are key differences: CP-DFT obtains the full pair density whereas APDFT obtains the system- and spherically-averaged pair density, and CP-DFT identifies an exact effective potential within a ground-state KS scheme, the CP-KS potential.

In the context of CP-DFT, authors in Ref. [270] present the *blue electron approximation*. In this rudimentary approximation the CP potential is simply generated by adding $1/|\mathbf{r}' - \mathbf{r}|$ to the external potential. This impurity potential represents the repulsion due to the missing electron as if it were a classical point particle at position \mathbf{r} . That is, an electron that is distinguishable from all others (painted blue). The authors use a variant of this approximation to produce surprisingly accurate results. At zero temperature, the correlation

energy errors for the uniform gas, simple two electron ions, including H^- , and the hydrogen dimer are below 20%, even as the bond is stretched. Achieving a smooth accurate binding energy curve for H_2 while being in a spin singlet still remains a major challenge to accomplish in KS-DFT. Moreover, as the temperature is increased, the blue electron approximation for the uniform gas becomes more accurate, which is intuitively consistent with the classical limit.

In this paper, we consider only zero temperature and extend the work of [270] by generalizing to spin-decomposed systems and presenting a formally exact theory for CP-DFT. We give many exact conditions of the CP density and the corresponding CP-KS potential. We illustrate examples using simple 3D and 1D systems. In the latter, we use 1D exponential interactions to mimic the 3D Coulomb interaction [29]. We explore the *blue electron approximation* presented in [270] and demonstrate that this approximation satisfies several of these key exact conditions. We discuss failures of the blue electron approximation and suggest improvements.

The paper is structured as follows. In Section 3.3 we review DFT and define relevant quantities. In Section 3.4 we present CP-DFT as a formally exact theory. In Section 3.5 we discuss additional relevant properties of the CP density and CP-KS potential. In Section 3.6 we explore approximations in light of exact conditions. Finally, in Section 3.7 we summarize results and discuss future directions of CP-DFT.

3.3 Background and Notation

3.3.1 DFT fundamentals

We consider non-relativistic purely electronic problems, and use Hartree atomic units (a.u.) throughout the paper. Begin with the variational principle for the exact N -electron ground-state energy

$$E = \min_{\Psi} \langle \Psi | \hat{H} | \Psi \rangle , \quad (3.2)$$

where \hat{H} is the N -electron Hamiltonian and the search is over all antisymmetric, normalized many-body wavefunctions Ψ [240]. We consider Hamiltonians of the form

$$\hat{H} = \hat{T} + \hat{V}_{ee} + \hat{V} , \quad (3.3)$$

where \hat{T} is the usual total kinetic energy operator, \hat{V}_{ee} is the two-body electron-electron repulsion operator, and \hat{V} is the one-body, possibly spin-dependent, total external potential for the system.

DFT replaces the central role of the one-body spin-dependent potentials, $v(x)$, with the ground-state spin densities $n(x)$, where

$$n(x) = N \int dx_2 \cdots dx_N |\Psi(x, x_2, \dots, x_N)|^2 . \quad (3.4)$$

Here $x_i = (\mathbf{r}_i, \sigma_i)$ incorporates both spatial and spin coordinates and $\int dx$ is shorthand to denote the integral over all space and sum over both spins and N is the total number of electrons in the system. The total density $n(\mathbf{r})$ is given by a sum over spin densities, $n(\mathbf{r}) = \sum_{\sigma} n(x)$. From the HK theorem generalized to spin DFT [423, 425], there is a one-to-one correspondence between $\{n(x)\}$ and $\{v(x)\}$. From the variational principle, the

ground-state energy of a system of N -electrons and external potentials $v(x)$ is

$$E = \min_n \left(F[n] + \int dx v(x) n(x) \right) \quad (3.5)$$

where $F[n]$ is the universal part of the Hohenberg-Kohn functional, defined as

$$F[n] = \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle, \quad (3.6)$$

where the minimization is over all antisymmetric wavefunctions that yield spin-densities $n(x)$ [240].

In the KS scheme, there exists an effective one-body potential, $v_s(x)$, whose corresponding ground-state spin densities for a system of non-interacting electrons match those of the physical interacting system. The total energy of the real system is given in terms of KS quantities:

$$E = \min_n \left(T_s[n] + \int dx v(x) n(x) + E_H[n] + E_{XC}[n] \right), \quad (3.7)$$

where T_s is the kinetic energy of the KS non-interacting electrons, E_H is the Hartree energy, $n = n_\uparrow + n_\downarrow$ is the total density, and E_{XC} is the XC energy. We denote the KS wavefunction as $\Phi_s[n]$, which we assume here is a single Slater determinant, as is typical. The KS wavefunction minimizes the kinetic energy functional,

$$T_s[n] = \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} | \Psi \rangle = T[\Phi_s[n]], \quad (3.8)$$

where

$$T[\Psi] = -\frac{1}{2} \sum_{i=1}^N \langle \Psi | \nabla_i^2 | \Psi \rangle. \quad (3.9)$$

The Euler-Lagrange equation corresponding to Eq. (3.7) is

$$\frac{\delta T_s}{\delta n(x)} + v_s(x) = 0, \quad (3.10)$$

and all potentials are undetermined up to a constant. The KS spin-dependent potential is

$$v_s(x) = v(x) + \int d^3r \frac{n(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} + v_{xc}[n](x), \quad (3.11)$$

with

$$v_{xc}[n](x) = \frac{\delta E_{xc}[n]}{\delta n(x)}. \quad (3.12)$$

The KS orbitals $\phi_{i\sigma}$ satisfy the KS eigenvalue equation

$$\left[-\frac{1}{2} \nabla^2 + v_s(x) \right] \phi_i(x) = \epsilon_{i\sigma} \phi_i(x). \quad (3.13)$$

3.3.2 Adiabatic connection

In the adiabatic connection approach to KS-DFT [164, 228, 156], we can modulate the interaction strength λ of Coulomb-interacting electrons such that

$$F^\lambda[n] = \min_{\Psi^\lambda \rightarrow n} \langle \Psi^\lambda | \hat{T} + \lambda \hat{V}_{ee} | \Psi^\lambda \rangle. \quad (3.14)$$

For $\lambda = 1$, we have our real, physical system and ground-state density n . For $\lambda = 0$, we have the KS system, because we have turned the electron-electron interaction off, but kept the same ground-state density n . We can also consider the limit as $\lambda \rightarrow \infty$, which is the strictly correlated electron (SCE) limit where the kinetic energy becomes negligible, see section 3.5.4. In all cases, the ground-state density remains fixed to that of the physical system n . This

implies that the external potential $v^\lambda[n]$ is λ -dependent. The potential $v^\lambda[n]$ corresponds to the unique one-body potential for which $n(\mathbf{r})$ is the ground-state density for a system of Coulomb-interacting electrons with interaction strength λ . As convention, if λ is absent from the notation then $\lambda = 1$ is assumed.

3.3.3 Pair densities and XC holes

It is natural to discuss the spin-decomposed pair densities $P^\lambda(x, x')$ of the λ -dependent wavefunctions defined in Eq. (A.31):

$$P^\lambda(x, x') = N(N-1) \int dx_3 \cdots dx_N |\Psi^\lambda(x, x', x_3, \dots, x_N)|^2, \quad (3.15)$$

which is the probability density of finding an electron of spin σ at \mathbf{r} and a second electron of spin σ' at \mathbf{r}' for interaction strength λ . The total pair density is the spin-summed quantity:

$$P^\lambda(\mathbf{r}, \mathbf{r}') = \sum_{\sigma\sigma'} P^\lambda(x, x'), \quad (3.16)$$

which is the joint probability of finding an electron at \mathbf{r} and a second electron at \mathbf{r}' for interaction strength λ . For $\lambda = 0$, $\Psi^{\lambda=0} = \Phi_s[n]$ and the pair density is

$$P^{\lambda=0}(x, x') = n(x) n(x') - |\gamma_s(x, x')|^2, \quad (3.17)$$

where γ_s is the KS (first-order) density matrix:

$$\gamma_s(x, x') = \delta_{\sigma\sigma'} \sum_{i=1}^{N_\sigma} \phi_i^*(x) \phi_i(x'), \quad (3.18)$$

The exchange hole is defined as

$$n_x(x, x') = -\frac{|\gamma_s(x, x')|^2}{n(x)}, \quad (3.19)$$

such that

$$P^{\lambda=0}(x, x') = n(x)(n(x') + n_x(x, x')). \quad (3.20)$$

Checking the normalization allows us to deduce

$$\int dx' n_x(x, x') = -1. \quad (3.21)$$

For $\lambda > 0$, we can generalize Eq. (3.20) by introducing the λ -dependent correlation hole $n_c^\lambda(x, x')$:

$$P^\lambda(x, x') = n(x)(n(x') + n_x(x, x') + n_c^\lambda(x, x')), \quad (3.22)$$

where $n_c^{\lambda=0}(x, x') = 0$ and normalization dictates

$$\int dx' n_c^\lambda(x, x') = 0. \quad (3.23)$$

The XC hole, $n_{xc}^\lambda(x, x')$, is simply the sum of the exchange and correlation holes, $n_{xc}^\lambda(x, x') = n_x(x, x') + n_c^\lambda(x, x')$. Analogously, we can express the spin-summed pair density as

$$P^\lambda(\mathbf{r}, \mathbf{r}') = n(\mathbf{r})(n(\mathbf{r}') + n_{xc}^\lambda(\mathbf{r}, \mathbf{r}')), \quad (3.24)$$

where

$$n_{xc}^\lambda(\mathbf{r}, \mathbf{r}') = \left(\sum_{\sigma, \sigma'} n(x) n_{xc}^\lambda(x, x') \right) / n(\mathbf{r}). \quad (3.25)$$

This quantity is especially of interest in DFT since it determines the XC energy in the adiabatic connection integral [63]:

$$E_{\text{xc}} = \frac{1}{2} \int_0^1 d\lambda \int d^3r \int d^3r' \frac{n(\mathbf{r}) n_{\text{xc}}^\lambda(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (3.26)$$

3.4 Theory

3.4.1 CP-DFT

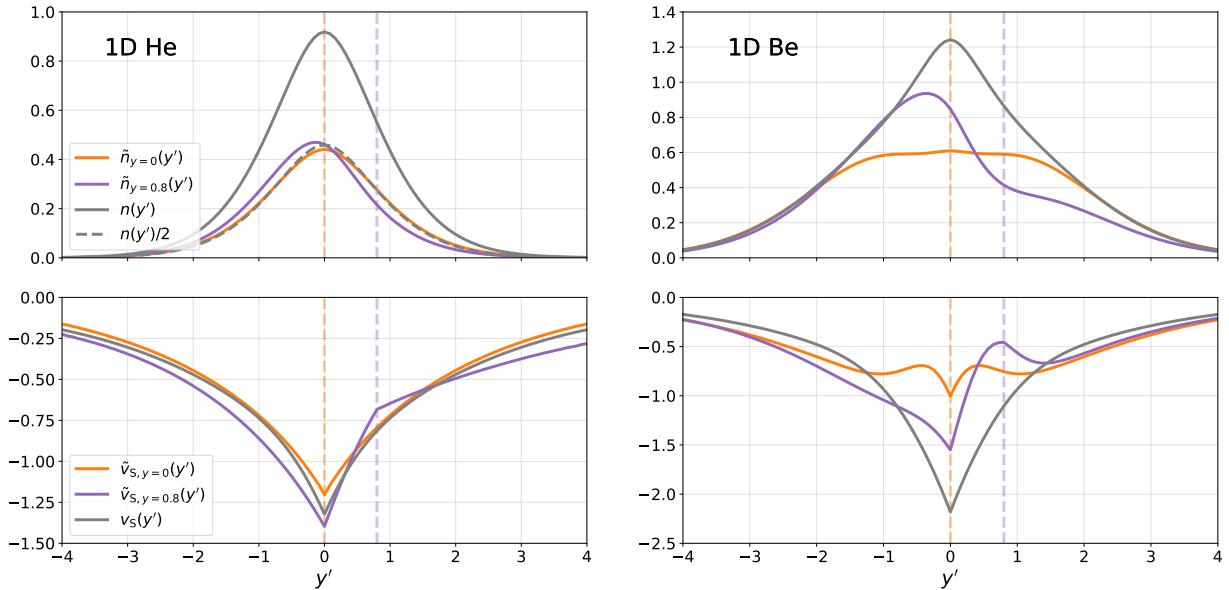


Figure 3.1: Exact CP densities and potentials in 1D He (left) and 1D Be (right): $\tilde{v}_{s,y}(y')$ is the CP-KS potential with corresponding CP density $\tilde{n}_y(y')$, $v_s(y')$ the KS potential with corresponding ground-state density $n(y')$. Quantities are plotted for reference positions $y = 0$ and $y = 0.8$.

To simplify the discussion, we will first consider the case of $\lambda = 1$ and pure DFT, i.e. not spin decomposed. The subsequent sections will introduce the adaption for different interaction strengths λ and spins σ .

We define the total conditional probability (CP) density density $\tilde{n}_{\mathbf{r}}(\mathbf{r}')$ as

$$\tilde{n}_{\mathbf{r}}(\mathbf{r}') \equiv \frac{P(\mathbf{r}, \mathbf{r}')}{n(\mathbf{r})} = n(\mathbf{r}') + n_{\text{xc}}(\mathbf{r}, \mathbf{r}'), \quad (3.27)$$

which corresponds to the conditional probability density of finding an electron at \mathbf{r}' , given an electron at \mathbf{r} . The subscript in $\tilde{n}_{\mathbf{r}}(\mathbf{r}')$ emphasizes the parametric dependence on the reference position \mathbf{r} . In Figure 3.1 we provide example exact densities for a 1-dimensional (1D) model of the He and Be atoms. The 1D model we use throughout mimics reality by replacing 3D Coulomb interactions, $1/|\mathbf{r}' - \mathbf{r}|$, with 1D exponentials of the form $A \exp(-\kappa|y - y'|)$, where $y, y' \in \mathbb{R}$ and $A, \kappa > 0$ are parameters fitted to a soft Coulomb potential, see [29, 251, 28, 249] for additional information. The purpose of using 1D models here and throughout the paper is to provide simple and illustrative examples which are qualitatively similar to 3D reality. The 1D “exact” density results are obtained from high accuracy density matrix renormalization group (DMRG) calculations using the ITensor library [124] with an energy convergence threshold of 10^{-7} Hartree. Associated single-particle potentials were obtained using a modified version of the KS-inversion algorithm outlined in [123]. The code used to perform KS-inversion is publicly available at [304].

In Figure 3.1 we plot the total ground-state density $n(y')$ for 1D He, which is spatially symmetric and analogous to a doubly occupied 1s orbital. The exact KS potential $v_s(y')$ corresponding to this density is also plotted. We show the CP density $\tilde{n}_y(y')$ for two different reference points, $y = 0$ and $y = 0.8$. Subtracting the ground-state density $n(y')$ from the CP density yields the ($\lambda = 1$) XC hole for a given reference position. We also show the exchange CP density which, for 2 electrons in a spin singlet, is simply half the total density. Thus differences of CP densities from this density are the correlation contributions which are very tiny when $y = 0$. For $y = 0.8$, correlation causes the CP density to be lower in the region of the electron and greater on the opposite side of the nucleus.

The upper right panel of the figure is the same as the left, but for the Be atom. Here the 1D Be ground-state density is analogous to a $1s^22s^2$ configuration. Now the CP density integrates to 3 electrons and is not closely approximated by half the density. However the same effect occurs when the reference electron is off-center and charge is pushed to the other side of the nucleus.

The CP density follows several properties. For all reference points \mathbf{r} , it must normalize to $N - 1$ electrons,

$$\int d^3r' \tilde{n}_{\mathbf{r}}(\mathbf{r}') = N - 1. \quad (3.28)$$

In Figure 3.1, taking the area under the curve of any 1D He CP density yields 1 electron. Similarly, for any 1D Be CP density the integrated area yields 3 electrons. Moreover, since the pair density is a symmetrical function of \mathbf{r} and \mathbf{r}' , the CP densities satisfy a complementary principle:

$$\tilde{n}_{\mathbf{r}}(\mathbf{r}') = \frac{n(\mathbf{r}')}{n(\mathbf{r})} \tilde{n}_{\mathbf{r}'}(\mathbf{r}), \quad (3.29)$$

which is simply Bayes' theorem, and may be amenable to modern machine-learning methods [400]. In Fig. 1, for Be, this means the ratio of where the purple curve intersects the orange vertical line to where the orange curve hits the purple vertical line equals the ratio of densities (gray curve) at the two points. For He, since the orange curve is almost identical to the grey dashed line, any CP density (purple curve) passes through (almost) the same value at the origin, namely $n(0)/2$.

Next we turn to the CP potentials that generate CP densities. For a given reference position \mathbf{r} and CP density $\tilde{n}_{\mathbf{r}}(\mathbf{r}')$, we define the CP potential, $\tilde{v}_{\mathbf{r}}(\mathbf{r}')$, as the unique one-body potential whose total ground-state density for $(N - 1)$ Coulomb interacting electrons yields the total CP density $\tilde{n}_{\mathbf{r}}(\mathbf{r}')$. Uniqueness of the CP potential (up to a constant shift) is guaranteed by

the HK theorem. If it is known, the CP potential for this $(N - 1)$ electron auxiliary system is simply

$$\tilde{v}_{\mathbf{r}}(\mathbf{r}') = v[\tilde{n}_{\mathbf{r}}](\mathbf{r}') \quad (3.30)$$

where $v[n](\mathbf{r})$ is the unique one-body potential for which $n(\mathbf{r})$ is the Coulomb-interacting ground-state density. In Figure 3.1 the exact CP potentials (and corresponding CP densities) are plotted for He atom at reference points $y = 0.0$ and $y = 0.8$. Notice that in the $y = 0.8$ case, the potential is asymmetric and contains a kink at the reference position. The kink in the CP potential exists due to the 1D exponential interaction used. However, for standard 3D Coulomb interactions, we instead obtain a kink (or cusp) in the CP density at the reference position due to the electron-electron cusp condition, see section 3.5.2. A general feature in any dimension is that the CP potential is less negative in the region of the reference point, as it must push electrons away.

We can generalize this HK map to a system of $(N - 1)$ electrons with interaction strength $\alpha \geq 0$ and ground-state density $\tilde{n}_{\mathbf{r}}$. In analogy, we follow Section 3.3.2, but throughout we are careful to use α to denote the interaction strength of the associated $(N - 1)$ electron auxiliary system, while λ will be used exclusively to denote the interaction strength of the total N electron system with fixed ground-state density n . That is, α and λ are independent variables in general. For a given α , we have

$$\tilde{v}_{\mathbf{r}}^{\alpha}(\mathbf{r}') = v^{\alpha}[\tilde{n}_{\mathbf{r}}](\mathbf{r}') , \quad (3.31)$$

so that $\tilde{v}_{s,\mathbf{r}}(\mathbf{r}') = \tilde{v}_{\mathbf{r}}^{\alpha=0}(\mathbf{r}')$. In Figure 3.1 exact CP-KS potentials (and corresponding CP densities) are plotted for 1D He and Be atoms. In the case of the 1D He atom, clearly the CP potential is the CP-KS potential. In the Be atom $y = 0.8$ case, we see a bump and kink in the CP-KS potential at the reference position, resulting in a small dip in the CP density

in the region near the reference position.

The CP-KS potential can be expressed as

$$\tilde{v}_{s,\mathbf{r}}(\mathbf{r}') = \tilde{v}_{\mathbf{r}}^{\alpha}(\mathbf{r}') + \alpha v_{\text{HX}}[\tilde{n}_{\mathbf{r}}](\mathbf{r}') + v_{\text{C}}^{\alpha}[\tilde{n}_{\mathbf{r}}](\mathbf{r}'), \quad (3.32)$$

where v_{HX} and v_{C}^{α} are the usual KS-DFT Hartree-exchange and correlation potentials, respectively [60]. As in standard KS-DFT, the CP-KS potential can be found self-consistently. Following Eq. (3.13), for each reference point \mathbf{r} the corresponding CP-KS orbitals $\tilde{\phi}_{\mathbf{r},i}(\mathbf{r}')$ satisfy the following CP-KS eigenvalue equation

$$\left[-\frac{1}{2}\nabla'^2 + \tilde{v}_{s,\mathbf{r}}(\mathbf{r}') \right] \tilde{\phi}_{\mathbf{r},i}(\mathbf{r}') = \epsilon_{\mathbf{r},i} \tilde{\phi}_{\mathbf{r},i}(\mathbf{r}'), \quad (3.33)$$

where $\sum_{i=1}^{N-1} |\tilde{\phi}_{\mathbf{r},i}(\mathbf{r}')|^2 = \tilde{n}_{\mathbf{r}}(\mathbf{r}')$.

Notice that by construction, the CP-DFT approach is formally exact: knowledge of an exact CP potential, $\tilde{v}_{\mathbf{r}}^{\alpha}(\mathbf{r}')$, and a corresponding exact XC potential, $\tilde{v}_{\text{XC}}^{\alpha}(\mathbf{r}')$, allows a self-consistent KS calculation for an exact CP density $\tilde{n}_{\mathbf{r}}(\mathbf{r}')$.

By construction, CP densities are nowhere negative and normalized to $N - 1$. Thus if their kinetic energy is finite (and we know of no counterexample), they are members of the usual set of well-behaved densities, \mathcal{I}_N of Ref. [255]. However, this is insufficient to guarantee non-interacting v -representability, just as in standard DFT [255, 109]. In practice, CP-DFT calculations use explicit approximations for both v_{XC} and $\tilde{v}_{\mathbf{r}}$, guaranteeing that all CP densities explored are v -representable by construction.

One may question whether a CP density $\tilde{n}_{\mathbf{r}}(\mathbf{r}')$ contains nodes whenever $\mathbf{r} = \mathbf{r}'$ due the exclusion principle, which might make the kinetic energy diverge. However, CP densities are derived from the pair density which is a spin-summed quantity, Eq. (3.16), so we avoid such fermionic nodes. In the spin adaptation of CP-DFT, we must be more careful, see

Section 3.4.2.

To extract the XC energy from Eq. (3.26) we need the λ -dependent XC holes, $n_{\text{xc}}^\lambda(\mathbf{r}, \mathbf{r}')$, defined in Eq. (3.24), so we write

$$\tilde{n}_{\mathbf{r}}^\lambda(\mathbf{r}') \equiv \frac{P^\lambda(\mathbf{r}, \mathbf{r}')}{n(\mathbf{r})} = n(\mathbf{r}') + n_{\text{xc}}^\lambda(\mathbf{r}, \mathbf{r}'), \quad (3.34)$$

for the conditional probability density of finding an electron at \mathbf{r}' , given an electron at \mathbf{r} and for interaction strength λ . The λ -dependent CP density also normalizes to $N - 1$ electrons and satisfies the complementary principle in Eq. (3.29).

In analogy with previous formalism, for a given $\mathbf{r}, \lambda, \alpha$ and CP density, $\tilde{n}_{\mathbf{r}}^\lambda(\mathbf{r}')$, we define the CP potential, $\tilde{v}_{\mathbf{r}}^{\alpha, \lambda}(\mathbf{r}')$, as the unique one-body potential whose ground-state density for α -strength interacting $(N - 1)$ electrons yields the CP density, $\tilde{n}_{\mathbf{r}}^\lambda(\mathbf{r}')$. If it exists, the CP potential is

$$\tilde{v}_{\mathbf{r}}^{\alpha, \lambda}(\mathbf{r}') = v^\alpha[\tilde{n}_{\mathbf{r}}^\lambda](\mathbf{r}') \quad (3.35)$$

where

$$v^\alpha[\tilde{n}_{\mathbf{r}}^\lambda](\mathbf{r}') = -\frac{\delta F^\alpha[\tilde{n}_{\mathbf{r}}^\lambda]}{\delta \tilde{n}_{\mathbf{r}}^\lambda}. \quad (3.36)$$

The CP-KS potential is $\tilde{v}_{\text{s}, \mathbf{r}}^\lambda(\mathbf{r}')$, where $\tilde{v}_{\text{s}, \mathbf{r}}^\lambda(\mathbf{r}') = \tilde{v}_{\mathbf{r}}^{\alpha=0, \lambda}(\mathbf{r}')$ and

$$\tilde{v}_{\text{s}, \mathbf{r}}^\lambda(\mathbf{r}') = \tilde{v}_{\mathbf{r}}^{\alpha, \lambda}(\mathbf{r}') + \alpha v_{\text{HX}}[\tilde{n}_{\mathbf{r}}^\lambda](\mathbf{r}') + v_{\text{C}}^\alpha[\tilde{n}_{\mathbf{r}}^\lambda](\mathbf{r}'). \quad (3.37)$$

Again, we emphasize that, by construction, CP-DFT is formally exact: knowledge of an exact CP potential, if it exists, $\tilde{v}_{\mathbf{r}}^{\alpha, \lambda}$, and an exact XC potential, v_{xc}^α , allows a self-consistent CP-KS calculation for an exact CP density $\tilde{n}_{\mathbf{r}}^\lambda$. If such exact CP-KS calculations are carried out for

all $\lambda \in [0, 1]$ and all reference positions, \mathbf{r} , in space, and the exact total ground-state density n is known, then the exact XC energy can be obtained from Eq. (3.26).

3.4.2 Spin-adapted CP-DFT

While Section 3.4.1 presents a formally exact construction for *any* interacting electron system, it is useful in practice to have an analogous spin-decomposed formally exact framework. Below we omit α and λ dependencies for clarity, but the dependence should follow straightforwardly from the Section 3.4.1.

We define the spin-CP density $\tilde{n}_x(\mathbf{r}')$ as

$$\tilde{n}_x(\mathbf{r}') \equiv \sum_{\sigma'} \tilde{n}_x(x') , \quad (3.38)$$

where

$$\tilde{n}_x(x') = \frac{P(x, x')}{n(x)} = n(x') + n_{\text{xc}}(x, x') . \quad (3.39)$$

The spin-CP density $\tilde{n}_x(\mathbf{r}')$ has a natural interpretation: given an electron of spin σ at reference position \mathbf{r} , it is the probability density of finding *any* electron at position \mathbf{r}' . The spin-CP density normalizes to $N - 1$ electrons,

$$\int d^3r' \tilde{n}_x(\mathbf{r}') = N - 1 , \quad (3.40)$$

making it a natural object to use as a starting point for spin-adapted CP-DFT. A weighted

sum over spin-CP densities can be used to determine the total CP density:

$$\tilde{n}_{\mathbf{r}}(\mathbf{r}') = \sum_{\sigma} \frac{n_{\sigma}(\mathbf{r})}{n(\mathbf{r})} \tilde{n}_{\mathbf{r}\sigma}(\mathbf{r}'). \quad (3.41)$$

In the case of a spin-unpolarized system, we recover $\tilde{n}_{\uparrow\mathbf{r}}(\mathbf{r}') = \tilde{n}_{\downarrow\mathbf{r}}(\mathbf{r}') = \tilde{n}_{\mathbf{r}}(\mathbf{r}')$ as expected.

In Figure 3.2 we provide exact densities from a 1D model of the spin-polarized Li atom. For each plotted spin-CP and total CP density, taking the area under the curve yields 2 electrons. The total CP densities are plotted in Figure 3.2 using dashed curves. From Eq. (3.41), the total CP density can be obtained from a sum of the spin-CP densities, with each weighted by a fraction of density in that spin channel. In 1D Li, for spin-restricted KS orbitals we have the exchange-limit ($\lambda = 0$) relation $n_{\uparrow}(y') = \tilde{n}_{(0,\uparrow)}^{\lambda=0}(y') = \tilde{n}_{(y,\downarrow)}^{\lambda=0}(y')$. In the upper left panel of Figure 3.2 we see that indeed $n_{\uparrow}(y')$ closely approximates the spin-CP density, $\tilde{n}_{(0,\uparrow)}(y')$, with reference point at the origin, which is a high density region dominated by exchange. In the upper right panel, for the same reference position but opposite spin, we obtain a very similar spin-CP density, $\tilde{n}_{(0,\uparrow)}(y') \approx \tilde{n}_{(0,\downarrow)}(y')$, because, for this reference point, the 1s orbital, which is doubly occupied, is dominant. However, for a spin-down reference point the exchange limit spin-CP density is independent of the reference position, $\tilde{n}_{(y,\downarrow)}^{\lambda=0}(y') = n_{\uparrow}(y')$, which is similar to the case in 1D He (Figure 3.1) and correspondingly we see little change in the down spin-CP density, $\tilde{n}_{(y,\downarrow)}(y')$, as the reference position is changed from 0 to 0.8. This is not the case for $\sigma = \uparrow$, so we see large differences between $\tilde{n}_{(0.8,\uparrow)}(y')$ and $\tilde{n}_{(0.8,\downarrow)}(y')$.

The analog definition for the spin-CP potential, $\tilde{v}_x(\mathbf{r}')$, is straightforward: it is the unique one-body potential whose “accessible” ground-state density from ($N-1$) Coulomb-interacting electrons yields the spin-CP density, $\tilde{n}_x(\mathbf{r}')$, that is, $\tilde{v}_x(\mathbf{r}') = v[\tilde{n}_x](\mathbf{r}')$. Note that this is within DFT, not spin DFT, and an “accessible” ground-state is one that conserves total spin, e.g., the total spins of the $N-$ and ($N-1$)-electron systems cannot differ by more than 1/2 unit of angular momentum [110, 202].

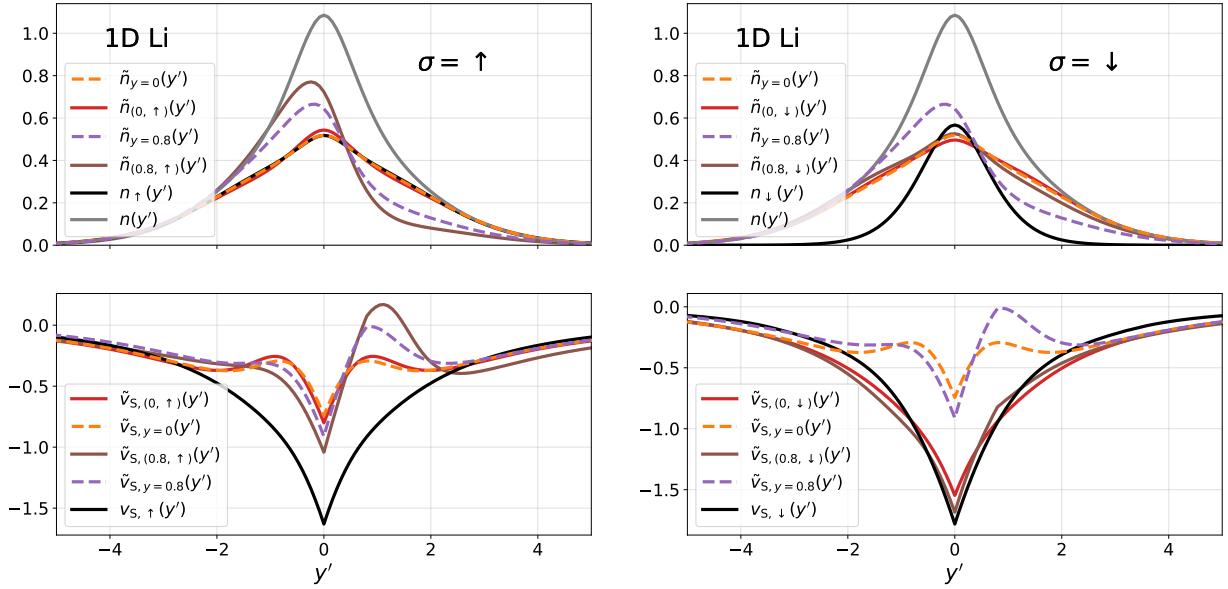


Figure 3.2: Exact CP densities and potentials in 1D Li with $N_\uparrow = 2, N_\downarrow = 1$. The CP-KS potential is $\tilde{v}_{s,y}(y')$ with corresponding total CP density $\tilde{n}_y(y')$ (dashed), the spin-CP-KS potential is $\tilde{v}_{s,x}(y')$ with corresponding spin-CP density $\tilde{n}_x(y')$, $v_{s,\sigma}(y')$ is the KS spin potential with corresponding ground-state spin density $n_\sigma(y')$, and $n(y')$ is the total ground-state density. Quantities are plotted for reference space-spin positions $x = (0, \uparrow)$, $x = (0.8, \uparrow)$ (left) and $x = (0, \downarrow)$, $x = (0.8, \downarrow)$ (right).

Practically, the main difference from Section 3.4.1 is that, for each reference point \mathbf{r} we solve two independent self-consistent CP-KS equations (one for each spin σ) to obtain corresponding spin-CP densities $\tilde{n}_x(\mathbf{r}') = \sum_{i=1}^{N-1} |\tilde{\phi}_{i,x}(\mathbf{r}')|^2$. Again, the ground-states should be accessible from a given spin configuration. For example, in 1D Li with $N_\uparrow = 2, N_\downarrow = 1$, if the reference electron is spin-down we must singly-occupy the lowest two CP-KS orbitals (with spin-up electrons) in the $(N-1)$ system to conserve spin. If the reference electron is spin-up, we doubly-occupy the lowest CP-KS orbital. This means that similar looking up and down spin-CP densities can have corresponding spin-CP-KS potentials that look quite different. This is why $\tilde{v}_{s,(0,\uparrow)}(y')$ looks so different from $\tilde{v}_{s,(0,\downarrow)}(y')$ in Figure 3.2.

Again, we emphasize that spin-CP densities should be a result from pure DFT with spin-restriction, not spin DFT. One may ask whether a spin-unrestricted CP-KS scheme could be used to obtain the fully spin-decomposed CP spin densities, $\tilde{n}_x(x')$, which correspond to the

conditional probability density of finding an electron at \mathbf{r}' of spin σ' , given an electron at \mathbf{r} of spin σ . However, these do not spatially integrate (normalize) to a predetermined integer in general,

$$\int d^3r' \tilde{n}_x(x') = N_{\sigma'} - \delta_{\sigma\sigma'} + \int d^3r' n_{\text{C}}(x, x') , \quad (3.42)$$

and so would require an ensemble of fractional particle numbers [103], presumably making approximations more complicated. Furthermore, these spin densities will contain nodes whenever $x' = x$ due to the exclusion principle. For instance, in the Li atom (with $N_{\uparrow} = 2$ and $N_{\downarrow} = 1$), the spin density $\tilde{n}_{\mathbf{r}\uparrow}^{\lambda=0}(\mathbf{r}'\uparrow)$ is non- v -representable, as it is a single-particle density which is not strictly positive [109]. Therefore, in general we do not recommend fully spin-decomposed CP-DFT.

3.4.3 CP-DFT with averaged quantities

The spin-CP densities $\tilde{n}_x^{\lambda}(\mathbf{r}')$ are clearly high-dimensional quantities: in general there is a λ and x dependence. In the following we define exact lower-dimensional quantities which can equivalently be used to extract relevant exact energies, such as the electron-electron repulsion potential energy $U_{\text{ee}} = E_{\text{H}} + E_{\text{xc}}$:

$$U_{\text{ee}} = \frac{1}{2} \int_0^1 d\lambda \int dx n(\mathbf{r}) \int d^3r' \frac{\tilde{n}_x^{\lambda}(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} . \quad (3.43)$$

This may be rewritten in terms of purely radial quantities:

$$U_{\text{ee}} = \frac{N}{2} \sum_{\sigma} \int du u \langle \tilde{n}_{\sigma}(u) \rangle \quad (3.44)$$

where $\mathbf{u} = \mathbf{r}' - \mathbf{r}$, $u = |\mathbf{r}' - \mathbf{r}|$, and the system and spherically-averaged CP density is

$$\langle \tilde{n}_\sigma(u) \rangle = \frac{1}{N} \int_0^1 d\lambda \int d^3r n(\mathbf{r}) \int d\Omega_{\mathbf{u}} \tilde{n}_{\sigma\mathbf{r}}^\lambda(\mathbf{r} + \mathbf{u}). \quad (3.45)$$

Note that this radial quantity integrates to $N-1$. The corresponding radial CP-KS potential is defined as $\tilde{v}_s[\langle \tilde{n}_\sigma \rangle](u)$. Knowledge of this much simpler functional also yields the XC energy, but now only a single 1D integral is required, instead of a 3D integral over \mathbf{r} . If useful approximations can be found directly for $\tilde{v}_s[\langle \tilde{n}_\sigma \rangle](u)$, CP-DFT calculations would be much more efficient. This approach is analogous to *average pair-density functional theory* [146, 147] which utilizes an effective radial potential, such as the Overhauser model [298, 96], to determine spherically- and system-averaged pair densities.

3.4.4 connection to exact factorization?

In this section we present an explicit differential equation for $\tilde{n}_x^\lambda(\mathbf{r}')$. We begin by following Ref. [136] and partition the λ -dependent N -electron Hamiltonian as

$$H^\lambda = H_1^\lambda + H_{N-1}^\lambda + \sum_{j=2}^N \frac{\lambda}{|\mathbf{r} - \mathbf{r}_j|}, \quad (3.46)$$

where H_1^λ is the Hamiltonian for a single electron

$$H_1^\lambda = -\frac{1}{2} \nabla_{\mathbf{r}}^2 + v^\lambda[n](\mathbf{r}) \quad (3.47)$$

and H_{N-1}^λ is the Hamiltonian for $(N-1)$ λ -strength interacting electrons

$$H_{N-1}^\lambda = \sum_{i=2}^N \left[-\frac{1}{2} \nabla_i^2 + v^\lambda[n](\mathbf{r}_i) \right] + \frac{1}{2} \sum_{i \neq j=2}^N \frac{\lambda}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (3.48)$$

We factorize the normalized ground-state wavefunction Ψ^λ of H^λ as

$$\Psi^\lambda(x, x_2, \dots, x_N) = \sqrt{\frac{n(x)}{N}} \tilde{\Psi}_x^\lambda(x_2, \dots, x_N). \quad (3.49)$$

Here, $\tilde{\Psi}_x^\lambda$ depends parametrically on x and is antisymmetric under interchange of the $N - 1$ electrons in its argument, but not under interchange of x with any x_i , $i \geq 2$, in general. Note that $\tilde{\Psi}_x^\lambda$ is not a ground-state wavefunction in general, but is uniquely defined by Eq. (3.49), and is often referred to as the *conditional amplitude* [136]. The ground-state wavefunction Ψ^λ yields the spin density $n(x)$, so by construction we have, for all x ,

$$\int dx_2 \cdots \int dx_N |\tilde{\Psi}_x^\lambda(x_2, x_3, \dots, x_N)|^2 = 1. \quad (3.50)$$

We also identify

$$\tilde{n}_x^\lambda(\mathbf{r}') = (N - 1) \sum_{\sigma'} \int dx_3 \cdots dx_N |\tilde{\Psi}_x^\lambda(x', x_3, \dots, x_N)|^2. \quad (3.51)$$

That is, the total density of $\tilde{\Psi}_x^\lambda$ is the exact spin-CP density \tilde{n}_x^λ . Using Eq. (3.49), we generalize Ref. [245] and find an effective equation for $\sqrt{n(x)}$,

$$\left[H_1^\lambda + v_{\text{eff}}^\lambda(x) \right] \sqrt{n(x)} = E^\lambda \sqrt{n(x)}, \quad (3.52)$$

where E^λ is the ground-state energy of H^λ and the one-body effective potential is

$$\begin{aligned} v_{\text{eff}}^\lambda(x) = & \lambda \int d^3r' \frac{\tilde{n}_x^\lambda(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} + \langle \tilde{\Psi}_x^\lambda | \hat{H}_{N-1}^\lambda | \tilde{\Psi}_x^\lambda \rangle \\ & + \frac{1}{2} \int dx_2 \cdots dx_N |\nabla_{\mathbf{r}} \tilde{\Psi}_x^\lambda|^2. \end{aligned} \quad (3.53)$$

It is often partitioned this way, where the first term is referred to as the *conditional potential* and the last term is the *kinetic potential* [136]. From Eq. (3.52) we also have

$$\begin{aligned}\hat{H}_1^\lambda \sqrt{n(x)} \tilde{\Psi}_x^\lambda &= (E^\lambda - v_{\text{eff}}^\lambda(x)) \sqrt{n(x)} \tilde{\Psi}_x^\lambda \\ &\quad - \nabla_{\mathbf{r}} \sqrt{n(x)} \cdot \nabla_{\mathbf{r}} \tilde{\Psi}_x^\lambda - \frac{\sqrt{n(x)}}{2} \nabla_{\mathbf{r}}^2 \tilde{\Psi}_x^\lambda.\end{aligned}\tag{3.54}$$

This yields an apparent Schrödinger equation for $\tilde{\Psi}_x^\lambda$ with Hamiltonian [135]:

$$\tilde{H}_x^\lambda = \hat{H}_{N-1}^\lambda + \sum_{j=2}^N \frac{\lambda}{|\mathbf{r} - \mathbf{r}_j|} + v_{\text{nuc}, x}^\lambda(x_2, \dots, x_N)\tag{3.55}$$

where

$$\begin{aligned}v_{\text{nuc}, x}^\lambda(x_2, \dots, x_N) &= \\ \frac{-2 \nabla_{\mathbf{r}} \sqrt{n(x)} \cdot \nabla_{\mathbf{r}} \tilde{\Psi}_x^\lambda - \sqrt{n(x)} \nabla_{\mathbf{r}}^2 \tilde{\Psi}_x^\lambda}{2 \sqrt{n(x)} \tilde{\Psi}_x^\lambda}\end{aligned}\tag{3.56}$$

which is a non-multiplicative potential and includes gradients of $\tilde{\Psi}_x^\lambda$ with respect to \mathbf{r} . The full equation reads:

$$\tilde{H}_x^\lambda \tilde{\Psi}_x^\lambda = v_{\text{eff}}^\lambda(x) \tilde{\Psi}_x^\lambda.\tag{3.57}$$

This is *not* a usual eigenvalue equation that you solve with given boundary conditions [142]. It is an inhomogenous differential equation satisfied by $\tilde{\Psi}_x^\lambda$, defined by Eq. (3.49). The total density of the solution $\tilde{\Psi}_x^\lambda$ is the exact spin-CP density $\tilde{n}_x^\lambda(\mathbf{r}')$, see Eq. (3.51). Eq. (3.57) is an example of the exact factorization technique, which is typically used in studying nuclear dynamics [7, 8, 13], but can also be applied to the pure electronic problem [363, 142, 212, 353]. The solution $\tilde{\Psi}_x^\lambda$ is not always the lowest eigenstate if one treats this as an inhomogenous eigenvalue problem, see Appendix B in Ref. [142]. That is, in general, $\tilde{\Psi}_x^\lambda$ does *not* correspond to the $(N - 1)$ -electron ground-state wavefunction in CP-DFT for CP potential

$\tilde{v}_x^\lambda(\mathbf{r}') = v[\tilde{n}_x^\lambda](\mathbf{r}')$ and interaction strength λ . Worse, $v_{\text{nuc},x}^\lambda$ depends on $N - 1$ coordinates simultaneously, so the usual theorems of DFT cannot be applied. For these reasons, the general relationship between CP-DFT and exact electron factorization (EEF) is subtle, contrary to the connections made between the EEF and DFT through Eq. (3.52) [213]. However, this analysis can still be useful to CP-DFT in some limits. In Section 3.5.3, we discuss the limit $|\mathbf{r}| \rightarrow \infty$, where $v_{\text{nuc},x}^\lambda$ vanishes everywhere, and we explore implications to CP-DFT.

3.5 Exact conditions

Throughout, it will be convenient to define a CP correction potential, $\Delta\tilde{v}_\mathbf{r}^{\alpha,\lambda}(\mathbf{r}')$, which is simply the difference between the CP potential and the external potential:

$$\begin{aligned}\Delta\tilde{v}_x^{\alpha,\lambda}(\mathbf{r}') &\equiv \tilde{v}_x^{\alpha,\lambda}(\mathbf{r}') - v^\lambda[n](\mathbf{r}') \\ &= \tilde{v}^\alpha[\tilde{n}_x^\lambda](\mathbf{r}') - v^\lambda[n](\mathbf{r}') .\end{aligned}\tag{3.58}$$

3.5.1 Two electron spin-singlet systems

For $N = 1$, $n_\mathbf{r}^\lambda(\mathbf{r}') = 0$ since the CP density must normalize to 0 and thus there is naturally no self-interaction error in CP-DFT for single electron systems [333]. If $N = 2$, the CP density has just one electron and the α -dependence and the exact v_{HXC} vanish. If the $N = 2$ electron system is unpolarized, i.e. a spin singlet, the CP densities are

$$\tilde{n}_\mathbf{r}^\lambda(\mathbf{r}') = \frac{2|\Psi^\lambda(\mathbf{r}, \mathbf{r}')|^2}{n(\mathbf{r})} ,\tag{3.59}$$

yielding

$$\tilde{v}_{\text{s},\mathbf{r}}^\lambda(\mathbf{r}') = \tilde{v}_\mathbf{r}^\lambda(\mathbf{r}') = \frac{1}{2} \frac{\nabla'^2 \Psi^\lambda(\mathbf{r}, \mathbf{r}')}{\Psi^\lambda(\mathbf{r}, \mathbf{r}')} + \epsilon_\mathbf{r}^\lambda ,\tag{3.60}$$

where $\epsilon_{\mathbf{r}}^\lambda$ is the ground-state energy of the CP-KS potential, $\tilde{v}_{s,\mathbf{r}}^\lambda(\mathbf{r}')$. Because the wavefunction satisfies the Schrödinger equation, we find

$$\Delta\tilde{v}_{\mathbf{r}}^\lambda(\mathbf{r}') + \Delta\tilde{v}_{\mathbf{r}'}^\lambda(\mathbf{r}) = \frac{\lambda}{|\mathbf{r}' - \mathbf{r}|} - E^\lambda + \epsilon_{\mathbf{r}'}^\lambda + \epsilon_{\mathbf{r}}^\lambda. \quad (3.61)$$

In the limit $|\mathbf{r}' - \mathbf{r}| \rightarrow 0$ we have to leading order:

$$\Delta\tilde{v}_{\mathbf{r}}^\lambda(\mathbf{r}') \rightarrow \frac{\lambda}{2|\mathbf{r}' - \mathbf{r}|}, \quad |\mathbf{r}' - \mathbf{r}| \rightarrow 0. \quad (3.62)$$

In the next section, we will see that this exact condition can be more generally derived from the electron-electron cusp condition. In the exchange limit ($\lambda = 0$) for two electron spin-singlet systems we have $\tilde{n}_{\mathbf{r}}^{\lambda=0}(\mathbf{r}') = n(\mathbf{r}')/2$ exactly, which is independent of the reference position. From uniqueness, the corresponding CP-KS potential is then

$$\tilde{v}_{s,\mathbf{r}}^{\lambda=0}(\mathbf{r}') = v_s[n](\mathbf{r}'). \quad (3.63)$$

In Figure 3.1 for 1D He we plot the CP density in the exchange limit (half the total density), which is quite close to the CP densities that include correlation.

3.5.2 Cusp condition

The generalized electron coalescence cusp condition requires

$$\left. \frac{\partial n_{xc}^\lambda(x, u)}{\partial u} \right|_{u=0} = \lambda \tilde{n}_x^\lambda(\mathbf{r}), \quad (3.64)$$

where $\mathbf{u} = \mathbf{r}' - \mathbf{r}$, $u = |\mathbf{r}' - \mathbf{r}|$, and the left-hand side has been spherically averaged over $\mathbf{r} + \mathbf{u}$ [63]. Setting $\lambda = 0$, we see that the cusp condition is a purely correlation-driven effect: the exchange hole does not contribute to the cusp. For generalized Coulomb systems the CP

spin density follows

$$\frac{\partial \tilde{n}_x^\lambda(u)}{\partial u} \Big|_{u=0} = \lambda \tilde{n}_x^\lambda(\mathbf{r}), \quad (3.65)$$

for $\mathbf{r} \neq \mathbf{R}_i$, where \mathbf{R}_i are nuclei positions. Following Kato's theorem [201], the cusp condition is satisfied in CP-DFT with a CP correction potential that has the following condition:

$$\Delta \tilde{v}_{\mathbf{r}}^{\alpha, \lambda}(\mathbf{r}') \rightarrow \frac{\lambda}{2|\mathbf{r}' - \mathbf{r}|}, \quad |\mathbf{r}' - \mathbf{r}| \rightarrow 0. \quad (3.66)$$

3.5.3 Long-range limits

When the reference position \mathbf{r} is sent to infinity, we can generalize Ref. [110] for arbitrary λ to deduce the leading-order term of the ground-state wavefunction of the λ -interacting N -electron system as

$$\lim_{|\mathbf{r}| \rightarrow \infty} \Psi^\lambda(x, x_2, \dots, x_N) = \sqrt{\frac{n(x)}{N}} \tilde{\Psi}_{\hat{x}}^{\lambda, N-1}(x_2, \dots, x_N), \quad (3.67)$$

where $\tilde{\Psi}_{\hat{x}}^{\lambda, N-1}$ is an accessible ground-state of the ionized $(N-1)$ -system and is parametrically dependent on $\hat{x} = (\hat{\mathbf{r}}, \sigma)$, where $\hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|$ is the direction and σ is the spin of the electron sent to infinity. The $\hat{\mathbf{r}}$ dependence occurs when the ionized $(N-1)$ -system ground-state is otherwise degenerate [110]. The wavefunction $\tilde{\Psi}_{\hat{x}}^{\lambda, N-1}$ is a solution to Eq. (3.57) at the large \mathbf{r} limit. For a finite number of degenerate $(N-1)$ ground-states we expect $\tilde{\Psi}_{\hat{x}}^{\lambda, N-1}$ to remain fixed for small changes in $\hat{\mathbf{r}}$, therefore the gradients in Eq. (3.56) vanish and $v_{\text{nuc}, x}^\lambda$ vanishes (see Refs. [145] and [144] for an in-depth discussion and exceptions). Taking $\alpha = \lambda$, we identify to leading order:

$$\Delta \tilde{v}_x^{\lambda=\alpha}[n](\mathbf{r}') \rightarrow \frac{\lambda}{|\mathbf{r}' - \mathbf{r}|}, \quad |\mathbf{r}| \rightarrow \infty. \quad (3.68)$$

That is, in the large reference position limit, the CP correction potential in Eq. (3.68) simply represents an impurity perturbation that breaks the possible degeneracy in the ionized $(N - 1)$ -system ground-state.

We continue this analysis and denote the fully spin-decomposed density of the wavefunction $\tilde{\Psi}_{\hat{x}}^{\lambda, N-1}$ as $\tilde{n}_{\hat{x}}^{\lambda, N-1}(x')$ and the total density of this wavefunction as $\tilde{n}_{\hat{x}}^{\lambda, N-1}(\mathbf{r}')$. The latter corresponds to the asymptotic limit of the spin-CP density to leading order:

$$\lim_{|\mathbf{r}'| \rightarrow \infty} \tilde{n}_x^{\lambda}(\mathbf{r}') = \tilde{n}_{\hat{x}}^{\lambda, N-1}(\mathbf{r}'). \quad (3.69)$$

From the complementary principal, for spin-CP densities we obtain:

$$\lim_{|\mathbf{r}'| \rightarrow \infty} \tilde{n}_x^{\lambda}(\mathbf{r}') = \frac{\sum_{\sigma'} n(x') \tilde{n}_{\hat{x}'}^{\lambda, N-1}(x)}{n(x)}. \quad (3.70)$$

For unpolarized systems or the total CP density we obtain

$$\lim_{|\mathbf{r}'| \rightarrow \infty} \frac{\tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}')}{n(\mathbf{r}')} = \frac{\tilde{n}_{\hat{\mathbf{r}}'}^{\lambda, N-1}(\mathbf{r})}{n(\mathbf{r})}. \quad (3.71)$$

That is, for a given reference position \mathbf{r} , the asymptotic behavior of the total CP density is the same as the asymptotic behavior of the total density of the system, up to a multiplicative constant. This can be seen in Figure 3.1. For Be, the ionized $N - 1$ system has a 1s orbital very similar to that of the neutral system, so for $y = 0$ the ratio appearing in Eq. (3.71) is roughly unity and the CP density and total density match far from the nucleus. For He and $y = 0$ the ratio is roughly 1/2 and we see that the CP density and half the total density overlap closely far away.

At large $|\mathbf{r}'|$ the highest occupied KS orbital dominates the density [245] and we have

$$\left[-\frac{1}{2} \nabla'^2 + v_s[n](\mathbf{r}') \right] \sqrt{n(\mathbf{r}')} = \epsilon_{\text{HOMO}} \sqrt{n(\mathbf{r}')}. \quad (3.72)$$

Since $\tilde{n}_{\mathbf{r}}^{\lambda}(\mathbf{r}')$ is simply proportional to $n(\mathbf{r}')$ at large $|\mathbf{r}'|$, from uniqueness (up to a constant shift) we have

$$\tilde{v}_{s,\mathbf{r}}^{\lambda}(\mathbf{r}') \rightarrow v_s[n](\mathbf{r}') , \quad |\mathbf{r}'| \rightarrow \infty . \quad (3.73)$$

The CP-KS potential approaches the original KS potential far from the system. This can also be seen in Figure 3.1, where the CP-KS potentials approach the original KS potential far from the nucleus. We can use Eq. (3.73) to determine a corresponding exact condition for the CP correction potential defined in Eq. (3.58). From Eqs. (3.32) and (3.58) we have

$$\begin{aligned} \tilde{v}_{s,\mathbf{r}}^{\lambda}(\mathbf{r}') &= v^{\lambda}[n](\mathbf{r}') + \Delta\tilde{v}_{\mathbf{r}}^{\alpha,\lambda}(\mathbf{r}') + v_{\text{HXC}}^{\alpha}[\tilde{n}_{\mathbf{r}}^{\lambda}](\mathbf{r}') \\ &= v_s[n](\mathbf{r}') - v_{\text{HXC}}^{\lambda}[n](\mathbf{r}') + \Delta\tilde{v}_{\mathbf{r}}^{\alpha,\lambda}(\mathbf{r}') + v_{\text{HXC}}^{\alpha}[\tilde{n}_{\mathbf{r}}^{\lambda}](\mathbf{r}') . \end{aligned} \quad (3.74)$$

The asymptotic limit $|\mathbf{r}'| \rightarrow \infty$ of the usual KS-DFT potentials are well-known to leading order [108]:

$$v_{\text{HXC}}^{\lambda}[n](\mathbf{r}') \rightarrow \frac{\lambda(N-1)}{r'} , \quad |\mathbf{r}'| \rightarrow \infty . \quad (3.75)$$

Since the CP density $\tilde{n}_{\mathbf{r}}^{\lambda}$ integrates to $N-1$ electrons:

$$v_{\text{HXC}}^{\alpha}[\tilde{n}_{\mathbf{r}}^{\lambda}](\mathbf{r}') \rightarrow \frac{\alpha(N-2)}{r'} , \quad |\mathbf{r}'| \rightarrow \infty . \quad (3.76)$$

From Eqs. (3.73) - (3.76) we obtain

$$\Delta\tilde{v}_{\mathbf{r}}^{\alpha,\lambda}(\mathbf{r}') \rightarrow \frac{N(\lambda-\alpha)+2\alpha-\lambda}{r'} , \quad |\mathbf{r}'| \rightarrow \infty . \quad (3.77)$$

3.5.4 Strictly correlated electron limit

When $\lambda \rightarrow \infty$, we approach the strictly correlated electron (SCE) limit. The Hamiltonian of Eq. (3.46) has the expansion [136]

$$H^\lambda \rightarrow \lambda(\hat{V}_{\text{ee}} + \hat{V}^{\text{SCE}}) + \mathcal{O}(\sqrt{\lambda}), \quad \lambda \rightarrow \infty, \quad (3.78)$$

where the last term is kinetic and is subleading. In this limit we have defined

$$\hat{V}^\lambda[n] \rightarrow \lambda \hat{V}^{\text{SCE}}[n] = \lambda \sum_{i=1}^N v^{\text{SCE}}[n](\mathbf{r}_i), \quad \lambda \rightarrow \infty, \quad (3.79)$$

where $\hat{V}^{\text{SCE}}[n]$ is the one-body potential that minimizes the classical potential potential energy operator $\hat{V}_{\text{ee}} + \hat{V}^{\text{SCE}}[n]$ and delivers $n(\mathbf{r})$ as the ground-state density. The ground-state wavefunction of such a Hamiltonian collapses into a distribution that can be expressed as [282]

$$|\Psi^{\text{SCE}}(x, x_2, \dots, x_N)|^2 = \frac{1}{N!} \sum_{\mathcal{P}} \int d^3s \frac{n(\mathbf{s})}{N} \prod_{i=1}^N \delta^{(3)}(\mathbf{r}_i - \mathbf{f}_{\mathcal{P}(i)}(\mathbf{s})), \quad (3.80)$$

where \mathcal{P} denotes a permutation of $1, \dots, N$, ensuring $|\Psi^{\text{SCE}}|^2$ is symmetric with respect to exchanging the coordinates of identical particles. The co-motion functions, $\mathbf{f}_i(\mathbf{r})$, dictate the positions of correlated electrons given an electron at position \mathbf{r} . The co-motion functions satisfy cyclic group properties, with $\mathbf{f}_1(\mathbf{r}) \equiv \mathbf{r}$, $\mathbf{f}_2(\mathbf{r}) \equiv \mathbf{f}(\mathbf{r})$, $\mathbf{f}_3(\mathbf{r}) = \mathbf{f}(\mathbf{f}(\mathbf{r}))$, and so on such that $\mathbf{f}_{N+1}(\mathbf{r}) = \mathbf{r}$. See Ref. [371] for additional properties. In the SCE limit, the pair density

becomes

$$\begin{aligned} P^{\text{SCE}}(\mathbf{r}, \mathbf{r}') &= \sum_{i \neq j=1} \int d^3s \frac{n(\mathbf{s})}{N} \delta^{(3)}(\mathbf{r} - \mathbf{f}_i(\mathbf{s})) \delta^{(3)}(\mathbf{r}' - \mathbf{f}_j(\mathbf{s})) \\ &= \sum_{i \neq j=1} \int d^3s \frac{n(\mathbf{s})}{N} \delta^{(3)}(\mathbf{f}_{N-i+2}(\mathbf{r}) - \mathbf{s}) \delta^{(3)}(\mathbf{f}_{N-j+2}(\mathbf{r}') - \mathbf{s}). \end{aligned} \quad (3.81)$$

We can neglect terms where neither i nor j are equal to 1 since the delta functions in the integral will not overlap for differing non-trivial co-motion functions. Evaluating the integral and applying cyclic properties we obtain

$$P^{\text{SCE}}(\mathbf{r}, \mathbf{r}') = n(\mathbf{r}) \sum_{i=2}^N \delta^{(3)}(\mathbf{r}' - \mathbf{f}_i(\mathbf{r})). \quad (3.82)$$

The CP density is then

$$\tilde{n}_x^{\text{SCE}}(\mathbf{r}') = \sum_{i=2}^N \delta^{(3)}(\mathbf{r}' - \mathbf{f}_i(\mathbf{r})). \quad (3.83)$$

Partitioning the SCE Hamiltonian in Eq. (3.78) like Eq. (3.46) and neglecting the subleading gradient terms allows us to identify a corresponding CP correction potential for $\alpha, \lambda \rightarrow \infty$,

$$\lambda \Delta \tilde{v}_x^{\text{SCE}}[n](\mathbf{r}') = \frac{\lambda}{|\mathbf{r}' - \mathbf{r}|}. \quad (3.84)$$

This is the potential required in this semiclassical limit. In this limit, we have simply fixed the “missing” electron at position \mathbf{r} as if it were a distinguishable particle and solve the resulting $(N - 1)$ electron system in the presence of this impurity potential, Eq. (3.84). In the context of classical statistical mechanics, this is equivalent to the Percus test particle procedure [310, 73, 21].

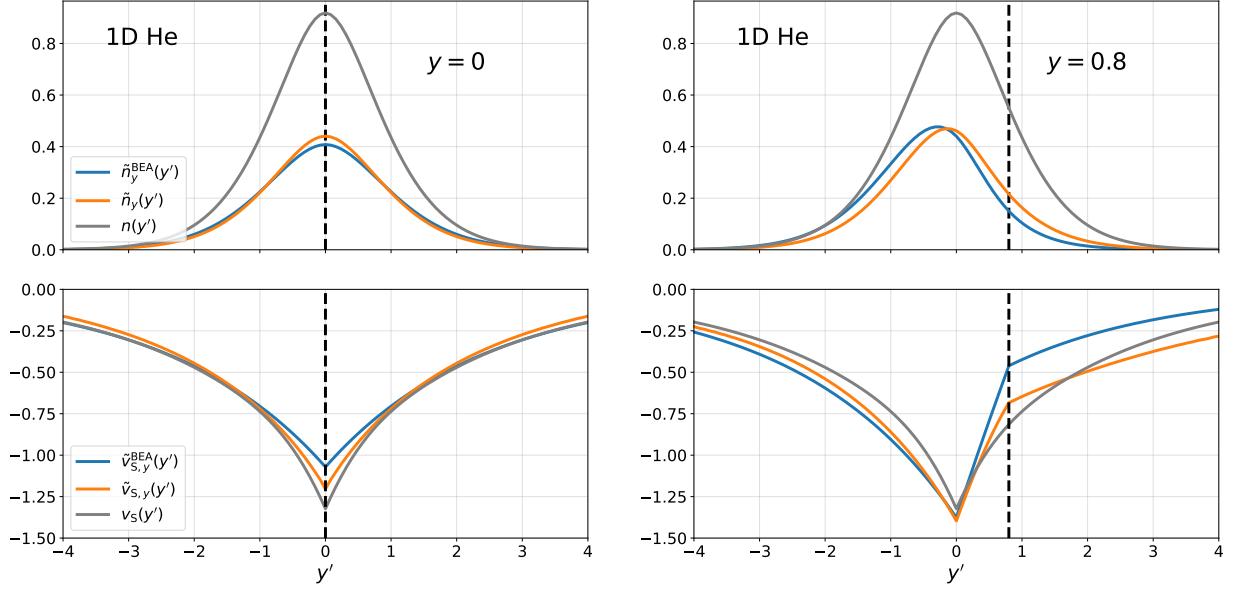


Figure 3.3: CP densities and potentials in 1D He ($\lambda = 1$): $\tilde{v}_{s,y}^{\text{BEA}}(y')$ is the 1D BEA approximation to the CP-KS potential with corresponding CP density $\tilde{n}_y^{\text{BEA}}(y')$, $\tilde{v}_{s,y}(y')$ is the exact CP-KS potential with corresponding exact CP density $\tilde{n}_y(y')$, $v_s(y')$ the exact KS potential with corresponding exact ground-state density $n(y')$, and $v(y')$ the external potential for our 1D He atom. Quantities are plotted for reference position $y = 0$ (left) and for $y = 0.8$ (right).

3.5.5 Dissociation limit

Consider a simple example of bond dissociation, the stretched H₂ bond. At the dissociation limit, the exact wavefunction has the Heitler-London [169] form ($\lambda \neq 0$):

$$\Psi^\lambda(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} (\phi_A(\mathbf{r}_1) \phi_B(\mathbf{r}_2) + \phi_B(\mathbf{r}_1) \phi_A(\mathbf{r}_2)) \quad (3.85)$$

where ϕ_A and ϕ_B are atomic H orbitals localized on each of the two protons. This yields a CP density:

$$\tilde{n}_\mathbf{r}^\lambda(\mathbf{r}') = n_B(\mathbf{r}'), \quad \mathbf{r} \text{ near } A, \quad (3.86)$$

and vice versa, and vanishes elsewhere. Thus the total electron-electron repulsion potential energy from Eq. (3.43) vanishes due to the lack of overlap, and each atomic region correctly yields the one-electron energy of a isolated hydrogen atom.

We can also consider arbitrarily long neutral H_N chains. In the spin-singlet case, where N is an even number, we obtain a generalized version of Eq. (3.85) at the dissociation limit ($\lambda \neq 0$):

$$\Psi^\lambda(\mathbf{r}_1, \dots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \sum_{\mathcal{P}} \prod_{i=1}^N \phi_i(\mathbf{r}_{\mathcal{P}(i)}), \quad (3.87)$$

where ϕ_i are atomic H orbitals localized on nuclei i , with position \mathbf{R}_i , in the enumerated H_N chain. The CP density is:

$$\tilde{n}_{\mathbf{r}}^\lambda(\mathbf{r}') = \sum_{i \neq j=1}^N n_i(\mathbf{r}'), \quad \mathbf{r} \text{ near } \mathbf{R}_j. \quad (3.88)$$

Again, the total electron-electron repulsion potential energy from Eq. (3.43) vanishes due to the lack of overlap, and each atomic region correctly yields a one-electron energy of a isolated hydrogen atom. The corresponding CP potential is

$$\tilde{v}_{\mathbf{r}}^{\lambda,\alpha}(\mathbf{r}') = - \sum_{i \neq j=1}^N \frac{1}{|\mathbf{r}' - \mathbf{R}_i|}, \quad \mathbf{r} \text{ near } \mathbf{R}_j. \quad (3.89)$$

Hence the CP correction potential is then

$$\Delta \tilde{v}_{\mathbf{r}}^{\lambda,\alpha}(\mathbf{r}') = \frac{1}{|\mathbf{r}' - \mathbf{R}_j|}, \quad \mathbf{r} \text{ near } \mathbf{R}_j. \quad (3.90)$$

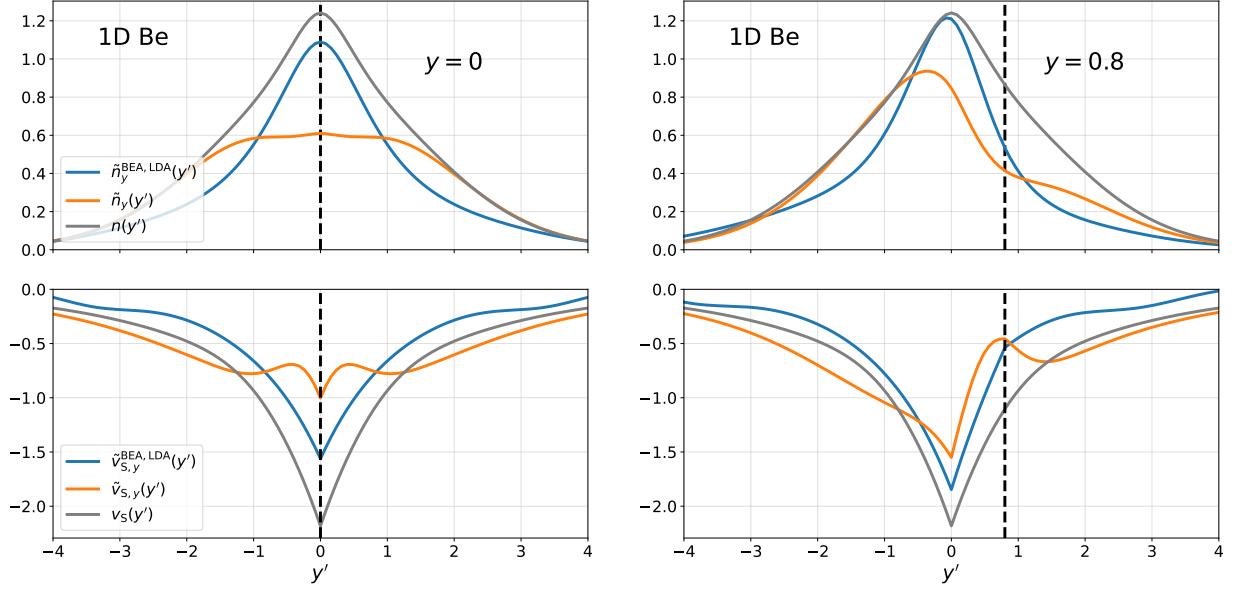


Figure 3.4: CP densities and potentials in 1D Be ($\lambda = 1$): $\tilde{v}_{s,y}^{\text{BEA,LDA}}(y')$ is the 1D BEA approximation to the CP-KS potential using the 1D LDA [29] approximation, see Eq. (3.96). The corresponding CP density is $\tilde{n}_{s,y}^{\text{BEA,LDA}}(y')$.

3.6 Blue electron approximation

In the *blue electron approximation* (BEA) we make the simple approximation

$$\Delta \tilde{v}_{\mathbf{r}}^{\lambda=\alpha, \text{BEA}}(\mathbf{r}') \equiv \frac{\lambda}{|\mathbf{r}' - \mathbf{r}|} \quad (3.91)$$

to the CP correction potential defined in Eq. (3.58). We call this the blue electron approximation because it corresponds to the classical result, Eq. (3.84). It is as if our reference electron was a distinguishable particle, painted blue for instance, fixed at position \mathbf{r} , yielding an impurity potential, Eq. (3.91). Using the BEA and setting $\lambda = \alpha$ in Eq. (3.74) we obtain

$$\tilde{v}_{s,\mathbf{r}}^{\lambda, \text{BEA}}(\mathbf{r}') = v_s[n](\mathbf{r}') + \frac{\lambda}{|\mathbf{r}' - \mathbf{r}|} + v_{\text{HXC}}^{\lambda}[\tilde{n}_{\mathbf{r}}^{\lambda}](\mathbf{r}') - v_{\text{HXC}}^{\lambda}[n](\mathbf{r}') . \quad (3.92)$$

While simple and semiclassical in origin, the BEA satisfies a surprising number of exact conditions discussed in this work.

In the exchange limit ($\lambda = 0$), the BEA CP correction potential and $v_{\text{HXC}}^{\lambda=0}$ terms vanish, leaving $\tilde{v}_{s,\mathbf{r}}^{\lambda=0,\text{BEA}}(\mathbf{r}') = v_s[n](\mathbf{r}')$. For two electron spin-singlet systems, this is exact, see Eq. (3.63), but not otherwise. Therefore, in general we do not expect the BEA to perform well for $N > 2$, unless exchange-limit corrections are also incorporated into the CP correction potential.

Taking a reference position \mathbf{r} that is far away from the system we see that BEA matches the exact asymptotic behavior in Eq. (3.68) by construction. Similarly, taking $\lambda = \alpha$ in Eq. (3.77), we obtain

$$\Delta\tilde{v}_{\mathbf{r}}^{\lambda=\alpha}(\mathbf{r}') \rightarrow \frac{\lambda}{r'}, \quad |\mathbf{r}'| \rightarrow \infty, \quad (3.93)$$

which matches the asymptotics of BEA in Eq. (3.92) if the exact asymptotic decay of $v_{\text{HXC}}^{\lambda}(\mathbf{r}')$, Eq.(3.75), is assumed. However, for approximate XC functionals, especially local and semilocal functionals, this condition is usually violated [226].

We can also show that BEA correctly dissociates neutral H_N chains. In the dissociation limit, the exact $v_{\text{HXC}}^{\lambda}(\mathbf{r}')$ terms will vanish and we obtain:

$$\tilde{v}_{s,\mathbf{r}}^{\lambda,\text{BEA}}(\mathbf{r}') = \frac{\lambda}{|\mathbf{r}' - \mathbf{r}|} - \sum_{i=1}^N \frac{1}{|\mathbf{r}' - \mathbf{R}_i|}. \quad (3.94)$$

In the absence of the first term, this KS potential would yield an N -fold degenerate ground-state eigenvalue which is equal to the H atom ground-state energy. For $\lambda \neq 0$ and when \mathbf{r} is near $\{\mathbf{R}_i\}$, the first term breaks some degeneracy: the ground-state eigenvalue has the same value as before but is now $(N - 1)$ -fold degenerate. The associated ground-state CP-KS

orbital, $\tilde{\phi}_{0,\mathbf{r}}$, is

$$\tilde{\phi}_{0,\mathbf{r}}(\mathbf{r}') = \frac{1}{\sqrt{N-1}} \sum_{i \neq j=1}^N \phi_i(\mathbf{r}'), \quad \mathbf{r} \text{ near } \mathbf{R}_j, \quad (3.95)$$

where ϕ_i are atomic H orbitals centered at position \mathbf{R}_i . The resulting CP density is the exact one, Eq. (3.89). If an approximate $v_{\text{HXC}}^\lambda(\mathbf{r}')$ is used, and there exists a self-interaction error for 1 electron systems, the atomic H orbitals in Eq. (3.95) are correspondingly approximate.

N	Symbol	$V_{\text{ee}}[n, \tilde{n}^{\text{BEA}}]$	$V_{\text{ee}}[n^{\text{LDA}}, \tilde{n}_x^{\text{BEA, LDA}}]$	$V_{\text{ee}}^{\text{HF}}$	$V_{\text{ee}}^{\text{exact}}$
2	He	0.659 (-0.031)	0.650 (-0.040)	0.722 (0.032)	0.690
2	Li^+	0.739 (-0.016)	0.735 (-0.020)	0.773 (0.018)	0.755
2	Be^{++}	0.779 (-0.013)	0.778 (-0.014)	0.802 (0.010)	0.792
3	Li	-	1.741 (0.094)	1.682 (0.035)	1.647
3	Be^+	-	1.971 (0.113)	1.881 (0.023)	1.858
4	Be	-	3.314 (0.126)	3.360 (0.172)	3.188

Table 3.1: Electron-electron repulsion energies, V_{ee} , for 1D systems using various methods. All energies are given in Hartree units. Errors with the exact are given in parenthesis. For $N = 3$ (spin-polarized systems), spin-CP-DFT calculations were performed self-consistently for each spin channel. 1D Hartree-Fock (HF) results are given as reference. Exact results are from 1D DMRG calculations [29]. The local BEA of the main text performs better for 3D Coulomb-interacting systems [270].

The BEA was proposed for 3D Coulomb-repelling electrons, for which it works well [270]. For 1D exponential repulsion, it is less accurate. In 1D the BEA takes the analog, $\Delta\tilde{v}_y^{\lambda, \text{BEA}}(y') = \lambda A \exp(-\kappa|y - y'|)$. For $\lambda = 1$ we plot the BEA CP potential and density for 1D He in Figure 3.3. We see that the BEA is reasonably accurate, but it is clear that the BEA potential, Eq. (3.91), is overly repulsive and overestimates the CP potential at the reference position. For $N > 2$ and $\lambda = 1$ we use the 1D LDA approximation (parameterized in [29])

and solve the CP-KS equations self-consistently with the following CP-KS potential:

$$\tilde{v}_{s,y}^{\text{BEA,LDA}}(y') = v(y') + A \exp(-\kappa|y - y'|) + v_{\text{HXC}}^{\text{LDA}}[\tilde{n}_y](y'). \quad (3.96)$$

We plot this potential for 1D Be in Figure 3.4. The BEA results here are worse than in 1D He, because of the exchange-limit deficiency previously discussed.

We can also apply this potential in the spin-CP-DFT case to spin-polarized systems, such as 1D Li in Figure 3.5. For $\sigma = \downarrow$, the BEA spin-CP density closely approximates the exact and the BEA does yield an accurate exchange-limit spin-CP density, $n_{\uparrow}^{\text{LDA}}(y')$, which is similar to the exact, $\tilde{n}_{(y,\downarrow)}^{\lambda=0}(y') = n_{\uparrow}(y')$.

With $\lambda = 1$ only we can calculate the total electron-electron repulsion energy, V_{ee} , given the total ground-state density n and CP density \tilde{n}_y . In our 1D analog we have:

$$V_{\text{ee}}[n, \tilde{n}_y] = \frac{1}{2} \int dy n(y) \int dy' \tilde{n}_y(y') A \exp(-\kappa|y - y'|). \quad (3.97)$$

In Table 3.1 we tabulate total electron-electron repulsion energies obtained from various methods. While we note that the BEA is an exceptionally crude approximation in 1D, for two-electron ions it yields energies that have comparable errors to Hartree-Fock (HF).

In 3D reality with Coulombic interactions, it can be seen from Eq. (3.66) that the BEA yields a CP density with a cusp that is too large by a factor of 2. To remedy this, we interpolate with a local density approximation

$$\Delta \tilde{v}_{\mathbf{r}}^{\lambda, \text{LBEA}}[n](\mathbf{r}') \equiv \frac{\lambda}{2|\mathbf{r} - \mathbf{r}'|} \left(1 + \text{Erf} \left(\frac{|\mathbf{r} - \mathbf{r}'|}{r_s(n(\mathbf{r}))} \right) \right), \quad (3.98)$$

where $r_s = (3/(4\pi n))^{1/3}$ is the Wigner-Seitz radius at the reference point \mathbf{r} . This *local blue electron approximation* (LBEA) will yield the exact cusp condition by Eq. (3.66) and satisfies the same discussed exact conditions as the BEA. In Figure 3.6 we plot the CP

densities and potentials for the (3D) He atom in the $(x', y', z' = 0)$ plane for reference position $\mathbf{r} = (0.563, 0, 0)$. We see that despite noticeable errors in the LBEA CP potential, the corresponding CP density is rather accurate. The exact results were obtained using exact diagonalizations in the recently developed *gausslet* basis set [347, 270]. In Ref. [270], using the LBEA in Eq. (3.98) with HF-calculated total densities, it was found that $E_{\text{XC}}^{\text{LBEA}} = -1.0736$ Ha (error of 0.007 Ha from the exact XC energy). See Ref. [270] for results on other two-electron ions, Hooke's atom, and H_2 dissociation. Within APDFT, the Overhauser model approximation, which represents a radial screened Coulomb interaction, is a close analog to the LBEA in CP-DFT. In fact, using the Overhauser model, the correlation energy results for two-electron ions in Ref. [146] are quite similar to the LBEA results of Ref. [270]. However, in non-radially symmetric systems or systems with more than two electrons, the Overhauser model and LBEA can produce quite different results. For example, in stretched H_2 the Overhauser model produces noticeable energy errors [148] whereas LBEA obtains the correct dissociation limit [270].

The CP-DFT approach can be used to address the uniform electron gas (UEG). Because the UEG is translationally invariant, the dependence on the reference position can be dropped. In Ref. [270] the LBEA with an added repulsive Gaussian potential term was used in CP-DFT calculations to accurately approximate the XC energy per particle of the UEG at all r_s values. In addition, the accuracy does not deteriorate as the temperature of the UEG is raised or if a more primitive method (Thomas-Fermi, as opposed to KS-DFT) is used to calculate CP densities. At zero temperature, the added Gaussian potential is needed at high densities where exchange dominates but is not captured well with the LBEA alone. However, it was also necessary to dampen this Gaussian as r_s is increased, yielding a somewhat empirical procedure but high accuracy results [270]. More recently, Ref. [309] considers an alternative approach where a portion of the CP density is fixed. Here the parallel component of the spin-CP density, $\tilde{n}_\sigma(\mathbf{r}', \sigma)$ in Eq. (3.38), is fixed to the exact exchange limit expression for the UEG (which is known analytically [101]). A spin-CP-DFT calculation is then performed for

the antiparallel component. As highlighted in Section 3.4.2, using spin densities in spin-CP-DFT is complicated, as these components do not integrate up to integer particle numbers in general, see Eq. (3.42). However, in the setting of the UEG this is not a concern, as the particle number is infinite (or exceedingly large, in a practical calculation). Furthermore, the parallel component of the spin-CP density is fixed to the exchange limit, which does yield an integer particle number. By construction, this approach recovers the correct exchange energy in the high-density limit, and yields sensible accuracy for all other r_s in the UEG [309].

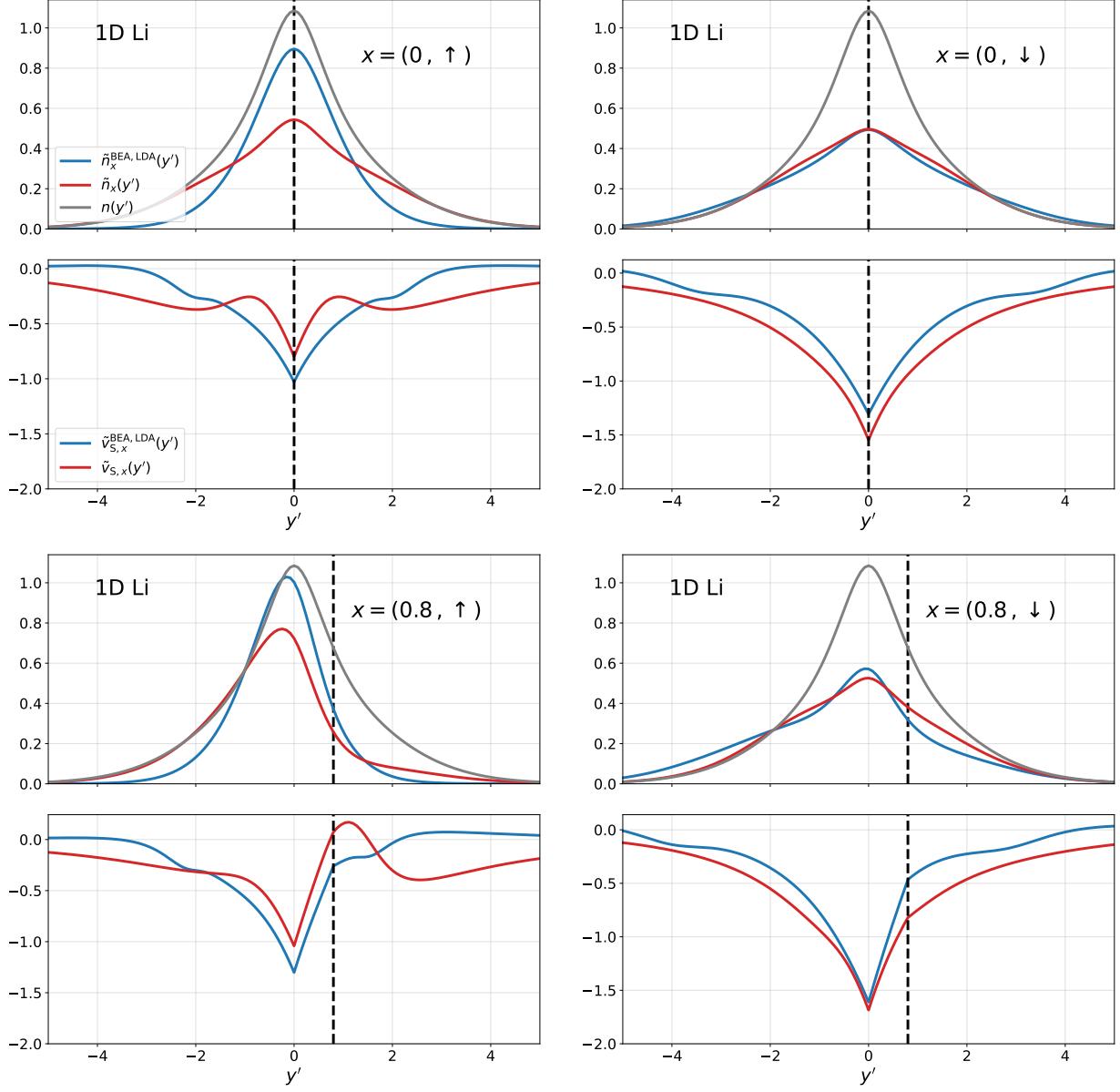


Figure 3.5: Spin-CP densities and potentials in 1D Li with $N_\uparrow = 2, N_\downarrow = 1$, and $\lambda = 1$: $\tilde{v}_{s,x}^{\text{BEA,LDA}}(y')$ is the 1D BEA approximation to the spin-CP-KS potential using the 1D LDA [29] approximation, see Eq. (3.96), with corresponding spin-CP density $\tilde{n}_{s,x}^{\text{BEA,LDA}}(y')$. The exact CP-KS potential is $\tilde{v}_{s,y}(y')$ with corresponding exact CP density $\tilde{n}_y(y')$, $v_s(y')$ the exact KS potential with corresponding exact ground-state density $n(y')$, and $v(y')$ the external potential for our 1D Li atom.

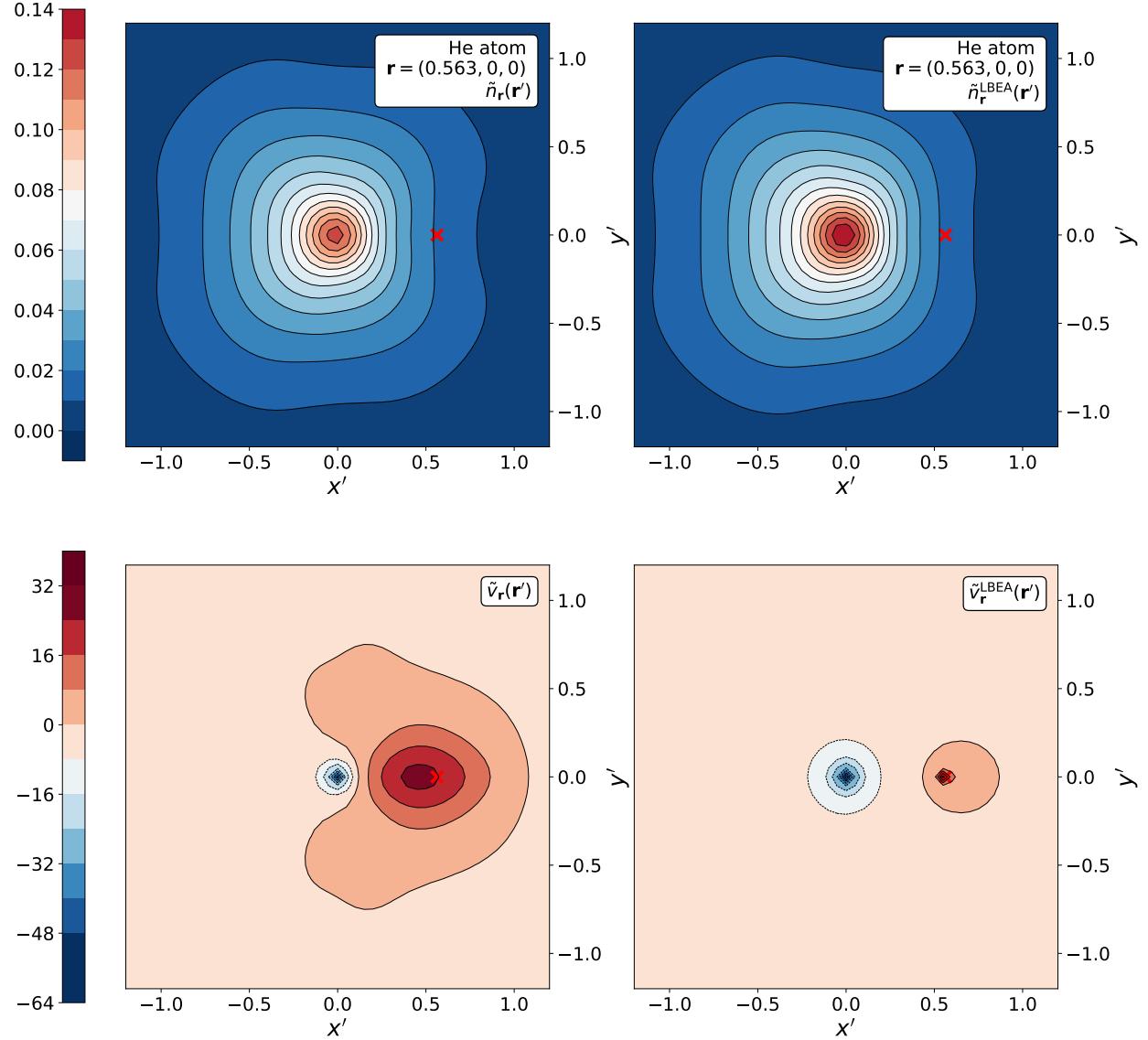


Figure 3.6: CP densities and potentials for (3D) He atom ($\lambda = 1$): contour plots for the exact CP density $\tilde{n}_r(\mathbf{r}')$ (top left) and corresponding exact CP-KS potential $\tilde{v}_r(\mathbf{r}')$ (bottom left) are plotted within the $(x', y', z' = 0)$ plane for reference position $\mathbf{r} = (0.563, 0, 0)$ (displayed as red cross). The corresponding LBEA results are plotted on the right.

3.7 Conclusions

In this work we present CP-DFT as a formally exact theory. CP-DFT is an approach to directly calculate CP densities $\tilde{n}_{\mathbf{r}}(\mathbf{r}')$ from ground-state calculations of $N - 1$ electrons in an external potential $\tilde{v}_{\mathbf{r}}(\mathbf{r}')$, known as the CP potential. In practice, the CP potential (or CP correction potential defined in Eq. (3.58)) used to obtain the CP density must be approximated. The XC energy can be extracted directly from the adiabatic connection formula using the ground-state density and CP densities. Notably, the CP-DFT approach bypasses the need for an XC energy functional approximation: only accurate ground-state densities and CP densities are needed to yield accurate XC energies within CP-DFT. We also introduce a formally exact spin-CP-DFT approach which can be used to obtain spin-CP densities. Similar to standard spin-DFT, we anticipate that the spin-CP-DFT approach may be more amenable to approximations that can address spin-polarized systems with more accuracy.

Throughout we present several exact conditions associated with CP densities and corresponding CP potentials, including: explicit expressions for two electron systems, long- and short-range asymptotic limits and conditions for general systems, the strictly correlated electron limit, and arbitrary neutral hydrogen chains at the dissociation limit.

We highlight an approximation, the *blue electron approximation* (BEA), which has semi-classical origins but, surprisingly, satisfies many exact conditions presented. An interpolated variation of BEA, the *local blue electron approximation* (LBEA), additionally satisfies the electron-electron cusp condition in resulting CP densities. For illustrative purposes, we provide select results for the BEA applied on 1D model systems for He, Li, and Be, as well as the LBEA applied on the (3D) He atom. Because the BEA was designed for 3D Coulomb-repelling systems, it works less well for 1D exponential repulsions, so we strongly recommend against using the 1D-mimic to study the efficacy of BEA. In previous work [270], the LBEA

was shown to yield usefully accurate results in several systems where standard DFT approximations can fail, such as single electron systems, stretched H_2 , and the hydrogen anion. In developing generalizeable CP-DFT approximations, it is clear that additional corrections are needed to obtain accurate CP and spin-CP densities in the exchange limit, which is not well captured using the BEA alone, except in two-electron singlet systems.

Exact conditions might play an important role in guiding construction or testing approximations in CP-DFT. In standard DFT, exact conditions on the energy functionals are most useful, rather than conditions on the densities or potentials. However, in CP-DFT, where an accurate CP density is the sole interest, only exact constraints on the CP densities and CP potentials are useful. Identifying further exact conditions on the CP density (or equivalently, the XC hole) and associating them to conditions on the CP potential may be helpful in the development of future CP-DFT approximations.

Overall, this work aims to provide a sound theoretical basis for CP-DFT and spin-CP-DFT. Simple examples are used throughout to demonstrate concepts and facilitate understanding. We hope that the content of this work will prompt future approximations and applications.

3.8 Acknowledgments

Work supported by DOE DE-SC0008696.

Chapter 4

Reassessing the role of exact conditions in density functional theory

This chapter is a reproduction of Ref. [303], which I co-authored with Kieron Burke.

4.1 Abstract

Exact conditions have long been used to guide the construction of density functional approximations. Nowadays hundreds of approximations are in common use, many of which neglect these conditions in their design. We analyze several well-known exact conditions and revive several obscure ones. Two crucial distinctions are drawn: that between necessary and sufficient conditions, and between all possible electronic densities and the subset of relevant Coulombic ground states. Simple search algorithms find violations of sufficient conditions while others construct densities that violate necessary conditions. We find that many empirical approximations satisfy many exact conditions for chemically relevant densities. We also find non-empirical approximations satisfy even more conditions than those enforced in

their construction.

4.2 Introduction

Modern density functional theory (DFT) calculations span many branches of the science of matter [296, 184, 343, 457]. In the standard Kohn-Sham approach [214], only the exchange-correlation (XC) energy need be approximated as a functional of the electronic (spin)-densities. Currently, hundreds of distinct XC approximations are available in standard DFT codes [266, 238], reflecting the immense difficulty in finding approximations that are generally accurate.

With the exact XC energy functional, Kohn-Sham DFT provides the exact ground-state density and energy for *any* electronic system. Exact conditions are known analytical properties of the exact functional and have played a vital role in the development of approximations [199]. The argument is that imposing exact conditions makes an approximation better resemble the exact functional, leading to improved generality. Typically, *non-empirical* functionals rely heavily on such conditions, eschew fitting to any chemical bonds, and work reasonably well for both materials and molecules. Such guiding principles led to a series of successful and widely used approximations, culminating in “Strongly Constrained and Appropriately Normed” (SCAN) [388] semilocal functional which attributes much of its success to the satisfaction of ‘all known’ (17) exact conditions that such a functional can satisfy.

On the other hand, many approximations tailored for molecular chemistry applications blatantly ignore exact conditions in their design [341]. Such approximations can be extremely accurate on comprehensive molecular benchmarks [140], where they are often more accurate than their more constrained counterparts. Typically, such chemically trained functionals behave poorly for materials.

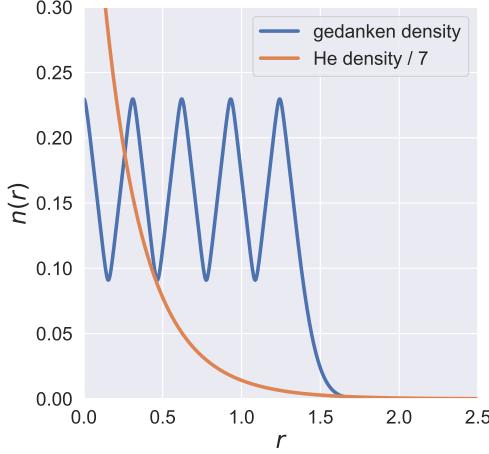


Figure 4.1: An unpolarized ground-state gedanken density with 2 electrons whose correlation energy is -21 mH in PBE, but +85 mH in LYP. For reference, the He atom density (divided by 7) is plotted.

We illustrate this difference with the correlation energy of the blue two-electron density in Fig 1, calculated with two generalized gradient approximations (GGA's). The first, the Perdew-Burke-Ernzerhof (PBE) correlation functional [313] adheres to many exact conditions and automatically satisfies the basic requirement that the correlation energy is never positive, yielding -21 mH. The second, the Lee-Yang-Parr (LYP) correlation functional [234], does not explicitly enforce many exact conditions and yields the nonsensical +85 mH. Yet LYP has been used successfully in over 100,000 chemical applications [60].

How can the success of these two seemingly disjoint design paradigms be rationalized? We resolve this paradox by reassessing the role of exact conditions in modern DFT approximations. To do this, we develop several new (and not so new) tools. We carefully parse the logic of exact conditions, finding that many enforced conditions are too strong for real matter. A computational scanning procedure finds violations, coupled with construction of corresponding reasonable (but not realistic) densities, as in Fig 1. Half a dozen exact conditions and hundreds of approximate functionals are analyzed. Several obscure conditions are revived and analyzed, while even well-known conditions yield surprising new twists. Finally, the role of exact conditions in density functional development is revisited.

4.3 Results

Begin with the correlation energy. In practice, approximations (denoted by tilde) have the form:

$$\tilde{E}_C[n] = \int d^3r n(\mathbf{r}) \tilde{\epsilon}_C[n](\mathbf{r}), \quad (4.1)$$

where $n(\mathbf{r})$ is an electronic density. While the developers define a conventional correlation energy per electron, $\tilde{\epsilon}_C[n](\mathbf{r})$, that is often implemented explicitly in DFT codes, other “gauges” exist yielding the same $\tilde{E}_C[n]$. For example, $\tilde{\epsilon}_C[n](\mathbf{r})$ and $\tilde{\epsilon}_C[n](\mathbf{r}) + \nabla^2(n^{2/3})/n$ yield identical $\tilde{E}_C[n]$ [325, 90]. Specifically, semilocal functionals can be written

$$\tilde{\epsilon}_C[n](\mathbf{r}) = \tilde{\epsilon}_C(r_s(\mathbf{r}), \zeta(\mathbf{r}), s(\mathbf{r}), \alpha(\mathbf{r}), q(\mathbf{r})), \quad (4.2)$$

where $r_s = (4\pi n/3)^{-1/3}$ is the Wigner-Seitz radius, $\zeta = (n_\uparrow - n_\downarrow)/n$ is the (dimensionless) spin polarization, $s = |\nabla n|/(2(3\pi^2)^{1/3}n^{4/3})$ is the (dimensionless) reduced gradient, $\alpha = (\tau - \tau^{VW})/\tau^{\text{unif}} \geq 0$ with $\tau = \sum_{i,\sigma}^{\text{occ.}} |\nabla \phi_{i,\sigma}|^2/2$, $\tau^{VW} = |\nabla n|^2/8n$, $\tau^{\text{unif}} = (3/20)(3\pi^2)^{2/3}n^{5/3}[(1 + \zeta)^{5/3} + (1 - \zeta)^{5/3}]$, and $q = \nabla^2 n/(4(3\pi^2)^{2/3}n^{5/3})$ is the reduced Laplacian. The local spin density approximations (LDA) depends only on r_s and ζ , generalized gradient approximations (GGAs) add dependence on s , while meta-generalized gradient approximations (MGGAs) can depend on all variables.

A simple exact condition is *correlation energy non-positivity*,

$$E_C[n] \leq 0, \quad (4.3)$$

which holds for any reasonable density, which we define as being positive, integrating to a finite quantity N , and have finite von Weizsäcker kinetic energy (\mathcal{I}_N of Ref. [255] or Eq. 34

of Ref. [429]). This is routinely enforced via

$$\tilde{\epsilon}_C[n](\mathbf{r}) \leq 0 \quad \text{for all } \mathbf{r} \text{ and any } n(\mathbf{r}). \quad (4.4)$$

Clearly, satisfying this local condition *guarantees* Eq. (4.3), but it is also excessive, i.e., not necessary. Moreover, starting from any $\tilde{\epsilon}_C[n](\mathbf{r})$ that satisfies Eq. (4.4), addition of $C \nabla^2(n^{2/3})/n$ violates it for sufficiently large C . If local violations of Eq. (4.4) *do* exist, then a counterexample density that violates the exact condition in Eq. (4.3) *might* be found. If it can be, the exact condition is violated for that density in any gauge. If no such counterexample can be found, the possibility that a gauge might be found that satisfies Eq. (4.4) remains open.

Returning to the LYP GGA, we found instances where $\epsilon_C^{\text{LYP}}(r_s, \zeta, s) > 0$ for $s \geq 1.74$, thus allowing the possibility of a violating *gedanken* density. Gedanken densities are thought experiment densities that need not be realistic [325]. So we construct a gedanken density that has large $s \geq 1.74$ values throughout its interior, violating the local condition in Eq. (4.4) (see Supplemental S1 for details). Importantly, we want such local violations to exist in *energetically relevant regions* of the density, that is, spatial regions that substantially contribute to the integral in Eq. 4.1. The gedanken density of Fig 4.1 is radial, nodeless, finite, continuous, differentiable to first and second order, and integrates to 2 electrons. It is a reasonable density and is also non-interacting v -representable, and when evaluated using the LYP correlation functional yields +85 mH. Thus, the LYP functional *can* violate correlation non-positivity.

But does LYP violate correlation non-positivity in the restricted space of realistic Coulombic densities, i.e, those ground-state densities of systems with Coulombic attractions to integer nuclear charges of small or no overall charge? The gedanken density of Fig 4.1 is not Coulombic: for instance, it lacks nuclear cusps as required by Kato's theorem [201].

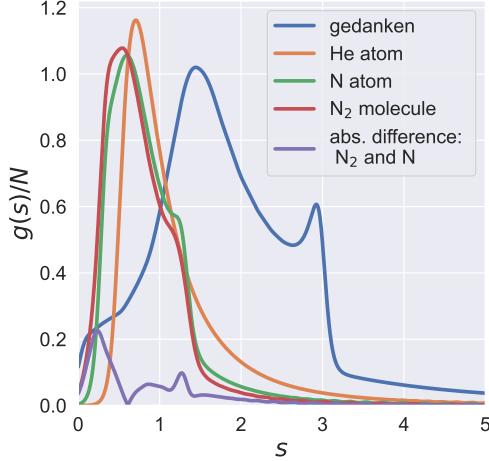


Figure 4.2: The distribution $g(s)$ is plotted for various ground-state densities: the gedanken density in Fig. 4.1, the He and N atoms, and the N_2 molecule. The absolute difference between the N_2 molecule and N atom distributions is also plotted. Details of the calculations can be found in Supplemental S2.

An important property of the gedanken density is that it has energetically relevant regions of the density with $s \geq 2$. The distribution

$$g(s) = \int d^3r n(\mathbf{r}) \delta^{(3)}(s - s(\mathbf{r})) , \quad (4.5)$$

was introduced in Refs. [476, 475], and $g(s)ds$ is the number of electrons in the system with reduced density gradient between s and $s + ds$, i.e, it is an analog of the density of states for energy levels. In Fig. 4.2, we plot $g(s)/N$ for various ground-state densities: the gedanken density in Fig. 4.1 and calculated densities for the He and N atoms and the N_2 molecule. Hartree-Fock densities are sufficiently accurate for our purposes. Unlike the Coulombic densities, the gedanken density $g(s)$ is centered around $s \approx 2$, as intended. For Coulombic systems [475], large $s > 2$ values are typically only found in the decaying tails of the density, and are energetically irrelevant. In molecular and extended systems, these tails (which may not even be present in periodic systems) are even less energetically relevant than their atomized counterparts [475] : in Fig. 4.2 the distribution $g(s)/N$ of N_2 is shifted to lower s than that of the atomized system (N atom) for $s > 1$. The electrons in the

bond between two (or more) atoms have smaller values of s which even vanishes at the bond center [475]. Although a single electron distributed across an infinitely separated chain of many protons has energetically relevant regions of large s , such a system is too far from neutral for our set.

Do large s values contribute importantly to energy differences, such as molecular binding energies? Valence electrons change considerably but their differences tend to also be energetically relevant only in regions of smaller s values : the difference $|g[\text{N}_2](s) - 2g[\text{N}](s)|$, is most prominent in regions $s < 2$ (see Fig. 4.2) [475]. In non-covalent bonding, s values up to ≈ 7 are relevant in binding energy differences for van der Waals (vdW) complexes [186, 287]. However, in these cases, typically a non-local correlation functional (such as DFT-D [71] or vdW-DF [154]) is employed which provides the bulk of the energetics that contributes to the binding of vdW complexes, whereas semilocal contributions are more relevant in structural aspects such as bonding distances and lattice constants [186].

Revisiting the LYP correlation energy example, we conjecture that no realistic Coulombic density ever yields a positive correlation energy. Such a density would need to have energetically relevant regions of the density with large $s > 1.74$, which is simply not observed in these systems. Over countless atomic and molecular densities, the LYP correlation functional has *not* yielded positive correlation energies [234, 278].

We perform an identical analysis on other popular approximations and tabulate the results in the first row of Table 4.1. For each exact condition (e.g. Eq. (4.3)), we check for violations of the corresponding local condition (e.g. Eq. (4.4)) for various semilocal approximations. To locate violations of such local conditions, we use a simple extensive grid search over applicable variables, $0 < r_s \leq 5, 0 \leq \zeta \leq 1, 0 \leq s \leq 5, 0 \leq \alpha \leq 5$, where the upper ranges are chosen to be reasonably large to encapsulate relevant regions of realistic Coulombic densities (see Supplemental S11 for more details). This brute-force numerical approach is necessary since analytical verification is often impossible with the complicated forms of modern XC

approximations. We utilize the Libxc [238] implementation of XC approximations to evaluate the XC energy densities in Eq. (4.2).

For a given local condition and approximation, if we find no violations in our grid search, then we assume that the approximation always satisfies the exact condition for any reasonable density (denoted by a \checkmark in Table 4.1). In other cases (denoted \checkmark^*), the corresponding local condition may be satisfied only for a restricted range of variable values. For example, for B3LYP correlation non-positivity, the local condition in Eq. 4.4 is satisfied whenever $s < 2.13$ and we display the bounds on the variable s that ensure satisfaction. In some cases (denoted \checkmark^{**}), we do not obtain a simple restricted range of variable values, but we find that local violations are exceedingly “rare” (less than 1% of the total number of permutations considered).

local condition	non-empirical			empirical				
	PBE	AM05	SCAN	B3LYP	CASE21	SOGGA11	M06	B97
E_c non-positivity (4.4)	\checkmark	\checkmark	\checkmark	$\checkmark^*, s < 2.13$	\checkmark	\checkmark	$\checkmark^*, \zeta = 0, s < 1.56$	$\checkmark^*, s < 1.42$
E_c scaling inequality (A.16)	\checkmark	\checkmark	\checkmark	$\checkmark^*, s < 2.15$	\checkmark	\checkmark^{**}	$\checkmark^*, \zeta = 0, s < 1.59$	$\checkmark^*, s < 1.52$
T_c upper bound (A.7)	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	$\checkmark^*, s < 1.36$	$\checkmark^*, \zeta = 0, s < 1.56$	$\checkmark^*, s < 1.62$
$U_c(\lambda)$ monotonicity (4.15)	\checkmark	\checkmark	\checkmark	$\checkmark^*, s < 1.82$	\checkmark	\times	$\checkmark^*, \zeta = 0, s < 1.56$	$\checkmark^*, s < 1.41$
LO extension to E_{xc} (A.18)	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	$\checkmark^*, 0.04 < s < 3.62$	$\checkmark^*, s < 4.46$
LO (A.10)	\checkmark	\checkmark	\checkmark	$\checkmark^*, s < 4.88$	\checkmark	$\checkmark^*, s < 4.98$	$\checkmark^*, 0.06 < s < 3.62$	$\checkmark^*, s < 4.43$

Table 4.1: For each condition, we assess if the local condition is satisfied (or partially satisfied) for an approximation (with more given in Supplemental S11).

The logic and concepts presented for correlation non-positivity generalize to other exact conditions on energy functionals. For instance, the *Lieb-Oxford (LO) bound [256] extension to XC energies* exact condition [312],

$$E_{xc}[n] \geq C_{\text{LO}} \int d^3r n \epsilon_{\text{x}}^{\text{unif}}[n](\mathbf{r}), \quad (4.6)$$

yields the local condition

$$\tilde{F}_{\text{xc}} \leq C_{\text{LO}}, \quad (4.7)$$

where we use $C_{\text{LO}} = 2.27$ in this work (although tighter bounds have been proven [246, 247]), $\tilde{F}_{\text{xc(C)}}[n](\mathbf{r}) \equiv \tilde{\epsilon}_{\text{xc(C)}}[n](\mathbf{r})/\epsilon_{\text{x}}^{\text{unif}}[n](\mathbf{r})$ is the exchange (correlation) enhancement factor with $\tilde{F}_{\text{xc}} = \tilde{F}_{\text{x}} + \tilde{F}_{\text{c}}$ and $\epsilon_{\text{x}}^{\text{unif}}[n](\mathbf{r}) = -(3/4\pi)(3\pi^2 n)^{1/3}$ is the exchange energy per particle of an unpolarized uniform electron gas. Many approximations enforce the local Eq. (A.18) to ensure Eq. 4.6.

Since the combined XC energy is the object of interest, some approximations fail to distinguish exchange and correlation. In the exact functional, one can extract E_{c} using uniform coordinate scaling [242, 243]:

$$E_{\text{c}}[n] = E_{\text{xc}}[n] - \lim_{\gamma \rightarrow \infty} \frac{E_{\text{xc}}[n_{\gamma}]}{\gamma}, \quad (4.8)$$

where $n_{\gamma}(\mathbf{r}) \equiv \gamma^3 n(\gamma \mathbf{r})$ and $\gamma > 0$. We apply this ‘‘conventional’’ partitioning to extract correlation energies where none have been defined or whose partitioning is ambiguous. E.g., for globals, Eq. (A.21) yields E_{c} of the semilocal form in Eq. (4.2), e.g., $\epsilon_{\text{c}}^{\text{B3LYP}} = 0.405\epsilon_{\text{c}}^{\text{LYP}} + 0.095\epsilon_{\text{c}}^{\text{VWN5}}$ [385]. This partitioning can differ from the developer’s intentions or rationalizations (see Supplemental S4). The LO bound applied to global hybrid functionals is discussed in Supplemental S5.

Besides Eq. (A.21), many other properties of the exact functional are written in terms of uniform coordinate scaling (or equivalently, through the adiabatic connection in DFT [164, 228, 156], see Supplemental S7). We simply list these ‘‘obscure’’ conditions and their local forms (further details and derivations can be found in the Supplemental S3). The *correlation*

uniform scaling inequality [242]

$$(\gamma - 1)E_C[n_\gamma] \geq \gamma(\gamma - 1)E_C[n] \quad (4.9)$$

has a corresponding local condition,

$$\frac{\partial \tilde{F}_C(r_s, \zeta, s, \alpha, q)}{\partial r_s} \geq 0. \quad (4.10)$$

The kinetic contribution to the correlation energy, T_C , is non-negative [242, 128]

$$T_C[n_\gamma] = \gamma \frac{dE_C[n_\gamma]}{d\gamma} - E_C[n_\gamma] \geq 0, \quad (4.11)$$

and shares the same local condition, Eq. A.16. The T_C upper bound [244, 241] reads

$$T_C[n_\gamma] \leq -\gamma \left(\frac{\partial E_C[n_\gamma]}{\partial \gamma} \Big|_{\gamma \rightarrow 0} \right) + E_C[n_\gamma], \quad (4.12)$$

with corresponding local condition,

$$\frac{\partial \tilde{F}_C}{\partial r_s} \leq \frac{\tilde{F}_C(\infty) - \tilde{F}_C}{r_s}, \quad (4.13)$$

where $\tilde{F}_C(\infty) = \tilde{F}_C(r_s \rightarrow \infty)$. Correlation energy adiabatic connection curves, $U_C(\lambda) = d(\lambda^2 E_C[n_{1/\lambda}])/d\lambda$, satisfy a *monotonicity condition* [243],

$$\frac{dU_C(\lambda)}{d\lambda} \leq 0, \quad (4.14)$$

with corresponding local condition [243]

$$\frac{\partial}{\partial r_s} \left(r_s^2 \frac{\partial \tilde{F}_C}{\partial r_s} \right) \geq 0. \quad (4.15)$$

The LO bound [256], often generalized as Eq. 4.6, is precisely

$$U_{\text{xc}}[n] \geq C_{\text{LO}} \int d^3r n \epsilon_{\text{x}}^{\text{unif}}[n](\mathbf{r}), \quad (4.16)$$

where $U_{\text{xc}}[n] = E_{\text{xc}}[n] - T_{\text{c}}[n]$ is the potential correlation energy. The corresponding local condition,

$$\tilde{F}_{\text{xc}} + r_s \frac{\partial \tilde{F}_{\text{c}}}{\partial r_s} \leq C_{\text{LO}}, \quad (4.17)$$

is *more* restrictive than the commonly used Eq. (A.18). Approximations satisfying Eq. (A.18) need not satisfy Eq. (A.10), such as B3LYP or SOGGA11 in Table 4.1. Results for a conjectured condition, $T_{\text{c}}[n] \leq -E_{\text{c}}[n]$ [242, 89, 128], are in Supplemental S6, S10, and S11.

This work does not provide a comprehensive study of all known exact conditions in DFT. A unified subset of several conditions (6) were chosen to illustrate the logic. But we have not touched on the self-interaction error [333], the asymptotic behavior of exchange and correlation potentials [15], or the flat-plane energy condition for fractional charges and spins [284, 79]. We expect our logic can be fruitfully applied to any exact condition in DFT.

We describe the conditions Eqs. A.12 - A.10 as obscure because, while proven several decades ago, none appear to be deliberately and generally enforced in modern approximations, even those that strive to satisfy as many exact conditions as possible. SCAN was designed to satisfy the correlation uniform scaling inequality (Eq. (A.12)), but only in extreme limits, $\gamma \rightarrow 0, \infty$ [388, 199]. The corresponding local condition in Eq. (A.16) *is* satisfied in SCAN, but adjustments of its parameters chosen to fit *appropriate norms* can produce violations (see Supplemental S9). Norms refer to properties of specific (but not bonded) reference systems, such as the uniform gas, the hydrogen atom, or noble gas dimers [388, 43]. So, by enforcing appropriate norms, SCAN satisfies *more* exact conditions than were explicitly included. In

other instances, such as PW91 [312] and LDA, the satisfaction of several of the obscure exact conditions was proven analytically after their publication [243] (Supplemental S8 is a proof that PBE satisfies Eq. A.12). We find (numerically) that the non-empirical PBE, AM05, and SCAN functionals, which were designed to satisfy a large set of exact conditions, also satisfy many additional exact conditions outside of the original set (Table 4.1). CASE21, a recent machine-learned empirical functional designed to adhere to select exact conditions [383], also satisfies these obscure conditions.

In Table 4.1, we also show that many empirical approximations satisfy local conditions in energetically relevant regions of realistic Coulombic densities, i.e., for s values that are not too large. When assessing our set of exact conditions on the set of coordinate-scaled HF densities for atoms H-Ar and their cations, we find that *all* are satisfied (see Supplemental S10), suggesting that these approximations will satisfy these conditions in the space of realistic Coulombic densities (possibly excepting the monotonicity condition in SOGGA11). This is intriguing because most such empirical approximations were designed *without* explicit adherence to these exact conditions. This finding appears to reinforce their importance in approximations: satisfaction of such esoteric exact conditions is hardly accidental. Furthermore, empirical approximations often employ ingredients, such as the dimensionless quantities s, ζ, α, q , which themselves were chosen to simplify satisfaction of exact conditions. In consequence, nearly all empirical approximations satisfy two simple exact constraints on the exchange energy: uniform coordinate scaling [242] and spin scaling [297].

Our results suggest a reassessment of the role of exact conditions in modern density functional development. Is it excessive to enforce strong local conditions to ensure the satisfaction of exact conditions? Flexible empirical model approximations also satisfy many exact conditions on the (highly relevant) space of realistic Coulombic densities and achieve high accuracy for molecular processes. When empirical functionals are locally constrained to satisfy conditions, their molecular benchmark performance is similar to the suboptimal

performance of their non-empirical counterparts, despite the advantage of training on molecular data [383, 99, 290], and the resulting functionals tend to closely mimic non-empirical counterparts, such as SCAN [99, 290].

So when is the enforcement of local conditions helpful? The answer appears to lie in the paucity of highly accurate data for solids. Empirical approximations are mostly developed for molecules, where copious benchmark data is now available, and are rarely even available in materials codes (as they typically perform poorly for solids). On the other hand, non-empirical functionals that dominate materials calculations include norms such as the uniform gas limit, which take the place of highly accurate data. By reducing to this limit, non-empirical functionals are guaranteed to yield moderately accurate results for solids. Moreover, the leading corrections in the asymptotic limit differ qualitatively between molecules and solids, because all molecules have turning surfaces at the Kohn-Sham HOMO energy, while few solids do [198]. This produces conflicting requirements on the gradient expansion of the approximation, as shown in the differences between PBE (good for atomic and molecular energies) and PBEsol (good for solid geometries and vibrations [323]). Such conflicts are resolved in SCAN, yielding improved results for both. Even in the presence of highly accurate data for solids, approximations that fit their data will worsen their results for molecules.

Even for molecular systems, non-empirical approximations like SCAN may also outperform empirical approximations with the same ingredients, especially when applied to new systems or new properties outside their training [199, 271, 197]. For instance, when evaluated on the artificial molecules in the MB16 benchmark [216, 140] (never used to parameterize empirical functionals), SCAN tends to outperform empirical approximations, including hybrids [199], suggesting a greater ability to extrapolate.

4.4 Conclusion

In conclusion, this work explores and revives several exact conditions in DFT and provides analysis on their satisfaction in approximations. Empirical-based approximations, which often employ flexible models that do not explicitly enforce exact conditions, typically violate exact conditions on some densities, but do not violate them for realistic Coulombic systems. As such approximations are trained on such systems, relevant exact conditions are implicitly enforced. Specific exact conditions (and appropriate norms) can be especially important when lacking reference data, where their enforcement (either approximate or exact) helps maintain the transferability of resulting approximations, e.g., from molecules to solids. These considerations may aid future approximations to achieve high accuracy that is maintained across many different classes of systems: molecules, solids, and everything in between, such as interfaces, clusters, and metal-organic frameworks.

4.5 Acknowledgments

Work supported by DOE DE-SC0008696. We thank John P. Perdew and Aaron D. Kaplan for useful discussions.

Chapter 5

Kohn-Sham equations as regularizer: building prior knowledge into machine-learned physics

This chapter is a production of Ref. [252], which I co-authored with Li Li, Stephan Hoyer, Ruoxi Sun, Ekin D. Cubuk, Patrick Riley, and Kieron Burke.

5.1 Abstract

Including prior knowledge is important for effective machine learning models in physics, and is usually achieved by explicitly adding loss terms or constraints on model architectures. Prior knowledge embedded in the physics computation itself rarely draws attention. We show that solving the Kohn-Sham equations when training neural networks for the exchange-correlation functional provides an implicit regularization that greatly improves generalization. Two separations suffice for learning the entire one-dimensional H_2 dissociation

tion curve within chemical accuracy, including the strongly correlated region. Our models also generalize to unseen types of molecules and overcome self-interaction error.

5.2 Introduction

Differentiable programming [34] is a general paradigm of deep learning, where parameters in the computation flow are trained by gradient-based optimization. Based on the enormous development in automatic differentiation libraries [58, 183, 301, 5], hardware accelerators [192] and deep learning [231], this emerging paradigm is relevant for scientific computing. It supports extremely strong physics prior knowledge and well-established numerical methods [182] and parameterizes the approximation by a neural network, which can approximate any continuous function [176]. Recent highlights include discretizing partial differential equations [31], structural optimization [178], sampling equilibrium configurations [295], differentiable molecular dynamics [366], differentiable programming tensor networks [254], optimizing basis sets in Hartree-Fock [395] and variational quantum Monte Carlo [171, 451].

Density functional theory (DFT), an approach to electronic structure problems, took an enormous step forward with the creation of the Kohn-Sham (KS) equations [214], which greatly improves accuracy from the original DFT [174, 402, 118]. The results of solving the KS equations are reported in tens of thousands of papers each year [190]. Given an approximation to the exchange-correlation (XC) energy, the KS equations are solved self-consistently. Results are limited by the quality of such approximations, and a standard problem of KS-DFT is to calculate accurate bond dissociation curves [387]. The difficulties are an example of strong correlation physics as electrons localize on separate nuclei [78].

Naturally, there has been considerable interest in using machine learning (ML) methods to improve DFT approximations. Initial work [253, 378] focused on the KS kinetic energy, as

a sufficiently accurate approximation would allow by-passing the solving of the KS equations [59, 250]. For XC, recent works focus on learning the XC potential (not functional) from inverse KS [187], and use it in the KS-DFT scheme [411, 365, 474, 288]. An important step forward was made last year, when it was shown that a neural network could find functionals using only three molecules, by training on both energies and densities [289], obtaining accuracy comparable to human-designed functionals, and generalizing to yield accurate atomization energies of 148 small molecules [93]. But this pioneering work does not yield chemical accuracy, nor approximations that work in the dissociation limit. Moreover, it uses gradient-free optimization which usually suffers from poor convergence behavior on the large number of parameters used in modern neural networks [105, 265, 355].

Here, we show that all these limitations are overcome by incorporating the KS equations themselves into the neural network training by backpropagating through their iterations – a *KS regularizer* (KSR) to the ML model. In a traditional KS calculation, the XC is given, the equations are cycled to self-consistency, and all previous iterations are ignored in the final answer. In other ML work, functionals are trained on either energies alone [137, 369, 44, 360], or even densities [365, 474, 283], but only after convergence. By incorporating the KS equations into the training, thereby learning the relation between density and energy at every iteration, we find accurate models with very little data and much greater generalizability.

Our results are illustrated in Figure 5.1, which is for a one-dimensional mimic of H_2 designed for testing electronic structure methods [29]. The distribution of curves of the ML model directly predicting E from geometries (direct ML) in (a) clearly fails to capture the physics. Next we demonstrate KSR with neural XC functionals from the first two rungs of Jacob’s ladder [326], by constraining the receptive field of the convolutional neural network [1]. The local density approximation (LDA) has a receptive field of just the current point, while the generalized gradient approximation (GGA) includes the nearest neighbor points, the minimal information for computing the spatial gradient of the density. In (b-c), the effect

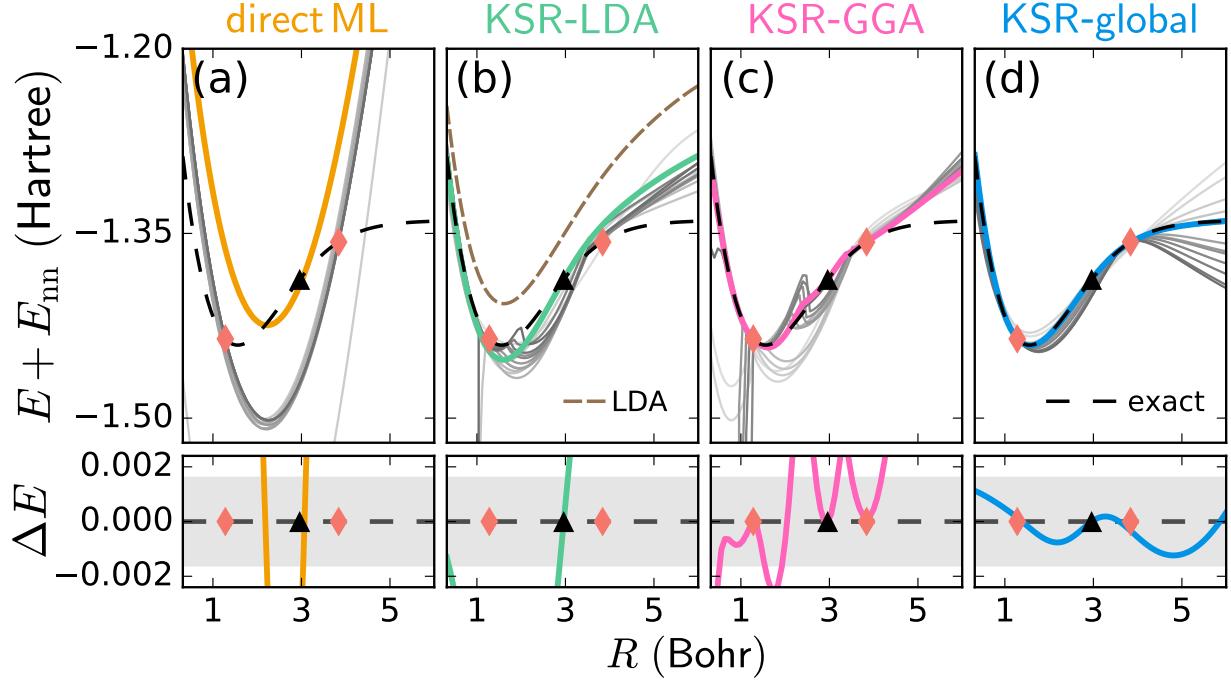


Figure 5.1: One-dimensional H_2 dissociation curves for several ML models trained from two molecules (red diamonds) with optimal models (highlighted in color) selected by the validation molecule at $R = 3$ (black triangles). The top panel shows energy (with E_{nn} , the nucleus-nucleus repulsion energy) with exact values shown by the black dashed line. The bottom panel shows difference from the exact curves with chemical accuracy in grey shadow. (a) directly predicts E from geometries and clearly fails to capture the physics from very limited data. (b-d) shows our method (KSR) with different inputs to the model to align with the first two rungs of Jacob’s ladder [326] (LDA and GGA) and then global (a fully non-local functional). Uniform gas LDA [29] is shown in brown. Grey lines denote 15 sampled functionals during training, with darker lines denoting later samples. Atomic units used throughout.

of the KSR yields reasonably accurate results in the vicinity of the data, but not beyond. The KSR-LDA behaves similar to the uniform gas LDA [29]. When an XC functional with a global receptive field is included in (d), chemical accuracy is achieved for all separations including the dissociation limit. Similar results can be achieved for H_4 , the one-electron self-interaction error can easily be made to vanish, and the interaction of a pair of H_2 molecules can be found without any training on this type of molecule (discussed below).

5.3 Results

Modern DFT finds the ground-state electronic density by solving the Kohn-Sham equations:

$$\left\{ -\frac{\nabla^2}{2} + v_s[n](\mathbf{r}) \right\} \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}). \quad (5.1)$$

The density is obtained from occupied orbitals $n(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2$. Here $v_s[n](\mathbf{r}) = v(\mathbf{r}) + v_H[n](\mathbf{r}) + v_{XC}[n](\mathbf{r})$ is the KS potential consisting of the external one-body potential and the density-dependent Hartree (H) and XC potentials. The XC potential $v_{XC}[n](\mathbf{r}) = \delta E_{XC}/\delta n(\mathbf{r})$ is the functional derivative of the XC energy functional $E_{XC}[n] = \int \epsilon_{XC}[n](\mathbf{r}) n(\mathbf{r}) d\mathbf{r}$, where $\epsilon_{XC}[n](\mathbf{r})$ is the XC energy per electron. The total electronic energy E is then given by the sum of the non-interacting kinetic energy $T_s[n]$, the external one-body potential energy $V[n]$, the Hartree energy $U[n]$, and XC energy $E_{XC}[n]$.

The KS equations are in principle exact given the exact XC functional [214, 431], which in practice is the only term approximated in DFT. From a computational perspective, the eigenvalue problem of Eq. (5.1) is solved repeatedly until the density converges to a fixed point, starting from an initial guess. We use linear density mixing [219] to improve convergence, $n_{k+1}^{(in)} = n_k^{(in)} + \alpha(n_k^{(out)} - n_k^{(in)})$. Figure 5.2(a) shows the unrolled computation flow. We approximate the XC energy per electron using a neural network $\epsilon_{XC,\theta}[n]$, where θ represents the trainable parameters. Together with the self-consistent iterations in Figure 5.2(b), the combined computational graph resembles a recurrent neural network [359] or deep equilibrium model [27] with additional fixed computational components. Density mixing improves convergence of KS self-consistent calculations and parallels the now common residual connections in deep neural networks [168] for efficient backpropagation.

If the neural XC functional were exact, KS self-consistent calculations would output the

exact density and the intermediate energies over iterations would converge to the exact energy. This intention can be translated into a loss function and the neural XC functional can be updated end-to-end by backpropagating through the KS self-consistent calculations. Throughout, experiments are performed in one dimension where accurate quantum solutions could be relatively easily generated via density matrix renormalization group (DMRG) [438]. The electron-electron repulsion is $A \exp(-\kappa|x - x'|)$, and attraction to a nucleus at $x = 0$ is $-A \exp(-\kappa|x|)$ [1]. We design the loss function as an expectation \mathbb{E} over training molecules,

$$L(\theta) = \underbrace{\mathbb{E}_{\text{train}} \left[\int dx (n_{\text{KS}} - n_{\text{DMRG}})^2 / N_e \right]}_{\text{density loss } L_n} + \underbrace{\mathbb{E}_{\text{train}} \left[\sum_{k=1}^K w_k (E_k - E_{\text{DMRG}})^2 / N_e \right]}_{\text{energy loss } L_E}, \quad (5.2)$$

where N_e is the number of electrons and w_k are non-negative weights. L_n minimizes the difference between the final density with the exact density. The gradient from L_n backpropagates through $v_{\text{XC},\theta}[n]$ in all KS iterations. However, if L_E only optimizes the final energy, no gradient flows through $E_{\text{XC},\theta}[n]$ except for the final iteration. To make backpropagation more efficient for $E_{\text{XC},\theta}[n]$, L_E optimizes the trajectory of energies over all iterations, which directly flows gradients to early iterations [18]. This makes the neural XC functional output accurate ϵ_{XC} at each iteration, and also drives the iterations to quickly converge to the exact energy. The optimal model is selected with minimal mean absolute energy per electron on the validation set.

Hundreds of useful XC functional approximations have been proposed [266]. Researchers typically design the symbolic form from physics intuition, with some (or no) fitting parameters. Here we build a neural XC functional with several differentiable components with physics intuition tailored for XC in Figure 5.2(c). A global convolution layer captures the long range interaction, $G(n(x), \xi_p) = \frac{1}{2\xi_p} \int dx' n(x') \exp(-|x - x'|/\xi_p)$. Note two special cases retrieve

known physics quantities, Hartree energy density $G(n(x), \kappa^{-1}) \propto \epsilon_H$ and electronic density $G(n(x), 0) = n(x)$. Global convolution contains multiple channels and ξ_p of each channel is trainable to capture interaction in different scales. Although the rectified linear unit [291] is popular, we use the sigmoid linear unit (SiLU) [106, 349] $f(x) = x/(1 + \exp(-x))$ because the infinite differentiability of SiLU guarantees the smoothness of v_{xc} , the first derivative, and the second and higher order derivatives of the neural network used in the L-BFGS training [258]. We do not enforce a specific choice of ϵ_{xc} (sometimes called a gauge [325]), but we do enforce some conditions, primarily to aid convergence of the algorithm. We require ϵ_{xc} to vanish whenever the density does, and that it be negative if at all possible. We achieved the former using the linearity of SiLU near the origin and turning off the bias terms in convolution layers. We softly impose the latter by a negative transform layer at the end, where a negative SiLU makes most output values negative. Finally, we design a self-interaction gate (SIG) that mixes in a portion of $-\epsilon_H$ to cancel the self-interaction error, $\epsilon_{xc}^{(out)} = \epsilon_{xc}^{(in)}(1 - \beta) - \epsilon_H \beta$. The portion is a gate function $\beta(N_e) = \exp(-(N_e - 1)^2/\sigma^2)$. When $N_e = 1$, then $\epsilon_{xc}^{(out)} = -\epsilon_H$. For more electrons, σ can be fixed or adjusted by the training algorithm to decide the sensitivity to N_e . For H_2 as $R \rightarrow \infty$, ϵ_{xc} tends to a superposition of the negative of the Hartree energy density at each nucleus and approaches half that for H_2^+ .

Now we dive deeper into the outstanding generalization we observed in a simple but not easy task: predicting the entire H_2 dissociation curve, as shown in Figure 5.1. It is not surprising that direct ML model completely fails. Neural networks are usually underdetermined systems as there are more parameters than training examples. Regularization is crucial to improve generalization [143, 224], especially when data is limited. Most existing works regularize models with particular physics prior knowledge by imposing *constraints* via feature engineering and preprocessing [92, 175], architecture design [403, 368, 215, 373] or physics-informed loss terms [348, 376]. Another strategy is to generate extra data for training using prior knowledge: in image classification problems, data are augmented by operations like flipping and cropping given the prior knowledge that labels are invariant to those opera-

tions [222]. KSR provides a natural data augmentation because although the exact densities and energies of only two separations are given, KSR samples different trajectories from an initial density to the exact density at each training step. More importantly, KSR focuses on learning an XC functional that can lead the KS self-consistent calculations to converge to the exact density from the initial density. Figure 5.3 visualizes the density trajectories sampled by KSR for one training separation $R = 3.84$. The functional with untrained parameters ($t = 0$) samples densities near the initial guess but soon learns to explore broadly and finds the trajectories toward the vicinity of the exact density.

In contrast, most existing ML functionals learn to predict the output of a single iteration from the exact density, which is a poor surrogate for the full self-consistent calculations [415]. These standard ML models have two major shortcomings. First, the exact density is unknown for new systems, so the model is not expected to behave correctly on unseen initial densities for KS calculations. Second, even if a model is trained on many densities for single iteration prediction, it is not guaranteed to converge the self-consistent calculations to a good solution [356]. On the other hand, since KSR allows the model access to all the KS iterations, it learns to optimize the entire self-consistent procedure to avoid the error accumulation from greedy optimization of single iterations. Further comparison for training neural XC functionals without or with “weaker” KSR is in the supplemental material [1].

Next we retrain our neural XC functional with KSR on $N_{\text{train}}/2$ examples each of H_2 and H_4 molecules. Figure 5.4 shows the prediction accuracy of KSR with both energy and density loss (full KSR), in comparison to KSR with only energy loss (energy only KSR) and direct ML model. We compute the energy mean absolute error on the holdout sets of H_2 ($R \in [0.4, 6]$) and H_4 ($R \in [1.04, 6]$). The average mean absolute error of H_2 and H_4 with various N_{train} is shown in Figure 5.4(a). Full KSR has the lowest error at minimum $N_{\text{train}} = 4$, reaching chemical accuracy at 6. As the size of the training set increases, energy only KSR reaches chemical accuracy at $N_{\text{train}} = 10$, but direct ML model never does (even

at 20). Then we test models on unseen types of molecules. In Figure 5.4(b), both KSR models have perfect prediction on H_2^+ ($R \in [0.64, 8.48]$) because of the self-interaction gate in the neural XC functionals, while direct ML models always have large errors. Finally we take a pair of equilibrium H_2 and separate them with $R = 0.16$ to 9.76 Bohr, denoted as H_2H_2 . KSR models generalize much better than ML for “zero-shot” prediction [281], where H_2H_2 has never been exposed to the model during training.

5.4 Conclusion

Why is the density important in training, and what use are the non-converged iterations? The density is the functional derivative of the energy with respect to the potential, so it gives the exact slope of the energy with respect to any change in the potential, including stretching (or compressing) the bond. Thus the density implicitly contains energetic information including the correct derivative at that point in the binding curve. KS iterations produce information about the functional in the vicinity of the minimum. During training, the network learns to construct a functional with both the correct minimum and all correct derivatives at this minimum. In the paradigm of differentiable programming, density is the hidden state carrying the information through the recurrent structure in Figure 5.2(a). Correct supervision from L_n greatly helps generalization from very limited data, see $N_{\text{train}} \leq 6$ in Figure 5.4. But as N_{train} increases, both KSR with and without L_n perform well in energy prediction. We show the solution of H_4 with $R = 4.32$ in Figure 5.5. With L_n , the density is clearly much accurate than KSR without L_n ($\int(n_{\text{KS}} - n_{\text{DMRG}})^2 dx = 9.2 \times 10^{-5}$ versus 9.8×10^{-2}). Then we compute the corresponding exact v_s using inverse KS method [187]. Both functionals do not reproduce the exact v_s . However, functional trained with L_n recovered most of the KS potential. Unlike previous works [365, 474, 288] that explicitly included the KS or XC potential into the loss function, our model never uses the exact KS potential.

In our KSR setup, the model aims at predicting ϵ_{xc} , from which the derived v_s yields accurate density. Therefore, predicting v_{xc} is a side product. We also address some concerns on training explicitly with v_{xc} . One artifact is that generating the exact v_s requires an additional inverse calculation, which is known to be numerically unstable [187]. [365] observe outliers while generating training v_{xc} from inverse KS. While v_{xc} is a fascinating and useful object for theoretical study, because its relation to the density is extremely delicate, it is far more practical to simply use the density to train on [289].

Differentiable programming blurs the boundary between physics computation and ML. Our results for KS-DFT serve as proof of principle for rethinking computational physics in this new paradigm. Although there is no explicit limitation of our algorithm to one dimension, we expect practical challenges with real molecules, which will require rewriting or extending a mature DFT code to support automatic differentiation. For example, our differentiable eigensolver for dense matrices [336] is not suitable for large problems, and will need to be replaced with methods for partial eigendecomposition of sparse matrices [235, 448]. Beyond density functionals, in principle all heuristics in DFT calculations, e.g., initial guess, density update, preconditioning, basis sets, even the entire self-consistent calculations as a meta-optimization problem [18], could be learned and optimized while maintaining rigorous physics – getting the best of both worlds.

5.5 Acknowledgments

The authors thank Michael Brenner, Sam Schoenholz, Lucas Wagner, and Hanjun Dai for helpful discussion. K.B. supported by NSF CHE-1856165, R.P. by DOE DE-SC0008696. Code is available at https://github.com/google-research/google-research/tree/master/jax_dft

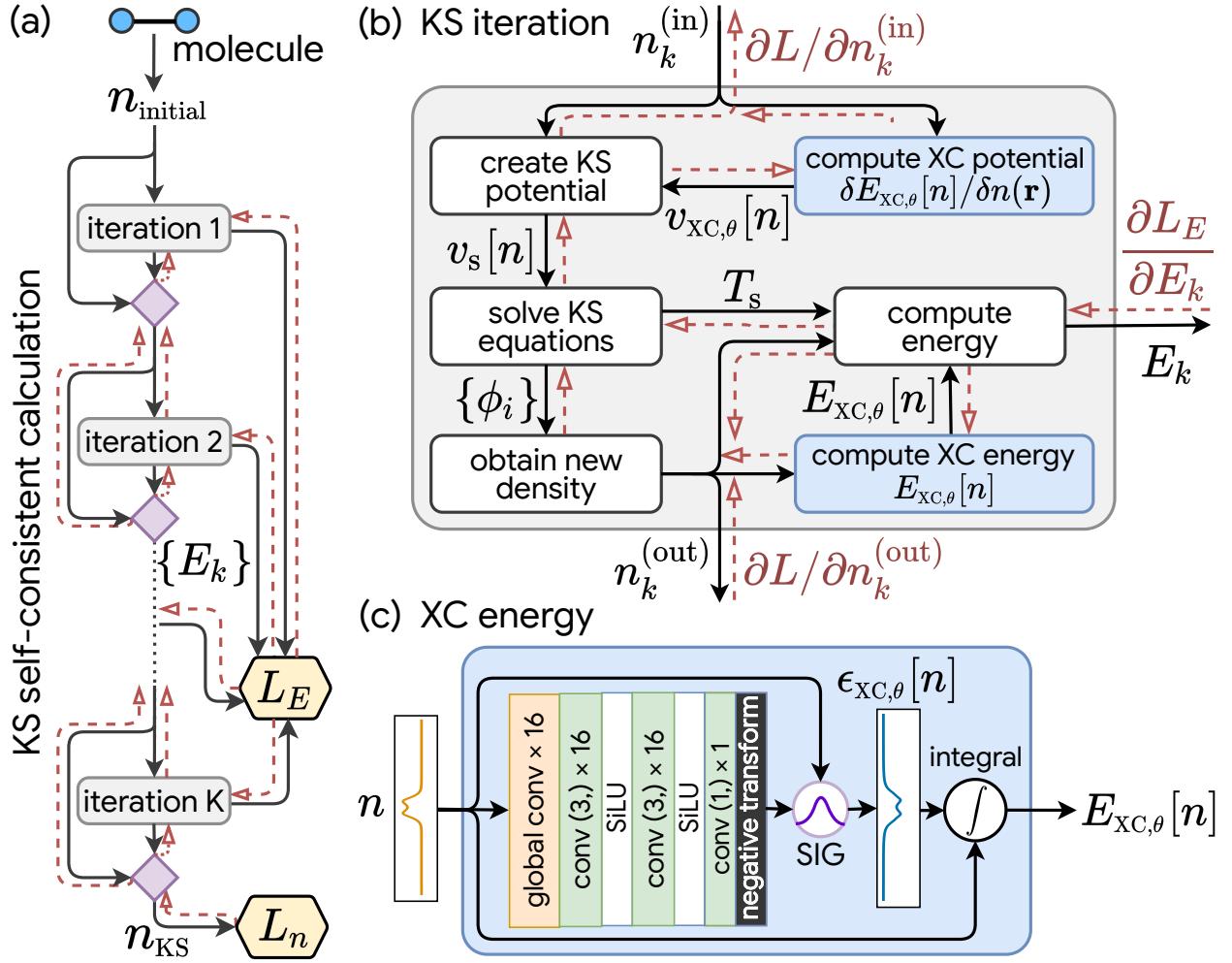


Figure 5.2: KS-DFT as a differentiable program. Black arrows are the conventional computation flow. The gradients flow along red dashed arrows to minimize the energy loss L_E and density loss L_n . (a) The high-level KS self-consistent calculations with linear density mixing (purple diamonds). (b) A single KS iteration produces $v_{\text{XC},\theta}[n]$ and $E_{\text{XC},\theta}[n]$ by invoking the XC energy calculation twice, once directly and once calculating a derivative using automatic differentiation. (c) The XC energy calculation using the global XC functional.

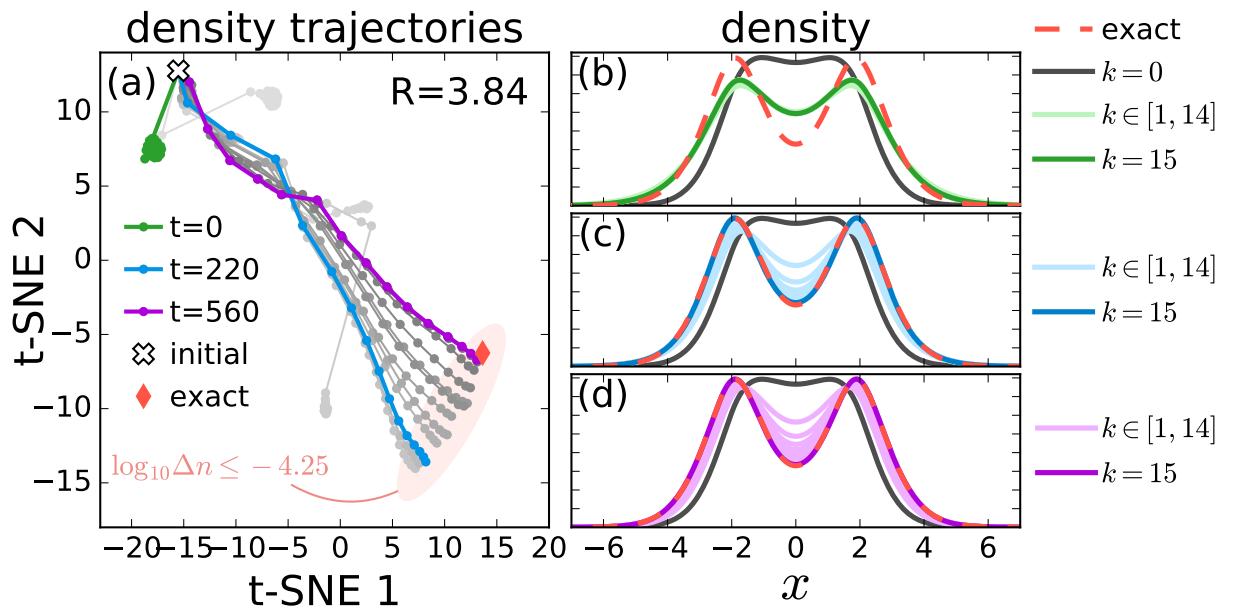


Figure 5.3: (a) t-SNE visualization [264] of density trajectories (grey dots) sampled by KSR during training for $R = 3.84$ from initial guess (cross) to exact density (red diamond). Darker trajectories denote later optimization steps t . Note t-SNE projection does not perfectly preserve the distance between densities. The light red ellipse illustrates the vicinity of the exact density within $\log_{10}(\int dx (n_{\text{KS}} - n_{\text{DMRG}})^2 / N_e) \leq -4.25$. Densities from each KS iteration in trajectories are plotted in the corresponding highlighted colors for (b) $t = 0$ untrained, (c) $t = 220$ optimal in Figure 5.1, and (d) $t = 560$ overfitting to training with bad generalization on validation.

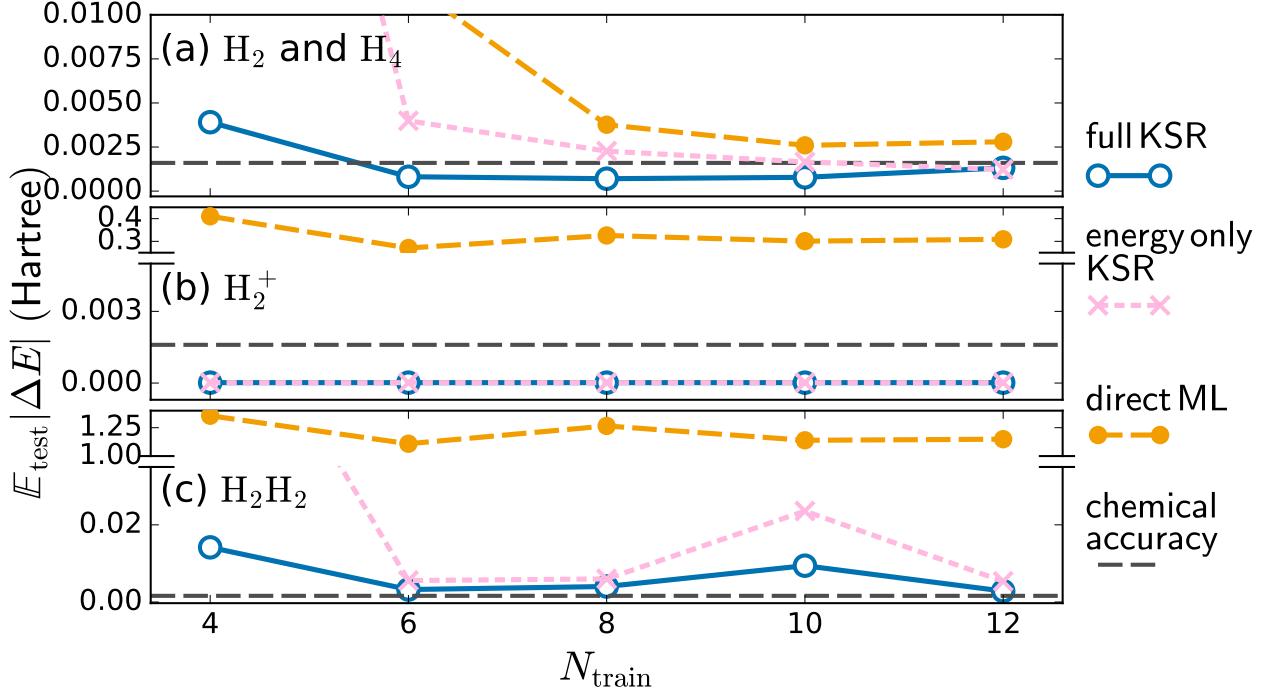


Figure 5.4: Test generalization of models as a function of the total number of training examples N_{train} : full KSR (blue), energy only KSR (pink) and direct ML (orange) on (a) holdout H_2 and H_4 , and unseen types of molecules (b) H_2^+ (c) H_2H_2 . Black dashed lines show chemical accuracy. See the supplemental material [1] for training details.

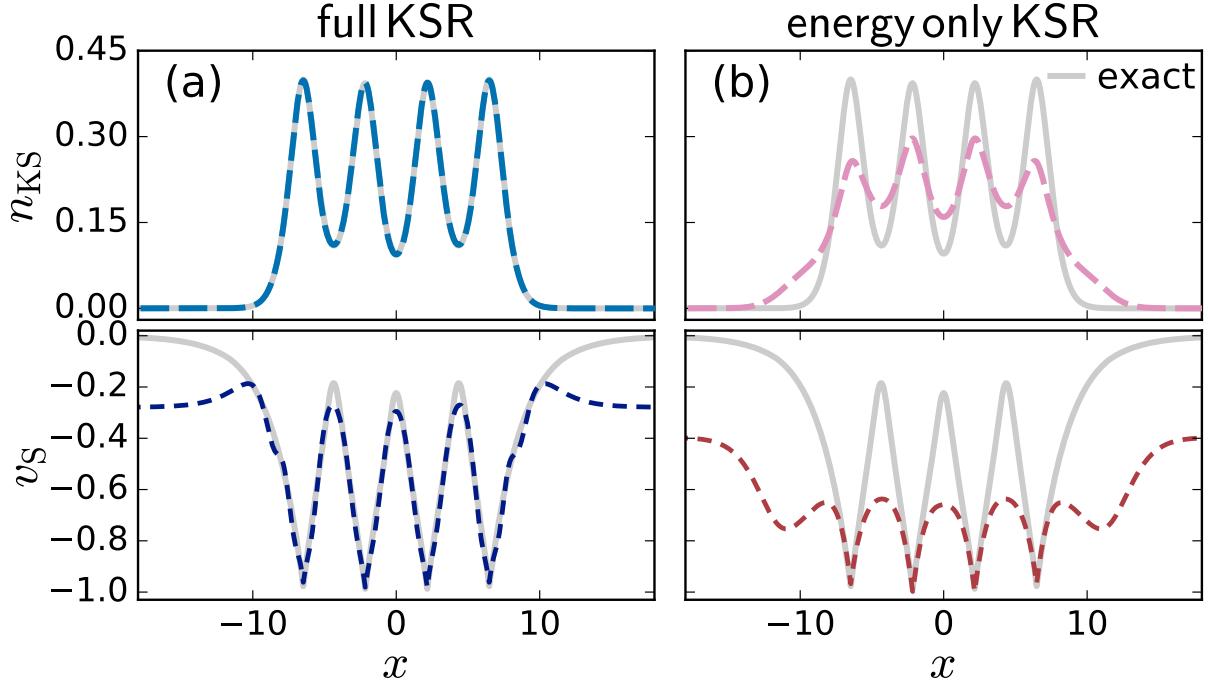


Figure 5.5: Density and KS potential of H_4 with $R = 4.32$ from neural XC functionals trained with (a) full KSR (blue) and (b) energy only KSR (pink) on training set of size $N_{\text{train}} = 20$. Exact curves are in grey. v_S are shifted by a constant for better comparison.

Chapter 6

How Well Does Kohn-Sham Regularizer Work for Weakly Correlated Systems?

This chapter is a reproduction of Ref. [195], which I co-authored with Bhupalee Kalita, Jielun Chen, Li Li, and Kieron Burke.

6.1 Abstract

Kohn-Sham regularizer (KSR) is a differentiable machine learning approach to finding the exchange-correlation functional in Kohn-Sham density functional theory (DFT) that works for strongly correlated systems. Here we test KSR for weak correlation. We propose spin-adapted KSR (sKSR) with trainable local, semilocal, and nonlocal approximations found by minimizing density and total energy loss. We assess the atoms-to-molecules generalizability by training on one-dimensional (1D) H, He, Li, Be, Be^{++} and testing on 1D hydrogen chains,

LiH, BeH₂, and helium hydride complexes. The generalization error from our semilocal approximation is comparable to other differentiable approaches, but our nonlocal functional outperforms any existing machine learning functionals, predicting ground-state energies of test systems with a mean absolute error of 2.7 milli-Hartrees.

6.2 Introduction

Determining the ground-state properties of many-electron systems is fundamental to molecular modeling problems in chemical and material sciences. However, solving the Schrödinger equation explicitly for more than a few hundred electrons is computationally intractable. Among several methods of approximation, Kohn-Sham density functional theory (KS-DFT or simply DFT) [174, 214], a method based on the electron density distribution rather than the many-electron wave function, provides chemically useful results with $O(N^3)$ scaling for an N -electron system [103]. DFT is formally exact, but the exchange-correlation (XC) energy, resulting from the quantum-mechanical interaction between electrons, must be approximated in practice. Hundreds of XC energy functional approximations have been formulated in the past few decades [266]. Functionals can be designed non-empirically, for example using physics and chemical-based intuition and satisfying known exact constraints [313], or can involve some fitting to reference data [468]. However, in any approach, these functional approximations do not yield chemical accuracy in general, that is, with errors less than 1.6 milli-Hartrees (mH) in atomic units (or 1 kcal/mol). Improving the accuracy of XC functional approximations often incurs additional computational cost in the practical DFT calculation [60]. However, there is no systematic way in general to develop and improve XC functional approximations.

In recent years, machine learning (ML) has been used to find better DFT approximations. Attempts have been made to enhance either the speed or accuracy of DFT. Some used ML

techniques to boost computational efficiency by approximating the non-interacting kinetic energy without solving the KS equations [379, 253, 59, 194]. In an effort to improve the accuracy of ML-DFT, a significant leap was achieved by Nagai et al. [289], who used a neural network (NN) model to approximate the XC functional and trained it with high accuracy coupled cluster (CCSD(T)) energies and densities of just three small molecules, while self-consistently solving the KS equations. This functional impressively generalized to 148 small molecules [93] to predict their energies and densities with accuracies comparable to human-designed functionals. However, the test set atomization energies were not chemically accurate. Also, they didn't have access to gradient information and were therefore limited to a gradient-free optimization scheme, which is inherently slow, often suffers poor convergence issues, and is difficult to scale to more complex NN models.

In DFT, many useful properties are extracted from the density, although an XC functional approximation need not produce accurate densities along with accurate energies [209]. In KS-DFT, we calculate the density self-consistently, and there is a nonlinear dependence of the XC functional on the density. Learning this relationship requires not only the ground truth mapping of the functional inputs to outputs but also how the functional performs in the underlying process. Hence the use of differentiable programming [34] becomes more intuitive [182]. With differentiable programming, conditioning the networks with physical insights becomes much simpler, and it can further help to ease the process of training.

Recently, Li et al. [252] made a valuable step in this direction by considering the entire DFT self-consistent calculation as a differentiable program. They implemented an end-to-end differentiable DFT code for 1-dimensional (1D) systems using JAX [58], a library that provides differentiation, vectorization, just-in-time compilation, and other composable transformations of Python and NumPy programs [163]. They parameterized the XC functional with an NN which incorporated non-local information about the density, along with known physical constraints. The self-consistent KS calculations were embedded into the training process by

backpropagating the gradients through the KS iterations. It was dubbed the Kohn-Sham regularizer (KSR). It could yield chemically accurate energies for uniformly separated 1D hydrogen chains at any separation by training on highly accurate energies and densities from only a few separations.

Following a similar approach, Kasim and Vinko [200] implemented an end-to-end differentiable DFT code in 3D for Gaussian-type orbitals and trained local and semi-local NN-based XC functional approximations, evaluating performance on small molecules. In another work, Dick et al. [100] constructed a semilocal XC functional that was carefully curated to account for several known exact conditions and pretrained to match SCAN, a popular meta-GGA functional [388]. While both of these works explore the generalizability of ML approximations for weakly correlated molecules with differentiable DFT codes, they do not incorporate global information, and their accuracy is limited to that of human-designed semilocal functionals. A slightly different approach involves introducing an ML correction term to a nonempirical or semi-empirical XC functional within a KS-DFT self-consistent framework [98, 72]. In such an approach, only a portion of the XC energy is approximated using ML and the functionals retain the characteristics of the baseline XC functional used. The recently proposed ML local hybrid functional, DM21 [211], addresses spin-symmetry breaking and delocalization error in DFT functionals. Consequently, it performs well on several main-group benchmark datasets and also correctly dissociates molecules. Unlike KSR, this functional is trained on large datasets of highly accurate reaction energies (not densities) in the loss function without explicitly supervising the self-consistent iterations.

[C3] Li et al. [252] explored the generalizability of KSR for a few strongly correlated systems with stretched bonds which is a completely different domain from most chemical applications of DFT. The aim there was to generate accurate binding energy curves (all the way to the dissociation limit) using the entire density (for the nonlocal approximation called global-KSR), using inputs at only two separations, for unpolarized hydrogen chains. The generalizability

was in finding the entire bond-dissociation energy curve of these chains. Moreover, only the total density was used and not the spin densities.

In the present work, we propose spin-polarized versions of local, semilocal, and nonlocal XC functional approximations within a differentiable spin-DFT implementation of KSR. We modify these approximations to predict XC energy densities using spin-densities as feature vectors while optimizing the NN parameters using total density and energy loss. Contrary to Ref. [252], we test the KSR approach in the domain of routine DFT calculations in chemistry, namely in and around equilibrium bond lengths. We find the remarkable result that training on energies and densities of a few atoms (and ions) alone produces accurate ground-state energies for equilibrium molecules (very reminiscent of the use of appropriate norms while avoiding using any covalent bond energies). We train and test on a variety of different elements, to obtain the generalizability relevant to chemistry. Almost all previous work in the chemical domain tests various approximate functional forms employing the standard ingredients locally [289, 100, 200]. Our work achieves high accuracy using the total density and is not limited to a specific set of human-chosen features.

6.3 Results

The practical implementation of DFT involves solving the Kohn-Sham (KS) equations to calculate the ground-state electron density,

$$\left\{ -\frac{1}{2} \nabla^2 + v_s[n](\mathbf{r}) \right\} \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}). \quad (6.1)$$

The electron density, $n(\mathbf{r})$, is the sum of the probability density over all occupied one-electron KS orbitals, $n(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2$. The KS potential, $v_s[n](\mathbf{r})$, contains the external one-body

potential, the Hartree potential, and the XC potentials,

$$v_s[n](\mathbf{r}) = v(\mathbf{r}) + v_h[n](\mathbf{r}) + v_{xc}[n](\mathbf{r}). \quad (6.2)$$

The XC potential is the functional derivative of the XC energy, $E_{xc}[n]$, with respect to the electron density [214], $v_{xc}[n](\mathbf{r}) = \delta E_{xc}[n](\mathbf{r})/\delta n(\mathbf{r})$. We can express $E_{xc}[n]$ in terms of an XC energy density per electron, $\epsilon_{xc}[n](\mathbf{r})$:

$$E_{xc}[n] = \int d^3r \epsilon_{xc}[n](\mathbf{r}) n(\mathbf{r}). \quad (6.3)$$

The ground-state energy is calculated from the self-consistent density by summing the non-interacting kinetic energy, T_s , the external potential energy, V , the Hartree energy, U , and the XC energy,

$$E_0 = T_s[n] + V[n] + U[n] + E_{xc}[n]. \quad (6.4)$$

The computational efficiency is also affected by the level of approximation used for the XC functional [326].

Density matrix renormalization group (DMRG) [438] can be used to efficiently generate highly accurate benchmark energies and densities for these 1D analog systems. We can address such systems using 1D KS-DFT calculations as well with suitable XC energy functional approximations, such as the 1D local spin-density approximation (LSDA) which was constructed in Ref. [30] from the 1D exponentially repelling uniform electron gas.

In essence, KSR is a ML-DFT regularization technique that utilizes a differentiable analog of the standard self-consistent DFT computational flow during training to train a suitable parameterized model for $E_{xc}[n] = E_{xc,\theta}[n]$, where θ are trainable parameters [252]. In this work, we consider NN-based (neural) XC models, but KSR as a regularization technique can

apply more broadly to any differentiable model choice. Knowledge of physical properties and constraints in the exact XC functional can help guide the construction of a neural XC approximation. The NN that parameterizes the XC functional in KSR is carefully curated to account for a few of the expected behaviors of the exact XC functional. Nonlocality is facilitated by adding a global convolution layer in $\epsilon_{\text{xc},\theta}[n]$ to help capture long-range interactions. The sigmoid linear unit (SiLU or Swish) [107, 350] activation function is used throughout because of its infinite differentiability. The KSR network is also complemented with a self-interaction gate (SIG) that partially cancels the self-interaction error by mixing in a portion of Hartree energy density to ϵ_{xc} .

In Ref. 252 several neural XC functional models were proposed: a local functional which only depends on the density at each point (KSR-LDA), a semi-local functional that uses local and gradient information about each point (KSR-GGA), and a global functional which utilized the global convolution layer and the SIG described above (KSR-global).

A main deficiency of the KSR technique in Ref. 252 is that it does not explicitly account for spin, and so may not generalize well for spin-polarized systems. Extending this technique and associated NN models to spin DFT requires a differentiable framework that can backpropagate through resulting spin densities. Spin is often incorporated in the neural XC functional using relative polarization, ζ , as a feature [289]. For up and down spin densities, $\{n_{\uparrow}, n_{\downarrow}\}$, $\zeta = (n_{\uparrow} - n_{\downarrow})/n$. While ζ can be introduced as an additional input channel to KSR neural ϵ_{xc} , its scale can be very different relative to n in general. Instead, we use up and down spin densities as input features, which have similar scales. The usual models and concepts for KSR can be extended to obtain a spin-adapted KSR (sKSR).

In sKSR-global, we have a global convolution layer that takes spin densities as inputs, and the kernel takes the form:

$$G(n_{\sigma}(x), \xi_p) = \frac{1}{2}\xi_p \int dx' n_{\sigma}(x') e^{-|x-x'|/\xi_p}, \quad (6.5)$$

where $\sigma \in \{\uparrow, \downarrow\}$ and ξ_p is a trainable parameter that represents an interaction scale. To keep the number of parameters comparable with KSR-global, we input each spin density to a global convolution layer consisting of 8 channels. We then concatenate the output on the channel dimension and input it to the latter convolution layers. For weakly correlated systems and greater generalizability, this approximation does not include any SIG. The rest of the network architecture is kept unchanged. sKSR-LDA and sKSR-GGA approximations to XC are devoid of global information. For sKSR-LDA, two convolution layers with filter size one and 8 channels map the spin-density to ϵ_{xc} at the same spatial point x . In sKSR-GGA, we specify the total density gradient explicitly as an additional input channel along with the spin-densities. Instead of using one convolution layer with filter size three, we use three convolution layers with filter size one and 8 channels each. The rest of the sKSR-LDA and sKSR-GGA architectures are also similar to KSR-LDA and KSR-GGA. Fig. 6.1(a) shows the comparative network structures for all three types of approximations. In all cases, the resulting ϵ_{xc} is symmetrized with respect to the input of the up and down densities:

$$\epsilon_{\text{xc}}^{\text{symm}}[n_{\uparrow}, n_{\downarrow}] = \frac{1}{2} \left[\epsilon_{\text{xc}}[n_{\uparrow}, n_{\downarrow}] + \epsilon_{\text{xc}}[n_{\downarrow}, n_{\uparrow}] \right]. \quad (6.6)$$

Our approximation replaces the ϵ_{xc} in a spin-polarized self-consistent KS-DFT framework. For spin-polarized systems we perform the above spin-unrestricted KS-DFT procedure, however for unpolarized systems we use spin-restricted KS-DFT to preserve spin-symmetry. Fig. 6.1(b) shows the conventional computational flow and the flow of the gradients during the self-consistent optimization. To train the neural XC functional, we use the following loss

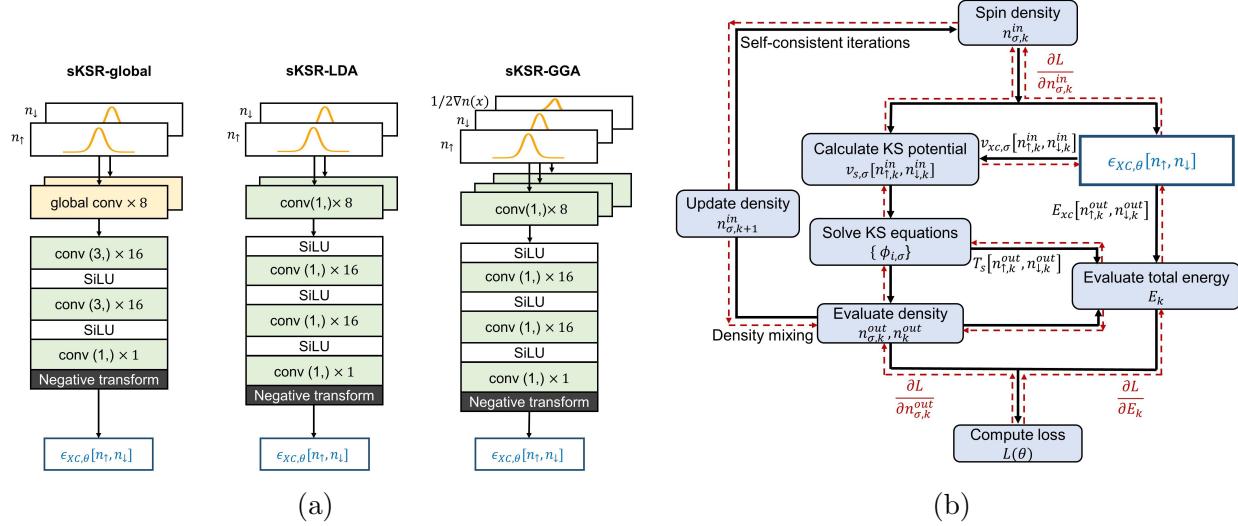


Figure 6.1: (a) sKSR-global, sKSR-LDA and sKSR-GGA architectures to calculate ϵ_{XC} from spin-densities. (b) sKSR – differentiable KS-DFT with spin-polarization. Black arrows refer to the conventional computational flow. The gradients flow along red-dashed arrows to minimize the loss during training.

function:

$$L(\theta) = \underbrace{\mathbb{E}_{\text{train}} [(E^{\text{sKSR}} - E^{\text{DMRG}})^2 / N_e]}_{\text{energy loss } L_E} + \underbrace{\mathbb{E}_{\text{train}} \left[\int dx (n^{\text{sKSR}} - n^{\text{DMRG}})^2 / N_e \right]}_{\text{density loss } L_n}, \quad (6.7)$$

where E^{sKSR} and n^{sKSR} are the converged total energy and total density obtained from the neural XC functional approximations, and E^{DMRG} and n^{DMRG} are the exact ground-state electronic energy and total density for each of the test systems. The total loss is evaluated as an expectation over training examples, where N_e is the number of electrons for a given training example. All quantities are in atomic units. We only consider the converged energy in the energy loss term rather than the energy trajectory throughout KS iterations, which was explored in Ref. 252. In this work we find that the self-consistent calculations converge quickly for the small atoms and ions used in training, and incorporating energy

loss from each KS iteration minimally affects the efficiency of the optimization process. The gradients are calculated based on the total loss with respect to the parameters, θ , through automatic differentiation. They are back-propagated across the self-consistent cycles and the parameters of the neural XC functionals are updated until the total loss is minimized.

As a simple consistency test, we pose the question: can KSR learn human-designed functionals from their observable results? Here we specifically investigate whether sKSR-LDA can learn the relatively simple but general human-designed 1D LSDA XC functional. Since our sKSR-LDA model utilizes hundreds of parameters, it is unclear whether training on just a few LSDA generated DFT results will yield a neural XC model that matches LSDA. We find that by training sKSR-LDA on LSDA-generated He and Li^{++} , we recover the LSDA XC functional almost exactly for unpolarized and fully polarized systems, see Figure 6.2. The sKSR-LDA model deviates at the high-density limit (low r_s limit) due to the limitation that our training densities only consist of $r_s > 0.5$.

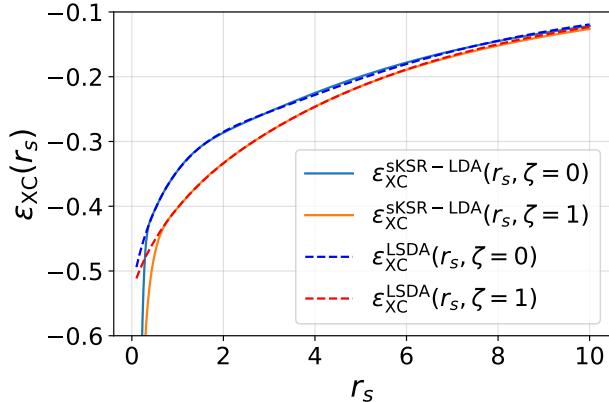


Figure 6.2: sKSR-LDA trained on 1D LSDA-calculated Li^{++} and He energies and densities. Here $r_s = 1/2n$ and $\epsilon_{\text{XC}}^{\text{unif}}$ corresponds to the XC energy density of the 1D uniform electron gas [30].

Next, we assess generalizability by training sKSR models using a few 1D atomic systems and testing on unseen 1D molecular systems. We trained all three models on DMRG energies and densities of H, He, Li, Be, and Be^{++} and validated on Be^+ . For training and validation

Table 6.1: Training, validation and test sets for generalizability experiment. The molecules in the test set refer to the relaxed structures.

Training		Validation		Testing	
H, He, Li	Be ⁺	H ₂ , H ₃ , H ₄ , H ₂ ⁺ , H ₃ ⁺			
Be, Be ⁺⁺		LiH, BeH ₂ , HeH ⁺			
		H-He-He-H ²⁺			
		He-H-H-He ²⁺			

details, see Supporting Information. The trained model was later used to calculate the properties of several molecules in their equilibrium ground-state or relaxed form (see Table 6.1). The errors in total energies, ionization, and atomization energies, as well as the average density losses for all three neural XC functional approximations, are reported in Table 6.2. Compared to LSDA, the mean absolute error (MAE) in sKSR-LDA calculated energies is reduced by a factor of three. On the other hand, sKSR-global is an order of magnitude higher in accuracy and yields total energies with an MAE of 2.7 mH, not so far from the chemical accuracy limit of 1.6 mH. The cumulative MAEs for the training, validation and test datasets are reported in Supporting Information.

Table 6.2: Total energy errors (in mH), density losses (in 10^{-4} Bohr $^{-1}$), and errors in ionization potentials for atoms and atomization energies in molecules (in mH) calculated using uniform gas LSDA [30], sKSR-LDA, sKSR-GGA, and sKSR-global respectively, for the training, validation, and test sets in Table 6.1.

Dataset	Symbol	LSDA			sKSR-LDA			sKSR-GGA			sKSR-global		
		ΔE	L_n	ΔIP									
Training	H	26.6	5.35	-26.6	4.51	0.55	-4.50	4.49	0.31	-4.49	0.85	0.33	-0.85
	He	41.4	2.89	-8.46	20.2	0.63	-21.3	7.49	0.24	-10.2	-0.69	0.03	0.62
	Li	33.7	5.02	16.6	-11.5	0.40	37.4	-12.0	1.37	20.2	-2.37	0.12	2.79
	Be	24.5	1.18	21.4	-23.5	1.03	12.1	-2.70	0.65	-5.29	1.16	0.07	-1.23
	Be ⁺⁺	55.3	0.75	-18.1	29.2	0.16	-46.1	6.55	0.49	-34.1	0.41	0.02	-1.43
MAE		36.3	3.04	18.3	17.8	0.56	24.3	6.65	0.16	14.8	1.10	0.12	1.38
Validation	Be ⁺	46.0	1.95	9.37	-11.3	0.12	40.5	-7.99	0.61	14.5	-0.07	0.03	0.49
		ΔAE			ΔAE			ΔAE			ΔAE		
Test	H ₂	34.04	1.82	19.2	19.5	0.35	-10.5	6.83	1.99	2.14	-0.73	0.07	2.43
	H ₃	35.6	1.93	44.3	0.45	0.21	13.1	-3.07	5.57	16.5	-3.56	3.22	6.11
	H ₄	32.3	3.82	74.3	7.66	1.59	10.4	-9.34	4.18	27.3	2.87	1.46	0.53
	H ₂ ⁺	19.6	6.68	7.09	2.78	0.71	1.73	1.68	1.71	2.81	-1.94	1.04	2.79
	H ₃ ⁺	31.2	0.78	22.1	20.6	1.87	-11.6	15.4	11.5	-6.44	-0.40	0.47	2.09
	LiH	30.9	3.72	29.5	-8.55	2.47	1.53	-16.6	3.86	9.14	-4.38	0.66	2.86
	BeH ₂	32.8	7.49	45.0	-27.8	5.5	13.4	-34.6	3.09	40.9	-5.07	1.29	7.93
	HeH ⁺	37.3	1.71	4.18	18.8	0.17	1.40	5.18	0.59	2.31	-1.60	0.13	0.91
	H-He-He-H ²⁺	36.7	14.7	46.1	5.00	6.00	35.5	-9.04	2.50	24.0	5.39	4.52	-6.77
	He-H-H-He ²⁺	46.1	7.40	36.7	19.9	6.48	20.6	4.35	4.75	10.6	0.79	5.47	-2.18
	MAE	33.6	5.00	32.9	13.1	2.53	12.0	10.6	3.98	14.2	2.67	1.83	3.46

The importance of spin in sKSR can be seen by comparing results with the original KSR-global model from Ref. [252]. For a valid comparison, we consider KSR-global without the SIG and train it with the sets from Table. 6.1, without adding the energy trajectory loss. The MAE in KSR-global predictions for total energies of the test molecules is 10.02 mH, comparable to sKSR-GGA, but much worse than sKSR-global (see Table. A2 in Supporting Information). sKSR-global also converges more quickly than KSR-global, reaching lower training losses with fewer training steps (see Fig. A7 in Supporting Information).

The size of our dataset is practically limited by the chemical space provided by 1D and the associated exponential interaction. Even though we are dealing with a much smaller dataset, we trained the sKSR models on the ground-state energies and densities of 5 atomic systems only and did not include any molecules, contrary to results in Ref. 289 and Ref. 200 which train on derived quantities, such as atomization and ionization energies, and include molecules in training.

Using sKSR-global, the predicted densities of each molecule have little noticeable error, see Fig. 6.3(a). The corresponding XC potentials are shown in Fig. 6.3(b). For all unpolarized systems, we run restricted KS calculations, and the up and down XC potentials match, while for polarized systems (Li, Be^+ , H_2^+ , and H_3 only) we run unrestricted KS calculations. The sKSR-LDA and sKSR-GGA total densities and XC potentials for the test set are included in the Supporting Information. The comparison to exact XC potentials is not expected to be as precise as potentials are extremely sensitive to densities. However, for each of these examples, we see that the sKSR-global XC potential closely mimics the exact XC potential, even though we did not include XC potentials in the training. Furthermore, seemingly large deviations in the XC potentials can result in similar resulting densities. For example, this can be seen in the case of BeH_2 where the XC potentials are noticeably different but the resulting densities are very similar. The KS potentials are reasonably accurate for the test set (see Supporting Information). Note that similar to the exact XC potentials, the sKSR-global

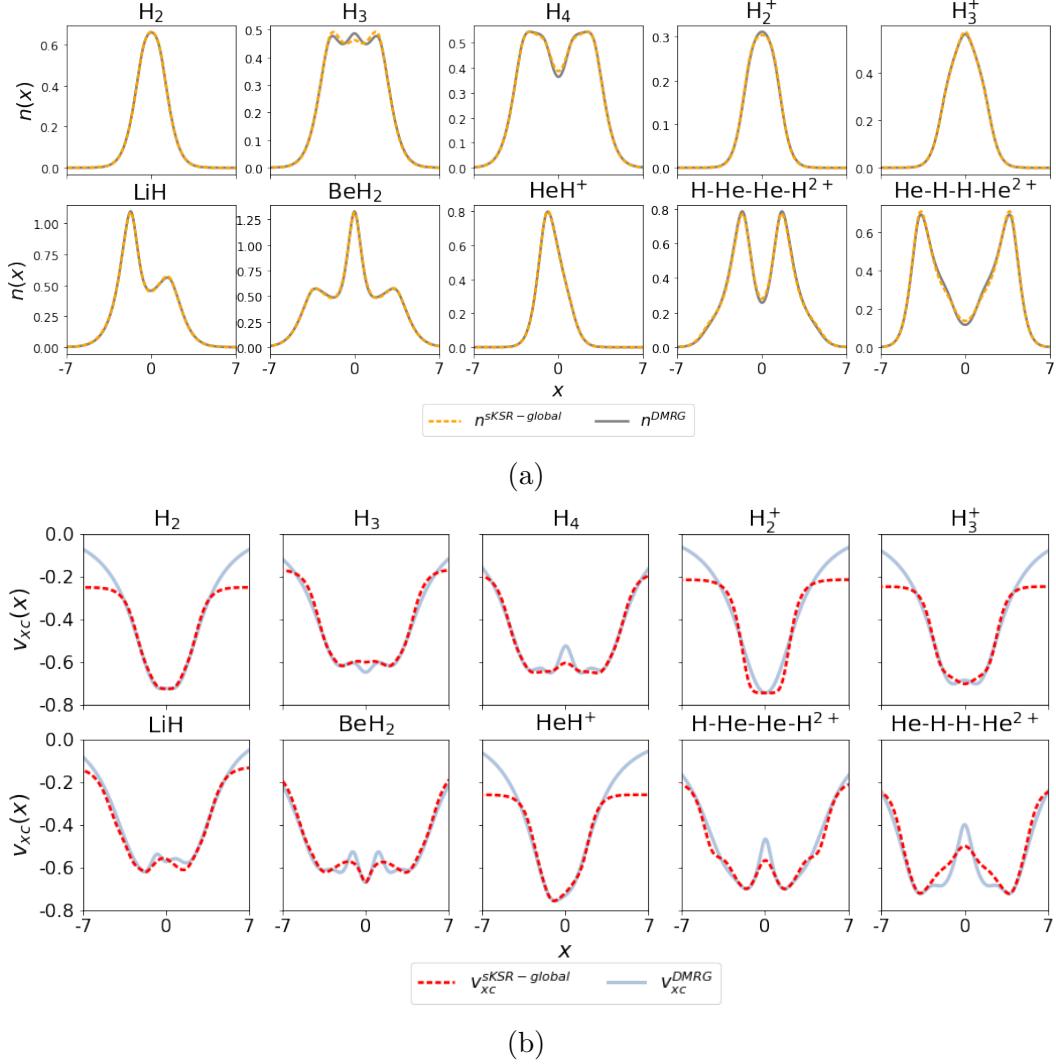


Figure 6.3: (a) The densities obtained using sKSR-global (orange dashes) and the exact ground-state densities (gray), (b) average XC potentials calculated from sKSR-global approximation (red dashes) to ϵ_{xc} and their exact counterparts calculated with DMRG (light blue) for the test molecules in Table. 6.1 at equilibrium separations. The sKSR potentials are shifted by a constant for a better comparison with the exact XC potentials. sKSR-global was trained on H, He, Li, Be, and Be^{++} and validated on Be^+ . Note that, in general, these 1D densities and XC potentials can differ even qualitatively from their 3D analogs.

XC potentials are smooth, due to the use of a smooth activation function.

We can use these potentials to validate the known theoretical properties of the exact XC potentials for different test systems, compare with other XC approximations, and utilize them to introduce corrections to existing local and semilocal approximations. Similarly, sKSR-global can also produce quite accurate spin-densities even though we did not incorporate spin-densities in the loss function while training the XC functionals (see Fig. A1 in Supporting Information).

A very interesting question is: how does our weakly-correlated sKSR behave for strongly-correlated systems? We answer this by studying the paradigm case of the H_2 binding curve in Fig. 6.4, where the sKSR-global curve remains highly accurate up to at least 3 Bohr. Just as with all single-particle methods, the restricted calculation yields energy that is far too high in the dissociated limit. On the other hand, an unrestricted calculation, which breaks spin-symmetry beyond about 4 Bohr, does dissociate correctly, but at the price of poor spin densities and a kink in the binding energy curve. Fig. A6 in Supporting Information shows analogous features for sKSR-LDA and sKSR-GGA, and also shows the accuracy of the total density of the unrestricted solutions at large separations. Fig. 6.4 also shows the result of a KSR-global calculation (i.e., total density only), but trained just on atoms. While it naturally dissociates correctly, it is much less accurate. Of course, the KSR-global of Ref. 252 is chemically accurate for the entire curve because its training included a stretched bond.

In many cases, the predictability of sKSR can extend well beyond the equilibrium bond distance. Fig. 6.5 shows the complete dissociation energy curve of LiH obtained from restricted calculation. Near equilibrium, sKSR-LDA and sKSR-GGA underestimate the binding energy but perform better than LSDA. As the bond is stretched, sKSR-GGA and sKSR-LDA quickly deviate from the expected trajectory. However, sKSR-global performs well throughout, extending its predictive accuracy well beyond the equilibrium bond distance. We show the total density and the XC potential of stretched LiH at 5.92 Bohr in Fig. 6.6. LSDA

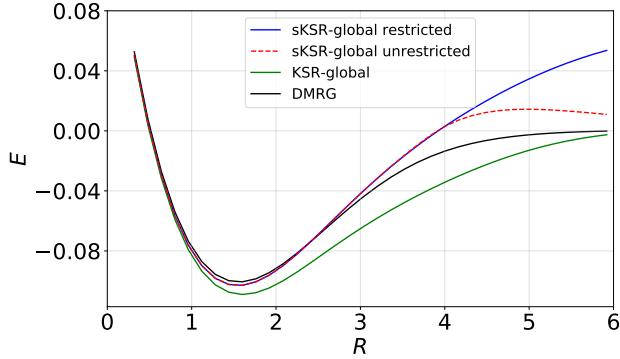


Figure 6.4: The binding energy curve of H_2 molecule calculated based on the total energy prediction for H_2 molecule and the energy of the individual H atoms. sKSR-global was evaluated using restricted KS (blue) and unrestricted KS (red dashes) scheme. The DMRG (black) and KSR-global (green) results are also shown. All the neural approximations, with and without spin, are trained on the dataset given in Table. 6.1.

largely overestimates the total energy of the stretched molecule, but its density remains reasonably accurate. The XC potentials calculated from neural XC functional approximations are comparable, with sKSR-global closely approximating the exact behavior. A comparison of the sKSR-global and the exact total density and XC potential of stretched LiH with respect to the atomic contributions from Lithium and Hydrogen is included in the Supporting Information.

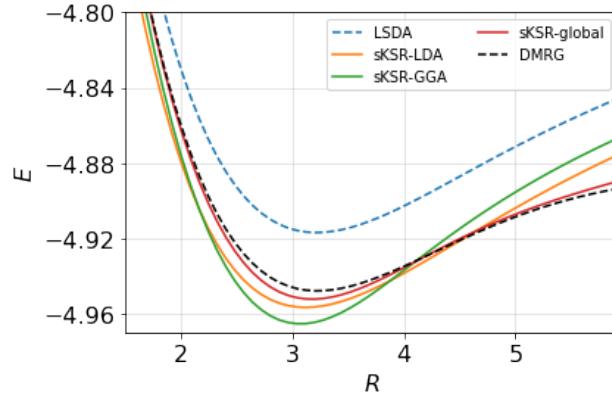


Figure 6.5: The complete dissociation energy curve of LiH molecule generated with sKSR-LDA (orange), sKSR-GGA (green) and sKSR-global(red). The DMRG (black dashes) and the uniform gas LSDA (blue dashes) results are also shown. The neural XC functional approximations were trained and validated on atoms and ions given in Table. 6.1.

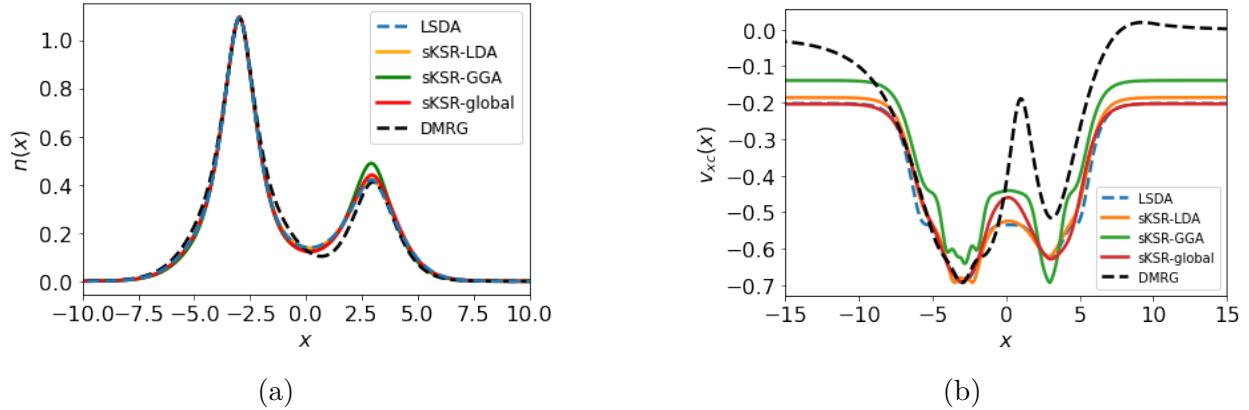


Figure 6.6: (a) The total density and (b) the average XC potentials of LiH at a bond-distance of 5.92 Bohr calculated with the three neural XC functionals as well as uniform-gas LSDA. The exact (DMRG) average XC potentials are included for comparison.

The approximate total energy of a molecule can have two types of error contributions: the error due to the approximate functional and the error arising from the self-consistent density [428]. For most XC functionals, the total density calculated from the self-consistent solution of the KS equations works as an excellent approximation to the exact density for most systems. Hence, the density-driven error is often negligible. However, some approximations can have significant density-driven errors [210]. For our test molecules, the errors in the self-consistent densities were trivial and consequently had minimal impacts on the atomization energy errors. The functional and density-driven errors in our neural XC functional approximations are reported for the hydrogen molecule in the supplementary information section.

6.4 Conclusion

We found that sKSR-global achieves remarkable accuracy and generalization in a very data-efficient manner by including the self-consistent KS equations into the training. sKSR-global predicts the ground-state energy of ten unseen molecules in equilibrium with a mean absolute error of 2.7 mH (~ 1.7 kcal/mol) when trained with just five atoms and ions. Hence, a

nonlocal XC functional approximation trained on atomic energies and densities can generate predictions for weakly-correlated molecules with near chemical accuracy. An extension of this work can lead to an ML functional that is applicable across a broad chemical spectrum without using an exceedingly large training set. The end-to-end differentiable implementation also ensures smooth and reasonable XC potentials. In addition, sKSR-global trained on atoms can adequately describe a molecule with a stretched bond. Combining differentiable programming with inherent physical intuition thus takes us one step closer to a generalizable, chemically accurate ML XC functional.

The application of the current sKSR algorithm is limited to 1D systems and our test set does not include real 3D molecules. However, the methods presented are transferable to 3D and we anticipate that the characteristic performance is not unique to 1D systems, as these systems tend to mimic their 3D analogs [430]. The low-dimensional examples are useful for quick and rigorous assessment of the quality of an approximation. Besides, the predictions from the local and semilocal approximation explored in our study are consistent with the 3D differentiable formulations in Ref 200 and Ref. 100.

6.5 Acknowledgments

This work is supported by National Science Foundation, grant no. DGE-1633631 (B. K.), CHE-1856165 (B. K., K. B.), and Department of Energy, grant no. DE-SC0008696 (R. P.).

Chapter 7

Large scale quantum chemistry with tensor processing units

This chapter is a reproduction of Ref. [307], which I co-authored with John Kozlowski, Ruyi Song, Jackson Beall, Martin Ganahl, Markus Hauru, Adam GM Lewis, Yi Yao, Shrestha Basu Mallick, Volker Blum, and Guifre Vidal.

7.1 Abstract

We demonstrate the use of Google’s cloud-based Tensor Processing Units (TPUs) to accelerate and scale up conventional (cubic-scaling) density functional theory (DFT) calculations. Utilizing 512 TPU cores, we accomplish the largest such DFT computation to date, with 247848 orbitals, corresponding to a cluster of 10327 water molecules with 103270 electrons, all treated explicitly. Our work thus paves the way towards accessible and systematic use of conventional DFT, free of any system-specific constraints, at unprecedeted scales.

7.2 Introduction

Computational methods for quantum chemistry and quantum physics have proven to be invaluable tools in modern scientific research and technological innovation. The application space of such methods is vast, ranging from the prediction of novel high-temperature superconductors [104] to the acceleration of drug discovery [69]; from the study of catalytic processes for e.g. CO₂ sequestration [455] and plastic recycling [189] to the design of nanomaterials [56], solar cells [126], and batteries [416].

In the landscape of quantum-based computational methods, density functional theory (DFT) especially stands out due to its ability to produce accurate results for a wide range of systems at a relatively low computational cost [190]. Accordingly, an impressive amount of computational research utilizes DFT calculations each year. For instance, the US National Energy Research Scientific Computing Center (NERSC) reported that nearly 30% of their supercomputer resources in 2018 were spent on DFT calculations alone [25]. Widespread research and development effort is continuously devoted towards optimizing the performance and accuracy of DFT calculations, giving rise to a plethora of open-source and commercial DFT software packages [394]. Several packages can leverage specialized hardware, such as general-purpose graphics processing units (GPUs), for most of the workload [180, 374, 20, 125, 141, 134, 158]. However, in conventional DFT implementations, i.e., without specific sparsity assumptions for the density matrix or Hamiltonian matrix, the computational cost scales as the third power of the number N of orbitals used to describe the system (referred to $O(N^3)$ DFT throughout this work), and this cubic scaling often makes simulating large systems, such as protein-ligand complexes or metal-organic frameworks [447], prohibitively expensive.

Google’s Tensor Processing Units (TPUs) are application-specific integrated circuits originally designed to accelerate large-scale machine learning workloads [191, 192, 58, 127, 6]. By leveraging the JAX library [58, 127, 6], it is nevertheless possible to repurpose TPUs

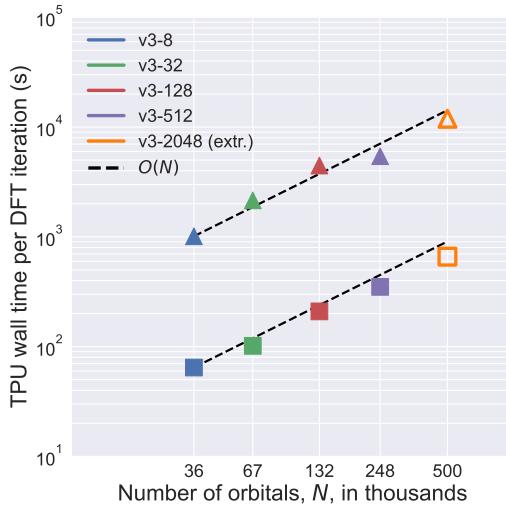


Figure 7.1: TPU v3 wall time for $O(N^3)$ density matrix purification, Eqs. (7.5)-(7.7), as a function of the number N of orbitals, for clusters of water molecules, both in single (squares) and double (triangles) precision. A full TPU v3 pod with 2048 cores and 32 TB of memory is expected to handle $N \sim 500\,000$ orbitals in our current implementation (extrapolated results necessitated by temporary resource unavailability).

for other computational tasks [45, 432, 300, 261, 262, 260, 181, 285, 166, 248, 157, 132]. In this work, we demonstrate the use of TPUs as quantum chemistry supercomputers by accelerating the $O(N^3)$ computational bottleneck of DFT approaches which use an auxiliary single-particle kinetic energy approximation, such as Kohn-Sham (KS) [174, 214] and generalized KS (gKS) [370] DFT, where gKS admits hybrid DFT functionals. This enables the systematic study of quantum chemistry problems at unprecedented scales. As a concrete demonstration, we performed end-to-end $O(N^3)$ DFT calculations on large clusters of water molecules, reaching a total of $N = 247\,848$ DFT orbitals, corresponding to 10 327 water molecules with 103 270 electrons, see Fig. 7.1 and Table 7.1. To our knowledge, this is the largest $O(N^3)$ DFT calculation to date, with the previously largest computation consisting of a single $O(N^3)$ DFT iteration with $N \approx 230\,000$ orbitals on Fujitsu’s K computer [165].

Some variants of DFT, most notably linear-scaling DFT [345, 57, 380, 418], avoid the $O(N^3)$ bottleneck altogether and can thus reach an even larger number of orbitals. However, these variants rely on additional approximations and conditions, such as truncating density matrix

Number of orbitals	Number of atoms	Number of electrons	TPU configuration	TPU wall time (s)		Relative energy per molecule (mHa)
				FP32	FP64	
35 544	4 443	14 810	v3-8	65	1 012	0.934
65 668	8 211	27 370	v3-32	102	2 150	0.531
131 544	16 443	54 810	v3-128	209	4 465	0.291
247 848	30 981	103 270	v3-512	350	5 434	0

Table 7.1: Tabulated results in Fig. 7.1, including also number of atoms and electrons. Wall times for the matrix purification step are shown both for single (FP32) and double (FP64) precision. Energies are relative to the largest calculation, $E[(\text{H}_2\text{O})_{N_{\text{mol}}}] / N_{\text{mol}} - E[(\text{H}_2\text{O})_{10327}] / 10327$, where N_{mol} is the total number of water molecules. In this sequence, we used a number of TPU cores that grows roughly as N^2 . As a result, walltimes are seen to roughly scale linearly in N , instead of the expected $O(N^3)$ scaling.

elements [80], or on special properties of only a subset of density functionals (such as semilocal density functional approximations). In turn, this results in restricted applicability, with e.g. linear-scaling DFT being suitable for insulating systems but not for metals or systems with a small energy gap [345]. In practice, conventional $O(N^3)$ DFT is a more preferable choice since it alleviates technical complexity and problem space restrictions associated with current lower-scaling methods, greatly extending the domain of problems to which DFT can be applied reliably and with relative ease.

There are many aspects that go into an $O(N^3)$ DFT calculation. Throughout we focus on atom-centered basis sets with all electrons treated explicitly, that is we do not consider e.g. plane waves or pseudopotentials. At a high level, one can identify two main computational steps: (a) building the DFT Hamiltonian matrix (with cost $O(N)$ to $O(N^2)$) and (b) computing the ground state density matrix ($O(N^3)$), see Fig. 7.2.

(a) *DFT Hamiltonian build*: Given a choice of N atom-centered basis functions $\chi_i(\mathbf{r})$, one needs to compute the DFT Hamiltonian matrix H and the overlap matrix S , with coefficients given by

$$H_{ij} = \langle \chi_i | \mathcal{H} | \chi_j \rangle, \quad S_{ij} = \langle \chi_i | \chi_j \rangle, \quad (7.1)$$

where \mathcal{H} represents the DFT Hamiltonian in the continuum and each matrix coefficient requires computing one or several integrals. Over the past few decades much effort has been devoted to optimizing the build of the $N \times N$ matrix H . Naively, the computational time here scales as $O(N^4)$, however, in many implementations the scaling is effectively reduced to $O(N^2)$ due to two-electron integral screening methods. The scaling can be further reduced to almost $O(N)$ if other strategies, such as fast multipole methods [437] or fast fourier transform based methods [418], such as the Ewald method for periodic systems, are employed. In this work we do *not* attempt to accelerate the Hamiltonian or overlap matrix build times with TPUs. Instead, we simply use a well-established all-electron DFT package, the *Fritz Haber Institute ab initio molecular simulations package* (FHI-aims) [180, 48, 49], which we run using CPUs.

(b) *Density matrix purification*: The pair of matrices H and S define a generalized eigenvalue problem, the so-called KS equations,

$$H |\phi_\alpha\rangle = e_\alpha S |\phi_\alpha\rangle, \quad (7.2)$$

with $|\phi_\alpha\rangle$ and e_α the KS orbitals and energies. Our goal is to compute the ground state density matrix

$$D \equiv \sum_{\alpha=1}^N \theta(\mu - e_\alpha) |\phi_\alpha\rangle \langle \phi_\alpha|, \quad (7.3)$$

where $\theta(x)$ is the step function and μ is the chemical potential, chosen such that $\sum_{\alpha=1}^N \theta(\mu - e_\alpha) = N_e$, for N_e the number of electrons in the system. The density matrix D can be obtained by solving the KS equations (7.2) using standard linear algebra libraries, such as LAPACK [16] or Intel MKL [3]. An alternative route, which we follow in this work, is to use a density matrix purification scheme [412, 207]. First, by computing the inverse square root

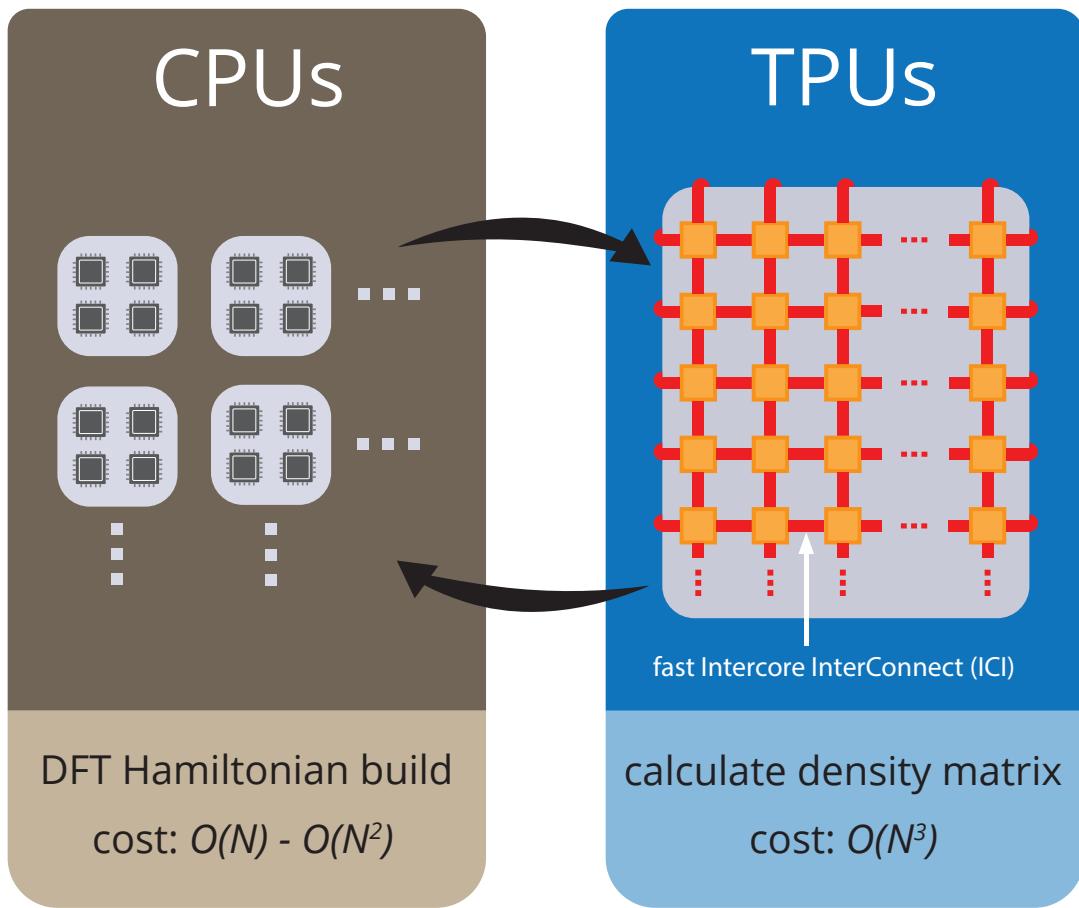


Figure 7.2: The two main steps of our implementation of an $O(N^3)$ DFT computation, the Hamiltonian build and computing the ground state density matrix, which we run on CPUs and TPUs, respectively. The DFT code FHI-aims [48, 49] is used to set up the Hamiltonian and the ELSI library [456, 472] is used to facilitate the integration of the TPU-based solver to FHI-aims.

of S ,

$$S \mapsto S^{-\frac{1}{2}} \quad (7.4)$$

we can write the Hamiltonian in an orthonormal basis,

$$H \mapsto \tilde{H} \equiv S^{-\frac{1}{2}} H S^{-\frac{1}{2}}. \quad (7.5)$$

and re-express (7.2) as a standard eigenvalue problem $\tilde{H} |\tilde{\phi}_\alpha\rangle = e_\alpha |\tilde{\phi}_\alpha\rangle$, where $|\tilde{\phi}_\alpha\rangle \equiv S^{1/2} |\phi_\alpha\rangle$. Next we compute the density matrix \tilde{D} ,

$$\tilde{H} \rightarrow \tilde{D} \equiv \theta(\mu I - \tilde{H}) = \sum_{\alpha=1}^{N_e} \theta(\mu - e_\alpha) |\tilde{\phi}_\alpha\rangle \langle \tilde{\phi}_\alpha|, \quad (7.6)$$

and finally re-express it in the original basis,

$$\tilde{D} \mapsto D \equiv S^{-\frac{1}{2}} \tilde{D} S^{-\frac{1}{2}}. \quad (7.7)$$

The transformation in Eq. (7.6) is obtained using a standard density matrix purification scheme that is suitable for TPUs, namely the hole-particle canonical purification scheme [412], which we elaborate on later in the paper.

If no further modifications are made (e.g., density matrix truncations in linear-scaling DFT), then the cost of computing D , whether by solving Eq. (7.2) or performing the four matrix transformations in Eqs. (7.4)-(7.7), scales as $O(N^3)$. This constitutes what is known as the *cubic wall* of DFT.

The density matrix D is used to derive several important quantities. The real-space electron

density $n(\mathbf{r})$ is given by

$$n(\mathbf{r}) = \sum_{i,j}^N \chi_i(\mathbf{r}) D_{ij} \chi_j(\mathbf{r}), \quad (7.8)$$

which can be computed on a real-space grid [48]. The sum of occupied KS eigenvalues, given by $\text{Tr}(H D) = \text{Tr}(\tilde{H} \tilde{D})$, is also used to compute the total ground-state energy. Additionally, the *energy weighted density matrix* Q ,

$$Q = D H D, \quad (7.9)$$

is also useful to compute atomic forces analytically [48].

7.3 Results

7.3.1 DFT with TPUs

The main result of our work is the successful use of TPUs to perform the four matrix transformations (7.4)-(7.7), thereby tackling the $O(N^3)$ computational bottleneck of DFT. We employed TPUs of the third generation, denoted v3. A single TPU v3 core contains two matrix multiply units (MXUs) to formidably accelerate matrix-matrix multiplication (matmul), resulting in about 10 teraFLOPS (floating point operations per second) of measured single-core matmul performance in single precision. Importantly for our purposes, matmuls are also available in double precision using a software-emulated 57-bit floating point format. In this approach, utilized algorithms require many more single precision floating point operations when operating in our emulated double precision than in single precision, and as a result matmuls in double precision take $\sim 11 \times$ longer than in single precision.

The smallest available TPU configuration consists of 8 TPU v3 cores with a total of 128 GB of dedicated high bandwidth memory (HBM), controlled by a single host with 48 CPU cores. The largest configuration is a pod with 2048 TPU v3 cores and 32 TB of HBM, controlled by 256 hosts. Given a choice of configuration, the available TPU cores are directly connected to nearest neighbors in a 2D torus network through fast inter-core interconnects (ICIs), see Fig. 7.2. The ICIs are critical to maintaining high performance when distributing matmuls and other dense linear algebra operations over all available TPU cores. In this work we used the JAX library [58, 127, 6] to write *single program multiple data* (SPMD) code and executed it on configurations made of p TPU cores, denoted v3- p , for $p = 8, 32, 128$ and 512 .

The TPU hardware architecture is especially suited for dense large-scale matmuls, which we perform in distributed form using the SUMMA algorithm [417], as recently demonstrated in Ref. [248]. Here it was shown that for sufficiently large matrices a v3-512 TPU can perform dense matmuls at near-optimal efficiency: the performance per TPU core (measured in single-precision FLOPS) is maintained at roughly 93% of the single TPU core maximum performance [248]. It is important to emphasize that TPUs are often ill-suited for other tasks, and hence the algorithms utilized in this work and those in Ref. [248] had to be picked carefully and may differ from more conventional choices used in CPUs or GPUs. The use of DM purification algorithms, rather than direct diagonalization, is especially attractive for TPUs since all steps can be evaluated from a series of matmuls. Clearly, transformations (7.5) and (7.7) require large-scale matmuls. Transformations (7.4) and (7.6) are implemented by an iteration involving matrix polynomials of small degree, where each polynomial requires a short sequence of matrix additions and multiplications. Specifically, the matrix inverse square root in (7.4) is implemented using a standard Newton-Schulz iteration [172], whereas for the density matrix purification in (7.6) we implemented the hole-particle canonical purification scheme [412]. Further algorithm details can be found in the Supporting Information.

For benchmarking purposes, we have performed end-to-end DFT computations on a sequence

of increasingly large water clusters with geometries obtained from standard molecular dynamics simulations (see Supporting Information). We leverage the DFT code FHI-aims [48, 49] to set up and drive calculations using CPUs, then use the TPUs to tackle the $O(N^3)$ dense linear algebra bottlenecks (7.4)-(7.7). We also utilize the *ELectronic Structure Infrastructure* (ELSI) library [456, 472] to facilitate the integration of FHI-aims and the TPU solver. In particular, the DFT Hamiltonian build time and associated computational scaling and parallelization are dictated exclusively by the FHI-aims code, which uses numeric atom-centered orbitals (NAOs) with an explicit finite spatial extent, and a truncated multipole expansion to accomplish low prefactor and efficient scaling of the Hamiltonian matrix build. While the computational time required to build the DFT Hamiltonian may vary greatly between different systems with the same total number of orbitals N (due to possible differences in the resulting sparsity in the systems), the computational time required for the $O(N^3)$ DM purification step performed on the TPU has much less variability since dense matrix operations are assumed, which do not utilize any sparsity present (see Supporting Information for more discussion). Thus, we emphasize that the TPU wall times, which are reported throughout only for water clusters, are fairly robust with respect to different systems with the same total number of orbitals.

Throughout this work we perform all-electron calculations using the PBE exchange-correlation functional and utilize an NAO basis set such that each H_2O molecule contributes 10 electrons, represented by 24 orbitals (5 for each hydrogen atom and 14 for the oxygen atom). First we consider a single TPU board with 8 TPU v3 cores controlled by a host with 48 CPU cores, and we run FHI-aims on the host. Fig. 7.3 shows the wall time for a single DFT iteration (including both Hamiltonian build on CPUs and density matrix purification on TPUs) as a function of the size of the water cluster, which ranges from a few thousand to $N \approx 50\,000$ orbitals. When using the TPU solver in single precision (green curve) we see that the $O(N)$ Hamiltonian build on 48 CPU cores takes longer than the $O(N^3)$ density matrix purification run on 8 TPU cores, thus shifting the bottleneck. Using the TPU solver in double precision

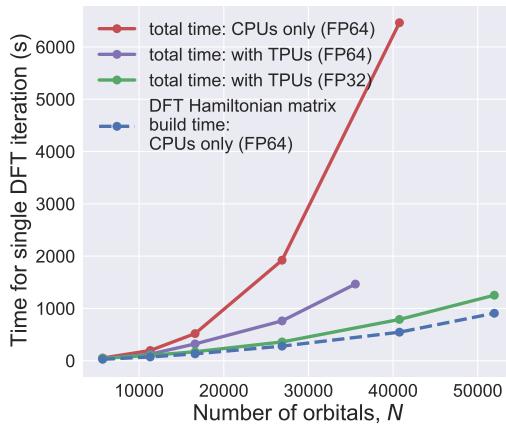


Figure 7.3: Wall times for a single DFT iteration on water clusters, using a TPU board composed of a CPU host (48 CPU cores) and 8 TPU cores with 128 GB of HBM. Green and purple curves correspond to using single and double precision in the TPU solver, respectively. The dashed blue curve corresponds to the CPU time spent on FHI-aims (always in double precision), and should be subtracted from the other curves in order to obtain the time spent on the TPU solvers. For reference, in red we also plot the time required for a CPU-only computation using the *Eigenvalue soLvers for Petaflop Applications* (ELPA), a highly parallelized eigensolver library [223, 267], run on 48 CPU cores.

is an order of magnitude slower and saturates the TPU’s HBM for $N \approx 36\,000$ orbitals.

Then we consider larger TPU configurations, of up to 512 TPU v3 cores, to perform end-to-end DFT computations on larger clusters, of up to 10 327 water molecules (or $N = 247\,848$ orbitals). Fig. 7.1 shows the TPU wall time for the $O(N^3)$ density matrix purification for one DFT iteration. These include 350 (5 434) seconds for a density matrix purification in single (double) precision on the largest cluster, demonstrating feasibility of DFT computations at that scale of a quarter of a million orbitals.

7.3.2 Dynamic precision on TPUs

In our implementation, early DFT iterations are treated with single precision and later ones in double precision. This *dynamic precision* approach allows us to cut down on the use of double precision matmuls on TPUs (which are significantly slower than single precision

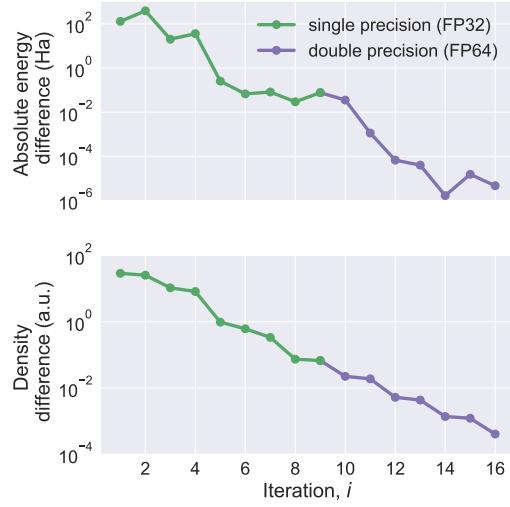


Figure 7.4: Convergence trajectory of an end-to-end dynamic precision DFT calculation on a $(\text{H}_2\text{O})_{10327}$ cluster. The absolute total energy differences between subsequent DFT iterations, i and $i - 1$, are plotted (top). The corresponding difference in real-space densities within the L^1 norm is plotted (bottom).

ones) without sacrificing accuracy of the final converged DFT result. Our criteria to switch precision is based on relative density changes, using the L^1 norm, defined as $L^1[f(\mathbf{r})] \equiv \int d^3r |f(\mathbf{r})|$, and relative energy changes:

$$\frac{1}{N_e} L^1[n^{[i]}(\mathbf{r}) - n^{[i-1]}(\mathbf{r})] < \epsilon \quad \text{and} \quad (7.10)$$

$$|E^{[i]} - E^{[i-1]}|/|E^{[i]}| < \epsilon, \quad (7.11)$$

where $n^{[i]}(\mathbf{r})$ is the real-space density at DFT iteration i , $E^{[i]}$ is the corresponding total ground-state energy, and we use $\epsilon = 5 \times 10^{-7}$ for single precision.

Fig. 7.4 shows the convergence trajectory of a dynamic precision DFT calculation for the largest cluster we have considered. We are able to converge such a DFT calculation to a fairly tight convergence threshold using first 9 single precision DFT iterations, followed by 7 double precision DFT iterations. In a smaller cluster, $(\text{H}_2\text{O})_{1481}$ with $N = 35\,544$ orbitals, a smaller number of double precision iterations are required for convergence, resulting in

an overall DFT calculation time that is under 5 hours on a single TPU board (v3-8), see Supporting Information.

7.4 Conclusion

This work has successfully demonstrated that TPUs can both *accelerate* and *scale up* DFT computations. Significant *acceleration* is already achieved using only a single TPU board with 8 TPU v3 cores, see Fig. 3. For instance, an end-to-end dynamic precision DFT computation with $N = 35\,544$ orbitals consisting of 12 iterations in single precision and 4 iterations in double precision yields converged results in under 5 hours. For context, using double precision only and the highly optimized ELPA $O(N^3)$ solver with 48 CPUs, the same water cluster calculation required 20 hours to achieve 16 DFT iterations.

In order to *scale up* the size of DFT computations while retaining high performance two main ingredients are involved: (i) a larger amount of high bandwidth memory, scaling as $O(N^2)$, to be able to store dense $N \times N$ matrices; (ii) a larger number of cores, with state-of-the-art inter-core connectivity, to more effectively execute the $O(N^3)$ floating point operations involved in the required distributed matrix transformations. As shown in Fig. 1, by using a number of cores that scales as $O(N^2)$ and commensurate amounts of HBM, we can scale up to $N = 500\,000$ orbitals with wall times that only grow proportional to N .

Once the main ingredients (i) and (ii) are satisfied, the Input/Output (IO) time, i.e. the end-to-end communication time between the TPU and CPU, can become an important (and possibly limiting) factor. This IO step is not exclusive to TPUs, for instance, it is also relevant and analogous in GPU setups that require communication with CPUs. This step is parallelizable, such that by using a number of TPU and CPU cores that scales as $O(N^2)$, the total IO time remains constant $O(1)$. However, in this context, obtaining

such optimality will depend on specific implementation details of the DFT code utilized, such as the matrix distribution pattern on the CPU processor grid. Our current prototype implementation is modular and generalizeable and IO times scale unfavorably as $O(N^2)$, even becoming the rate-limiting step in some cases (see Supporting Information). Less-general (but straightforward) engineering approaches are expected to be much more optimal, but are beyond the scope of this work which aims to demonstrate the use of TPUs in a more general context.

We emphasize that other hardware accelerators, most notably GPUs, can also accelerate and scale up DFT computations in a similar manner as discussed above, and Ref. [95] recently presented a useful and positive development in this direction. Modern distributed GPU configurations are expected to achieve similar performance to TPUs in this regard, however, a direct comparison is complicated by the highly diverse nature of distributed GPU configurations found in practice. On the Summit supercomputer configuration, it has been reported that 432 distributed Nvidia V100 GPUs can perform matmuls for dense $N > 500\,000$ matrices with a performance per GPU (measured in FLOPS) that is roughly 85% of the single V100 GPU maximum performance [170].

Here we have focused, for simplicity, on applying DFT to clusters of water molecules. More complicated systems may present additional difficulties. For instance, protein-ligand complexes often require more elaborate schemes, such as including solvation to facilitate the convergence of the DFT iteration [239, 357]. Work in progress shows that our TPU-based large-scale DFT computations can also successfully address protein-ligand complexes with explicit solvents, as well as in a variety of other large systems, including DNA segments, carbon nanotubes, and graphene surfaces. In addition to single-point DFT energy calculations, analytical forces can also be extracted from the TPU-calculated *energy-weighted density matrix*, enabling large-scale geometry optimization or Ab initio molecular dynamics calculations.

DFT is a highly successful quantum-based method, but it is ultimately a consistent-field approximation, which may not be accurate enough for certain applications. Fortunately, TPUs can also accelerate and scale up other, more accurate quantum chemistry approaches where the computational bottleneck is again given by dense linear algebra operations. For example, in density matrix renormalization group (DMRG) [438] calculations, TPUs can be used to reach an unprecedentedly large bond dimension $D = 65\,536$ [132]. Similarly, we anticipate that TPUs will thrive in other methods such as coupled cluster [117] and Møller–Plesset perturbation theory [422]. Even when applying such higher-level methods, large-scale DFT may still be a crucial piece in simulations that require a quantum-mechanically treated region that embeds a subsystem treated with higher-level correlated methods [390].

To conclude, in this work we have successfully repurposed TPUs as quantum chemistry supercomputers by tackling the $O(N^3)$ computational bottleneck of density functional theory. We demonstrated performance and scalability with a water cluster with $N = 247\,848$ orbitals, which to our knowledge is the largest $O(N^3)$ DFT computation to date. We remark that cloud-based TPUs, and other hardware accelerators such as GPUs, are more accessible and affordable than traditional supercomputer resources. Our work thus paves the way towards accessible and straightforward use of quantum chemistry computational methods for much larger systems than were previously possible.

Competing Interests: V.B. received compensation as an advisor from Google during part of this work. V.B. is also a board member of MS1P e.V., the non-profit organization that licenses the FHI-aims electronic structure code used in this work. He does not receive any financial gains from this position.

7.5 Acknowledgments

The authors would like to thank Toru Shiozaki and Garnet Kin-Lic Chan for suggesting to investigate the use of TPUs to accelerate mean-field quantum chemistry methods (by accelerating matrix multiplications for a diagonalization-free construction of the density matrix, as in the density matrix purification used in this paper), and Toru Shiozaki, Garnet Kin-Lic Chan, Chase Roberts and Stefan Leichenauer for previous exploratory work in this direction. Also, special thanks to Xing Zhang and Garnet Kin-Lic Chan for work adjusting PySCF, as part of an on-going integration of our TPU solver, to be described elsewhere, and to David Bowler, Tsuyoshi Miyazaki, and Jun-ichi Iwata for help documenting the largest DFT computation run on the K computer. Finally, the authors would also like to thank Giuseppe M. J. Barca, Anudhyan Boral, Michael Brenner, Kieron Burke, Rafael Gomez-Bombarelli, JW Feng, Philipp Furche, Andreas Goeller, Stephan Hoyer, Olivier Lacombe, Stefan Leichenauer, Lin Lin, Ruben Martin Romo, Todd Martinez, Anders M. N. Niklasson, Nicholas Rubin, Zak Stone, Matthias Tan, Keiran Thompson, Edward Valeev, and Jae Yoo for useful discussions and comments. Research supported with Cloud TPUs from Google’s TPU Research Cloud (TRC). Sandbox is a team within the Alphabet family of companies, which includes Google, Verily, Waymo, X, and others. G.V. is a CIFAR fellow in the Quantum Information Science Program, a Distinguished Invited Professor at the Institute of Photonic Sciences (ICFO), and a Distinguished Visiting Research Chair at Perimeter Institute. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Research, Innovation, and Science. R.S. and Y.Y. were partially supported by the National Science Foundation (NSF), USA under Award No. 1450280.

Bibliography

- [1] See Supplemental Material below for information about 1d model systems; computational details of DMRG; KS calculations; training, validation and test; neural networks; training without KSR and training with weaker KSR. It includes Refs. [29, 124, 58, 167, 138, 258, 421, 365].
- [2] Defined through Gaussian implementation.
- [3] *Intel Math Kernel Library. Reference Manual*. Intel Corporation, Santa Clara, USA, 2009. ISBN 630813-054US.
- [4] https://github.com/pedensor/dft_exconditions, 2023.
- [5] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [6] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, page 265–283, USA, 2016. USENIX Association.
- [7] A. Abedi, N. T. Maitra, and E. K. U. Gross. Exact factorization of the time-dependent electron-nuclear wave function. *Phys. Rev. Lett.*, 105:123002, Sep 2010.
- [8] A. Abedi, N. T. Maitra, and E. K. U. Gross. Correlated electron-nuclear dynamics: Exact factorization of the molecular wavefunction. *The Journal of Chemical Physics*, 137(22), 2012.
- [9] C. Adamo and V. Barone. Toward reliable adiabatic connection models free from adjustable parameters. *Chem. Phys. Lett.*, 274(1-3):242, 1997.

- [10] C. Adamo and V. Barone. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.*, 108(2):664, 1998.
- [11] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110(13):6158, 1999.
- [12] R. D. Adamson, P. M. W. Gill, and J. A. Pople. Empirical density functionals. *Chem. Phys. Lett.*, 284(1–2):6, 1998.
- [13] F. Agostini and E. Gross. Ultrafast dynamics with the exact factorization. *The European Physical Journal B*, 94(9):1–14, 2021.
- [14] W. Ai, W.-H. Fang, and N. Q. Su. The role of range-separated correlation in long-range corrected hybrid functionals. *J. Phys. Chem. Lett.*, 12:1207–1213, 2021.
- [15] C.-O. Almbladh and U. von Barth. Exact results for the charge and spin densities, exchange-correlation potentials, and density-functional eigenvalues. *Physical Review B*, 31(6):3231, 1985.
- [16] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999.
- [17] J. B. Anderson. Quantum chemistry by random walk. h 2p, h+3 d3h 1a1, h2 3Σ+u, h4 1Σ+g, be 1s. *The Journal of Chemical Physics*, 65(10):4121–4127, nov 1976.
- [18] M. Andrychowicz, M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. De Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in neural information processing systems*, pages 3981–3989, 2016.
- [19] J. A. Anta and A. A. Louis. Probing ion-ion and electron-ion correlations in liquid metals within the quantum hypernetted chain approximation. *Phys. Rev. B*, 61:11400–11410, May 2000.
- [20] E. Aprà, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtli, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha,

A. Otero-de-la Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, A. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, and R. J. Harrison. Nwchem: Past, present, and future. *The Journal of Chemical Physics*, 152(18):184102, 2020.

- [21] A. J. Archer, B. Chacko, and R. Evans. The standard mean-field treatment of inter-particle attraction in classical dft is better than one might expect. *The Journal of Chemical Physics*, 147(3):034501, 2017.
- [22] R. Armiento and A. E. Mattsson. Functional designed to include surface effects in self-consistent density functional theory. *Phys. Rev. B*, 72:085108, Aug 2005.
- [23] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Saunders College Publishing, 1976.
- [24] A. Austin, G. A. Petersson, M. J. Frisch, F. J. Dobek, G. Scalmani, and K. Throssell. A density functional with spherical atom dispersion terms. *J. Chem. Theory Comput.*, 8(12):4989–5007, 2012.
- [25] B. Austin. Nersc-10 workload analysis (data from 2018). https://portal.nersc.gov/project/m888/nersc10/workload/N10_Workload_Analysis.latest.pdf. Accessed: 11-20-2022.
- [26] B. M. Austin, D. Y. Zubarev, and W. A. Lester. Quantum monte carlo and related approaches. *Chemical Reviews*, 112(1):263–288, 2012.
- [27] S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [28] T. E. Baker, K. Burke, and S. R. White. Accurate correlation energies in one-dimensional systems from small system-adapted basis functions. *Physical Review B*, 97(8):085139, 2018.
- [29] T. E. Baker, E. M. Stoudenmire, L. O. Wagner, K. Burke, and S. R. White. One-dimensional mimicking of electronic structure: The case for exponentials. *Physical Review B*, 91(23):235141, 2015.
- [30] T. E. Baker, E. M. Stoudenmire, L. O. Wagner, K. Burke, and S. R. White. One-dimensional mimicking of electronic structure: The case for exponentials. *Phys. Rev. B*, 91:235141, Jun 2015.
- [31] Y. Bar-Sinai, S. Hoyer, J. Hickey, and M. P. Brenner. Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences*, page 201814058, 2019.

- [32] R. J. Bartlett and M. Musial. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.*, 79:291–352, Feb 2007.
- [33] A. P. Bartók and J. R. Yates. Regularized SCAN functional. *J. Chem. Phys.*, 150(16):161101, 2019.
- [34] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic differentiation in machine learning: a survey. *The Journal of Machine Learning Research*, 18(1):5595–5637, Jan. 2017.
- [35] A. D. Becke. Correlation energy of an inhomogeneous electron gas: A coordinate-space model. *J. Chem. Phys.*, 88(2):1053–1062, 1988.
- [36] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098–3100, Sep 1988.
- [37] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98(7):5648, 1993.
- [38] A. D. Becke. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.*, 98(2):1372, 1993.
- [39] A. D. Becke. Thermochemical tests of a kinetic-energy dependent exchange-correlation approximation. *Int. J. Quantum Chem.*, 52(S28):625–632, 1994.
- [40] A. D. Becke. Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing. *J. Chem. Phys.*, 104(3):1040, 1996.
- [41] A. D. Becke. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.*, 107(20):8554, 1997.
- [42] A. D. Becke. A new inhomogeneity parameter in density-functional theory. *J. Chem. Phys.*, 109(6):2092–2098, 1998.
- [43] A. D. Becke. Density-functional theory versus density-functional fits. *The Journal of Chemical Physics*, 2022.
- [44] J. Behler and M. Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters*, 98(14):146401, 2007.
- [45] F. Belletti, D. King, K. Yang, R. Nelet, Y. Shafi, Y.-F. Shen, and J. Anderson. Tensor processing units for financial monte carlo. In *Proceedings of the 2020 SIAM Conference on Parallel Processing for Scientific Computing*, pages 12–23, 2020.
- [46] Y. A. Bernard, Y. Shao, and A. I. Krylov. General formulation of spin-flip time-dependent density functional theory using non-collinear kernels: Theory, implementation, and benchmarks. *J. Chem. Phys.*, 136(20):204103, 2012.

[47] D. I. Bilc, R. Orlando, R. Shaltaf, G.-M. Rignanese, J. Íñiguez, and P. Ghosez. Hybrid exchange-correlation functional for accurate prediction of the electronic and structural properties of ferroelectric oxides. *Phys. Rev. B*, 77:165107, Apr 2008.

[48] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler. Ab initio molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications*, 180(11):2175–2196, 2009.

[49] V. Blum, S. Kokott, M. Rossi, and M. Scheffler. FHI-AIMS. <https://fhi-aims.org/>. Accessed: 11-20-2022.

[50] A. D. Boese, A. Chandra, J. M. L. Martin, and D. Marx. From ab initio quantum chemistry to molecular dynamics: The delicate case of hydrogen bonding in ammonia. *J. Chem. Phys.*, 119(12):5965, 2003.

[51] A. D. Boese, N. L. Doltsinis, N. C. Handy, and M. Sprik. New generalized gradient approximation functionals. *J. Chem. Phys.*, 112(4):1670, 2000.

[52] A. D. Boese and N. C. Handy. A new parametrization of exchange-correlation generalized gradient approximation functionals. *J. Chem. Phys.*, 114(13):5497, 2001.

[53] A. D. Boese and N. C. Handy. New exchange-correlation density functionals: The role of the kinetic-energy density. *J. Chem. Phys.*, 116(22):9559, 2002.

[54] A. D. Boese and J. M. L. Martin. Development of density functionals for thermochemical kinetics. *J. Chem. Phys.*, 121(8):3405, 2004.

[55] M. Bonitz, T. Dornheim, Z. A. Moldabekov, S. Zhang, P. Hamann, H. Kahlert, A. Filinov, K. Ramakrishna, and J. Vorberger. Ab initio simulation of warm dense matter. *Physics of Plasmas*, 27(4):042710, 2020.

[56] G. Boschetto, T. Xu, M. Yehya, J. Thireau, A. Lacampagne, B. Charlot, T. Gil, and A. Todri-Sanial. Exploring 1d and 2d nanomaterials for health monitoring wearable devices. In *2021 IEEE International Conference on Flexible and Printable Sensors and Systems (FLEPS)*, pages 1–4. IEEE, 2021.

[57] D. R. Bowler and T. Miyazaki. Calculations for millions of atoms with density functional theory: linear scaling shows its potential. *Journal of Physics: Condensed Matter*, 22(7):074207, 2010.

[58] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, and S. Wanderman-Milne. JAX: composable transformations of Python+NumPy programs. <http://github.com/google/jax>, 2018.

[59] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller. Bypassing the kohn-sham equations with machine learning. *Nature communications*, 8(1):1–10, 2017.

[60] K. Burke. Perspective on density functional theory. *J. Chem. Phys.*, 136:150901, 2012.

- [61] K. Burke, A. Cancio, T. Gould, and S. Pittalis. Atomic correlation energies and the generalized gradient approximation. *ArXiv e-prints*, September 2014.
- [62] K. Burke, J. Perdew, and D. Langreth. Is the local spin density approximation exact for short-wavelength fluctuations? *Phys. Rev. Lett.*, 73:1283, 1994.
- [63] K. Burke, J. P. Perdew, and M. Ernzerhof. Why semilocal functionals work: Accuracy of the on-top pair density and importance of system averaging. *The Journal of Chemical Physics*, 109(10):3760–3771, 1998.
- [64] M. T. Caldeira and R. Custodio. Partial combination of composite strategy and the B3LYP functional for the calculation of enthalpies of formation. *J. Mol. Model.*, 25(3):62, Feb 2019.
- [65] A. Cancio, G. P. Chen, B. T. Krull, and K. Burke. Fitting a round peg into a round hole: Asymptotically correcting the generalized gradient approximation for correlation. *J. Chem. Phys.*, 149(8):084116, 2018.
- [66] A. C. Cancio and M. Y. Chou. Beyond the local approximation to exchange and correlation: The role of the Laplacian of the density in the energy density of Si. *Phys. Rev. B*, 74:081202, Aug 2006.
- [67] J. Carmona-Espíndola, J. L. Gázquez, A. Vela, and S. B. Trickey. Global hybrid exchange energy functional with correct asymptotic behavior of the corresponding potential. *Theor. Chem. Acc.*, 135(5):120, 2016.
- [68] J. Carmona-Espíndola, J. L. Gázquez, A. Vela, and S. B. Trickey. Generalized gradient approximation exchange energy functional with near-best semilocal performance. *J. Chem. Theory Comput.*, 15(1):303–310, 2019.
- [69] C. N. Cavasotto, N. S. Adler, and M. G. Aucar. Quantum chemical approaches in structure-based virtual screening and lead optimization. *Frontiers in chemistry*, 6:188, 2018.
- [70] T. Chachiyo and H. Chachiyo. Understanding electron correlation energy through density functional theory. *Comput. Theor. Chem.*, 1172:112669, 2020.
- [71] D. Chakraborty, K. Berland, and T. Thonhauser. Next-generation nonlocal van der waals density functional. *Journal of Chemical Theory and Computation*, 16(9):5893–5911, 2020.
- [72] Y. Chen, L. Zhang, H. Wang, and W. E. DeePKS: A Comprehensive Data-Driven Approach toward Chemically Accurate Density Functional Theory. *Journal of Chemical Theory and Computation*, 17(1):170–181, jan 2021.
- [73] J. Chihara. Unified description of metallic and neutral liquids and plasmas. *Journal of Physics: Condensed Matter*, 3(44):8715–8744, nov 1991.

- [74] L. Chiodo, L. A. Constantin, E. Fabiano, and F. Della Sala. Nonuniform scaling applied to surface energies of transition metals. *Phys. Rev. Lett.*, 108:126402, Mar 2012.
- [75] J. Čížek. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *The Journal of Chemical Physics*, 45(11):4256–4266, dec 1966.
- [76] A. J. Cohen and N. C. Handy. Assessment of exchange correlation functionals. *Chem. Phys. Lett.*, 316(1–2):160, 2000.
- [77] A. J. Cohen and N. C. Handy. Dynamic correlation. *Mol. Phys.*, 99(7):607, 2001.
- [78] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Insights into current limitations of density functional theory. *Science*, 321(5890):792–794, 2008.
- [79] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Challenges for density functional theory. *Chemical reviews*, 112(1):289–320, 2012.
- [80] D. Cole, C.-K. Skylaris, E. Rajendra, A. Venkitaraman, and M. Payne. Protein-protein interactions from linear-scaling first-principles quantum-mechanical calculations. *Europhysics Letters*, 91(3):37004, 2010.
- [81] R. Colle and O. Salvetti. Approximate calculation of the correlation energy for the closed shells. *Theor. Chim. Acta*, 37(4):329, 1975.
- [82] L. A. Constantin, E. Fabiano, and F. Della Sala. Improving atomization energies of molecules and solids with a spin-dependent gradient correction from one-electron density analysis. *Phys. Rev. B*, 84:233103, Dec 2011.
- [83] L. A. Constantin, E. Fabiano, and F. Della Sala. Semilocal dynamical correlation with increased localization. *Phys. Rev. B*, 86:035130, Jul 2012.
- [84] L. A. Constantin, E. Fabiano, S. Laricchia, and F. Della Sala. Semiclassical neutral atom as a reference system in density functional theory. *Phys. Rev. Lett.*, 106:186406, May 2011.
- [85] L. A. Constantin, E. Fabiano, and F. D. Sala. Spin-dependent gradient correction for more accurate atomization energies of molecules. *J. Chem. Phys.*, 137(19):194105, 2012.
- [86] L. A. Constantin, A. Terentjevs, F. Della Sala, P. Cortona, and E. Fabiano. Semiclassical atom theory applied to solid-state physics. *Phys. Rev. B*, 93:045126, Jan 2016.
- [87] P. Cortona. Note: Theoretical mixing coefficients for hybrid functionals. *J. Chem. Phys.*, 136(8):086101, 2012.
- [88] H. Cox, A. L. Baskerville, V. Syrjanen, and M. Melgaard. The bound state stability of the hydride ion in hf theory. *Advances in Quantum Chemistry*, 2020.

[89] S. Crisostomo, R. Pederson, J. Kozlowski, B. Kalita, A. C. Cancio, K. Datchev, A. Wasserman, S. Song, and K. Burke. Seven useful questions in density functional theory. *Letters in Mathematical Physics*, 113(2):42, 2023.

[90] F. G. Cruz, K.-C. Lam, and K. Burke. Exchange- correlation energy density from virial theorem. *The Journal of Physical Chemistry A*, 102(25):4911–4917, 1998.

[91] G. I. Csonka, J. P. Perdew, and A. Ruzsinszky. Global hybrid functionals: A look at the engine under the hood. *J. Chem. Theory Comput.*, 6(12):3688, 2010.

[92] E. D. Cubuk, A. D. Sendek, and E. J. Reed. Screening billions of candidates for solid lithium-ion conductors: A transfer learning approach for small data. *The Journal of chemical physics*, 150(21):214701, 2019.

[93] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople. Gaussian-2 theory for molecular energies of first-and second-row compounds. *The Journal of chemical physics*, 94(11):7221–7230, 1991.

[94] E. E. Dahlke and D. G. Truhlar. Improved density functionals for water. *J. Phys. Chem. B*, 109(33):15677, 2005.

[95] S. Das, P. Motamarri, V. Subramanian, D. M. Rogers, and V. Gavini. Dft-fe 1.0: A massively parallel hybrid cpu-gpu density functional theory code using finite-element discretization. *arXiv preprint arXiv:2203.07820*, 2022.

[96] B. Davoudi, M. Polini, R. Asgari, and M. Tosi. Self-consistent overhauser model for the pair distribution function of an electron gas in dimensionalities $d= 3$ and $d= 2$. *Physical Review B*, 66(7):075110, 2002.

[97] J. M. del Campo, J. L. Gázquez, S. B. Trickey, and A. Vela. Non-empirical improvement of PBE and its hybrid PBE0 for general description of molecular properties. *J. Chem. Phys.*, 136(10):104108, 2012.

[98] S. Dick and M. Fernandez-Serra. Machine learning accurate exchange and correlation functionals of the electronic density. *Nature Communications*, 11(1):3509, 2020.

[99] S. Dick and M. Fernandez-Serra. Highly accurate and constrained density functional obtained with differentiable programming. *Physical Review B*, 104(16):L161109, 2021.

[100] S. Dick and M. Fernandez-Serra. Using differentiable programming to obtain an energy and density-optimized exchange-correlation functional, 2021.

[101] P. A. M. Dirac. Note on exchange phenomena in the Thomas atom. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26(03):376–385, 1930.

[102] T. Dornheim, S. Groth, and M. Bonitz. The uniform electron gas at warm dense matter conditions. *Physics Reports*, 744:1–86, 2018.

[103] R. M. Dreizler and E. K. U. Gross. *Density Functional Theory: An Approach to the Quantum Many-Body Problem*. Springer–Verlag, Berlin, 1990.

- [104] D. Duan, H. Yu, H. Xie, and T. Cui. Ab initio approach and its impact on superconductivity. *Journal of Superconductivity and Novel Magnetism*, 32(1):53–60, 2019.
- [105] J. C. Duchi, M. I. Jordan, M. J. Wainwright, and A. Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015.
- [106] S. Elfwing, E. Uchibe, and K. Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 107:3–11, 2018.
- [107] S. Elfwing, E. Uchibe, and K. Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 107:3–11, 2018. Special issue on deep reinforcement learning.
- [108] E. Engel, J. Chevary, L. Macdonald, and S. Vosko. Asymptotic properties of the exchange energy density and the exchange potential of finite systems: relevance for generalized gradient approximations. *Zeitschrift für Physik D Atoms, Molecules and Clusters*, 23(1):7–14, 1992.
- [109] H. Englisch and R. Englisch. Hohenberg–Kohn theorem and non-V-representable densities. *Physica A: Statistical Mechanics and its Applications*, 121(1-2):253–268, 1983.
- [110] M. Ernzerhof, K. Burke, and J. Perdew. Long-range asymptotic behavior of ground-state wavefunctions. *J. Chem. Phys.*, 105:2798, 1996.
- [111] M. Ernzerhof and G. E. Scuseria. Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *J. Chem. Phys.*, 110(11):5029, 1999.
- [112] R. Evans. The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids. *Advances in Physics*, 28(2):143–200, 1979.
- [113] E. Fabiano, L. A. Constantin, P. Cortona, and F. Della Sala. Global hybrids from the semiclassical atom theory satisfying the local density linear response. *J. Chem. Theory Comput.*, 11(1):122–131, 2015.
- [114] E. Fabiano, L. A. Constantin, and F. Della Sala. Generalized gradient approximation bridging the rapidly and slowly varying density regimes: A PBE-like functional for hybrid interfaces. *Phys. Rev. B*, 82:113104, Sep 2010.
- [115] E. Fabiano, L. A. Constantin, and F. Della Sala. Testing the broad applicability of the PBEint GGA functional and its one-parameter hybrid form. *Int. J. Quantum Chem.*, 113(5):673–682, 2013.
- [116] E. Fabiano, P. E. Trevisanutto, A. Terentjevs, and L. A. Constantin. Generalized gradient approximation correlation energy functionals based on the uniform electron gas with gap model. *J. Chem. Theory Comput.*, 10(5):2016–2026, 2014.

[117] B. S. Fales, E. R. Curtis, K. G. Johnson, D. Lahana, S. Seritan, Y. Wang, H. Weir, T. J. Martínez, and E. G. Hohenstein. Performance of coupled-cluster singles and doubles on modern stream processing architectures. *Journal of Chemical Theory and Computation*, 16(7):4021–4028, 2020.

[118] E. Fermi. Statistical method to determine some properties of atoms. *Rend. Accad. Naz. Lincei*, 6(602-607):5, 1927.

[119] E. Fermi. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente (a statistical method for the determination of some atomic properties and the application of this method to the theory of the periodic system of elements). *Zeitschrift für Physik A Hadrons and Nuclei*, 48:73–79, 1928.

[120] R. P. Feynman, N. Metropolis, and E. Teller. Equations of state of elements based on the generalized fermi-thomas theory. *Phys. Rev.*, 75:1561–1573, May 1949.

[121] M. Filatov and W. Thiel. A new gradient-corrected exchange-correlation density functional. *Mol. Phys.*, 91(5):847, 1997.

[122] M. Filatov and W. Thiel. A nonlocal correlation energy density functional from a Coulomb hole model. *Int. J. Quantum Chem.*, 62(6):603, 1997.

[123] K. Finzel, P. W. Ayers, and P. Bultinck. A simple algorithm for the kohn-sham inversion problem applicable to general target densities. *Theoretical Chemistry Accounts*, 137(3):6, 2018.

[124] M. Fishman, S. R. White, and E. M. Stoudenmire. The ITensor software library for tensor network calculations, 2020.

[125] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Cariati, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision A.1, 2016. Gaussian Inc. Wallingford CT 2009.

[126] J. M. Frost, K. T. Butler, F. Brivio, C. H. Hendon, M. Van Schilfgaarde, and A. Walsh. Atomistic origins of high-performance in hybrid halide perovskite solar cells. *Nano letters*, 14(5):2584–2590, 2014.

- [127] R. Frostig, M. Johnson, and C. Leary. Compiling machine learning programs via high-level tracing. 2018.
- [128] D. Frydel, W. M. Terilla, and K. Burke. Adiabatic connection from accurate wavefunction calculations. *The Journal of Chemical Physics*, 112(12):5292–5297, 2000.
- [129] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun. Accurate and numerically efficient r^2 SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.*, 11(19):8208–8215, 2020.
- [130] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun. Correction to "Accurate and numerically efficient r^2 SCAN meta-generalized gradient approximation". *J. Phys. Chem. Lett.*, 11(21):9248–9248, 2020.
- [131] L. Gagliardi, D. G. Truhlar, G. Li Manni, R. K. Carlson, C. E. Hoyer, and J. L. Bao. Multiconfiguration pair-density functional theory: A new way to treat strongly correlated systems. *Accounts of chemical research*, 50(1):66–73, 2017.
- [132] M. Ganahl, J. Beall, M. Hauru, A. G. Lewis, J. H. Yoo, Y. Zou, and G. Vidal. Density matrix renormalization group with tensor processing units. *arXiv preprint arXiv:2204.05693*, 2022.
- [133] A. García, N. Papior, A. Akhtar, E. Artacho, V. Blum, E. Bosoni, P. Brandimarte, M. Brandbyge, J. I. Cerdá, F. Corsetti, R. Cuadrado, V. Dikan, J. Ferrer, J. Gale, P. García-Fernández, V. M. García-Suárez, S. García, G. Huhs, S. Illera, R. Korytár, P. Koval, I. Lebedeva, L. Lin, P. López-Tarifa, S. G. Mayo, S. Mohr, P. Ordejón, A. Postnikov, Y. Pouillon, M. Pruneda, R. Robles, D. Sánchez-Portal, J. M. Soler, R. Ullah, V. W.-z. Yu, and J. Junquera. Siesta: Recent developments and applications. *The Journal of Chemical Physics*, 152(20):204108, 2020.
- [134] L. Genovese, B. Videau, M. Ospici, T. Deutsch, S. Goedecker, and J.-F. Méhaut. Daubechies wavelets for high performance electronic structure calculations: The bigdft project. *Comptes Rendus Mécanique*, 339(2):149–164, 2011.
- [135] S. Giarrusso. (private communication).
- [136] S. Giarrusso, S. Vuckovic, and P. Gori-Giorgi. Response potential in the strong-interaction limit of density functional theory: Analysis and comparison with the coupling-constant average. *Journal of Chemical Theory and Computation*, 14(8):4151–4167, 2018. PMID: 29906106.
- [137] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning- Volume 70*, pages 1263–1272. JMLR. org, 2017.
- [138] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[139] L. Goerigk and S. Grimme. A general database for main group thermochemistry, kinetics, and noncovalent interactions – assessment of common and reparameterized (meta-)GGA density functionals. *J. Chem. Theory Comput.*, 6(1):107, 2010.

[140] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme. A look at the density functional theory zoo with the advanced gmtkn55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics*, 19(48):32184–32215, 2017.

[141] X. Gonze, F. Jollet, F. Abreu Araujo, D. Adams, B. Amadon, T. Appelcourt, C. Audouze, J.-M. Beuken, J. Bieder, A. Bokhanchuk, E. Bousquet, F. Bruneval, D. Caliste, M. Côté, F. Dahm, F. Da Pieve, M. Delaveau, M. Di Gennaro, B. Dorado, C. Espejo, G. Geneste, L. Genovese, A. Gerossier, M. Giantomassi, Y. Gillet, D. Hamann, L. He, G. Jomard, J. Laflamme Janssen, S. Le Roux, A. Levitt, A. Lherbier, F. Liu, I. Lukačević, A. Martin, C. Martins, M. Oliveira, S. Poncé, Y. Pouillon, T. Rangel, G.-M. Rignanese, A. Romero, B. Rousseau, O. Rubel, A. Shukri, M. Stankovski, M. Torrent, M. Van Setten, B. Van Troeye, M. Verstraete, D. Waroquiers, J. Wiktor, B. Xu, A. Zhou, and J. Zwanziger. Recent developments in the abinit software package. *Computer Physics Communications*, 205:106–131, 2016.

[142] X. Gonze, J. S. Zhou, and L. Reining. Variations on the “exact factorization” theme. *The European Physical Journal B*, 91(10):1–21, 2018.

[143] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.

[144] P. Gori-Giorgi and E. J. Baerends. Asymptotic nodal planes in the electron density and the potential in the effective equation for the square root of the density. *The European Physical Journal B*, 91(7):160, 2018.

[145] P. Gori-Giorgi, T. Gál, and E. J. Baerends. Asymptotic behaviour of the electron density and the kohn–sham potential in case of a kohn–sham homo nodal plane. *Molecular Physics*, 114(7–8):1086–1097, 2016.

[146] P. Gori-Giorgi and A. Savin. Simple model for the spherically and system-averaged pair density: Results for two-electron atoms. *Physical Review A*, 71(3):032513, 2005.

[147] P. Gori-Giorgi and A. Savin. Kohn-sham calculations combined with an average pair-density functional theory. *International Journal of Modern Physics B*, 21(13n14):2449–2459, 2007.

[148] P. Gori-Giorgi and A. Savin. Range separation combined with the overhauser model: Application to the h₂ molecule along the dissociation curve. *International Journal of Quantum Chemistry*, 109(9):1950–1961, 2009.

[149] P. Gori-Giorgi, M. Seidl, and G. Vignale. Density-functional theory for strongly interacting electrons. *Phys. Rev. Lett.*, 103:166402, Oct 2009.

[150] F. Graziani, M. P. Desjarlais, R. Redmer, and S. B. Trickey, editors. *Frontiers and Challenges in Warm Dense Matter*, volume 96 of *Lecture Notes in Computational Science and Engineering*. Springer International Publishing, 2014.

[151] S. Grimme. Accurate calculation of the heats of formation for large main group compounds with spin-component scaled MP2 methods. *J. Phys. Chem. A*, 109(13):3067–3077, 2005.

[152] S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787, 2006.

[153] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.*, 132(15):154104, 2010.

[154] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of chemical physics*, 132(15):154104, 2010.

[155] S. Groth, T. Dornheim, T. Sjostrom, F. D. Malone, W. M. C. Foulkes, and M. Bonitz. Ab initio exchange-correlation free energy of the uniform electron gas at warm dense matter conditions. *Phys. Rev. Lett.*, 119:135001, Sep 2017.

[156] O. Gunnarsson and B. Lundqvist. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B*, 13:4274, 1976.

[157] E. Gustafson, B. Holzman, J. Kowalkowski, H. Lamm, A. Y. Li, G. Perdue, S. V. Isakov, O. Martin, R. Thomson, J. Beall, M. Ganahl, G. Vidal, and E. Peters. Large scale multi-node simulations of z2 gauge theory quantum circuits using google cloud platform. In *2021 IEEE/ACM Second International Workshop on Quantum Computing Software (QCS)*, pages 72–79, Los Alamitos, CA, USA, nov 2021. IEEE Computer Society.

[158] M. Hacene, A. Anciaux-Sedrakian, X. Rozanska, D. Klahr, T. Guignon, and P. Fleurat-Lessard. Accelerating vasp electronic structure calculations using graphic processing units. *Journal of computational chemistry*, 33(32):2581–2589, 2012.

[159] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.*, 109(15):6264, 1998.

[160] N. C. Handy and A. J. Cohen. A dynamical correlation functional. *J. Chem. Phys.*, 116(13):5411–5418, 2002.

[161] N. C. Handy and D. J. Tozer. The development of new exchange-correlation functionals: 3. *Mol. Phys.*, 94(4):707, 1998.

[162] J.-P. Hansen and I. R. McDonald, editors. *Theory of Simple Liquids*. Academic Press, Oxford, fourth edition edition, 2013.

- [163] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020.
- [164] J. Harris and R. Jones. The surface energy of a bounded electron gas. *J. Phys. F*, 4:1170, 1974.
- [165] Y. Hasegawa, J.-I. Iwata, M. Tsuji, D. Takahashi, A. Oshiyama, K. Minami, T. Boku, H. Inoue, Y. Kitazawa, I. Miyoshi, and M. Yokokawa. Performance evaluation of ultra-large-scale first-principles electronic structure calculation code on the k computer. *Int. J. High Perform. Comput. Appl.*, 28(3):335–355, aug 2014.
- [166] M. Hauru, A. Morningstar, J. Beall, M. Ganahl, A. Lewis, and G. Vidal. Simulation of quantum physics with tensor processing units: brute-force computation of ground states and time evolution. *arXiv preprint arXiv:2111.10466*, 2021.
- [167] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [168] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [169] W. Heitler and F. London. Interaction between neutral atoms and homopolar binding according to quantum mechanics. *Z. Physik*, 44:455, 1927.
- [170] T. Herault, Y. Robert, G. Bosilca, and J. Dongarra. Generic matrix multiplication for multi-gpu accelerated distributed-memory platforms over parsec. In *2019 IEEE/ACM 10th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems (ScalA)*, pages 33–41. IEEE, 2019.
- [171] J. Hermann, Z. Schätzle, and F. Noé. Deep-neural-network solution of the electronic schrödinger equation. *Nature Chemistry*, pages 1–7, 2020.
- [172] N. J. Higham. Stable iterations for the matrix square root. *Numerical Algorithms*, 15(2):227–242, 1997.
- [173] W.-M. Hoe, A. J. Cohen, and N. C. Handy. Assessment of a new local exchange functional OPTX. *Chem. Phys. Lett.*, 341(3):319–328, 2001.
- [174] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, Nov 1964.
- [175] J. Hollingsworth, L. Li, T. E. Baker, and K. Burke. Can exact conditions improve machine-learned density functionals? *The Journal of Chemical Physics*, 148(24):241743, 2018.

[176] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

[177] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu, and T. Frauenheim. Dftb+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of Chemical Physics*, 152(12):124101, 2020.

[178] S. Hoyer, J. Sohl-Dickstein, and S. Greydanus. Neural reparameterization improves structural optimization. *arXiv preprint arXiv:1909.04240*, 2019.

[179] C. D. Hu and D. C. Langreth. A spin dependent version of the Langreth–Mehl exchange-correlation functional. *Phys. Scr.*, 32(4):391, 1985.

[180] W. P. Huhn, B. Lange, V. W.-z. Yu, M. Yoon, and V. Blum. Gpu acceleration of all-electron electronic structure theory using localized numeric atom-centered basis functions. *Computer Physics Communications*, 254:107314, 2020.

[181] F. Huot, Y.-F. Chen, R. Clapp, C. Boneti, and J. Anderson. High-resolution imaging on tpus. *arXiv preprint arXiv:1912.08063*, 2019.

[182] M. Innes, A. Edelman, K. Fischer, C. Rackauckas, E. Saba, V. B. Shah, and W. Tebbutt. A differentiable programming system to bridge machine learning and scientific computing. July 2019.

[183] M. Innes, E. Saba, K. Fischer, D. Gandhi, M. C. Rudilosso, N. M. Joy, T. Karmali, A. Pal, and V. Shah. Fashionable modelling with flux. *CoRR*, 2018.

[184] A. Jain, Y. Shin, and K. A. Persson. Computational predictions of energy materials using density functional theory. *Nature Reviews Materials*, 1(1):1–13, 2016.

[185] S. Jana, K. Sharma, and P. Samal. Improving the performance of Tao–Mo non-empirical density functional with broader applicability in quantum chemistry and materials science. *J. Phys. Chem. A*, 123(29):6356–6369, 2019.

[186] T. Jenkins, K. Berland, and T. Thonhauser. Reduced-gradient analysis of van der waals complexes. *Electronic Structure*, 3(3):034009, 2021.

[187] D. S. Jensen and A. Wasserman. Numerical methods for the inverse problem of density functional theory. *International Journal of Quantum Chemistry*, 118(1):e25425, 2018.

[188] Y. Jin and R. J. Bartlett. Accurate computation of x-ray absorption spectra with ionization potential optimized global hybrid functional. *J. Chem. Phys.*, 149(6):064111, 2018.

[189] G. O. Jones, A. Yuen, R. J. Wojtecki, J. L. Hedrick, and J. M. Garcia. Computational and experimental investigations of one-step conversion of poly (carbonate)s into value-added poly (aryl ether sulfone)s. *Proceedings of the National Academy of Sciences*, 113(28):7722–7726, 2016.

[190] R. O. Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of modern physics*, 87(3):897, 2015.

[191] N. Jouppi, D. Yoon, G. Kurian, S. Li, N. Patil, J. Laudon, C. Young, and D. Patterson. A domain-specific supercomputer for training deep neural networks. *Communications of the ACM*, 63:67–78, 06 2020.

[192] N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers, et al. In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the 44th Annual International Symposium on Computer Architecture*, ISCA ’17, pages 1–12, New York, NY, USA, 2017. Association for Computing Machinery.

[193] S. Kais, D. Herschbach, N. Handy, C. Murray, and G. Laming. Density functionals and dimensional renormalization for an exactly solvable model. *The Journal of chemical physics*, 99(1):417–425, 1993.

[194] B. Kalita, L. Li, R. J. McCarty, and K. Burke. Learning to Approximate Density Functionals. *Accounts of Chemical Research*, 54(4):818–826, feb 2021.

[195] B. Kalita, R. Pederson, J. Chen, L. Li, and K. Burke. How well does kohn–sham regularizer work for weakly correlated systems? *The Journal of Physical Chemistry Letters*, 13(11):2540–2547, 2022.

[196] J. K. Kang and C. B. Musgrave. Prediction of transition state barriers and enthalpies of reaction by a new hybrid density-functional approximation. *J. Chem. Phys.*, 115(24):11040–11051, 2001.

[197] B. Kanungo, P. M. Zimmerman, and V. Gavini. A comparison of exact and model exchange–correlation potentials for molecules. *The Journal of Physical Chemistry Letters*, 12(50):12012–12019, 2021.

[198] A. D. Kaplan, S. J. Clark, K. Burke, and J. P. Perdew. Calculation and interpretation of classical turning surfaces in solids. *npj Computational Materials*, 7(1):1–10, 2021.

[199] A. D. Kaplan, M. Levy, and J. P. Perdew. Predictive power of the exact constraints and appropriate norms in density functional theory. *arXiv preprint arXiv:2207.03855*, 2022.

[200] M. F. Kasim and S. M. Vinko. Learning the exchange-correlation functional from nature with fully differentiable density functional theory. *Phys. Rev. Lett.*, 127:126403, Sep 2021.

[201] T. Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Communications on Pure and Applied Mathematics*, 10(2):151–177, 1957.

[202] J. Katriel and E. R. Davidson. Asymptotic behavior of atomic and molecular wave functions. *Proceedings of the National Academy of Sciences*, 77(8):4403–4406, 1980.

[203] M. Kaupp, M. Renz, M. Parthey, M. Stolte, F. Würthner, and C. Lambert. Computational and spectroscopic studies of organic mixed-valence compounds: where is the charge? *Phys. Chem. Chem. Phys.*, 13:16973–16986, 2011.

[204] T. W. Keal and D. J. Tozer. The exchange-correlation potential in Kohn–Sham nuclear magnetic resonance shielding calculations. *J. Chem. Phys.*, 119(6):3015, 2003.

[205] T. W. Keal and D. J. Tozer. A semiempirical generalized gradient approximation exchange-correlation functional. *J. Chem. Phys.*, 121(12):5654–5660, 2004.

[206] T. W. Keal and D. J. Tozer. Semiempirical hybrid functional with improved performance in an extensive chemical assessment. *J. Chem. Phys.*, 123(12):121103, 2005.

[207] J. Kim and Y. Jung. A perspective on the density matrix purification for linear scaling electronic structure calculations. *Int. J. Quantum Chem.*, 2015.

[208] M.-C. Kim, E. Sim, and K. Burke. Communication: Avoiding unbound anions in density functional calculations. *J. Chem. Phys.*, 134(17):171103, 2011.

[209] M.-C. Kim, E. Sim, and K. Burke. Understanding and reducing errors in density functional calculations. *Phys. Rev. Lett.*, 111:073003, Aug 2013.

[210] Y. Kim, S. Song, E. Sim, and K. Burke. Halogen and chalcogen binding dominated by density-driven errors. *The Journal of Physical Chemistry Letters*, 10(2):295–301, December 2019.

[211] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, and A. J. Cohen. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science*, 374(6573):1385–1389, 2021.

[212] J. Kocák, E. Kraisler, and A. Schild. Charge-transfer steps in density functional theory from the perspective of the exact electron factorization. *The Journal of Physical Chemistry Letters*, 12(12):3204–3209, 2021.

[213] J. Kocák, E. Kraisler, and A. Schild. The geometric potential of the exact electron factorization: Meaning, significance and application. *arXiv:2010.14885*, April 2022.

[214] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, Nov 1965.

[215] R. Kondor, H. T. Son, H. Pan, B. Anderson, and S. Trivedi. Covariant compositional networks for learning graphs. *arXiv preprint arXiv:1801.02144*, 2018.

- [216] M. Korth and S. Grimme. “mindless” dft benchmarking. *Journal of chemical theory and computation*, 5(4):993–1003, 2009.
- [217] M. Krack and A. M. Köster. An adaptive numerical integrator for molecular integrals. *The Journal of chemical physics*, 108(8):3226–3234, 1998.
- [218] E. Kraisler, G. Makov, and I. Kelson. Ensemble v -representable ab initio density-functional calculation of energy and spin in atoms: A test of exchange-correlation approximations. *Phys. Rev. A*, 82:042516, Oct 2010.
- [219] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54(16):11169–11186, Oct 1996.
- [220] J. B. Krieger, J. Chen, G. J. Iafrate, and A. Savin. *Construction of An Accurate Self-interaction-corrected Correlation Energy Functional Based on An Electron Gas with A Gap*, pages 463–477. Springer US, Boston, MA, 1999.
- [221] J. B. Krieger, J. Chen, and S. Kurth. Construction and application of an accurate self-interaction-corrected correlation energy functional based on an electron gas with a gap. *AIP Conf. Proc.*, 577(1):48–69, 2001.
- [222] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [223] P. Küs, A. Marek, S. Köcher, H.-H. Kowalski, C. Carbogno, C. Scheurer, K. Reuter, M. Scheffler, and H. Lederer. Optimizations of the eigensolvers in the elpa library. *Parallel Computing*, 85:167–177, 2019.
- [224] J. Kukačka, V. Golkov, and D. Cremers. Regularization for deep learning: A taxonomy. *arXiv preprint arXiv:1710.10686*, 2017.
- [225] S. Kurth, J. P. Perdew, and P. Blaha. Molecular and solid-state tests of density functional approximations: LSD, GGAs, and meta-GGAs. *Int. J. Quantum Chem.*, 75(4-5):889–909, 1999.
- [226] K. Laidig. Density functional methods of the spatial distribution of electronic charge. *Chem. Phys. Lett.*, 225:285, Apr 1994.
- [227] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*, 45(4):255, oct 1950.
- [228] D. Langreth and J. Perdew. The exchange-correlation energy of a metallic surface. *Solid State Commun.*, 17:1425, 1975.
- [229] D. C. Langreth and M. J. Mehl. Easily implementable nonlocal exchange-correlation energy functional. *Phys. Rev. Lett.*, 47:446, Aug 1981.

[230] J. Larsen. *Foundations of High-Energy-Density Physics: Physical Processes of Matter at Extreme Conditions*. Cambridge University Press, 2017.

[231] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[232] C. Lee and R. G. Parr. Exchange-correlation functional for atoms and molecules. *Phys. Rev. A*, 42:193–200, Jul 1990.

[233] C. Lee, W. Yang, and R. G. Parr. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785, Jan 1988.

[234] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785–789, Jan 1988.

[235] T. H. Lee. Adjoint method for design sensitivity analysis of multiple eigenvalues and associated eigenvectors. *AIAA Journal*, 45(8):1998–2004, 2007.

[236] S. Lehtola and M. A. Marques. Many recent density functionals are numerically unstable. *arXiv preprint arXiv:2206.14062*, 2022.

[237] S. Lehtola and M. A. L. Marques. Meta-local density functionals: A new rung on Jacob’s ladder. *J. Chem. Theory Comput.*, 17(2):943–948, 2021.

[238] S. Lehtola, C. Steigemann, M. J. Oliveira, and M. A. Marques. Recent developments in libxc – a comprehensive library of functionals for density functional theory. *SoftwareX*, 7:1–5, 2018.

[239] G. Lever, D. J. Cole, N. D. Hine, P. D. Haynes, and M. C. Payne. Electrostatic considerations affecting the calculated homo–lumo gap in protein molecules. *Journal of Physics: Condensed Matter*, 25(15):152101, 2013.

[240] M. Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v -representability problem. *Proceedings of the National Academy of Sciences of the United States of America*, 76(12):6062–6065, 1979.

[241] M. Levy. Density-functional exchange correlation through coordinate scaling in adiabatic connection and correlation hole. *Physical Review A*, 43(9):4637, 1991.

[242] M. Levy and J. P. Perdew. Hellmann-feynman, virial, and scaling requisites for the exact universal density functionals. shape of the correlation potential and diamagnetic susceptibility for atoms. *Physical Review A*, 32(4):2010, 1985.

[243] M. Levy and J. P. Perdew. Tight bound and convexity constraint on the exchange-correlation-energy functional in the low-density limit, and other formal tests of generalized-gradient approximations. *Physical Review B*, 48(16):11638, 1993.

[244] M. Levy and J. P. Perdew. Density functionals for exchange and correlation energies: Exact conditions and comparison of approximations. *International journal of quantum chemistry*, 49(4):539–548, 1994.

[245] M. Levy, J. P. Perdew, and V. Sahni. Exact differential equation for the density and ionization energy of a many-particle system. *Phys. Rev. A*, 30:2745–2748, Nov 1984.

[246] M. Lewin, E. H. Lieb, and R. Seiringer. Floating wigner crystal with no boundary charge fluctuations. *Phys. Rev. B*, 100:035127, Jul 2019.

[247] M. Lewin, E. H. Lieb, and R. Seiringer. Improved lieb-oxford bound on the indirect and exchange energies. *arXiv preprint arXiv:2203.12473*, 2022.

[248] A. G. Lewis, J. Beall, M. Ganahl, M. Hauru, S. B. Mallick, and G. Vidal. Large-scale distributed linear algebra with tensor processing units. *Proceedings of the National Academy of Sciences*, 119(33):e2122762119, 2022.

[249] L. Li, T. E. Baker, S. R. White, and K. Burke. Pure density functional for strong correlation and the thermodynamic limit from machine learning. *Phys. Rev. B*, 94:245129, Dec 2016.

[250] L. Li, T. E. Baker, S. R. White, K. Burke, et al. Pure density functional for strong correlation and the thermodynamic limit from machine learning. *Physical Review B*, 94(24):245129, 2016.

[251] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke. Kohn-sham equations as regularizer: Building prior knowledge into machine-learned physics. *Physical review letters*, 126(3):036401, 2021.

[252] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke. Kohn-sham equations as regularizer: Building prior knowledge into machine-learned physics. *Phys. Rev. Lett.*, 126:036401, Jan 2021.

[253] L. Li, J. C. Snyder, I. M. Pelaschier, J. Huang, U.-N. Nirajan, P. Duncan, M. Rupp, K.-R. Müller, and K. Burke. Understanding machine-learned density functionals. *International Journal of Quantum Chemistry*, 116(11):819–833, 2016.

[254] H.-J. Liao, J.-G. Liu, L. Wang, and T. Xiang. Differentiable programming tensor networks. *Physical Review X*, 9(3):031041, 2019.

[255] E. H. Lieb. Density functionals for coulomb systems. *Int. J. Quantum Chem.*, 24(3):243–277, 1983.

[256] E. H. Lieb and S. Oxford. Improved lower bound on the indirect coulomb energy. *International Journal of Quantum Chemistry*, 19(3):427–439, 1981.

[257] C. Y. Lin, M. W. George, and P. M. W. Gill. EDF2: A density functional for predicting molecular vibrational frequencies. *Aust. J. Chem.*, 57(4):365–370, 2004.

[258] D. C. Liu and J. Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[259] L. Lu, H. Hu, H. Hou, and B. Wang. An improved B3LYP method in the calculation of organic thermochemistry and reactivity. *Comput. Theor. Chem.*, 1015:64, 2013.

[260] T. Lu, Y.-F. Chen, B. Hechtman, T. Wang, and J. Anderson. Large-scale discrete fourier transform on tpus. *IEEE Access*, 9:93422–93432, 2021.

[261] T. Lu, T. Marin, Y. Zhuo, Y.-F. Chen, and C. Ma. Accelerating mri reconstruction on tpus. In *2020 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–9. IEEE, 2020.

[262] T. Lu, T. Marin, Y. Zhuo, Y.-F. Chen, and C. Ma. Nonuniform fast fourier transform on tpus. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 783–787. IEEE, 2021.

[263] B. J. Lynch, P. L. Fast, M. Harris, and D. G. Truhlar. Adiabatic connection for kinetics. *J. Phys. Chem. A*, 104(21):4811, 2000.

[264] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[265] N. Maheswaranathan, L. Metz, G. Tucker, D. Choi, and J. Sohl-Dickstein. Guided evolutionary strategies: Augmenting random search with surrogate gradients. In *International Conference on Machine Learning*, pages 4264–4273. PMLR, 2019.

[266] N. Mardirossian and M. Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics*, 115(19):2315–2372, 2017.

[267] A. Marek, V. Blum, R. Johann, V. Havu, B. Lang, T. Auckenthaler, A. Heinecke, H.-J. Bungartz, and H. Lederer. The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science. *Journal of Physics: Condensed Matter*, 26(21):213201, 2014.

[268] J. T. Margraf, C. Kunkel, and K. Reuter. Towards density functional approximations from coupled cluster correlation energy densities. *J. Chem. Phys.*, 150(24):244116, 2019.

[269] A. E. Mattsson, R. Armiento, J. Paier, G. Kresse, J. M. Wills, and T. R. Mattsson. The AM05 density functional applied to solids. *J. Chem. Phys.*, 128(8):084714, 2008.

[270] R. J. McCarty, D. Perchak, R. Pederson, R. Evans, Y. Qiu, S. R. White, and K. Burke. Bypassing the energy functional in density functional theory: Direct calculation of electronic energies from conditional probability densities. *Physical Review Letters*, 125(26):266401, 2020.

[271] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko. Density functional theory is straying from the path toward the exact functional. *Science*, 355(6320):49–52, 2017.

[272] D. Mejia-Rodriguez and S. B. Trickey. Deorbitalization strategies for meta-generalized-gradient-approximation exchange-correlation functionals. *Phys. Rev. A*, 96:052512, Nov 2017.

[273] D. Mejia-Rodriguez and S. B. Trickey. Deorbitalized meta-GGA exchange-correlation functionals in solids. *Phys. Rev. B*, 98:115161, Sep 2018.

[274] D. Mejía-Rodríguez and S. B. Trickey. Meta-GGA performance in solids at almost GGA cost. *Phys. Rev. B*, 102:121109, Sep 2020.

[275] G. Menconi, P. J. Wilson, and D. J. Tozer. Emphasizing the exchange-correlation potential in functional development. *J. Chem. Phys.*, 114(9):3958, 2001.

[276] N. D. Mermin. Thermal properties of the inhomogenous electron gas. *Phys. Rev.*, 137:A: 1441, 1965.

[277] P. D. Mezei, G. I. Csonka, and M. Kállay. Simple modifications of the SCAN meta-generalized gradient approximation functional. *J. Chem. Theory Comput.*, 14(5):2469–2479, 2018.

[278] B. Miehlich, A. Savin, H. Stoll, and H. Preuss. Results obtained with the correlation energy density functionals of becke and lee, yang and parr. *Chemical Physics Letters*, 157(3):200–206, 1989.

[279] B. Miehlich, A. Savin, H. Stoll, and H. Preuss. Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. *Chem. Phys. Lett.*, 157(3):200, 1989.

[280] S. K. Min, A. Abedi, K. S. Kim, and E. K. U. Gross. Is the molecular berry phase an artifact of the born-oppenheimer approximation? *Phys. Rev. Lett.*, 113:263004, Dec 2014.

[281] A. Mirhoseini, A. Goldie, M. Yazgan, J. Jiang, E. Songhori, S. Wang, Y.-J. Lee, E. Johnson, O. Pathak, S. Bae, et al. Chip placement with deep reinforcement learning. *arXiv preprint arXiv:2004.10746*, 2020.

[282] A. Mirtschink, M. Seidl, and P. Gori-Giorgi. Energy densities in the strong-interaction limit of density functional theory. *Journal of chemical theory and computation*, 8(9):3097–3107, 2012.

[283] J. R. Moreno, G. Carleo, and A. Georges. Deep learning the hohenberg-kohn maps of density functional theory. *Physical Review Letters*, 125(7):076402, 2020.

[284] P. Mori-Sánchez, A. J. Cohen, and W. Yang. Discontinuous nature of the exchange-correlation functional in strongly correlated systems. *Phys. Rev. Lett.*, 102:066403, Feb 2009.

[285] A. Morningstar, M. Hauru, J. Beall, M. Ganahl, A. G. Lewis, V. Khemani, and G. Vidal. Simulation of quantum many-body dynamics with tensor processing units: Floquet prethermalization. *PRX Quantum*, 3(2):020331, 2022.

[286] M. Mostafanejad, M. D. Liebenthal, and A. E. DePrince III. Global hybrid multiconfiguration pair-density functional theory. *Journal of chemical theory and computation*, 16(4):2274–2283, 2020.

[287] É. D. Murray, K. Lee, and D. C. Langreth. Investigation of exchange energy density functional accuracy for interacting molecules. *Journal of Chemical Theory and Computation*, 5(10):2754–2762, 2009.

[288] R. Nagai, R. Akashi, S. Sasaki, and S. Tsuneyuki. Neural-network kohn-sham exchange-correlation potential and its out-of-training transferability. *The Journal of chemical physics*, 148(24):241737, 2018.

[289] R. Nagai, R. Akashi, and O. Sugino. Completing density functional theory by machine learning hidden messages from molecules. *npj Computational Materials*, 6(1):1–8, 2020.

[290] R. Nagai, R. Akashi, and O. Sugino. Machine-learning-based exchange correlation functional with physical asymptotic constraints. *Physical Review Research*, 4(1):013106, 2022.

[291] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, ICML’10, page 807–814, Madison, WI, USA, 2010. Omnipress.

[292] R. Neumann and N. C. Handy. Investigations using the Becke–Roussel exchange functional. *Chem. Phys. Lett.*, 246(4):381–386, 1995.

[293] A. M. N. Niklasson. Expansion algorithm for the density matrix. *Phys. Rev. B*, 66(15):155115, 2002.

[294] A. M. N. Niklasson, C. J. Tymczak, and M. Challacombe. Trace resetting density matrix purification in O(N) self-consistent-field theory. *The Journal of Chemical Physics*, 118(19):8611–8620, 2003.

[295] F. Noé, S. Olsson, J. Köhler, and H. Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

[296] J. K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard. Density functional theory in surface chemistry and catalysis. *Proceedings of the National Academy of Sciences*, 108(3):937–943, 2011.

[297] G. Oliver and J. Perdew. Spin-density gradient expansion for the kinetic energy. *Phys. Rev. A*, 20:397, 1979.

[298] A. Overhauser. Pair-correlation function of an electron gas. *Canadian journal of physics*, 73(11-12):683–686, 1995.

[299] A. H. R. Palser and D. E. Manolopoulos. Canonical purification of the density matrix in electronic-structure theory. *Phys. Rev. B*, 58(19):12704–12711, 1998.

[300] Z. Pan and P. Mishra. Hardware acceleration of explainable machine learning using tensor processing units. *arXiv preprint arXiv:2103.11927*, 2021.

[301] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037, 2019.

[302] B. Patra, S. Jana, L. A. Constantin, and P. Samal. Relevance of the Pauli kinetic energy density for semilocal functionals. *Phys. Rev. B*, 100:155140, Oct 2019.

[303] R. Pederson and K. Burke. Reassessing the role of exact conditions in density functional theory. *arXiv preprint arXiv:2303.01766*, 2023.

[304] R. Pederson, J. Chen, L. Li, and R. McCarthy. 1D density functional theory solver. https://github.com/pederson/DFT_1d, 2022.

[305] R. Pederson, J. Chen, S. R. White, K. Burke, et al. Conditional probability density functional theory. *Physical Review B*, 105(24):245138, 2022.

[306] R. Pederson, B. Kalita, and K. Burke. Machine learning and density functional theory. *Nature Reviews Physics*, 4(6):357–358, 2022.

[307] R. Pederson, J. Kozlowski, R. Song, J. Beall, M. Ganahl, M. Hauru, A. G. M. Lewis, Y. Yao, S. B. Mallick, V. Blum, and G. Vidal. Large scale quantum chemistry with tensor processing units. *Journal of Chemical Theory and Computation*, 19(1):25–32, 2023.

[308] L. S. Pedroza, A. J. R. da Silva, and K. Capelle. Gradient-dependent density functionals of the Perdew–Burke–Ernzerhof type for atoms, molecules, and solids. *Phys. Rev. B*, 79:201106, May 2009.

[309] D. Perchak, R. J. McCarty, and K. Burke. Correlation energy of the uniform electron gas determined by ground-state conditional probability density functional theory. *Phys. Rev. B*, 105:165143, Apr 2022.

[310] J. K. Percus. Approximation methods in classical statistical mechanics. *Phys. Rev. Lett.*, 8:462–463, Jun 1962.

[311] J. P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B*, 33:8822, Jun 1986.

[312] J. P. Perdew. In P. Ziesche and H. Eschrig, editors, *Proceedings of the 75. WE-Heraeus-Seminar and 21st Annual International Symposium on Electronic Structure of Solids*, page 11, Berlin, 1991. Akademie Verlag.

- [313] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, Oct 1996. *ibid.* **78**, 1396(E) (1997).
- [314] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, Oct 1996.
- [315] J. P. Perdew, K. Burke, and M. Ernzerhof. Errata: Generalized gradient approximation made simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.*, 78:1396, Feb 1997.
- [316] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46:6671, Sep 1992.
- [317] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 48:4978, Aug 1993.
- [318] J. P. Perdew, M. Ernzerhof, and K. Burke. Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics*, 105(22):9982–9985, 1996.
- [319] J. P. Perdew, M. Ernzerhof, K. Burke, and A. Savin. On-top pair-density interpretation of spin density functional theory, with applications to magnetism. *International journal of quantum chemistry*, 61(2):197–205, 1997.
- [320] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Phys. Rev. Lett.*, 82:2544, Mar 1999.
- [321] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun. Workhorse semilocal density functional for condensed matter physics and quantum chemistry. *Phys. Rev. Lett.*, 103:026403, Jul 2009.
- [322] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun. Erratum: Workhorse semilocal density functional for condensed matter physics and quantum chemistry [Phys. Rev. Lett. 103, 026403 (2009)]. *Phys. Rev. Lett.*, 106:179902, Apr 2011.
- [323] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100:136406, Apr 2008.
- [324] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100:136406, Apr 2008.

[325] J. P. Perdew, A. Ruzsinszky, J. Sun, and K. Burke. Gedanken densities and exact constraints in density functional theory. *The Journal of Chemical Physics*, 140(18):–, 2014.

[326] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *The Journal of Chemical Physics*, 123(6):062201, 2005.

[327] J. P. Perdew, A. Savin, and K. Burke. Escaping the symmetry dilemma through a pair-density interpretation of spin-density functional theory. *Phys. Rev. A*, 51:4531–4541, Jun 1995.

[328] J. P. Perdew, A. Savin, and K. Burke. Escaping the symmetry dilemma through a pair-density interpretation of spin-density functional theory. *Physical Review A*, 51(6):4531, 1995.

[329] J. P. Perdew and J. Sun. The lieb-oxford lower bounds on the coulomb energy, their importance to electron density functional theory, and a conjectured tight bound on exchange. *arXiv preprint arXiv:2206.09974*, 2022.

[330] J. P. Perdew, J. Tao, V. N. Staroverov, and G. E. Scuseria. Meta-generalized gradient approximation: Explanation of a realistic nonempirical density functional. *J. Chem. Phys.*, 120(15):6898, 2004.

[331] J. P. Perdew and Y. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45(23):13244–13249, Jun 1992.

[332] J. P. Perdew and Y. Wang. Pair-distribution function and its coupling-constant average for the spin-polarized electron gas. *Phys. Rev. B*, 46:12947–12954, Nov 1992.

[333] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048–5079, May 1981.

[334] K. Pernal, R. Podeszwa, K. Patkowski, and K. Szalewicz. Dispersionless density functional theory. *Phys. Rev. Lett.*, 103:263201, Dec 2009.

[335] F. Perrot. Gradient correction to the statistical electronic free energy at nonzero temperatures: Application to equation-of-state calculations. *Phys. Rev. A*, 20:586–594, Aug 1979.

[336] K. B. Petersen and M. S. Pedersen. The matrix cookbook, nov 2012. Version 20121115.

[337] R. Peverati and D. G. Truhlar. Communication: A global hybrid generalized gradient approximation to the exchange-correlation functional that satisfies the second-order density-gradient constraint and has broad applicability in chemistry. *J. Chem. Phys.*, 135(19):191102, 2011.

[338] R. Peverati and D. G. Truhlar. Exchange-correlation functional with good accuracy for both structural and energetic properties while depending only on the density and its gradient. *J. Chem. Theory Comput.*, 8(7):2310, 2012.

[339] R. Peverati and D. G. Truhlar. An improved and broadly accurate local approximation to the exchange-correlation density functional: The MN12-L functional for electronic structure calculations in chemistry and physics. *Phys. Chem. Chem. Phys.*, 14:13171, 2012.

[340] R. Peverati and D. G. Truhlar. M11-L: A local density functional that provides improved accuracy for electronic structure calculations in chemistry and physics. *J. Phys. Chem. Lett.*, 3(1):117, 2012.

[341] R. Peverati and D. G. Truhlar. Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2011):20120476, 2014.

[342] R. Peverati, Y. Zhao, and D. G. Truhlar. Generalized gradient approximation that recovers the second-order density-gradient expansion with optimized across-the-board performance. *J. Phys. Chem. Lett.*, 2(16):1991, 2011.

[343] C. J. Pickard, I. Errea, and M. I. Eremets. Superconducting hydrides under pressure. *Annual Review of Condensed Matter Physics*, 11:57–76, 2020.

[344] S. Pittalis, C. R. Proetto, A. Floris, A. Sanna, C. Bersier, K. Burke, and E. K. U. Gross. Exact conditions in finite-temperature density-functional theory. *Phys. Rev. Lett.*, 107:163001, Oct 2011.

[345] J. C. A. Prentice, J. Aarons, J. C. Womack, A. E. A. Allen, L. Andrinopoulos, L. Anton, R. A. Bell, A. Bhandari, G. A. Bramley, R. J. Charlton, R. J. Clements, D. J. Cole, G. Constantinescu, F. Corsetti, S. M.-M. Dubois, K. K. B. Duff, J. M. Escartín, A. Greco, Q. Hill, L. P. Lee, E. Linscott, D. D. O'Regan, M. J. S. Phipps, L. E. Ratcliff, A. R. Serrano, E. W. Tait, G. Teobaldi, V. Vitale, N. Yeung, T. J. Zuehlsdorff, J. Dziedzic, P. D. Haynes, N. D. M. Hine, A. A. Mostofi, M. C. Payne, and C.-K. Skylaris. The onetep linear-scaling density functional theory program. *The Journal of Chemical Physics*, 152(17):174111, 2020.

[346] E. I. Proynov and A. J. Thakkar. Is combining meta-GGA correlation functionals with the OPTX exchange functional useful? *Int. J. Quantum Chem.*, 106(2):436–446, 2006.

[347] Y. Qiu and S. R. White. Hybrid gausslet/gaussian basis sets. *The Journal of Chemical Physics*, 155(18):184107, 2021.

[348] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

- [349] P. Ramachandran, B. Zoph, and Q. V. Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017.
- [350] P. Ramachandran, B. Zoph, and Q. V. Le. Searching for activation functions, 2018.
- [351] M. Reiher, O. Salomon, and B. A. Hess. Reparameterization of hybrid functionals based on energy differences of states of different multiplicity. *Theor. Chem. Acc.*, 107(1):48–55, 2001.
- [352] M. Renz, K. Theilacker, C. Lambert, and M. Kaupp. A reliable quantum-chemical protocol for the characterization of organic mixed-valence compounds. *J. Am. Chem. Soc.*, 131(44):16292–16302, 2009.
- [353] R. Requist and E. K. U. Gross. Fock-space embedding theory: Application to strongly correlated topological phases. *Physical Review Letters*, 127(11):116401, 2021.
- [354] J. Rey and A. Savin. Virtual space level shifting and correlation energies. *Int. J. Quantum Chem.*, 69(4):581–590, 1998.
- [355] L. M. Rios and N. V. Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293, 2013.
- [356] S. Ross and D. Bagnell. Efficient reductions for imitation learning. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 661–668, 2010.
- [357] E. Rudberg. Difficulties in applying pure Kohn–Sham density functional theory electronic structure methods to protein molecules. *Journal of Physics: Condensed Matter*, 24(7):072202, 2012.
- [358] E. Rudberg, E. H. Rubensson, P. Salek, and A. Kruchinina. Ergo: An open-source program for linear-scaling electronic structure calculations. *SoftwareX*, 7:107–111, 2018.
- [359] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [360] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters*, 108(5):058301, 2012.
- [361] A. Ruzsinszky, G. I. Csonka, and G. E. Scuseria. Regularized gradient expansion for atoms, molecules, and solids. *J. Chem. Theory Comput.*, 5(4):763, 2009.
- [362] R. Sarmiento-Pérez, S. Botti, and M. A. L. Marques. Optimized exchange and correlation semilocal functional for the calculation of energies of formation. *J. Chem. Theory Comput.*, 11(8):3844–3850, 2015.

- [363] A. Schild and E. K. U. Gross. Exact single-electron approach to the dynamics of molecules in strong laser fields. *Phys. Rev. Lett.*, 118:163202, Apr 2017.
- [364] H. L. Schmider and A. D. Becke. Optimized density functionals from the extended G2 test set. *J. Chem. Phys.*, 108(23):9624, 1998.
- [365] J. Schmidt, C. L. Benavides-Riveros, and M. A. Marques. Machine learning the physical nonlocal exchange–correlation functional of density-functional theory. *The journal of physical chemistry letters*, 10(20):6425–6431, 2019.
- [366] S. S. Schoenholz and E. D. Cubuk. JAX M.D.: End-to-end differentiable, hardware accelerated, molecular dynamics in pure python. <https://github.com/google/jax-md>, <https://arxiv.org/abs/1912.04232>, 2019.
- [367] N. E. Schultz, Y. Zhao, and D. G. Truhlar. Density functionals for inorganometallic and organometallic chemistry. *J. Phys. Chem. A*, 109(49):11127, 2005.
- [368] K. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature communications*, 10(1):1–10, 2019.
- [369] K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in neural information processing systems*, pages 991–1001, 2017.
- [370] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy. Generalized kohn-sham schemes and the band-gap problem. *Physical Review B*, 53(7):3764, 1996.
- [371] M. Seidl, S. Di Marino, A. Gerolin, L. Nenna, K. J. Giesbertz, and P. Gori-Giorgi. The strictly-correlated electron functional for spherically symmetric systems revisited. *arXiv preprint arXiv:1702.05022*, 2017.
- [372] M. Seidl, P. Gori-Giorgi, and A. Savin. Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities. *Physical Review A*, 75(4):042511, 2007.
- [373] S. Seo and Y. Liu. Differentiable physics-informed graph networks. *arXiv preprint arXiv:1902.02950*, 2019.
- [374] S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. L. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder Jr., C. Song, A. V. Titov, I. S. Ufimtsev, L.-P. Wang, and T. J. Martínez. Terachem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *WIREs Computational Molecular Science*, 11(2):e1494, 2021.
- [375] Y. Shao, M. Head-Gordon, and A. I. Krylov. The spin–flip approach within time-dependent density functional theory: Theory and applications to diradicals. *J. Chem. Phys.*, 118(11):4807–4818, 2003.

[376] R. Sharma, A. B. Farimani, J. Gomes, P. Eastman, and V. Pande. Weakly-supervised deep learning of heat transport via physics informed loss. *arXiv preprint arXiv:1807.11374*, 2018.

[377] J. C. Smith, A. Pribam-Jones, and K. Burke. Exact thermal density functional theory for a model system: Correlation components and accuracy of the zero-temperature exchange-correlation approximation. *Phys. Rev. B*, 93:245131, Jun 2016.

[378] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke. Finding density functionals with machine learning. *Physical review letters*, 108(25):253002, 2012.

[379] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke. Finding density functionals with machine learning. *Phys. Rev. Lett.*, 108:253002, Jun 2012.

[380] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal. The SIESTA method for ab initio order-n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.

[381] D. Spaangberg. Ergoscf xyz - water clusters. <http://www.ergoscf.org/xyz/h2o.php>. Accessed: 11-20-2022.

[382] Z. M. Sparrow, B. G. Ernst, T. K. Quady, and R. A. DiStasio. Uniting nonempirical and empirical density functional approximation strategies using constraint-based regularization. *J. Phys. Chem. Lett.*, 13(30):6896–6904, 2022.

[383] Z. M. Sparrow, B. G. Ernst, T. K. Quady, and R. A. DiStasio Jr. Case21: Uniting non-empirical and semi-empirical density functional approximation strategies using constraint-based regularization. *arXiv preprint arXiv:2109.12560*, 2021.

[384] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.*, 119(23):12129, 2003.

[385] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of physical chemistry*, 98(45):11623–11627, 1994.

[386] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98(45):11623, 1994.

[387] E. M. Stoudenmire, L. O. Wagner, S. R. White, and K. Burke. One-dimensional continuum electronic structure with the density-matrix renormalization group and its implications for density-functional theory. *Phys. Rev. Lett.*, 109:056402, Aug 2012.

[388] J. Sun, A. Ruzsinszky, and J. P. Perdew. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.*, 115:036402, Jul 2015.

[389] J. Sun, A. Ruzsinszky, and J. P. Perdew. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.*, 115:036402, Jul 2015.

[390] Q. Sun and G. K.-L. Chan. Quantum embedding theories. *Accounts of chemical research*, 49(12):2705–2712, 2016.

[391] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, et al. Recent developments in the pyscf program package. *The Journal of chemical physics*, 153(2):024109, 2020.

[392] M. Swart, M. Solá, and F. M. Bickelhaupt. A new all-round density functional based on spin states and s_n2 barriers. *J. Chem. Phys.*, 131(9):094103, 2009.

[393] D. N. Tahchieva, D. Bakowies, R. Ramakrishnan, and O. A. von Lilienfeld. Torsional potentials of glyoxal, oxaryl halides, and their thiocarbonyl derivatives: Challenges for popular density functional approximations. *J. Chem. Theory Comput.*, 14(9):4806–4817, 2018.

[394] L. Talirz, L. M. Ghiringhelli, and B. Smit. Trends in atomistic simulation software usage. *Living Journal of Computational Molecular Science*, 3(1):1–12, 2021.

[395] T. Tamayo-Mendoza, C. Kreisbeck, R. Lindh, and A. Aspuru-Guzik. Automatic differentiation in quantum chemistry with applications to fully variational Hartree-Fock. *ACS Cent Sci*, 4(5):559–566, May 2018.

[396] J. Tao. An accurate MGGA-based hybrid exchange-correlation functional. *J. Chem. Phys.*, 116(6):2335–2337, 2002.

[397] J. Tao and Y. Mo. Accurate semilocal density functional for condensed-matter physics and quantum chemistry. *Phys. Rev. Lett.*, 117:073001, Aug 2016.

[398] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, 91:146401, Sep 2003.

[399] A. J. Thakkar and S. P. McCarthy. Toward improved density functionals for the correlation energy. *J. Chem. Phys.*, 131(13):134109, 2009.

[400] S. Theodoridis. *Machine learning: a Bayesian and optimization perspective*. Academic press, 2015.

[401] L. H. Thomas. The calculation of atomic fields. *Math. Proc. Camb. Phil. Soc.*, 23(05):542–548, 1927.

[402] L. H. Thomas. The calculation of atomic fields. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 23, pages 542–548. Cambridge University Press, 1927.

[403] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

[404] V. Tognetti, P. Cortona, and C. Adamo. The performances of a parameter-free local correlation functional: The Ragot–Cortona model. *Chem. Phys. Lett.*, 439(4–6):381, 2007.

[405] V. Tognetti, P. Cortona, and C. Adamo. Increasing physical constraints and improving performances in a parameter-free GGA functional. *Chem. Phys. Lett.*, 460(4–6):536, 2008.

[406] V. Tognetti, P. Cortona, and C. Adamo. A new parameter-free correlation functional based on an average atomic reduced density gradient analysis. *J. Chem. Phys.*, 128(3):034101, 2008.

[407] J. Toulouse, A. Savin, and C. Adamo. Validation and assessment of an accurate approach to the correlation problem in density functional theory: The Kriger–Chen–Iafrate–Savin model. *J. Chem. Phys.*, 117(23):10465–10473, 2002.

[408] D. J. Tozer and N. C. Handy. The development of new exchange-correlation functionals. *J. Chem. Phys.*, 108(6):2545, 1998.

[409] D. J. Tozer and N. C. Handy. Development of new exchange-correlation functionals. 2. *J. Phys. Chem. A*, 102(18):3162, 1998.

[410] D. J. Tozer, N. C. Handy, and W. H. Green. Exchange-correlation functionals from ab initio electron densities. *Chem. Phys. Lett.*, 273(3–4):183, 1997.

[411] D. J. Tozer, V. E. Ingamells, and N. C. Handy. Exchange-correlation potentials. *The Journal of Chemical Physics*, 105(20):9200–9213, 1996.

[412] L. A. Truflandier, R. M. Dianzinga, and D. R. Bowler. Communication: Generalized canonical purification for density matrix minimization. *The Journal of Chemical Physics*, 144(9):091102, 2016.

[413] T. Tsuneda, T. Suzumura, and K. Hirao. A new one-parameter progressive Colle–Salvetti-type correlation functional. *J. Chem. Phys.*, 110(22):10664, 1999.

[414] T. Tsuneda, T. Suzumura, and K. Hirao. A reexamination of exchange energy functionals. *J. Chem. Phys.*, 111(13):5656–5667, 1999.

[415] G. Tucker, A. Mnih, C. J. Maddison, J. Lawson, and J. Sohl-Dickstein. Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models. In *Advances in Neural Information Processing Systems*, pages 2627–2636, 2017.

[416] A. Urban, D.-H. Seo, and G. Ceder. Computational understanding of li-ion batteries. *npj Computational Materials*, 2(1):1–13, 2016.

[417] R. A. Van De Geijn and J. Watts. SUMMA: scalable universal matrix multiplication algorithm. *Concurrency: Practice and Experience*, 9(4):255–274, 1997.

[418] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Computer Physics Communications*, 167(2):103–128, 2005.

[419] P. Verma and D. G. Truhlar. HLE16: A local Kohn–Sham gradient approximation with good performance for semiconductor band gaps and molecular excitation energies. *J. Phys. Chem. Lett.*, 8(2):380–387, 2017.

[420] P. Verma and D. G. Truhlar. HLE17: An improved local exchange–correlation functional for computing semiconductor band gaps and molecular excitation energies. *J. Phys. Chem. C*, 121(13):7144–7154, 2017.

[421] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . . Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[422] L. Vogt, R. Olivares-Amaya, S. Kermes, Y. Shao, C. Amador-Bedolla, and A. Aspuru-Guzik. Accelerating resolution-of-the-identity second-order Møller-Plesset quantum chemistry calculations with graphical processing units. *The Journal of Physical Chemistry A*, 112(10):2049–2057, 2008.

[423] U. von Barth and L. Hedin. A local exchange–correlation potential for the spin polarized case. *Journal of Physics C: Solid State Physics*, 5(13):1629, 1972.

[424] T. V. Voorhis and G. E. Scuseria. A novel form for the exchange–correlation energy functional. *J. Chem. Phys.*, 109(2):400, 1998.

[425] S. H. Vosko and J. P. Perdew. Theory of the spin susceptibility of an inhomogeneous electron gas via the density functional formalism. *Can. J. of Phys.*, 53:1385, 1975.

[426] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211, 1980.

[427] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of physics*, 58(8):1200–1211, 1980.

[428] S. Vuckovic, S. Song, J. Kozlowski, E. Sim, and K. Burke. Density functional analysis: The theory of density-corrected dft. *Journal of Chemical Theory and Computation*, 15(12):6636–6646, November 2019. PMID: 31682433.

[429] L. O. Wagner, T. E. Baker, M. Stoudenmire, E., K. Burke, and S. R. White. Kohn-sham calculations with the exact functional. *Phys. Rev. B*, 90:045109, Jul 2014.

[430] L. O. Wagner, E. Stoudenmire, K. Burke, and S. R. White. Reference electronic structure calculations in one dimension. *Phys. Chem. Chem. Phys.*, 14(24):8581–8590, 2012.

[431] L. O. Wagner, E. M. Stoudenmire, K. Burke, and S. R. White. Guaranteed convergence of the kohn-sham equations. *Physical review letters*, 111(9):093003, 2013.

[432] Q. Wang, M. Ihme, Y.-F. Chen, and J. Anderson. A tensorflow simulation framework for scientific computing of fluid flows on tensor processing units. *Computer Physics Communications*, page 108292, 2022.

[433] Y. Wang, X. Jin, H. S. Yu, D. G. Truhlar, and X. He. Revised M06-L functional for improved accuracy on chemical reaction barrier heights, noncovalent interactions, and solid-state physics. *Proc. Natl. Acad. Sci. U. S. A.*, 114(32):8487–8492, 2017.

[434] Y. Wang, P. Verma, X. Jin, D. G. Truhlar, and X. He. Revised M06 density functional for main-group and transition-metal chemistry. *Proc. Natl. Acad. Sci. U. S. A.*, 115(41):10257–10262, 2018.

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

[436] A. J. White and L. A. Collins. Fast and universal kohn-sham density functional theory algorithm for warm dense matter to hot dense plasma. *Phys. Rev. Lett.*, 125:055002, Jul 2020.

[437] C. A. White, B. G. Johnson, P. M. Gill, and M. Head-Gordon. Linear scaling density functional calculations via the continuous fast multipole method. *Chemical Physics Letters*, 253(3-4):268–278, 1996.

[438] S. R. White. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.*, 69(19):2863–2866, Nov 1992.

[439] S. R. White. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B*, 48:10345–10356, Oct 1993.

[440] S. R. White. Hybrid grid/basis set discretizations of the schrödinger equation. *The Journal of Chemical Physics*, 147(24):244102, 2017.

[441] S. R. White and E. M. Stoudenmire. Multisliced gausslet basis sets for electronic structure. *Phys. Rev. B*, 99:081110, Feb 2019.

[442] K. W. Wiitala, T. R. Hoye, and C. J. Cramer. Hybrid density functional methods empirically optimized for the computation of ^{13}C and ^1H chemical shifts in chloroform solution. *J. Chem. Theory Comput.*, 2(4):1085–1092, 2006.

[443] L. C. Wilson. Development of non-local Wigner-like correlation-energy density functionals through coordinate scaling requirements and evaluation of constraint obedience in the construction of density functionals. *Chem. Phys.*, 181(3):337 – 353, 1994.

[444] L. C. Wilson and S. Ivanov. A new Wigner-like correlation-energy functional from coordinate scaling requirements. *Int. J. Quantum Chem.*, 69(4):523, 1998.

[445] L. C. Wilson and M. Levy. Nonlocal Wigner-like correlation-energy density functional through coordinate scaling. *Phys. Rev. B*, 41:12930, Jun 1990.

[446] P. J. Wilson, T. J. Bradley, and D. J. Tozer. Hybrid exchange-correlation functional determined from thermochemical data and ab initio potentials. *J. Chem. Phys.*, 115(20):9233, 2001.

[447] H. B. Wu and X. W. D. Lou. Metal-organic frameworks and their derived materials for electrochemical energy storage and conversion: Promises and challenges. *Science Advances*, 3(12):eaap9252, 2017.

[448] H. Xie, J.-G. Liu, and L. Wang. Automatic differentiation of dominant eigensolver and its applications in quantum physics. *Physical Review B*, 101(24):245139, 2020.

[449] X. Xu and W. A. Goddard. The extended Perdew–Burke–Ernzerhof functional with improved accuracy for thermodynamic and electronic properties of molecular systems. *J. Chem. Phys.*, 121(9):4068, 2004.

[450] X. Xu and W. A. Goddard. The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proc. Natl. Acad. Sci. U. S. A.*, 101(9):2673, 2004.

[451] L. Yang, Z. Leng, G. Yu, A. Patel, W.-J. Hu, and H. Pu. Deep learning-enhanced variational monte carlo method for quantum many-body physics. *Phys. Rev. Research*, 2:012039, Feb 2020.

[452] H. S. Yu, X. He, S. L. Li, and D. G. Truhlar. 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.*, 7:5032–5051, 2016.

[453] H. S. Yu, X. He, and D. G. Truhlar. 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.*, 12(3):1280–1293, 2016.

[454] H. S. Yu, W. Zhang, P. Verma, X. He, and D. G. Truhlar. Nonseparable exchange-correlation functional for molecules, including homogeneous catalysis involving transition metals. *Phys. Chem. Chem. Phys.*, 17:12146–12160, 2015.

[455] J. Yu, L.-H. Xie, J.-R. Li, Y. Ma, J. M. Seminario, and P. B. Balbuena. Co₂ capture and separations using MOFs: computational and experimental studies. *Chemical reviews*, 117(14):9674–9754, 2017.

[456] V. W.-z. Yu, C. Campos, W. Dawson, A. Garcia, V. Havu, B. Hourahine, W. P. Huhn, M. Jacquelin, W. Jia, M. Keceli, et al. Elsi—an open infrastructure for electronic structure solvers. *Computer Physics Communications*, 256:107459, 2020.

[457] L. Zeng, S. B. Jacobsen, D. D. Sasselov, M. I. Petaev, A. Vanderburg, M. Lopez-Morales, J. Perez-Mercader, T. R. Mattsson, G. Li, M. Z. Heising, et al. Growth model interpretation of planet size distribution. *Proceedings of the National Academy of Sciences*, 116(20):9723–9728, 2019.

[458] Q. Zhao, M. Levy, and R. G. Parr. Applications of coordinate-scaling procedures to the exchange-correlation energy. *Phys. Rev. A*, 47:918–922, Feb 1993.

[459] Y. Zhao, N. González-García, and D. G. Truhlar. Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods. *J. Phys. Chem. A*, 109(9):2012–2018, 2005.

[460] Y. Zhao, B. J. Lynch, and D. G. Truhlar. Development and assessment of a new hybrid density functional model for thermochemical kinetics. *J. Phys. Chem. A*, 108(14):2715, 2004.

[461] Y. Zhao, B. J. Lynch, and D. G. Truhlar. Multi-coefficient extrapolated density functional theory for thermochemistry and thermochemical kinetics. *Phys. Chem. Chem. Phys.*, 7:43–52, 2005.

[462] Y. Zhao, N. E. Schultz, and D. G. Truhlar. Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. *J. Chem. Phys.*, 123(16):161103, 2005.

[463] Y. Zhao, N. E. Schultz, and D. G. Truhlar. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theory Comput.*, 2(2):364, 2006.

[464] Y. Zhao and D. G. Truhlar. Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: The mPW1B95 and mPWB1K models and comparative assessments for hydrogen bonding and van der waals interactions. *J. Phys. Chem. A*, 108(33):6908, 2004.

[465] Y. Zhao and D. G. Truhlar. Benchmark databases for nonbonded interactions and their use to test density functional theory. *J. Chem. Theory Comput.*, 1(3):415–432, 2005.

[466] Y. Zhao and D. G. Truhlar. Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions. *J. Phys. Chem. A*, 109(25):5656, 2005.

[467] Y. Zhao and D. G. Truhlar. Density functional for spectroscopy: No long-range self-interaction error, good performance for Rydberg and charge-transfer states, and better performance on average than B3LYP for ground states. *J. Phys. Chem. A*, 110(49):13126, 2006.

[468] Y. Zhao and D. G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *The Journal of Chemical Physics*, 125(19):194101, 2006.

[469] Y. Zhao and D. G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.*, 125(19):194101, 2006.

[470] Y. Zhao and D. G. Truhlar. Exploring the limit of accuracy of the global hybrid meta density functional for main-group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.*, 4(11):1849, 2008.

[471] Y. Zhao and D. G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, 120(1-3):215, 2008.

[472] V. W. zhe Yu, F. Corsetti, A. García, W. P. Huhn, M. Jacquelin, W. Jia, B. Lange, L. Lin, J. Lu, W. Mi, A. Seifitokaldani, Álvaro Vázquez-Mayagoitia, C. Yang, H. Yang, and V. Blum. Elsi: A unified software interface for kohn–sham electronic structure solvers. *Computer Physics Communications*, 222:267–285, 2018.

[473] J. Zheng, Y. Zhao, and D. G. Truhlar. The DBH24/08 database and its use to assess electronic structure model chemistries for chemical reaction barrier heights. *J. Chem. Theory Comput.*, 5(4):808, 2009.

[474] Y. Zhou, J. Wu, S. Chen, and G. Chen. Toward the exact exchange–correlation potential: A three-dimensional convolutional neural network construct. *The journal of physical chemistry letters*, 10(22):7264–7269, 2019.

[475] A. Zupan, K. Burke, M. Ernzerhof, and J. P. Perdew. Distributions and averages of electron density parameters: Explaining the effects of gradient corrections. *The Journal of chemical physics*, 106(24):10184–10193, 1997.

[476] A. Zupan, J. P. Perdew, K. Burke, and M. Causà. Density-gradient analysis for density functional theory: Application to atoms. *International journal of quantum chemistry*, 61(5):835–845, 1997.

Appendix A

Supplemental Info for Chapter 4

This chapter is a reproduction of the supplemental info in Ref. [303].

A.1 Gedanken density details

The Gedanken density used in the main text has the radial form

$$n(r) \propto \begin{cases} C - \frac{A}{\pi} \cos^{-1} \left((1 - \eta) \sin \left(\frac{2\pi r}{T} - \frac{\pi}{2} \right) \right) & r \leq r_p \\ e^{-ar^2 + br + c} & r > r_p \end{cases} \quad (\text{A.1})$$

where the density is a damped triangle wave starting from the origin where $\eta \geq 0$ controls the smoothness of the waveform (where $\eta = 0$ produces a triangle wave), $T > 0$ is the period of the waveform, $N_p \geq 0$ is an integer that controls the number of peaks in the waveform, $A > 0$ is the amplitude, and $C > 0$ is an offset. At $r_p = (N_p - 3/4)T$, the density transitions to a Gaussian, where a , b , and c are determined by ensuring continuity and first and second derivative continuity. The final gedanken density is then normalized to the desired number

of electrons (2 here). Specific variable values used for the gedanken density in the main text are provided in Table A.1. In Fig. A.1 we plot this gedanken density and its corresponding Kohn-Sham (KS) potential,

$$v_s(r) = \frac{1}{2} \frac{\frac{d^2}{dr^2}(r\sqrt{n(r)})}{r\sqrt{n(r)}}. \quad (\text{A.2})$$

C	0.2387324146
A	0.1679968844
η	0.05
T	0.3105085788
N_p	5
a	22.0154308155
b	51.4622187780
c	2.2×10^{-14}

Table A.1: Variable values used in the example gedanken density.

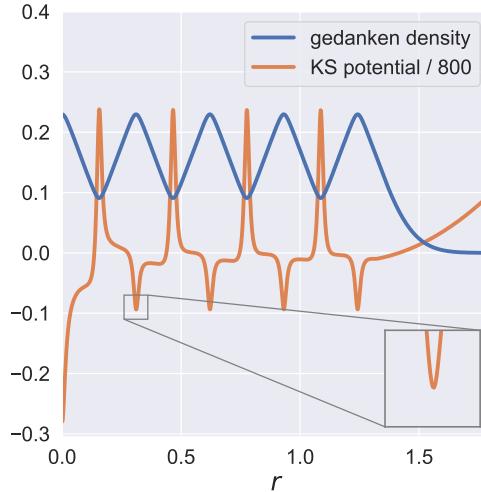


Figure A.1: The example Gedanken density considered in the main text and its corresponding KS potential. The potential is continuous everywhere.

In Table A.2, the exchange energy for the gedanken density is given for different exchange approximations, including B88 exchange [36], which was explicitly designed to the correct large r behavior for Coulombic systems. Two-electron densities, such as the gedanken density, follow a tight Lieb-Oxford (LO) bound for exchange, $E_X[n] \geq 1.174 E_X^{\text{LDA}}[n]$ [325]. For

the gedanken density, $1.174 E_x^{\text{LDA}} = -0.925$ and we see that PBE and B88 exchange violate this exact condition. On the other hand, SCAN explicitly enforces this tight bound condition and does not violate it.

exact	SCAN	LDA	PBE	B88
-0.832	-0.898	-0.788	-1.062	-1.136

Table A.2: Exchange energies (in atomic units) computed on the two-electron Gedanken density.

A.2 computational details

All atomic calculations were performed using the PySCF code [391] with the aug-ccpvdz basis set. The distribution $g(s)$ was computed using a fermi distribution smoothening with fictitious temperature $T = 0.05$ (following Ref. [476]). For some systems, increased radial grids (up to 500) were used in the Gauss-Chebyshev grid scheme [217] to maintain high fidelity of $g(s)$ at large s values. Further details can be found in our public code [4].

A.3 Local conditions derivations and relations

We derive several of the local conditions used in the main text which have not been previously reported in the literature to our knowledge. We start with the *T_c upper bound exact condition* [244]

$$T_c[n_\gamma] \leq -\gamma \left(\frac{\partial E_c[n_\gamma]}{\partial \gamma} \Big|_{\gamma \rightarrow 0} \right) + E_c[n_\gamma], \quad (\text{A.3})$$

where [128]

$$T_C[n_\gamma] = \gamma \frac{\partial E_C[n_\gamma]}{\partial \gamma} - E_C[n_\gamma]. \quad (\text{A.4})$$

For the semilocal approximations considered here,

$$\tilde{E}_C[n_\gamma] = \gamma \int d^3r n(\mathbf{r}) \epsilon_x^{\text{unif}}[n](\mathbf{r}) \tilde{F}_C(r_s/\gamma, \zeta, s, \alpha, q). \quad (\text{A.5})$$

This is due to the fact that the dimensionless quantities ζ , s , α , and q are all scale-invariant, i.e. $\zeta[n_\gamma](\mathbf{r}) = \zeta[n](\gamma\mathbf{r})$, $s[n_\gamma](\mathbf{r}) = s[n](\gamma\mathbf{r})$, and so on. Thus the γ -dependence in the approximation only shows up in the local quantity, $r_s[n_\gamma](\mathbf{r}) = r_s[n](\gamma\mathbf{r})/\gamma$. For compactness, in the following we drop the explicit dependence on scale-invariant quantities. Substituting in Eq. (A.5) into Eq. (A.3) and enforcing the inequality on the integrands (that is, the local enforcement the exact condition) yields

$$\gamma \frac{\partial \tilde{F}_C(r_s/\gamma)}{\partial \gamma} \geq -\tilde{F}_C(\infty) + \tilde{F}_C(r_s/\gamma). \quad (\text{A.6})$$

Using the chain rule and rearranging we obtain the following local condition

$$\frac{\partial \tilde{F}_C}{\partial r_s} \leq \frac{\tilde{F}_C(\infty) - \tilde{F}_C}{r_s}. \quad (\text{A.7})$$

The limit $\tilde{F}_C(\infty)$ is the *strictly correlated electron* [372] limit, which is discussed further in Section A.7.

The LO bound [256] exact condition reads

$$U_{\text{xc}}[n] \geq C_{\text{LO}} \int d^3r n \epsilon_x^{\text{unif}}[n](\mathbf{r}), \quad (\text{A.8})$$

where $U_{\text{xc}}[n] = E_{\text{xc}}[n] - T_C[n]$. For the semilocal approximations considered here, utilizing

Eq. (A.4) evaluated at $\gamma = 1$ we have

$$\tilde{U}_{\text{xc}}[n] = \int d^3r n(\mathbf{r}) \epsilon_{\text{x}}^{\text{unif}}[n](\mathbf{r}) \left[\tilde{F}_{\text{xc}} + r_s \frac{\partial \tilde{F}_{\text{c}}}{\partial r_s} \right]. \quad (\text{A.9})$$

Enforcing the LO bound locally yields the following local condition,

$$\left(1 + r_s \frac{\partial}{\partial r_s} \right) \tilde{F}_{\text{xc}} \leq C_{\text{LO}}. \quad (\text{A.10})$$

It is also known that $T_{\text{c}}[n] \geq 0$, and Eq. (A.4) yields the following condition

$$\gamma \frac{\partial E_{\text{c}}[n_{\gamma}]}{\partial \gamma} - E_{\text{c}}[n_{\gamma}] \geq 0. \quad (\text{A.11})$$

However one can show that this condition is automatically satisfied under the *correlation uniform scaling inequality* exact condition [242],

$$\begin{aligned} E_{\text{c}}[n_{\gamma}] &\geq \gamma E_{\text{c}}[n] \quad (\gamma > 1), \\ E_{\text{c}}[n_{\gamma}] &\leq \gamma E_{\text{c}}[n] \quad (\gamma < 1). \end{aligned} \quad (\text{A.12})$$

Let $\gamma > 0$ be arbitrary, $\epsilon > 0$, with $\gamma_+ = 1 + \epsilon/\gamma$ and $\gamma_- = 1 - \epsilon/\gamma$, such that $\epsilon/\gamma \ll 0$ (infinitesimal). From Eq. (A.12) we have

$$E_{\text{c}}[n_{\gamma\gamma_+}] - E_{\text{c}}[n_{\gamma\gamma_-}] \geq \gamma_+ E_{\text{c}}[n_{\gamma}] - \gamma_- E_{\text{c}}[n_{\gamma}]. \quad (\text{A.13})$$

Simplifying yields

$$\gamma \frac{E_{\text{c}}[n_{\gamma+\epsilon}] - E_{\text{c}}[n_{\gamma-\epsilon}]}{2\epsilon} \geq E_{\text{c}}[n_{\gamma}]. \quad (\text{A.14})$$

Applying the definition of a derivative and identifying $T_C[n_\gamma]$ we obtain T_C *non-negativity*:

$$T_C[n_\gamma] = \gamma \frac{\partial E_C[n_\gamma]}{\partial \gamma} - E_C[n_\gamma] \geq 0. \quad (\text{A.15})$$

Enforcing this condition locally yields the following local condition for approximations

$$\frac{\partial \tilde{F}_C(r_s, \zeta, s, \alpha, q)}{\partial r_s} \geq 0, \quad (\text{A.16})$$

which is the same one that corresponds to Eq. (A.12).

The two inequalities in Eq. (A.12) are equivalent. For instance, let $\gamma' < 1$ be arbitrary (but strictly positive) and take $\gamma = 1/\gamma' > 1$. Take $n \mapsto n_{\gamma'}$ in Eq. (A.12) and we have $E_C[n_{\gamma'}] \geq \gamma E_C[n_\gamma]$. Substituting for γ and rearranging we have $E_C[n_{\gamma'}] \leq \gamma' E_C[n]$ with $\gamma' < 1$.

Finally, we note that if the local conditions in Eqs. (A.16) and (A.7) are satisfied, then we have

$$\tilde{F}_{xc} \leq \tilde{F}_{xc} + r_s \frac{\partial \tilde{F}_C}{\partial r_s} \leq \tilde{F}_x + \tilde{F}_C(\infty). \quad (\text{A.17})$$

Since the exchange energy follows the simple scaling relation, $E_x[n_\gamma] = \gamma E_x[n]$, \tilde{F}_x needs to be scale-invariant and thus independent of r_s . The rightmost side of the inequality, $\tilde{F}_x + \tilde{F}_C(\infty)$, is then the maximum value that \tilde{F}_{xc} can take (assuming all other variables besides r_s are fixed). Therefore, if an approximation satisfies Eqs. (A.16) and (A.7), then the local condition

$$\tilde{F}_{xc} \leq C_{\text{LO}}, \quad (\text{A.18})$$

which is the standard one that corresponds to the LO bound on E_{xc} , will imply Eq. (A.10). In

Table 1 in the main text, indeed we see that functionals like PBE, AM05, PBE, and CASE21, which simultaneously satisfy Eqs. (A.16), (A.7), and (A.18), automatically satisfy the LO in Eq. (A.8).

We also remark that an enhancement factor \tilde{F}_C that satisfies the conditions in Eqs.(A.16), (A.7), and (A.18) is monotonic and Lipschitz continuous in r_s , i.e., the derivative is bounded, $0 \leq \partial\tilde{F}_C/\partial r_s \leq K$, where K is a finite constant known as the Lipschitz constant. Such a property of the enhancement factor may help assuage possible issues during numerical integration [236]. However, this property is only with respect to the r_s variable and is clearly not sufficient to ensure stability.

A.4 conventional exchange and correlation partitioning

We define a global hybrid functional approximation as

$$\tilde{E}_{\text{xc}}^{\text{hyb}}[n] = \tilde{E}_{\text{xc}}[n] + \tilde{a}E_{\text{x}}[n], \quad (\text{A.19})$$

where $\tilde{a} > 0$, $E_{\text{x}}[n]$ is the exact exchange energy, and $\tilde{E}_{\text{xc}}[n]$ is the remaining semilocal density functional that can be expressed as

$$\tilde{E}_{\text{xc}}[n] = \int d^3r n(\mathbf{r})\tilde{\epsilon}_{\text{xc}}(r_s, \zeta, s, \alpha, q). \quad (\text{A.20})$$

Note the difference between α and other usual definitions for the mixing parameter in global hybrids.

In hybrid XC functionals and other available approximations, the partitioning of the exchange and correlation may be ambiguous or not defined. In these cases, we use the following

conventional partitioning

$$E_C[n] = E_{XC}[n] - \lim_{\gamma \rightarrow \infty} \frac{E_{XC}[n_\gamma]}{\gamma}, \quad (A.21)$$

which holds for the exact functional [243]. In global hybrid functionals, the exact exchange contributions cancel out in Eq. (A.21), since $E_X[n_\gamma] = \gamma E_X[n]$, and do not contribute to the correlation energy. Here the conventional partitioning for the correlation energy can be expressed as a semilocal density functional with the following correlation energy per electron

$$\tilde{\epsilon}_C[n](\mathbf{r}) = \tilde{\epsilon}_{XC}(r_s, \zeta, s, \alpha, q) - \lim_{\gamma \rightarrow \infty} \frac{\tilde{\epsilon}_{XC}(r_s/\gamma, \zeta, s, \alpha, q)}{\gamma}. \quad (A.22)$$

In many cases, the conventional partitioning is consistent with the one defined by the authors of the approximation (if one exists). This agreement occurs whenever an approximation satisfies: $\tilde{E}_X[n_\gamma] = \gamma \tilde{E}_X[n]$ and $\lim_{\gamma \rightarrow \infty} \tilde{E}_C[n_\gamma]/\gamma \rightarrow 0$. The latter is satisfied when $\tilde{E}_C[n_\gamma]$ goes to a finite constant as $\gamma \rightarrow \infty$, but also for approximations of the form

$$\tilde{E}_C[n] = \int d^3r n(\mathbf{r}) \epsilon_C^{\text{unif}}[n] \tilde{G}(\zeta, s, \alpha, q), \quad (A.23)$$

where $\epsilon_C^{\text{unif}}[n]$ is a suitable parameterization for the correlation energy per electron of the uniform gas that has logarithmic divergence in the high-density limit (e.g. PW92 [331] or VWN [427]) and $\tilde{G}(\zeta, s, \alpha, q)$ depends only on dimensionless variables. In general, the conventional partitioning we use may differ from the developer's intentions or rationalizations. For instance, approximations that consider a portion of exact exchange to be part of the correlation energy will not be consistent with our conventional partitioning, since such energy contributions will cancel out in Eq. (A.21).

Throughout, conventional partitioning is utilized whenever the exchange and correlation

partitioning is not available in the Libxc library [238]. In our analysis, we do not consider range-separated hybrid functionals, double-hybrid functionals, or functionals that contain non-local correlation. Therefore, the correlation energy functional is always expressed in the standard semilocal form considered in this work.

A.5 Lieb-Oxford bound for hybrid functionals

A global hybrid functional satisfies the XC energy LO bound when

$$\tilde{E}_{\text{xc}}[n] \geq C_{\text{LO}} E_{\text{x}}^{\text{LDA}}[n] - \tilde{a} E_{\text{x}}[n]. \quad (\text{A.24})$$

The exchange energy alone follows a tighter bound

$$E_{\text{x}}[n] \geq C_{\text{LO}}^{\text{x}} E_{\text{x}}^{\text{LDA}}[n], \quad (\text{A.25})$$

with coefficient $C_{\text{LO}}^{\text{x}} < C_{\text{LO}}$. Therefore, we can ensure the LO bound with a semilocal functional satisfying

$$\begin{aligned} \tilde{E}_{\text{xc}}[n] &\geq (C_{\text{LO}} - \tilde{a} C_{\text{LO}}^{\text{x}}) E_{\text{x}}^{\text{LDA}}[n] \\ &\geq C_{\text{LO}} E_{\text{x}}^{\text{LDA}}[n] - \tilde{a} E_{\text{x}}[n]. \end{aligned} \quad (\text{A.26})$$

Thus, it is sufficient (but not necessary) that $\tilde{E}_{\text{xc}}[n]$ satisfy a LO-like bound with coefficient $C_{\text{LO}} - \tilde{a} C_{\text{LO}}^{\text{x}}$. A larger coefficient could result in a violation of the XC energy LO bound in Eq. (A.24). To see this, let $\Delta > 0$ and let n_{LO}^{x} be a density such that $E_{\text{x}}[n_{\text{LO}}^{\text{x}}] = C_{\text{LO}} E_{\text{x}}^{\text{LDA}}[n_{\text{LO}}^{\text{x}}]$, then $(C_{\text{LO}} - \tilde{a} C_{\text{LO}}^{\text{x}} + \Delta) E_{\text{x}}^{\text{LDA}}[n_{\text{LO}}^{\text{x}}] < C_{\text{LO}} E_{\text{x}}^{\text{LDA}}[n_{\text{LO}}^{\text{x}}] - \tilde{a} E_{\text{x}}[n_{\text{LO}}^{\text{x}}]$. A smaller coefficient would also ensure Eq. (A.24), but it would be over-restrictive.

The corresponding local condition is straightforward

$$\tilde{F}_{\text{xc}} \leq C_{\text{LO}} - \tilde{a} C_{\text{LO}}^{\text{x}}. \quad (\text{A.27})$$

The corresponding local condition for the LO bound involving $\tilde{U}_{\text{xc}}^{\text{hyb}}[n]$ is found by using the conventional partitioning in Eq. (A.22) and applying Eq. (A.4) to yield

$$\tilde{F}_{\text{xc}} + r_s \frac{\partial \tilde{F}_{\text{C}}}{\partial r_s} \leq C_{\text{LO}} - \tilde{a} C_{\text{LO}}^{\text{x}}. \quad (\text{A.28})$$

In practice, C_{LO} and C_{LO}^{x} are not known precisely and need to be approximated with proven (but not optimal) bounds. To give the most benefit of the doubt when assessing approximations, we use $C_{\text{LO}}^{\text{x}} = 1.174$ (the conjectured tight exchange LO coefficient [329]) and $C_{\text{LO}} = 2.27$ (which is the same value we use to evaluate non-hybrids).

A.6 Local condition for the conjecture: $T_{\text{C}} \leq -E_{\text{C}}$

While unproven, it has been conjectured [242, 89, 128] that

$$T_{\text{C}}[n] \leq -E_{\text{C}}[n] \quad (\text{conjecture}). \quad (\text{A.29})$$

One can employ the definition in Eq. (A.4) and arrive at the following local condition

$$\frac{\partial \tilde{F}_{\text{C}}}{\partial r_s} \leq \frac{\tilde{F}_{\text{C}}}{r_s} \quad (\text{conjecture}). \quad (\text{A.30})$$

In the following Sections A.10 and A.11 we explore the satisfaction of Eqs. (A.29) and (A.30), respectively, in approximate functionals.

A.7 Relation to adiabatic connection

Uniform coordinate scaling is closely related to the adiabatic connection in DFT, which has long been an illuminating concept for rationalizing and improving density functional approximations [318]. The formalism developed has also revealed many useful exact conditions.

In the adiabatic connection formalism [164, 228, 156], we insert a variable coupling constant $\lambda \geq 0$ for Coulomb-interacting electrons with

$$F^\lambda[n] = \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} + \lambda \hat{V}_{ee} | \Psi \rangle , \quad (\text{A.31})$$

where \hat{T} is the usual total kinetic energy operator, \hat{V}_{ee} is the two-body electron-electron repulsion operator, and the minimization is over all antisymmetric wavefunctions that yield the density n . For $\lambda = 1$ and ground-state density $n(\mathbf{r})$, we have our real physical system. Taking $\lambda = 0$, we have the KS system, and $F^{\lambda=0}[n] = T_s[n]$, where $T_s[n]$ is the kinetic energy of the non-interacting KS system with ground-state density $n(\mathbf{r})$. In all cases $\lambda \geq 0$, the ground-state density remains fixed to that of the physical system $n(\mathbf{r})$.

The adiabatic connection is directly coupled to uniform coordinate scaling by

$$F^\lambda[n] = \lambda^2 F[n_{1/\lambda}] . \quad (\text{A.32})$$

This relation also extends to any other energy functional component, e.g., $E_C^\lambda[n] = \lambda^2 E_C[n_{1/\lambda}]$. Therefore, exact conditions written in terms of uniform coordinate scaling can be recast in terms of adiabatic connection quantities.

Eq. (A.12) is rewritten as

$$\begin{aligned} E_c^\lambda[n] &\geq \lambda E_c[n] \quad (\lambda < 1), \\ E_c^\lambda[n] &\leq \lambda E_c[n] \quad (\lambda > 1). \end{aligned} \tag{A.33}$$

Using

$$\frac{\partial E_c[n_{1/\lambda}]}{\partial(1/\lambda)} = -\frac{\partial E_c^\lambda[n]}{\partial\lambda} + 2\frac{E_c^\lambda[n]}{\lambda} \tag{A.34}$$

we recast the exact condition in Eq. (A.3) as

$$\frac{\partial E_c^\lambda[n]}{\partial\lambda} \geq \left(-\frac{\partial E_c^\lambda[n]}{\partial\lambda} + 2\frac{E_c^\lambda[n]}{\lambda} \right) \Big|_{\lambda\rightarrow\infty}, \tag{A.35}$$

and applying L'Hospital's rule we obtain

$$\frac{\partial E_c^\lambda[n]}{\partial\lambda} \geq \frac{\partial E_c^\lambda[n]}{\partial\lambda} \Big|_{\lambda\rightarrow\infty}. \tag{A.36}$$

This equation can also be rewritten as a simple statement in the *strictly correlated electron* (SCE) [372] limit ($\lambda \rightarrow \infty$),

$$\langle \Psi^\lambda[n] | \hat{V}_{ee} | \Psi^\lambda[n] \rangle \geq \langle \Psi^{\lambda\rightarrow\infty}[n] | \hat{V}_{ee} | \Psi^{\lambda\rightarrow\infty}[n] \rangle. \tag{A.37}$$

In the SCE limit, the kinetic energy component is subleading and $\Psi^{\lambda\rightarrow\infty}[n]$ minimizes \hat{V}_{ee} , thus Eq. (A.37) is clear.

A.8 Analytical derivation: PBE satisfies the correlation uniform scaling inequality exact condition

The PBE correlation energy functional has the form [313]

$$E_C^{\text{PBE}}[n] = \int d^3r n(\mathbf{r}) (\epsilon_C^{\text{PW92}}(r_s, \zeta) + H(r_s, \zeta, t)), \quad (\text{A.38})$$

where ϵ_C^{PW92} is the PW92 [331] parameterized correlation energy per electron of the uniform gas and $H(r_s, \zeta, t)$ is defined in Ref. [313] along with the dimensionless gradient t . For simplicity, we set all positive constants in H to unity, as the final conclusion will not depend on their specific values. With this,

$$H = \ln \left[1 + t^2 \left[\frac{1 + At^2}{1 + At^2 + A^2t^4} \right] \right] \geq 0, \quad (\text{A.39})$$

where

$$A = [\exp(-\epsilon_c^{\text{PW92}}) - 1]^{-1} \geq 0. \quad (\text{A.40})$$

In the following, we show that PBE correlation satisfies Eq. (A.16), and thus the correlation uniform scaling inequality exact condition in Eq. (A.12). Since PW92 already satisfies Eq. (A.16), proving the following condition,

$$\frac{\partial}{\partial r_s} \left(\frac{H(r_s, \zeta, t)}{\epsilon_x^{\text{unif}}(r_s)} \right) \geq 0, \quad (\text{A.41})$$

is sufficient to ensure that PBE satisfies the exact condition.

To show this, we start with

$$\frac{\partial}{\partial r_s} \left(\frac{H(r_s, \zeta, t)}{\epsilon_x^{\text{unif}}(r_s)} \right) = \frac{\left(\frac{\partial}{\partial r_s} H \right)}{\epsilon_x^{\text{unif}}} + \frac{\left(\frac{\partial}{\partial r_s} \epsilon_x^{\text{unif}} \right) H}{(\epsilon_x^{\text{unif}})^2}. \quad (\text{A.42})$$

Since $\epsilon_x^{\text{unif}} = -(3/4)(3/2\pi)^{2/3}r_s^{-1}$, the second term in Eq. (A.42) is positive. Next we evaluate the first term,

$$\frac{\partial H}{\partial r_s} = \frac{1}{1+x} \times \frac{\partial x}{\partial A} \times \frac{\partial A}{\partial \epsilon_c^{\text{PW92}}} \times \frac{\partial \epsilon_c^{\text{PW92}}}{\partial r_s}, \quad (\text{A.43})$$

where

$$x = t^2 \left[\frac{1 + At^2}{1 + At^2 + A^2 t^4} \right] \geq 0. \quad (\text{A.44})$$

The intermediate derivatives are derived analytically:

$$\frac{\partial x}{\partial A} = \frac{-t^4 A(t^2 A + 2)}{(t^4 A^2 + t^2 A + 1)^2} \leq 0, \quad (\text{A.45})$$

$$\frac{\partial A}{\partial \epsilon_c^{\text{PW92}}} = \frac{\exp(-\epsilon_c^{\text{PW92}})}{(\exp(-\epsilon_c^{\text{PW92}}) - 1)^2} \geq 0, \quad (\text{A.46})$$

and

$$\frac{\partial \epsilon_c^{\text{PW92}}}{\partial r_s} \geq 0. \quad (\text{A.47})$$

Therefore,

$$\frac{\partial H}{\partial r_s} \leq 0, \quad (\text{A.48})$$

and the first term in Eq. (A.42) is also positive and thus the exact condition is satisfied.

A.9 Readjusting parameters in the SCAN functional

While the local condition in Eq. (A.16) is satisfied in the published SCAN functional, reasonable adjustments of the parameters designated for the fitting of the “appropriate norms” can result in violations of the local condition. Specifically, we adjust the $b_{1c} = 0.02858$ parameter in SCAN, which was fit to match the correlation energy of the $Z \rightarrow \infty$ limit of two-electron ions (one of the five appropriate norms in SCAN, see the supplementary material in Ref. [388]). By evaluating SCAN analytically, we find that using $b_{1c} > 0.2$ results in violations of Eq. (A.16). Under this single parameter modification, the other exact conditions that SCAN satisfies are still satisfied by construction. Therefore, by virtue of fitting to various appropriate norms, the SCAN functional satisfies more exact conditions than were explicitly enforced.

A.10 exact conditions on atomic system densities

In this assessment, we first calculate Hartree-Fock (HF) densities and orbitals for neutral atoms H-Ar and their cations. The fixed HF densities and orbitals are then used to evaluate the energies (non-self-consistently) from different DFT approximations. HF densities are used because they provide high quality densities and an equal footing across different approximations. We also performed separate calculations using self-consistent densities and observed marginal differences. The absolute errors from the experimental ionization energies are provided in Fig. A.2. In addition to established approximations, we also test a very simple modified B3LYP (MOD-B3LYP) which satisfies the correlation exact conditions discussed for *any* density,

$$\epsilon_c^{\text{MOD-B3LYP}}(r_s, \zeta, s) = \Theta(s - 1.82) \epsilon_c^{\text{B3LYP}}(r_s, \zeta, s), \quad (\text{A.49})$$

where $\epsilon_c^{\text{B3LYP}} = 0.405\epsilon_c^{\text{LYP}} + 0.095\epsilon_c^{\text{VWN5}}$ [385]. The step function eliminates the local condition violations found in Table 1 of the main text. As argued in the main text, energy contributions from such large s values are less relevant in Coulombic systems, and indeed in Fig. A.2 we see that B3LYP and MOD-B3LYP have MAEs that differ only by 0.1 kcal/mol. Our modified functional is constructed for demonstration purposes only.

For each HF density (a total of 35 atomic systems), we scale the density n_γ with $\gamma \in [0.01, 2]$ (50 evenly spaced values) and evaluate whether an exact condition is satisfied. Indeed, in Fig. A.3 we see that the exact conditions tested are all satisfied within our set of functional approximations and atomic systems. We also test the conjecture $T_c \leq -E_c$ and find instances of violation for PBE and M08-HX.

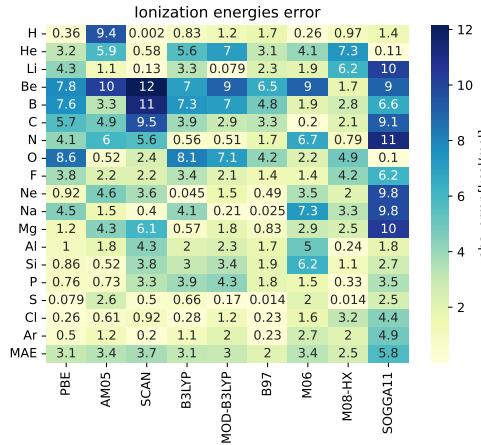


Figure A.2: Ionization energy errors for atomic systems H-Ar. All approximations are evaluated on HF densities and orbitals.

A.11 assessment of local conditions across available approximations

We utilize an exhaustive grid search to determine whether local conditions are satisfied for a given approximation. For GGAs, we consider 10000 evenly spaced values of $r_s \in [0.0001, 5]$,

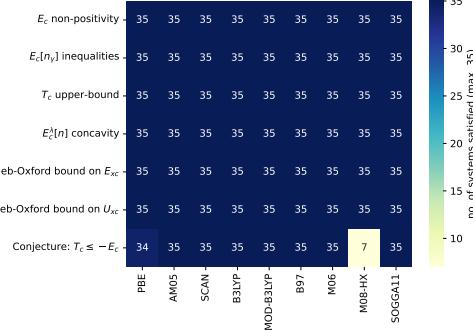


Figure A.3: The number of HF atomic densities tested which satisfy a given exact condition. A total of 35 atomic systems are tested: H-Ar and their cations.

500 evenly spaced values of $s \in [0, 5]$, and 100 evenly spaced values of $\zeta \in [0, 1]$. For MGGAs, we consider 5000 evenly spaced values of $r_s \in [0.0001, 5]$, 100 evenly spaced values of $s \in [0, 5]$, 20 evenly spaced values of $\zeta \in [0, 1]$, and 100 evenly spaced values of $\alpha \in [0, 5]$ or $q \in [-10, 10]$. The number of values checked per variable is less in MGGAs to alleviate computational effort due to the combinatorial nature of the exhaustive search.

In determining whether local conditions are satisfied, a reasonable tolerance threshold of at most ± 0.001 is employed to approximately account for numerical errors arising from the numerical precision used and the finite difference method used to calculate numerical derivatives (further details can be found in our public code [4]). However, the numerical errors introduced are not guaranteed to be within the tolerances used.

In the following tables below, we report the fraction of local condition violations found in our exhaustive search. That is, we divide the number of violations found by the total number of permutations considered in the extensive grid search parameter space. If we find 0 such violations, then we conclude that the corresponding exact condition is satisfied for any reasonable density.

Table A.3: GGA functionals: numerical assessment of corresponding local conditions.

	$E_c[n]$ non-positivity	$E_c[n_\gamma]$ uniform scaling inequality	$T_c[n]$ upper bound	$U_c(\lambda)$ monotonicity	LO extension to E_{xc}	LO	conjecture: $T_c \leq -E_c$
ACGGAP [65, 61]	0	0	0	0	—	—	0.414
ACGGA [65, 61]	0	0	0	0	—	—	0
AM05 [22, 269]	0	0	0	0	0	0	0
APBE [84]	0	0	0	0	0	0	0.004
B97-D [152]	0.632	0.503	0.559	0.608	0.268	0.183	0.633
B97-GGA1 [76]	0.636	0.514	0.557	0.612	0.390	0.317	0.639
BEEFVDW [435]	0	0	0	0	0.003	0.013	0
BMK [54]	0.627	0.304	0.648	0.621	—	—	0.616
CCDF [268]	0	0	0	0	—	—	0
CHACHIYO [70]	0	0	0.044	0	0.217	0.217	0.010
CS1 [160, 346]	0.604	0.204	0.528	0.530	—	—	0.601
EDF1 [12]	0.605	0.245	0.002	0.231	0.162	0.203	0.527
FT97 [122, 121]	0	0	1e-05	0.003	—	—	0
GAM [454]	0.598	0.459	0.560	0.578	0.145	0.083	0.596
GAPC [116]	0.004	0.011	2e-04	0.005	—	—	0.015
GAPLOC [116]	4e-04	2e-04	2e-04	0.005	—	—	0.033
HCTH-120 [51]	0.495	0.310	0.327	0.450	0.065	0.061	0.507
HCTH-147 [51]	0.467	0.290	0.298	0.422	0.113	0.093	0.478
HCTH-407P [50]	0.536	0.445	0.428	0.508	0.105	0.075	0.543
HCTH-407 [52]	0.481	0.382	0.365	0.450	0.112	0.079	0.489
HCTH-93 [159]	0.435	0.196	0.258	0.386	0.266	0.237	0.446
HCTH-A [159]	0.493	0.289	0.355	0.454	0	0	0.501
HCTH-P14 [275]	0	0	0	0	0	0.029	0
HCTH-P76 [275]	0.986	0.929	0.999	0.991	0.011	0.005	0.978
HLE16 [419]	0.481	0.305	0.364	0.447	0.477	0.473	0.487
HYB-TAU-HCTH [53]	0.615	0.439	0.520	0.585	—	—	0.620
KT1 [204]	0.791	0.434	0.096	0.169	0.153	0.076	0.664
KT2 [204]	0.832	0.477	0.106	0.177	0.156	0.077	0.686
KT3 [205]	0.862	0.461	0.120	0.192	0.164	0.077	0.678
LM [229, 179]	0	0.119	0.464	0.384	—	—	0
LYPR [14]	0.320	0.113	0.801	0.590	—	—	0.438
LYP [233, 279]	0.576	0.218	0.003	0.203	—	—	0.511
MGGAC [302]	0	0	0	0	—	—	0.007
MOHLYP2 [473]	0.576	0.174	0.002	0.193	0.340	0.337	0.509
MOHLYP [367]	0.243	0.092	0	0.092	0.048	0.096	0.328
MPWLYP1W [94]	0.500	0.190	4e-07	0.168	0.003	0.004	0.474
N12 [338]	0	0	0	0	0.150	0.170	0
NCAP [68]	0.455	0.299	0.029	0.277	0.231	0.207	0.403
OBLYP-D [139]	0.595	0.246	0.002	0.243	0.019	0.020	0.505
OP-B88 [413]	0	2e-04	6e-04	0.002	—	—	7e-04
OP-G96 [413, 414]	0	2e-04	6e-04	0.002	—	—	7e-04
OP-PBE [413, 414]	0	2e-04	6e-04	0.002	—	—	7e-04
OP-PW91 [413, 414]	0	0.001	0.001	0.002	—	—	0.001
OP-XALPHA [413, 414]	0	2e-04	6e-04	0.002	—	—	7e-04
OPBE-D [139]	0	0	0	0	0.009	0.009	0.006
OPTC [77]	3e-06	0	0	0	—	—	0
OPWLYP-D [139]	0.596	0.247	7e-04	0.240	0.019	0.022	0.507
P86-FT [311]	0.454	0.298	0.027	0.276	—	—	0.402
P86VWN-FT [311]	0.447	0.297	0.026	0.275	—	—	0.389
P86VWN [311]	0.447	0.297	0.026	0.275	—	—	0.389
P86 [311]	0.454	0.298	0.027	0.276	—	—	0.403
PBE-JRGX [308]	0	0	0	0	—	—	0.006
PBE-MOL [97]	0	0	0	0	0	0	0.004
PBE-SOL [324]	0	0	0	0	0	0	0.006
PBE-VWN [218, 314, 315]	0	0	0	0	—	—	6e-04
PBE1W [94]	0	0	0	0	0	0	0
PBEFE [362]	0	0	0	0	0	0	0.006
PBEINT [114]	0	0	0	0	0	0	0.005
PBELOC [83]	0	0	0	0.003	—	—	0.271
PBELYP1W [94]	0.414	0.157	174 0	0.141	0	0	0.427
PBE [314, 315]	0	0	0	0	0	0	0.005

Table A.4: GGA functionals: numerical assessment of corresponding local conditions.

	$E_c[n]$ non-positivity	$E_c[n_\gamma]$ uniform scaling inequality	$T_c[n]$ upper bound	$U_c(\lambda)$ monotonicity	LO extension to E_{xc}	LO	conjecture: $T_c \leq -E_c$
PW91 [312, 316, 317]	3e-06	0	0	0	0	0	0
Q2D [74]	0	0.041	0.012	0.032	0	0	0.002
REGTPSS [321]	0	0	0	0	—	—	0.406
REVTCA [405]	0	0	0.003	0.051	—	—	0.024
RGE2 [361]	0	0	0	0	0	0	0.005
SCAN-E0 [389]	0	0	0	0	—	—	0
SG4 [86]	0	0.048	0.327	0.385	8e-04	0.007	0.050
SOGGA11 [342]	0	0.003	0.064	0.229	0	1e-04	0.002
SPBE [392]	0	0	0	0	—	—	0
TAU-HCTH [53]	0.595	0.491	0.492	0.567	—	—	0.603
TCA [406]	0	0	0	0	—	—	0
TH-FCFO [410]	0.223	0.772	0.233	0.237	0.787	0.228	0.226
TH-FCO [410]	0.200	0.795	0.211	0.214	0.781	0.205	0.203
TH-FC [410]	0.988	0.009	0.994	0.996	0.822	0.990	0.989
TH-FL [410]	0	1.000	0	0	0.498	0	0
TH1 [408]	0.215	0.780	0.224	0.225	0.295	0.220	0.218
TH2 [409]	0.061	0.935	0.070	0.070	0.323	0.065	0.063
TH3 [161]	0.217	0.781	0.217	0.219	0.284	0.218	0.218
TH4 [161]	0.103	0.894	0.106	0.106	0.238	0.105	0.105
TM-LYP [399]	0.575	0.209	0.122	0.173	—	—	0.565
TM-PBE [399]	0	0	0	0	—	—	0.509
W94 [443]	0	0	0	3e-05	—	—	0
WI0 [444]	0.614	0.004	0.002	0.014	—	—	0.603
WI [444]	0.900	0.008	0.008	0.023	—	—	0.896
WL [445]	0.590	0.166	0.595	0.581	—	—	0.377
XLYP [450]	0.576	0.218	0.003	0.203	0.013	0.019	0.511
XPBE [449]	0	0	0	0	0	0	0
ZPBEINT [82]	0	0.020	0.343	0.299	—	—	2e-04
ZPBESOL [82]	0	0.013	0.375	0.263	—	—	2e-04
ZVPBEINT [85]	0	0.093	0.273	0.226	—	—	0.001
ZVPBELOC [113]	0	3e-04	0.199	0.112	—	—	0.075
ZVPBESOL [85]	0	0.080	0.299	0.212	—	—	0.001

Table A.5: MGGA functionals: numerical assessment of corresponding local conditions.

	$E_C[n]$ non-positivity	$E_C[n_\gamma]$ uniform scaling inequality	$T_C[n]$ upper bound	$U_C(\lambda)$ monotonicity	LO extension to E_{xc}	LO	conjecture: $T_c \leq -E_c$
B88 [35]	0	0	0	0	—	—	0
B94 [39]	0	0	0	0	—	—	0
BC95 [40]	0	0	0	0	—	—	0
CC06 [66]	0	0	0	0	0	0	0
CS [81, 233]	0.352	0.166	0.252	0.275	—	—	0.481
HLE17 [420]	0	0	0	0	0.092	0.093	0
HTAPW [237]	0	0	0	0	—	—	0
KCISK [354, 220, 221, 225, 407]	0	0	0	0.012	—	—	0.041
KCIS [354, 220, 221, 225, 407]	0	0	0	0	—	—	0
LP90 [232]	0	0.963	0.998	0.990	0	0	0
M06-L [469, 471]	0.700	0.661	0.705	0.696	0.228	0.181	0.698
M11-L [340]	0.385	0.194	0.004	0.153	0.425	0.456	0.478
MN12-L [339]	0.424	0.266	0.022	0.227	0.048	0.086	0.506
MN15-L [453]	0.462	0.184	3e-04	0.156	2e-04	0.006	0.594
OTPSS-D [139]	0	0	0	0	0	0	0.007
PKZB [320]	0	0	0	0	0	0	0.006
R2SCANL [274, 129, 130]	0	0	0	0	0	0	0
R2SCAN [129, 130]	0	0	0	0	0	0	0
REVM06-L [433]	0.777	0.702	0.835	0.783	5e-05	7e-05	0.767
REVSCAN [277]	0.212	0.063	0	0.084	0	0	0.288
REVTM [185]	0	0	0	0	0	0	0.329
REVTPSS [321, 322]	0	0	0	0	0	0	0.424
RSCAN [33]	0	0.008	0.101	0.160	0	0	0
SCANL [272, 273, 389]	0	0	0	0	0	0	0
SCAN [389]	0	0	0	0	0	0	0
TM [397]	0	0	0	0	0	0	7e-04
TPSSLOC [83]	0	0	0	0.003	—	—	0.280
TPSSLYP1W [94]	0.408	0.179	0	0.138	0	0	0.420
TPSS [398, 330]	0	0	0	0	0	0	0.008
VSXC [424]	0.298	0.084	0.190	0.260	—	—	0.294
ZLP [458]	0	0.943	0.998	0.981	0.343	0.285	0

Table A.6: Hybrid GGA functionals: numerical assessment of corresponding local conditions.

	$E_c[n]$ non-positivity	$E_c[n_\gamma]$ uniform scaling inequality	$T_c[n]$ upper bound	$U_c(\lambda)$ monotonicity	LO extension to E_{xc}	LO	conjecture: $T_c \leq -E_c$
APBE0 [113]	0	0	0	0	0	0	0.004
APF [24]	0	0	0	0	0	0	0
B1LYP [9]	0.576	0.218	0.003	0.203	0	0	0.511
B1PW91 [9]	0	0	0	0	0	0	0
B1WC [47]	0	0	0	0	0	0	0.005
B3LYP-MCM1 [64]	0.753	0.282	0.681	0.455	0	0	0.590
B3LYP-MCM2 [64]	0.558	0.210	4e-04	0.187	0	0	0.504
B3LYP3 [386]	0.457	0.174	0	0.154	0	0	0.451
B3LYP5 [386]	0.457	0.174	0	0.154	0	0	0.451
B3LYPS [351]	0.414	0.159	0	0.137	0	5e-05	0.442
B3LYP [386]	0.414	0.159	0	0.137	0	8e-05	0.442
B3P86 [2]	0.074	0.141	0	0.209	0	0	0.265
B3PW91 [37]	0	0	0	0	0	0	0
B5050LYP [375]	0.457	0.174	0	0.154	0	0	0.451
B97-1P [76]	0.666	0.463	0.634	0.649	0.056	0.042	0.663
B97-1 [159]	0.698	0.476	0.686	0.687	0.002	0.001	0.693
B97-2 [446]	0.665	0.479	0.605	0.643	0.065	0.048	0.666
B97-3 [206]	0.612	0.464	0.601	0.601	0.089	0.050	0.608
B97-K [54]	0.330	0.004	0.520	0.342	0	0	0.291
B97 [41]	0.645	0.359	0.612	0.626	0.003	0.003	0.641
BHANDHLYP [38, 2]	0.576	0.218	0.003	0.203	0	0	0.511
BHANDH [38, 2]	0.576	0.218	0.003	0.203	0	0	0.511
BLYP35 [352, 203]	0.576	0.218	0.003	0.203	0	0	0.511
CAP0 [67]	0	0	0	0	0.080	0.081	0.005
CASE21 [382]	0	0	0	0	0	0	0.259
EDF2 [257]	0.362	0.132	0	0.125	0	0	0.398
HAPBE [113]	0	0	0	0	0	0	0.012
HFLYP [233, 279]	0.576	0.218	0.003	0.203	0	0	0.511
HPBEINT [115]	0	0	0	0	0	0	0.005
KMLYP [196]	0.121	0.052	0	0.054	0	0	0.279
MB3LYP-RC04 [404]	0.359	0.143	0.001	0.126	0	1e-04	0.375
MPW1K [263]	0	0	0	0	0	0	0
MPW1LYP [10]	0.576	0.218	0.003	0.203	0	0	0.511
MPW1PBE [10]	0	0	0	0	0	0	0.005
MPW1PW [10]	0	0	0	0	0	0	0
MPW3LYP [464]	0.462	0.178	0	0.153	0	0	0.465
MPW3PW [10]	0	0	0	0	0	0	0
MPWLYP1M [367]	0.576	0.218	0.003	0.203	3e-04	3e-04	0.511
O3LYP [173, 77]	0.457	0.174	0	0.154	0.017	0.028	0.451
PBE-2X [393]	0	0	0	0	0	0	0.005
PBE-MOL0 [97]	0	0	0	0	0	0	0.004
PBE-MOLB0 [97]	0	0	0	0	0	0	0.005
PBE-SOL0 [97]	0	0	0	0	0	0	0.006
PBE0-13 [87]	0	0	0	0	0	0	0.005
PBE38 [153]	0	0	0	0	0	0	0.005
PBE50 [46]	0	0	0	0	0	0	0.005
PBEB0 [97]	0	0	0	0	0	0	0.005
PBEH [11, 111]	0	0	0	0	0	0	0.005
QTP17 [188]	0.407	0.156	0	0.135	0	0	0.439
REVB3LYP [259]	0.438	0.168	0	0.145	0	0	0.453
SB98-1A [364]	0.803	0.435	0.903	0.809	0	0	0.786
SB98-1B [364]	0.267	0.049	0.055	0.215	0.013	0.019	0.278
SB98-1C [364]	0.635	0.363	0.593	0.615	0.004	0.004	0.633
SB98-2A [364]	0	0	0	0	0.002	0.012	0
SB98-2B [364]	0.478	0.180	0.327	0.432	2e-05	4e-04	0.484
SB98-2C [364]	0.632	0.371	0.586	0.611	0.001	0.002	0.629
SOGGA11-X [337]	0	0.042	0.176	0.169	0	0.003	0
WC04 [442]	0	0	176	0	0.123	0.240	0
WP04 [442]	0	0	0	0	0.349	0.468	0
X3LYP [450]	0.462	0.178	0	0.153	0	0	0.465

Table A.7: Hybrid MGGA functionals: numerical assessment of corresponding local conditions.

	$E_c[n]$ non-positivity	$E_c[n_\gamma]$ uniform scaling inequality	$T_c[n]$ upper bound	$U_c(\lambda)$ monotonicity	LO extension to E_{xc}	LO	conjecture: $T_c \leq -E_c$
B0KCIS [407]	0	0	0	0	0	0	0
B86B95 [40]	0	0	0	0	0	9e-04	0
B88B95 [40]	0	0	0	0	0	9e-04	0
B98 [42]	0.202	0.004	0.145	0.119	0	0	0.172
BB1K [460]	0	0	0	0	0	1e-03	0
BR3P86 [292]	0.290	0.336	0.005	0.247	0	6e-04	0.305
DLDf [334]	0.543	0.455	0.523	0.532	0.298	0.241	0.540
EDMGGAH [396]	0.351	0.166	0.254	0.276	0.015	0.013	0.481
M05-2X [463]	0.745	0.696	0.760	0.743	0.070	0.042	0.741
M05 [462]	0.748	0.559	0.767	0.729	0.075	0.073	0.731
M06-2X [471]	0.681	0.613	0.720	0.682	0.058	0.036	0.674
M06-HF [467]	0.298	0.232	0.382	0.313	0.049	0.105	0.284
M06 [471]	0.419	0.229	0.257	0.375	0.329	0.330	0.423
M08-HX [470]	0.039	0.008	0	0.007	0.070	0.259	0.136
M08-SO [470]	0.079	0.019	0	0.015	0.149	0.260	0.196
MN15 [452]	0.785	0.612	0.396	0.631	0.288	0.276	0.860
MPW1B95 [464]	0	0	0	0	0	9e-04	0
MPW1KCIS [459]	0	0	0	0	0	0	0
MPWB1K [464]	0	0	0	0	0	1e-03	0
MPWKCIS1K [459]	0	0	0	0	0	0	0
PBE1KCIS [465]	0	0	0	0	0	0	0
PW6B95 [466]	0	0	0	0	0	9e-04	0
PW86B95 [40]	0	0	0	0	0	9e-04	0
PWB6K [466]	0	0	0	0	0	1e-03	0
REVM06 [434]	0.676	0.591	0.685	0.670	0.003	0.002	0.670
REVTPSSH [91]	0	0	0	0	0	0	0.424
TPSS0 [151]	0	0	0	0	0	0	0.008
TPSS1KCIS [461]	0	0	0	0	0	0	0
TPSSH [384]	0	0	0	0	0	0	0.008
X1B95 [464]	0	0	0	0	0	0	0
XB1K [464]	0	0	0	0	0	4e-05	0

Appendix B

Supplemental Info for Chapter 5

This chapter is a reproduction of the supplemental info in Ref. [252].

B.1 1D Model systems

In 1D, we utilize exponential Coulomb interactions to mimic the standard 3D Coulomb potential,

$$v_{\text{exp}}(x) = A \exp(-\kappa|x|), \quad (\text{B.1})$$

where $A = 1.071295$ and $\kappa^{-1} = 2.385345$ [29]. Within this model, the external one-body potential for a nuclei of atomic number Z and position x' is represented by $-Z v_{\text{exp}}(x - x')$. The external potential for arbitrary molecular systems and geometries is modeled as

$$v(x) = - \sum_j Z_j v_{\text{exp}}(x - x_j). \quad (\text{B.2})$$

For example, a 1D H_2 molecule at separation $R = 4$ can be represented by $v(x) = -v_{\text{exp}}(x - 2) - v_{\text{exp}}(x + 2)$. The repulsion between electrons at positions x and x' is given by the two-body potential $v_{\text{ee}}(x - x') = v_{\text{exp}}(x - x')$. We represent all systems on a 1D grid of $m = 513$ points each separated by a distance $h = 0.08$. The center grid point is at the origin, $x = 0$, and the range of grid points is $x \in \{-20.48, \dots, 20.48\}$. For consistency, all nuclei positions reside on grid points in calculations. In this convention, all molecules in this work are either symmetric about the origin $x = 0$ or $x = 0.04$, depending on the separation between nuclei.

B.2 DMRG calculation details

The real-space interacting Hamiltonian for a 1D system of lattice spacing h becomes in second quantized notation,

$$H = \frac{5}{4h^2} \sum_{j,\sigma} n_{j\sigma} - \frac{2}{3h^2} \sum_{\langle i,j \rangle, \sigma} c_{i\sigma}^\dagger c_{j\sigma} + \frac{1}{24h^2} \sum_{\langle\langle i,j \rangle\rangle, \sigma} c_{i\sigma}^\dagger c_{j\sigma} + \sum_j v(x_j) n_j + \sum_{ij} v_{\text{ee}}(x_i - x_j) n_i n_j, \quad (\text{B.3})$$

where the operator $c_{j\sigma}^\dagger$ creates (and $c_{j\sigma}$ annihilates) an electron of spin σ on site j , $n_{j\sigma} = c_{j\sigma}^\dagger c_{j\sigma}$, and $n_j = n_{j\uparrow} + n_{j\downarrow}$. The single and double brackets below the sums indicate sums over nearest and next nearest neighbors, respectively. The hopping term coefficients are determined by the 4-th order central finite difference approximation to the second derivative. The Hamiltonian is solved using DMRG to obtain highly accurate ground-state energies and densities. Calculations are performed using the ITensor library [124] with an energy convergence threshold of 10^{-7} Ha.

B.3 KS calculation details

B.3.1 Local Density Approximation

In our 1D model the electron repulsion is an exponential interaction. To implement a local density approximation (LDA) for this interaction we use Ref. [29] which provides the exponentially repelling uniform gas exchange energy analytically and an accurate parameterized model for the correlation energy. We use this specific implementation for all LDA calculations.

B.3.2 Initial density

We solve the Schrödinger equation of the non-interacting system with the external potential $v(x)$ defined in Eq. B.2,

$$\left\{ -\frac{\nabla^2}{2} + v(x) \right\} \phi_i(x) = \epsilon_i \phi_i(x). \quad (\text{B.4})$$

The density is the square sum of all the occupied orbitals $n(x) = \sum_i |\phi_i(x)|^2$. In all the KS self-consistent calculations presented in this work, we use the density of the non-interacting system with external potential $v(x)$ as the initial density.

B.3.3 Linear density mixing

Linear density mixing is a well-known strategy to improve the convergence of the KS self-consistent calculation,

$$n_{k+1}^{(\text{in})} = n_k^{(\text{in})} + \alpha(n_k^{(\text{out})} - n_k^{(\text{in})}). \quad (\text{B.5})$$

In this work, we apply an exponential decay on the mixing factor $\alpha = 0.5 \times 0.9^{k-1}$.

B.3.4 XC potential from automatic differentiation

The XC functional $\epsilon_{\text{XC}}[n] : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a mapping from density $n \in \mathbb{R}^m$ to XC energy density $\epsilon_{\text{XC}} \in \mathbb{R}^m$, where m is the number of grids. Instead of hand-deriving the functional derivative, we use automatic differentiation in JAX [58] to compute

$$v_{\text{XC}} = \frac{\delta E_{\text{XC}}}{\delta n} = \frac{\delta \int n \epsilon_{\text{XC}}[n] dx}{\delta n}. \quad (\text{B.6})$$

The XC energy density $\epsilon_{\text{XC}}[n]$ is denoted as `xc_energy_density_fn`. It takes the `density`, a float array with size m as input argument and returns a float array with size m . This function can be a conventional physics XC functional (e.g., LDA) or a neural XC functional.

The XC energy E_{XC} can be computed by the following function,

```

1 def get_xc_energy(density, xc_energy_density_fn, grids):
2     """Gets the xc energy by discretizing the following integral.
3
4     E_xc = \int density * xc_energy_density_fn(density) dx.
5
6     Args:
7         density: Float numpy array with shape (num_grids,).
8         xc_energy_density_fn: function takes density and returns float numpy
9             array
10            with shape (num_grids,).
11
12    Returns:
13        Float.
14    """

```

```

15     return jnp.dot(xc_energy_density_fn(density), density) * utils.get_dx(
16         grids)

```

Then the functional derivative $v_{\text{xc}} = \delta E_{\text{xc}} / \delta n$ can be computed via `jax.grad` function,

```

1 def get_xc_potential(density, xc_energy_density_fn, grids):
2     """Gets xc potential.
3
4     The xc potential is derived from xc_energy_density through automatic
5     differentiation.
6
7     Args:
8         density: Float numpy array with shape (num_grids,).
9         xc_energy_density_fn: function takes density and returns float numpy
10        array
11        with shape (num_grids,).
12
13     Returns:
14         Float numpy array with shape (num_grids,).
15
16     """
17     return jax.grad(get_xc_energy)(
18         density, xc_energy_density_fn, grids) / utils.get_dx(grids)

```

The `jax.grad` function computes the gradient of `get_xc_energy` function with respect to the first argument `density` using automatic differentiation. Both functions can be found in the `scf` module in the JAX-DFT library.

B.3.5 Symmetry

The training molecules used in this paper are symmetric to their centers. We define the symmetry operation on functions on the grids $\mathcal{S} : \mathbb{R}^m \rightarrow \mathbb{R}^m$. It flips a function at the

center and averages with itself. In each KS iteration, we enforce the symmetry on the XC functional, $\epsilon_{\text{xc}}[n] \rightarrow \mathcal{S}(\epsilon_{\text{xc}}[\mathcal{S}(n)])$. So E_{xc} and v_{xc} are transformed as

$$E_{\text{xc}} = \int n \epsilon_{\text{xc}}[n] dx \rightarrow \int n \mathcal{S}(\epsilon_{\text{xc}}[\mathcal{S}(n)]) dx \quad (\text{B.7})$$

$$v_{\text{xc}} = \frac{\delta \int n \epsilon_{\text{xc}}[n] dx}{\delta n} \rightarrow \frac{\delta \int n \mathcal{S}(\epsilon_{\text{xc}}[\mathcal{S}(n)]) dx}{\delta n}. \quad (\text{B.8})$$

Before the output of each KS iteration, $n_k^{(\text{out})} \rightarrow \mathcal{S}(n_k^{(\text{out})})$.

We note that applying the symmetry restriction does not change the model performance of the molecules around equilibrium. However, because both stretched H_2 and H_4 have vanishing KS gaps, the KS self-consistent calculation is difficult to converge. The gradient information for the KSR relies on the stable output of the KS self-consistent calculation. In realistic 3D DFT codes, enforcing symmetry is one common strategy to improve the convergence of KS self-consistent calculations. Therefore, we applied symmetry restriction so that KSR can obtain stable gradient information in the stretched limit cases.

B.4 Training, validation and test

B.4.1 Weights in trajectory loss

We use $w_k = 0.9^{K-k} H(k-10)$, where H is the Heaviside function and K is the total number of iterations.

B.4.2 Number of KS iterations

We first run KS calculations with the standard uniform gas LDA functional [29]. The number of iterations to converge the largest separation for different molecules are around 8, 25, 5, 6

for H_2 , H_4 , H_2^+ , H_2H_2 respectively. During the training of neural XC functionals with KSR, we use a fixed number of iterations that is greater than or equal to the estimation from LDA for each type of molecules so it is sufficient for convergence. The number of iterations for different molecules are listed in Table B.1.

Table B.1: Number of iterations for different molecules.

	H_2	H_4	H_2^+	H_2H_2
train	15	40	—	—
validation	15	40	—	—
test	15	40	5	10

B.4.3 Dataset for learning H_2 dissociation from two molecules

H_2 dissociation curves in Figure 1 are trained from exact densities and energies of two molecules. One is a compressed H_2 ($R = 1.28$) and the other is a stretched H_2 ($R = 3.84$) molecule. The optimal checkpoint is selected by a validation molecule with $R = 2.96$.

B.4.4 Dataset for learning and predicting several types of molecules

Training molecules

The distances between nearby atoms for training molecules used in Figure 4 are listed in Table B.2.

Validation molecules

The distances between nearby atoms for validation molecules used in Figure 4 are listed in Table B.3.

Table B.2: Distances between nearby atoms for training molecules used in Figure 4.

N_{train}	H_2					H_4				
4		1.28		3.84			2.08	3.36		
6	0.48	1.28		3.84		1.28	2.08	3.36		
8	0.48	1.28	3.04	3.84		1.28	2.08	3.36	4.48	
10	0.48	1.28	3.04	3.84	4.64	1.28	2.08	3.36	4.00	4.48
12	0.48	1.28	2.40	3.04	3.84	4.64	1.28	2.08	3.04	3.36
14	0.48	1.28	2.40	3.04	3.52	3.84	4.64	1.28	2.08	3.68
16	0.48	1.28	1.76	2.40	3.04	3.52	3.84	4.64	2.08	2.56
18	0.48	1.28	1.76	2.40	3.04	3.52	3.84	4.16	4.64	3.04
20	0.48	0.80	1.28	1.76	2.40	3.04	3.52	3.84	4.16	4.64
							1.28	1.76	2.08	2.56
								3.04	3.36	3.68
								4.00	4.48	4.80

Table B.3: The distances between nearby atoms for validation molecules used in Figure 4. The validation set is fixed for calculations with $4 \leq N_{\text{train}} \leq 20$.

Molecule R	
H_2	1.68, 2.96, 4.40, 5.52
H_4	1.84, 2.64, 3.28, 5.04

Test molecules

The distances between nearby atoms for test molecules used in Figure 4 are listed in Table B.4.

Molecule R	
H_2	0.40, 0.56, 0.72, 0.88, 1.04, 1.20, 1.36, 1.52, 1.84, 2.00, 2.16, 2.32, 2.48, 2.64, 2.80, 3.12, 3.28, 3.44, 3.60, 3.76, 3.92, 4.08, 4.24, 4.56, 4.72, 4.88, 5.04, 5.20, 5.36, 5.68, 5.84, 6.00
H_4	1.04, 1.20, 1.36, 1.52, 1.68, 2.00, 2.16, 2.32, 2.48, 2.80, 2.96, 3.12, 3.44, 3.60, 3.76, 3.92, 4.08, 4.24, 4.40, 4.56, 4.72, 4.88, 5.20, 5.36, 5.52, 5.68, 5.84, 6.00
H_2^+	0.64, 0.80, 0.96, 1.12, 1.28, 1.44, 1.60, 1.76, 1.92, 2.08, 2.24, 2.40, 2.48, 2.56, 2.64, 2.72, 2.88, 3.04, 3.20, 3.36, 3.52, 3.68, 3.84, 4.00, 4.16, 4.32, 4.48, 4.64, 4.80, 4.96, 5.12, 5.28, 5.44, 5.60, 5.76, 5.92, 6.08, 6.24, 6.40, 6.56, 6.72, 6.88, 7.04, 7.20, 7.36, 7.52, 7.68, 7.84, 8.00, 8.16, 8.32, 8.48
H_2H_2	0.16, 0.48, 0.80, 1.12, 1.44, 1.76, 2.08, 2.40, 2.72, 3.04, 3.36, 3.68, 4.00, 4.32, 4.64, 4.96, 5.28, 5.60, 5.92, 6.24, 6.56, 6.88, 7.20, 7.52, 7.84, 8.16, 8.48, 8.80, 9.12, 9.44, 9.76

Table B.4: The distances between nearby atoms for test molecules used in Figure 4. The test set is fixed for calculations with $4 \leq N_{\text{train}} \leq 20$.

Test errors

We extend the plot of test errors in Figure 4 to $N_{\text{train}} = 20$ in Figure B.1 and list all the numerical values in Table B.5.

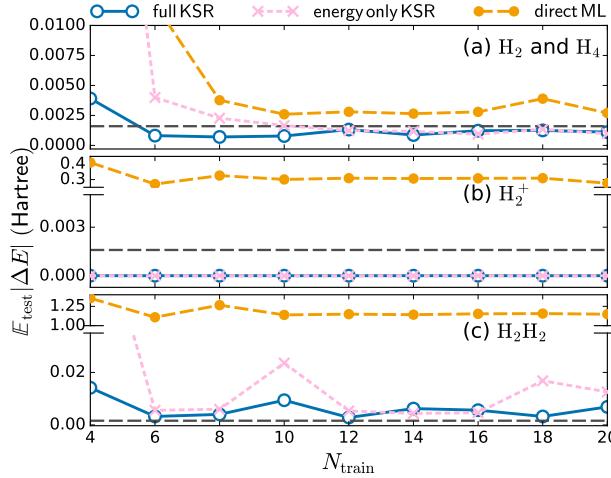


Figure B.1: Test generalization of ML models as a function of the total number of training examples N_{train} : full KSR (blue), energy only KSR (pink) and direct ML (orange) on (a) holdout H_2 and H_4 , and unseen types of molecules (b) H_2^+ (c) H_2H_2 . Black dashed lines show chemical accuracy. All the numerical values are listed in Table B.5.

Table B.5: Numerical values of test errors plotted in Figure B.1.

Model	Molecule	N_{train}								
		4	6	8	10	12	14	16	18	20
KSR $L_n + L_E$	H_2 and H_4	3.91×10^{-3}	8.15×10^{-4}	7.04×10^{-4}	7.82×10^{-4}	1.32×10^{-3}	8.64×10^{-4}	1.23×10^{-3}	1.25×10^{-3}	1.11×10^{-3}
	H_2^+	1.71×10^{-5}								
	H_2H_2	1.42×10^{-2}	3.23×10^{-3}	4.02×10^{-3}	9.41×10^{-3}	2.83×10^{-3}	6.23×10^{-3}	5.62×10^{-3}	3.23×10^{-3}	6.87×10^{-3}
KSR L_E	H_2 and H_4	4.82×10^{-2}	3.99×10^{-3}	2.27×10^{-3}	1.66×10^{-3}	1.23×10^{-3}	1.17×10^{-3}	9.37×10^{-4}	1.35×10^{-3}	9.73×10^{-4}
	H_2^+	1.71×10^{-5}								
	H_2H_2	9.19×10^{-2}	5.56×10^{-3}	5.99×10^{-3}	2.36×10^{-2}	5.35×10^{-3}	4.38×10^{-3}	4.67×10^{-3}	1.68×10^{-2}	1.26×10^{-2}
ML	H_2 and H_4	4.95×10^{-2}	1.18×10^{-2}	3.77×10^{-3}	2.60×10^{-3}	2.80×10^{-3}	2.65×10^{-3}	2.79×10^{-3}	3.89×10^{-3}	2.70×10^{-3}
	H_2^+	4.10×10^{-1}	2.71×10^{-1}	3.26×10^{-1}	3.01×10^{-1}	3.09×10^{-1}	3.07×10^{-1}	3.08×10^{-1}	3.09×10^{-1}	2.76×10^{-1}
	H_2H_2	1.36	1.11	1.27	1.14	1.15	1.14	1.15	1.16	1.15

Dissociation curve of H_2 , H_4 , H_2^+ and H_2H_2

Figure B.2 shows the dissociation curve of H_2 , H_4 , H_2^+ and H_2H_2 . Curves of KSR-global (blue) are computed from the neural XC functional trained from the full KSR with $N_{\text{train}} = 8$ (four H_2 molecules and four H_4 molecules) in Figure B.1. KSR fits H_2 and H_4 well even in the stretched limit. KSR perfectly predicts H_2^+ because of self-interaction gate in the neural XC functional. Although H_2H_2 has never been exposed to the model during training, KSR performs well at small distances ($R < 3$) and at large distances ($R > 8$) but slightly overbinds around $R = 5$.

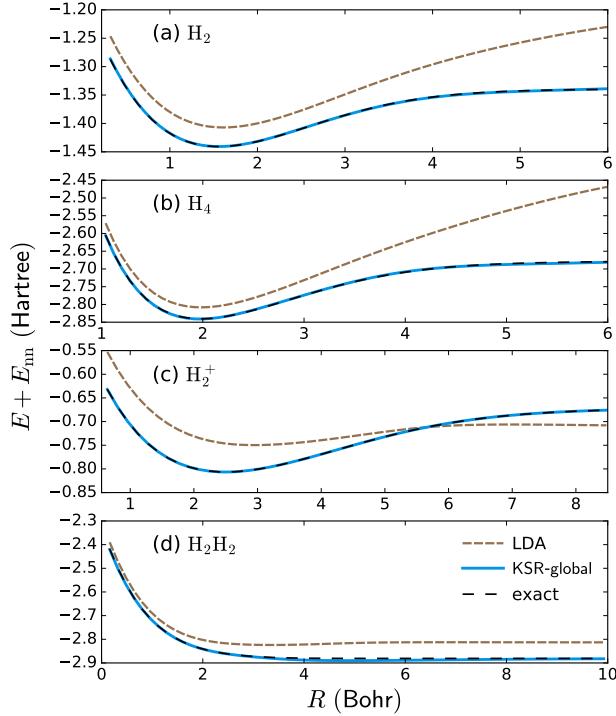


Figure B.2: Dissociation curves of (a) H_2 , (b) H_4 , (c) H_2^+ , and (d) H_2H_2 . Dashed black lines are the exact curves. Dashed brown lines denote the results computed from uniform gas LDA and blue solid lines denote the results computed from KSR-global.

B.5 Neural networks

B.5.1 Architecture

Figure B.3 illustrates the model architectures used in Figure 1. In the ML model that directly predicts energy from geometry (Figure B.3(a)), we first solve Eq. B.4 to obtain a density for a particular molecular geometry. We use this density as a smooth representation of the geometry. The first few layers (global conv-conv-SiLU-conv-SiLU) are identical to the KSR-global in Figure B.3(d). Next, we use convolution layers with 128 channels and dense layers to increase the capacity of the model. Finally, a dense layer with a single unit outputs the scalar E .

The KSR-LDA and KSR-GGA approaches do not have the global convolution layer be-

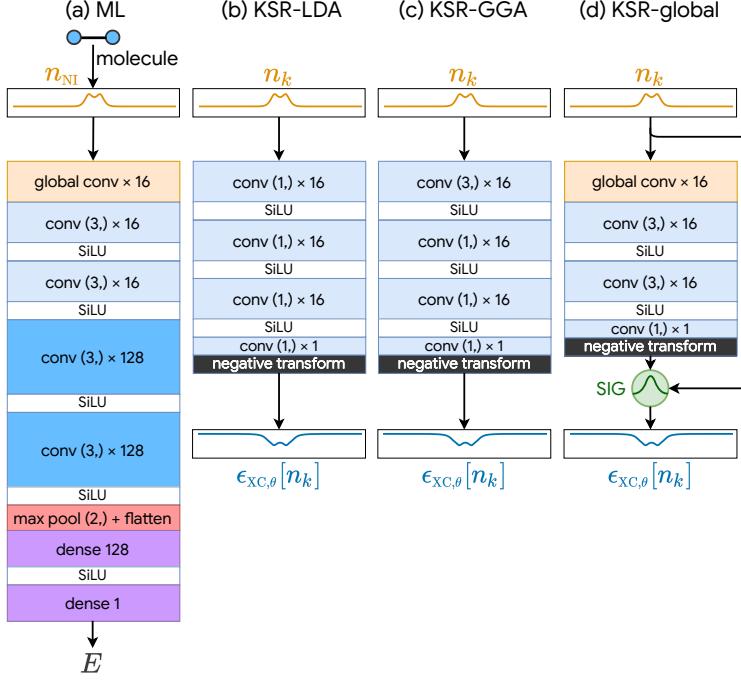


Figure B.3: Model architectures of (a) the ML model that directly predicts energy E from geometry, (b) the neural LDA with KSR, (c) the neural GGA with KSR, and (d) the neural global functional.

cause the use of global information violates the local and semi-local approximation. The first layer of KSR-LDA is a convolution with filter size 1. It mimics the physics of the standard LDA approach by mapping the density value to the XC energy density at the same point, $\epsilon_{\text{XC},\theta}^{\text{LDA}} : \mathbb{R}^1 \rightarrow \mathbb{R}^1$. KSR-GGA uses a convolution layer with filter size 3 to map the density values of three nearby points to the XC energy density at the center point, $\epsilon_{\text{XC},\theta}^{\text{GGA}} : \mathbb{R}^3 \rightarrow \mathbb{R}^1$. The XC energy density in the entire space is also computed pointwise, $\epsilon_{\text{XC}} = \{\epsilon_{\text{XC},\theta}^{\text{GGA}}[n(x_{-1}, x_0, x_1)], \dots, \epsilon_{\text{XC},\theta}^{\text{GGA}}[n(x_{m-2}, x_{m-1}, x_m)]\} \in \mathbb{R}^m$. The remaining network structure of KSR-LDA and KSR-GGA is identical to KSR-global, except for self-interaction gate (SIG).

B.5.2 Layers

Convolution

Filter weights in 1D convolution are initialized by He normal initialization [167]. The stride is one. The edges are padded with zero to ensure that the size of the output spatial dimension is the same as the size of the unpadded input spatial dimension. There is no bias term.

Global convolution

Global convolution contains multiple channels to capture the interaction in different scales. The operation in each channel is

$$G(n(x), \xi_p) = \frac{1}{2\xi_p} \int dx' n(x') \exp(-|x - x'|/\xi_p). \quad (\text{B.9})$$

where ξ_p is trainable and controls the scale of the interaction. We parameterize $\xi_p = a + (b - a) \cdot \sigma(\eta_p)$ using the sigmoid function $\sigma(x) = 1/(1 + \exp(-x))$ to bound $\xi_p \in (a, b)$. η_p is initialized using the normal distribution $\mathcal{N}(0, 10^{-4})$. For the 16-channel global convolution layer used in this work, we have $\eta_1 \equiv 0$ to preserve the input density and the rest $\eta_p \in (0.1, \kappa^{-1})$ are trainable.

Dense

The dense layer is only used in the ML model (Figure B.3(a)). The weights are initialized using Glorot normal initialization [138] and bias terms are initialized using the normal distribution $\mathcal{N}(0, 10^{-4})$.

B.5.3 Checkpoint selection

Each calculation is repeated with 25 random seeds. The model is trained by L-BFGS [258] implemented in SciPy [421] `scipy.optimize.fmin_l_bfgs_b` with `factr=1,m=20,pgtol=1e-14`. Parameter checkpoints are saved every 10 steps until L-BFGS stops. The optimal checkpoint is the checkpoint with the lowest average energy error per electron on validation sets, $\mathbb{E}_{\{\mathcal{S}_{\text{val}}\}} \mathbb{E}_{\mathcal{M} \in \mathcal{S}_{\text{val}}} |(E - E_{\text{DMRG}})|/N_e$, where E is the final energy from KS calculations, E_{DMRG} is the exact energy, and N_e is the number of electrons. The validation sets $\{\mathcal{S}_{\text{val}}\}$ and the molecules \mathcal{M} in each sets are listed in Table B.3.

B.6 Training a neural XC functional without KS regularization

[365] proposed a neural XC functional that can be used in a KS self-consistent calculation in the inference stage. Unlike our work, which trains the network through KS calculations, they train the network in a single-step. The training set contains 12800 molecules and the validation set contains 6400 molecules. Here, the molecules are exact solutions of one-dimensional two-electron problems in the external potential of up to three random nuclei (Equation 4 in [365]). The exact v_{xc} for each molecule is computed by an inverse KS method. They input exact ground state density and train the network to predict XC energy per length e_{xc} that minimizes the loss function: a weighted combination of the mean square errors (MSE) of the XC energy, XC potential, its numerical spatial derivative, and the difference between the XC energy and the integral over the potential (Equation 5 in [365]),

$$L(\theta) = \alpha \text{MSE}(E_{\text{xc}}) + \beta \text{MSE}(v_{\text{xc}}) + \gamma \text{MSE}\left(\frac{dv_{\text{xc}}(x)}{dx}\right) + \delta \text{MSE}\left(E_{\text{xc}} - \int dx v_{\text{xc}}(x)n(x)\right), \quad (\text{B.10})$$

where $\alpha = 1.0$, $\beta = 100.0$, $\gamma = 10.0$, and $\delta = 1.0$ is the weights used in [365].

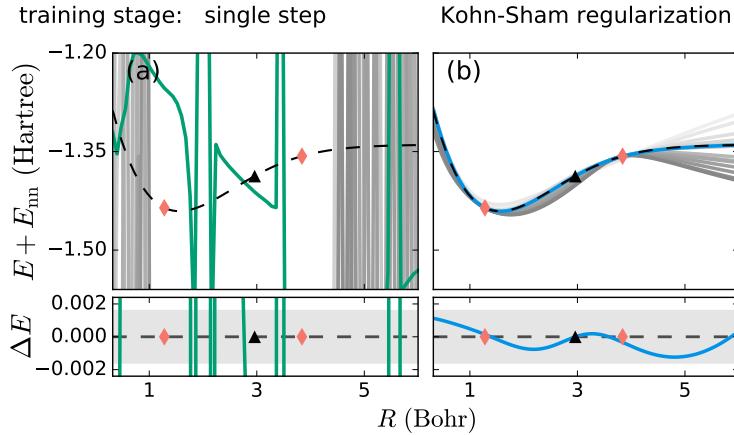


Figure B.4: Training the neural XC functional using (a) single-step and (b) Kohn-Sham regularization. Both functionals use Kohn-Sham self-consistent calculations in the inference stage.

It is natural to pose the question: does the generalization from the two H_2 training molecules in Figure 1 result from using KS self-consistent calculations in the inference stage rather than the training stage? This is a reasonable concern because the XC energy is usually a small portion of the total energy. To justify this concern, we first use inverse KS to get the exact v_{xc} on the two H_2 molecules used in the H_2 experiment. Then, we take the model architecture¹ and loss function in [365] and attempt to learn the entire dissociation curve of H_2 from two molecules. Figure B.4 compares the results from KS self-consistent calculations using functionals trained on (a) single-step and (b) Kohn-Sham regularization. It is not surprising that even though both approaches use KS self-consistent calculations in the inference stage, the model trained on a single-step fails to generalize in the small training set limit (1/6400 training set size to the original paper). The neural XC functional is a many-to-many mapping, which is very hard to learn with limited data. Moreover, KS self-consistent calculations start with an initial density that is not the exact ground state density. It is clearly out of the interpolation region for the model that has only seen exact densities of two molecules.

¹The only difference is that this model predicts $\epsilon_{\text{xc}}[n](x)$ rather than $e_{\text{xc}}[n](x)$ in [365]. The relation between them is $e_{\text{xc}}[n](x) = \epsilon_{\text{xc}}[n](x) \cdot n(x)$

We would like to emphasize that this comparison aims to show that using KS calculations in *training* – Kohn-Sham regularizer – is crucial to the generalization. A single-step model could work well as reported in [365] with a larger training set and exact v_{XC} .

B.7 Training a neural XC functional with “weaker” Kohn-Sham regularization

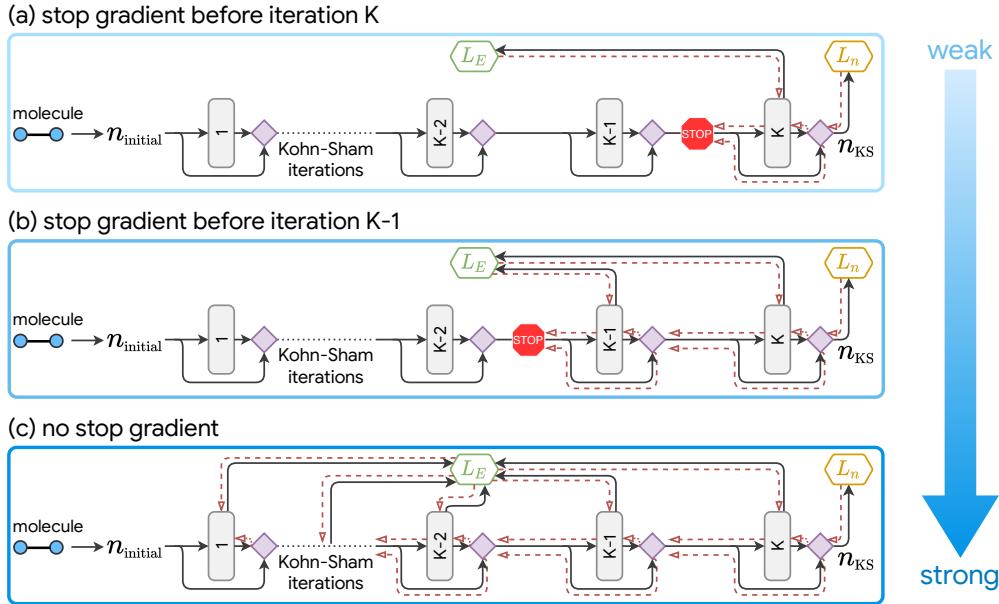


Figure B.5: Computational graph with different KSR strength. (a) stop gradient before iteration K . (b) stop gradient before iteration $K-1$. (c) no stop gradient. This is the same computation graph used in the main text.

Unlike other methods that build physics prior knowledge to the model through constraint, KSR “augments” densities for the model during training. Thus, there is no single coefficient to explicitly control the strength of the regularization. A straightforward idea to control the KSR strength is to change the total number of iterations K in the KS self-consistent calculations. However, a small K may not be sufficient to converge KS calculations. Thus it is ambiguous to understand whether the worse performance is from weaker regularization or un converged KS calculations. Here we design an approach to control the strength of KSR

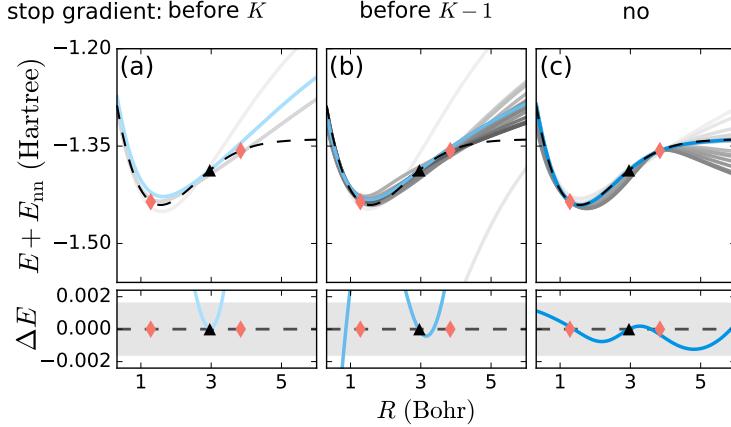


Figure B.6: H_2 dissociation curves trained from two molecules (red diamonds) with a corresponding computational graph shown in Figure B.5.

by stopping the gradient flow in the backpropagation and keeping K fixed.

Stop gradient is a common operation in differentiable programming. It acts as an identity in the forward pass so it does not affect the KS calculations. In the backpropagation, it sets the gradient passing through it to zero. As shown in Figure B.5, we stop gradient before a certain KS iteration $k = k^*$ so all the previous iterations $k < k^*$ have no access to the gradient information. Since the gradient may still flow into the iteration k from L_E through its energy output E_k , we also stop the gradient on E_k for $k < k^*$. To simplify the graph, we remove the arrows between E_k to L_E for $k < k^*$. In Figure B.5(a), the neural XC functional is updated only from the gradient information flowing through the final iteration. (b) is similar to (a) but has access to the gradient flowing through the last two iterations. No stop gradient is applied to (c) and it is identical to the computational graph we used in the main text. We repeat the same experiment in Figure 1. Figure B.6 shows the H_2 dissociation curves trained with three stop gradient setting in Figure B.5. In Figure B.6(a), L-BFGS converges quickly as there is no sufficient gradient information for training. By including the gradient information in the $K - 1$ -th iteration, the distribution of the dissociation curves predicted by the model during training get closer to the true curve in (b). For comparison, we place the distribution of dissociation curves from model without stop gradient in (c), previously

shown in Figure 1(d), where the physics of the true dissociation curve is captured.

Appendix C

Supplemental Info for Chapter 7

This chapter is a reproduction of the supplemental info in Ref. [307].

C.1 orthogonalization details

The original set of N basis functions $\chi_i(\mathbf{r})$, corresponding in our current FHI-aims implementation to numeric atom-centered orbitals (NAOs), are not orthogonal, in the sense that the overlap matrix S , with coefficients

$$S_{ij} = \langle \chi_i | \chi_j \rangle = \int d^3r \chi_i(\mathbf{r}) \chi_j(\mathbf{r}) \quad (\text{C.1})$$

is not the identity matrix, but some non-trivial, positive definite matrix. To identify the linear combinations of orbitals that are occupied in the ground state of the system, we need to account for overlaps between the orbitals. This can be done using the Löwdin

decomposition, where the Hamiltonian matrix H , with coefficients

$$H_{ij} = \langle \chi_i | \mathcal{H} | \chi_j \rangle = \int d^3r \chi_i(\mathbf{r}) \mathcal{H}(\mathbf{r}) \chi_j(\mathbf{r}), \quad (\text{C.2})$$

is transformed into an orthonormal basis as $H \mapsto \tilde{H} = S^{-\frac{1}{2}} H S^{-\frac{1}{2}}$. The transformed Hamiltonian \tilde{H} can then be purified, as described in the next section, to yield the density matrix \tilde{D} , which is then transformed back into the original orbitals basis,

$$\tilde{D} \mapsto D = S^{-\frac{1}{2}} \tilde{D} S^{-\frac{1}{2}}. \quad (\text{C.3})$$

For this purpose, the inverse square root $S^{-\frac{1}{2}}$ of the overlap matrix is needed. To compute it efficiently on TPUs, we need an algorithm for computing the matrix inverse square root for which the computational bottleneck reduces to repeated matrix multiplications. As discussed in [172], this can be achieved using Newton-Schulz iterations. The iteration

$$X_{[n+1]} = \frac{1}{2} X_{[n]} (3I - X_{[n]}^2), \quad X_{[0]} = A, \quad (\text{C.4})$$

converges to the matrix sign function of A , which turns positive eigenvalues to $+1$ and negative eigenvalues to -1 . Notice that two matrix multiplications are needed for each iteration. These matrix multiplications can be executed very quickly when distributed over a set of TPUs. The inverse square root can in turn be cast as a sign function as

$$\text{sgn} \left(\begin{bmatrix} 0 & S \\ I & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & S^{\frac{1}{2}} \\ S^{-\frac{1}{2}} & 0 \end{bmatrix}. \quad (\text{C.5})$$

Applying the iteration (C.4) to the block matrix in (C.5) results in the Denman-Beavers iteration for computing the inverse square root.

In single (double) precision, the above procedure typically converges in about 35-50 (65-90)

iterations, depending on how small the absolute value of the smallest (in absolute value) eigenvalue of matrix A is. In order to further accelerate this computation, we introduce the pre-conditioning polynomial iteration (which we described and justified in Sect. III.D of [248] in the related context of the matrix sign function for singular values),

$$X_{[n+1]} = aX_{[n]}(I - \frac{4}{27}a^2X_{[n]}^2), \quad X_{[0]} = A, \quad (C.6)$$

where $a = \frac{3}{2}\sqrt{3} - s_-$ for some choice of small $s_- > 0$. [Notice that for $a = 3/2$, that is $s_- = 3(\sqrt{3}-1)/2$ we recover (C.4).] This pre-conditioning polynomial accelerates the growth of small eigenvalues of A , until they become of size at least s_- . From then on, the regular Newton-Schulz iteration is used to bring all the positive eigenvalues to 1, with quadratic convergence. For $s_- = 0.1$, in single (double) precision we need 15-20 (35) iterations of the pre-conditioning polynomial and 10 (10) iterations of the regular polynomial, for a total of 25-30 (45) iterations.

The inverse of S is of course very sensitive to poor conditioning of S , which for the overlap matrix corresponds to (nearly) linearly dependent orbitals, a common occurrence when dealing with large molecules. The danger of instability and poor accuracy due to small eigenvalues of S is especially pressing if operating in low numerical precision. Consequently we always compute $S^{-\frac{1}{2}}$ in double precision. Because TPUs do not operate natively in double precision but rather rely on software emulation, this incurs a significant time cost. However, in a full DFT simulation that cost gets amortized: the overlap matrix does not change between DFT iterations (only the Hamiltonian does), and thus we only need to compute $S^{-\frac{1}{2}}$ once at the first iteration, write the result to disk, and reread and use it at all the successive iterations with negligible cost.

C.2 purification details

By *density matrix purification* we mean the map from a Hermitian matrix \tilde{H} of linear size N to a certain density matrix \tilde{D} , itself a projector into the subspace corresponding to the N_e smallest (or most negative) eigenvalues of \tilde{H} , where N_e is the number of electrons in the system. [As in the rest of the paper, the tilde in the Hamiltonian \tilde{H} and density matrix $\tilde{D}^{(k)}$ denotes that these matrices are expressed in an orthonormalized basis of orbitals, as described in the previous section.] In symbols, let

$$\tilde{H} = V\Sigma V^H \tag{C.7}$$

be the ascendingly sorted eigendecomposition of \tilde{H} , and let $\rho \equiv \text{diag}(1_1, 1_2, \dots, 1_{N_e}, 0_{N_e+1}, 0_{N_e+2}, \dots, 0_N)$ be a diagonal matrix with N_e 1's followed by $N - N_e$ 0's on the main diagonal. Then \tilde{D} is defined by

$$\tilde{D} \equiv V\rho V^H. \tag{C.8}$$

It follows identically that $\tilde{D}^2 = \tilde{D}$, so that \tilde{D} is indeed a projector.

With the decomposition (C.7) in hand, \tilde{D} is trivially computed by the manual substitution $\Sigma \rightarrow \rho$. Unfortunately an efficient algorithm for Hermitian eigendecomposition is not presently available in a distributed-TPU context. Instead, we turn to matrix-multiplication based purification algorithms originally developed in the context of linear scaling methods (see [207] for a review; note that since our matrices are dense and are not truncated, our implementations scale as N^3 despite the name).

Density matrix purification algorithms can be divided into two classes [299] by the manner in which N_e is specified. In *grand canonical purification*, a so-called *chemical potential* μ is given, and \tilde{D} found so that the $N_e + 1$ 'th most negative entry of Σ is the first to exceed μ .

This can be achieved by shifting the spectrum of \tilde{H} by μ so that the latter divides negative from positive eigenvalues, and then computing a polar decomposition using the methods of [248].

In the *canonical purification* used in this work, N_e is instead specified directly. Compared to the grand canonical case this is more directly relevant to computations of molecular electronic structure, where μ is unknown but the number N_e of electrons is provided.

Various algorithms for canonical purification have been proposed in the literature. The original scheme is presented in [299], and is variously referred to as *canonical purification* (in which case other algorithms are given a different name), *trace-preserving purification*, or the *Palser and Manolopoulos scheme*. The *trace-resetting* schemes proposed in [293, 294] are probably most common in practical use. We use the *generalized* or *hole-particle* scheme presented in [412]. In our TPU experiments this iteration yields performance comparable to that of [293, 294], but avoids certain branching conditionals which are awkward to phrase efficiently on the TPU.

All such schemes work by first mapping the input \tilde{H} to some initial $X_{[0]}$ with eigenvectors unchanged but eigenvalues bound in $[0, 1]$, and then repeatedly applying a matrix-multiplication based iteration which also preserves eigenvectors. This iteration is chosen so that the eigenvalues of its fixed point $X_{[\infty]}$ are *exactly* either 0 or 1 with $\text{Tr}(X_{[\infty]}) = N_e$; $X_{[\infty]}$ then satisfies (C.8) and it is thus equal to \tilde{D} , up to numerical error. In practice, the number of purification iterations required for convergence varies across different Hamiltonians and the numerical precision desired. Relevant factors include the size of the energy gap and the fraction of occupied to unoccupied states. Typically, less than 50 purification iterations are required [412]. Calculations performed to double precision tend to require more purification iterations, up to twice as many as the same calculation performed to single precision.

Details of the specific iteration we use are given in [412]. It can be reproduced by the

initialization

$$X_{[0]} = \beta_1 I + \beta_2 (\mu I - \tilde{H}), \quad (\text{C.9a})$$

$$\mu = \frac{\text{Tr} \tilde{H}}{N}, \quad (\text{C.9b})$$

$$\beta_1 = \frac{k/N}{e_+ - \mu}, \quad (\text{C.9c})$$

$$\beta_2 = \frac{1 - k/N}{\mu - e_-}, \quad (\text{C.9d})$$

where e_+ and e_- are estimates of the largest and smallest eigenvalues of \tilde{H} obtained by e.g. the Gershgorin circle theorem. Note that in practice we use the slightly more complicated initialization referred to as *HPCP+* in [412], which gives moderately improved performance when N_e is far from $N/2$. In either case, the iterate $X_{[n+1]}$ is found from its predecessor $X_{[n]}$ via

$$X'_{[n]} = I - X_{[n]}, \quad (\text{C.10a})$$

$$X_{[n+1]} = X_{[n]} + 2 \left(X_{[n]}^2 X'_{[n]} - \frac{\text{Tr}(X_{[n]}^2 X'_{[n]})}{\text{Tr}(X_{[n]} X'_{[n]})} X_{[n]} X'_{[n]} \right). \quad (\text{C.10b})$$

Once \tilde{D} is found, its counterpart in the non-orthogonal basis, D , is found by applying (C.3).

In Fig. C.1 we demonstrate the computational scaling of density matrix purification on TPUs using dense random Hermitian matrices and single precision. In this benchmark we scale both the system size (dimension of the matrix, N) and the number of TPU v3 cores used. Starting with a single TPU board, consisting of 8 TPU v3 cores, we can systematically scale up to hundreds (or thousands) of TPU v3 cores. Using a full TPU v3 pod (consisting of 2048 TPU v3 cores), we project that we can address dense systems of $N = 500\,000$ orbitals within 30 minutes using single precision.

For dense linear algebra, the computational scaling here is cubic. If suitable sparsity is assumed, and the density matrix is correspondingly truncated, sparse linear algebra can be

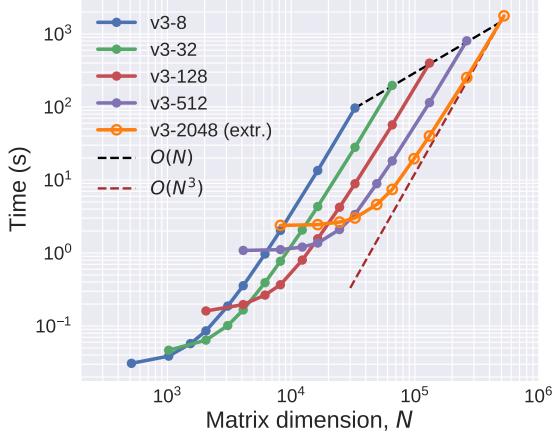


Figure C.1: Average wall times for TPU density matrix purification (normalized to a total of 50 purification iterations) in single precision using dense random Hermitian matrices of dimension N . Open-circle data points (v3-2048 results) are a linear extrapolation from v3-512 results.

used to obtain linear scaling. Such linear scaling approaches have been implemented and used in practice within computational quantum chemistry packages, however their practical application is limited to systems whose density matrix has a sufficient sparsity structure to ensure accurate results.

C.3 FHI-aims TPU integration details

We outline the practical details of integrating a TPU-based density matrix solver with the CPU-based DFT package FHI-aims. The platform and integration described here is a prototype. Its main purpose is to illustrate, in actual end-to-end DFT computations, the viability of accelerating the $O(N^3)$ bottleneck using TPUs.

The software ELSI [456, 472] is used to facilitate the connection between FHI-aims and the TPU by providing an interface and abstraction in which FHI-aims, or other codes, such as Siesta [133] and DFTB+ [177], can utilize external eigensolvers launched within ELSI. We implement in-house routines to launch the TPU-based density matrix purification (instead of

an eigensolver) using the ELSI standard. In all calculations, we use an off-the-shelf FHI-aims code with *no* modifications (version 210226).

In the course of a DFT calculation, FHI-aims utilizes the following matrices: the overlap matrix S , the DFT Hamiltonian H , and the density matrix D . FHI-aims distributes each matrix across CPU processes and memory using a 2D block-cyclic distribution pattern. On the other hand, in our current TPU implementation which utilizes SUMMA (Scalable Universal Matrix Multiplication Algorithm), our TPU-based solver requires matrices to be distributed across a TPU processor grid as 2D blocks in a *checkerboard* distribution, see [248] for more details. This poses a practical matrix communication challenge between the CPU-based and TPU-based schemes since their matrix distribution patterns differ in both cyclicity and the number of processors. A simple solution to communicate such matrices between CPU and TPU is to serialize and transfer the respective matrices and deserialize and redistribute them in the desired scheme. Specifically, we utilize available MPI processes on the CPU to serialize (and deserialize) to a network disk using the ELSI IO module and compressed sparse column (CSC) format with no cyclicity. Double precision is used throughout. We note that each process (CPU and TPU) calculates where its data should be within the serialized CSC representation of the whole matrix and reads (writes) to (from) only that portion of the matrix representation on the centralized network drive. That is, our implementation incurs some algorithmic overhead but only communicates the data that is needed.

This is not the most performant solution, however, it is generalizable, makes use of existing tooling within ELSI, and avoids the complexity of the various distribution patterns. Due to the use of the CSC format, serializing (writing) dense matrices to disk is especially costly and dominates the total CPU-TPU communication time for large system sizes. For transparency, in Fig. C.2 we plot the average observed total end-to-end CPU-TPU communication time (excluding the TPU density matrix purification time) incurred in our current implementation. There are, however, several ways to further optimize the current integra-

tion. For instance, we are currently using an off-the-shelf network file system (NFS) share, and replacing it with a different implementation of a portable operating system interface (POSIX) compliant distributed file system (one designed for high-performance applications) would result in a much higher throughput and would not require any changes to our code. In addition, further algorithmic optimizations are likely possible.

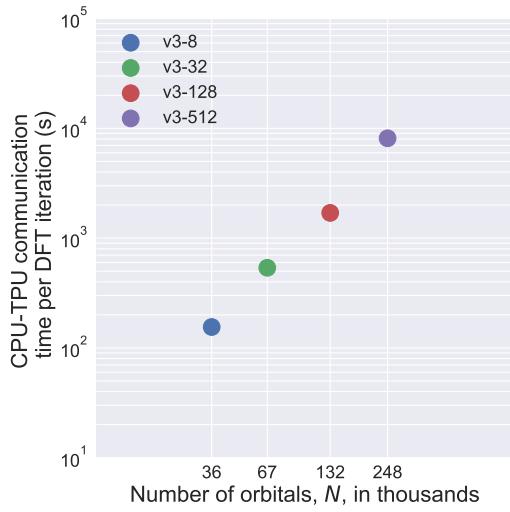


Figure C.2: Total end-to-end CPU-TPU communication time per DFT iteration (excluding the TPU density matrix purification time) incurred in current integration with FHI-aims.

Within FHI-aims all calculations are performed using the all-electron “light defaults” numeric atom-centered basis set [48]. This results in 5 basis functions per H atom and 14 basis functions per O atom in the water cluster calculations. All calculations are non-periodic with open boundary conditions and utilize the PBE [313] XC functional. The geometries of water clusters are directly obtained from Ref. [381] which were generated by taking spherical cutouts of varying radii from a large molecular dynamics simulation of bulk water at standard pressure and an average temperature of 300K (further details can be found in Ref. [358]).

In our implementation with FHI-aims, hybrid functionals can also be used without any modification, but simply result in longer DFT Hamiltonian build times on CPUs. Analytical forces are also available from FHI-aims using TPU-computed energy-weighted density

matrices, which are also communicated using the above scheme and facilitated using ELSI.

C.4 dynamic precision on smaller water clusters

The dynamic precision approach illustrated in Fig. 4 of the main text for 10 327 water molecules, when applied to smaller systems, allows for a larger part of the computation to be performed in single precision. For instance, an end-to-end converged DFT calculation on $(\text{H}_2\text{O})_{1481}$ cluster with $N = 35\,544$ orbitals required 11 iterations in single precision and 4 iterations in double precision, with an overall time of under 5 hours on a single TPU (v3-8) board, see Fig. C.3.

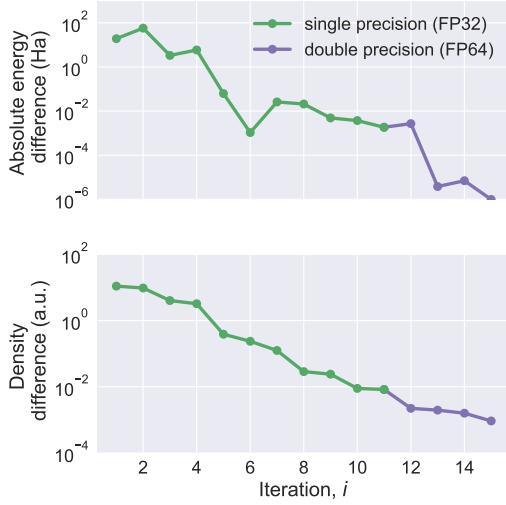


Figure C.3: Convergence trajectory of an end-to-end dynamic precision DFT calculation on a $(\text{H}_2\text{O})_{1481}$ cluster with $N = 35\,544$ orbitals. The absolute total energy differences between subsequent DFT iterations, i and $i - 1$, are plotted (top). The corresponding difference in real-space densities within the L^1 norm is plotted (bottom).