Orbital-Free Density Functional Theory with Continuous Normalizing Flows

Alexandre de Camargo
Department of Chemistry and Chemical Biology
McMaster University

Ricky T. Q. Chen
FAIR, Meta

Rodrigo A. Vargas-Hernández
Department of Chemistry and Chemical Biology
McMaster University
vargashr@mcmaster.ca

Abstract

Orbital-free density functional theory (OF-DFT) provides an alternative approach for calculating the molecular electronic energy, relying solely on the electron density. In OF-DFT, both the ground-state density is optimized variationally to minimize the total energy functional while satisfying the normalization constraint. In this work, we introduce a novel approach by parameterizing the electronic density with a normalizing flow ansatz, which is also optimized by minimizing the total energy functional. Our model successfully replicates the electronic density for a diverse range of chemical systems, including a one-dimensional diatomic molecule, specifically Lithium hydride with varying interatomic distances, as well as comprehensive simulations of hydrogen and water molecules, all conducted in Cartesian space.

1 Introduction

Unlike Kohn-Sham density functional theory (KS-DFT), which relies on molecular orbitals [1], Orbital-free density functional theory (OF-DFT) offers a distinctive computational approach within quantum chemistry and condensed matter physics.

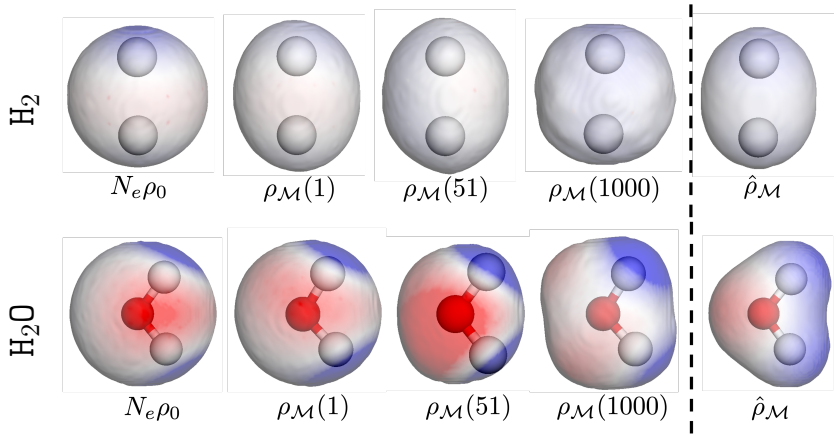


Figure 1: MEP for $\rho_{\mathcal{M}}(i)$ computed at different iterations i. $\hat{\rho}_{\mathcal{M}}$ and ρ_0 are defined in the main text.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

It finds its roots in the Hohenberg-Kohn theorems [2, 3], which unequivocally establish that the ground-state properties of a many-electron system can be ascertained through the minimization of an energy functional, denoted as $E[\rho]$, that operates solely on the electron density, ρ. This approach renders OF-DFT especially advantageous when grappling with vast and intricate systems, making it a valuable tool with wide-ranging applications in materials science [4].

For a system with N electrons subjected to an external potential (v_{ext}), the total energy functional is defined as $E[\rho] = F[\rho] + \int v_{\text{ext}}(\boldsymbol{x})\rho(\boldsymbol{x})\mathrm{d}\,\boldsymbol{x}$ [2, 3]. $F[\rho]$ represents the universal functional, with its precise form remaining unknown. Common approximations rely on kinetic ($T[\rho]$) and potential ($V[\rho]$) energy contributions, the specific forms of which also remain elusive. Numerous proposed functionals aim to approximate $T[\rho]$ [4–7], and machine learning (ML) algorithms have been employed for this purpose [8–12]. In contrast, within the context of KS-DFT, $T[\rho]$ is approximated as the sum of non-interacting particles [1, 3]. Researchers are continuously refining OF-DFT methods to enhance accuracy and broaden their applicability.

In OF-DFT, the ground-state density is found by solving a constrained optimization problem,

$$\min_{\rho(\boldsymbol{x})} E[\rho(\boldsymbol{x})] - \mu\left(\int \rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - \mathrm{N_e}\right) \text{ s.t. } \rho(\boldsymbol{x}) \geq 0, \tag{1}$$

where μ, also known as the chemical potential, acts as the Lagrange multiplier associated with the normalization constraint on the total number of particles ($\mathrm{N_e}$). These constraints, which enforce both positivity and normalization, ensure the attainment of physically valid solutions. Typically, conventional methods for solving for ρ involve self-consistent procedures based on functional derivatives, leading to the Euler equation $\delta E[\rho(\boldsymbol{x})]/\delta\rho(\boldsymbol{x}) - \mu = 0$ [3].

To represent molecular densities in real space, grid-based models are commonly employed [13–17]. An alternative real-space formulation was proposed by Chan et al. [18], which involves expressing the density as a linear combination of Gaussian basis functions (η_i), denoted as $\varphi(\boldsymbol{x}) = \sum_i C_i\eta_i(\boldsymbol{x})$, and ensuring positivity through $\rho_{\mathcal{M}}(\boldsymbol{x}) = \varphi^2(\boldsymbol{x})$. Both approaches, mesh and basis set, do not guarantee the normalization constraint; however, the latter method assists in the computation of essential integrals required for estimating functional derivatives and other relevant quantities.

In the present work, we present an alternative constrain-free approach to solve for the ground-state density using normalizing flows (NF).

2 Proposed Method

Our method consists of parametrizing the electronic density $\rho(\boldsymbol{x})$ with a Normalizing Flow. NFs are probabilistic models capable of transforming a simple density (ρ_0) into a potentially more complex target distribution, using the change of variables formula,

$$\rho_\phi(\mathbf{x}) = \rho_0(\mathbf{z})\left|\det\nabla_{\mathbf{z}}T_\phi(\mathbf{z})\right|^{-1}, \tag{2}$$

where $T_\phi(\cdot)$[1] is a bijective transformation, and $\rho_0(\mathbf{z})$ is the base distribution of the flow-based model. Eq. (2) guarantees the preservation of volume in the density transformation. To have a valid flow, we must have access to samples from ρ_0 and its probability density function (PDF) should be differentiable for optimization. A common approach is to parametrize $\mathrm{T}(\cdot)$ through a composition of functions; $T_\phi(\cdot) = T_K(\cdot) \circ \cdots \circ T_1(\cdot)$ [19–21]. These composable transformations can be considered as a flow discretized over time. A continuous-time formulation was proposed by Chen et al. [22], referred to as continuous normalizing flows (CNF). This framework is centered around the computation of the log density and $\mathrm{T}(\cdot)$ through an ordinary differential equation (ODE),

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\mathbf{z}(t) = g_\phi(\mathbf{z}(t), t), \text{ and } \tfrac{\mathrm{d}}{\mathrm{d}t}\log\rho(\mathbf{z}(t)) = -\nabla\cdot g_\phi, \tag{3}$$

where "$\nabla\cdot$" denotes the divergence operator. In this work, $g_\phi(\cdot)$ is a neural network composed of three hidden layers each with 512 neurons and the hyperbolic tangent activation function. Other architectures were tested but found to be sub-optimal.

Here, we reframe the OF-DFT variational problem as a Lagrangian-free optimization problem for molecular densities in the real space by parameterizing $\rho_{\mathcal{M}}(\boldsymbol{x})$ with a CNF ($\rho_\phi(\boldsymbol{x})$). Since the molecular density normalizes to $\mathrm{N_e}$, we define $\rho_{\mathcal{M}}(\boldsymbol{x})$ as, $\rho_{\mathcal{M}}(\boldsymbol{x}) := \mathrm{N_e}\,\rho_\phi(\boldsymbol{x})$, guaranteeing the satisfaction of the normalization constrain. $\rho_\phi(\boldsymbol{x})$ is also known as *shape factor* [3, 23].

[1] $T_\phi : \mathbb{R}^D \to \mathbb{R}^D$ an invertible and differentiable transformation.

The value of any functional that conforms to the total energy functional ($E[\rho_{\mathcal{M}}]$) can be rewritten in terms of an expectation over ρ_ϕ,

$$F[\rho_{\mathcal{M}}] = \int f(\boldsymbol{x}, \rho_{\mathcal{M}}, \nabla\, \rho_{\mathcal{M}})\, \rho_{\mathcal{M}}(\boldsymbol{x}) \mathrm{d}\, \boldsymbol{x} = (\mathrm{N_e})^p\, \mathbb{E}_{\rho_\phi}[f(\boldsymbol{x}, \rho_\phi, \nabla \rho_\phi)], \tag{4}$$

where $(\mathrm{N_e})^p$ is the constant factor related to the number of electrons, and $f(\boldsymbol{x}, \rho_\phi, \nabla\rho_\phi)$ is the integrand of the functional $F[\rho_{\mathcal{M}}]$. All functionals are estimated with Monte Carlo, where the samples are drawn from $\rho_0(\mathbf{z})$ and transformed by a CNF, $\boldsymbol{x} = T_\phi(\mathbf{z}) := \mathbf{z} + \int_{t_0}^{T} g_\phi(\mathbf{z}(t), t)\mathrm{d}t$.

Similar to any variational protocol in many-body physics, the minimization of total energy can be done through gradient-based algorithms using automatic differentiation (AD) [24–27]. AD for OF-DFT with grids was proposed by Tan et al. [28]. For all the results presented here, $\rho_0(\mathbf{z})$ is a Gaussian distribution with an identity covariance matrix. We found the `RMSProp` [29] algorithm with a learning rate of 3×10^{-4} to be the most appropriate optimizer. Code was developed using `JAX` Ecosystem [30, 31]. The code is available in the following GitHub Repository.

3 Results

In this section, we present the results for a one-dimensional model for diatomic models, specifically `LiH`, and the simulations for hydrogen and water molecules.

3.1 1-D model for diatomic molecules

Based on [32] work, we considered a one-dimensional model for diatomic molecules where the total energy functional is defined as,

$$E[\rho_{\mathcal{M}}] = T[\rho_{\mathcal{M}}] + V_{\mathrm{H}}[\rho_{\mathcal{M}}] + V_{\text{e-N}}[\rho_{\mathcal{M}}] + E_{\mathrm{X}}[\rho_{\mathcal{M}}]. \tag{5}$$

The total kinetic energy is estimated by the sum of the Thomas-Fermi (TF) (Eq. 6) and Weizsäcker (Eq. 7) functionals [3]; $T[\rho_{\mathcal{M}}] = T_{\mathrm{TF}}[\rho_{\mathcal{M}}] + T_{\mathrm{W}}[\rho_{\mathcal{M}}]$.

$$T_{\mathrm{TF}}[\rho_{\mathcal{M}}] = \frac{3}{10}(3\pi^2)^{\frac{2}{3}} \int (\rho_{\mathcal{M}}(\boldsymbol{x}))^{5/3}\, \mathrm{d}\, \boldsymbol{x} \tag{6}$$

and

$$T_{\mathrm{W}}[\rho_{\mathcal{M}}] = \frac{\lambda}{8} \int \frac{(\nabla\, \rho_{\mathcal{M}}(\boldsymbol{x}))^2}{\rho_{\mathcal{M}}} \mathrm{d}\, \boldsymbol{x}, \tag{7}$$

where the phenomenological parameter λ was set to 0.2 [18]. For one-dimensional systems T_{TF} is $T_{\mathrm{TF}}[\rho_{\mathcal{M}}] = \pi^2/24 \int (\rho_{\mathcal{M}}(x))^3\, \mathrm{d}x$ [32] and T_{W} and E_{X} have the same analytic form as in Eqs. 7 and 11 respectively. We rewrite $T_{\mathrm{W}}[\rho_{\mathcal{M}}]$ in terms of the score function; $T_{\mathrm{W}}[\rho_{\mathcal{M}}] = \frac{\lambda}{8} \int (\nabla \log \rho_{\mathcal{M}}(x))^2\, \rho_{\mathcal{M}}(x) \mathrm{d}x$, and in the CNF framework it can be computed by solving,

$$\tfrac{\mathrm{d}}{\mathrm{d}t} \nabla \log \rho_\phi = -\nabla^2 g_\phi - (\nabla \log \rho_\phi)^T \left(\nabla g_\phi(\mathbf{z}(t), t)\right), \tag{8}$$

where ∇^2 is the Laplacian operator. This allows us to make use of memory-efficient gradients [33] for optimizing $T_{\mathrm{W}}[\rho_{\mathcal{M}}]$.

The Hartree ($V_{\mathrm{H}}[\rho_{\mathcal{M}}]$) potential and the external potential ($V_{\text{e-N}}[\rho_{\mathcal{M}}]$) functionals both are defined by a soft version [32],

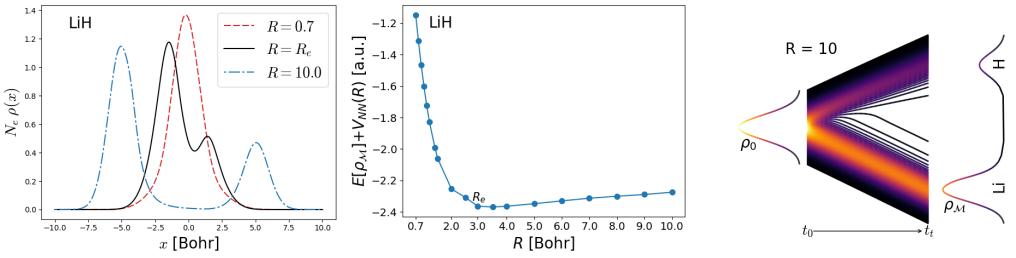
$$V_{\mathrm{H}}[\rho_{\mathcal{M}}] = \int\int v_{\mathrm{H}}(x)\, \rho_{\mathcal{M}}(x)\, \rho_{\mathcal{M}}(x') \mathrm{d}x\mathrm{d}x' = \int\int \tfrac{\rho_{\mathcal{M}}(x)\, \rho_{\mathcal{M}}(x')}{\sqrt{1+|x-x'|^2}} \mathrm{d}x\mathrm{d}x', \tag{9}$$

$$V_{\text{e-N}}[\rho_{\mathcal{M}}] = \int v_{\text{e-N}}(x)\, \rho_{\mathcal{M}}(x) \mathrm{d}x = -\int \left(\tfrac{Z_\alpha}{\sqrt{1+|x-R/2|^2}} + \tfrac{Z_\beta}{\sqrt{1+|x+R/2|^2}} \right) \rho_{\mathcal{M}}(x) \mathrm{d}x. \tag{10}$$

We only consider the Dirac exchange functional [3],

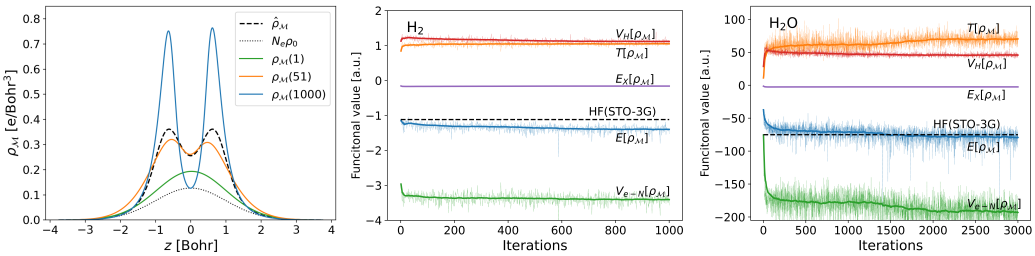
$$E_{\mathrm{X}}[\rho_{\mathcal{M}}] = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \int {\rho_{\mathcal{M}}}^{4/3}(\boldsymbol{x})\ \mathrm{d}\, \boldsymbol{x}\,. \tag{11}$$

We use the Lithium hydride molecule (`LiH`) as an example of diatomic molecule modeling in one dimension. For `LiH`, we utilize[1]: $Z_\alpha = 3$, $Z_\beta = 1$, $N_e = 2$ (valence electrons), and various nuclear distances R. In Fig. 2a, we present $\rho_{\mathcal{M}}$ for various R values, each one corresponding to the training of independent CNF models using only 2,000 iterations. Additionally, we display the potential energy surface curve in Fig. 2b, with the equilibrium bond distance found at $R_e = 2.95$ Bohr [32]. By exploring different R values, we validate the capability of our CNF ansatz for $\rho_{\mathcal{M}}$ to parameterize diverse chemical scenarios, ranging from strong nuclear interactions ($R < R_e$) to bond-breaking ($R \gg R_e$). The latter case is found to be challenging for existing DFT methodologies [34]. From Fig. 2a, it is evident that $\rho_{\mathcal{M}}$ does not distribute electron density along the inter-atomic axis. Fig. 2c we



(a) LiH density (b) Potential energy surface (c) $T_\phi(x(t))$

Figure 2: (a) Ground state electronic density of LiH for various inter-atomic distances R. (b) The value of the potential energy as a function of R is computed with $E[\rho_{\mathcal{M}}] + V_{\text{NN}}(R)$[2]. (c) The change of $x(t)$ and $\rho_\phi(x(t))$ through the transformation T_ϕ (Eq. 3). For all simulations, ρ_0 is a Gaussian distribution with $\sigma = 1$.



(a) H_2 density (b) H_2 training (c) H_2O training

Figure 3: (a) A cross-section of $\rho_{\mathcal{M}}(\texttt{i})$ at various iterations i. (b, c) The value of each density functional throughout the optimization, all values computed using an exponential moving average.

illustrate the transformation ρ_0 to $\rho_{\mathcal{M}}$ through the ODE (Eq. 3). Notably, for $R = 10$, $\rho_{\mathcal{M}}$ exhibits a bimodal distribution, with a higher concentration of electron density nearer to the Li nucleus.

3.2 3-D simulations

To demonstrate the capability to use CNF for real-space simulations, we considered the optimization of H_2 and H_2O. For both chemical systems, we considered the same $E[\rho_{\mathcal{M}}]$ where $v_{\text{e-N}}(\boldsymbol{x})$; $v_{\text{e-N}}(\boldsymbol{x}) = -\sum_i \frac{Z_i}{\|\boldsymbol{x}-\mathbf{R}_i\|}$[1]. For these simulations, no soft approximation was considered for the V_H and $V_{\text{e-N}}$ functionals.

Fig.1 illustrates the change in $\rho_{\mathcal{M}}$ through the optimization procedure for H_2 and H_2O. To further illustrate the change of $\rho_{\mathcal{M}}$ from ρ_0 by minimizing $E[\rho_{\mathcal{M}}]$, we computed the molecular electrostatic potential (MEP); $V_{\text{MEP}}(\boldsymbol{x}) = v_{\text{e-N}}(\boldsymbol{x}) - \int \rho_{\mathcal{M}}(\boldsymbol{x}')/\|\boldsymbol{x}-\boldsymbol{x}'\|\mathrm{d}\,\boldsymbol{x}'$. For both systems, ρ_0 is a multivariate Gaussian distribution located at zero, and the last layer of g_ϕ was initialized to zero. We compare our results with Hartree-Fock (HF) with STO-3G basis set ($\hat{\rho}_{\mathcal{M}}$). We monitored the value of $\int \rho_{\mathcal{M}}(\boldsymbol{x})\mathrm{d}\,\boldsymbol{x}$ using Becke's integration Scheme [14]; however, we found no significant difference w.r.t. $\mathrm{N_e}$, $\left|\int \rho_{\mathcal{M}}(\boldsymbol{x})\mathrm{d}\,\boldsymbol{x} - \mathrm{N_e}\right| < 10^{-4}$.

For the hydrogen molecule, the CNF model took 50 iterations to transport the density mass towards the H nuclei (Fig. 1), $R = 1.4$ Bohr. After 1,000 iterations we observe no significant difference in $E[\rho_{\mathcal{M}}]$ (Fig. 3b) and we observe a higher density around the nuclei compared to $\hat{\rho}_{\mathcal{M}}$, Fig. 3a.

For the optimization of the water molecule (Fig. 3c), our procedure was able to describe the two electron pairs, characteristic of this chemical system. We can also observe the change on ρ_0, a fully radial symmetric density, to a density with a higher electron density closer to the O atom. We set the maximum number of iterations to 3,000 with a 512 batch size, however, a ρ_0 closer to $\rho_{\mathcal{M}}$ could reduce the number of iterations.

[1] Z_i is the atomic number of atom i.
[2] $V_{\text{NN}}(R) = Z_\alpha Z_\beta/\sqrt{1+R^2}$

4 Summary

We have introduced a novel numerical procedure for parametrizing the ground-state energy of molecular densities using normalizing flows within the OF-DFT framework. This approach optimizes $\rho_{\mathcal{M}}$ variationally by minimizing the total energy $E[\rho_{\mathcal{M}}]$, marking it as the first constraint-free method that ensures both normalization and positivity (Eq. 1). Our study encompasses two types of simulations: one involving one-dimensional diatomic molecules and another involving a comprehensive real-space simulation of the H_2 and H_2O systems.

References

[1] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[2] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[3] R. G. Parr and W. Yang. Density functional theory of atoms and molecules. In K. Fukui and B. Pullman, editors, *Horizons of Quantum Chemistry*, pages 5–15, Dordrecht, 1980. Springer Netherlands. ISBN 978-94-009-9027-2.

[4] W. C. Witt, B. G. Del Rio, J. M. Dieterich and E. A. Carter. Orbital-free density functional theory for materials research. *Journal of Materials Research*, 33(7):777–795, 2018.

[5] S. Kumar, E. L. Borda, B. Sadigh, S. Zhu, S. Hamel, B. Gallagher, V. Bulatov, J. Klepeis and A. Samanta. Accurate parameterization of the kinetic energy functional. *The Journal of Chemical Physics*, 156(2), 2022.

[6] V. Bach and L. Delle Site. *Many-Electron Approaches in Physics, Chemistry and Mathematics*. Springer, 2014.

[7] E. A. Carter. Challenges in modeling materials properties without experimental input. *Science*, 321(5890):800–803, 2008.

[8] 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.

[9] 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):872, 2017.

[10] R. Meyer, M. Weichselbaum and A. W. Hauser. Machine learning approaches toward orbital-free density functional theory: Simultaneous training on the kinetic energy density functional and its functional derivative. *Journal of chemical theory and computation*, 16(9):5685–5694, 2020.

[11] F. Imoto, M. Imada and A. Oshiyama. Order-n orbital-free density-functional calculations with machine learning of functional derivatives for semiconductors and metals. *Physical Review Research*, 3(3):033198, 2021.

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

[13] J. J. Mortensen, L. B. Hansen and K. W. Jacobsen. Real-space grid implementation of the projector augmented wave method. *Physical review B*, 71(3):035109, 2005.

[14] A. D. Becke. A multicenter numerical integration scheme for polyatomic molecules. *The Journal of chemical physics*, 88(4):2547–2553, 1988.

[15] L. O. Wagner, EM Stoudenmire, K. Burke and S. R. White. Reference electronic structure calculations in one dimension. *Physical Chemistry Chemical Physics*, 14(24):8581–8590, 2012.

[16] H. Jiang, H. U. Baranger and W. Yang. Density-functional theory simulation of large quantum dots. *Physical Review B*, 68(16):165337, 2003.

[17] L.-Z. Bu and W. Wang. Efficient single-grid and multi-grid solvers for real-space orbital-free density functional theory. *Computer Physics Communications*, 290:108778, 2023.

[18] G. K.-L. Chan, A. J. Cohen and N. C. Handy. Thomas-Fermi-Dirac-von Weizsäcker models in finite systems. *The Journal of Chemical Physics*, 114(2):631–638, 2001.

[19] G. Papamakarios, E. Nalisnick, D. Jimenez Rezende, S. Mohamed and B. Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *The Journal of Machine Learning Research*, 22(1):2617–2680, 2021.

[20] D. Rezende and S. Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

[21] I. Kobyzev, S. JD Prince and M. A Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence*, 43(11): 3964–3979, 2020.

[22] R. T. Chen, Y. Rubanova, J. Bettencourt and D. K. Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[23] R. G. Parr and L. J. Bartolotti. Some remarks on the density functional theory of few-electron systems. *The Journal of Physical Chemistry*, 87(15):2810–2815, 1983. doi: 10.1021/j100238a023.

[24] J. M. Arrazola, S. Jahangiri, Delgado D., J. Ceroni, J. Izaac, A. Száva, U. Azad, R. A. Lang, Z. Niu, O. Di Matteo, R. Moyard, J. Soni, M. Schuld, R. A. Vargas-Hernández, T. Tamayo-Mendoza, C. Yen-Yu Lin, A. Aspuru-Guzik and N. Killoran. Differentiable quantum computational chemistry with pennylane, 2023.

[25] 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. doi: 10.1103/PhysRevLett.127.126403.

[26] T. Tamayo-Mendoza, C. Kreisbeck, R. Lindh and A. Aspuru-Guzik. Automatic differentiation in quantum chemistry with applications to fully variational hartree–fock. *ACS Central Science*, 4(5):559–566, 2018. doi: 10.1021/acscentsci.7b00586. PMID: 29806002.

[27] R. A. Vargas–Hernández, K. Jorner, R. Pollice and A. Aspuru–Guzik. Inverse molecular design and parameter optimization with Hückel theory using automatic differentiation. *The Journal of Chemical Physics*, 158(10):104801, 03 2023. ISSN 0021-9606. doi: 10.1063/5.0137103.

[28] C. W. Tan, C. J. Pickard and W. C. Witt. Automatic differentiation for orbital-free density functional theory. *The Journal of Chemical Physics*, 158(12), 2023.

[29] T. Tieleman, G. Hinton et al. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

[30] J. Bradbury, R. Frostig, P.Hawkins, M.J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne and Q. Zhang. JAX: composable transformations of Python+NumPy programs.

[31] J. Heek, A.Levskaya, A. Oliver, M. Ritter, B. Rondepierre, A. Steiner and M. van Zee. Flax: A neural network library and ecosystem for JAX.

[32] J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K.R. Müller and K. Burke. Orbital-free bond breaking via machine learning. *The Journal of chemical physics*, 139(22), 2013.

[33] C. Shuangshuang, D. Sihao, K. Yiannis and B. Mårten. Learning continuous normalizing flows for faster convergence to target distribution via ascent regularizations. In *The Eleventh International Conference on Learning Representations*, 2023.

[34] P. Mori-Sánchez, A. J. Cohen and W. Yang. Discontinuous nature of the exchange-correlation functional in strongly correlated systems. *Physical review letters*, 102(6):066403, 2009.