



UvA-DARE (Digital Academic Repository)

Learning Neural Free-Energy Functionals with Pair-Correlation Matching

Dijkman, J.; Dijkstra, M.; van Roij, R.; Welling, M.; van de Meent, J.-W.; Ensing, B.

DOI

[10.1103/PhysRevLett.134.056103](https://doi.org/10.1103/PhysRevLett.134.056103)

Publication date

2025

Document Version

Final published version

Published in

Physical Review Letters

License

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment>)

[Link to publication](#)

Citation for published version (APA):

Dijkman, J., Dijkstra, M., van Roij, R., Welling, M., van de Meent, J.-W., & Ensing, B. (2025). Learning Neural Free-Energy Functionals with Pair-Correlation Matching. *Physical Review Letters*, 134(5), Article 056103. <https://doi.org/10.1103/PhysRevLett.134.056103>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Learning Neural Free-Energy Functionals with Pair-Correlation Matching

Jacobus Dijkman^{1,2}, Marjolein Dijkstra³, René van Roij⁴, Max Welling²,
Jan-Willem van de Meent², and Bernd Ensing^{1,5}


¹*Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, The Netherlands*

²*Informatics Institute, University of Amsterdam, The Netherlands*

³*Soft Condensed Matter & Biophysics, Debye Institute for Nanomaterials Science, Utrecht University, The Netherlands*

⁴*Institute for Theoretical Physics, Utrecht University, The Netherlands*

⁵*AI4Science Laboratory, University of Amsterdam, The Netherlands*

 (Received 21 May 2024; accepted 10 January 2025; published 7 February 2025; corrected 6 March 2025)

The intrinsic Helmholtz free-energy functional, the centerpiece of classical density functional theory, is at best only known approximately for 3D systems. Here we introduce a method for learning a neural-network approximation of this functional by exclusively training on a dataset of radial distribution functions, circumventing the need to sample costly heterogeneous density profiles in a wide variety of external potentials. For a supercritical Lennard-Jones system with planar symmetry, we demonstrate that the learned neural free-energy functional accurately predicts inhomogeneous density profiles under various complex external potentials obtained from simulations.

DOI: [10.1103/PhysRevLett.134.056103](https://doi.org/10.1103/PhysRevLett.134.056103)

Inhomogeneous many-body systems play a profound role in both science and technology, with examples spanning from p - n junctions in semiconducting solid-state devices to phenomena like hydrogen bubbles in electrolyzers, or gas adsorption in porous materials [1–6]. Density functional theory (DFT) is a powerful theoretical framework to describe the thermodynamic equilibrium properties and the structure of such systems, relying solely on the one-body density profile $\rho(\mathbf{r})$ [7]. Classical DFT (cDFT) relies on the existence of an excess free-energy functional $\mathcal{F}_{\text{exc}}[\rho]$, which describes the nonideal contribution to the total intrinsic free-energy functional $\mathcal{F}[\rho]$ and encompasses the interparticle interactions. However, the main bottleneck of cDFT is that this functional is in general unknown, and hence, one has to rely on developing accurate approximations for the excess free-energy functionals.

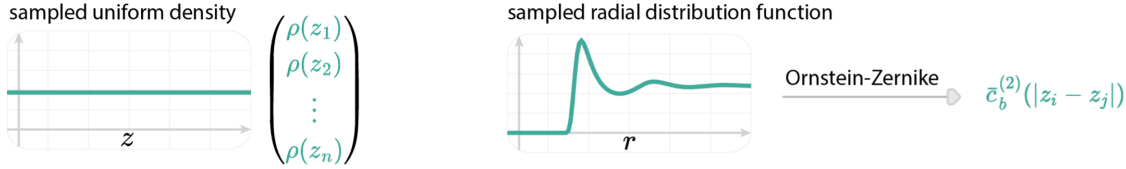
Historically, the field of cDFT has emphasized development of analytical approximations for $\mathcal{F}_{\text{exc}}[\rho]$, often based on thermodynamics and (direct) pair correlations derived from *approximate* closures of the Ornstein-Zernike equation of bulk systems [8]. For instance, the highly successful fundamental measure theory (FMT) for hard spheres is deeply connected to the Percus-Yevick closure [9], and many functionals for systems with soft van der Waals or Coulombic interactions build on mean-field and mean-spherical approximations [8,10].

In recent years, there has been a resurgence of cDFT developments facilitated by machine learning (ML) methods, which employ virtually *exact* thermodynamic and structural data obtained from explicit many-body simulations to learn data-driven representations of the

excess free-energy functional $\mathcal{F}_{\text{exc}}[\rho]$. In the classical regime, the first machine-learned cDFTs focused on supercritical Lennard-Jones fluids, for which explicit approximate functional forms for $\mathcal{F}_{\text{exc}}[\rho]$ were fitted to density profiles in external fields obtained from simulations, both for 1D [11] and 3D systems in planar geometry [12]. Recent work, once again leveraging simulations of density profiles in a variety of external potentials, has shown that a neural approximation of the functional derivative $\delta\mathcal{F}_{\text{exc}}/\delta\rho$ for hard-sphere systems outperforms FMT [13] in accurately estimating inhomogeneous density profiles.

In this Letter, we introduce a neural free-energy functional that we train using *pair-correlation matching*, a novel optimization objective that matches the Hessian of the neural approximation to pair correlations of particles. We show that pair-correlation matching yields a neural functional that accurately predicts the excess free energy for a one-component system of interacting particles. The differentiable nature of this neural functional allows access to, and learning from, various structural and thermodynamic properties by utilizing the first and second functional derivatives of $\mathcal{F}_{\text{exc}}[\rho]$. Unlike previous ML approaches to cDFT [11–14], our neural functional is trained by directly learning particle correlations from radial distribution functions sampled from short simulations of homogeneous bulk systems (illustrated in Fig. 1), rather than inferring them from a costly dataset of inhomogeneous densities. We demonstrate that this neural free-energy functional can be applied in the cDFT framework to achieve accurate estimates for inhomogeneous density profiles in external fields without ever having seen any inhomogeneous

(a) Generate training data from MC simulations

■ simulation data


(b) Train the neural network with pair-correlation matching

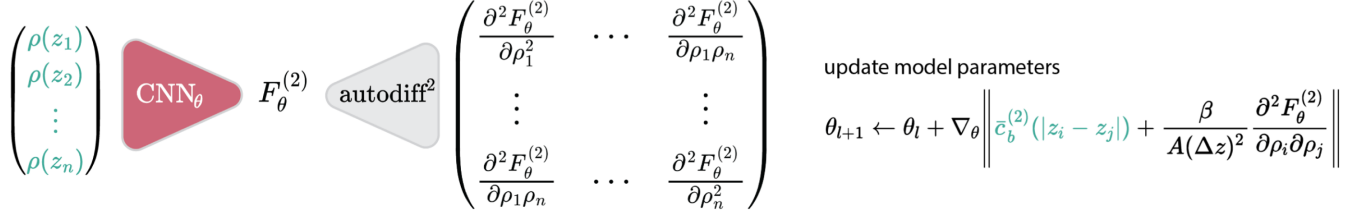

 (c) Classical DFT at fixed $\beta\mu$ and $\beta V^{\text{ext}}(z)$


FIG. 1. (a) Bulk densities in planar geometry $\rho(z_i) = \rho_b$ and radial distribution functions $g(r)$ are sampled from Monte Carlo simulations of homogeneous bulk systems of Lennard-Jones particles. Each $g(r)$ is converted to the second functional derivative of the excess free energy $\delta^2 \mathcal{F}_{\text{exc}} / \delta \rho(z_i) \delta \rho(z_j)$ by employing the Ornstein-Zernike equation. (b) Through automatic differentiation (autodiff²), the neural functional $F_\theta^{(2)}$ is optimized to fit the Hessian of the model output with respect to input density profiles to $\delta^2 \mathcal{F}_{\text{exc}} / \delta \rho(z_i) \delta \rho(z_j)$. (c) The optimized model can then be applied in cDFT to obtain nonuniform equilibrium density profiles through automatic differentiation (autodiff) and the free energy $F_\theta^{(2)}$ for a system of Lennard-Jones particles subjected to arbitrary external potentials.

densities during training. Simultaneously, the neural free-energy functional provides accurate estimates of the excess free energy.

Central to classical DFT is the grand canonical equilibrium density

$$\rho_0(\mathbf{r}) = \frac{1}{\Lambda^3} \exp \left(\beta \mu - \beta \left. \frac{\delta \mathcal{F}_{\text{exc}}[\rho]}{\delta \rho(\mathbf{r})} \right|_{\rho=\rho_0} - \beta V_{\text{ext}}(\mathbf{r}) \right), \quad (1)$$

with $\beta = 1/k_B T$, Λ the thermal wavelength, μ the chemical potential, and $V_{\text{ext}}(\mathbf{r})$ the external potential. This self-consistency relation can be leveraged to find $\rho_0(\mathbf{r})$ through recursive Picard iteration [15–17]. Since $\rho_0(\mathbf{r})$ is dependent on $\mathcal{F}_{\text{exc}}[\rho]$ through the first functional derivative $\delta \mathcal{F}_{\text{exc}}[\rho] / \delta \rho(\mathbf{r})$, previous approaches to leveraging machine learning for cDFT [11–13] involve training a model to capture $\delta \mathcal{F}_{\text{exc}} / \delta \rho(\mathbf{r})$, which can be derived from sampled inhomogeneous equilibrium density profiles and employing Eq. (1).

We train a convolutional neural network to directly learn the excess free energy $\mathcal{F}_{\text{exc}}[\rho]$, enabling the straightforward calculation of functional derivatives by (auto)differentiating

the neural functional with respect to its inputs. We focus on 3D systems in a planar geometry, where the excess free-energy of a system of area A is a functional of the density $\rho(z)$, which is constant across any plane parallel to the xy plane, i.e., $\rho(z) = \rho(x, y, z)$ with $\rho(x, y, z) = \rho(x', y', z)$ for all $(x, y), (x', y')$ within the confines of A . We represent the excess free-energy functional $\mathcal{F}_{\text{exc}}[\rho]$ as a neural network $F_\theta^{(2)}(\rho_1, \dots, \rho_n)$, with ρ_i the density at grid point z_i for $i \in \{1, \dots, n\}$, network parameters θ , and where the upper index “(2)” indicates that the neural network is optimized for pair correlations of bulk systems obtained from simulations.

The functional derivative $\delta \mathcal{F}_{\text{exc}} / \delta \rho(z_i)$ is defined as the limit of the partial derivative $\lim_{\Delta z \rightarrow 0} (1/\Delta z) \partial \mathcal{F}_{\text{exc}} / \partial \rho_i$ with $\rho_i = \rho(z_i)$ and Δz the (uniform) grid spacing. We leverage this relationship by employing automatic differentiation (autodiff) [18,19] to approximate the first and second functional derivatives of $\mathcal{F}_{\text{exc}}[\rho]$ on a finite grid by $(1/\Delta z) \partial F_\theta^{(2)} / \partial \rho_i$ and $(1/\Delta z)^2 \partial^2 F_\theta^{(2)} / \partial \rho_i \partial \rho_j$, respectively. Here we note that the second functional derivative of $\mathcal{F}_{\text{exc}}[\rho]$ in planar geometry is related to the laterally integrated direct correlation function $\bar{c}_b^{(2)}(|z_i - z_j|)$ through

$$\begin{aligned} \frac{-\beta}{A} \frac{\delta^2 \mathcal{F}_{\text{exc}}[\rho]}{\delta \rho(z_i) \delta \rho(z_j)} &= \bar{c}_b^{(2)}(|z_i - z_j|) \\ &\equiv \int_{|z_i - z_j|}^{\infty} dr 2\pi r c_b^{(2)}(r), \end{aligned} \quad (2)$$

where the direct correlation function $c_b^{(2)}(r)$ of homogeneous bulk fluids at density ρ_b is obtained from simulated radial distribution functions $g(r)$ through the bulk Ornstein-Zernike equation

$$c_b^{(2)}(r) = \frac{1}{2\pi^2} \int_0^{\infty} \frac{\sin(kr)}{kr} \left(\frac{\hat{h}(k)}{1 + \rho_b \hat{h}(k)} \right) k^2 dk, \quad (3)$$

with $\hat{h}(k)$ the Fourier transform of the total correlation function $h(r) = g(r) - 1$ [8]. Thus we train our neural network such that its Hessian $\partial^2 F_{\theta}^{(2)}/\partial \rho_i \partial \rho_j$ represents well the laterally integrated direct correlation function obtained from homogeneous bulk simulations (see End Matter for further details). We will refer to this approach as pair-correlation matching.

We illustrate this methodology for systems interacting with a Lennard-Jones potential truncated at $r_{\text{cut}} = 4\sigma$ with σ the particle diameter, shifted upward by $\epsilon_{\text{cut}} = 0.98 \times 10^{-3}\epsilon$ with ϵ the well depth, at a temperature $k_B T/\epsilon = 2$, i.e., above the critical point. Setting $\Lambda = \sigma$ throughout, we perform grand-canonical Monte Carlo (GCMC) simulations of homogeneous bulk systems at a variety of chemical potentials $\beta\mu \in [-4, 0.5]$, resulting in bulk densities $\rho_b \sigma^3 \in [0.02, 0.67]$. We employ 10^9 trial moves in a cubic box with an edge length $L = 10\sigma$ (hence with area $A = L^2$ and volume $V = L^3$) and apply periodic boundary conditions. We sample $g(r)$ and ρ_b and convert each $g(r)$ to $\bar{c}_b^{(2)}(z)$ using Eqs. (2) and (3) with a grid spacing $\Delta z = \sigma/32$. We create a data set of 800 combinations of $\bar{c}_b^{(2)}(z)$ and ρ_b , from which we learn the parameters θ of $F_{\theta}^{(2)}$. The neural network architecture of $F_{\theta}^{(2)}$ is a six-layer convolutional neural network with periodic padding [20] to match the system's periodic boundary conditions. Each convolution uses a kernel size of 3 with a dilation rate of 3. The number of channels per hidden layer is configured as follows: [16, 16, 32, 32, 64, 64], such that θ consists of 24.4 K parameters (see End Matter for further details).

To evaluate the accuracy of our neural excess free-energy functional $F_{\theta}^{(2)}$, we compare it to the van der Waals-like mean-field approximation $F_{\text{exc}}^{\text{MF}}$, which treats the attractions of Lennard-Jones particles as a perturbation on the hard-sphere system, as implemented in `PyDFTIj` [1]. We use the White-Bear mark II version of FMT for the excess free energy of the hard-sphere system [21]. Additionally, we compare to $F_{\theta}^{(1)}$, which is a neural free-energy functional trained by minimization of the error between

$(1/\Delta z)\partial F_{\theta}^{(1)}/\partial \rho_i$ and $\delta \mathcal{F}_{\text{exc}}/\delta \rho(z_i)$ rather than by pair-correlation matching. This neural functional has the same neural network architecture as $F_{\theta}^{(2)}$ and is trained on a dataset of 800 nonuniform densities, subjected to the same set of chemical potentials as before (see the Supplemental Material, S. 2 [22]). Note that the free-energy functional $F_{\theta}^{(1)}$ introduced in this Letter is different from the neural direct correlation functional introduced by Sammüller *et al.* [13], since the gradient of $F_{\theta}^{(1)}$ supplies a global estimate of $\delta \mathcal{F}_{\text{exc}}/\delta \rho(z_i)$ for all $i \in \{1, \dots, n\}$, in contrast to a local estimate at one position z_i (see the Supplemental Material, S. 8 [22]). By approximating $\delta \mathcal{F}_{\text{exc}}/\delta \rho(z_i)$ by the gradient $(1/\Delta z)\partial F_{\theta}^{(n)}/\partial \rho_i$ for $n = 1$ and 2, both neural functionals are applied in Picard iterations [15–17] to obtain DFT estimates for the equilibrium density profiles of inhomogeneous systems according to Eq. (1).

The DFT results for an exemplary external potential at $\beta\mu = 0$ are shown in Fig. 2(a). We observe that the neural functionals $F_{\theta}^{(1)}$ and $F_{\theta}^{(2)}$ provide similar estimates that closely agree with simulation data. For the same external potential, we evaluate the accuracy of DFT estimates for the free energy for a range of chemical potentials $-4 < \beta\mu < 6$ [Fig. 2(b)]. We compare with the excess free energy obtained from GCMC simulations through thermodynamic integration (see the Supplemental Material, S. 4 [22]). We observe that both neural functionals outperform $F_{\text{exc}}^{\text{MF}}$ within the range of μ values in the training set, exhibiting good agreement with simulations. The DFT estimates are shown until the Picard iterations diverge. We observe that $F_{\theta}^{(1)}$ diverges rapidly when extrapolating beyond the training set, even earlier than $F_{\text{exc}}^{\text{MF}}$. In contrast, $F_{\theta}^{(2)}$ converges to a solution far beyond the trained μ range.

For a more detailed comparison of the accuracy of the free-energy functionals for various inhomogeneous systems, we performed separate DFT calculations for 150 distinct external potentials, evaluated across the range $-4 < \beta\mu < 6$. Both $F_{\theta}^{(1)}$ and $F_{\theta}^{(2)}$ functionals show excellent agreement with simulated data for density estimates [Fig. 2(c)] and free-energy estimates [Fig. 2(d)], outperforming $F_{\text{exc}}^{\text{MF}}$ across all evaluated external potentials. Within the training range, $F_{\theta}^{(1)}$ exhibits marginally lower prediction errors than $F_{\theta}^{(2)}$, while $F_{\theta}^{(2)}$ demonstrates superior accuracy when extrapolating beyond the training range. Additional experiments show that the predictive accuracy of $F_{\theta}^{(2)}$ is more sensitive to highly inhomogeneous densities compared to $F_{\theta}^{(1)}$, though this effect is reduced when incorporating higher bulk densities into the training set of $F_{\theta}^{(2)}$ (see the Supplemental Material, S. 7 [22]). Although we might expect that training on bulk data provides significantly less information about the underlying functional than training on inhomogeneous data, these results

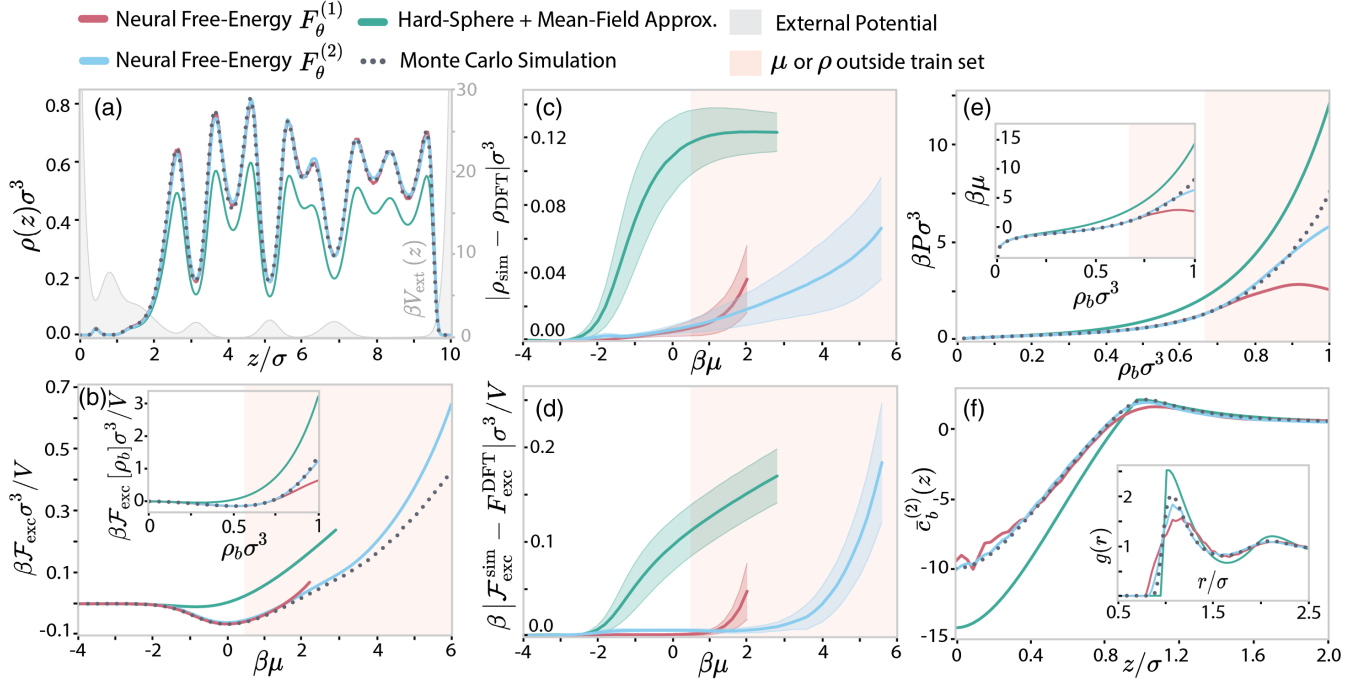


FIG. 2. Evaluation of neural free-energy functionals $F_\theta^{(1)}$ and $F_\theta^{(2)}$, where $F_\theta^{(1)}$ is optimized by matching inhomogeneous one-body densities and $F_\theta^{(2)}$ by pair-correlation matching in the homogeneous bulk. (a) Density profiles of a Lennard-Jones system in a planar geometry characterized by an external potential (shown in gray) at a chemical potential of $\beta\mu = 0$ obtained from DFT using $F_\theta^{(1)}$ and $F_\theta^{(2)}$ and the mean-field approximation $F_{\text{exc}}^{\text{MF}}$, along with the simulated density profile. (b) Comparison of the free-energy estimates using $F_\theta^{(1)}$, $F_\theta^{(2)}$, and $F_{\text{exc}}^{\text{MF}}$ for the specific external potential shown in (a), evaluated across the range $-4 < \beta\mu < 6$. Inset shows the bulk excess free-energy obtained from DFT and simulations across a range of bulk densities $0 < \rho_b\sigma^3 < 1$. (c) Mean absolute error $(1/n) \sum |\rho_{\text{sim}}(z_i) - \rho_{\text{DFT}}(z_i)|$ between density profiles obtained from DFT using $F_\theta^{(1)}$, $F_\theta^{(2)}$, and $F_{\text{exc}}^{\text{MF}}$ and densities sampled from simulation. (d) Absolute error of the excess free-energy from DFT and simulations. (c),(d) Data is shown for 150 distinct external potentials, evaluated across the range $-4 < \beta\mu < 6$, with steps of $\Delta\beta\mu = 0.2$. The area of the mean \pm standard deviation is colored. The error is shown up to the point where the Picard iterations stop to converge to a solution within 1000 iterations. (e) The bulk pressure and chemical potential obtained from DFT and simulations across a range of bulk densities $0 < \rho_b\sigma^3 < 1$. (f) The laterally integrated direct correlation function $\bar{c}_b^{(2)}(z)$ at $\rho_b\sigma^3 = 0.67$ and the radial distribution function $g(r)$ obtained from simulation and DFT.

show that pair-correlation matching works remarkably well in practice.

Additionally, the free-energy functionals can be applied in a uniform density setting to obtain access to the bulk pressure

$$P(\rho_b) = \left(\frac{\delta\mathcal{F}_{\text{exc}}}{\delta\rho} \Big|_{\rho_b} + k_B T \right) \rho_b - \mathcal{F}_{\text{exc}}[\rho_b]/V, \quad (4)$$

and the chemical potential μ following Eq. (1). Again, we find excellent agreement of both neural functionals with simulations within the training set, and superior agreement of $F_\theta^{(2)}$ at higher densities, as shown in Fig. 2(e). Lastly, we demonstrate that $F_\theta^{(2)}$ compared to both $F_\theta^{(1)}$ and $F_{\text{exc}}^{\text{MF}}$ provides accurate estimates for the laterally integrated direct correlation function $\bar{c}_b^{(2)}(z)$ and the radial distribution function $g(r)$ as shown in Fig. 2(f) for $\rho_b\sigma^3 = 0.67$. Here, we approximate $\bar{c}_b^{(2)}(z)$ by $[-\beta/A(\Delta z)^2] \partial^2 F_\theta / \partial \rho_i \partial \rho_j$

$z = |z_i - z_j|$. To derive the radial distribution function, we first numerically calculate $c_b^{(2)}(r)$ from $\bar{c}_b^{(2)}(z)$ using

$$-\frac{1}{2\pi} \left(\frac{1}{z} \frac{d\bar{c}_b^{(2)}(z)}{dz} \right) \Big|_{z=r} = c_b^{(2)}(r). \quad (5)$$

We then obtain $g(r)$ using the Ornstein-Zernike equation [Eq. (3)]. To suppress artifacts stemming from numerical transformations in the region for $g(r) \rightarrow 0$, we apply a noise-reducing filter (see the Supplemental Material, S. 5 [22]).

Our results lead to the surprising observation that the neural free-energy functional $F_\theta^{(2)}$ is robust and predicts accurate nonuniform density profiles, even substantially beyond its training range and solely by training on pair correlation functions of homogeneous bulk systems. This training on homogeneous bulk systems offers an alternative to existing training schemes of classical DFT, which so far have all been based on training on density

profiles of inhomogeneous fluids in a variety of external fields [11–14]. For pairwise systems, this alternative could be particularly useful if pair correlations are the only data available (e.g., from scattering experiments on homogeneous fluids), if the external potential leads to computationally expensive 3D density profiles, or if spatial heterogeneity is coupled to orientational anisotropy (e.g., in liquid crystals, nonspherical molecules, or patchy particles [14]). The presently proposed training on homogeneous bulk systems can also be a computationally inexpensive addition to the existing training protocols on inhomogeneous fluids, where an optimal combination may be found between training on homogeneous and inhomogeneous states. In this sense, the present study can be seen as an extreme case of training on pair correlations in the bulk only.

In conclusion, this Letter introduces a generic machine learning approach to obtain classical free-energy functionals through pair-correlation matching, through which we attain a neural free-energy functional that enables simultaneous and direct access to both the excess free energy and the density of a classical supercritical system of interacting particles in any external environment.

Acknowledgments—The authors would like to thank Kieron Burke, Florian Sammüller, Matthias Schmidt, and Robert Evans for valuable discussions. The authors acknowledge the University of Amsterdam Data Science Centre for financial support. M. D. acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Program (Grant Agreement No. ERC-2019-ADG 884902 SoftML). J. W. M. acknowledges funding from the European Union’s Horizon Framework Programme (Grant Agreement ID No. 101120237).

[1] E. d. A. Soares, A. G. Barreto Jr., and F. W. Tavares, Classical density functional theory reveals structural information of H₂ and CH₄ fluids adsorbed in MOF-5, [arXiv:2303.11384](https://arxiv.org/abs/2303.11384).

[2] F. Guo, Y. Liu, J. Hu, H. Liu, and Y. Hu, Classical density functional theory for gas separation in nanoporous materials and its application to CH₄/H₂ separation, *Chem. Eng. Sci.* **149**, 14 (2016).

[3] F. Guo, Y. Liu, J. Hu, H. Liu, and Y. Hu, Fast screening of porous materials for noble gas adsorption and separation: A classical density functional approach, *Phys. Chem. Chem. Phys.* **20**, 28193 (2018).

[4] F. Guo, Y. Liu, J. Hu, H. Liu, and Y. Hu, Screening of porous materials for toxic gas adsorption: Classical density functional approach, *Ind. Eng. Chem. Res.* **59**, 14364 (2020).

[5] J. Fu, Y. Tian, and J. Wu, Classical density functional theory for methane adsorption in metal-organic framework materials, *AIChE J.* **61**, 3012 (2015).

[6] Y. Liu and H. Liu, Classical density functional theory for fluids adsorption in MOFs, in *Metal-Organic Frameworks*, edited by F. Zafar and E. Sharmin (IntechOpen Limited, London, 2016).

[7] R. Evans, The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids, *Adv. Phys.* **28**, 143 (1979).

[8] J.-P. Hansen and I. R. McDonald, in *Theory of Simple Liquids* (Elsevier, New York, 2013).

[9] R. Roth, Fundamental measure theory for hard-sphere mixtures: A review, *J. Phys. Condens. Matter* **22**, 063102 (2010).

[10] P. Cats, R. Evans, A. Härtel, and R. Van Roij, Primitive model electrolytes in the near and far field: Decay lengths from DFT and simulations, *J. Chem. Phys.* **154**, 124504 (2021).

[11] S. C. Lin and M. Oettel, A classical density functional from machine learning and a convolutional neural network, *SciPost Phys.* **6**, 025 (2019).

[12] P. Cats, S. Kuipers, S. De Wind, R. Van Damme, G. M. Coli, M. Dijkstra, and R. Van Roij, Machine-learning free-energy functionals using density profiles from simulations, *APL Mater.* **9**, 031109 (2021).

[13] F. Sammüller, S. Hermann, D. De Las Heras, and M. Schmidt, Neural functional theory for inhomogeneous fluids: Fundamentals and applications, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2312484120 (2023).

[14] A. Simon, J. Weimar, G. Martius, and M. Oettel, Machine learning of a density functional for anisotropic patchy particles, *J. Chem. Theory Comput.* **20**, 1062 (2024).

[15] R. Roth, Introduction to Density Functional Theory of Classical Systems: Theory and Applications, Lecture Notes (2006), https://bytebucket.org/knepley/cdft-git/wiki/papers/Lecture_Notes_on_DFT_Roland_Roth.pdf.

[16] M. Edelmann and R. Roth, A numerical efficient way to minimize classical density functional theory, *J. Chem. Phys.* **144**, 074105 (2016).

[17] J. Mairhofer and J. Gross, Numerical aspects of classical density functional theory for one-dimensional vapor-liquid interfaces, *Fluid Phase Equilib.* **444**, 1 (2017).

[18] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, Automatic differentiation in machine learning: A survey, *J. Mach. Learn. Res.* **18**, 1 (2018), <https://www.jmlr.org/papers/v18/17-468.html>.

[19] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, Automatic differentiation in PyTorch, *NIPS Autodiff Workshop* (2017), <https://openreview.net/forum?id=BJJsrnfCZ>.

[20] A. Alguacil, W. G. Pinto, M. Bauerheim, M. C. Jacob, and S. Moreau, Effects of boundary conditions in fully convolutional networks for learning spatio-temporal dynamics, [arXiv:2106.11160](https://arxiv.org/abs/2106.11160).

[21] H. Hansen-Goos and R. Roth, Density functional theory for hard-sphere mixtures: The White Bear version mark II, *J. Phys. Condens. Matter* **18**, 8413 (2006).

[22] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.134.056103> for details on the presented methods and additional results, with references to Ref. [23].

[23] F. Sammüller and M. Schmidt, Neural density functionals: Local learning and pair-correlation matching, *Phys. Rev. E* **110**, L032601 (2024).

Correction: A statement of thanks has been added to the Acknowledgments.

End Matter

End matter: Details of pair-correlation matching—In this work, we train a convolutional neural network to directly approximate the excess free energy $\mathcal{F}_{\text{exc}}[\rho]$, enabling the straightforward calculation of functional derivatives by (auto)differentiating the neural functional with respect to its inputs. Rather than approximating the free energy directly, we train the neural functional by optimizing an objective that matches the Hessian of the network to a direct correlation function that is computed from short simulations of homogeneous bulk systems. We refer to this approach as pair-correlation matching.

We focus on 3D systems in a planar geometry, where the excess free energy of a system of area A is a functional of the density $\rho(z)$, which is constant across any plane parallel to the xy plane, i.e., $\rho(z) = \rho(x, y, z)$ with $\rho(x, y, z) = \rho(x', y', z)$ for all $(x, y), (x', y')$ within the confines of A . We represent the excess free-energy functional $\mathcal{F}_{\text{exc}}[\rho]$ as a neural network $F_{\theta}^{(2)}(\rho_1, \dots, \rho_n)$, with ρ_i the density at grid point z_i for $i \in \{1, \dots, n\}$, network parameters θ , and where the upper index “(2)” indicates that the neural network is optimized for pair correlations of bulk systems as obtained from simulations.

The functional derivative $\delta\mathcal{F}_{\text{exc}}/\delta\rho$ evaluated at z_i is defined as the limit of the partial derivative $\lim_{\Delta z \rightarrow 0} (1/\Delta z) \partial\mathcal{F}_{\text{exc}}/\partial\rho_i$ with $\rho_i = \rho(z_i)$ and Δz the (uniform) grid spacing. We leverage this relationship by employing autodiff to approximate the first and second functional derivatives of $\mathcal{F}_{\text{exc}}[\rho]$ on a finite grid by $(1/\Delta z) \partial F_{\theta}^{(2)}/\partial\rho_i$ and $(1/\Delta z)^2 \partial^2 F_{\theta}^{(2)}/\partial\rho_i \partial\rho_j$, respectively. We target the approximation of the second functional derivative in homogeneous bulk systems with bulk densities ρ_b as our primary optimization objective, i.e., we minimize the difference between $(1/\Delta z)^2 \partial^2 F_{\theta}^{(2)}/\partial\rho_i \partial\rho_j|_{\rho_b}$ and $\delta^2\mathcal{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)|_{\rho_b}$.

To train our neural functional according to this optimization objective, we require ground-truth examples of $\delta^2\mathcal{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)|_{\rho_b}$. We obtain these examples by sampling the radial distribution function $g(r)$ from simulations of homogeneous bulk systems and applying the Ornstein-Zernike equation to obtain $\delta^2\mathcal{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)|_{\rho_b}$ from $g(r)$. Key to this transformation is the two-body direct correlation function, which is defined in terms of the second functional derivative as

$$c^{(2)}(\mathbf{r}, \mathbf{r}') = -\beta \frac{\delta^2\mathcal{F}_{\text{exc}}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}, \quad (\text{A1})$$

for systems with arbitrary geometry. In a uniform and isotropic bulk fluid, $c^{(2)}(\mathbf{r}, \mathbf{r}') = c_b^{(2)}(r)$ only depends on the distance between two points,

$$r = |\mathbf{r} - \mathbf{r}'| = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}. \quad (\text{A2})$$

In such a system, we obtain the direct correlation function $c_b^{(2)}(r)$ from the radial distribution function $g(r)$ by first calculating the total correlation function $h(r) = g(r) - 1$ and then applying the bulk Ornstein-Zernike equation

$$c_b^{(2)}(r) = \frac{1}{2\pi^2} \int_0^\infty \frac{\sin(kr)}{kr} \left(\frac{\hat{h}(k)}{1 + \rho\hat{h}(k)} \right) k^2 dk, \quad (\text{A3})$$

with $\hat{h}(k)$ the Fourier transform of the total correlation function $h(r)$. Since we consider systems in planar geometry, we integrate Eq. (A1) with respect to the x and y coordinates and define polar coordinates $R = \sqrt{(x - x')^2 + (y - y')^2}$ to obtain an expression for the laterally integrated direct correlation function $\bar{c}_b^{(2)}(z)$ in terms of $c_b^{(2)}(r)$, i.e.,

$$\begin{aligned} \bar{c}_b^{(2)}(|z - z'|) &= \frac{1}{A} \int dx dy \int dx' dy' c_b^{(2)}(|\mathbf{r} - \mathbf{r}'|) \\ &= \int_0^\infty dR 2\pi R c_b^{(2)} \left[\sqrt{R^2 + (z - z')^2} \right] \\ &= \int_{|z-z'|}^\infty dr 2\pi r c_b^{(2)}(r) \\ &= \frac{-\beta \delta^2\mathcal{F}_{\text{exc}}[\rho]}{A \delta\rho(z)\delta\rho(z')}. \end{aligned} \quad (\text{A4})$$

Here we assume that $g(r)$ has converged to unity and that $c_b^{(2)}(r)$ is sufficiently short ranged such that it has essentially decayed to zero at the distance of half the box size, $r = L/2$. For the cubic systems with edge length $L = 10\sigma$, we found this condition to hold for bulk densities $\rho_b\sigma^3 < 0.67$.

With Eq. (A4), we have arrived at an expression that enables us to construct a training set, by computing $\delta^2\mathcal{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)|_{\rho_b}$ from sampled $g(r)$ through $c_b^{(2)}(r)$:

$$\frac{-\beta \delta^2\mathcal{F}_{\text{exc}}[\rho]}{A \delta\rho(z_i)\delta\rho(z_j)} = \int_{|z_i-z_j|}^\infty dr 2\pi r c_b^{(2)}(r). \quad (\text{A5})$$

We now construct an optimization objective that minimizes the distance between $(1/\Delta z)^2 \partial^2 F_{\theta}^{(2)}/\partial\rho_i \partial\rho_j$ and $\delta^2\mathcal{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)$, i.e.,

$$L(\theta) = \frac{1}{nm} \sum_{i,j} \left(\bar{c}_b^{(2)}(|z_i - z_j|) + \frac{\beta}{A(\Delta z)^2} \frac{\partial^2 F_{\theta}^{(2)}}{\partial\rho_i \partial\rho_j} \right)^2, \quad (\text{A6})$$

where n denotes the number of grid points and m denotes the number of Hessian rows fitted per loss evaluation. To reduce computational cost during training, we compute

a uniformly sampled batch of $m = 10$ rows of the Hessian $\{[\delta^2 F_\theta^{(2)}/\delta\rho_i\delta\rho_j]_{i=1}^n\}_{j=1}^m$ per loss evaluation. Optimizing for $\delta^2 \mathcal{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)$ means we lack information on the integration constant $\delta\mathcal{F}_{\text{exc}}/\delta\rho_0 = C$, akin to missing the integration constant C when integrating df/dx to find $f(x)$, e.g.,

$$f(x) = \int dx \frac{df}{dx} + C. \quad (\text{A7})$$

Therefore, we also apply an additional offset loss term during training that minimizes the difference between the uniform $\delta\mathcal{F}_{\text{exc}}/\delta\rho(z_i)|_{\rho_b}$ corresponding to bulk densities ρ_b and $\partial F_\theta^{(2)}/\partial\rho_i|_{\rho_b}$, given as

$$L_{\text{offset}} = \frac{1}{n} \sum_i \left(\frac{1}{\Delta z} \frac{\partial F_\theta^{(2)}}{\partial\rho_i} - \frac{\delta\mathcal{F}_{\text{exc}}}{\delta\rho(z_i)} \Big|_{\rho_b} \right)^2, \quad (\text{A8})$$

where n denotes the number of grid points and $\delta\mathcal{F}_{\text{exc}}/\delta\rho(z_i)$ is obtained from sampled bulk densities according to Eq. (1). This implies that the neural functional $F_\theta^{(2)}$ actually also learns the bulk density ρ_b (or actually its deviation from the ideal-gas density) at given μ and is thus expected to be able to accurately represent the bulk relation $\mu(\rho_b)$.

It is important to mention that we similarly cannot obtain the integration constant $\mathcal{F}_{\text{exc}}[0] = C$ from $\delta\mathcal{F}_{\text{exc}}/\delta\rho(z_i)$ in the case we seek access to $\mathcal{F}_{\text{exc}}[\rho]$ after training. However, since we know the low-density limit $\mathcal{F}_{\text{exc}}[0] = 0$, we can correct the output of our neural network functional such that $F_\theta^{(2)}[\rho] - C$ approximates $\mathcal{F}_{\text{exc}}[\rho]$, with $C = F_\theta^{(2)}[0]$.

The training procedure of pair-correlation matching is summarized in Algorithm 1. Here, we represent $c_b^{(2)}(z_i)$ numerically as $c_{b,i}^{(2)}$ with $c_{b,i}^{(2)} = c_{b,n-i}^{(2)}$ due to periodic boundary conditions. Similarly, we represent the sampled bulk density $\rho_b(z_i)$ numerically as $\rho_{b,i}$, where $\rho_{b,i} = \rho_{b,j}$, $\forall i, j \in n$.

We construct a dataset from simulations of 10^9 trial moves in a cubic box with an edge length of 10σ subject to periodic boundary conditions and a $\sigma/32$ grid spacing. All simulations are carried out at distinct chemical potentials $\beta\mu \in [-4, 0.5]$, resulting in a maximum bulk density of $\rho_b\sigma^3 = 0.67$ and a minimum bulk density of $\rho_b\sigma^3 = 0.02$

ALGORITHM 1. Pair-correlation matching.

Data: $\mathcal{D} = \{ \{c_{b,i}^{(2),0}, \rho_{b,i}^0\}_{i=1}^n, \dots, \{c_{b,i}^{(2),D}, \rho_{b,i}^D\}_{i=1}^n \}$ containing D pair-correlation functions $\{c_{b,i}^{(2)}\}_{i=1}^n$; D bulk density profiles $\{\rho_{b,i}\}_{i=1}^n$; loss scaling factor $\alpha = 1/1000$; loss scaling factor $\beta = 1/32$; number of uniformly sampled Hessian rows $m = 10$.

Result: trained neural functional $F_\theta^{(2)}(\{\rho_i\}_{i=1}^n)$.

for epoch do

for each $\{c_{b,i}^{(2)}, \rho_{b,i}\}_{i=1}^n$ **in** \mathcal{D} **do**

compute NN output $F_\theta^{(2)}(\{\rho_{b,i}\}_{i=1}^n)$;

compute $\{\partial F_\theta^{(2)}/\partial\rho_i\}_{i=1}^n$ with autodiff;

uniformly sample batch

$\mathcal{B} = \{\partial F_\theta^{(2)}/\partial\rho_i\}_{i=1}^m$ from $\{\partial F_\theta^{(2)}/\partial\rho_i\}_{i=1}^n$;

for each $\partial F_\theta^{(2)}/\partial\rho_i$ **in** \mathcal{B} **do**

compute $\{\partial^2 F_\theta^{(2)}/\partial\rho_i\partial\rho_j\}_{j=1}^n$ with autodiff;

compute $\{\delta^2 \mathcal{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)\}_{j=1}^n$ from $\{c_{b,i}^{(2)}\}_{i=1}^n$ with Eq. (A5);

end

$L(\theta) =$

$(1/nm) \sum_{i=1}^m \sum_{j=1}^n [\delta^2 \mathcal{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j) -$

$(1/\Delta z)^2 \partial^2 F_\theta^{(2)}/\partial\rho_i\partial\rho_j]^2$;

compute $\{\delta\mathcal{F}_{\text{exc}}/\delta\rho(z_i)\}_{i=1}^n$ with Eq. (1);

$L_{\text{offset}}(\theta) =$

$(1/n) \sum_{i=1}^n (\delta\mathcal{F}_{\text{exc}}/\delta\rho(z_i) - (1/\Delta z) \partial F_\theta^{(2)}/\partial\rho_i)^2$;

$L(\theta) = \alpha L(\theta) + \beta L_{\text{offset}}(\theta)$;

update parameters

$\theta \leftarrow \text{Optimizer}(\theta, \nabla_\theta L(\theta))$;

end

within the training set. Here, we verify that $c_b^{(2)}(r)$ has essentially decayed to zero at $r = L/2$ for bulk densities $\rho_b\sigma^3 < 0.67$.

We employ a convolutional neural network with periodic and dilated convolutions, each with a kernel size of 3, a dilation of 2, and 6 layers. The number of channels per layer is set to $N_{\text{channels}} = [16, 16, 32, 32, 64, 64]$, applying average pooling with kernel size 2 after each layer. This network takes as input an array of $n = 320$ values of $\{\rho_i\}_{i=1}^n$ per data point and produces a single scalar output, $F_\theta^{(2)}$. The model is trained for 180 epochs, taking approximately ~ 30 minutes on an Nvidia RTX 4070 GPU.