Supplementary Material: Learning Neural Free-Energy Functionals with Pair-Correlation Matching

Jacobus Dijkman,[1,2] Marjolein Dijkstra,[3] René van Roij,[4]
Max Welling,[2] Jan-Willem van de Meent,[2] 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*

(Dated: January 14, 2025)

CONTENTS

1. Classical Density Functional Theory 1
2. Training on the One-Body Direct Correlation Function 1
3. Construction of External Potentials 2
4. Obtaining the Excess Free Energy from Simulation 2
5. Numerical Errors in the Radial Distribution Function 3
6. More Examples of Density Estimates 4
7. Limitations of Pair-Correlation Matching 4
8. Local Learning and Pair-Correlation Matching 5
9. Extending the Training Set for Neural Functionals 9
10. Neural Functionals with Increased Resolution 10
11. Sampling Inhomogeneous Densities in Planar Geometry Systems vs. Arbitrary 3D Systems 10
References 11

1. CLASSICAL DENSITY FUNCTIONAL THEORY

Classical density functional theory (cDFT) is a grand-canonical framework that relies on the fact that the variational grand potential $\Omega[\rho]$ of a classical one-component many-body system at a given temperature T is uniquely determined by the particle density $\rho(\mathbf{r})$ via

$$\Omega[\rho] = \mathscr{F}[\rho] + \int \mathrm{d}r \rho(\mathbf{r}) \left(V_{\text{ext}}\left(\mathbf{r}\right) - \mu\right), \tag{S.1}$$

with $\mathscr{F}[\rho]$ representing the intrinsic Helmholtz free-energy functional, $V_{\text{ext}}(\mathbf{r})$ the external potential, and μ the chemical potential. For a given particle-particle interaction and temperature, the unique density functional $\mathscr{F}[\rho]$ determines the thermodynamic and structural equilibrium properties of a system for any chemical potential and external potential. Within cDFT, it is conventional to split the intrinsic free-energy functional into an ideal and excess contribution, $\mathscr{F}[\rho] = \mathscr{F}_{\text{id}}[\rho] + \mathscr{F}_{\text{exc}}[\rho]$. The ideal-gas contribution is exactly known as

$$\mathscr{F}_{\text{id}}[\rho] = \frac{1}{\beta} \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r}) \big(\ln \rho(\mathbf{r}) \Lambda^3 - 1 \big), \tag{S.2}$$

with $\beta = 1/k_B T$ and Λ the thermal wavelength.

Mathematical proofs exist [1] stating that (i) the equilibrium density profile, denoted here as $\rho_0(\mathbf{r})$, minimizes $\Omega[\rho]$, and (ii) the equilibrium grand potential equals $\Omega[\rho_0]$. Clearly, once $\mathscr{F}[\rho]$ for the system of interest is known, the Euler-Lagrange equation $\delta\Omega[\rho]/\delta\rho(\mathbf{r})|_{\rho_0} = 0$ can be solved to find $\rho_0(\mathbf{r})$ and $\Omega[\rho_0]$. The Euler-Lagrange equation takes the form

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

This self-consistency relation can be leveraged to find $\rho_0(\mathbf{r})$ through Picard iteration [2–4].

2. TRAINING ON THE ONE-BODY DIRECT CORRELATION FUNCTION

Instead of using pair-correlation matching, we can train a neural free-energy functional $F_\theta^{(1)}$ by minimizing the error between $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)$ and $(1/\Delta z)\partial F_\theta^{(1)}/\partial\rho_i$, as illustrated in Fig. S.1. We denote this neural functional as $F_\theta^{(1)}$, as the first functional derivative of the excess free energy is associated with the one-body direct correlation function by $c^{(1)}(x,y,z) = -\beta\delta\mathscr{F}_{\text{exc}}/\delta\rho(x,y,z)$. Relating to previous work, this neural functional is different from the neural functional of the one-body direct correlation function developed by Sammüller *et al.* [5], as expanded upon in Section 8 of the Supplementary Material.

We obtain the first functional derivative of the excess free energy from the equilibrium density by rearranging Eq. (S.3):

$$-\beta\frac{\delta\mathscr{F}_{\text{exc}}}{\delta\rho(x,y,z)} = \ln\Lambda^3\rho(x,y,z) + \beta V_{ext}(x,y,z) - \beta\mu. \quad \text{(S.4)}$$

We consider 3D systems in a planar geometry, where the excess free energy of a system with 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. For such a system, we can write

$$-\beta\frac{\delta\mathscr{F}_{\text{exc}}}{\delta\rho(z)} = A(\ln\Lambda^3\rho(z) + \beta V_{ext}(z) - \beta\mu). \quad \text{(S.5)}$$

This means that we can obtain $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z)$ by sampling equilibrium densities from simulation. We sample density profiles of inhomogeneous systems of Lennard-Jones particles above the critical point at a temperature $k_BT/\epsilon = 2$. 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 conducted at distinct chemical potentials $\beta\mu \in [-4, 0.5]$ and a maximum local density of $\rho(z)\sigma^3 = 0.67$. The choice of external potentials is explained in Section 3 of the Supplementary Material. We use the same convolutional neural network architecture as for the neural functional $F_\theta^{(2)}$: 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. The model is trained for 5000 epochs, requiring approximately 100 minutes on an Nvidia RTX 4070 GPU. The training procedure is summarized in Algorithm 1.

Since this method is dependent on the set of inhomogeneous densities included in the train dataset, the accuracy of the $F_\theta^{(1)}$ functional could plausibly be improved further by iterating on the design of the training dataset. We have not tested this extensively, since our primary interest in this work is in the pair correlation matching methodology.

Algorithm 1: Training on $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z)$

Data: train dataset $\mathcal{D} = \{\{\rho_i^0\}_{i=1}^n, \ldots, \{\rho_i^D\}_{i=1}^n\}$ consisting of D density profiles $\{\rho_i\}_{i=1}^n$ evaluated at gridpoints $\{z_i\}_{i=1}^n$.
Result: trained neural network model $F_\theta^{\text{exc}}(\{\rho_i\}_{i=1}^n)$.
for epoch **do**
  for each $\{\rho_i\}_{i=1}^n$ in $\mathcal{D}$ **do**
    generate model output scalar $F_\theta^{(1)}(\{\rho_i\}_{i=1}^n)$;
    compute $\{\partial F_\theta^{(1)}/\partial\rho_i\}_{i=1}^n$ with autodiff;
    compute $\{\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)\}_{i=1}^n$ with Eq. (S.5);
    $L_\theta = \frac{1}{n}\sum_{i=1}^n(\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i) - (1/\Delta z)\partial F_\theta^{(1)}/\partial\rho_i)^2$;
    update parameters $\theta \leftarrow \text{Optimizer}(\theta, \nabla_\theta L_\theta)$;
  end
end

3. CONSTRUCTION OF EXTERNAL POTENTIALS

The inhomogeneous density profiles used in this work, both for performance evaluation and training of the neural functional $F_\theta^{(1)}$, were generated via MC simulations subjected to various external potentials. These external potentials consist of randomized variations of well-potentials and Gaussian potentials.

The form of the well-potentials is adapted from Cats *et al.* [6], where

$$\beta V_{\text{well}}(z) = \begin{cases} 0 & \text{for} \quad |z| \in [-w\frac{L}{2}, w\frac{L}{2}], \\ s\left(\frac{|z|-w\frac{L}{2}}{(1-w)\frac{L}{2}}\right)^p & \text{for} \quad |z| > w\frac{L}{2}. \end{cases} \quad \text{(S.6)}$$

Here, s represents the dimensionless strength characterizing the potential at $|z| = L/2$, uniformly sampled with $s \sim \mathcal{U}(40, 60)$; w denotes the width of the central part of the slit ($\beta V_{\text{ext}} = 0$) and was uniformly sampled with $w \sim \mathcal{U}(0.4, 0.9)$; p characterizes the steepness of the potential and was uniformly sampled with $p \sim \mathcal{U}(2, 9)$.

The Gaussian potentials were constructed as a sum of Gaussians, expressed as

$$\beta V_{\text{Gauss}}(z) = \sum_{i=1}^N h_i \exp\left(-\frac{(z-\mu_i)^2}{2\sigma_i^2}\right), \quad \text{(S.7)}$$

where the number of Gaussians N is randomly chosen between $N = 0$ and $N = 10$; the mean of the Gaussians μ_i is uniformly sampled from $\mu \sim \mathcal{U}(0, L)$; the standard

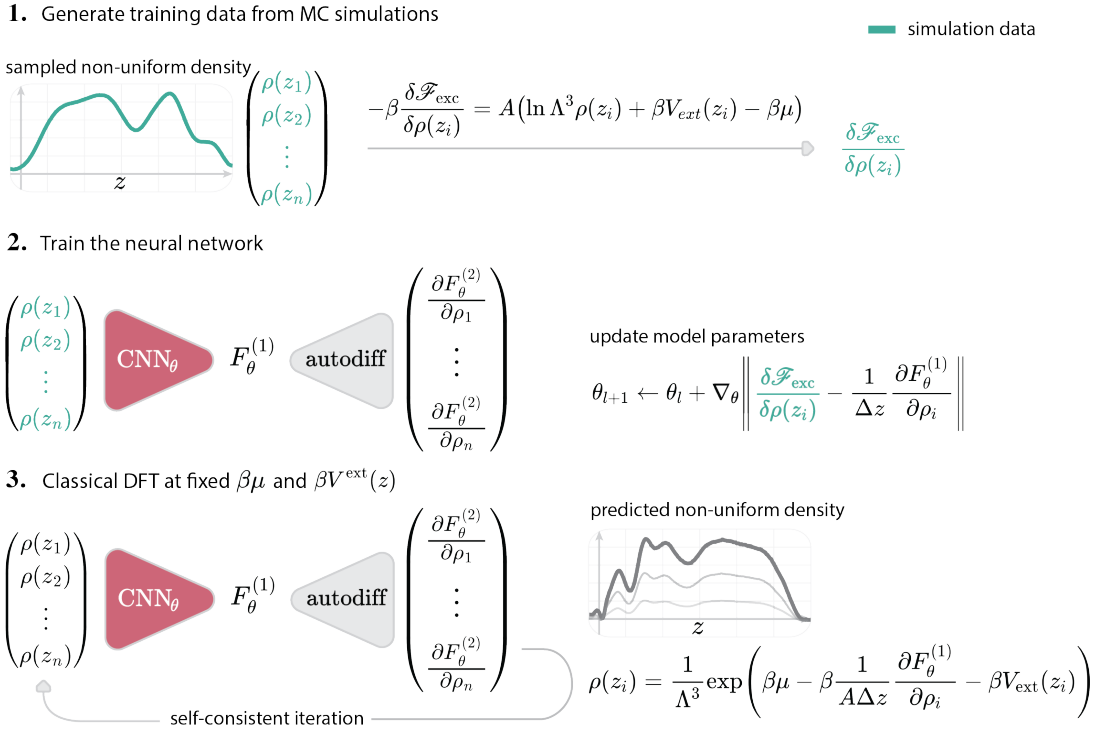

FIG. S.1: The neural free-energy functional $F_\theta^{(1)}$ is trained by fitting the gradient of the model output with respect to the input to the first functional derivative of the excess free energy as obtained from simulation, after which it can be applied within the classical Density Functional Theory (cDFT) framework. **1.** Non-uniform densities are sampled from Monte Carlo simulations of Lennard-Jones particles subjected to inhomogeneous external potentials, from which $\delta\mathscr{F}_{\rm exc}/\delta\rho(z_i)$ is derived. **2.** Through automatic differentiation (autodiff), the neural functional is optimized to fit the gradient of the model output with respect to input density profiles to $\delta\mathscr{F}_{\rm exc}/\delta\rho(z_i)$. **3.** The optimized model can then be applied in cDFT to obtain the non-uniform equilibrium densities and the excess free energy for a system of Lennard-Jones particles subjected to inhomogeneous external potentials.

deviation of the Gaussians σ_i is uniformly sampled from $\sigma \sim \mathcal{U}(0, L/10) + L/100$; the height of the Gaussians h_i is randomly sampled from a squared normal distribution, $h \sim \mathcal{U}^2(\mu = 0, \sigma^2 = 1)$.

The potentials constructed from a combination of well-potentials and Gaussian potentials were simply constructed by summing the individual potentials

$$\beta V_{\rm ext}(z) = \beta\left(V_{\rm well}(z) + V_{\rm Gauss}(z)\right). \tag{S.8}$$

For the results shown in Fig. 2 of the main text, simulations were performed for 150 distinct external potentials, of which 50 densities were generated using pure well-potentials, 50 were generated using a set of Gaussian potentials and 50 were generated using a combination of well-potentials and Gaussian potentials.

To train the neural functional $F_\theta^{(1)}$, a distinct dataset was constructed by using 200 pure well-potentials; 500 Gaussian potentials; 300 combinations of well-potentials and Gaussian potentials; 100 systems without an external potential. All simulations were conducted at a randomly selected chemical potential $\beta\mu \in [-4, 0.5]$.

4. OBTAINING THE EXCESS FREE ENERGY FROM SIMULATION

The excess Helmholtz free energy of the homogeneous bulk fluid (at a fixed volume V and temperature T) can be obtained from grand-canonical simulation via the equilibrium grand potential $\Omega_{eq}(\mu)$. The latter can be obtained by thermodynamic integration

$$\Omega_{eq}(\mu) = -\int_{-\infty}^{\mu} {\rm d}\mu' N(\mu'), \tag{S.9}$$

where $N(\mu') = V\rho_b(\mu')$ is the simulated (average) number of particles at chemical potential μ' in this homogeneous bulk system. Using the thermodynamic relation $\Omega_{eq} = \mathscr{F} - \mu N$ and $\mathscr{F} = \mathscr{F}_{id} + \mathscr{F}_{exc}$ with $\mathscr{F}_{id} = Nk_BT(\ln(N\Lambda^3/V) - 1)$, we arrive at $\mathscr{F}_{exc} = \Omega_{eq} + \mu(N)N - \mathscr{F}_{id}$. Here $\mu(N)$ is the inverse of $N(\mu)$, and we can identify $\mathscr{F}_{exc}$ with the excess free-energy functional evaluated at the homogeneous bulk, $\mathscr{F}_{exc}[\rho_b]$.

To calculate the integral of Eq. (S.9), we sample the number of particles for a system in steps of $\Delta\beta\mu = 0.2$, for a range of $\beta\mu = -4$ (corresponding to a bulk density of $\rho_b\sigma^3 = 0.02$) up to the target chemical potential. The

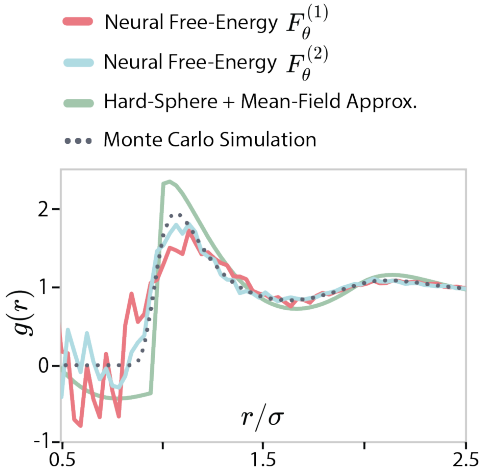

FIG. S.2: Unfiltered estimates of the radial distribution function for $\rho_b\sigma^3 = 0.67$. The radial distribution estimates from the neural functionals $F_\theta^{(1)}$, $F_\theta^{(2)}$ and $F_{\text{exc}}^{\text{MF}}$, which represents the analytical approximation of the White-Bear mark II version of Fundamental Measure Theory (FMT) combined with a mean-field approximation for the attractive part of the Lennard-Jones potential, are compared with simulation results.

resulting excess free energy from simulations is favorably compared with its neural-network representations $F_\theta^{(1)}$ and $F_\theta^{(2)}$ in Fig. 2 of the main text.

5. NUMERICAL ERRORS IN THE RADIAL DISTRIBUTION FUNCTION

The derivation of $g(r)$ from $c^{(2)}(z)$ involves several extensive transformations, which can introduce sensitivity to numerical errors. This numerical instability is most pronounced when $g(r) \to 0$, as shown in Fig. S.2.

To reduce the numerical errors in this region, we firstly average the bulk $\delta^2 F_\theta/\delta\rho(z_i)\delta\rho(z_j)$ values across rows of the neural functional Hessian, since $\delta^2 F_\theta/\delta\rho(z_i)\delta\rho(z_j)$ should be exactly the same as $\delta^2 F_\theta/\delta\rho(z_j)\delta\rho(z_i)$, due to the symmetry of a bulk system. In practice, small numerical differences remain between the Hessian rows for the neural functionals $F_\theta^{(1)}$, $F_\theta^{(2)}$ after training, which can influence the result of the extensive transformation from $c^{(2)}(z)$ to $g(r)$. Therefore averaging $\delta^2 F_\theta/\delta\rho(z_i)\delta\rho(z_j)$ across the Hessian reduces numerical errors when calculating $g(r)$. In addition, we slightly adapt $g(r)$ by filtering it as

$$g_{\text{filter}}(r_i) = \begin{cases} 0 & \text{for all } r_i < r_0, \\ g(r_i) & \text{otherwise,} \end{cases} \tag{S.10}$$

where r_0 is the largest interparticle distance r_i for which $g(r_i) \leq 0$.

6. MORE EXAMPLES OF DENSITY ESTIMATES

In addition to Figure 2a in the main text, we present additional examples of DFT predictions for various external potentials, as shown in Fig. S.3. In this Figure, the columns depict increasing chemical potentials from left to right: $\beta\mu = -3$ for Fig. S.3a and S.3d; $\beta\mu = 0$ for Fig. S.3b and S.3e; and $\beta\mu = 3$ for Fig. S.3c and S.3f.

The external potentials shown in Fig. S.3a–c are generated using a set of Gaussian potentials (Fig. S.3a and S.3c), and a combination of Gaussian potentials and a well-potential (Fig. S.3b), as described in Section 3 of the Supplementary Material. We observe that both neural functionals $F_\theta^{(1)}$ and $F_\theta^{(2)}$ provide accurate predictions for $\beta\mu = -3$ and $\beta\mu = 0$, both outperforming $F_{\text{exc}}^{\text{MF}}$. However, $F_{\text{exc}}^{\text{MF}}$ and $F_\theta^{(1)}$ both fail to converge to a solution for $\beta\mu = 3$ (Fig. S.3c and Fig. S.3f), which lies far outside the training set range of $\beta\mu \in [-4, 0.5]$. While $F_\theta^{(2)}$ is also trained with (uniform) densities up to $\beta\mu = 0.5$, it is still capable of producing relatively accurate predictions for $\beta\mu = 3$.

Additionally, we test the performance of the neural functionals on a number of atypical external potentials that are in no way related to the family of external potentials used elsewhere in this work (Fig. S.3d–f). Here we observe that both neural functionals $F_\theta^{(1)}$ and $F_\theta^{(2)}$ still perform well for these potentials for $\beta\mu = -3$ and $\beta\mu = 0$. Again, $F_{\text{exc}}^{\text{MF}}$ and $F_\theta^{(1)}$ fail to converge to a solution for $\beta\mu = 3$, whereas $F_\theta^{(2)}$ provides a relatively accurate estimate (Fig. S.3f).

7. LIMITATIONS OF PAIR-CORRELATION MATCHING

Given the surprising ability to accurately predict inhomogeneous density profiles by learning from merely bulk systems, it is natural to ask whether there are limits to the regimes where pair correlation can be applied successfully. In particular, we would like to understand to what extent inhomogeneities can be described within this approach and how far we can extrapolate away from the bulk using the pair-correlation matching approach. Moreover, since the pair-correlation matching approach relies on the inhomogeneity of the pair-correlation function for $F_\theta^{(2)}$ to approximate inhomogeneous densities, it is reasonable to question whether including pair-correlation functions at higher bulk densities in the training set would increase the performance of the pair-correlation matching approach for predicting highly inhomogeneous densities.

To investigate the extent to which pair-correlation matching is able to accurately describe inhomogeneities, we explore 4 types of external potentials for which a clear trend is visible between systematically increasing their

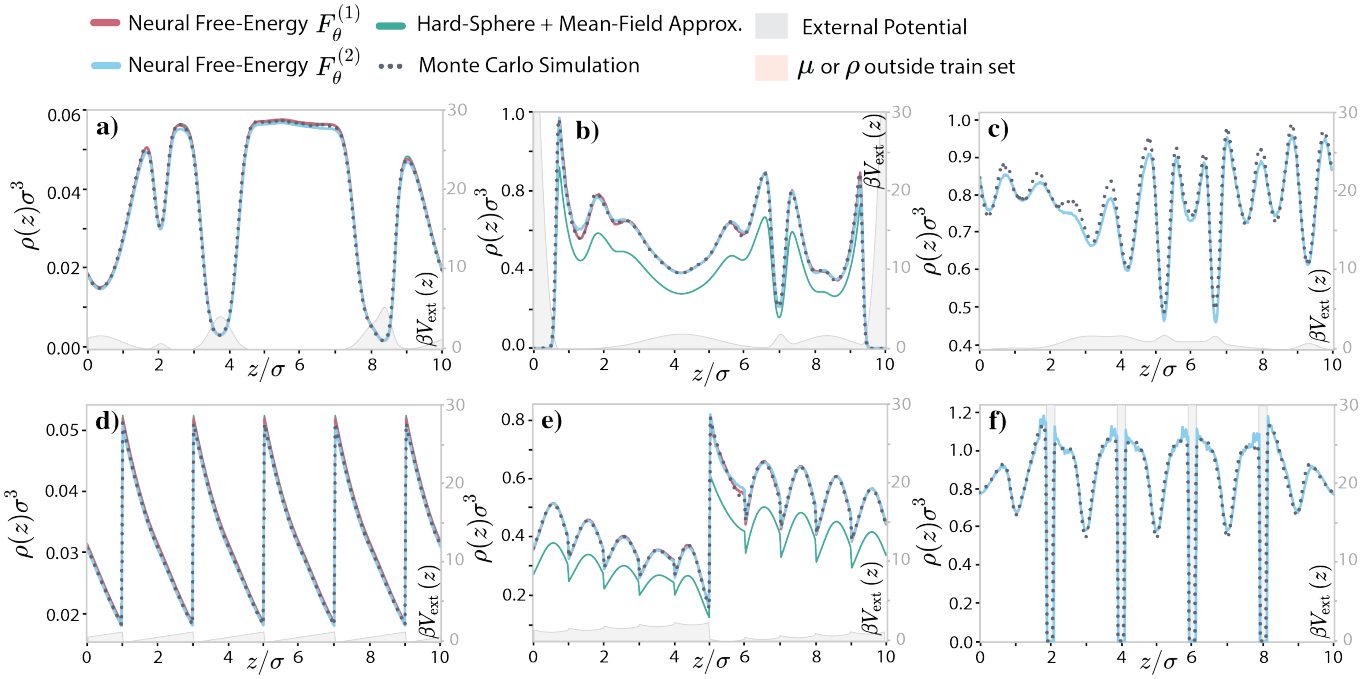

FIG. S.3: Evaluation of the neural free-energy functionals $F_\theta^{(1)}$ and $F_\theta^{(2)}$, where $F_\theta^{(1)}$ is optimized based on one-body correlation functions while $F_\theta^{(2)}$ is optimized using pair-correlation matching. Inhomogeneous density profiles are shown at six different external potentials (gray filling, right axes) through classical DFT comparing the functionals $F_\theta^{(1)}$, $F_\theta^{(2)}$ and $F_{\text{exc}}^{\text{MF}}$, which represents the analytical approximation of the White-Bear mark II version of Fundamental Measure Theory (FMT) combined with a mean-field approximation for the attractive part of the Lennard-Jones potential. The chemical potentials are given by **a/d)** $\beta\mu = -3$, **b/e)** $\beta\mu = 0$, and **c/f)** $\beta\mu = 3$, the latter being far beyond the training set $\beta\mu \in [-4, 0.5]$ where only $F_\theta^{(2)}$ gives a converged solution.

inhomogeneity and the impact on prediction accuracy. Figure S.4 illustrates these scenarios: Increasing the slit-pore width (Fig. S.4a-c); Increasing the steepness of a hyperbolic tangent potential (Fig. S.4d-f); Increasing the height of a Gaussian potential (Fig. S.4g-i); Increasing the wave number of a sinusoidal potential (Fig. S.4j-l). All systems are subjected to $\beta\mu = 0$.

We additionally investigate to what extent the predictive power of $F_\theta^{(2)}$ in highly inhomogeneous systems can be improved by extending the bulk density range of the train set. We compare with neural functionals $F_{\theta,\uparrow}^{(1)}$, and $F_{\theta,\uparrow}^{(2)}$ which have been trained with inhomogeneous densities or bulk pair-correlation functions in the range $-4 < \beta\mu < 3$ respectively, as expanded upon in Section 9. All neural functionals investigated in this section are trained on inhomogeneous densities and pair-correlation functions with a resolution of $\Delta z = \sigma/100$, for which the details can be found in Section 10.

Fig. S.4a-c demonstrates the effect of an increasingly wide slit-pore on prediction errors. Fig. S.4a shows that this increase in inhomogeneity affects the $F_\theta^{(2)}$-type functionals more than the $F_\theta^{(1)}$-type functionals. Fig. S.4b shows the density profile for a barrier width of 0.625σ, where $F_{\theta,\uparrow}^{(2)}$ exhibited the largest errors across the range shown in Fig. S.4b. A closer examination of the most erroneous region (Fig. S.4c) reveals that the $F_{\theta,\uparrow}^{(2)}$ functional underestimated the peaks and valleys of the density profile near the barrier.

Fig. S.4d-f shows the effect of an increasingly steep wall potential on the prediction error. As seen in Fig. S.4d, the error of the $F_\theta^{(2)}$-type functionals are affected by increasing steepness of the potential. The predicted DFT densities are shown for the external potential with the highest $F_{\theta,\uparrow}^{(2)}$ error, at $p = 4.6$ in Fig. S.4e-f. As expected by the similarity in external potential, the prediction error produced in the region near the wall shown in S.4f is similar to that in Fig. S.4c.

Fig. S.4g-i shows the effect of an increasingly high Gaussian peak potential on the prediction error of the neural functionals. Since this potential is perhaps less extreme than the other potentials, the error across the domain is also lower. However, we do still observe a small discrepancy between the increase of the Gaussian height between the $F_\theta^{(1)}$-type and $F_\theta^{(1)}$-type functionals. The DFT predictions at the Gaussian potential with the highest $F_{\theta,\uparrow}^{(2)}$ prediction error (with a Gaussian height of $5.9/\beta$) are shown in Fig. S.4h-i.

Fig. S.4j-l shows the effect of increasing the wave number of a sinusoidal potential on the prediction error of the neural functionals. This scenario proved to be the most

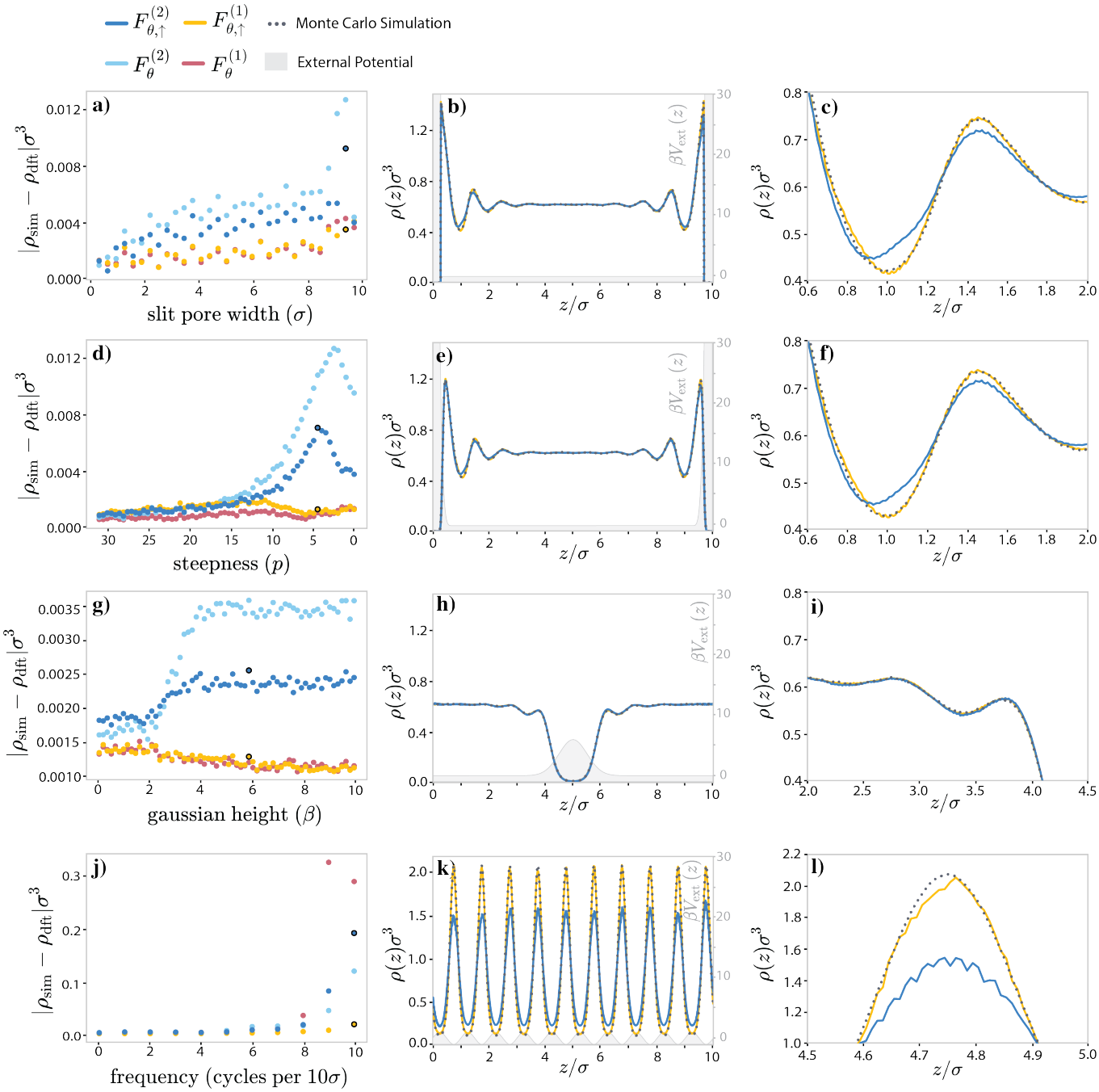

FIG. S.4: The performance of neural functionals trained with pair-correlation matching in comparison to neural functionals trained with a dataset of inhomogeneous densities on a selection of external potentials with a large degree of inhomogeneity, specifically selected to highlight settings where neural functional predictions deviate from Monte Carlo data when increasing the inhomogeneity. All systems are subjected to $\beta\mu = 0$. Neural functional $F^{(1)}_{\theta}$ (red) is trained with inhomogeneous densities within the range $-4 < \beta\mu < 0.5$; Neural functional $F^{(2)}_{\theta}$ (light blue) is trained with bulk pair-correlation functions in the range $-4 < \beta\mu < 0.5$; Neural functional $F^{(1)}_{\theta,\uparrow}$ (yellow) is trained with inhomogeneous densities within the range $-4 < \beta\mu < 3$; Neural functional $F^{(2)}_{\theta,\uparrow}$ (dark blue) is trained with bulk pair-correlation functions in the range $-4 < \beta\mu < 3$. **a/d/g/j)** mean absolute error $\frac{1}{n}\sum|\rho_{\mathrm{sim}}(z_i) - \rho_{\mathrm{dft}}(z_i)|$ of DFT predictions and MC simulations of particle densities for respectively a slit-pore potential with increasing width; a wall potential with increasing steepness indicated by p as specified in S.6; a Gaussian potential with increasing height; a sine potential with increasing frequency. Errors corresponding to the examples shown in the two rightmost columns are highlighted with a black circle. **b/e/h/k)** comparison of DFT predictions at the external potential that induced the largest mean error of the $F^{(2)}_{\theta,\uparrow}$ functional at respectively a width of 9.4σ; steepness parameter $p = 4.6$; Gaussian height of $5.9/\beta$; sine wavenumber equal to $2\pi/\sigma$. **c/f/i/l)** Zoom-in of the most erroneous region of the DFT estimate of the $F^{(2)}_{\theta,\uparrow}$ functional the system shown in b/e/h/k).

challenging for all functionals. Focusing on the $F_\theta^{(2)}$-type functionals, we see that the prediction error increases significantly at higher wave numbers. Fig. S.4k-l shows the DFT predictions at a sinusoidal potential with a wave number equal to $2\pi/\sigma$, which induced the highest $F_{\theta,\uparrow}^{(2)}$ prediction error. In Fig S.4j, we can also see that the $F_\theta^{(1)}$ functional shows the highest observed prediction error across all functionals for periods getting as small as σ, whereas larger periods (so smaller wave numbers) yield better results for all functionals. This can be attributed by the fact that the $F_\theta^{(1)}$ functional has probably not seen these types of density fluctuations in its train set of $-4 < \beta\mu < 0.5$. Indeed, including inhomogeneous densities up to $\beta\mu = 3$ incorporates state points with higher density variations in the train set of $F_{\theta,\uparrow}^{(1)}$ and lowers the prediction error.

These results suggest that, while pair-correlation matching is generally robust, we do observe a degradation of accuracy when predicting densities for highly inhomogeneous external potentials. As might be expected, increasing the range of bulk densities used during training significantly increases the predictive performance in highly inhomogeneous external potentials. These observations also suggest that adding bulk data for $\beta\mu > 3$ could improve the performance further yet.

8. LOCAL LEARNING AND PAIR-CORRELATION MATCHING

In Sammüller *et al.* [5], a local learning approach was proposed for learning the one-body direct correlation function directly. Since $c^{(1)}(x,y,z) = -\beta\delta\mathscr{F}_{\text{exc}}/\delta\rho(x,y,z)$, this is equivalent to learning the first functional derivative of the excess free energy directly. In this approach, the neural functional maps a local neighbourhood of density $[\rho(z_{i-w}),\dots,\rho(z_i),\dots\rho(z_{i+w})]$ around position z_i to a single value $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)$ at position z_i, as illustrated in Fig. S.5c. This *local* neural functional is trained on $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)$ obtained from sampled inhomogeneous densities, according to Eq. S.5. We will indicate this neural functional by $\delta F^{(1)}/\delta\rho(z_i)_\theta$. These notations and illustrations are adapted from Sammüller *et al.* [5] for consistency with the rest of this work and we refer to the original work for further details on this approach.

This approach is different from the $F_\theta^{(1)}$ functional introduced in this work in two respects. Firstly, $F_\theta^{(1)}$ is a neural functional for the excess free energy whereas the method of Sammüller *et al.* [5] learns a neural functional for the *functional derivative* of the excess free energy, as illustrated in Fig. S.5. Therefore, the $F_\theta^{(1)}$ functional is optimized by fitting the *gradient* of the neural network output $(1/\Delta z)\partial F_\theta^{(1)}/\partial\rho_i$ to $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)$, whereas $\delta F^{(1)}/\delta\rho(z_i)_\theta$ is optimized by fitting the neural network output to $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)$. Secondly, $F_\theta^{(1)}$ is a *global* functional, meaning that it accepts the full density field $[\rho(z_1),\dots,\rho(z_n)]$ as input and estimates $[\mathscr{F}_{\text{exc}}/\delta\rho(z_1),\dots,\mathscr{F}_{\text{exc}}/\delta\rho(z_n)]$ for the full system. In contrast, $\delta F^{(1)}/\delta\rho(z_i)_\theta$ is a *local* functional, which takes as input a local neighborhood of density $[\rho(z_{i-w}),\dots,\rho(z_i),\dots\rho(z_{i+w})]$ and estimates $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)$ at position z_i.

Similar to the local approach for learning $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)$ from inhomogeneous densities, pair-correlation matching can also be performed within this local learning scheme, as introduced by Sammüller and Schmidt [7] in response to an earlier version of this work. Here, the objective is to approximate $\delta^2\mathscr{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)$ for $j \in 1,\dots,n$ by the gradient of the neural functional $\delta F^{(2)}/\delta\rho(z_i)_\theta$, as illustrated in Fig. S.5d. This is different from the *global* pair-correlation matching approach used in this work, where the neural functional $F_\theta^{(2)}$ takes as input the density in the entire system $[\rho(z_1),\dots,\rho(z_n)]$, and is optimized by fitting the Hessian of the neural network to $\delta^2\mathscr{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)$ for all $i,j \in 1,\dots,n$, with n as the total number of gridpoints across the system.

As is evident in the results by Sammüller and Schmidt [7], the local version of pair-correlation matching does not seem to match the predictive capabilities of the global pair-correlation method introduced in this work. Here we implement this local version of pair-correlation matching to further investigate this discrepancy.

We implement similar *local* neural functionals to the approach by Sammüller *et al.* [5], a multi-layer perceptron (MLP) with 3 hidden-layers, each with 512 nodes. The neural network takes as input bulk densities of resolution $\Delta z = \sigma/100$. The input of the neural network is a local window of bulk densities $[\rho_b(z_{i-w}),\dots,\rho_b(z_i),\dots,\rho_b(z_{i+w})]$, a window of 3.5σ on both sides of z_i, meaning $w = 350$ gridpoints on both sides of $\rho(z_i)$. We train $\delta F^{(2)}/\delta\rho(z_i)_\theta$ on bulk pair-correlation functions in range $-4 < \beta\mu < 0.5$, and $\delta F^{(2)}/\delta\rho(z_i)_{\theta,\uparrow}$ on pair-correlation functions in range $-4 < \beta\mu < 3$. We compare with *global* neural functional $F_\theta^{(2)}$ trained on pair-correlation functions in range $-4 < \beta\mu < 0.5$ and $F_{\theta,\uparrow}^{(2)}$, trained in range $-4 < \beta\mu < 3$. We expand upon training neural functionals with pair-correlation matching in range $-4 < \beta\mu < 3$ in Section 9. Both $F_\theta^{(1)}$ and $F_\theta^{(2)}$ are trained on data with grid-spacing $\Delta z = \sigma/100$ as well, as detailed in Section 10.

We compare the accuracy of local and global neural functionals trained using pair-correlation matching on the same set of highly inhomogeneous external potentials as discussed in Section 7. We see that across all external potentials investigated, the local version of pair-correlation matching yields higher errors and a larger error increase with increase in inhomogeneity of the potential, as shown in Fig. S.6. At this stage it is not clear what is the reason behind this discrepancy in pre-

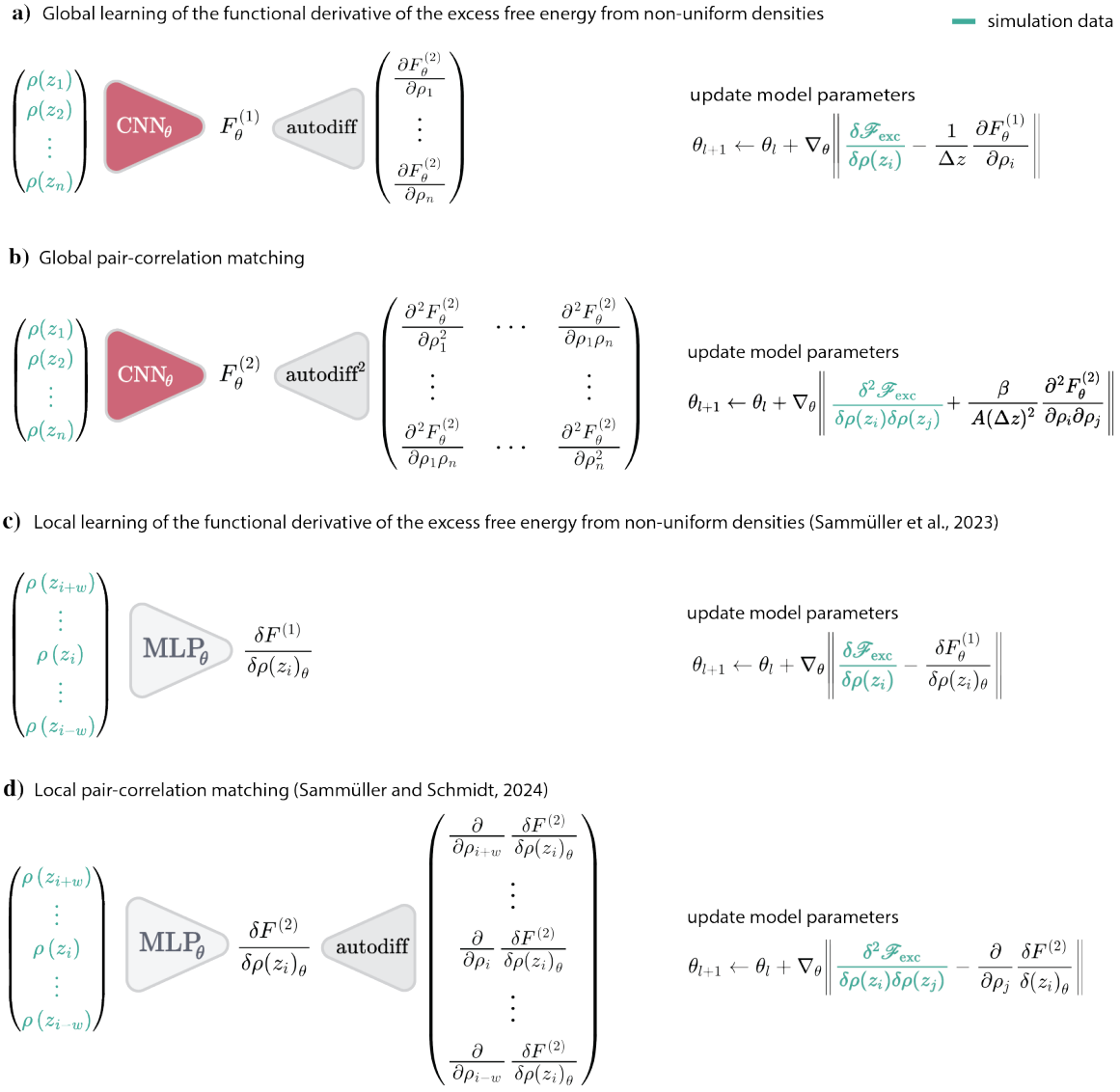

FIG. S.5: Overview of the global neural free-energy functionals introduced in this work and the local neural functionals introduced by Sammüller *et al.* [5] and Sammüller and Schmidt [7]. **a)** a neural free energy functional trained by fitting the gradient of a convolutional neural network to the first functional derivative of the excess free energy obtained from Monte Carlo simulations, as introduced in this work. **b)** a neural free energy functional trained with *global* pair-correlation matching, as introduced in this work. **c)** a local neural functional $\delta F^{(1)}/\delta\rho(z_i)_\theta$ trained by fitting the output of the neural network to the first functional derivative of the excess free energy, as obtained from Monte Carlo simulations. As introduced by Sammüller *et al.* [5], this approach uses a multi-layer perceptron (MLP) type neural network to estimate $\delta\mathscr{F}_{\text{exc}}/\delta\rho(z_i)$ at position z_i from a local range of density $[\rho(z_{i-w}), \dots, \rho(z_i), \dots \rho(z_{i+w})]$ around position z_i. **d)** pair-correlation matching applied in the local learning scheme [7]. Here, a neural network approximation for the functional derivative of the excess free energy $\delta F^{(2)}/\delta\rho(z_i)_\theta$ is optimized using a local adaptation of pair-correlation matching. Whereas pair-correlation matching introduced in this work fits the Hessian of the neural free-energy to the second functional derivative of the excess free-energy, here the gradient of the neural functional $\delta F^{(2)}/\delta\rho(z_i)_\theta$ is used to approximate a local range of $\delta^2\mathscr{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)$ for $j = i-w, \dots, i+w$.

dictive power between pair-correlation matching applied to global and local learning schemes.

9. EXTENDING THE TRAINING SET FOR NEURAL FUNCTIONALS

Since the pair-correlation matching approach relies on the inhomogeneity of the pair-correlation function for $F_\theta^{(2)}$ to approximate inhomogeneous densities, it is reasonable to question whether including pair-correlation

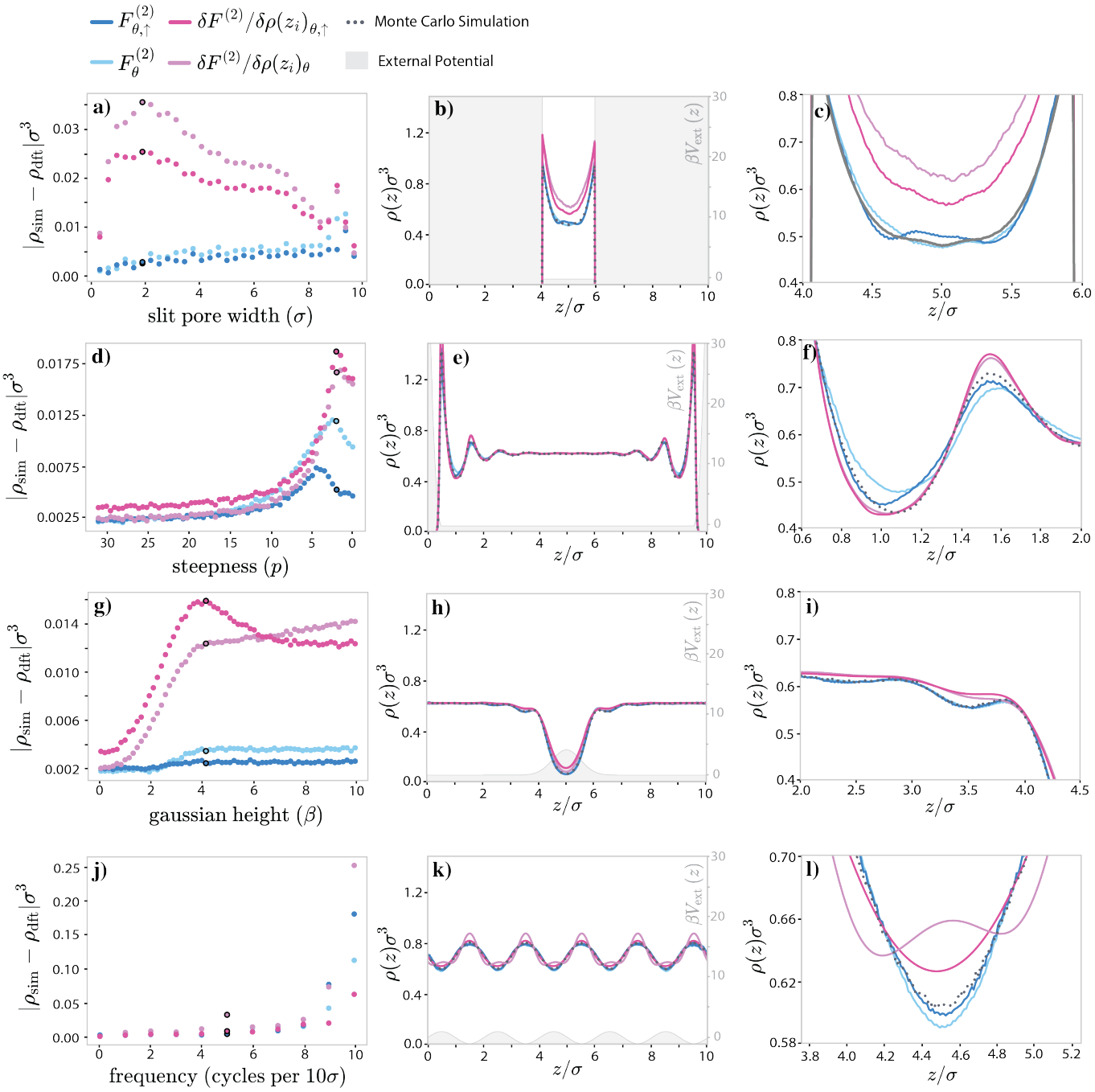

FIG. S.6: A comparison of the performance of neural functionals $\delta F^{(2)}/\delta\rho(z_i)_\theta$ trained with a local adaptation of pair-correlation matching against neural free-energy functionals $F^{(2)}_\theta$ trained with the global pair-correlation approach on a selection of external potentials with a large degree of inhomogeneity, specifically selected to highlight settings where neural functional predictions deviate from Monte Carlo data when increasing the inhomogeneity. All systems are subjected to $\beta\mu = 0$. Neural functional $\delta F^{(2)}/\delta\rho(z_i)_\theta$ (light purple) is trained with bulk pair-correlation functions in the range $-4 < \beta\mu < 0.5$; Neural functional $\delta F^{(2)}/\delta\rho(z_i)_{\theta,\uparrow}$ (dark purple) is trained with bulk pair-correlation functions in the range $-4 < \beta\mu < 3$; Neural functional $F^{(2)}_\theta$ (light blue) is trained with bulk pair-correlation functions in the range $-4 < \beta\mu < 0.5$; Neural functional $F^{(2)}_{\theta,\uparrow}$ (dark blue) is trained with bulk pair-correlation functions in the range $-4 < \beta\mu < 3$. **a/d/g/j)** mean absolute error $\frac{1}{n}\sum|\rho_{\mathrm{sim}}(z_i) - \rho_{\mathrm{dft}}(z_i)|$ of DFT predictions and MC simulations of particle densities for respectively a slit-pore potential with increasing width; a wall potential with increasing steepness indicated by p as specified in S.6; a Gaussian potential with increasing height; a sine potential with increasing frequency. Errors corresponding to the examples shown in the two rightmost columns are highlighted with a black circle. **b/e/h/k)** comparison of DFT predictions at samples with a large error of the $\delta F^{(2)}/\delta\rho(z_i)_{\theta,\uparrow}$ functional at respectively a width of 1.9σ; steepness parameter $p = 2.1$; Gaussian height of $4.2/\beta$; sine frequency of 0.5 cycles per 1σ. **c/f/i/l)** Close-up of the most erroneous regions of the DFT estimate of the $\delta F^{(2)}/\delta\rho(z_i)_{\theta,\uparrow}$ functional the system shown in b/e/h/k).

functions at higher bulk densities in the training set would increase the performance of the pair-correlation matching approach for predicting highly inhomogeneous densities. Therefore, the results of Section 7 and Section 8 include a comparison with neural functional $F^{(2)}_{\theta,\uparrow}$, which has been trained on direct correlation functions in range $-4 < \beta\mu < 3$. This is an extension of the train set of the neural functional $F^{(2)}_{\theta}$ for which the train set sits within the range $-4 < \beta\mu < 0.5$, as detailed in the main paper. The train set of $F^{(2)}_{\theta,\uparrow}$ consisted of 1000 direct correlation functions obtained from radial distribution functions sampled in cubic systems of size $(10\sigma)^3$, as well as 134 direct correlation functions in the range $0.5 < \beta\mu < 3$ obtained from from cubic systems of size $(20\sigma)^3$.

To obtain $c^{(2)}_b(r)$ through Eq. 8 of the main paper, we assume that $g(r)$ has converged to unity at 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 $\rho_b\sigma^3 < 0.67$ corresponding to $\beta\mu < 0.5$; for cubic systems with edge length $L = 20\sigma$, we found the condition to hold for bulk densities $\rho_b\sigma^3 < 0.82$ corresponding to $\beta\mu < 3$.

Since the direct correlation function is short ranged [8], we combine the direct correlation functions obtained from systems of size $(20\sigma)^3$ by cutting the tail of the computed $c^{(2)}(r)$ at 5σ, where $c^{(2)}(r) \to 0$, such that all computed $c^{(2)}(r)$ have range $0 \leq r \leq 5\sigma$ and are all used to compute $\delta^2\mathscr{F}_{\text{exc}}/\delta\rho(z_i)\delta\rho(z_j)|_{\rho_b}$ within a planar system of size $L = 10\sigma$. In bulk, the range $-4 < \beta\mu < 3$ corresponds to $0.02 < \rho_b\sigma^3 < 0.82$, exceeding the maximum bulk density $\rho_b\sigma^3 = 0.67$ for the range $-4 < \beta\mu < 0.5$.

In Section 8, we additionally compare with neural functional $F^{(1)}_{\theta,\uparrow}$, for which the train set of inhomogeneous densities similarly been extended from samples in the range $-4 < \beta\mu < 0.5$ to $-4 < \beta\mu < 3$. The dataset consists of 800 train samples. The inhomogeneous densities contained in this extended dataset are induced by the same type of external potentials as in the rest of this work, as detailed in Section 4.

These neural functionals $F^{(1)}_{\theta,\uparrow}$ and $F^{(2)}_{\theta,\uparrow}$ were both trained at a resolution of $\Delta z = \sigma/100$ and the neural network architecture detailed in Section 11.

10. NEURAL FUNCTIONALS WITH INCREASED RESOLUTION

The detailed comparisons of Sections 7 and 8 were performed with neural functionals $F^{(1)}_{\theta}$ and $F^{(2)}_{\theta}$ that were trained with respectively inhomogeneous densities and bulk direct correlation functions of grid-spacing $\Delta z = \sigma/100$, and therefore estimate densities with this resolution of $\Delta z = \sigma/100$ as well. This is an increase in resolution of the training data from a grid-spacing of $\sigma/32$, as used in the main paper. Together with this increase in resolution, the number of layers in the convolutional neural network of the neural functionals $F^{(1)}_{\theta}$ and $F^{(2)}_{\theta}$ was increased from 6 to 8, with the number of channels per layer as $N_{\text{channels}} = [32, 32, 32, 32, 64, 64, 64, 64]$, applying average-pooling with kernel size 2 after each layer. These adaptations were applied to optimize the predictive performance of these functionals and as an attempt to limit the effect of numerical errors, such as accumulated in the numerical transformation from the $g(r)$ to $\bar{c}^{(2)}(|z - z'|)$, on the estimates of the neural functionals. Additionally, we changed the parameter α in the Picard iteration [2–4] from 0.1 to 0.01 for better convergence. Lastly, the loss factors $\alpha = 1/1000$ and $\beta = 1/32$ of the $F^{(2)}_{\theta}$ functional were changed to $\alpha = 1$ and $\beta = 1$ (see the End Matter of the main paper), to place more importance on the correct offset of the predicted density. This slightly reduced the error of the $F^{(2)}$ functional for lower densities.

All neural functionals $F^{(1)}_{\theta}$, $F^{(1)}_{\theta}$, $F^{(1)}_{\theta,\uparrow}$ and $F^{(2)}_{\theta,\uparrow}$ as discussed in Sections 7 and 8, were trained with these adaptations for higher resolution neural functionals. The neural functional $F^{(1)}_{\theta}$ with increased resolution was trained on 800 inhomogeneous densities in range range $-4 < \beta\mu < 0.5$. The $F^{(2)}_{\theta}$ functional with increased resolution was trained on 1000 bulk direct correlation functions in range $-4 < \beta\mu < 0.5$. This difference in dataset size was due to the fact that no bulk direct correlation functions had to be held in the test set, as the performance of the neural functional was validated on inhomogeneous densities instead.

The neural functional $F^{(1)}_{\theta,\uparrow}$ with increased resolution was trained on 800 inhomogeneous densities in range range $-4 < \beta\mu < 3$. The $F^{(2)}_{\theta,\uparrow}$ functional with increased resolution was trained on 1000 direct correlation functions in range $-4 < \beta\mu < 0.5$ and 134 direct correlation functions in the range $0.5 < \beta\mu < 3$. These functionals are further detailed in Section 9.

11. SAMPLING INHOMOGENEOUS DENSITIES IN PLANAR GEOMETRY SYSTEMS VS. ARBITRARY 3D SYSTEMS

The density profiles in planar geometry used in this study require approximately 1 hour of CPU time per density profile. Extrapolating these times naively to full three-dimensional density profiles, it would require an impractical $A/(\Delta x\Delta y) \cdot 1$ hours = 102400 hours to generate each density with similar accuracy for a resolution of $\Delta x = \Delta y = \Delta z = \sigma/32$ as used in this work. To illustrate this, we compare a 1D slice from a 3D density profile with a 1D density profile in planar geometry, both sampled within the same number of MC steps (10^9 trial moves with 10^7 equilibration moves and 4 decorrelation cycles) (Fig. S.7). These density profiles are constructed

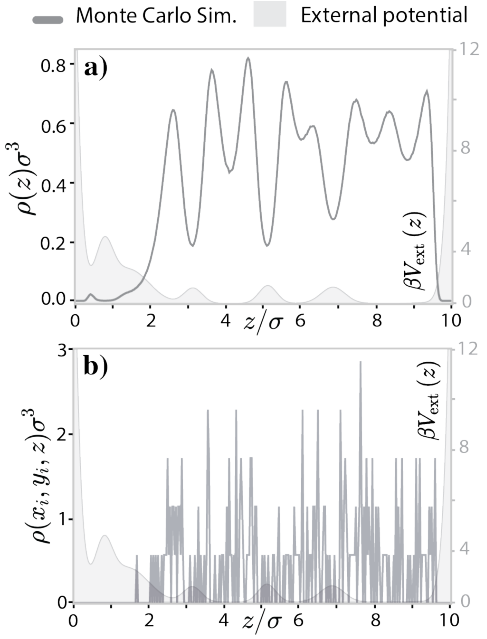

FIG. S.7: Comparison between sampling densities $\rho(z)$ in planar geometry and $\rho(x, y, z)$ in arbitrary three-dimensional geometry. **a)** $\rho(z)$ in an external potential $V_{\mathrm{ext}}(z)$ (shown in gray) sampled from a MC simulation with 10^9 trial moves. **b)** Slice of $\rho(x, y, z)$ at x_i for the same laterally symmetric external potential $V_{\mathrm{ext}}(z)$, also sampled from a MC simulation with 10^9 trial moves.

by binning particle positions in intervals of $4 \cdot N_{particles}$ trial moves throughout the MC simulation. After completion of the simulation, only a very low number of particles has been counted in each bin of the 3D histogram that constructs the density of Fig. S.7b. Many bins remain empty and many contain only a few particles, creating the discrete peaks of Fig. S.7b.

This illustrates that much longer sampling times are necessary to sample accurate three-dimensional density profiles. The results presented in this paper provide a compelling alternative: the radial distribution functions sampled for this work were already obtained from 3D bulk systems, after which they were numerically transformed into direct correlation functions in planar geometry. This means that it is likely that the same dataset of radial distribution functions can be used when extending this approach to arbitrary three-dimensional systems, with the only difference that a numerical transformation to the *radially symmetric* pair-correlation function $c^{(2)}(r)$ needs to be applied.

[1] 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).

[2] R. Roth, Introduction to Density Functional Theory of Classical Systems: Theory and Applications, Lecture Notes (2006).

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

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

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

[6] 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**, 10.1063/5.0042558 (2021), arxiv:2101.01942.

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

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