Supplementary Information for
"Neural density functionals: Local learning and pair-correlation matching"

Florian Sammüller and Matthias Schmidt
*Theoretische Physik II, Physikalisches Institut, Universität Bayreuth, D-95447 Bayreuth, Germany**
(Dated: July 31, 2024)

I. LOCAL INHOMOGENEOUS ONE-BODY LEARNING

A. Training data

Training of the neural correlation functional proceeds as in Ref. [1] with GCMC simulation data for inhomogeneous randomized external potentials. We consider the truncated ($r_c = 2.5\sigma$) Lennard-Jones fluid at constant supercritical [2] temperature $k_BT/\epsilon = 1.5$. We reuse the training data set of Ref. [1] consisting of 500 simulation results in a box of length 20σ in the inhomogeneous x-direction and lateral lengths of 10σ. The inhomogeneities in the considered training set are substantial and result in local density maxima of up to $\rho(x)\sigma^3 \lesssim 6$, which nevertheless poses no difficulty during sampling (in contrast to accessing the radial distribution function $g(r)$ for large bulk densities, cf. Sec. III A). The data set is deposited in Zenodo [1].

One-body direct correlation profiles are calculated pointwise from the sampled density profiles $\rho(x)$ according to

$$c_1^{\mathrm{ref}}(x;[\rho]) = \ln \rho(x) + \beta V_{\mathrm{ext}}(x) - \beta\mu, \qquad (1)$$

where the external potential $V_{\mathrm{ext}}(x)$, chemical potential μ and inverse temperature $\beta = 1/(k_BT)$ are known input quantities of the simulations. The logarithm in Eq. (1) is understood as $\ln(\rho(x)\Lambda^3)$, where the thermal wavelength Λ is set to the particle size σ, which defines our unit of length.

B. Neural network and training procedure

The neural network used for the local representation of the one-body direct correlation functional $c_1(x;[\rho])$ possesses a straightforward multilayer perceptron (MLP) architecture, see Fig. S5. We use three hidden layers with 512 nodes and smooth non-linear activation functions (Gaussian error linear units, "GELU"). The input layer consists of 701 nodes, as the density profile is provided in a window of size 3.5σ around the location of interest with a discretization interval of $\Delta x = 0.01\sigma$.

* Florian.Sammueller@uni-bayreuth.de

Training the neural network on the basis of inhomogeneous one-body learning amounts to minimizing the loss

$$L_{\mathrm{inhom}} = \sum_{i,j} \left(c_1(x_i;[\rho_j]) - c_1^{\mathrm{ref}}(x_i;[\rho_j])\right)^2, \qquad (2)$$

where the neural network prediction $c_1(x_i;[\rho_j])$ is compared to the reference data $c_1^{\mathrm{ref}}(x_i;[\rho_j])$ for each discretized location x_i of every inhomogeneous system j in the training set. Code, simulation data and trained models are openly available [3].

C. Applications

For the prediction of density profiles, rearranging Eq. (1) yields

$$\rho(x) = \exp\left(-\beta(V_{\mathrm{ext}}(x) - \mu) + c_1(x;[\rho])\right), \qquad (3)$$

which constitutes the standard Euler-Lagrange equation of DFT. Given an analytical or neural representation of $c_1(x;[\rho])$, Eq. (3) can be solved self-consistently for $\rho(x)$ with a standard Picard iteration [1].

Accessing structural information on the pair-correlation level is facilitated by functional differentiation of $c_1(x;[\rho])$, which yields the two-body direct correlation functional

$$c_2(x,x';[\rho]) = \frac{\delta c_1(x;[\rho])}{\delta\rho(x')}. \qquad (4)$$

The functional derivative can be evaluated efficiently for general density input $\rho(x)$ by automatic differentiation and normalization with the discretization interval Δx. Using automatic differentiation is particularly suitable and natural in case of a neural-network-based representation of $c_1(x;[\rho])$.

II. BULK PAIR-CORRELATIONS

The considered neural functionals operate on the level of direct correlations in planar geometry. In the following, we lay out conversions between radial and planar geometry in direct and Fourier space. The Ornstein-Zernike equation is utilized to obtain total correlation functions from neural predictions and, inversely, to convert radial distribution functions $g(r)$ from direct measurement in simulation to corresponding direct correlation functions for use in the pair-correlation matching, see Sec. III A.

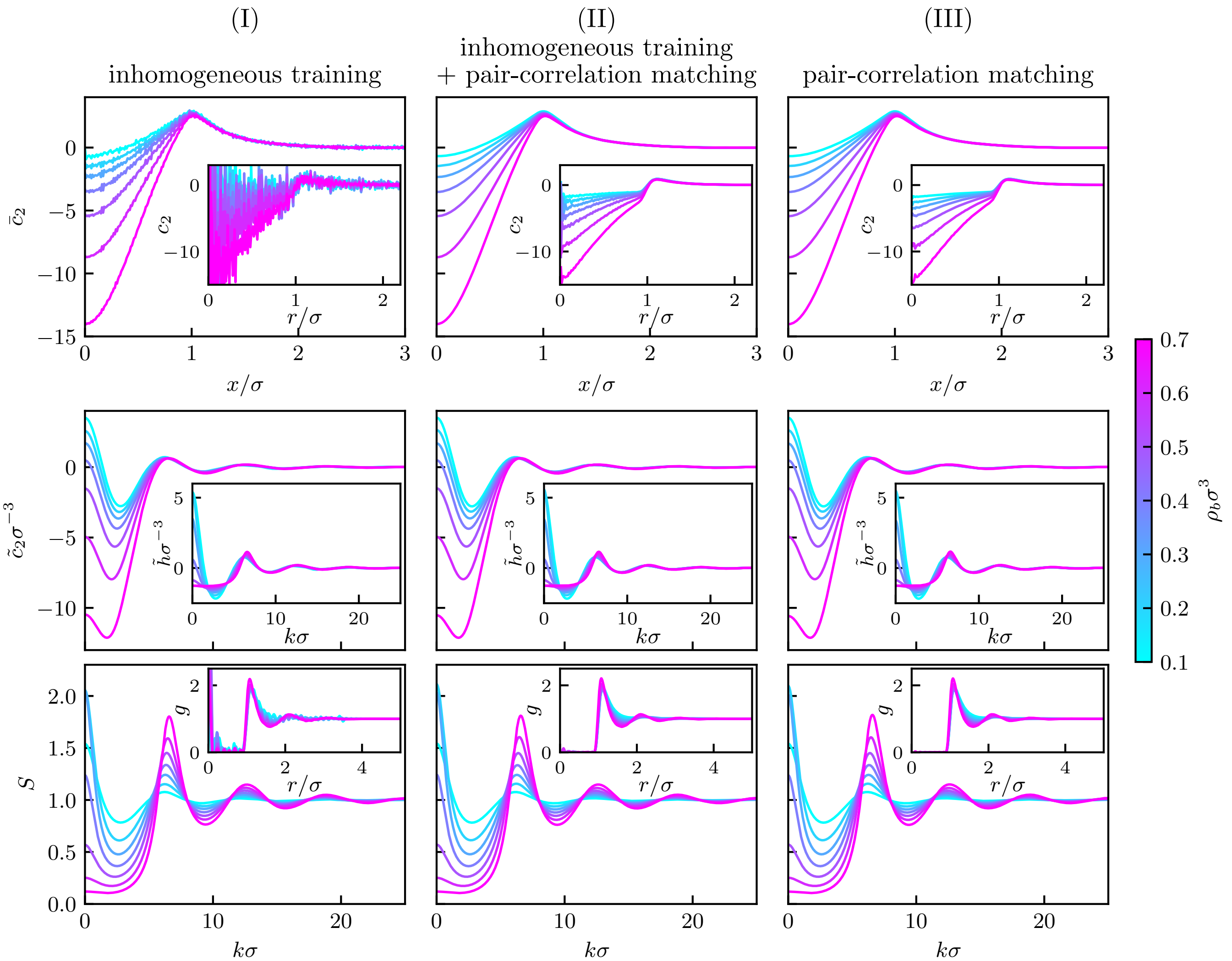

FIG. S1. Comparison of predictions of neural functionals for the truncated Lennard-Jones fluid at temperature $k_BT/\epsilon = 1.5$ and bulk densities $\rho_b\sigma^3 = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$. Results are shown for training with purely inhomogeneous reference data (I, left column), with additional pair-correlation regularization (II, middle column) as well as with pure pair-correlation matching (III, right column). We depict the planar two-body direct correlation function $\bar{c}_2(x;\rho_b)$ and its radial representation $c_2(r;\rho_b)$ (first row), the direct and total correlation functions $\tilde{c}_2(k;\rho_b)$ and $\tilde{h}(k;\rho_b)$ in Fourier space (second row), as well as the static structure factor $S(k;\rho_b)$ and the radial distribution function $g(r;\rho_b)$ (third row).

A. Planar and radial geometry

The planar projection of a generic radially symmetric function $f(\mathbf{r}) = f(|\mathbf{r}|) = f(r)$ is given by

$$\bar{f}(x) = \int \mathrm{d}y \int \mathrm{d}z\, f(\sqrt{x^2+y^2+z^2}) = 2\pi \int_x^\infty \mathrm{d}r\, r f(r), \tag{5}$$

where we indicate quantities in the reduced planar geometry by an overbar. By differentiating Eq. (5), one arrives at the inverse transformation from planar to radial geometry via

$$f(r) = -\frac{1}{2\pi r} \left.\frac{\partial \bar{f}(x)}{\partial x}\right|_{x=r}, \tag{6}$$

where we have assumed radial symmetry for $f(r)$. Applying a one-dimensional Fourier transform to the planar projection of a radially symmetric function yields directly the radially symmetric representation in Fourier space (indicated by a tilde),

$$\begin{aligned}\int \mathrm{d}x\, \bar{f}(x)\exp(ik_x x) &= \int \mathrm{d}\mathbf{r}\, f(r)\exp(i\mathbf{k}\cdot\mathbf{r})\\ &= \tilde{f}(|\mathbf{k}|) = \tilde{f}(k),\end{aligned} \tag{7}$$

where $\int \mathrm{d}\mathbf{r}$ denotes a volume integral over the entire domain.

Transforming a radially symmetric function $f(r)$ in real space to its radially symmetric Fourier representation $\tilde{f}(k)$ is facilitated by a radial Fourier(-Hankel) transform,

$$\tilde{f}(k) = \frac{4\pi}{k} \int_0^\infty \mathrm{d}r\, r f(r) \sin(kr). \tag{8}$$

The inverse transform of Eq. (8) is identical up to a prefactor,

$$f(r) = \frac{1}{2\pi^2 r} \int_0^\infty \mathrm{d}k\, k \tilde{f}(k) \sin(kr). \tag{9}$$

The numerical evaluation of Eqs. (8) and (9) can be facilitated either via a fast Fourier transform (FFT) or via a discrete sine transform (DST). In both cases, particular care must be taken to account correctly for the discretization and the implied periodicity of the radially symmetric function to be transformed.

B. Direct and total correlation functions

For the considered two-body direct correlation function $\bar{c}_2(x;\rho_b)$, as obtained from the neural functional by autodifferentiation (4) and network evaluation in bulk with constant density input $\rho(x) = \rho_b$, Eq. (6) gives access to the more common radial representation $c_2(r;\rho_b)$, as shown in Figs. 1 and S1. Due to the required numerical differentiation and the division by r in Eq. (6), the results for $c_2(r;\rho_b)$ are prone to numerical artifacts in particular for small values of r.

Commencing with a neural functional prediction for $\bar{c}_2(x;\rho_b)$, a one-dimensional Fourier transform yields the one-body direct correlation function $\tilde{c}_2(k;\rho_b)$ in radial geometry according to Eq. (7). The total correlation function $\tilde{h}(k;\rho_b)$ can then be calculated algebraically in Fourier space via the Ornstein-Zernike equation

$$\tilde{h}(k;\rho_b) = \frac{\tilde{c}_2(k;\rho_b)}{1 - \rho_b \tilde{c}_2(k;\rho_b)}. \tag{10}$$

From $\tilde{h}(k;\rho_b)$, one obtains the static structure factor

$$S(k;\rho_b) = 1 + \rho_b \tilde{h}(k;\rho_b), \tag{11}$$

and the transformation (9) back to direct space gives access to the radial distribution function

$$g(r;\rho_b) = h(r;\rho_b) + 1. \tag{12}$$

Neural functional predictions of the quantities in Eqs. (10)–(12) are shown in Fig. S1 at different bulk densities $\rho_b\sigma^3 = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$, and 0.7.

III. PAIR-CORRELATION MATCHING AND REGULARIZATION

A. Preparation of simulation data

We have performed 500 simulations of the Lennard-Jones fluid in a cubic box of size $(20\sigma)^3$ at different chemical potentials chosen uniformly in the range $-7 < \mu/\epsilon < 0$. This yields accurately sampled radial distribution functions $g(r;\rho_b)$ in $0 \le r/\sigma \le 10$ for bulk densities $0.01 \lesssim \rho_b\sigma^3 \lesssim 0.72$. Accessing larger bulk densities is hindered by the fact that $g(r)$ still displays non-negligible oscillations for $\rho_b\sigma^3 \gtrsim 0.72$ for the maximal radial distance $r = 10\sigma$ in the considered simulation box. This stands in contrast to the sampling of inhomogeneous density profiles, where much larger local density maxima pose no substantial difficulty, see Sec. I A.

From the simulation data for $g(r;\rho_b)$, we calculate:

- $h(r;\rho_b)$ via the definition (12) of the total pair-correlation function,
- $\tilde{h}(k;\rho_b)$ via the radial Fourier transform (8),
- $\tilde{c}_2(k;\rho_b)$ via the Ornstein-Zernike equation (10),
- $c_2(r;\rho_b)$ via the inverse radial Fourier transform (9),
- $\bar{c}_2(x;\rho_b)$ via the planar projection (5).

The planar bulk two-body direct correlation function $\bar{c}_2(x;\rho_b)$ can be compared directly to the functional derivative (4) of the one-body direct correlation functional in planar geometry,

$$\bar{c}_2(x;\rho_b) = \left.\frac{\delta c_1(0;[\rho])}{\delta \rho(x)}\right|_{\rho=\rho_b}. \tag{13}$$

Automatic differentiation is used to evaluate the functional derivative of the neural-network-based representation of $c_1(x;[\rho])$ in Eq. (13).

B. Loss functions and training procedure

Pair-correlation matching is implemented via the loss

$$\begin{aligned} L_{\mathrm{pc}} = &\sum_{i,j} \left(\bar{c}_2(x_i;\rho_{b,j}) - \bar{c}_2^{\mathrm{ref}}(x_i;\rho_{b,j})\right)^2 \\ &+ \sum_j \left(\mu_{\mathrm{exc}}(\rho_{b,j}) - \mu_{\mathrm{exc}}^{\mathrm{ref}}(\rho_{b,j})\right)^2, \end{aligned} \tag{14}$$

where $\bar{c}_2(x_i;\rho_{b,j})$ is calculated for the given bulk density input $\rho_{b,j}$ of bulk simulation j according to Eq. (13). The result is compared directly to the pre-calculated simulation reference $\bar{c}_2^{\mathrm{ref}}(x_i;\rho_{b,j})$, see Sec. III A, for each discretized spatial coordinate x_i. To fix the remaining integration constant, the excess chemical potential $\mu_{\mathrm{exc}}(\rho_b) = -k_B T c_1(x;\rho_b)$ of the neural network prediction is compared to the simulation reference $\mu_{\mathrm{exc}}^{\mathrm{ref}}$, which serves as an additive contribution to L_{pc}.

In general, we consider a linear combination of both loss functions,

$$L = \alpha_{\mathrm{inhom}} L_{\mathrm{inhom}} + \alpha_{\mathrm{pc}} L_{\mathrm{pc}}, \tag{15}$$

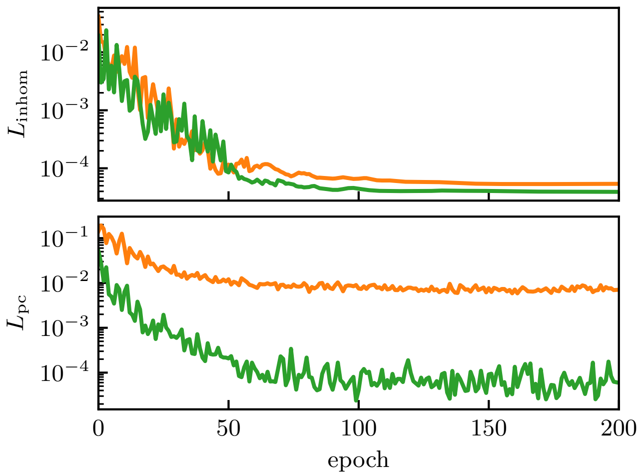

FIG. S2. Training dynamics with pair-correlation regularization (II, green) results in lower loss values L_{inhom} and L_{pc} as compared to pure inhomogeneous one-body learning without regularization (I, orange). We show the L_{inhom} and L_{pc} as a function of the training epoch.

which is used for the backpropagation and for the adaptation of the trainable parameters of the neural network. The constant factors α_{inhom} and α_{pc} control the relative influence of inhomogeneous one-body matching (2) and pair-correlation matching (14) by weighting the respective loss terms. Three choices are considered:

(I) $\alpha_{\mathrm{inhom}} = 1$, $\alpha_{\mathrm{pc}} = 0$:
purely inhomogeneous one-body learning as in Ref. [1],

(II) $\alpha_{\mathrm{inhom}} = 1$, $\alpha_{\mathrm{pc}} = 0.01$:
inhomogeneous one-body learning with pair-correlation regularization as introduced in this work,

(III) $\alpha_{\mathrm{inhom}} = 0$, $\alpha_{\mathrm{pc}} = 1$:
pure pair-correlation matching inspired by Ref. [4].

C. Training dynamics and bulk results

Fig. S2 shows a comparison of the training dynamics for the individual loss terms L_{inhom} and L_{pc} for inhomogeneous one-body learning with (II) and without (I) pair-correlation regularization. The magnitude of both loss terms is decreased for the case of added pair-correlation regularization.

In Fig. S1, we depict results for bulk pair-correlation functions for each training strategy (I)–(III). Purely inhomogeneous one-body training (I) yields slightly noisy derivatives $\bar{c}_2(x;\rho_b)$, which hamper the performance of challenging numerical transformations, e.g. to obtain the two-body direct correlation function $c_2(r;\rho_b)$ in radial geometry or the radial distribution function $g(r)$. This problem is resolved both with the added pair-correlation regularization (II) as well as with pure pair-correlation matching (III). Although the latter strategy enables accurate bulk predictions due to the supply of bulk data during training, it provides no reasonable extrapolation to inhomogeneous scenarios, as seen e.g. in inaccurate predictions of density profiles (Fig. 1) and of inhomogeneous two-body correlations (Fig. S3).

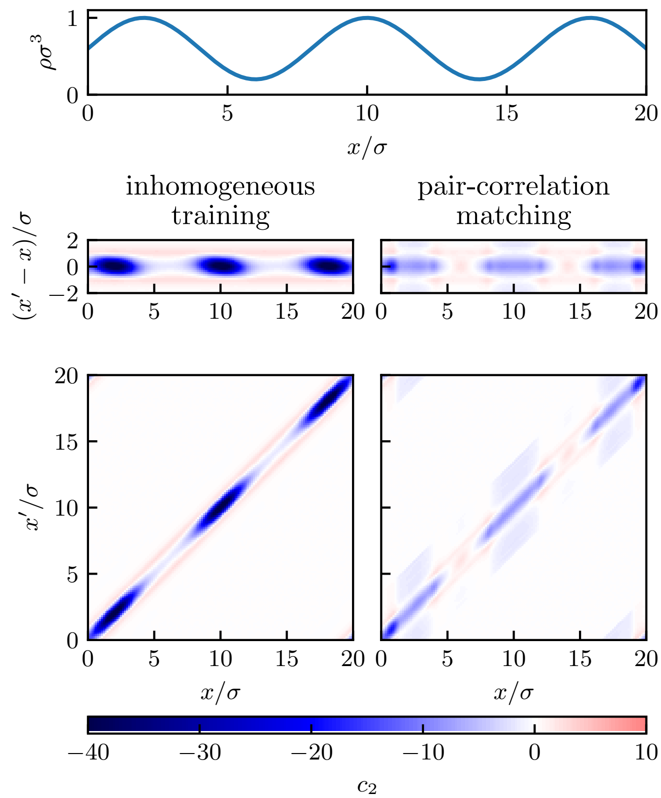

FIG. S3. Two-body direct correlation functional $c_2(x, x'; [\rho])$ for an exemplary inhomogeneous density profile $\rho(x)$ as input. We show results for neural functionals that have been trained via inhomogeneous one-body learning (I) and with pure pair-correlation matching (III). The results of the pair-correlation-matched neural functional violate the expected exchange symmetry $c_2(x, x'; [\rho]) = c_2(x', x; [\rho])$.

IV. INHOMOGENEOUS TWO-BODY CORRELATIONS AND NOETHER SUM RULES

Autodifferentiation also gives access to $c_2(x, x'; [\rho])$ for inhomogeneous density input according to Eq. (4), which enables further quality assessments of the neural functionals. As the two-body direct correlation functional formally arises as a second-order functional derivative of the excess free energy functional $F_{\mathrm{exc}}[\rho]$, testing the expected exchange symmetry $c_2(x, x'; [\rho]) = c_2(x', x; [\rho])$ serves as a first consistency check. Additionally, testing

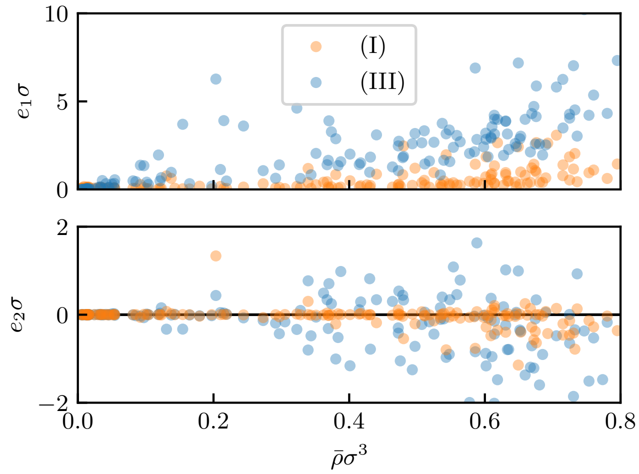

FIG. S4. Test of the Noether sum rules (16) and (17) as obtained from neural functionals trained on the basis of inhomogeneous one-body profiles (I, orange) and via bulk pair-correlation matching (III, blue). The respective errors e_1 and e_2 of the predictions are considerably larger in the latter case.

the validity of the sum rules

$$\nabla c_1(\mathbf{r};[\rho]) = \int \mathrm{d}\mathbf{r}'\, c_2(\mathbf{r},\mathbf{r}';[\rho])\nabla'\rho(\mathbf{r}'), \tag{16}$$

$$\int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r}) \int \mathrm{d}\mathbf{r}'\,\rho(\mathbf{r}')\nabla c_2(\mathbf{r},\mathbf{r}';[\rho]) = 0, \tag{17}$$

which result from thermal Noether invariance [5], constitutes a further valuable test of the fitness of a neural functional [1].

We depict in Fig. S3 results for $c_2(x,x';[\rho])$ with inhomogeneous density input as obtained from neural functionals that have been trained solely with inhomogeneous one-body data (I) and with bulk pair-correlations (III), respectively. The results from the pair-correlation-matched neural functional do not display the expected symmetry of the two-body direct correlation functional upon exchanging the spatial coordinates. We recall that within the local learning strategy, the exchange symmetry is not automatically enforced, unlike in Ref. [4] where the global excess free energy acts as the output of their neural network.

In Fig. S4, we show the discrepancy of the Noether sum rules (16) and (17) as defined via $e_1 = \|\partial_x c_1(x;[\rho]) - \int \mathrm{d}x'\, c_2(x,x';[\rho])\partial_{x'}\rho(x')\|_\infty$ and $e_2 = \int \mathrm{d}x\,\rho(x)\int \mathrm{d}x'\,\rho(x')\partial_x c_2(x,x';[\rho])$. The results of the neural functional from pure pair-correlation matching (III) show considerable deviations from these sum rules.

V. NEURAL FREE ENERGY METHODS

As an alternative to working with one-body direct correlations, we also consider neural networks that give immediate access to the excess free energy functional $F_{\mathrm{exc}}[\rho]$. We base the construction of these neural free energy functionals on the parametrization

$$F_{\mathrm{exc}}[\rho] = \int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r}) f_{\mathrm{exc}}(\mathbf{r};[\rho]), \tag{18}$$

which provides a well-defined localization of the excess free energy via the uniquely determined quantity

$$f_{\mathrm{exc}}(\mathbf{r};[\rho]) = -k_B T \int_0^1 \mathrm{d}a\, c_1(\mathbf{r};[a\rho]). \tag{19}$$

A. Local learning of $f_{\mathrm{exc}}(\mathbf{r};[\rho])$

As a first approach, it is hence natural to construct a local neural-network-based representation of $f_{\mathrm{exc}}(\mathbf{r};[\rho])$. We specialize to the considered planar geometry and reuse the MLP architecture of the neural correlation functional as described in Sec. I B and shown in Fig. S5.

Training is facilitated by inhomogeneous one-body matching (2), where one determines

$$c_1(\mathbf{r};[\rho]) = -\beta f_{\mathrm{exc}}(\mathbf{r};[\rho]) - \int \mathrm{d}\mathbf{r}'\,\rho(\mathbf{r}')\frac{\delta\beta f_{\mathrm{exc}}(\mathbf{r};[\rho])}{\delta\rho(\mathbf{r}')}, \tag{20}$$

with the functional derivative of $f_{\mathrm{exc}}(\mathbf{r};[\rho])$ being obtained by automatic differentiation within the training loop. To be able to evaluate the integral in a local learning scheme, the arguments $\mathbf{r}$ and $\mathbf{r}'$ have been swapped according to the expected symmetry $\delta f_{\mathrm{exc}}(\mathbf{r};[\rho])/\delta\rho(\mathbf{r}') = \delta f_{\mathrm{exc}}(\mathbf{r}';[\rho])/\delta\rho(\mathbf{r})$ that arises from Eq. (19) and $c_2(\mathbf{r},\mathbf{r}';[\rho]) = c_2(\mathbf{r}',\mathbf{r};[\rho])$. Crucially, this symmetry is not enforced intrinsically for an untrained neural functional for $f_{\mathrm{exc}}(\mathbf{r};[\rho])$, but it rather arises only after sufficient training. Utilizing the interchangeability of $\mathbf{r}$ and $\mathbf{r}'$ in Eq. (20) for the calculation of one-body direct correlations during training is nevertheless valid, as this leaves the true minimum of the matching condition (2) unchanged.

Applying the local neural functional for the prediction of free energy values is straightforward via the numerical evaluation of Eq. (18) for the considered density profile $\rho(x)$. For the prediction of equilibrium density profiles, $c_1(x;[\rho])$ is obtained in the considered planar geometry from Eq. (20) for use in the self-consistent solution of Eq. (3).

B. Convolutional neural network for $F_{\mathrm{exc}}[\rho]$

As an alternative to the local learning of $f_{\mathrm{exc}}(\mathbf{r};[\rho])$, we use a convolutional neural network (CNN) to represent the global quantity $F_{\mathrm{exc}}[\rho]$ while retaining the specific structure of Eq. (18). This approach bears similarity to the neural network construction in Ref. [4], but crucially differs in the utilization of a specific parametrization, Eq. (18), and in the chosen architecture, which is described in the following and illustrated in Fig. S6.

The CNN consists of 8 convolutional layers with kernel size 11, [16, 16, 32, 32, 64, 64, 16, 1] filters and dilation rates of [1, 2, 4, 8, 16, 32, 64, 1]. These hyperparameters are chosen to enable the propagation of non-local density information in a range that is similar to the window widths of the MLPs used in the previous local learning methods. We use cyclic padding for the convolutions, as is appropriate for the periodic boundary conditions of the considered systems, and employ no pooling or other coarse-graining layers, thus keeping full spatial resolution throughout the network. As a consequence of this fully convolutional architecture, the network remains applicable to virtually arbitrary system sizes and hence facilitates predictions “beyond-the-box” [1], as is also the case for the local learning approaches.

To arrive at the value of $F_{\rm exc}[\rho]$ in the spirit of Eq. (18), the last layer is multiplied pointwise with the input density profile $\rho(\mathbf{r})$ and integrated over the whole domain. In particular, this implies that this last convolutional layer acts as a representation of the one-body profile $f_{\rm exc}(\mathbf{r};[\rho])$.

For application in the self-consistent calculation of density profiles, automatic differentiation of the $F_{\rm exc}[\rho]$-CNN with respect to the input density $\rho(\mathbf{r})$ yields the complete one-body direct correlation profile $c_1(\mathbf{r};[\rho])$ to be used in Eq. (3).

[1] F. Sammüller, S. Hermann, D. de las Heras, and M. Schmidt, Neural functional theory for inhomogeneous fluids: Fundamentals and applications, Proc. Natl. Acad. Sci. **120**, e2312484120 (2023); simulation data and models deposited in Zenodo, doi:10.5281/zenodo.8380004.

[2] N. B. Wilding, Critical-point and coexistence-curve properties of the Lennard-Jones fluid: A finite-size scaling study, Phys. Rev. E **52**, 602 (1995).

[3] Code, simulation data, and neural functionals are available at: https://github.com/sfalmo/local-pair-matching.

[4] J. Dijkman, M. Dijkstra, R. van Roij, M. Welling, J.-W. van de Meent, and B. Ensing, Learning Neural Free-Energy Functionals with Pair-Correlation Matching, arXiv:2403.15007.

[5] S. Hermann and M. Schmidt, Noether's theorem in statistical mechanics, Commun. Phys. **4**, 176 (2021).

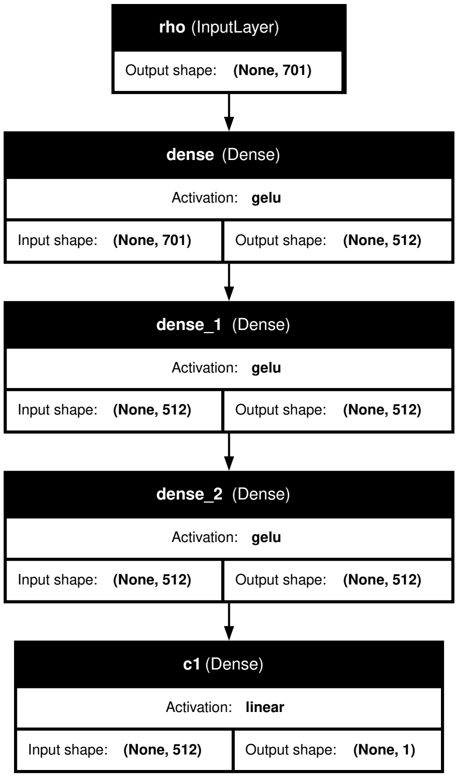

FIG. S5. Architecture of the MLP for local learning of $c_1(x; [\rho])$, cf. Sec. I B, and $f_{\mathrm{exc}}(x; [\rho])$, cf. Sec. V A. The input and output shapes indicate the variable batch size ("None") and the number of nodes for each layer.

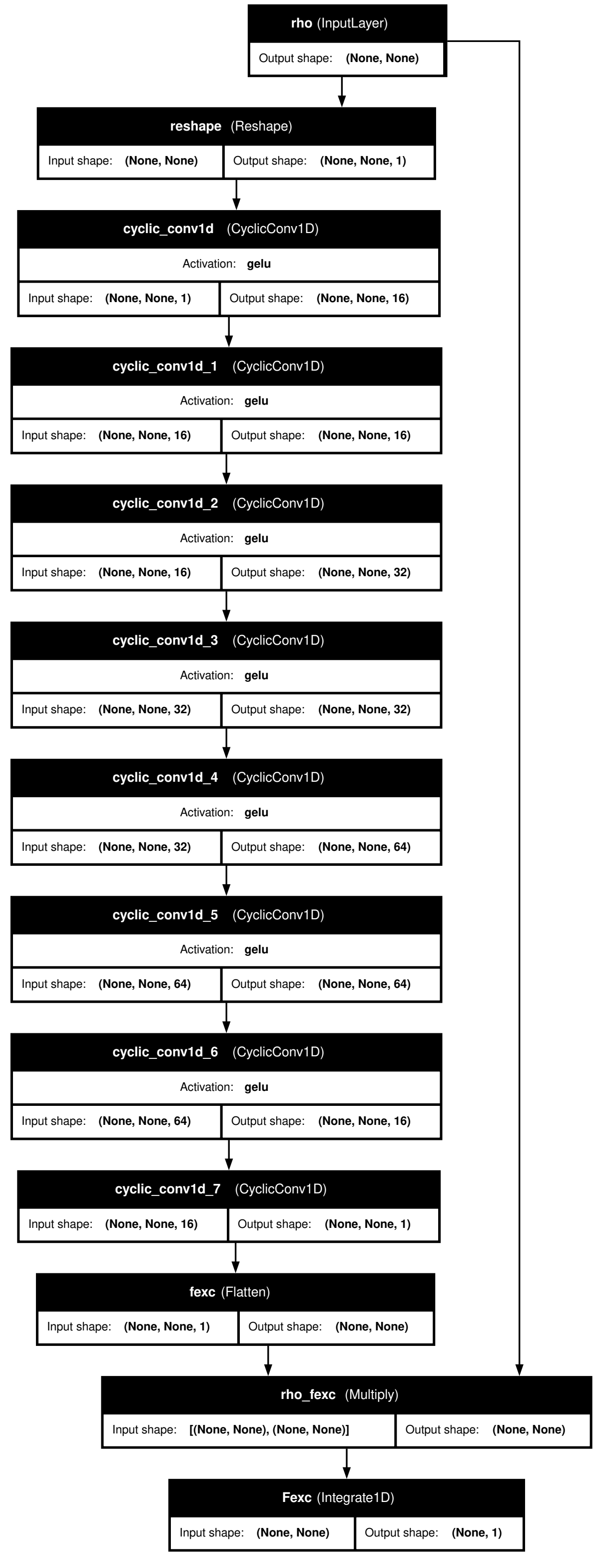

FIG. S6. Architecture of the CNN for $F_{\mathrm{exc}}[\rho]$, cf. Sec. V B. The input and output shapes indicate the batch size, the number of spatial bins ("None" for variable size) and the number of channels.