

Multivariate hyperdensity functional theory for inhomogeneous equilibrium fluids: From first principles to simulation-based machine learning

Florian Sammüller* and Matthias Schmidt†

Theoretische Physik II, Physikalisches Institut, Universität Bayreuth, D-95447 Bayreuth, Germany

(Dated: 1 September 2026)

Hyperdensity functional theory [Sammüller *et al.*, Phys. Rev. Lett. **133**, 098201 (2024)] facilitates the investigation of the equilibrium behavior of a general order parameter or statistical mechanical observable in spatially inhomogeneous classical many-body systems. The approach is based on applying the exact Mermin-Evens classical density functional mapping to an extended ensemble. Here we present the multivariate generalization for investigating simultaneously the properties and interrelations of several different hyperobservables of choice. The resulting framework gives rise to a systematic characterization and prediction scheme for general many-body phenomena. All pertinent equilibrium averages, variances, and covariances constitute universal density functionals, as we demonstrate explicitly. Associated one-body hyperfluctuation profiles quantify the degree of correlation of the local density with first- and second-order combinations of hyperobservables. These multivariate hyperfluctuation profiles are accessible in many-body simulations and they satisfy exact hyper-Ornstein-Zernike equations, which we derive from the minimization principle in the extended multivariate ensemble. The formal structure of the theory integrates naturally with supervised machine learning, which renders all hyperdensity functionals accessible in practice via training of neural networks on simulation data. We demonstrate all salient techniques using the illustrative case of clustering in confined hard rod fluids, thereby choosing the total number of particles and the largest cluster size as the representative hyperobservables of interest. Our numerical methodology enables the efficient and successful prediction of all statistical quantities induced by the chosen hyperobservables, which we verify via comparison to test data and which we attribute to the tight interplay of first-principles and machine-learning concepts that our general approach combines.

I. INTRODUCTION

Classical density functional theory [1–5] is a powerful approach for treating the collective behavior of many-body systems. The theory is based on using the one-body density distribution $\rho(\mathbf{r})$, where $\mathbf{r}$ indicates position, as its central variable. The formally exact Mermin-Evens minimization principle [1, 6] ascertains that the grand potential functional $\Omega[\rho]$ is minimized by the true equilibrium density distribution and that the value of the functional is the true equilibrium grand potential. Going beyond the density “profile” $\rho(\mathbf{r})$, the local compressibility $\chi_\mu(\mathbf{r}) = \partial\rho(\mathbf{r})/\partial\mu$, i.e., the partial derivative of the density profile with respect to the chemical potential μ , was put forward [7–12], along with more general fluctuation profiles [13–15], in order to quantify and rationalize complex interfacial behavior, such as drying of a solvent near substrates and around solutes [7–12, 15].

Within classical density functional theory the complexity of the many-body problem is tucked away in the excess (over ideal gas) intrinsic free energy functional $F_{\text{exc}}[\rho]$, which constitutes the nontrivial contribution to $\Omega[\rho]$ and which is unknown in general. Recent progress on the basis of machine learning has propelled the decade-long quest for the construction of reliable approximations for $F_{\text{exc}}[\rho]$ to a new and arguably very promising level. Both classically [16–33] and within the quantum counter-

part of the theory [34–41], as well as for nonequilibrium situations [42, 43] based on power functional theory [5], a range of different methods has been put forward. In particular, representing the one-body direct correlation functional [1, 2, 4] locally by a standard feed-forward neural network [26–33] proves to be a versatile and computationally straightforward approach to use simulation results as training data for supervised machine learning of the underlying functional mapping; we refer the reader to Ref. [29] for generalizations of this local learning technique and the relationship to other methods [24, 44].

Tight integration of the fundamental statistical mechanics turns out to be a crucial aspect in many recent machine learning methods as well as in conceptual and numerical workflows. Classical density functional theory in particular provides a rigorous mathematical footing that naturally suits the incorporation of neural networks and modern computational methods such as automatic differentiation. Unlike in other common applications of machine learning techniques, the target mapping that is represented by a neural network does not constitute a quantity that is of immediate and direct interest in applications. Instead, focusing on the arguably abstract yet fundamental one-body direct correlation functional allows one to keep the entire formal structure of classical density functional theory intact, thereby enabling a consistent and unified starting point for the determination of a wide range of desired quantities. Exploiting underlying theoretical principles is hence crucial for conceiving accurate and physically interpretable machine learning methodologies. Vice versa, the inclusion of machine learning triggers strong incentive to pursue fundamental

* Florian.Sammueler@uni-bayreuth.de

† Matthias.Schmidt@uni-bayreuth.de

theory development, as we demonstrate in the present work.

The recent hyperdensity functional theory [45, 46] constitutes a conceptual step beyond standard classical density functional theory [1–4], as it allows one to address the equilibrium average of a general, i.e., virtually arbitrary, observable of interest. Such a “hyperobservable” is represented by a classical phase space function $\hat{A}$ and the theory facilitates to formulate its average as a universal density functional $A[\rho]$. That the density profile alone is enough to determine the thermal mean of *any* quantity $\hat{A}$ for a given fluid type, as specified by its intrinsic Hamiltonian and for prescribed temperature T , is an immediate mathematical consequence of the Mermin-Evans map [1, 6].

Hyperdensity functional theory [45, 46] renders this merely formal theorem operational for the purpose of addressing concrete physical problems and phenomena. This is achieved by formulating a route toward expressing $A[\rho]$ as a practically accessible density functional. For a chosen hyperobservable $\hat{A}$, the theory gives rise to a corresponding hyperfluctuation profile $\chi_A(\mathbf{r})$, which correlates local density fluctuations with the instantaneous value of $\hat{A}$ for each microstate of the many-body phase space. Furthermore, a hyperdirect correlation functional $c_A(\mathbf{r}, [\rho])$ yields $A[\rho]$ upon functional line integration and this object is amenable to supervised machine learning. An exact hyper-Ornstein-Zernike identity interrelates the one-body profiles $\rho(\mathbf{r})$, $\chi_A(\mathbf{r})$, and $c_A(\mathbf{r})$ in the form of a nontrivial integral equation. In addition, hyperdensity functional theory is deeply connected with the hyperforce framework [46–48], which arises as a consequence of gauge invariance [49, 50] and Noether’s theorem in statistical mechanics [51–58]; see Ref. [59] for a viewpoint and Ref. [60] for a popular account of these developments.

Here we investigate ramifications of the hyperdensity functional approach for the general multivariate case of a set of observables $\{\hat{A}_\ell\}$, where ℓ enumerates all items. The multivariate extension is important as it allows one to study systematically the statistical interrelationships between several observables that are simultaneously of interest. In particular, the generalization yields exhaustive insight into the characteristics of higher-order correlation functions and fluctuations, which is already pertinent for the case of a single observable and which constitutes a significant expansion of the concepts laid out in Refs. [45, 46]. Quantifying the emergent correlation structure forms a crucial prerequisite for identifying and rationalizing the causation of physical mechanisms in any considered many-body system. Importantly, the multivariate hyperdensity functional formalism thereby rests firmly on first principles, which determines uniquely the specific forms and mutual relationships of the relevant correlation measures.

We exploit this formal structure of the theory to develop a corresponding multistage machine learning scheme for the iterative training of a hierarchy of neural hyperdirect correlation functionals. These provide

full quantitative closure of the theory, enabling fast and accurate predictions and opening up routes toward carrying out inverse design and high-throughput analysis tasks. An indispensable conceptual and computational feature is the identification and utilization of Gâteaux derivatives, which perform directional functional differentiation to capture the nonlocal correlations in key parts of the framework. We leverage the immediate access to this important type of functional derivative via modern automatic differentiation techniques and develop efficient evaluation schemes in our numerical workflow.

The paper is organized as follows. We present the formulation of the multivariate hyperdensity functional theory in Sec. II, which encompasses a description of the extended statistical mechanical ensemble in Sec. II A, the definition of hyperobservables and their associated hyperfluctuation profiles in Sec. II B, the derivation of first- and second-order hyper-Ornstein-Zernike equations and hyperdirect correlation functionals in Sec. II C, a description of averages and covariances of hyperobservables expressed as density functionals in Sec. II D, and a practically relevant reformulation of the hyper-Ornstein-Zernike equations in terms of Gâteaux derivatives in Sec. II E. Consequences and applications of the theory are laid out in Sec. III. We discuss our multistage protocol for acquiring multivariate neural hyperdensity functionals via supervised machine learning in Sec. III A and apply the framework for investigating the clustering statistics of the hard rod fluid in Sec. III B. We give conclusions and an outlook to future extensions and applications in Sec. IV.

II. THEORY

A. Extended ensemble

The system of interest is specified by its Hamiltonian H . We consider an extended ensemble that is characterized by a set $\{\hat{A}_\ell\}$, where the index $\ell = 1, \dots, L$ enumerates different observables $\hat{A}_\ell$, which may be virtually arbitrary phase space functions. For terminological disambiguation to other (generic) observables, we will refer to the specific set $\{\hat{A}_\ell\}$ that pertains to the following derivations as “hyperobservables”. The extended grand ensemble is defined by the probability distribution function $f = \exp(-\beta(H - \mu N) + \sum_\ell \lambda_\ell \hat{A}_\ell) / \Xi$, where the normalization constant is the partition sum $\Xi = \text{Tr} \exp(-\beta(H - \mu N) + \sum_\ell \lambda_\ell \hat{A}_\ell)$, with N denoting the particle number and $\beta = 1/(k_B T)$ with the Boltzmann constant k_B . The operator Tr denotes the standard classical trace over phase space such that averages are defined as $\langle \cdot \rangle = \text{Tr} \cdot f$. The grand potential of the extended ensemble is $\Omega = -k_B T \ln \Xi$.

The equilibrium distribution function f in the extended ensemble contains in its exponent a sum over all hyperobservables, with the coupling constants λ_ℓ determining the strength of the contribution of the ℓ th hyper-

observable. However, we are interested ultimately only in the case of vanishing coupling, $\lambda_\ell = 0$ for all ℓ , such that the equilibrium statistical physics that corresponds to the original Hamiltonian H under consideration is restored. Hence, when derivatives with respect to λ_ℓ are considered, it is generally implied that these are evaluated at $\lambda_\ell = 0$. For higher-order derivatives, we take the corresponding convention that the evaluation at vanishing coupling occurs after the variation has been performed for all involved coupling parameters.

An analogous way to view the extended ensemble is to consider instead an extended Hamiltonian $H_A = H - \sum_\ell \lambda_\ell \hat{A}_\ell / \beta$ which enters the standard Boltzmann factor $\exp(-\beta(H_A - \mu N))$. This setup leads to the identical partition sum and grand potential as described above.

B. Hyperobservables and hyperfluctuation profiles

We wish to address the thermal expectation value A_ℓ of the ℓ th hyperobservable as well as the covariance $A_{\ell\ell'}$ of the ℓ th and the ℓ' th hyperobservable, and hence define

$$A_\ell = \langle \hat{A}_\ell \rangle, \quad (1)$$

$$A_{\ell\ell'} = \text{cov}(\hat{A}_\ell, \hat{A}_{\ell'}), \quad (2)$$

where the covariance of two phase space functions $\hat{X}$ and $\hat{Y}$ is defined in the common way, $\text{cov}(\hat{X}, \hat{Y}) = \langle \hat{X}\hat{Y} \rangle - XY$, with the averages $X = \langle \hat{X} \rangle$ and $Y = \langle \hat{Y} \rangle$. Note that for $\ell = \ell'$, Eq. (2) reduces to $A_{\ell\ell} = \text{var}(\hat{A}_\ell)$, where the variance of a phase space function $\hat{X}$ is defined as $\text{var}(\hat{X}) = \langle \hat{X}^2 \rangle - X^2$.

Equivalently, A_ℓ and $A_{\ell\ell'}$ can be expressed in terms of first and second derivatives of the grand potential with respect to the coupling parameters, which is straightforward to show via differentiating the elementary definition of Ω :

$$A_\ell = - \left. \frac{\partial \beta \Omega}{\partial \lambda_\ell} \right|_{V_{\text{ext}}}, \quad (3)$$

$$A_{\ell\ell'} = - \left. \frac{\partial^2 \beta \Omega}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_{V_{\text{ext}}}. \quad (4)$$

In the notation we have indicated that the form of the external potential $V_{\text{ext}}(\mathbf{r})$ is kept fixed upon building the parametric derivatives; also μ and T are fixed (and suppressed in the notation). By inserting Eq. (3) into Eq. (4) one can straightforwardly show

$$A_{\ell\ell'} = \left. \frac{\partial A_\ell}{\partial \lambda_{\ell'}} \right|_{V_{\text{ext}}} = \left. \frac{\partial A_{\ell'}}{\partial \lambda_\ell} \right|_{V_{\text{ext}}}, \quad (5)$$

which is reminiscent of a Maxwell identity. The exchange symmetry of the indices ℓ and ℓ' in Eq. (5) is facilitated by exchanging the order of the two parametric derivatives acting on Ω in Eq. (4).

We define the following sets of first- and second-order hyperfluctuation profiles:

$$\chi_\ell(\mathbf{r}) = \text{cov}(\hat{\rho}(\mathbf{r}), \hat{A}_\ell), \quad (6)$$

$$\chi_{\ell\ell'}(\mathbf{r}) = \text{triv}(\hat{\rho}(\mathbf{r}), \hat{A}_\ell, \hat{A}_{\ell'}), \quad (7)$$

where $\hat{\rho}(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i)$ is the standard one-body density “operator” with $\delta(\cdot)$ indicating the Dirac distribution and the sum running over all particles $i = 1, \dots, N$ that are currently in the system. The first-order hyperfluctuation profile $\chi_\ell(\mathbf{r})$ is in direct generalization of the univariate case [45, 46]. The second-order hyperfluctuation profile $\chi_{\ell\ell'}(\mathbf{r})$ is given in Eq. (7) in correlator form as the triple covariance (“trivariance”), which we define for three phase space functions $\hat{X}$, $\hat{Y}$, and $\hat{Z}$ as

$$\begin{aligned} \text{triv}(\hat{X}, \hat{Y}, \hat{Z}) &= \langle \hat{X}\hat{Y}\hat{Z} \rangle - XYZ \\ &\quad - \text{cov}(\hat{X}, \hat{Y})Z \\ &\quad - \text{cov}(\hat{Y}, \hat{Z})X \\ &\quad - \text{cov}(\hat{X}, \hat{Z})Y, \end{aligned} \quad (8)$$

where $Z = \langle \hat{Z} \rangle$, which can alternatively be written as

$$\begin{aligned} \text{triv}(\hat{X}, \hat{Y}, \hat{Z}) &= \langle \hat{X}\hat{Y}\hat{Z} \rangle + 2XYZ \\ &\quad - \langle \hat{X}\hat{Y} \rangle Z - \langle \hat{Y}\hat{Z} \rangle X - \langle \hat{X}\hat{Z} \rangle Y. \end{aligned} \quad (9)$$

More generally, the hyperfluctuation profiles can be defined as joint cumulants of $\hat{\rho}(\mathbf{r})$ with combinations of hyperobservables $\hat{A}_\ell, \hat{A}_{\ell'}, \dots$, whereby the number of considered hyperobservables determines the order of the hyperfluctuation profile. This definition is in line with the covariance and trivariance expressions (6) and (7), and it allows one in principle to continue the hierarchy of hyperfluctuation profiles. In the same fashion, the averages (1) and covariances (2) form the two lowest-order members of the hierarchy of joint cumulants of hyperobservables. This general structure originates from the term $\exp(\sum_\ell \lambda_\ell \hat{A}_\ell)$ in the generalized Boltzmann factor, which renders the grand potential a multivariate cumulant generating function. That the parametric derivative expressions (3) and (4) hold confirms the fact that both A_ℓ and $A_{\ell\ell'}$ constitute joint cumulants.

Similar considerations for the hyperfluctuation profiles give rise to

$$\chi_\ell(\mathbf{r}) = \left. \frac{\partial \rho(\mathbf{r})}{\partial \lambda_\ell} \right|_{V_{\text{ext}}}, \quad (10)$$

$$\chi_{\ell\ell'}(\mathbf{r}) = \left. \frac{\partial^2 \rho(\mathbf{r})}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_{V_{\text{ext}}}, \quad (11)$$

where the partial derivatives are performed again upon keeping μ, T , and $V_{\text{ext}}(\mathbf{r})$ fixed. Inserting Eq. (10) into Eq. (11) yields

$$\chi_{\ell\ell'}(\mathbf{r}) = \left. \frac{\partial \chi_\ell(\mathbf{r})}{\partial \lambda_{\ell'}} \right|_{V_{\text{ext}}} = \left. \frac{\partial \chi_{\ell'}(\mathbf{r})}{\partial \lambda_\ell} \right|_{V_{\text{ext}}}, \quad (12)$$

whereby the labels ℓ and ℓ' may again be exchanged due to rearranging the order of differentiation in Eq. (11), in analogy to Eq. (5).

The first-order hyperfluctuation profiles, as defined either in covariance form (6) or as parametric derivatives (10), are a generalization of the local compressibility [8, 10, 12–15, 61, 62] and the local thermal susceptibility [13–15]. These latter one-body profiles have been used extensively to address the physics of solvophobicity and hydrophobicity. Wilding and Turci [61] have recently introduced the second-order pressure derivative of the density profile, $\partial^2 \rho(\mathbf{r})/\partial p^2$, for their investigation of surface criticality; our second-order hyperfluctuation profiles [Eqs. (7) and (11)] can be seen as a formally analogous generalization.

As a precursor to the functional dependencies that we seek to identify, we express the first- and second-order hyperfluctuation profiles as the following functional derivatives:

$$\chi_\ell(\mathbf{r}) = -\frac{\delta A_\ell}{\delta \beta V_{\text{ext}}(\mathbf{r})}, \quad (13)$$

$$\chi_{\ell\ell'}(\mathbf{r}) = -\frac{\delta A_{\ell\ell'}}{\delta \beta V_{\text{ext}}(\mathbf{r})}, \quad (14)$$

where we recall that A_ℓ is the average (1) of the hyperobservable $\hat{A}_\ell$ and $A_{\ell\ell'}$ is the covariance (2) of $\hat{A}_\ell$ and $\hat{A}_{\ell'}$. Here $\delta/\delta \beta V_{\text{ext}}(\mathbf{r})$ denotes the functional derivative with respect to the scaled external potential $\beta V_{\text{ext}}(\mathbf{r})$; trivially one has $\delta/\delta \beta V_{\text{ext}}(\mathbf{r}) = \beta^{-1} \delta/\delta V_{\text{ext}}(\mathbf{r})$ as $\beta = \text{const.}$ Both Eqs. (13) and (14) follow from functional differentiation of Eqs. (3) and (4) with respect to $-\beta V_{\text{ext}}(\mathbf{r})$, recalling $\rho(\mathbf{r}) = \delta\Omega/\delta V_{\text{ext}}(\mathbf{r})$, exchanging the order of derivatives, and identifying the hyperfluctuation profiles via Eqs. (10) and (11). Alternatively, one can start from the definition (1) of A_ℓ , utilize the general identity $-\delta\langle\hat{X}\rangle/\delta \beta V_{\text{ext}}(\mathbf{r}) = \text{cov}(\hat{\rho}(\mathbf{r}), \hat{X})$ for any phase space function $\hat{X}$, which can be verified by explicit calculation of the functional derivative [14], and identify the covariance expression (6) of $\chi_\ell(\mathbf{r})$. Similarly, using the definition (2) of $A_{\ell\ell'}$ and calculating the functional derivatives of the individual averages that contribute to the covariance then allows one to identify the trivariance expression (7) of $\chi_{\ell\ell'}(\mathbf{r})$.

The thus developed relationships demonstrate the prowess of the extended ensemble to generate meaningful statistical mechanical quantities via parametric derivatives [Eqs. (3), (4), (10), and (11)], which are associated to averages and covariances [Eqs. (1) and (2)] as well as to functional derivatives [Eqs. (13) and (14)] in the *standard* grand ensemble. All relationships are based on the elementary definition of the extended ensemble as described by its grand potential and we recall that we are ultimately interested only in the limit of vanishing coupling constants λ_ℓ .

C. Hyper-Ornstein-Zernike equations and hyperdirect correlation functionals

We turn to the density functional description and assume that the extended Hamiltonian $H_A = H - \sum_\ell \lambda_\ell \hat{A}_\ell/\beta$ is sufficiently well-behaved for small values of λ_ℓ such that the corresponding grand potential density functional $\Omega[\rho]$ exists. Then the Mermin-Evans minimization principle [1, 6] of classical density functional theory ascertains that:

$$\left. \frac{\delta\Omega[\rho]}{\delta\rho(\mathbf{r})} \right|_{H_A} = 0 \quad (\text{min}), \quad (15)$$

where $\delta/\delta\rho(\mathbf{r})$ denotes the functional derivative with respect to $\rho(\mathbf{r})$. The extended Hamiltonian H_A , including the form of the external potential $V_{\text{ext}}(\mathbf{r})$, is kept fixed together with μ and T upon building the density functional derivative. Equality in the relationship (15) is achieved for the true equilibrium density profile $\rho(\mathbf{r}) = \langle\hat{\rho}(\mathbf{r})\rangle$.

The grand potential density functional $\Omega[\rho]$ splits into several distinct additive contributions according to

$$\Omega[\rho] = F_{\text{id}}[\rho] + F_{\text{exc}}[\rho] + \int d\mathbf{r} \rho(\mathbf{r}) [V_{\text{ext}}(\mathbf{r}) - \mu], \quad (16)$$

where the ideal gas free energy functional is given explicitly as $F_{\text{id}}[\rho] = k_B T \int d\mathbf{r} \rho(\mathbf{r}) [\ln(\rho(\mathbf{r}) \Lambda^d) - 1]$ with the thermal de Broglie wavelength Λ and spatial dimensionality d ; we set $\Lambda = 1$ in the following. The intrinsic excess free energy functional $F_{\text{exc}}[\rho]$ accounts for the nontrivial interparticle interactions and it is hence unknown in general. Note that in our extended ensemble setting, the interaction effects can be interpreted to consist of both the unperturbed interparticle interaction potential $u(\mathbf{r}^N)$ and the coupling to the extension, i.e., $u(\mathbf{r}^N) - \sum_\ell \lambda_\ell \hat{A}_\ell/\beta$ in total. Therefore, $F_{\text{exc}}[\rho]$ depends implicitly on the coupling constants $\{\lambda_\ell\}$, which will be crucial in subsequent derivations.

Inserting the splitting (16) of $\Omega[\rho]$ into Eq. (15) leads to the standard Euler-Lagrange equation

$$c_1(\mathbf{r}, [\rho]) = \ln \rho(\mathbf{r}) + \beta V_{\text{ext}}(\mathbf{r}) - \beta \mu, \quad (17)$$

where $c_1(\mathbf{r}, [\rho]) = -\delta \beta F_{\text{exc}}[\rho]/\delta \rho(\mathbf{r})$ is the one-body direct correlation functional [1–5]. The algebraic form of Eq. (17) is thereby unchanged irrespective of working in the standard grand canonical or in the extended ensemble.

As the minimization principle (15) holds true for any (permissible) form of H_A , we retain a valid identity by differentiating Eq. (15) with respect to λ_ℓ and hence

$$\left. \frac{\partial}{\partial \lambda_\ell} \frac{\delta\Omega[\rho]}{\delta\rho(\mathbf{r})} \right|_{H_A} = 0. \quad (18)$$

Analogous to Eq. (18) we can differentiate the Euler-Lagrange equation (17) with respect to λ_ℓ , thereby keeping $V_{\text{ext}}(\mathbf{r})$ and the thermodynamic conditions μ, T fixed, but monitoring the changes in the equilibrium density

profile $\rho(\mathbf{r})$ upon changing λ_ℓ ; we recall the relationship (10) of $\chi_\ell(\mathbf{r})$ as is relevant for the use of the chain rule of differentiation. The result after rearrangement is the following (first-order) hyper-Ornstein-Zernike relation [45, 46]:

$$c_{1,\ell}^A(\mathbf{r}, [\rho]) = \frac{\chi_\ell(\mathbf{r})}{\rho(\mathbf{r})} - \int d\mathbf{r}' \chi_\ell(\mathbf{r}') c_2(\mathbf{r}, \mathbf{r}', [\rho]). \quad (19)$$

The (first-order) one-body hyperdirect correlation functional $c_{1,\ell}^A(\mathbf{r}, [\rho])$ is thereby given as:

$$c_{1,\ell}^A(\mathbf{r}, [\rho]) = \left. \frac{\partial c_1(\mathbf{r}, [\rho])}{\partial \lambda_\ell} \right|_\rho, \quad (20)$$

where the parametric derivative is taken at fixed density profile as is indicated in the notation. Note that the hyperdirect correlation functional $c_{1,\ell}^A(\mathbf{r}, [\rho])$ only tracks the *direct* effect of varying λ_ℓ , whereas changes of the equilibrium density profile $\rho(\mathbf{r})$ that are induced by changing λ_ℓ are captured by the second term on the right-hand side of Eq. (19), which introduces coupling via the standard two-body direct correlation density functional $c_2(\mathbf{r}, \mathbf{r}', [\rho]) = \delta c_1(\mathbf{r}, [\rho]) / \delta \rho(\mathbf{r}')$. Crucially, $c_{1,\ell}^A(\mathbf{r}, [\rho])$ as defined via Eq. (20) carries density functional dependence, as (i) $c_1(\mathbf{r}, [\rho])$ is a universal density functional and (ii) the minimization principle continues to apply in the extended ensemble. In Sec. II E, we demonstrate an alternative formulation of the hyper-Ornstein-Zernike equation (19) based on directional functional (Gâteaux) derivatives.

In order to address the coupling between different hyperobservables we go one order further and hence differentiate Eq. (18) with respect to a second parameter $\lambda_{\ell'}$ at fixed thermodynamic state point μ, T and external potential $V_{\text{ext}}(\mathbf{r})$. Building this second (mixed) parametric derivative of the Euler-Lagrange equation (17) is analogous to building the first derivative of the first-order hyper-Ornstein-Zernike relation (19) with respect to $\lambda_{\ell'}$. The result is in both cases the following second-order hyper-Ornstein-Zernike relation:

$$\begin{aligned} c_{1,\ell\ell'}^A(\mathbf{r}, [\rho]) &= \frac{\chi_{\ell\ell'}(\mathbf{r})}{\rho(\mathbf{r})} - \frac{\chi_\ell(\mathbf{r})\chi_{\ell'}(\mathbf{r})}{\rho(\mathbf{r})^2} \\ &\quad - \int d\mathbf{r}' \left[\chi_{\ell'}(\mathbf{r}') c_{2,\ell}^A(\mathbf{r}, \mathbf{r}') + \chi_\ell(\mathbf{r}') c_{2,\ell'}^A(\mathbf{r}, \mathbf{r}') \right] \\ &\quad - \int d\mathbf{r}' d\mathbf{r}'' \chi_\ell(\mathbf{r}') \chi_{\ell'}(\mathbf{r}'') c_3(\mathbf{r}, \mathbf{r}', \mathbf{r}'') \\ &\quad - \int d\mathbf{r}' \chi_{\ell\ell'}(\mathbf{r}') c_2(\mathbf{r}, \mathbf{r}'). \end{aligned} \quad (21)$$

We have organized Eq. (21) such that the terms on the right-hand side express the second-order one-body hyperdirect correlation functional

$$\begin{aligned} c_{1,\ell\ell'}^A(\mathbf{r}, [\rho]) &= \left. \frac{\partial^2 c_1(\mathbf{r}, [\rho])}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_\rho \\ &= \left. \frac{\partial c_{1,\ell}^A(\mathbf{r}, [\rho])}{\partial \lambda_{\ell'}} \right|_\rho, \end{aligned} \quad (22)$$

where the second equality follows from Eq. (20).

For brevity of notation we have dropped the density functional dependences of $c_{2,\ell}^A(\mathbf{r}, \mathbf{r}', [\rho])$, $c_2(\mathbf{r}, \mathbf{r}', [\rho])$, and $c_3(\mathbf{r}, \mathbf{r}', \mathbf{r}'', [\rho])$ on the right-hand side of Eq. (21), but emphasize that these are indeed all density functionals. The first-order two-body hyperdirect correlation functional is thereby defined as:

$$c_{2,\ell}^A(\mathbf{r}, \mathbf{r}', [\rho]) = \left. \frac{\partial c_2(\mathbf{r}, \mathbf{r}', [\rho])}{\partial \lambda_\ell} \right|_\rho, \quad (23)$$

in analogy to the one-body version (20). We can readily obtain the alternative form

$$c_{2,\ell}^A(\mathbf{r}, \mathbf{r}', [\rho]) = \frac{\delta c_{1,\ell}^A(\mathbf{r}, [\rho])}{\delta \rho(\mathbf{r}')} \quad (24)$$

by using the standard definition $c_2(\mathbf{r}, \mathbf{r}', [\rho]) = \delta c_1(\mathbf{r}, [\rho]) / \delta \rho(\mathbf{r}')$ in Eq. (23), reordering the parametric and the functional derivatives and then identifying $c_{1,\ell}^A(\mathbf{r}, [\rho])$ via Eq. (20). All integral expressions in the second-order hyper-Ornstein-Zernike equation (21) can be treated as directional functional derivatives using the Gâteaux calculus laid out in Sec. II E.

D. Hyperobservables as density functionals

We aim to find density functional expressions for both the mean of the ℓ th hyperobservable, A_ℓ , and for the global covariance of the ℓ th and the ℓ' th hyperobservable, $A_{\ell\ell'}$. We start with the relation (3) of A_ℓ and express the parametric derivative on the right-hand side as the following parametric derivative of the excess free energy functional:

$$\left. \frac{\partial \Omega[\rho]}{\partial \lambda_\ell} \right|_{V_{\text{ext}}} = \left. \frac{\partial F_{\text{exc}}[\rho]}{\partial \lambda_\ell} \right|_\rho, \quad (25)$$

where the form of the density profile $\rho(\mathbf{r})$ is kept fixed upon building the derivative of $F_{\text{exc}}[\rho]$. Equation (25) follows from using the chain rule and the splitting (16) of $\Omega[\rho]$. The arising terms may be simplified by identifying the minimization principle (15) as well as eliminating contributions independent of λ_ℓ , which leads to the compact and arguably surprising result (25). We refer the reader to Refs. [45, 46] for further details of this derivation.

Using relation (25) in Eq. (3) gives rise to the thermal average in density functional form:

$$A_\ell[\rho] = - \left. \frac{\partial \beta F_{\text{exc}}[\rho]}{\partial \lambda_\ell} \right|_\rho. \quad (26)$$

Due to being generated via a parametric derivative of $F_{\text{exc}}[\rho]$ for the extended Hamiltonian, the thermal average $A_\ell[\rho]$ inherits the property of being a universal density functional, as indicated in the notation.

In practice it is useful to substantiate Eq. (26) by expressing the excess free energy in terms of the functional line integral [1, 2]

$$\begin{aligned} F_{\text{exc}}[\rho] &= -k_{\text{B}}T \int \mathcal{D}[\rho] c_1(\mathbf{r}, [\rho]) \\ &= -k_{\text{B}}T \int d\mathbf{r} \rho(\mathbf{r}) \int_0^1 da c_1(\mathbf{r}, [a\rho]). \end{aligned} \quad (27)$$

The second line in Eq. (27) constitutes a linear parametrization of the general functional line integral $\int \mathcal{D}[\rho]$ whereby $a\rho(\mathbf{r})$, $0 \leq a \leq 1$, is a scaled version of the “target” density profile $\rho(\mathbf{r})$ that appears as the functional argument on the left hand side. Inserting Eq. (27) into Eq. (26), then moving the parametric derivative with respect to λ_ℓ inside of the integral and using the definition (20) of the first-order one-body hyperdirect correlation functional $c_{1,\ell}^A(\mathbf{r}, [\rho])$ yields:

$$A_\ell[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) \int_0^1 da c_{1,\ell}^A(\mathbf{r}, [a\rho]). \quad (28)$$

From either Eq. (26) or Eq. (28), one can further conclude that:

$$c_{1,\ell}^A(\mathbf{r}, [\rho]) = \frac{\delta A_\ell[\rho]}{\delta \rho(\mathbf{r})}, \quad (29)$$

which reveals that the first-order one-body hyperdirect correlation functional is a measure of (direct) changes of the thermal mean A_ℓ upon local changes in the density profile $\rho(\mathbf{r})$.

Equation (28) is of a form which facilitates explicit access to $A_\ell[\rho]$ provided that $c_{1,\ell}^A(\mathbf{r}, [\rho])$ is known. As demonstrated in Refs. [45, 46] for a single hyperobservable and in Sec. III for the multivariate case, a practical route to obtain the required hyperdirect correlation functionals $c_{1,\ell}^A(\mathbf{r}, [\rho])$ proceeds via simulation-based machine learning of neural functionals [27, 28].

We next turn to the covariance $A_{\ell\ell'}$ and recall its relationship (4) to the second mixed parametric derivative of the grand potential at fixed form of the external potential. We apply this operation to the grand potential density functional (16), which is analogous to building the derivative of Eq. (25) with respect to $\lambda_{\ell'}$. The result is:

$$\begin{aligned} \left. \frac{\partial^2 \Omega[\rho]}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_{V_{\text{ext}}} &= \left. \frac{\partial^2 F_{\text{exc}}[\rho]}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_\rho \\ &+ \int d\mathbf{r}' \chi_{\ell'}(\mathbf{r}') \frac{\delta}{\delta \rho(\mathbf{r}')} \left. \frac{\partial F_{\text{exc}}[\rho]}{\partial \lambda_\ell} \right|_\rho \\ &+ \frac{\delta}{\delta \rho(\mathbf{r}')} \int d\mathbf{r} \chi_\ell(\mathbf{r}) \left. \frac{\delta \Omega[\rho]}{\delta \rho(\mathbf{r})} \right|_{V_{\text{ext}}}, \end{aligned} \quad (30)$$

where we address each individual contribution in the following.

The third term on the right-hand side of Eq. (30) can be simplified by moving the density functional derivative

inside of the integral. Then applying the product rule of differentiation gives two terms, both of which vanish, one due to the minimization principle (15) and one due to proto-form (18) of the first-order hyper-Ornstein-Zernike equation. The second term on the right-hand side of Eq. (30) can be rewritten upon reordering the derivatives and successively identifying $c_1(\mathbf{r}, [\rho])$ in its standard form and $c_{1,\ell}^A(\mathbf{r}, [\rho])$ via Eq. (20). Recalling the second parametric derivative relationship (4), we obtain the desired covariance as a density functional in the following form:

$$A_{\ell\ell'}[\rho] = - \left. \frac{\partial^2 \beta F_{\text{exc}}[\rho]}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_\rho + \int d\mathbf{r} \chi_{\ell'}(\mathbf{r}) c_{1,\ell}^A(\mathbf{r}, [\rho]), \quad (31)$$

which holds per construction also for interchanged labels ℓ and ℓ' as a nontrivial symmetry of the second term on the right-hand side.

Alternatively, one can rewrite the second term on the right-hand side of Eq. (31) by multiplying the first-order hyper-Ornstein-Zernike relation (19) by $\chi_{\ell'}(\mathbf{r})$, exchanging indices ℓ and ℓ' , and then integrating over $\mathbf{r}'$. Insertion into Eq. (31) gives the density functional form:

$$\begin{aligned} A_{\ell\ell'}[\rho] &= - \left. \frac{\partial^2 \beta F_{\text{exc}}[\rho]}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_\rho + \int d\mathbf{r} \frac{\chi_\ell(\mathbf{r}) \chi_{\ell'}(\mathbf{r})}{\rho(\mathbf{r})} \\ &- \int d\mathbf{r} d\mathbf{r}' \chi_\ell(\mathbf{r}) \chi_{\ell'}(\mathbf{r}') c_2(\mathbf{r}, \mathbf{r}', [\rho]), \end{aligned} \quad (32)$$

which is equivalent to Eq. (31). Again it is useful to express the present second parametric derivative of the excess free energy functional via functional line integration. We hence write $F_{\text{exc}}[\rho]$ as the standard functional line integral (27) over $c_1(\mathbf{r}, [\rho])$, exchange the orders of derivatives and integrals and identify the second-order one-body hyperdirect correlation functional $c_{1,\ell\ell'}^A(\mathbf{r}, [\rho])$ according to Eq. (22). The result is the following functional line integral:

$$- \left. \frac{\partial^2 \beta F_{\text{exc}}[\rho]}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_\rho = \int d\mathbf{r} \rho(\mathbf{r}) \int_0^1 da c_{1,\ell\ell'}^A(\mathbf{r}, [a\rho]), \quad (33)$$

with the scaled density profile $a\rho(\mathbf{r})$, $0 \leq a \leq 1$, constituting the functional argument of the second-order hyperdirect correlation functional.

Insertion of Eq. (33) into the covariance density functional (31) yields:

$$\begin{aligned} A_{\ell\ell'}[\rho] &= \int d\mathbf{r} \rho(\mathbf{r}) \int_0^1 da c_{1,\ell\ell'}^A(\mathbf{r}, [a\rho]) \\ &+ \int d\mathbf{r} \chi_{\ell'}(\mathbf{r}) c_{1,\ell}^A(\mathbf{r}, [\rho]), \end{aligned} \quad (34)$$

where again the indices ℓ and ℓ' in the second term can be interchanged with no harm. From Eq. (32), we further

obtain the equivalent form

$$A_{\ell\ell'}[\rho] = \int d\mathbf{r}\rho(\mathbf{r}) \int_0^1 da c_{1,\ell\ell'}^A(\mathbf{r}, [a\rho]) + \int d\mathbf{r} \frac{\chi_\ell(\mathbf{r})\chi_{\ell'}(\mathbf{r})}{\rho(\mathbf{r})} - \int d\mathbf{r}d\mathbf{r}' \chi_\ell(\mathbf{r})\chi_{\ell'}(\mathbf{r}') c_2(\mathbf{r}, \mathbf{r}', [\rho]). \quad (35)$$

In Appendix A, we verify numerically the equivalence of Eq. (34) in its original form and with exchanged indices ℓ and ℓ' as well as Eq. (35), using the neural functional workflow laid out in Sec. III. Equations (31) and (34) further reveal that

$$c_{1,\ell\ell'}^A(\mathbf{r}, [\rho]) = \frac{\delta A_{\ell\ell'}[\rho]}{\delta \rho(\mathbf{r})} - \int d\mathbf{r}' \chi_{\ell'}(\mathbf{r}') c_{1,\ell}^A(\mathbf{r}, [\rho]). \quad (36)$$

We note the appearance of the second term on the right-hand side due to the chain rule, which might not be gleaned at first sight from the first-order result (29).

As an alternative derivation, the form (34) of the covariance density functional can be obtained from the relationship $A_{\ell\ell'} = \partial A_\ell / \partial \lambda_{\ell'}$, see Eq. (5), applied to the density functional $A_\ell[\rho]$ as given in Eq. (28). One hence needs to differentiate $A_\ell[\rho]$ with respect to $\lambda_{\ell'}$ upon fixing μ , T , and $V_{\text{ext}}(\mathbf{r})$. Differentiating the parametric dependence of $c_{1,\ell}^A(\mathbf{r}, [a\rho])$ on the right-hand side of Eq. (28) with respect to the coupling parameter generates the functional line integral over $c_{1,\ell\ell'}^A(\mathbf{r}, [\rho])$ in Eq. (34). The second term stems from the changes of the target density profile that forms the upper bound of the functional line integral in Eq. (28). Performing the derivative with respect to $\lambda_{\ell'}$ cancels the functional line integration, evaluates its integrand at the target density, and gives according to Eq. (10) a factor $\chi_{\ell'}(\mathbf{r})$ from the chain rule. The result is the second term in Eq. (34), which we reemphasize is no longer a functional line integral: While $c_{1,\ell}^A(\mathbf{r}, [\rho])$ is indeed a density functional, no integration over the functional argument takes place. Instead, the hyperdirect correlation functional is merely evaluated at density $\rho(\mathbf{r})$, and the result enters the standard position integral together with the hyperfluctuation profile $\chi_{\ell'}(\mathbf{r})$. We show in Sec. III that the density functional $A_{\ell\ell'}[\rho]$ is, analogous to $A_\ell[\rho]$, explicitly accessible via a multistage supervised machine learning workflow based on simulation data of a considered fluid in inhomogeneous environments.

The above process of expressing parametric derivatives of $\Omega[\rho]$ with respect to (multiple) coupling parameters $\lambda_\ell, \lambda_{\ell'}, \dots$ in terms of density functionals can be applied to generate a hierarchy of increasing complexity. As a representative example, differentiating Eq. (34) by $\lambda_{\ell''}$

leads to

$$A_{\ell\ell'\ell''}[\rho] = \int d\mathbf{r}\rho(\mathbf{r}) \int_0^1 da c_{1,\ell\ell'\ell''}^A(\mathbf{r}, [a\rho]) + \int d\mathbf{r} \chi_{\ell''}(\mathbf{r}) c_{1,\ell\ell'}^A(\mathbf{r}, [\rho]) + \int d\mathbf{r} [\chi_{\ell'\ell''}(\mathbf{r}) c_{1,\ell}^A(\mathbf{r}, [\rho]) + \chi_{\ell'}(\mathbf{r}) c_{1,\ell\ell''}^A(\mathbf{r}, [\rho])] + \int d\mathbf{r}d\mathbf{r}' \chi_{\ell'}(\mathbf{r}) \chi_{\ell''}(\mathbf{r}') c_{2,\ell}^A(\mathbf{r}, \mathbf{r}', [\rho]). \quad (37)$$

where $A_{\ell\ell'\ell''} = \text{triv}(\hat{A}_\ell, \hat{A}_{\ell'}, \hat{A}_{\ell''}) = \partial A_{\ell\ell'} / \partial \lambda_{\ell''}$ and $c_{1,\ell\ell'\ell''}^A(\mathbf{r}, [\rho]) = \partial c_{1,\ell\ell'}^A(\mathbf{r}, [\rho]) / \partial \lambda_{\ell''}$.

For completeness, the case of a single hyperobservable $\hat{A} = \hat{A}_1$ of interest is contained in the current theory simply via setting the number of hyperobservables $L = 1$. Then the sole remaining variance is $A_{11} = \text{var}(\hat{A})$. The second-order hyperfluctuation profile acquires corresponding meaning as $\chi_{11}(\mathbf{r}) = \text{triv}(\hat{\rho}(\mathbf{r}), \hat{A}, \hat{A}) = \text{cov}(\hat{\rho}(\mathbf{r}), \hat{A}^2) - 2\chi_1(\mathbf{r})A$. The second-order hyper-Ornstein-Zernike equation (21) can then be simplified as follows:

$$c_{1,11}^A(\mathbf{r}, [\rho]) = \frac{\chi_{11}(\mathbf{r})}{\rho(\mathbf{r})} - \frac{\chi_1(\mathbf{r})^2}{\rho(\mathbf{r})^2} - 2 \int d\mathbf{r}' c_{2,1}^A(\mathbf{r}, \mathbf{r}', [\rho]) \chi_1(\mathbf{r}') - \int d\mathbf{r}' d\mathbf{r}'' c_3(\mathbf{r}, \mathbf{r}', \mathbf{r}'', [\rho]) \chi_1(\mathbf{r}') \chi_1(\mathbf{r}'') - \int d\mathbf{r}' c_2(\mathbf{r}, \mathbf{r}', [\rho]) \chi_{11}(\mathbf{r}'), \quad (38)$$

which extends the univariate theory of Refs. [45, 46] to the second order.

E. Hyper-Ornstein-Zernike equations and Gâteaux derivatives

While the derivations of the previous sections give a complete account of the multivariate hyperdensity functional theory, we reexamine in the following a formalization of the functional calculus, which enables a significant reduction in complexity for both theoretical and practical considerations. An equivalent reformulation of integral expressions in the hyper-Ornstein-Zernike equations (19) and (21) in terms of generalized directional functional derivatives is laid out here. This sheds further light onto their underlying algebraic structure and proves to be crucial in subsequent numerical applications.

Starting with the first-order hyper-Ornstein-Zernike equation (19), we recall that the integral term on its right-hand side arises from the derivation via Eq. (18) due to the chain rule of functional differentiation. Fixing

the free position argument $\mathbf{r}$, the general structure of this prototypical expression is

$$\mathcal{D}_\psi \mathcal{F}[\rho] = \int d\mathbf{r}' \psi(\mathbf{r}') \frac{\delta \mathcal{F}[\rho]}{\delta \rho(\mathbf{r}')} \quad (39)$$

with generic function $\psi(\mathbf{r}')$ and generic functional $\mathcal{F}[\rho]$ (not to be confused with the intrinsic free energy functional $F[\rho]$). The second term in Eq. (19) is recovered explicitly by setting $\psi(\mathbf{r}') = \chi_\ell(\mathbf{r}')$ and $\mathcal{F}[\rho] = c_1(\mathbf{r}, [\rho])$ in Eq. (39).

By definition of the functional derivative [63], Eq. (39) may equivalently be expressed as the following Gâteaux differential:

$$\mathcal{D}_\psi \mathcal{F}[\rho] = \left. \frac{d\mathcal{F}[\rho + \epsilon\psi]}{d\epsilon} \right|_{\epsilon=0}, \quad (40)$$

where ϵ is a scalar parameter. Equation (40) makes apparent that terms of the form (39) constitute (generalized) directional derivatives of the functional $\mathcal{F}$ with respect to variations along ψ , evaluated at the functional argument ρ . The corresponding directional (Gâteaux) derivative operator $\mathcal{D}_\psi$ is introduced for succinct notation.

Evaluating terms of the form of the right-hand side of Eq. (39) via directional differentiation according to Eq. (40) is both straightforward and highly beneficial from a practical standpoint in numerical work. Automatic differentiation techniques facilitate immediate implementation of Eq. (40); we give an example in the programming language Julia in Appendix B. Utilizing this approach circumvents entirely the need to evaluate both the full functional derivative $\delta \mathcal{F}[\rho]/\delta \rho(\mathbf{r}')$ as well as the spatial integral in Eq. (39). Instead, only the variation with respect to a given perturbation $\epsilon\psi(\mathbf{r}')$ must be evaluated in Eq. (40), which leads to a substantial reduction in both computational and memory demand. As we demonstrate below, besides the strikingly compact code, the benefit lies in excellent performance. The approach is similar to calculating Jacobian-vector products directly, without intermediate evaluation of the entire Jacobian, as is key strategy in various numerical routines that arise particularly in machine learning disciplines [64].

In the second-order hyper-Ornstein-Zernike equation (21), the integral term featuring $c_3(\mathbf{r}, \mathbf{r}', \mathbf{r}'', [\rho])$ is of type

$$\mathcal{D}_{\psi_1} \mathcal{D}_{\psi_2} \mathcal{F}[\rho] = \int d\mathbf{r}' \int d\mathbf{r}'' \psi_1(\mathbf{r}') \psi_2(\mathbf{r}'') \frac{\delta^2 \mathcal{F}[\rho]}{\delta \rho(\mathbf{r}') \delta \rho(\mathbf{r}'')}, \quad (41)$$

which corresponds to a second-order Gâteaux derivative that is readily computed by a two-step application of Eq. (40). A code example which facilitates the efficient evaluation of such generalized vector-Hessian-vector products is given in Appendix B.

For completeness, we spell out the reformulated versions of the first- and second-order hyper-Ornstein-Zernike equations (19) and (21), where the arising in-

tegrals are reexpressed in terms of Gâteaux differentials:

$$c_{1,\ell}^A(\mathbf{r}, [\rho]) = \frac{\chi_\ell(\mathbf{r})}{\rho(\mathbf{r})} - \mathcal{D}_{\chi_\ell} c_1(\mathbf{r}, [\rho]), \quad (42)$$

$$\begin{aligned} c_{1,\ell\ell'}^A(\mathbf{r}, [\rho]) &= \frac{\chi_{\ell\ell'}(\mathbf{r})}{\rho(\mathbf{r})} - \frac{\chi_\ell(\mathbf{r})\chi_{\ell'}(\mathbf{r})}{\rho(\mathbf{r})^2} \\ &\quad - \mathcal{D}_{\chi_{\ell'}} c_{1,\ell}^A(\mathbf{r}, [\rho]) - \mathcal{D}_{\chi_\ell} c_{1,\ell'}^A(\mathbf{r}, [\rho]) \\ &\quad - \mathcal{D}_{\chi_\ell} \mathcal{D}_{\chi_{\ell'}} c_1(\mathbf{r}, [\rho]) - \mathcal{D}_{\chi_{\ell\ell'}} c_1(\mathbf{r}, [\rho]). \end{aligned} \quad (43)$$

We demonstrate below that Eqs. (42) and (43) constitute not merely formal rewritings, but rather crucial tools for dealing with the hyper-Ornstein-Zernike structure in numerical work.

III. APPLICATIONS

A. Simulation and machine learning perspective

In the following, we lay out our multistage approach for machine learning the direct and hyperdirect correlation functionals based on simulation data for inhomogeneous fluid equilibria. In Fig. 1 the general workflow is illustrated for a concrete application scenario, see Sec. III B. The neural functionals provide full access to averages $A_\ell[\rho]$ and covariances $A_{\ell\ell'}[\rho]$ as well as to spatially resolved first- and second-order hyperfluctuation profiles $\chi_\ell(\mathbf{r})$ and $\chi_{\ell\ell'}(\mathbf{r})$, as demonstrated below.

Let us first consider a specific inhomogeneous system of interest at fixed thermodynamic state point μ , T and with prescribed external potential $V_{\text{ext}}(\mathbf{r})$ to reexamine the accessibility of the different terms that arise in the hyper-Ornstein-Zernike relations (19) and (21), cf. also their Gâteaux forms (42) and (43). For capturing the right-hand side of the first-order hyper-Ornstein-Zernike relation (19), one requires to have results for the density profile $\rho(\mathbf{r})$ and the hyperfluctuation profiles $\chi_\ell(\mathbf{r})$ that arise in this system. The former quantity is straightforward to acquire in standard grand canonical Monte Carlo simulations and the latter is available via sampling the covariance (6) and crucially without having to represent the extended ensemble directly. From these results, the first-order one-body hyperdirect correlation function $c_{1,\ell}^A(\mathbf{r})$ that pertains to the specific system under investigation can be constructed explicitly via the first-order hyper-Ornstein-Zernike equation. For this, one either evaluates Eq. (19), where the standard two-body direct correlation functional $c_2(\mathbf{r}, \mathbf{r}', [\rho])$ can be obtained via (automatic) differentiation of a neural (or, if available, analytical) one-body direct correlation functional [27, 28]. Alternatively and more efficiently, Eq. (42) facilitates an equivalent calculation using automatic directional derivatives of $c_1(\mathbf{r}, [\rho])$.

For each system in the simulation data set, the measured quantities $\rho(\mathbf{r})$ and $\chi_\ell(\mathbf{r})$ as well as the density functional $c_1(\mathbf{r}, [\rho])$ are hence required in order to evaluate the corresponding profiles $c_{1,\ell}^A(\mathbf{r})$. Construction of

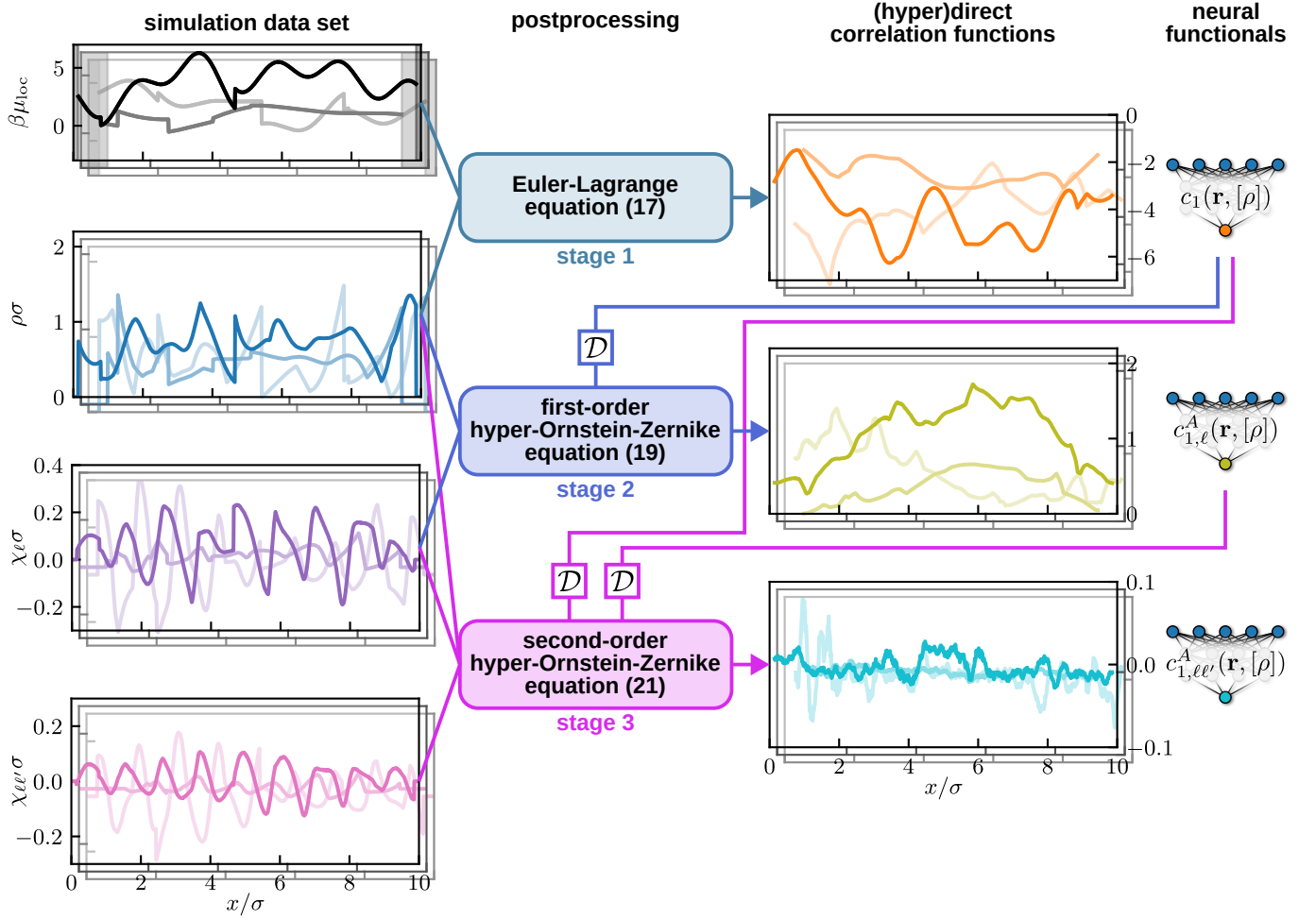


FIG. 1. Multistage workflow for machine learning multivariate hyperdensity functionals. A grand canonical Monte Carlo simulation data set (left column) of equilibria of a considered fluid type serves as ground truth. Inhomogeneity is induced in each simulation by a randomly chosen (scaled) local chemical potential $\beta\mu_{\text{loc}}(\mathbf{r}) = \beta\mu - \beta V_{\text{ext}}(\mathbf{r})$ and the resulting set of one-body profiles $\rho(\mathbf{r})$, $\chi_\ell(\mathbf{r})$, and $\chi_{\ell\ell'}(\mathbf{r})$ is sampled. We depict exemplary data for the one-dimensional hard rod fluid where $\mathbf{r} = x$. As hyperobservables of choice, we select the total number of particles, $\hat{A}_1$, and the number of particles in the largest cluster, $\hat{A}_2$, see the text for a detailed specification. Postprocessing of the simulation data occurs in three subsequent stages via the Euler-Lagrange equation (17), the first-order hyper-Ornstein-Zernike equation (19), and the second-order hyper-Ornstein-Zernike equation (21). The latter two equations are thereby implemented via Eqs. (42) and (43) with the symbol $\mathcal{D}$ denoting the evaluation of directional functional (Gâteaux) derivatives. The data set features hyperfluctuation and hyperdirect correlation profiles for all pertinent index combinations; we only show representative results for $\ell = 2$ and $\ell' = 1$. The resulting (hyper)direct correlation profiles are the target quantities for machine learning corresponding neural (hyper)density functionals $c_1(\mathbf{r}, [\rho])$, $c_{1,\ell}^A(\mathbf{r}, [\rho])$, and $c_{1,\ell\ell'}^A(\mathbf{r}, [\rho])$, which are fit for application in both standard and hyperdensity functional theory. We show the workflow for neural functional predictions in Fig. 2 as well as benchmark results in Fig. 3.

a neural functional via training on the thus acquired data set is feasible to obtain an accurate representation of the universal density functional mapping $c_{1,\ell}^A(\mathbf{r}, [\rho])$. Expressing the average $A_\ell[\rho]$ as a density functional is then viable by employing the neural hyperdirect correlation functional in the functional line integral (28). In Refs. [45, 46], applications of the first-order theory are laid out for a single hyperobservable, confirming the practical accessibility of all quantities as well as the possibility to machine learn the underlying functional mapping.

We turn to the second-order theory and consider

the right-hand side of the second-order hyper-Ornstein-Zernike relation, cf. Eqs. (21) and (43), where the density profile $\rho(\mathbf{r})$ and the hyperfluctuation profiles $\chi_\ell(\mathbf{r})$ and $\chi_{\ell\ell'}(\mathbf{r})$ are available in analogy to the first-order case. The second-order hyperfluctuation profile $\chi_{\ell\ell'}(\mathbf{r})$ can be obtained via the trivariate correlator (7), which must hence be sampled in all simulations. While the trivariate does converge more slowly than covariances or standard averages, we emphasize that for scalar observables $\hat{A}_\ell$ and $\hat{A}_{\ell'}$ of interest, $\chi_{\ell\ell'}(\mathbf{r})$ is still a one-body profile depending only on a single position argument $\mathbf{r}$. This is

beneficial from a practical point of view as it keeps the sampling feasible with standard histogram methods.

While both Eqs. (21) and (43) are equivalent formulations of the second-order hyper-Ornstein-Zernike relation, their numerical treatment differs drastically such that only the latter Eq. (43) constitutes the computationally feasible route, as we illustrate in the following. Dealing with Eq. (21) would formally require to evaluate all integral expressions on its right-hand side, thereby utilizing $c_2(\mathbf{r}, \mathbf{r}', [\rho])$, $c_3(\mathbf{r}, \mathbf{r}', \mathbf{r}'', [\rho])$, and $c_{2,\ell}^A(\mathbf{r}, \mathbf{r}', [\rho])$ as generalized integral kernels. The two- and three-body direct correlation functionals $c_2(\mathbf{r}, \mathbf{r}', [\rho])$ and $c_3(\mathbf{r}, \mathbf{r}', \mathbf{r}'', [\rho])$ can be obtained in principle from automatic differentiation of $c_1(\mathbf{r}, [\rho])$ [27, 28, 30], and $c_{2,\ell}^A(\mathbf{r}, \mathbf{r}', [\rho])$ follows from automatic differentiation of $c_{1,\ell}^A(\mathbf{r}, [\rho])$ according to Eq. (24). In (quasi)-one-dimensional geometries and on the first-order level of the theory, the evaluation of Eq. (19) remains feasible via this direct route, as demonstrated in Refs. [45, 46]. However, if spatial resolution is increased, e.g. when working in multiscale settings [27] or in more general geometries [65–68], the explicit evaluation of the full objects $c_2(\mathbf{r}, \mathbf{r}', [\rho])$, $c_3(\mathbf{r}, \mathbf{r}', \mathbf{r}'', [\rho])$, and $c_{2,\ell}^A(\mathbf{r}, \mathbf{r}', [\rho])$ becomes increasingly expensive. Concretely, for a discretized spatial grid with n grid points, $c_2(\mathbf{r}, \mathbf{r}', [\rho])$ comprises n^2 values which need to be computed and stored in a naive approach. On the second-order level of the theory, this issue becomes prohibitively expensive. For $c_3(\mathbf{r}, \mathbf{r}', \mathbf{r}'', [\rho])$, which appears in Eq. (21), n^3 entries would need to be evaluated and stored, thereby using computationally demanding second-order (e.g. forward-over-reverse) automatic differentiation techniques [64]. To circumvent this issue, we utilize the reformulation of the integral expressions in the hyper-Ornstein-Zernike equations (19) and (21) in terms of directional functional derivatives, see Eqs. (42) and (43). Exploiting the Gâteaux calculus enables a numerically efficient route toward evaluating all arising terms in both hyper-Ornstein-Zernike equations; we refer to Appendix B for implementation details.

Analogous to the first-order case, the resulting data set for $c_{1,\ell\ell'}^A(\mathbf{r})$ evaluated with the second-order hyper-Ornstein-Zernike equation across all simulated systems provides the basis for extracting the underlying density functional relationship $c_{1,\ell\ell'}^A(\mathbf{r}, [\rho])$ via a neural network. The covariances $A_{\ell\ell'}[\rho]$ are then accessible as genuine density functionals by evaluating either Eq. (34) or Eq. (35), whereby the second-order one-body hyperdirect correlation functional enters in the functional line integration, and the remaining functionals are available from either first-order hyperdensity $[c_{1,\ell}^A(\mathbf{r}, [\rho])]$ or standard density functional theory $[c_2(\mathbf{r}, \mathbf{r}', [\rho])]$. Similarly, the first-order hyperfluctuation profiles $\chi_\ell(\mathbf{r})$ that appear in Eq. (21) may be obtained by solution of Eq. (19), see also the account given in Refs. [45, 46]. We recall the illustration of the multistage machine learning of neural hyperdensity functionals in Fig. 1, and exemplify the workflow for the investigation of clustering statistics in the hard rod fluid, as laid out in Sec. III B.

From the standpoint of having all occurring density functionals available, either as analytic approximation or via representation in terms of neural networks, both hyper-Ornstein-Zernike relations (19) and (21) form defining integral equations for the first- and second-order hyperfluctuation profiles. Specifically, for a system with known density profile, as follows from solution of the Euler-Lagrange equation (17), the first-order hyper-Ornstein-Zernike relation (19) is an integral equation for $\chi_\ell(\mathbf{r})$, in which the specific form of $c_{1,\ell}^A(\mathbf{r}, [\rho])$, evaluated at the equilibrium density profile $\rho(\mathbf{r})$, constitutes a fixed inhomogeneity. Having all $\chi_\ell(\mathbf{r})$ obtained in this way, the second-order hyper-Ornstein-Zernike relation (21) then forms an integral equation for the second-order hyperfluctuation profiles $\chi_{\ell\ell'}(\mathbf{r})$. These appear both locally [first term on the right-hand side of Eq. (21)] and inside of the integration over $c_2(\mathbf{r}, \mathbf{r}', [\rho])$ [last term on the right-hand side of Eq. (21)]. All other terms are fixed and constitute the inhomogeneity that ultimately generates the resulting equilibrium shape of $\chi_{\ell\ell'}(\mathbf{r})$. The prediction workflow from neural hyperdensity functionals to averages, covariances, and hyperfluctuation profiles is illustrated in Fig. 2.

B. Case study: Fluctuations of particle number and cluster size in confined hard rod fluids

For a practical application of our framework, we consider the one-dimensional hard rod fluid and wish to investigate its cluster statistics. We recall that the Percus functional [69] permits exact analytic treatment in standard density functional theory and that neural functionals provide highly accurate surrogates [28, 33] in numerical work. As shown in Refs. [45, 46], univariate (first-order) hyperdensity functional theory gives access to a single complex observable of choice in the form of a tractable neural hyperdensity functional.

We here consider the following multivariate case. As the first hyperobservable, we choose the total number of particles within the system, $\hat{A}_1 = N$, which we recall is a fluctuating quantity in the grand ensemble. For the second hyperobservable $\hat{A}_2$, the size of the largest cluster is chosen, which is an order parameter commonly used to investigate gel-forming systems [70–72]. Two particles i and j are thereby defined as bonded if their distance is within a predefined cutoff, $|x_i - x_j| < x_c$, where x_i and x_j are the positions of the particles. The cutoff distance $x_c = 1.2\sigma$ is chosen in the following, where σ is the particle diameter. A cluster comprises of particles that are either directly or transitively bonded, and $\hat{A}_2$ is the number of particles in the largest cluster for a given microstate $x^N = x_1, \dots, x_N$. The calculation of $\hat{A}_2$, which is hence indeed a (phase space) function of the configuration x^N , is performed with a clustering algorithm that remains applicable for arbitrary spatial dimensionality; code is provided in Ref. [73]. We are ultimately interested in making the averages $A_1[\rho]$ and $A_2[\rho]$, variances $A_{11}[\rho]$

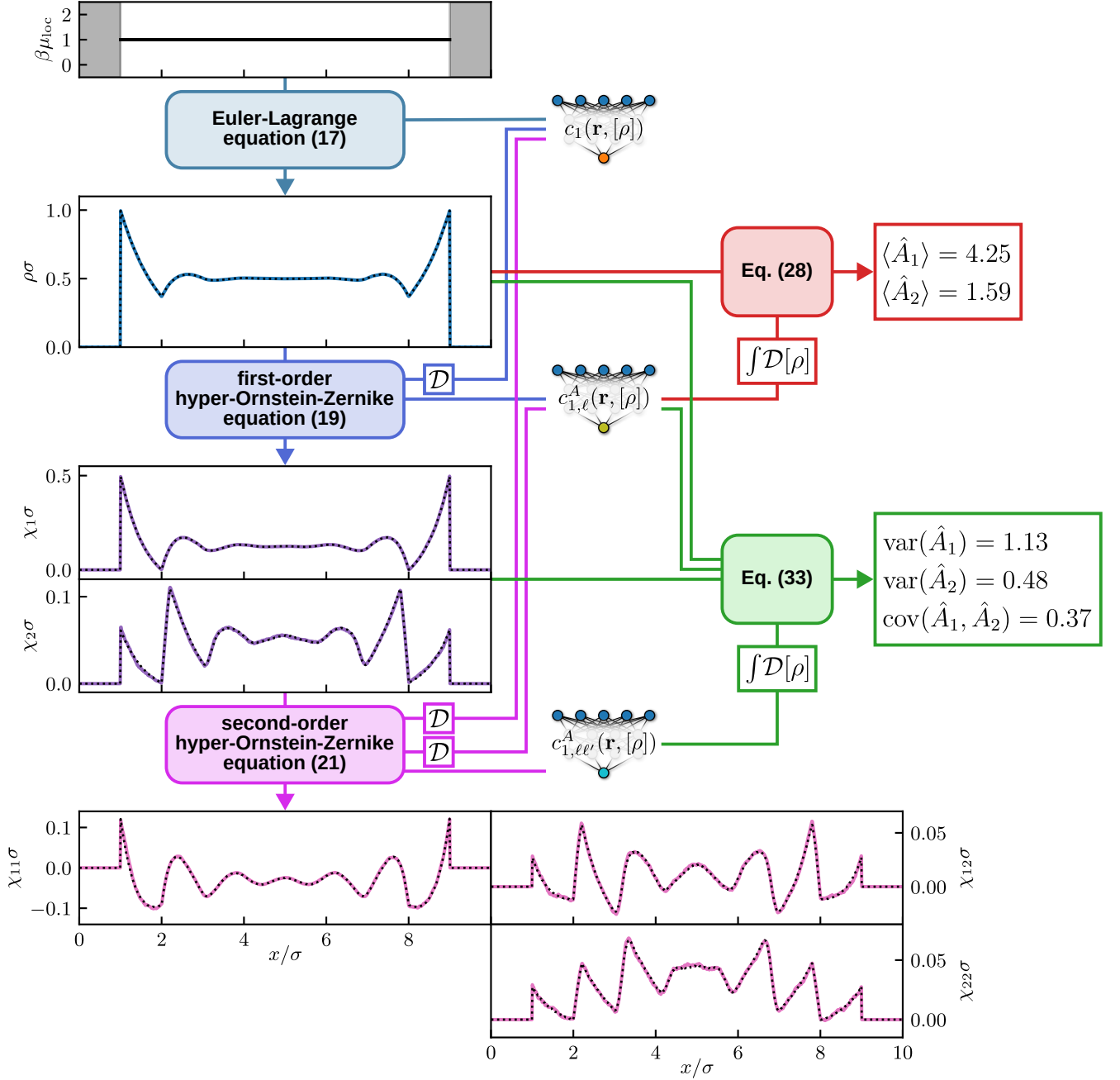


FIG. 2. Prediction workflow for evaluating averages and covariances of chosen hyperobservables using neural hyperdensity functionals. As a prototypical case, we depict the one-dimensional ($\mathbf{r} = x$) hard rod fluid under hard-wall confinement with walls at positions $x/\sigma = 1$ and $x/\sigma = 9$. The chosen hyperobservables are the total number of particles, $\hat{A}_1$, and the number of particles in the largest cluster, $\hat{A}_2$, see text. Via standard solution of the Euler-Lagrange equation (17) with the neural functional $c_1(\mathbf{r}, [\rho])$, the density profile $\rho(x)$ corresponding to the imposed (scaled) local chemical potential $\beta\mu_{loc}(x)$ is obtained. The associated first-order hyperfluctuation profiles $\chi_\ell(x)$, where $\ell = 1, 2$ for the depicted case, follow from solution of the first-order hyper-Ornstein-Zernike equation (19), where the neural hyperdirect correlation function $c_{1,\ell}^A(\mathbf{r}, [\rho])$ is invoked and $c_2(\mathbf{r}, \mathbf{r}', [\rho])$ is obtained from automatic differentiation of the neural functional $c_1(\mathbf{r}, [\rho])$, as indicated by the symbol $\mathcal{D}$. Using functional line integration ($\int \mathcal{D}[\rho]$) of $c_{1,\ell}^A(\mathbf{r}, [\rho])$ via Eq. (28), the averages $A_\ell = \langle \hat{A}_\ell \rangle$ are evaluated. Eq. (34) yields the variances $A_{\ell\ell} = \text{var}(\hat{A}_\ell)$ and covariances $A_{\ell\ell'} = \text{cov}(\hat{A}_\ell, \hat{A}_{\ell'})$, where $\ell, \ell' = 1, 2$, via functional line integration of $c_{1,\ell\ell'}^A(\mathbf{r}, [\rho])$ together with the evaluation of the integrals in Eq. (34) incorporating $\chi_{\ell'}(\mathbf{r})$ and $c_{1,\ell}^A(\mathbf{r}, [\rho])$. Additionally, the second-order hyperfluctuation profiles $\chi_{\ell\ell'}(x)$, are obtained from solution of the second-order hyper-Ornstein-Zernike equation (21), where integral terms that do not involve $\chi_{\ell\ell'}(x)$ are evaluated using directional functional derivatives, see Eq. (43). The neural predictions for all one-body profiles match simulation results (black dotted lines), which serve as independent reference.

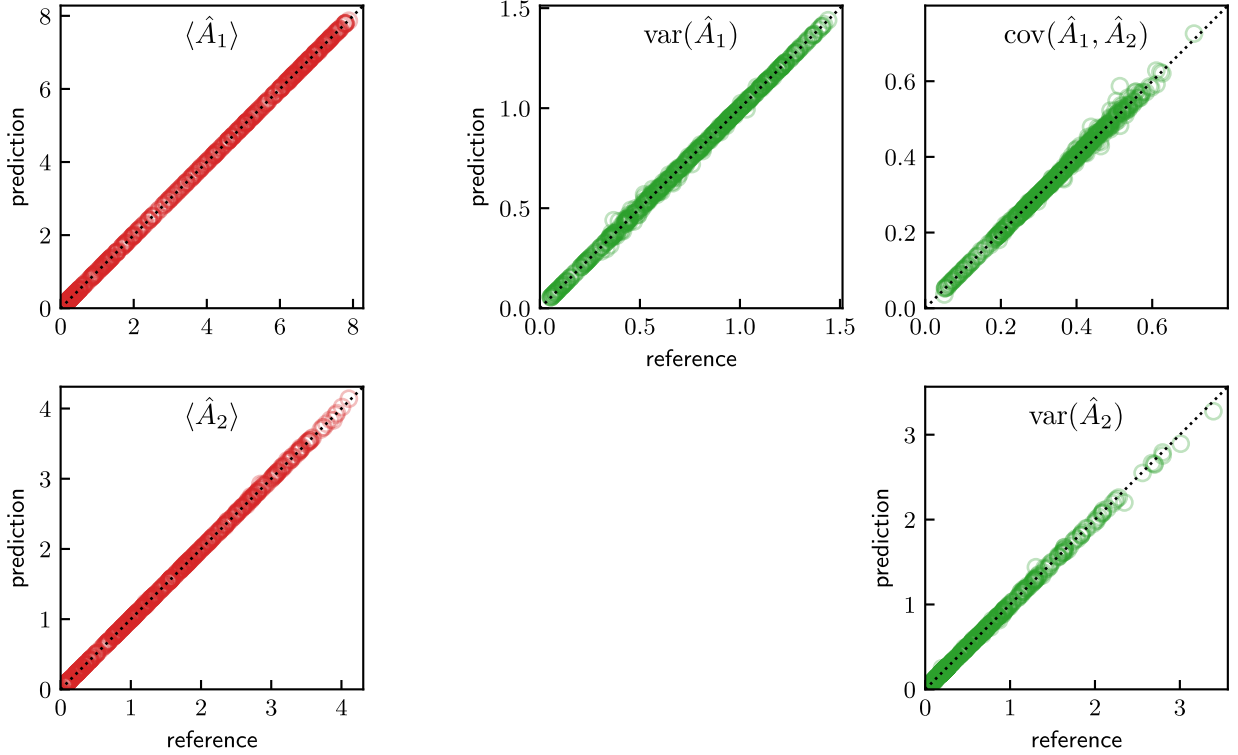


FIG. 3. Predictions for a test data set of 512 simulations that have not been included in the training; see Fig. 1 for an illustration of the underlying machine learning protocol and Fig. 2 for the prediction scheme. As before, the chosen hyperobservables are the total number of particles, $\hat{A}_1$, and the number of particles in the largest cluster, $\hat{A}_2$, see text. All relevant averages $\langle \hat{A}_1 \rangle$ and $\langle \hat{A}_2 \rangle$, variances $\text{var}(\hat{A}_1)$ and $\text{var}(\hat{A}_2)$, and the covariance $\text{cov}(\hat{A}_1, \hat{A}_2)$ are recovered accurately across the entire test data set via the neural multivariate hyperdensity functional framework.

and $A_{22}[\rho]$, as well as the covariance $A_{12}[\rho] = A_{21}[\rho]$ explicitly accessible via neural hyperdensity functionals.

In order to generate a suitable training data set, we employ grand canonical Monte Carlo simulations and sample all of the required one-body profiles $\rho(\mathbf{r})$, $\chi_1(\mathbf{r})$, $\chi_2(\mathbf{r})$, $\chi_{11}(\mathbf{r})$, $\chi_{22}(\mathbf{r})$, and $\chi_{12}(\mathbf{r}) = \chi_{21}(\mathbf{r})$ using standard histogram-counting methods. The first- and second-order hyperfluctuation profiles are respectively calculated via the covariance and trivariate expressions (6) and (7), for which all required averages are sampled. In total, 512 individual simulations are carried out at randomized scaled chemical potentials chosen uniformly in the range $-5.0 < \beta\mu < 5.0$. In each simulation, a randomly generated external potential is imposed, which consists of hard walls, linear segments, and Fourier modes. Typical external potential landscapes and corresponding results for a selection of pertinent one-body profiles are shown in Fig. 1. The machine learning of all neural (hyper)density functionals proceeds via the multistage protocol laid out in Sec. III A and illustrated in Fig. 1. At each stage, the corresponding (hyper)direct correlation function $c_1(\mathbf{r})$, $c_{1,\ell}^A(\mathbf{r})$, or $c_{1,\ell\ell'}^A(\mathbf{r})$ is calculated via post-processing of simulation data, thereby utilizing the availability of neural functionals of the previous stage. For each of the thus obtained correlation functions, a neural

network is trained to represent locally [27, 28, 45] the underlying functional dependence on the density profile.

The training of all neural functionals proceeds successfully, as is expected on the basis of the formal existence and uniqueness of the involved functional mappings. We find consistently and reproducibly that the mean squared error loss function that we employ is minimized effectively upon standard optimization based on the underlying data set. We take this as data-driven verification of the existence of all universal functionals that we seek to extract from the simulation results.

Employing the neural functionals for predicting $A_\ell[\rho]$ and $A_{\ell\ell'}[\rho]$ via Eqs. (28) and (34) yields values for the respective averages and covariances that closely match the simulation reference. Besides the results for the exemplary case of hard-wall confinement reported in Fig. 2, we test in Fig. 3 the neural predictions for a randomized data set of 512 simulations that have not been included in the training. The agreement of all considered averages and covariances is very satisfactory across the entire data set. Additionally, we demonstrate in Fig. 2 the accessibility of the complete set of first- and second-order hyperfluctuation profiles, which are again validated against reference simulation data. In conclusion, all quantities of interest can be predicted consistently with neural (hyper)density

functionals; we recall Appendix A for a numerical validation of equivalent routes for calculating covariances. In line with previous work on standard neural density functionals [27, 30], the range of applicability thereby far exceeds the underlying simulation data and allows for extrapolation to unseen scenarios, such as the exemplary case of pure hard-wall confinement shown in Fig. 2, which has not been featured in the randomized training data set.

The specific choice of hyperobservables permits further identifications and consistency checks involving the arising hyperfluctuation profiles. As $\hat{A}_1 = N$, the first-order hyperfluctuation profile $\chi_1(\mathbf{r})$ attains corresponding physical meaning as the (scaled) local compressibility $k_B T \chi_\mu(\mathbf{r})$; a well-known quantity used in previous work [8, 10, 12–14] to rationalize particle number fluctuations occurring in particular at substrates or solutes and close to phase transitions. Returning to the exemplary case of hard-wall confinement (see Fig. 2), $\chi_1(\mathbf{r})$ shows typical enhanced fluctuations in the vicinity of the walls and oscillations at increasing distance from the walls follow closely the shape of the density profile. The hard-wall contact value is $\chi_1(0^+) = \rho_b$ [14, 46], where ρ_b is the density of the bulk fluid far away from the wall. In the present system, the bulk density may be approximated by the plateau value of $\rho(\mathbf{r})$ in the center of the box, thus verifying the contact theorem numerically. Choosing the total number of particles N as a hyperobservable constitutes a special case, as the corresponding hyperdirect correlation functional reduces analytically to $c_1^A(\mathbf{r}) = 1$ [13, 14, 45, 46], which allows treatment already with standard (neural) density functional methods [30]; we elaborate on further consequences below.

The behavior of $\chi_2(\mathbf{r})$, which localizes fluctuations of the size of the largest cluster, is remarkably different to that of $\chi_1(\mathbf{r})$. The largest values of $\chi_2(\mathbf{r})$ are encountered away from the wall, at a distance approximately equal to the pairwise cluster cutoff criterion $x_c = 1.2\sigma$, indicating that (i) a particle being present at this location is favorable for the occurrence of large clusters in the system and (ii) microscopic details of the chosen hyperobservable are deeply imprinted in the corresponding hyperfluctuation profile. The hypercontact theorem $\chi_2(0^+) = A_2^b/V$ [46] connects the hard-wall contact value $\chi_2(0^+)$ to the mean size of the largest cluster per volume, A_2^b/V , in the bulk fluid.

The predicted second-order hyperfluctuation profiles $\chi_{11}(\mathbf{r})$, $\chi_{22}(\mathbf{r})$, and $\chi_{12}(\mathbf{r}) = \chi_{21}(\mathbf{r})$ generally show more pronounced and rapid oscillations, as is consistent with the increased correlation level that is probed. An interpretation of these quantities can be based on the trivariance definition (7). For given hyperobservables $\hat{A}_\ell$ and $\hat{A}_{\ell'}$, the second-order hyperfluctuation profile $\chi_{\ell\ell'}(\mathbf{r})$ measures how a particle being present at location $\mathbf{r}$ influences thermal fluctuations of the *combination* of observables $\hat{A}_\ell \hat{A}_{\ell'}$. The specific form of the trivariance, cf. Eq. (8), thereby ensures that effects of mere covariances of $\hat{A}_\ell$ and $\hat{A}_{\ell'}$ are excluded, as these are already captured

by the first-order hyperfluctuation profiles. A similar strategy for “distilling” higher-order fluctuation has been considered by Wilding and Turci [61], who argue that doing so is favorable for capturing near-critical surface phase behavior in their solute-solvent system. For our present scenario, one may glean from the results shown in Fig. 2 that particle number-number fluctuations decrease toward the middle and that they are negative, i.e. anticorrelated, in large parts of the system. Incorporating the largest cluster size changes this behavior, with $\chi_{22}(\mathbf{r})$ being positive everywhere and tending to increase toward the middle of the system. Sharp maxima arise for $\chi_{22}(\mathbf{r})$ at discrete multiples of the cluster cutoff distance x_c from the wall. This demonstrates the capability of the second-order hyperfluctuation profiles to resolve increasing levels of *microscopic* structuring while remaining firmly grounded on the one-body level.

Spatial integration of the density operator over the entire system domain yields $\int d\mathbf{r} \hat{\rho}(\mathbf{r}) = N$ by definition. Besides the trivial relation $\int d\mathbf{r} \rho(\mathbf{r}) = \langle N \rangle$, the integrated hyperfluctuation profiles may therefore be reduced to:

$$\int d\mathbf{r} \chi_\ell(\mathbf{r}) = \text{cov}(N, \hat{A}_\ell), \quad (44)$$

$$\int d\mathbf{r} \chi_{\ell\ell'}(\mathbf{r}) = \text{triv}(N, \hat{A}_\ell, \hat{A}_{\ell'}), \quad (45)$$

which follows from using their covariance and trivariance definitions (6) and (7). Hence, when choosing $\hat{A}_1 = N$, Eq. (44) implies the identity $\int d\mathbf{r} \chi_1(\mathbf{r}) = A_{11}$. Despite the apparent redundancy, the relation provides a valuable consistency check for the present conceptual investigation; indeed we find the sum rule to be verified numerically. Note also that Eq. (45) yields $\int d\mathbf{r} \chi_{11}(\mathbf{r}) = \text{triv}(N, N, N) = \langle N^3 \rangle - \langle N \rangle^3 - 3 \text{var}(N) \langle N \rangle$, which allows one to quantify third-order particle number fluctuations in the present context. We emphasize that the level of redundancy and the arising simplifications depend in general on the specific choice of hyperobservables, and that treating the total number of particles N is already implicit in the theory by integration of hyperfluctuation profiles according to Eqs. (44) and (45).

IV. CONCLUSIONS

Based on the recent hyperdensity functional theory for soft matter systems [45, 46], we have developed a first-principles-rooted approach to investigate the thermal equilibrium behavior of multiple variables that are simultaneously of interest. The multivariate hyperdensity functional theory allows one to systematically study correlations and causal relationships of several order parameters in inhomogeneous fluids, soft matter, and general classical many-body systems featuring nontrivial internal and external interactions. Our framework yields a concrete and practically feasible route toward expressing averages $A_\ell[\rho]$ and covariances $A_{\ell\ell'}[\rho]$ of a set of general

“hyperobservables” $\{\hat{A}_\ell\}$, $\ell = 1, \dots, L$, as universal density functionals. Crucially, these encapsulate only intrinsic properties emerging from the interparticle interactions and they hence remain applicable to any specific physical situation as determined by the external environment.

The theory rests formally on the Mermin-Evans classical density functional mapping [1, 2, 4], whereby the fundamental minimization principle is considered in an extended ensemble that is induced by the set of hyperobservables $\{\hat{A}_\ell\}$. Utilizing this extended ensemble determines uniquely first- and second-order hyperfluctuation profiles $\chi_\ell(\mathbf{r})$ and $\chi_{\ell\ell'}(\mathbf{r})$ that are associated with the respective hyperobservables, cf. their definitions (10) and (11) as parametric derivatives of the density profile as well as the equivalent formulation in terms of the covariance (6) and trivariance (7). The hyper-Ornstein-Zernike equations (19) and (21) incorporate $\chi_\ell(\mathbf{r})$ and $\chi_{\ell\ell'}(\mathbf{r})$ and they constitute defining integral equations for the first- and second-order hyperdirect correlation functionals $c_{1,\ell}^A(\mathbf{r}, [\rho])$ and $c_{1,\ell\ell'}^A(\mathbf{r}, [\rho])$, which both possess universal density functional dependence. Functional line integration of the hyperdirect correlation functionals, see Eqs. (28) and (34), respectively yields $A_\ell[\rho]$ and $A_{\ell\ell'}[\rho]$ in density functional form.

We have substantiated the multivariate hyperdensity functional framework by considering the accessibility of all required functional mappings from both the simulation and machine learning perspective. As is the case in standard classical density functional theory [27–30], all functional relationships are amenable to be extracted from a simulation data set of inhomogeneous fluid equilibria and to be represented with neural networks. In extension of the (univariate) first-order theory [45, 46], the structure of the pertinent equations naturally gives rise to a multistage protocol for machine learning neural (hyper)direct correlation functionals, as illustrated in Fig. 1. The hyper-Ornstein-Zernike equations (19) and (21) lie at the heart of this methodology and a reformulation to their equivalent Gâteaux forms (42) and (43), which feature (generalized) directional functional derivatives, turns out to be crucial for the numerical evaluation of all required expressions. We have applied this workflow to the investigation of clustering statistics in the hard rod fluid, whereby the two hyperobservables $\hat{A}_1$ and $\hat{A}_2$ of interest were chosen respectively as the total number of particles in the system and the number of particles in the largest cluster, defined via a transitive proximity criterion. Despite the simplicity of the system, the largest cluster size is an algorithmically defined many-body characteristic, making standard theoretical treatments arguably intractable in general. Using neural hyperdensity functionals, all pertinent averages $A_1[\rho]$ and $A_2[\rho]$, variances $A_{11}[\rho]$ and $A_{22}[\rho]$, as well as the covariance $A_{12}[\rho]$ were shown to be explicitly accessible to high precision. Additionally, all hyperfluctuation profiles $\chi_\ell(\mathbf{r})$ and $\chi_{\ell\ell'}(\mathbf{r})$, $\ell, \ell' = 1, 2$, could be determined accurately via solution of the hyper-Ornstein-Zernike equations (19) and (21), and these were shown to quantify

systematically the localized fluctuations associated with the chosen hyperobservables.

We conclude with several conceptual points.

(i) The hyperdensity functional framework can be extended naturally to multicomponent systems that feature distinct particle species, as is relevant when dealing, e.g., with fluid mixtures [32, 48]. The species-labelling of all pertinent quantities and their inter- and cross-coupling via species-resolved hyper-Ornstein-Zernike equations is laid out in Appendix C.

(ii) Irrespective of the number of particle species, the theoretical derivations can be pursued beyond the second-order level. Doing so gives rise to a hierarchy of higher-order hyper-Ornstein-Zernike equations that involve terms with increasing number of spatial integrals or, equivalently, with increasing derivative order. However, in clear distinction to other well-known hierarchical equations in statistical mechanics, such as the BBGKY hierarchy [4], the resulting set of hyper-Ornstein-Zernike equations does not necessitate a closure relation at the level of truncation in order to be solvable. Instead, all relevant expressions are explicitly accessible via neural functionals. This “inverted” hierarchical structure arises due to starting from the functional Mermin-Evans minimization principle at the one-body level and applying derivatives instead of integrating out degrees of freedom in the full many-body phase space picture.

(iii) In a pedestrian approach, one could be tempted to circumvent the present second-order theory by treating all arising averages via first-order hyperdensity functional theory. For instance, instead of determining the covariance $A_{\ell\ell'} = \text{cov}(\hat{A}_\ell, \hat{A}_{\ell'}) = \langle \hat{A}_\ell \hat{A}_{\ell'} \rangle - \langle \hat{A}_\ell \rangle \langle \hat{A}_{\ell'} \rangle$ via the second-order route laid out in this work, one may in principle consider the averages $\langle \hat{A}_\ell \hat{A}_{\ell'} \rangle$, $\langle \hat{A}_\ell \rangle$, and $\langle \hat{A}_{\ell'} \rangle$ individually via the first-order theory [45, 46] and build the covariance a posteriori. However, doing so has several severe drawbacks. Numerically, the route is susceptible to catastrophic cancellation errors, as the averages that enter the covariance are generally much larger than the covariance itself. Furthermore, the inherent theoretical structure that becomes apparent at the second-order level is entirely ignored. We hence argue that treating covariances with the present second-order route, thereby incorporating physical constraints via the second-order hyper-Ornstein-Zernike equation in integral form (21) or in directional functional derivative form (43) and via the second-order hyperdirect correlation functional (22), is both conceptually and practically beneficial as opposed to staying on the first-order level of the theory.

(iv) While we have kept our exemplary application in Sec. IIIB simple regarding the considered geometry and fluid type, the theory and workflow that we have developed are general and remain applicable to more complex systems. In Refs. [45, 46], the feasibility of machine learning univariate hyperdensity functionals has been demonstrated for the hard sphere fluid in planar three-dimensional geometry and we expect similar extensibility for the multivariate case. In this regard, we

emphasize that we deliberately employ (small) multi-layer perceptrons for extracting and representing all (hyper)direct correlation functionals. Despite their simple architecture, these neural networks already suffice in our applications as universal approximators of the required functional relationships. However, we see much potential in tuning the neural network architecture to the application under consideration, in particular when dealing with complex systems and geometries, as becomes pertinent, e.g., for increased spatial resolution or when including orientational degrees of freedom. Utilizing neural networks with built-in invariance or equivariance under specific symmetry operations may prove to be crucial for addressing these advanced applications, as exploited recently for constructing (standard) neural density functionals in two- and three-dimensional systems [65–67].

In future work, it may be interesting to consider a range of further theoretical developments and application scenarios. It might be valuable to investigate the feasibility of deriving *analytical* hyperdensity functionals. This is clearly a significant challenge, albeit a well-posed one, whereby the availability of highly accurate neural hyperdensity functionals might serve as a useful aid and benchmark for finding workable analytic approximations. Additional theoretical developments could encompass the inclusion of closure relations in the spirit of integral equation theories, which could serve as an alternative route to solving the hyper-Ornstein-Zernike equations. Turning to practical applications, the systematic study of inhomogeneous fluids near solvents and substrates hinges on a thorough description of surface-induced physical mechanisms that encompass enhanced local fluctuations, lyo- and hydrophobicity, as well as wetting and drying transitions, which constitute pertinent problems to be tackled via the present theory. Bui and Cox [74] have laid out a variant of the hyperdensity functional framework to investigate charged and dielectric systems under spatial confinement [75]. Their recent work concerns the generalization to azeotropic mixtures [76] and ab initio models of water [77]; it would be interesting to expand upon their findings via the inclusion of multiple species-resolved observables of interest. Applications to further (complex) three-dimensional fluids featuring bespoke order parameters are a natural and highly challenging goal. Pertinent examples include the investigation of colloidal equilibrium gels [70–72], where the largest cluster size as investigated in this work is one of several central order parameters for quantifying the onset of gelation. Further soft matter systems that necessitate the use of tailor-made order parameters for describing their emerging microstructure are two-dimensional systems and fluids with orientation-dependent interactions, where e.g. the hexatic order parameter arises as a key quantity. One could also envisage *designing* a set of suitable order parameters by analyzing multiple hyperobservables and their corresponding hyperfluctuation profiles in light of desired target properties. Utilizing multivariate hyperdensity functional theory to deal with such intricate matters consti-

tutes a highly valuable work program to be investigated in the future.

DATA AVAILABILITY

The data that support the findings of this article are openly available [73].

ACKNOWLEDGMENTS

We thank Silas Robitschko, Sophie Hermann, and Johanna Müller for useful discussions. This work is supported by the DFG (Deutsche Forschungsgemeinschaft) under Project No. 551294732.

Appendix A: Comparison of covariance calculations

Figure 4 shows a comparison of alternative implementations for calculating variances and covariances. We evaluate Eq. (34) in its original form and with exchanged indices ℓ and ℓ' in the second term on its right-hand side as well as Eq. (35). All numerical results agree with each other, thus confirming the equivalence of the above expressions.

Appendix B: Implementation of directional (Gâteaux) derivatives via automatic differentiation in Julia

Listing 1 provides an implementation in the programming language Julia for calculating first- and second-order directional (Gâteaux) derivatives generically and efficiently for a given functional $\mathcal{F}[\rho]$ using forward-mode automatic differentiation.

Appendix C: Multivariate hyperdensity functional theory for mixtures

For systems that are composed of several different components, much of the above theoretical structure persists and can be formulated straightforwardly, as we sketch in the following. We label the different species by $\alpha = 1, \dots, m$, where m is the total number of components. To facilitate book-keeping it is useful to introduce index sets $\mathcal{N}_\alpha$ that contain the labels i of all particles that belong to species α . The frequently occurring summation over all particles of a certain species α can then be written compactly as $\sum_{i \in \mathcal{N}_\alpha}$. An example that employs this convention is the partial density operator of species α , written as $\hat{\rho}_\alpha(\mathbf{r}) = \sum_{i \in \mathcal{N}_\alpha} \delta(\mathbf{r} - \mathbf{r}_i)$.

The standard mixture form of the Hamiltonian is $H = \sum_\alpha \sum_{i \in \mathcal{N}_\alpha} \mathbf{p}_i^2 / (2m_\alpha) + u(\mathbf{r}^N) + \sum_\alpha \sum_{i \in \mathcal{N}_\alpha} V_{\text{ext}}^{(\alpha)}(\mathbf{r}_i)$, where $u(\mathbf{r}^N)$ is the interparticle interaction potential

of the mixture, $V_{\text{ext}}^{(\alpha)}(\mathbf{r})$ is the external potential that acts on species α at position $\mathbf{r}$, and m_α is the mass of particles of species α ; the sum $\sum_\alpha$ runs over all species. The partition sum of the extended ensemble is $\Xi = \text{Tr} \exp(-\beta(H - \sum_\alpha \mu_\alpha N_\alpha) + \sum_\ell \lambda_\ell \hat{A}_\ell)$, where μ_α denotes the chemical potential and N_α the number of particles of species α . The corresponding extended grand potential is then $\Omega = -k_B T \ln \Xi$. Thermal averages in the extended ensemble are then built according to $\langle \cdot \rangle = \text{Tr} \cdot \exp(-\beta(H - \sum_\alpha \mu_\alpha N_\alpha) + \sum_\ell \lambda_\ell \hat{A}_\ell)$. An equivalent interpretation is that of an extended Hamiltonian $H_A = H - \sum_\ell \lambda_\ell \hat{A}_\ell / \beta$, which is coupled to particle baths, such

that its Boltzmann factor is $\exp(-\beta(H_A - \sum_\alpha \mu_\alpha N_\alpha))$. This leads to identical expressions for the partition sum and the grand potential as given above.

We follow the same structure as the one-component multivariate hyperdensity functional theory of Sec. II in its following generalization to mixtures. As in the one-component version, we consider a set $\{\hat{A}_\ell\}$, where the index $\ell = 1, \dots, L$ enumerates the different observables $\hat{A}_\ell$, which can, but need not, be linked to specific species α . Hence for generality, we keep the organization into species distinct from the choice of the observables. In any practical application, there can of course be physically motivated interrelations. As before, we aim at describing averages and covariances of the form (1) and (2), respectively.

In generalization of Eqs. (3)–(11), we obtain the following identities:

$$A_\ell = -\frac{\partial \beta \Omega}{\partial \lambda_\ell} \Big|_{\{V_{\text{ext}}\}}, \quad (\text{C1})$$

$$A_{\ell\ell'} = -\frac{\partial^2 \beta \Omega}{\partial \lambda_\ell \partial \lambda_{\ell'}} \Big|_{\{V_{\text{ext}}\}}, \quad (\text{C2})$$

where all species-wise external potentials are kept fixed in the parametric differentiation, as indicated by the shorthand notation $\{V_{\text{ext}}\} = \{V_{\text{ext}}^{(\alpha)}(\mathbf{r})\}$.

The partial (i.e. species-resolved) first- and second-order hyperfluctuation profiles have the following form:

$$\chi_\ell^{(\alpha)}(\mathbf{r}) = \text{cov}(\hat{\rho}_\alpha(\mathbf{r}), \hat{A}_\ell), \quad (\text{C3})$$

$$\chi_{\ell\ell'}^{(\alpha)}(\mathbf{r}) = \text{triv}(\hat{\rho}_\alpha(\mathbf{r}), \hat{A}_\ell, \hat{A}_{\ell'}). \quad (\text{C4})$$

We recall that the partial density profile $\rho_\alpha(\mathbf{r}) = \langle \hat{\rho}_\alpha(\mathbf{r}) \rangle$ of species α can be obtained as the following functional derivative: $\rho_\alpha(\mathbf{r}) = \delta \Omega / \delta V_{\text{ext}}^{(\alpha)}(\mathbf{r})$. Similarly, functionally differentiating the mean A_ℓ and the covariance $A_{\ell\ell'}$ by the (scaled) external potential for species α yields:

$$\chi_\ell^{(\alpha)}(\mathbf{r}) = -\frac{\delta A_\ell}{\delta \beta V_{\text{ext}}^{(\alpha)}(\mathbf{r})}, \quad (\text{C5})$$

$$\chi_{\ell\ell'}^{(\alpha)}(\mathbf{r}) = -\frac{\delta A_{\ell\ell'}}{\delta \beta V_{\text{ext}}^{(\alpha)}(\mathbf{r})}. \quad (\text{C6})$$

Equivalently, the species-labelled first- and second-order hyperfluctuation profiles can be obtained as the following parametric derivatives of the density profile:

$$\chi_\ell^{(\alpha)}(\mathbf{r}) = \frac{\partial \rho_\alpha(\mathbf{r})}{\partial \lambda_\ell} \Big|_{\{V_{\text{ext}}\}}, \quad (\text{C7})$$

$$\chi_{\ell\ell'}^{(\alpha)}(\mathbf{r}) = \frac{\partial^2 \rho_\alpha(\mathbf{r})}{\partial \lambda_\ell \partial \lambda_{\ell'}} \Big|_{\{V_{\text{ext}}\}}. \quad (\text{C8})$$

The Euler-Lagrange equations for the mixture are species-dependent and for species α given by:

$$c_1^{(\alpha)}(\mathbf{r}, [\{\rho\}]) = \ln \rho_\alpha(\mathbf{r}) + \beta V_{\text{ext}}^{(\alpha)}(\mathbf{r}) - \beta \mu_\alpha, \quad (\text{C9})$$

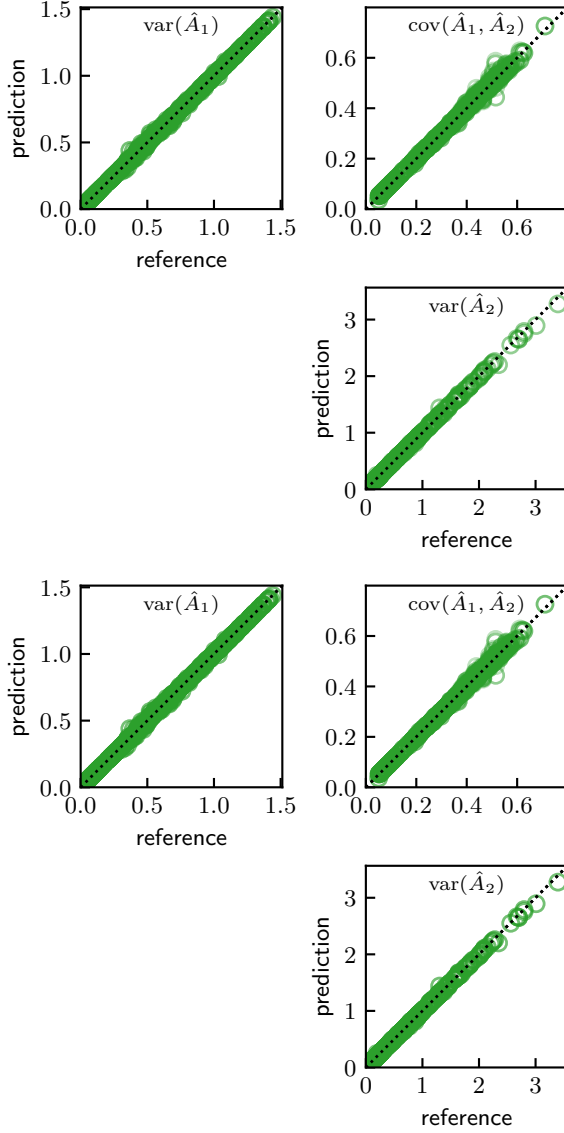


FIG. 4. Results for variances and covariances evaluated via Eq. (34), where ℓ and ℓ' in the second term on the right-hand side is exchanged (top), and via Eq. (35) (bottom). Both routes yield numerically equivalent results across the entire test data set, confirming the equivalence of both expressions and the exchange symmetry of the indices ℓ and ℓ' .

Listing 1. Implementation of first-order and second-order directional (Gâteaux) derivatives in the programming language Julia using the package ForwardDiff [78] for forward-mode automatic differentiation.

```
using ForwardDiff: derivative

# First-order Gateaux derivative of F at rho in direction psi
function Gateaux(F, rho, psi)
    derivative(epsilon -> F(rho + epsilon * psi), 0)
end

# Second-order Gateaux derivative of F at rho in directions psi1 and psi2
function Gateaux(F, rho, psi1, psi2)
    derivative(epsilon -> Gateaux(F, rho + epsilon * psi2, psi1), 0)
end
```

where $c_1^{(\alpha)}(\mathbf{r}, [\{\rho\}])$ is the one-body direct correlation functional for species α , which depends functionally on the set of all partial density profiles, for which the shorthand notation $\{\rho\} = \{\rho_\alpha(\mathbf{r})\}$ is introduced. The Euler-Lagrange equations (C9) constitute a set of m coupled equations, with one equation for each component α . Parametrically differentiating Eq. (C9) with respect to λ_ℓ yields the following first-order hyper-Ornstein-Zernike relation:

$$c_{1,\ell}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]) = \frac{\chi_\ell^{(\alpha)}(\mathbf{r})}{\rho_\alpha(\mathbf{r})} - \sum_{\alpha'} \int d\mathbf{r}' \chi_\ell^{(\alpha')}(\mathbf{r}') c_2^{(\alpha\alpha')}(\mathbf{r}, \mathbf{r}', [\{\rho\}]). \quad (\text{C10})$$

We have reordered the terms such that the right-hand side of Eq. (C10) contains all contributions that are accessible as averages in simulations, which encompasses the partial density profiles and $\chi_\ell^{(\alpha)}(\mathbf{r})$ via Eq. (C4). The species-resolved two-body direct correlation functionals may be obtained from analytical or neural functional representations.

The left hand side of Eq. (C10) is the (first-order) one-body hyperdirect correlation functional, as is specific to species α and observable ℓ , defined as

$$c_{1,\ell}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]) = \left. \frac{\partial c_1^{(\alpha)}(\mathbf{r}, [\{\rho\}])}{\partial \lambda_\ell} \right|_{\{\rho\}}. \quad (\text{C11})$$

The first-order two-body hyperdirect correlation functional is then defined as

$$c_{2,\ell}^{A,(\alpha\alpha')}(\mathbf{r}, \mathbf{r}', [\{\rho\}]) = \left. \frac{\partial c_2^{(\alpha\alpha')}(\mathbf{r}, \mathbf{r}', [\{\rho\}])}{\partial \lambda_\ell} \right|_{\{\rho\}} \quad (\text{C12})$$

$$= \frac{\delta c_{1,\ell}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}])}{\delta \rho_{\alpha'}(\mathbf{r}')}, \quad (\text{C13})$$

where the latter version is obtained from considering the relationship between one- and two-body direct correlation functionals for a mixture: $c_2^{(\alpha\alpha')}(\mathbf{r}, \mathbf{r}', [\{\rho\}]) = \delta c_1^{(\alpha)}(\mathbf{r}, [\{\rho\}]) / \delta \rho_{\alpha'}(\mathbf{r}')$.

Building the second parametric derivative of the species-resolved Euler-Lagrange equation (C9), or anal-

ogously differentiating the first-order hyper-Ornstein-Zernike relation (C10) with respect to $\lambda_{\ell'}$ gives the following second-order hyper-Ornstein-Zernike relation:

$$\begin{aligned} c_{1,\ell\ell'}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]) &= \frac{\chi_{\ell\ell'}^{(\alpha)}(\mathbf{r})}{\rho_\alpha(\mathbf{r})} - \frac{\chi_\ell^{(\alpha)}(\mathbf{r}) \chi_{\ell'}^{(\alpha)}(\mathbf{r})}{\rho_\alpha(\mathbf{r})^2} \\ &\quad - \sum_{\alpha'} \int d\mathbf{r}' \chi_{\ell'}^{(\alpha')}(\mathbf{r}') c_{2,\ell}^{A,(\alpha\alpha')}(\mathbf{r}, \mathbf{r}', [\{\rho\}]) \\ &\quad - \sum_{\alpha'} \int d\mathbf{r}' \chi_\ell^{(\alpha')}(\mathbf{r}') c_{2,\ell'}^{A,(\alpha\alpha')}(\mathbf{r}, \mathbf{r}', [\{\rho\}]) \\ &\quad - \sum_{\alpha'\alpha''} \int d\mathbf{r}' d\mathbf{r}'' \chi_\ell^{(\alpha')}(\mathbf{r}') \chi_{\ell'}^{(\alpha'')}(\mathbf{r}'') \\ &\quad \quad \times c_3^{(\alpha\alpha'\alpha'')}(\mathbf{r}, \mathbf{r}', \mathbf{r}'', [\{\rho\}]) \\ &\quad - \sum_{\alpha'} \int d\mathbf{r}' \chi_{\ell\ell'}^{(\alpha')}(\mathbf{r}') c_2^{(\alpha\alpha')}(\mathbf{r}, \mathbf{r}', [\{\rho\}]). \end{aligned} \quad (\text{C14})$$

Despite the increase in complexity over the first-order version (C10), here again the expressions on the right-hand side are available via sampling in simulations and numerical neural or approximate analytical functional methods. In analogy to Eq. (22), the second-order one-body hyperdirect correlation functional for species α is

$$c_{1,\ell\ell'}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]) = \left. \frac{\partial^2 c_1^{(\alpha)}(\mathbf{r}, [\{\rho\}])}{\partial \lambda_\ell \partial \lambda_{\ell'}} \right|_{\{\rho\}} \quad (\text{C15})$$

$$= \left. \frac{\partial c_{1,\ell}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}])}{\partial \lambda_{\ell'}} \right|_{\{\rho\}}. \quad (\text{C16})$$

where the form (C16) is obtained from inserting Eq. (C11) into Eq. (C15).

As in the one-component multivariate theory, we can functionally integrate in the space of partial densities to retrieve the mean of the ℓ th observable,

$$A_\ell[\{\rho\}] = \sum_\alpha \int d\mathbf{r} \rho_\alpha(\mathbf{r}) \int_0^1 da c_{1,\ell}^{A,(\alpha)}(\mathbf{r}, [\{a\rho\}]), \quad (\text{C17})$$

and the covariance of the ℓ th and ℓ' th observable,

$$A_{\ell\ell'}[\{\rho\}] = \sum_{\alpha} \int d\mathbf{r} \rho_{\alpha}(\mathbf{r}) \int_0^1 da c_{1,\ell\ell'}^{A,(\alpha)}(\mathbf{r}, [\{a\rho\}]) \\ + \sum_{\alpha} \int d\mathbf{r} \chi_{\ell'}^{(\alpha)}(\mathbf{r}) c_{1,\ell}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]), \quad (\text{C18})$$

with the second integral in Eq. (C18) being equivalent for exchanged labels ℓ and ℓ' . In the parameterized functional line integrals of both Eqs. (C17) and (C18), all partial density profiles are scaled simultaneously, such that $\{a\rho\} = \{a\rho_{\alpha}(\mathbf{r})\}$ with identical scaling factor $0 \leq a \leq 1$ for all species $\alpha = 1, \dots, m$. Other parameterizations of the functional line integrals are nevertheless conceivable.

For completeness, Eq. (C18) can be rewritten in the following form:

$$A_{\ell\ell'}[\{\rho\}] = \sum_{\alpha} \int d\mathbf{r} \rho_{\alpha}(\mathbf{r}) \int_0^1 da c_{1,\ell\ell'}^{A,(\alpha)}(\mathbf{r}, [\{a\rho\}]) \\ + \sum_{\alpha} \int d\mathbf{r} \frac{\chi_{\ell}^{(\alpha)}(\mathbf{r}) \chi_{\ell'}^{(\alpha)}(\mathbf{r})}{\rho_{\alpha}(\mathbf{r})} \\ - \sum_{\alpha\alpha'} \int d\mathbf{r} d\mathbf{r}' \chi_{\ell}^{(\alpha)}(\mathbf{r}) \chi_{\ell'}^{(\alpha')}(\mathbf{r}') c_2^{(\alpha\alpha')}(\mathbf{r}, \mathbf{r}', [\{\rho\}]), \quad (\text{C19})$$

analogously to Eq. (35).

The reformulation of integrals in the species-resolved hyper-Ornstein-Zernike equations (C10) and (C14) in terms of directional functional derivatives remains feasible by defining the species-aware Gâteaux differential

as

$$\mathcal{D}_{\{\psi\}} \mathcal{F}[\{\rho\}] = \sum_{\alpha} \int d\mathbf{r}' \psi_{\alpha}(\mathbf{r}') \frac{\delta \mathcal{F}[\{\rho\}]}{\delta \rho_{\alpha}(\mathbf{r}')} \\ = \left. \frac{d\mathcal{F}[\{\rho_{\alpha} + \epsilon \psi_{\alpha}\}]}{d\epsilon} \right|_{\epsilon=0} \quad (\text{C20})$$

for a generic functional $\mathcal{F}[\{\rho\}]$, which may now depend in general on the entire set of partial density profiles $\{\rho\} = \{\rho_{\alpha}(\mathbf{r})\}$. The directional functional derivative is then induced by the set $\{\psi\} = \{\psi_{\alpha}(\mathbf{r})\}$ of (generic) species-wise functions, as made explicit in the second line of Eq. (C20).

Using Eq. (C20), both multispecies hyper-Ornstein-Zernike equations (C10) and (C14) attain respectively their following compact Gâteaux forms:

$$c_{1,\ell}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]) = \frac{\chi_{\ell}^{(\alpha)}(\mathbf{r})}{\rho_{\alpha}(\mathbf{r})} - \mathcal{D}_{\{\chi_{\ell}\}} c_1^{(\alpha)}(\mathbf{r}, [\{\rho\}]), \quad (\text{C21})$$

$$c_{1,\ell\ell'}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]) = \frac{\chi_{\ell\ell'}^{(\alpha)}(\mathbf{r})}{\rho_{\alpha}(\mathbf{r})} - \frac{\chi_{\ell}^{(\alpha)}(\mathbf{r}) \chi_{\ell'}^{(\alpha)}(\mathbf{r})}{\rho_{\alpha}(\mathbf{r})^2} \\ - \mathcal{D}_{\{\chi_{\ell'}\}} c_{1,\ell}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]) \\ - \mathcal{D}_{\{\chi_{\ell}\}} c_{1,\ell'}^{A,(\alpha)}(\mathbf{r}, [\{\rho\}]) \\ - \mathcal{D}_{\{\chi_{\ell}\}} \mathcal{D}_{\{\chi_{\ell'}\}} c_1^{(\alpha)}(\mathbf{r}, [\{\rho\}]) \\ - \mathcal{D}_{\{\chi_{\ell\ell'}\}} c_1^{(\alpha)}(\mathbf{r}, [\{\rho\}]), \quad (\text{C22})$$

in formal analogy to the single-species versions (42) and (43).

-
- [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. Evans, Density functionals in the theory of non-uniform fluids, in *Fundamentals of Inhomogeneous Fluids*, edited by D. Henderson (Marcel Dekker, 1992) pp. 85–175.
 - [3] R. Evans, M. Oettel, R. Roth, and G. Kahl, New developments in classical density functional theory, *J. Phys.: Condens. Matter* **28**, 240401 (2016).
 - [4] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids: With Applications to Soft Matter*, 4th ed. (Elsevier/Academic Press, Amsterdam, 2013).
 - [5] M. Schmidt, Power functional theory for many-body dynamics, *Rev. Mod. Phys.* **94**, 015007 (2022).
 - [6] N. D. Mermin, Thermal properties of the inhomogeneous electron gas, *Phys. Rev.* **137**, A1441 (1965).
 - [7] R. Evans and N. B. Wilding, Quantifying density fluctuations in water at a hydrophobic surface: Evidence for critical drying, *Phys. Rev. Lett.* **115**, 016103 (2015).
 - [8] R. Evans and M. C. Stewart, The local compressibility of liquids near non-adsorbing substrates: A useful measure of solvophobicity and hydrophobicity?, *J. Phys.: Condens. Matter* **27**, 194111 (2015).
 - [9] B. Chacko, R. Evans, and A. J. Archer, Solvent fluctuations around solvophobic, solvophilic, and patchy nanostructures and the accompanying solvent mediated interactions, *J. Chem. Phys.* **146**, 124703 (2017).
 - [10] R. Evans, M. C. Stewart, and N. B. Wilding, A unified description of hydrophilic and superhydrophobic surfaces in terms of the wetting and drying transitions of liquids, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23901 (2019).
 - [11] M. K. Coe, R. Evans, and N. B. Wilding, Density depletion and enhanced fluctuations in water near hydrophobic solutes: Identifying the underlying physics, *Phys. Rev. Lett.* **128**, 045501 (2022).
 - [12] N. B. Wilding, R. Evans, and F. Turci, What is the best simulation approach for measuring local density fluctuations near solvo-/hydrophobes?, *J. Chem. Phys.* **160**, 164103 (2024).
 - [13] T. Eckert, N. C. X. Stuhlmüller, F. Sammüller, and M. Schmidt, Fluctuation profiles in inhomogeneous fluids, *Phys. Rev. Lett.* **125**, 268004 (2020).
 - [14] T. Eckert, N. C. X. Stuhlmüller, F. Sammüller, and

- M. Schmidt, Local measures of fluctuations in inhomogeneous liquids: Statistical mechanics and illustrative applications, *J. Phys.: Condens. Matter* **35**, 425102 (2023).
- [15] M. K. Coe, R. Evans, and N. B. Wilding, Understanding the physics of hydrophobic solvation, *J. Chem. Phys.* **158**, 034508 (2023).
- [16] T. Santos-Silva, P. I. C. Teixeira, C. Anquetil-Deck, and D. J. Cleaver, Neural-network approach to modeling liquid crystals in complex confinement, *Phys. Rev. E* **89**, 053316 (2014).
- [17] S.-C. Lin and M. Oettel, A classical density functional from machine learning and a convolutional neural network, *SciPost Phys.* **6**, 025 (2019).
- [18] S.-C. Lin, G. Martius, and M. Oettel, Analytical classical density functionals from an equation learning network, *J. Chem. Phys.* **152**, 021102 (2020).
- [19] 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).
- [20] C. Qiao, X. Yu, X. Song, T. Zhao, X. Xu, S. Zhao, and K. E. Gubbins, Enhancing gas solubility in nanopores: A combined study using classical density functional theory and machine learning, *Langmuir* **36**, 8527 (2020).
- [21] P. Yatsyshin, S. Kalliadasis, and A. B. Duncan, Physics-constrained Bayesian inference of state functions in classical density-functional theory, *J. Chem. Phys.* **156**, 074105 (2022).
- [22] A. Malpica-Morales, P. Yatsyshin, M. A. Durán-Olivencia, and S. Kalliadasis, Physics-informed Bayesian inference of external potentials in classical density-functional theory, *J. Chem. Phys.* **159**, 104109 (2023).
- [23] X. Fang, M. Gu, and J. Wu, Reliable emulation of complex functionals by active learning with error control, *J. Chem. Phys.* **157**, 214109 (2022).
- [24] 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, *Phys. Rev. Lett.* **134**, 056103 (2025).
- [25] M. M. Kelley, J. Quinton, K. Fazel, N. Karimitari, C. Sutton, and R. Sundararaman, Bridging electronic and classical density-functional theory using universal machine-learned functional approximations, *J. Chem. Phys.* **161**, 144101 (2024).
- [26] A. T. Bui and S. J. Cox, Learning classical density functionals for ionic fluids, *Phys. Rev. Lett.* **134**, 148001 (2025).
- [27] 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).
- [28] F. Sammüller, S. Hermann, and M. Schmidt, Why neural functionals suit statistical mechanics, *J. Phys.: Condens. Matter* **36**, 243002 (2024).
- [29] F. Sammüller and M. Schmidt, Neural density functionals: Local learning and pair-correlation matching, *Phys. Rev. E* **110**, L032601 (2024).
- [30] F. Sammüller, M. Schmidt, and R. Evans, Neural density functional theory of liquid-gas phase coexistence, *Phys. Rev. X* **15**, 011013 (2025).
- [31] S. M. Kampa, F. Sammüller, M. Schmidt, and R. Evans, Metadensity functional theory for classical fluids: Extracting the pair potential, *Phys. Rev. Lett.* **134**, 107301 (2025).
- [32] S. Robitschko, F. Sammüller, M. Schmidt, and R. Evans, Learning the bulk and interfacial physics of liquid-liquid phase separation with neural density functionals, *J. Chem. Phys.* **163**, 161101 (2025).
- [33] F. Sammüller and M. Schmidt, Determining the chemical potential via universal density functional learning, *Phys. Rev. Lett.* **136**, 068202 (2026).
- [34] R. Nagai, R. Akashi, S. Sasaki, and S. Tsuneyuki, Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability, *J. Chem. Phys.* **148**, 241737 (2018).
- [35] R. Nagai, R. Akashi, and O. Sugino, Completing density functional theory by machine learning hidden messages from molecules, *npj Comput. Mater.* **6**, 43 (2020).
- [36] J. Schmidt, C. L. Benavides-Riveros, and M. A. L. Marques, Machine learning the physical nonlocal exchange-correlation functional of density-functional theory, *J. Phys. Chem. Lett.* **10**, 6425 (2019).
- [37] Y. Zhou, J. Wu, S. Chen, and G. Chen, Toward the exact exchange-correlation potential: A three-dimensional convolutional neural network construct, *J. Phys. Chem. Lett.* **10**, 7264 (2019).
- [38] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation, *Nat. Comput. Sci.* **2**, 367 (2022).
- [39] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, Kohn-Sham equations as regularizer: Building prior knowledge into machine-learned physics, *Phys. Rev. Lett.* **126**, 036401 (2021).
- [40] J. Gedeon, J. Schmidt, M. J. P. Hodgson, J. Wetherell, C. L. Benavides-Riveros, and M. A. L. Marques, Machine learning the derivative discontinuity of density-functional theory, *Mach. Learn.: Sci. Technol.* **3**, 015011 (2022).
- [41] R. Pederson, B. Kalita, and K. Burke, Machine learning and density functional theory, *Nat. Rev. Phys.* **4**, 357 (2022).
- [42] D. de las Heras, T. Zimmermann, F. Sammüller, S. Hermann, and M. Schmidt, Perspective: How to overcome dynamical density functional theory, *J. Phys.: Condens. Matter* **35**, 271501 (2023).
- [43] T. Zimmermann, F. Sammüller, S. Hermann, M. Schmidt, and D. de las Heras, Neural force functional for non-equilibrium many-body colloidal systems, *Mach. Learn.: Sci. Technol.* **5**, 035062 (2024).
- [44] K. Ram, J. Dijkman, R. van Roij, J.-W. van de Meent, B. Ensing, M. Welling, and D. Cremers, Learned free-energy functionals from pair-correlation matching for dynamical density functional theory, *Phys. Rev. E* **112**, 045314 (2025).
- [45] F. Sammüller, S. Robitschko, S. Hermann, and M. Schmidt, Hyperdensity functional theory of soft matter, *Phys. Rev. Lett.* **133**, 098201 (2024).
- [46] F. Sammüller and M. Schmidt, Why hyperdensity functionals describe any equilibrium observable, *J. Phys.: Condens. Matter* **37**, 083001 (2025).
- [47] S. Robitschko, F. Sammüller, M. Schmidt, and S. Hermann, Hyperforce balance via thermal Noether invariance of any observable, *Commun Phys* **7**, 103 (2024).
- [48] J. Matthes, S. Robitschko, J. Müller, S. Hermann, F. Sammüller, and M. Schmidt, Gauge invariance and hyperforce correlation theory for equilibrium fluid mixtures, *J. Chem. Phys.* **164**, 044903 (2026).
- [49] J. Müller, S. Hermann, F. Sammüller, and M. Schmidt,

- Gauge invariance of equilibrium statistical mechanics, *Phys. Rev. Lett.* **133**, 217101 (2024).
- [50] J. Müller, F. Sammüller, and M. Schmidt, Why gauge invariance applies to statistical mechanics, *J. Phys. A: Math. Theor.* **58**, 125003 (2025).
 - [51] S. Hermann and M. Schmidt, Noether’s theorem in statistical mechanics, *Commun. Phys.* **4**, 176 (2021).
 - [52] S. Hermann and M. Schmidt, Why Noether’s theorem applies to statistical mechanics, *J. Phys.: Condens. Matter* **34**, 213001 (2022).
 - [53] S. Hermann and M. Schmidt, Variance of fluctuations from Noether invariance, *Commun. Phys.* **5**, 276 (2022).
 - [54] S. Hermann and M. Schmidt, Force balance in thermal quantum many-body systems from Noether’s theorem, *J. Phys. A: Math. Theor.* **55**, 464003 (2022).
 - [55] F. Sammüller, S. Hermann, D. de las Heras, and M. Schmidt, Noether-constrained correlations in equilibrium liquids, *Phys. Rev. Lett.* **130**, 268203 (2023).
 - [56] S. Hermann, F. Sammüller, and M. Schmidt, Noether invariance theory for the equilibrium force structure of soft matter, *J. Phys. A: Math. Theor.* **57**, 175001 (2024).
 - [57] S. M. Tschoop, F. Sammüller, S. Hermann, M. Schmidt, and J. M. Brader, Force density functional theory in- and out-of-equilibrium, *Phys. Rev. E* **106**, 014115 (2022).
 - [58] F. Sammüller, S. Hermann, and M. Schmidt, Comparative study of force-based classical density functional theory, *Phys. Rev. E* **107**, 034109 (2023).
 - [59] B. Rotenberg, Symmetry spotted in statistical mechanics, *Physics* **17**, 163 (2024).
 - [60] J. L. Miller, Gauge invariance applies to statistical mechanics too, *Phys. Today* **78**, 11 (2025).
 - [61] N. B. Wilding and F. Turci, Critical surface phase behavior governs hydrophobic attraction between extended solutes, *J. Chem. Phys.* **163**, 144901 (2025).
 - [62] N. B. Wilding, R. Evans, and F. Turci, An emergent interfacial length scale governs hydrophobic attraction across all molecular scales, not a crossover (2026), *chemRxiv*:15004612.
 - [63] R. Parr and Y. Weitao, *Density-Functional Theory of Atoms and Molecules*, International Series of Monographs on Chemistry (Oxford University Press, 1994).
 - [64] 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).
 - [65] F. Glitsch, J. Weimar, and M. Oettel, Neural density functional theory in higher dimensions with convolutional layers, *Phys. Rev. E* **111**, 055305 (2025).
 - [66] J. Weimar, M. Oettel, and A. Simon, Cubic-equivariant neural density functional theory for three-dimensional lattice fluids (2026), *arXiv*:2608.08137 [cond-mat.stat-mech].
 - [67] B. Cheng, Equivariant learning of a transferable three-dimensional classical density functional (2026), *arXiv*:2608.13506 [cond-mat.stat-mech].
 - [68] S. M. Kampa, M. Schmidt, and F. Sammüller, Spherical metadensity functional learning for inhomogeneous classical fluids (2026), to appear in *J. Chem. Theory Comput.*, *arXiv*:2606.14370 [cond-mat.soft].
 - [69] J. K. Percus, Equilibrium state of a classical fluid of hard rods in an external field, *J. Stat. Phys.* **15**, 505 (1976).
 - [70] S. Saw, N. L. Ellegaard, W. Kob, and S. Sastry, Computer simulation study of the phase behavior and structural relaxation in a gel-former modeled by three-body interactions, *J. Chem. Phys.* **134**, 164506 (2011).
 - [71] B. A. Lindquist, R. B. Jadrich, D. J. Milliron, and T. M. Truskett, On the formation of equilibrium gels via a macroscopic bond limitation, *J. Chem. Phys.* **145**, 074906 (2016).
 - [72] F. Sammüller, D. de las Heras, and M. Schmidt, Inhomogeneous steady shear dynamics of a three-body colloidal gel former, *J. Chem. Phys.* **158**, 054908 (2023).
 - [73] Code, data and models available at: <https://github.com/sfalmo/multi-hyper>.
 - [74] A. T. Bui and S. J. Cox, A first-principles approach to electromechanics in liquids, *J. Phys.: Condens. Matter* **37**, 285101 (2025).
 - [75] A. T. Bui and S. J. Cox, Dielectrocapillarity for exquisite control of fluids, *Nat. Commun.* **17**, 2661 (2026).
 - [76] K. L. Y. Zhou, A. T. Bui, and S. J. Cox, Roles of bulk and surface thermodynamics in the selective adsorption of a confined azeotropic mixture, *J. Phys. Chem. B* **130**, 4455–4466 (2026).
 - [77] A. T. Bui and S. J. Cox, A unified machine-learning framework for ab initio multiscale modeling of liquids, *Proc. Natl. Acad. Sci. U.S.A.* **123**, e2610049123 (2026).
 - [78] J. Revels, M. Lubin, and T. Papamarkou, Forward-mode automatic differentiation in Julia (2016), *arXiv*:1607.07892.