

RESEARCH ARTICLE | NOVEMBER 09 2023

Gaussian representation of coarse-grained interactions of liquids: Theory, parametrization, and transferability

Jaehyeok Jin ; Jisung Hwang ; Gregory A. Voth  



J. Chem. Phys. 159, 184105 (2023)

<https://doi.org/10.1063/5.0160567>





Lock-in Amplifier



Boxcar Averager



Zurich Instruments

Boost Your Optics and Photonics Measurements

[Find out more](#)

Gaussian representation of coarse-grained interactions of liquids: Theory, parametrization, and transferability

Cite as: J. Chem. Phys. 159, 184105 (2023); doi: 10.1063/5.0160567

Submitted: 2 June 2023 • Accepted: 6 October 2023 •

Published Online: 9 November 2023



Jaehyeok Jin,^{1,2}  Jisung Hwang,³  and Gregory A. Voth^{1,a)} 

AFFILIATIONS

¹ Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, 5735 S. Ellis Ave., Chicago, Illinois 60637, USA

² Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, USA

³ Department of Statistics, The University of Chicago, 5747 S. Ellis Ave., Chicago, Illinois 60637, USA

^{a)} Author to whom correspondence should be addressed: gavoth@uchicago.edu

ABSTRACT

Coarse-grained (CG) interactions determined via bottom-up methodologies can faithfully reproduce the structural correlations observed in fine-grained (atomistic resolution) systems, yet they can suffer from limited extensibility due to complex many-body correlations. As part of an ongoing effort to understand and improve the applicability of bottom-up CG models, we propose an alternative approach to address both accuracy and transferability. Our main idea draws from classical perturbation theory to partition the hard sphere repulsive term from effective CG interactions. We then introduce Gaussian basis functions corresponding to the system's characteristic length by linking these Gaussian sub-interactions to the local particle densities at each coordination shell. The remaining perturbative long-range interaction can be treated as a collective solvation interaction, which we show exhibits a Gaussian form derived from integral equation theories. By applying this numerical parametrization protocol to CG liquid systems, our microscopic theory elucidates the emergence of Gaussian interactions in common phenomenological CG models. To facilitate transferability for these reduced descriptions, we further infer equations of state to determine the sub-interaction parameter as a function of the system variables. The reduced models exhibit excellent transferability across the thermodynamic state points. Furthermore, we propose a new strategy to design the cross-interactions between distinct CG sites in liquid mixtures. This involves combining each Gaussian in the proper radial domain, yielding accurate CG potentials of mean force and structural correlations for multi-component systems. Overall, our findings establish a solid foundation for constructing transferable bottom-up CG models of liquids with enhanced extensibility.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0160567>

I. INTRODUCTION

With advances in computer architecture, coarse-grained (CG) models^{1–11} have gained in popularity since they can significantly extend the spatiotemporal scales of computer simulations, enabling one to explore the long length and time-scale behavior in systems that are difficult or infeasible for conventional atomistic simulations.^{12–18} With many other available CG approaches, bottom-up CG methodologies, in which CG models and interactions are determined based on fine-grained (FG) information, are often employed to reproduce the essential structural correlations at the atomistic level.^{8,10,19,20}

Despite its efficiency and ability to reproduce accurate descriptions of structural correlations, CG modeling faces several significant bottlenecks that limit its applicability to a wider range of systems. One such obstacle has been the transferability problem,^{7,10,11,21–23} where the bottom-up CG interactions cannot be determined as a single kind of interaction term because they are parametrized directly from the reference trajectories and vary over different conditions, e.g., temperature,^{24–26} pressure,^{27,28} phase,²⁹ and composition.³⁰ This issue can be further explained from a statistical mechanical point of view that the thermodynamically-consistent bottom-up CG interactions approximate the many-body CG potentials of mean force (PMFs) in terms of CG variables.^{31–33} Since the many-body

CG PMF is a free energy quantity, not purely an energetic quantity as with the FG potential,^{25,26,34} the effective CG interactions are expected to vary with different thermodynamic conditions. Therefore, CG interactions obtained via a bottom-up approach at a given state point often differ from those at other state points. Furthermore, the many-body nature of the CG interactions limits the design principles for cross-interactions between other CG systems. Even though there has been increasing interest in and attempts to resolve the transferability problem, we note that there has only been limited success and applicability over a wide range of systems to date.^{26,35–50}

In this paper, we provide theoretical and practical approaches that can serve as an alternative framework to resolve the transferability issue. From CG interactions determined by bottom-up methodologies such as multiscale coarse-graining (MS-CG)^{31,33,51–53} and relative entropy minimization (REM),^{54–56} we develop a new parametrization strategy that allows Gaussian basis sets to represent the given CG interaction. In particular, we focus on CG liquids here to derive a stronger connection between our proposed theory and the classical theories of liquids. Namely, we build our theory and parametrization strategy based on the classical perturbation theory of liquids,^{57–62} in which the hard sphere-like repulsive interactions primarily determine the liquid structure. In contrast, the long-range interactions can be treated as smaller attractions, i.e., perturbations. Recent systematic studies have demonstrated that the hard sphere description of molecular CG models can describe the dynamic properties, e.g., self-diffusion coefficients, indicating the fidelity of classical perturbation theory in CG models.^{63–65} Then, we treat the effective perturbative sub-interactions as Gaussians inspired by the classical density functional theoretical approaches.^{66–69} Since pair interactions at a certain distance correspond to the specific local density in that region, it is conceivable that the effective Gaussian interaction can modulate the changes in the local density. In the mesoscale regime, this can be further formulated into the many-body dissipative particle dynamics^{70,71} or the Gaussian equivalent representation^{72–76} governed by classical field theory. A microscopic interpretation of this mesoscale physics was first suggested by Stillinger, who proposed that the soft nature of the pair interactions can be represented by a Gaussian core under the canonical partition function.⁷⁷ While this Gaussian core model introduces a physical link to the Helmholtz free energy functional of the system via the mean-field ansatz, a key assumption in Stillinger's model is that the system has soft interactions, which allow for modeling "soft" polymer systems.^{78,79} To date, numerous polymer simulations have been performed using the Gaussian core model.^{80–84} At the microscopic level, particularly notable contribution is the Gaussian model potential developed by Berne and Pechukas in the early 1970s to account for the orientational dependence of pair interactions.⁸⁵

Among these microscopic and mesoscopic regimes where CG modeling takes place, it has been shown that the Gaussian models or processes can effectively represent the reduced system.^{86–88} As discussed by Milano *et al.*,⁸⁶ a key assumption in these approaches is that the probability distribution of the coarsened (or renormalized) variables can be written as a sum of Gaussian distributions. In this light, the combination of hard-core repulsive and Gaussian-like interactions has been widely applied to liquids, e.g.,

water.⁸⁹ Chaimovich and Shell parametrized the CG interaction of water to assess its anomalous behaviors across a range of densities and temperatures.⁹⁰ In pursuit of capturing two characteristic length scales, the CG interaction was represented as a combination of Lennard-Jones (LJ) and Gaussian interaction forms. By employing the REM, Chaimovich and Shell observed linear changes in interactions with changes in temperature and density, thereby extending the range of temperatures and density conditions from the earlier study conducted by Johnson *et al.*⁹¹ Recently, Shen *et al.* demonstrated the utility of the Gaussian repulsion in conjunction with the REM framework,³⁰ which was then extended to polymer field theory.⁹² In more complex biomolecules, top-down CG approaches have adopted these relatively simple Gaussian interactions to emulate the complicated protein behaviors^{93,94} and protein-protein interactions.^{13,14,16} Despite the success of the aforementioned Gaussian-based models for CG modeling, it remains unclear why these Gaussian representations are needed and how to interpret them from the bottom-up perspective. Based on these theoretical foundations, in this paper, we present a physical link for understanding molecular interactions in terms of Gaussian basis sets.

Our primary focus is to determine an effective reduction of this many-body CG PMF using a combination of hard sphere repulsion and Gaussian perturbative terms in order to further reduce the representation of the CG interactions. In particular, the ultimate goal of this study would be to go beyond what has been reported in the literature by utilizing the Gaussian representation to solve the transferability problem. Once we can relate the intensity, mean, and variance of each Gaussian with the molecular densities at the local region, then the essential variables that determine the Gaussian basis can be written as functions of the given thermodynamic state points, commonly referred to as "equations of state." The physical principles for determining the equations of state for molecular CG interactions should be consistent with those at other spatiotemporal scales. At a microscopic level, a much simpler analogy is given by hard spheres, where the system properties can be expressed as a function of the packing density.^{95–98} At the larger mesoscopic level, the complex equations of state can be generalized to understand the many-body dissipative interactions in terms of the local free energy.^{70,71} Therefore, by interpreting the perturbative interactions, especially with Gaussian basis sets, as a function of system variables, we will examine if the present strategy can provide a new method for transferring bottom-up CG interactions to different state points and even across different composition state points. Thus, the end goal of this work will be to infer the equations of state underlying the sub-interactions of the many-body CG PMF as a function of the system variables. Finally, we will illustrate how the specific sub-interaction strength is system density-dependent and leads to equations of state that scale linearly with the interaction strength parameter.

The remainder of this paper is organized as follows. In Sec. II, we provide a detailed sketch of the underpinning theory. In Sec. III, we first apply the developed framework for liquid CG models (methanol and acetonitrile) to validate our theory. Then, later in the section, we extrapolate these simplified interactions to different thermodynamic state points and compositions while maintaining their simple Gaussian descriptions.

II. THEORY

A. Pair decomposition of the CG interaction

We consider a CG system consisting of N particles with a configuration denoted as $\mathbf{R}^N = \{\mathbf{R}_1, \dots, \mathbf{R}_N\}$. In this context, the effective many-body CG PMF, U , should be a function of $\mathbf{R}^N$, i.e., $U(\mathbf{R}^N)$. However, practically determining this high-dimensional $U(\mathbf{R}^N)$ for molecular systems is immensely challenging. As a result, it is often approximated in a pairwise additive manner¹⁰

$$U(\mathbf{R}^N) \approx \sum_{I > J} U_{CG}(R_{IJ}), \quad (1)$$

where the pairwise CG interaction U_{CG} solely depends on the pairwise distance R_{IJ} . Thus, $U_{CG}(R)$ here can be understood as the pair potential resulting from a pair-additive approximation to the many-body CG PMF, $U(\mathbf{R}^N)$, for CG models involving more than two sites. For the sake of convenience, we will refer to the CG pair potential $U_{CG}(R)$ as the (approximated) many-body CG PMF in CG variables, as is commonly used in the literature.^{7–11} While various atomistic interactions lead to diverse forms of CG interactions for molecular systems, classical perturbation theory suggests that the effective interactions acting on homogeneous systems (in this case liquids) can be decomposed into a hard sphere-like (or “hard-core”) reference interaction and long-range perturbative interactions.^{57–62} To note, pioneering work from Weeks, Chandler, and Andersen showed that the short-ranged hard sphere interaction primarily governs the structure of liquids, and the long-range interaction acts as a perturbation to the short-range interaction.^{60–62} Namely, for the given CG interactions, $U_{CG}(R)$, this suggests

$$U_{CG}(R) = U_{HS}(R) + U'(R), \quad (2)$$

where $U_{HS}(R)$ is the reference repulsion, and $U'(R)$ is the long-range perturbative interaction. For non-hard sphere systems, the reference hard-core repulsion is often approximated by the following inverse power law (IPL) form^{99–101}

$$U_{HS}(R) = \left(\frac{\alpha_{HS}}{R} \right)^{\beta_{HS}}, \quad (3)$$

where α_{HS} and β_{HS} determine the hardness of the hard-core region of the interaction. The choice of IPL functional in Eq. (3) is reasonable in many molecular CG systems⁵³ due to the strongly repulsive nature of the underlying atomistic nuclei when they overlap, which is not the case for smoothly decaying Gaussians.⁶¹ With the repulsive term in place, the remaining long-range interactions can be accurately described using Gaussian basis functions, since determining these functions can be convexly optimized. However, it is unclear how this decomposition can be understood in terms of local densities, an issue we will explore in the remainder of this subsection and Sec. II B.

The local density of particle I , ρ_I , at a pair distance R is defined as

$$\rho_I(R) = \frac{n_I(R)}{V_R} = \frac{1}{V_R} \sum_J \delta(R - R_{IJ}), \quad (4)$$

where the function δ represents a discrete indicator function, i.e., $\delta(R - R_{IJ})$ is 1 when $R = R_{IJ}$ and 0 otherwise. The normalization factor V_R in the denominator corresponds to the volume of

a sphere with a radius R . In practical applications, Eq. (4) can be implemented by introducing a smooth weight function $\omega(R)$ to approximate $\delta(R - R_{IJ})$, consistent with the weighted density approximation in density functional theory¹⁰² or many-body dissipative particle dynamics:⁷⁰ $\rho_I(R) \approx \sum_J \omega(R_{IJ})/V_R$. Several options for $\omega(R_{IJ})$ have been proposed in the CG modeling literature, such as the Lucy function,^{103–105} a sigmoid function with a hyperbolic tangent,^{45,47,106–108} or a sixth-order polynomial.^{46,48,109}

Equation (4) can alternatively be written as a structural average using the pair correlation function, or radial distribution function (RDF), $g(R)$:

$$\rho_I(R) = \frac{1}{V_R} 4\pi\rho \int_0^R r'^2 g(r') dr'. \quad (5)$$

Furthermore, we can relate the (reduced) pair correlation function in Eq. (5) to an effective Helmholtz free energy by the reversible work theorem¹¹⁰

$$g(R) = e^{-\beta U_2(R)}, \quad (6)$$

where $U_2(R)$ denotes the pair PMF (PPMF).¹¹¹

By definition, the PPMF $U_2(R)$ essentially reflects a two-particle free energy, where its force component (known as pair mean force¹⁹), $-\nabla U_2(R)$, determines the average force on two particles separated by R . However, if the target system is no longer two particles in vacuum, as in the case for condensed phase liquids as studied in this work, the pair mean force encompasses environment-mediated forces as well as direct contributions from pair interactions $-\nabla U_{CG}(R)$. These additional forces contribute as averaged correlated forces arising from the environment and constitute higher-order correlations due to the solvation effect.

A paradigmatic example that underscores the non-trivial contribution of higher-order environment effects is, as introduced in Ref. 19, the identity coarse-graining (no explicit coarse-graining) applied to a single-site LJ fluid. In this prime example, the non-trivial difference between $U_{LJ}(R)$ and $-k_B T \ln g(R)$ arises from solvation forces and can be extended to even more complex molecular systems.^{112,113}

A rigorous decomposition of the pair mean force can be achieved by way of the Yvon–Born–Green (YBG) integral equation.¹¹⁴ The YBG hierarchy describes the mean force acting on particle 1 when the distance between particles 1 ($\mathbf{R}_1$) and 2 ($\mathbf{R}_2$) is $|\mathbf{R}_1 - \mathbf{R}_2| = R$:

$$-\nabla_1 U_2(\mathbf{R}_1, \mathbf{R}_2) = -\nabla_1 U_{CG}(\mathbf{R}_1, \mathbf{R}_2) + \int d\mathbf{R}_3 (-\nabla_1 U_{CG}(\mathbf{R}_1, \mathbf{R}_3)) \rho_{3|2}(\mathbf{R}_3|\mathbf{R}_1, \mathbf{R}_2). \quad (7)$$

In Eq. (7), $\rho_{3|2}(\mathbf{R}_3|\mathbf{R}_1, \mathbf{R}_2)$ represents the conditional triplet distribution defined as

$$\rho_{3|2}(\mathbf{R}_3|\mathbf{R}_1, \mathbf{R}_2) := \rho \frac{g^{(3)}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3)}{g(\mathbf{R}_1, \mathbf{R}_2)}. \quad (8)$$

Equation (7) elucidates that the pair mean force $-\nabla_1 U_2(\mathbf{R}_1, \mathbf{R}_2)$ is composed of a direct part involving the effective CG force $-\nabla_1 U_{CG}(\mathbf{R}_1, \mathbf{R}_2)$ and an indirect contribution $\int d\mathbf{R}_3 (-\nabla_1 U_{CG}(\mathbf{R}_1, \mathbf{R}_3)) \rho_{3|2}(\mathbf{R}_3|\mathbf{R}_1, \mathbf{R}_2)$. When rearranged, this equation reveals that the effective CG interaction for CG

simulation should account for the pair mean force and the negative of the solvation contribution to avoid *overcounting* the indirect contribution. Since the YBG equation and the Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy provide a rigorous and transparent interpretation of force balance stemming from various correlations,^{115–119} several coarse-graining methods based on the generalized YBG equation can construct high fidelity CG model by accurately elucidating the role of many-body correlations in determining forces and potentials.^{120–124}

In this work, we propose an alternative approach that aims to decompose the effective CG PMF $U_{CG}(R)$ into the PPMF and indirect (many-body) contributions:

$$U_{CG}(R) = U_2(R) + U_{MB}(R), \quad (9)$$

where $U_{MB}(R)$ denotes the higher-order environment-mediated contributions to the CG PMF in order to correct the solvation effect for the direct contribution (i.e., PPMF) $U_2(R)$. The complex correlations between the indirect particles in the solvation shell can result in such a many-body solvation term.¹²⁵ Yet, this indirect solvation potential is expressed in pairwise basis sets, so hereafter we would like to denote it as effective collective interaction. Then, it appears that Eq. (9) may indicate that the effective CG pair interaction can be presented as a combination of the pairwise interaction and the effective (many-body) collective interaction, akin to the many-body expansion. However, we intend to draw a distinction between these two approaches. This is because the standard expression for the many-body expansion is formulated for the overall CG energy in $\mathbf{R}^N$ space:¹²⁶

$$U(\mathbf{R}^N) = \sum_{I>J} U^{(2)}(R_{IJ}) + \sum_{IJK} U^{(3)}(\theta_{IJK}, R_{IJ}, R_{IK}) + \sum_{IJKL} U^{(4)}(\theta_1, \theta_2, \psi) + \dots, \quad (10)$$

where $U^{(n)}(\dots)$ is the n -body non-bonded interactions. While Eq. (10) provides an improved description in terms of many-body correlations compared to Eq. (1), it necessitates the inclusion of higher-order interactions, e.g., Stillinger–Weber interaction, which are no longer pairwise interactions. Notably, the many-body projection theory, built upon Eq. (10), effectively addresses this limitation by implicitly projecting higher-order interactions to lower-order ones, as outlined in Refs. 127 and 128. For instance, in the case of three-body interactions, Refs. 127 and 128 demonstrated that the three-body interaction $U^{(3)}(\theta_{IJK}, R_{IJ}, R_{IK})$ can be projected onto pairwise basis sets by integrating over additional configurational variables θ_{IJK} and R_{IK} using a conditional probability at the fixed distance R_{IJ} , $p(\theta_{IJK}, R_{IK}|R_{IJ})$, resulting in the expression

$$U(\mathbf{R}^N) = \sum_{I>J} U^{(2)}(R_{IJ}) + \sum_{I>J} C_3 \int d\theta_{IJK} dR_{IK} \times p(\theta_{IJK}, R_{IK}|R_{IJ}) U^{(3)}(\theta_{IJK}, R_{IJ}, R_{IK}). \quad (11)$$

Hence, for interactions $U^{(n)}(\dots)$ defined based on the many-body expansion [Eq. (10)], $U^{(2)}$ represents the pairwise interaction derived from the CG system, and $U^{(n>2)}$ describes the n -body interactions. However, it is important to highlight that the decomposition

between U_2 and U_{MB} in Eq. (9) is rather notably different, as U_2 also incorporates the higher-order (solvation) effect. The central conclusion of this subsection is that the decomposition scheme presented in Eq. (9) differs from the many-body expansion or projection theory. Since these two decompositions are derived differently, a possible future direction of interest would be to further understand the correspondence between these distinct decomposition approaches.

According to the YBG equation [Eq. (7)], $U_{MB}(R)$ should be the negative counterpart of the solvation potential up to a constant, i.e.,

$$U_{MB}(R) = \int d\mathbf{R}_1 \int d\mathbf{R}_3 (\nabla_1 U_{CG}(\mathbf{R}_1, \mathbf{R}_3)) \rho_{3|2}(\mathbf{R}_3|\mathbf{R}_1, \mathbf{R}_2) + (\text{const.}) \quad (12)$$

Even though this functional form implies that $U_{MB}(R)$ entails higher-order correlations beyond pairwise correlation, it involves complex integration over two configuration variables. As a result, the subsequent subsection (Sec. II C) will introduce a novel approach to determine $U_{MB}(R)$ as an alternative method.

However, prior to determining the functional form of $U_{MB}(R)$, classical perturbation theory can still be applied to the decomposed interaction form that incorporates both direct and indirect parts. Employing this approach, Eq. (9), a perturbative treatment of the CG system leads to

$$U_{CG}(R) = U_{HS}(R) + U'_2(R) + U_{MB}(R), \quad (13)$$

where the perturbative term is read as $U'_{CG}(R) = U'_2(R) + U_{MB}(R)$. Here, we only focus on the PPMF that is linked via the reversible work theorem to assess Eq. (5). Since the hard-core repulsive interaction is not directly involved in the CG PMF that gives rise to the non-zero number density, Eq. (13) implies that the shape of perturbative interactions primarily determines the local density beyond the hard-core regime.

B. Gaussian representation of CG interaction

We now introduce the Gaussian basis functions $G_i(\epsilon_i, \mu_i, \sigma_i) = \epsilon_i N(\mu_i, \sigma_i^2)$ to decompose $U'_2(R)$ into separate Gaussian sub-interactions. Without loss of generality, we can express $U'_2(R)$ as

$$U'_2(R) = \sum_i G_i(\epsilon_i, \mu_i, \sigma_i) = \sum_i \epsilon_i N(\mu_i, \sigma_i^2), \quad (14)$$

where ϵ_i , μ_i , and σ_i correspond to the magnitude, mean, and variance of the Gaussian basis functions, respectively. While the choice of Gaussian parameters might initially appear arbitrary, examining structural correlations allows us to minimize the number of Gaussian sub-interactions. Specifically, for liquids, Gaussian sub-interactions can be chosen to be distributed across each coordination shell. In other words, first, let us denote the distance of the i th coordination shell from the origin as $\bar{R}_i$. This $\bar{R}_i$ value can be directly estimated from the pair distance at the i th minimum on the RDF curve.¹¹⁰ Consequently, $\rho_I(R)$ from Eq. (5) can be expressed as

$$\rho_I(R) = \frac{4\pi\rho}{V_R} \left\{ \sum_i \int_{\bar{R}_{i-1}}^{\bar{R}_i} r'^2 e^{-\beta \sum_i G_i(\epsilon_i, \mu_i, \sigma_i)} dr' \right\}, \quad (15)$$

where we define $\bar{R}_0$ as 0 for clarity. A detailed derivation of Eq. (15) is provided in Appendix A. After identifying a set of Gaussians $\{G_i\}_i$ that minimizes the overlap between the adjacent Gaussians and is mainly distributed within each coordination shell, i.e., $\bar{R}_{i-1} < \mu_i < \bar{R}_i$ and $\sigma_i \leq (\bar{R}_i - \bar{R}_{i-1})/4$, covering at least 95%, we can approximate Eq. (15) as

$$\begin{aligned} \rho_I(R) &\approx \frac{4\pi\rho}{V_R} \left\{ \sum_{i(\bar{R}_{i-1} < R < \bar{R}_i)} \int_{\bar{R}_{i-1}}^{\bar{R}_i} r'^2 e^{-\beta G_i(\epsilon_i, \mu_i, \sigma_i)} dr' \right\} \\ &= \sum_{i(\bar{R}_{i-1} < R < \bar{R}_i)} \rho_I^i(\bar{R}_i, R), \end{aligned} \quad (16)$$

where the summation indices, $i(\bar{R}_{i-1} < R < \bar{R}_i)$, indicate that the i th Gaussian within the range $\bar{R}_{i-1} < r' < \bar{R}_i$ effectively contributes to the local density at the pair distance r' . Equation (16) is the central result of this paper, showing that the separated Gaussian interactions result in separable local densities. This finding can be intuitively understood by considering the relationship between the interaction potential and the distribution of particles. In regions where the potential is more attractive, more particles tend to be present, while in regions dominated by repulsive interactions, the particle density is lower. Therefore, introducing Gaussian basis sets to CG interactions provides a tractable way to understand the complex CG PMF as a linear combination of density-based metrics. Furthermore, as the Gaussian basis functions are closely related to the overall system density, we seek to determine how these Gaussian parameters vary for different system conditions. From the long-standing efforts to construct simple equations of state ranging from van der Waals to the mesoscopic systems,¹²⁹ we aim to determine an effective equation that can provide a general rule for predicting the behavior of each sub-interaction for different conditions.

C. Higher-order interaction in CG PMFs

While Gaussian basis sets introduced in Sec. II B can capture the average pair interactions over all configurations (equivalent to the Boltzmann inversion of the RDF), it is important also to consider higher-order environment-mediated interaction to avoid double-counting the solvation effect. Stated differently, even though a relatively simple reduction of the CG PMF by extracting the PPMF from its RDF could capture the local densities originating from pairwise ordering, relying solely on U'_2 from Eq. (13) would likely overlook and misrepresent the higher-order contributions to the many-body CG PMF. Therefore, it is essential to consider the non-trivial contribution of the effective collective interaction $U_{MB}(R)$. As discussed earlier, the analytical formulation of $U_{MB}(R)$ is quite complicated for complex systems as it involves the three-body distribution function via the YBG equation, resulting in limited predictability.¹²⁴ Nevertheless, our primary focus in this paper is on molecular liquids, and therefore, we provide an alternative derivation of $U_{MB}(R)$ from the Ornstein–Zernike integral equation.¹³⁰

For isotropic liquids, the Ornstein–Zernike equation describes the total correlation function $h(R) = g(R) - 1$ by dividing $h(R)$ into

a direct part, $c(R)$, and an indirect part arising from the propagation of interactions through the surrounding environments:

$$h(R) = c(R) + \rho \int c(|R - R'|) h(R') dR'. \quad (17)$$

To close the hierarchy presented in Eq. (17), a closure relationship between $h(R)$, $c(R)$, and $U_{CG}(R)$ is needed. This is achieved by introducing the bridge function $B(R)$, which can be expressed as an infinite summation of diagrams that are free of nodal circles.⁹⁵ The general form of the closure relationship is:

$$h(R) + 1 = \exp(-\beta U_{CG}(R) + h(R) - c(R) + B(R)). \quad (18)$$

Rearranging Eq. (18) in terms of the pair interaction potential $\beta U(R)$ yields

$$\beta U_{CG}(R) = g(R) - \ln g(R) - 1 - c(R) + B(R). \quad (19)$$

However, directly employing Eqs. (17) and (18) is not practically feasible for molecular systems due to the absence of a closed-form solution for $B(R)$.¹³⁰ Therefore, the commonly adopted approximation for the bridge function is the hypernetted chain (HNC) closure, which sets $B(R) = 0$.¹³¹ The HNC approximation works effectively for liquids and condensed matter systems with short-range interactions.⁹⁵

Under the HNC approximation with $B(R) = 0$, we can express $U_{CG}(R)$ as

$$U_{CG}(R) = -k_B T \ln g(R) + k_B T (g(R) - 1 - c(R)). \quad (20)$$

By recognizing that the first term on the right-hand side is the PPMF, i.e., $U_2(R) = -k_B T \ln g(R)$, the decomposition of Eq. (9) leads to

$$U_{MB}(R) = k_B T (g(R) - 1 - c(R)) := k_B T \ln y(R), \quad (21)$$

where $y(R)$ is the cavity correlation function defined as $y(R) := g(R)/e^{-\beta U_{CG}(R)}$.¹³² The right-hand side of Eq. (21), $\ln y(R)$, is known as the thermal potential $\omega(R)$,¹³³ which encompasses all indirect interactions of the system, as $y(R)$ describes the correlation between two fictitious cavities within the fluid.¹³⁴ Notably, our derivation formally establishes that the thermal potential from the cavity correlation function is equivalent to the collective solvation interaction.

Precisely determining $y(R)$ for simple and analytical systems can be achieved through several theoretical methodologies. A systematic effort can be undertaken using Henderson's method¹³⁵

$$y(R_{12}) = \exp(\beta \mu^{\text{res}}) \exp \left(\left[-\beta \sum_{j=2}^{N+1} U(R_{1j}) \right] \right), \quad (22)$$

where the ensemble average is taken under the constant NVT ensemble with the residual chemical potential μ^{res} that gives the zero-distance value of $y(0)$. Equation (22) or a direct simulation method¹³⁶ can offer numerical ways to compute the cavity correlation function. Yet, it is challenging to calculate the logarithm of the cavity function for complex molecular liquids and to simplify $\ln y(R)$ into an analytical form due to its convoluted nature. Nevertheless, the computed shape of $\ln y(R)$, as reported in Refs. 136–138,

strongly suggests that we can approximate it as a Gaussian-like interaction with a mean of zero. Our hypothesis gains further support from hard sphere theory, where one could expect to simplify $\ln y(R)$ as a function of pair distance and packing density η . This idea is substantiated by Meeron and Siegert,¹³⁹ where they derived $y(R)$ from the pair correlation of spherical non-interacting cavities. Subsequently, Grundke and Henderson demonstrated that $\ln y(R)$ can be approximated as a cubic polynomial in terms of R .¹⁴⁰

In this section, we present an alternative derivation that further substantiates the Gaussian hypothesis. As $R \rightarrow \infty$, $g(R) \rightarrow 1$ and $U(R) \rightarrow 0$, hence $y(R) \rightarrow 1$. Therefore, we further divide $y(R)$ into $y_0(R) + 1$, allowing the following asymptotic behavior: $y_0(R) \rightarrow 0$ as $R \rightarrow \infty$. By expanding $U_{MB}(R)$ using $y_0(R)$, we obtain

$$U_{MB}(R) = k_B T \ln(y_0(R) + 1) = k_B T \left[y_0(R) - \frac{y_0(R)^2}{2} + \frac{y_0(R)^3}{3} - \dots \right]. \quad (23)$$

Also, since $y(R)$ generally demonstrates a monotonically decreasing nature with a maximum at $R = 0$, i.e., $y'_0(0) = 0$ and $y''_0(0) < 0$,^{136–138} expanding $y_0(R)$ up to the second order gives

$$y_0(R) \approx y_0(0) + \frac{y''_0(0)}{2} R^2. \quad (24)$$

Substituting $y_0(R)$ into Eq. (21) yields

$$U_{MB}(R) \approx k_B T \left[y_0(0) \left(1 - \frac{y_0(0)}{2} + \frac{y_0^2(0)}{3} \right) + \frac{y''_0(0)}{2} R^2 \right], \quad (25)$$

where the contribution from higher order terms $\mathcal{O}(R^4)$ should be negligible based on Ref. 140. Then, we notice that Eq. (25) can be approximated as the following Gaussian function centered at $R = 0$:

$$U_{MB}(R) \approx k_B T y_0(0) \left(1 - \frac{y_0(0)}{2} + \frac{y_0^2(0)}{3} \right) \times \exp \left[-\frac{|y''_0(0)|}{y_0(0)} \frac{3}{6 - 3y_0(0) + 2y_0^2(0)} R^2 \right]. \quad (26)$$

Given that $y''_0(0) < 0$, this exponential function accurately captures Gaussian curvature up to the second order. While theoretically it is possible to determine $y_0(R)$ from CG simulations and estimate these values in principle using Henderson's method,¹³⁵ doing so for numerous thermodynamic state points throughout this study would be computationally demanding. Thus, we simplify Eq. (26) by employing a single Gaussian basis centered at $R = 0$:

$$U_{MB}(R) = A_{MB} \exp \left(-\frac{R^2}{2\delta_{MB}^2} \right), \quad (27)$$

where the interaction strength A_{MB} can be related to $\ln y(0) = \beta\mu^{\text{res}}$, i.e., $A_{MB} \approx k_B T (e^{\beta\mu^{\text{res}}} - 1) (7/6 - e^{\beta\mu^{\text{res}}}/2 + e^{2\beta\mu^{\text{res}}}/3)$. Similarly, δ_{MB}^2 can be understood as $\delta_{MB}^2 \approx (e^{\beta\mu^{\text{res}}} - 1) / |y''_0(0)| \cdot (7/6 - e^{\beta\mu^{\text{res}}}/2 + e^{2\beta\mu^{\text{res}}}/3)$.

In order to gain a deeper microscopic understanding of the relationship between A_{MB} and the system condition, one can assess this quantity using the Percus–Yevick equation of state, where the

excess chemical potential is expressed as a function of packing fraction (η)¹⁴¹

$$e^{\beta\mu^{\text{res}}} \approx \frac{(1 + 2\eta)^2}{(1 - \eta)^4}. \quad (28)$$

In turn, Eq. (28) suggests that A_{MB} will monotonically increase when the packing fraction increases, signifying that A_{MB} is related to the soft (maximum) repulsion between particle pairs and its dependence on the packing fraction. Since this repulsion can be understood in terms of a collective behavior, we would like to introduce a mesoscopic perspective to Eq. (27). One viable approach to understanding this link is with the dissipative particle dynamics, where the conserved interactions $U_{MB}(R)$ between each mesoscopic site are represented as a smoothly decaying function.^{70,71,142–145} A natural interpretation of $U_{MB}(R)$ is the mean-field interaction between the pairs of CG blobs under the influence of other CG blobs nearby. The alternative interpretation of Eq. (27) could be the Yukawa interaction, where $U_{MB}(R)$ aims to capture the long-range electrostatic interactions between the FG particles embedded in the CG entity. Thus, the width δ_{MB}^2 must be chosen carefully based on the molecular nature to account for such effects.

However, it is still unclear if our decomposition scheme and functional representation are valid in terms of microscopic bottom-up methodologies. Hence, we first check if the proposed decomposition scheme is consistent with conventional bottom-up CG methodologies by explicitly comparing each contribution from the given CG system. As a proof-of-concept, the effective CG interaction of methanol liquid at ambient conditions is decomposed by extracting the PPMF contribution from its RDF $U_2(R) = -k_B T \ln g(R)$ and $U_{MB}(R) = U_{CG}(R) - U_2(R)$, see Fig. 1. Here, the CG interaction was obtained via the force-matching algorithm (see Sec. II D for details).

Figure 1 validates our assumption about both Eqs. (14) and (17). The complex interaction profile in the CG PMF can now be understood as a combination of the PPMF from the RDF and the environment-mediated solvation potential. In particular,

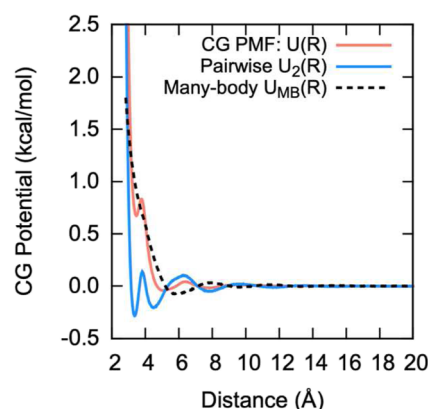


FIG. 1. Decomposition of the pair (blue line) and the multi-body solvation (black dashed) PMFs from the CG PMF (red line) of the CG methanol system. The PPMF contribution can be obtained by the reversible work theorem in Eq. (6), and the solvation term can be inferred from the mesoscopic description of liquids.

$U_{MB}(R)$, corresponding to the difference between $U_{CG}(R)$ and $U_2(R)$, exhibits a slowly decaying repulsion, which can be attributed to the nature of the solvation force. Due to the highly repulsive nature of the CG potential near methanol's first coordination shell, the repulsion from the neighboring methanol particles will drive the pair to draw closer, resulting in an attractive solvation force. However, $U_{MB}(R)$ needs to counteract the effective solvation effect, hence it must be repulsive and is consistent with the observation in Fig. 1. While there is a slight fluctuation after 7–8 Å, this contribution can be faithfully described by including Gaussians up to the second or third coordination shell according to Eq. (27).

Our understanding of the bottom-up CG PMF is consistent with the many-body dissipative particle dynamics framework. In the many-body dissipative particle dynamics, the pairwise interaction $U_2(R)$ becomes a local density-dependent interaction, which bridges the two different methodologies at distinct scales. We, therefore, propose an alternative perspective on the bottom-up CG PMF by combining the reversible work theorem from the microscopic regime with the many-body interactions from the mesoscopic treatment. However, parametrizing each sub-interaction shown in Fig. 1 is not trivial due to the ambiguously large contributions from the hard sphere repulsion at short ranges. Thus, we seek to parametrize the variables that determine the overall CG PMF via

$$U_{CG}(R) = U_{HS}(R) + U_2(R) + U_{MB}(R) \\ = \left(\frac{\alpha_{HS}}{R}\right)^{\beta_{HS}} + \sum_i \epsilon_i N(\mu_i, \sigma_i^2) + A_{MB} \exp\left(-\frac{R^2}{2\delta_{MB}^2}\right). \quad (29)$$

D. Parametrization strategy and multiscale coarse-graining

The proposed strategy for parametrizing the bottom-up CG interactions using Eq. (29) consists of three steps: (1) Determine the many-body CG PMF from the FG simulation (this is often approximated as pairwise CG potentials using force-matching or REM with spline basis sets), (2) parametrize the CG PMF from step (1) using the Gaussian representation, and (3) determine the equation of state for the parametrized sub-interaction variables. It is worth noting that we do not directly parametrize Eq. (29) using the FG trajectories because the systematic determination of only a handful of Gaussian parameters to the vast FG statistics may result in the overfitting of the high-dimensional CG PMF. Instead, having the B-spline functions as initial basis sets can amend this problem due to its expressiveness by matching the spline coefficients over different knots. Then, we seek to parametrize the Gaussian coefficients to the pairwise profile constructed by the B-splines. The advantages of using B-splines and related discussions on their usefulness for expressing CG interactions can be found in Refs. 53 and 146.

In practice, we first obtained the effective pairwise CG PMF using the MS-CG methodology.^{31,33,51–53} This methodology is based on the force-matching method that minimizes the force residuals $\chi^2[\mathbf{F}]$ between the CG particles I and the corresponding FG entities linked by the mapping operator $M_R^N: \mathbf{r}^n \rightarrow \mathbf{R}^N$ ($\mathbf{r}^n$ and $\mathbf{R}^N$ denote the FG and CG configurations, respectively). In this case, $\chi^2[\mathbf{F}]$ can be reformulated as

$$\chi^2[\mathbf{F}] = \frac{1}{3N} \left\langle \sum_{I=1}^N |\mathbf{F}_I(M_R^N(\mathbf{r}^n)) - \mathbf{f}_I(\mathbf{r}^n)|^2 \right\rangle. \quad (30)$$

The minimization target of Eq. (30) is the CG force $\mathbf{F}_I(M_R^N(\mathbf{r}^n))$, or $\mathbf{F}_I(\mathbf{R}^N)$, which is matched to the microscopic FG force component projected on the CG particle I , $\mathbf{f}_I(\mathbf{r}^n) := \sum_{i \in \mathbb{I}_I} \mathbf{f}_i(\mathbf{r}^n)$. Here, $\mathbb{I}_I$ is defined as a set of FG particles that are mapped to the CG particle I , and the CG force field $\mathbf{F}_I(M_R^N(\mathbf{r}^n))$ is expressed using the pairwise basis sets $\{\phi_2(R_{IJ})\}$ with unit vector $\hat{e}_{IJ}$, i.e.,

$$\mathbf{F}_I(M_R^N(\mathbf{r}^n)) = \sum_{J \neq I} \phi_2(R_{IJ}) \hat{e}_{IJ} = \sum_{J \neq I} \sum_k c_k u_k(R_{IJ}) \hat{e}_{IJ}. \quad (31)$$

The rightmost equality of Eq. (31) is satisfied by introducing the B-spline functions $\{u_k\}$ and their knot coefficient $\{c_k\}$, and the force-matching determines the $\{c_k\}$ via a least-square minimization. In our numerical parametrization, B-spline sets of the sixth order were chosen with a resolution of 0.2 Å by employing the publicly available version of the MSCGFM source code,¹⁴⁷ which is also accessible within the OpenMSCG program suite.¹⁴⁸ With the MS-CG PMFs in place, we can introduce the Gaussian basis sets to further reduce the CG interaction representation.

E. Simulation details

We have applied our theory to two different molecular liquids, methanol and acetonitrile, because they are anisotropic polar liquids with more than one characteristic length at the single-site CG resolution. Therefore, methanol and acetonitrile require more than one Gaussian basis function to fully recapitulate their interaction profiles, which provides solid validation of our methodology.

We first set up the FG trajectories to obtain the effective pairwise CG interactions for these liquids. Then, in order to examine how the Gaussian parameters change with temperature and bulk density, we used different temperatures and changed the bulk densities by adjusting the system box size from the equilibrium volume at 300 K and 1 atm. We opted to change the length of the simulation box as an indirect control of the CG pressure since the naïve estimate of the CG pressure using the virial expression is not identical to the FG counterpart due to the missing degrees of freedom, known as the “pressure representability” problem. For detailed analyses of this issue, we refer readers to Refs. 21, 29, and 149. Instead of calculating the CG pressure using a compatible observable expression,^{21,149} the reference CG pressure at each condition can be, in principle, deduced from the equilibrium volume of the system at the given temperature and pressure.

For methanol, we considered the following temperature range to account for temperature transferability: 300, 325, 350, 375, 400, 425, and 450 K. In addition, in order to span different volumes to account for pressure transferability, we varied the length of the cubic simulation box from 40.9 to 41.7 Å. These box lengths correspond to the overall densities of 0.7777, 0.7720, 0.7664, 0.7608, 0.7553, 0.7498, 0.7444, 0.7391, and 0.7338 g cm^{−3}, respectively. To note, at 1 atm and 300 K, the equilibrium box length was 41.163 Å for methanol.

Similarly, for acetonitrile, we used the temperatures of 250, 275, 300, 325, 350, and 375 K with box lengths ranging from 45.4 to

46.1 Å that correspond to the overall densities of 0.7285, 0.7237, 0.7189, 0.7142, 0.7095, 0.7049, 0.7003, and 0.6958 g cm⁻³. In this case, the equilibrium box length at 1 atm and 300 K was 45.798 Å. OPLS/AA force fields were selected as the atomistic force fields for simulating both liquid systems.^{150,151}

All FG simulations in this paper were performed with the following protocol using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) MD engine.^{152–154} First, the initial configurations at each cubic box were generated by the Packmol program,¹⁵⁵ and the FG topologies were created from the Visual Molecular Dynamics (VMD) suite.¹⁵⁶ Then, at the target temperature, the FG trajectories needed for the CG parametrization were produced using the constant NVT dynamics for 5 ns with Nosé–Hoover thermostat^{157,158} and coupling constant of 0.1 ps. Finally, CG simulations were performed from the last snapshot of the mapped FG trajectories using the obtained CG force fields under the constant NVT condition for 5 ns.

III. RESULTS

A. Methanol: Bottom-up CG interactions

The bottom-up CG interactions of methanol were first obtained by employing the MS-CG force-matching using Eqs. (30) and (31). Based on the settings for the FG simulation in Sec. II E, Fig. 2 illustrates the changes in the CG interactions over the different temperature and bulk density conditions, which are expected from the bottom-up nature. Since the many-body CG PMF under constant volume becomes the Helmholtz free energy, we expect the interactions to change linearly with the temperature (at the fixed density).^{25,26,34} However, as the temperature and volume change, the CG PMF is no longer reduced to the Helmholtz free energy due to the additional pressure term,²⁹ and the volume dependence on the pairwise CG interactions becomes relatively obscure. Nevertheless, it is apparent that the changes in temperature and volume result in non-negligible changes in the CG interactions.

Another important observation from the methanol PMFs is that mainly two sub-interactions vary with temperature and density. In particular, we observe one repulsive sub-interaction near 3.8 Å and one attractive sub-interaction at 4.8–5 Å. At other distances, we note that the hard-core repulsion at $R < 3$ Å and long-range potentials at $R > 7$ Å remain unchanged under variations in temperature and density. This observation further confirms our initial hypothesis that the Gaussian sub-interactions can effectively modulate changes in the local coordination shell. These changes provide an initial set of $\{\bar{R}_i\}_i$ and determine the number of basis functions needed for the reduced representation. We chose the $\{\bar{R}_i\}_i$ to be $\bar{R}_1 = 4.0$, $\bar{R}_2 = 6.5$, and $\bar{R}_3 = 8.5$ Å.

B. Methanol: Parametrization of Gaussian basis functions and equation of state

Based on the CG PMF analysis, we now build a Gaussian representation of the CG methanol model by parametrizing Eq. (29) to the CG PMFs shown in Fig. 2. As described in Sec. III A, we chose two Gaussian basis functions to parametrize the methanol interactions, with one Gaussian located from $\bar{R}_0$ to $\bar{R}_1$ and the other from $\bar{R}_1$ to $\bar{R}_2$. It is also possible to include a third Gaussian basis function to capture the long-range region near 8 Å. However, in order

to mitigate any potential numerical instability that could result from an additional Gaussian basis function and demonstrate the feasibility of this method, we chose to minimize the number of Gaussian basis functions used. To the best of our knowledge, this represents one of the earliest examples of this approach.

With two Gaussian basis functions, Eq. (29) is reduced to

$$U_{\text{MeOH}}(R) = a_{\text{MeOH}} \exp\left(-\frac{(R - b_{\text{MeOH}})^2}{c_{\text{MeOH}}^2}\right) + \left(\frac{d_{\text{MeOH}}}{R}\right)^{e_{\text{MeOH}}} - f_{\text{MeOH}} \exp\left(-\frac{(R - g_{\text{MeOH}})^2}{h_{\text{MeOH}}^2}\right) + k_{\text{MeOH}} \exp\left(-\frac{R^2}{l_{\text{MeOH}}^2}\right). \quad (32)$$

Rather than parametrizing all ten different parameters shown in Eq. (32), analyzing the PMFs from MS-CG can reduce the complexity of parametrizing these parameters. Interestingly, in Fig. 2, for all temperature and density conditions, we observe that the local maximum ascribed to the first attractive Gaussian is located at $R = 3.8$ Å, indicating $b_{\text{MeOH}} \approx 3.8$ Å. However, the widths and intensities of the first Gaussian basis vary with the different system conditions, and as such, values for a and c need to be determined. Additionally, at higher temperatures above 375 K, the effective potential values in the second Gaussian region, located near 4.8 Å, become less attractive as temperature increases and eventually become repulsive. The gradual change observed in our results indirectly indicates the contribution of collective many-body interactions near the second coordination shell. To account for the changes in the sign of the second Gaussian region, we interpret this as the mean-field exponential interaction not completely decaying at the second coordination shell. This is achieved by choosing a mean-field length of $l_{\text{MeOH}}^2 = 15$ Å². Finally, we determined the Gaussian parameters for the reduced interaction form (units omitted for clarity):

$$U_{\text{MeOH}}(R) = a_{\text{MeOH}} \exp\left(-\frac{(R - 3.8)^2}{c_{\text{MeOH}}^2}\right) + \left(\frac{d_{\text{MeOH}}}{R}\right)^{e_{\text{MeOH}}} - f_{\text{MeOH}} \exp\left(-\frac{(R - g_{\text{MeOH}})^2}{h_{\text{MeOH}}^2}\right) + k_{\text{MeOH}} \exp\left(-\frac{R^2}{15}\right). \quad (33)$$

To reiterate, our objective is twofold: to obtain parameters for Gaussian basis functions at varying temperatures and densities, as well as to establish the relationship between each parameter and state points to enable transferability. Since these sub-interactions are the renormalized representation of many-body quantities, there is no concise, analytical expression available for these parameters. Nevertheless, we aim to establish a simple correspondence between these variables, as they should correlate with local densities. This idea is akin to fitting the equations of state for thermodynamic properties in hard sphere liquids as a function of packing density.^{95–98} To note, previous studies of CG interactions under varying temperatures and densities have demonstrated a linear dependence. In this context, Izvekov *et al.* extended the MS-CG framework to describe density-dependent potentials,^{159,160} and Ref. 161 conducted a systematic study of how temperature and density variations impact pair

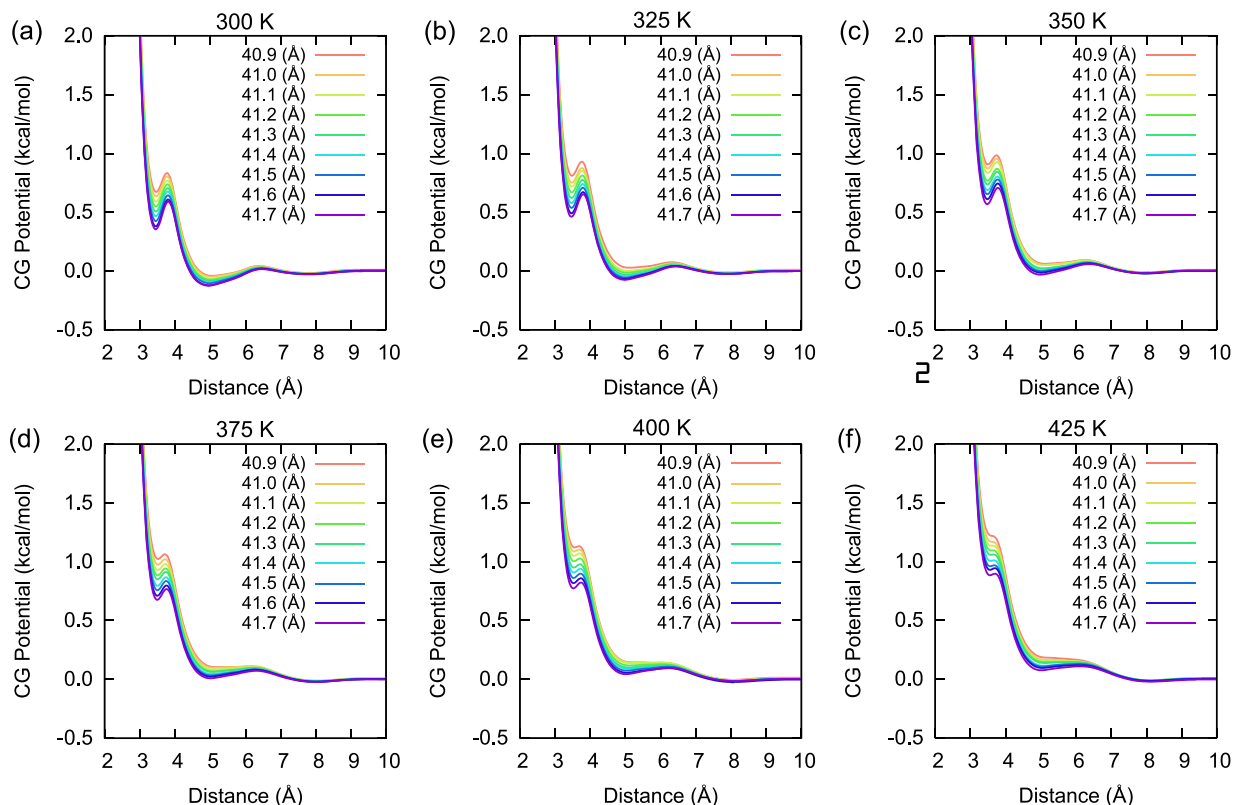


FIG. 2. Effective CG interactions of the methanol system for various density conditions at temperatures of (a) 300 K, (b) 325 K, (c) 350 K, (d) 375 K, (e) 400 K, and (f) 425 K. At each temperature, we considered the different pressure conditions determined by the system box size ranging from 40.9 Å (red line) to 41.7 Å (violet line).

potentials in CG liquids. However, these observations had certain limitations, making it challenging to extrapolate to different thermodynamic state points and phases. Therefore, we anticipate that deriving such equations of state through atomistic simulations will impart a systematic and predictable understanding of bottom-up CG interactions. An illustrative example of such an approach applied to molecular systems is presented in Ref. 162.

For stable numerical optimization, we enforced the following bounds inferred from the PMFs in Fig. 2 during optimization of the parameters in Eq. (33): $0 < a_{\text{MeOH}} < 2$ kcal/mol, $0 < c_{\text{MeOH}} < 0.25$ Å (since c should be smaller than the half of the well width), $3 < d_{\text{MeOH}} < 4$ Å, $5 < e_{\text{MeOH}}$ (to account for strongly repulsive nature), $0 < f_{\text{MeOH}} < 2$ kcal/mol, $4 < g_{\text{MeOH}} < 5.5$ Å (based on the second coordination shell), $0 < h_{\text{MeOH}} < 1$ Å, and $0 < k_{\text{MeOH}} < 2$ kcal/mol. All parametrization procedures were performed on MATLAB R2019b¹⁶³ using the curve fitting module with the trust-region-reflective least squares.¹⁶⁴

After the initial iteration of parametrization over all possible temperature-density conditions, we observed that the hard sphere diameter remained invariant under different densities at fixed temperatures. For example, at 300 K, the optimized d_{MeOH} values range from 3.167 to 3.170 Å with an average of 3.168 Å. This range is consistent with the conventional studies on the effective hard sphere diameter reported by Wilhelm that the hard sphere diameter is only

a function of the temperature.¹⁶⁵ Remarkably, we obtained a linear equation of state [Fig. 3(a)] by examining the average d_{MeOH} values over different temperatures, which allows us to parametrize $d_{\text{MeOH}}(\rho, T)$ as a function of temperature T as

$$d_{\text{MeOH}}(\rho, T) = 2.243 \times 10^{-4} \cdot T + 3.0790. \quad (34)$$

Also, under the conditions studied, the hard-core repulsive slope e_{MeOH} was found to marginally fluctuate around 15. Since the hard-core repulsive nature should depend on the molecule itself, it is reasonable to choose $e_{\text{MeOH}} = 15$ for the rest of the parametrization of the methanol system. Remarkably, this exponent value aligns with the exponent of the inverse power law potential derived from the LJ fluids using a different framework based on the virial-potential correlation.^{99,100,166–168} Reference 101 demonstrated that an exponent of 15 can faithfully reproduce the structure, dynamics, and heat capacity of the Kob–Andersen binary LJ liquid.¹⁶⁹ This congruence indirectly suggests that our approach can also effectively capture the inverse power law nature of molecular liquids. Subsequent subsections will explore how this value might vary for different molecules.

We then conducted the next iteration using the d values determined by Eq. (34). During each step, we prioritized optimizing the variables that contributed the most toward the overall CG

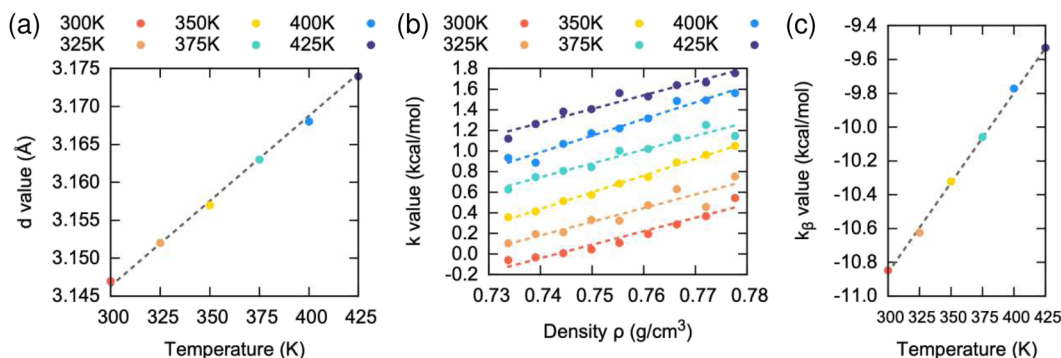


FIG. 3. Determination of simple equations of state describing the parametrized d and k variables for the Gaussian CG model. (a) Hard-core repulsive diameter at different temperatures from 300 K (red) to 425 K (navy). (b) Interaction strength of the collective solvation interaction term at different temperatures with its temperature-dependent y -intercept in panel (c).

interactions. This was to ensure that the residual contribution could be effectively captured in the subsequent iteration. In turn, in the second iteration, we investigated the relationship between the parametrized k_{MeOH} value and the system variables. As shown in Fig. 3(b), there exists a linear relationship between k_{MeOH} and system density, resulting in the equation of state for $k_{\text{MeOH}}(\rho, T)$ as

$$k_{\text{MeOH}}(\rho, T) = k_{\text{MeOH},\alpha} \cdot \rho + k_{\text{MeOH},\beta}(T), \quad (35)$$

which is expected from the effects of bulk volume on the macroscopic density. We report that this linear dependence shows a nearly identical slope, and the y -intercept ($k_{\text{MeOH},\beta}$) also reveals a linear relationship with temperature, as demonstrated in Fig. 3(c). As a result, k_{MeOH} values can be represented by a relatively simple equation of the form:

$$k_{\text{MeOH}}(\rho, T) = 14.574 \cdot \rho + k_{\text{MeOH},\beta}, \quad (36a)$$

$$k_{\text{MeOH},\beta}(T) = 1.0629 \times 10^{-2} \cdot T - 14.0489. \quad (36b)$$

While removing the uncertainty underlying the hard-core region, Fig. 4(a) delineates a consistent trend for the parametrized a_{MeOH} value, which becomes larger as temperature increases. The linear trend in the strength of the first repulsive Gaussian at the first coordination shell can be attributed to the tendency of molecules to repel each other more as the packing density increases. It should be noted that the changes with respect to density for a_{MeOH} are relatively minor across different temperatures. By using the averaged slope from this linear relationship, we further confirmed that the y -intercept, $a_{\text{MeOH},\beta}$, exhibits linear behavior as a function of temperature [Fig. 4(b)]:

$$a_{\text{MeOH}}(\rho, T) = 0.7593 \cdot \rho + a_{\text{MeOH},\beta}, \quad (37a)$$

$$a_{\text{MeOH},\beta}(T) = -1.5828 \times 10^{-3} \cdot T + 0.4923, \quad (37b)$$

which is consistent with our expectation that the molecules experience lower repulsion at higher temperatures. We also investigated the impact of temperature and density on the mean of the second

attractive Gaussian (g_{MeOH}) by analyzing its parametrized values in Figs. 4(c) and 4(d). Our results demonstrate that g_{MeOH} follows a similar linear equation of state as a_{MeOH} , being coupled to both density and temperature as

$$g_{\text{MeOH}}(\rho, T) = -2.2993 \cdot \rho + g_{\text{MeOH},\beta}, \quad (38a)$$

$$g_{\text{MeOH},\beta}(T) = -2.4694 \times 10^{-4} T + 7.5903. \quad (38b)$$

Note that we did not use the dataset at 300 K in the subsequent parametrization of the sub-interaction variables and the equations of state because of numerical instabilities during the parametrization procedure introduced by the overly negative second attractive well.

For the next step, using the a_{MeOH} , c_{MeOH} , d_{MeOH} , e_{MeOH} , g_{MeOH} , and k_{MeOH} values from the parametrized equations of state, we found the h_{MeOH} values to be less sensitive to temperature, as described by

$$h_{\text{MeOH}}(\rho) = 3.7797 \cdot \rho - 2.5381, \quad (39)$$

as shown in Fig. 5.

The last two steps of parametrization were conducted to determine the equations of state for f_{MeOH} and c_{MeOH} . First, we noticed that the f_{MeOH} values, derived from the intensity of the second attractive Gaussians, did not vary under different density conditions but demonstrated temperature dependence. Figure 6(b) clearly shows a trend of exponential increase in f_{MeOH} intensity with temperature, as evidenced by plotting the averaged f_{MeOH} values for each temperature. This observation is consistent with an Arrhenius-like behavior, which can be described as

$$f_{\text{MeOH}}(T) = 748.195 \cdot \exp(-7.877/17\beta) + 1.115 \times 10^{-1}, \quad (40)$$

where $\beta = (k_B T)^{-1}$, and the Boltzmann constant is k_B . In the subsequent step, we obtained the c_{MeOH} values, as shown in Figs. 6(c) and 6(d). Compared to other variables, c_{MeOH} demonstrates quite a different trend, and its physical picture is less clear than the other parameters. Still, the trend shown in Fig. 6(c) can be interpreted as a linear function with a slope that depends on temperature. As seen in Fig. 6(d), the parametrized relationship based on the Gaussian parametrization captures this trend via

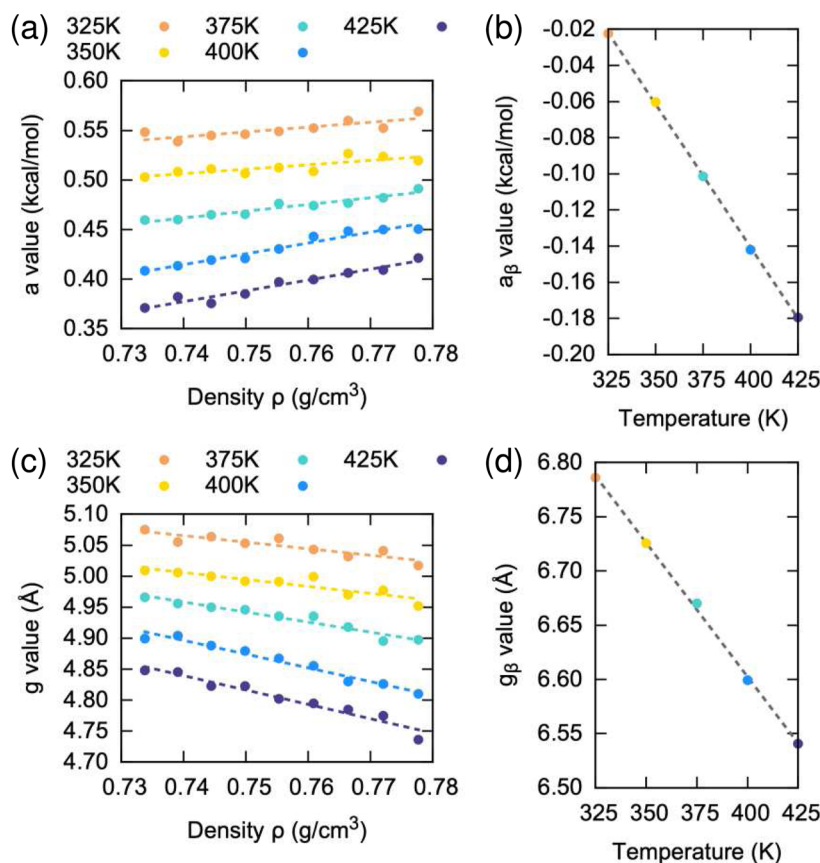


FIG. 4. Determination of the equations of state to parametrize the a and g variables for the Gaussian CG model. (a) and (b) Interaction strength of the short-range Gaussian repulsion term as a function of temperature and density, with temperatures ranging from 325 K (orange) to 425 K (navy). (c) and (d) Local minima for the long-range second Gaussian attraction as a function of temperature and density.

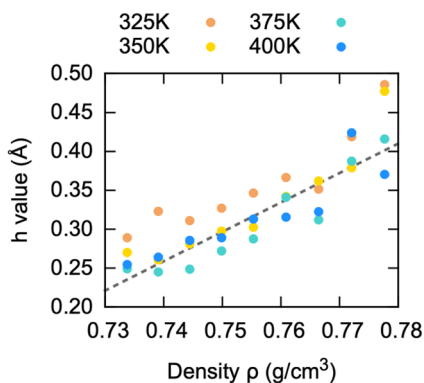


FIG. 5. Parametrized h variables (interaction width) for the attractive long-range Gaussian as a function of density.

$$c_{\text{MeOH}}(\rho, T) = (1.306 \times 10^{-2} T - 4.1739) \times (\rho - 0.758) + 0.35. \quad (41)$$

Finally, by combining Eqs. (34)–(41), the reduced Gaussian representation for bottom-up methanol CG models can be represented as

$$U_{\text{MeOH}}(R) = a_{\text{MeOH}}(\rho, T) \exp\left(-\frac{(R - 3.8)^2}{c_{\text{MeOH}}^2(\rho, T)}\right) + \left(\frac{d_{\text{MeOH}}(\rho, T)}{R}\right)^{15} - f_{\text{MeOH}}(T) \times \exp\left(-\frac{(R - g_{\text{MeOH}}(\rho, T))^2}{h_{\text{MeOH}}^2(\rho)}\right) + k_{\text{MeOH}}(\rho, T) \exp\left(-\frac{R^2}{15}\right). \quad (42)$$

C. Acetonitrile: Bottom-up CG interactions

A similar strategy as utilized for methanol can be implemented for acetonitrile. As detailed step-by-step instructions have been provided in Secs. III A and III B, in this section, we summarize

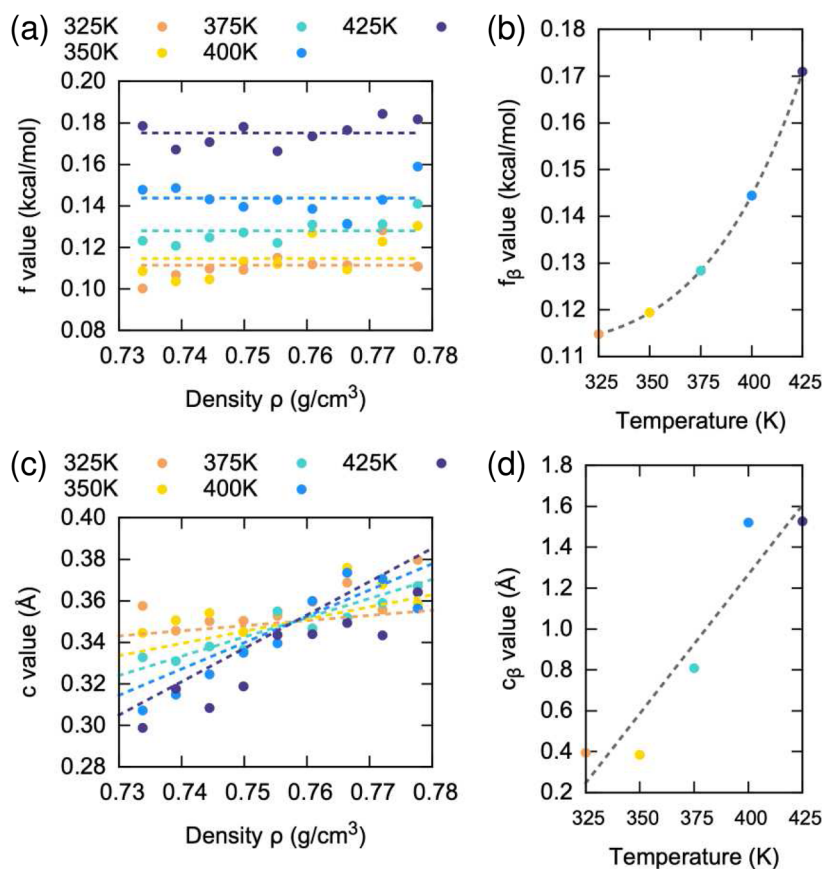


FIG. 6. Determination of the equations of state for the remaining variables (f , c) of the Gaussian CG model. (a) and (b) Interaction strengths for the long-range attractive Gaussian interaction as a function of temperature. (c) and (d) Interaction widths for the short-range repulsive Gaussian interaction as a function of temperature and density.

the essential findings obtained from the parametrization of the acetonitrile system and highlight its difference with respect to methanol. First, we identified the number of Gaussians and Gaussian intervals from the CG PMF obtained by multiscale force-matching, as depicted in Fig. 7.

Even though acetonitrile and methanol share a similar molecular backbone ($\text{CH}_3\text{-CX}$), their effective interactions are quite different. In CG acetonitrile, there is no repulsive well at short distances. However, a repulsive interaction is still present in the local region of 3.5–4 Å. The presence of a locally repulsive interaction near 3.5–4 Å is suggested by the decrease in slope observed after passing 3.8 Å from the hard-core repulsion at 3 Å. This change in slope is likely due to the attractive Gaussian at larger distances, which introduces Gaussian curvature and causes the initially steep slope from the hard-core repulsion to become relatively flat. Similar to methanol, we observed the signature of a negative Gaussian near the second coordination shell at 4.5–6 Å for acetonitrile. Since the changes in the CG potentials after the second coordination shell are quite negligible at different densities and temperatures, we also employ the two Gaussian basis functions discussed for the acetonitrile system.

D. Acetonitrile: Parametrization details for Gaussian basis functions

Using Eq. (29), we now parametrize the interaction parameters for acetonitrile and determine the equation of state for the simplified representation. From the distinct interaction profiles discussed in Sec. III C, we noticed that the inflection point near the first coordination shell is located near 3.8 Å, and hence $b_{\text{MeCN}} = 3.8$ Å. Moreover, the crossover of the second attractive Gaussian at the second coordination shell still exists at 4–6 Å, and we employed the same mean-field solvation length as $l_{\text{MeCN}}^2 = 15$ Å², resulting in the following interaction form:

$$U_{\text{MeCN}}(R) = a_{\text{MeCN}} \exp\left(-\frac{(R-3.8)^2}{c_{\text{MeCN}}^2}\right) + \left(\frac{d_{\text{MeCN}}}{R}\right)^{e_{\text{MeCN}}} - f_{\text{MeCN}} \exp\left(-\frac{(R-g_{\text{MeCN}})^2}{h_{\text{MeCN}}^2}\right) + k_{\text{MeCN}} \exp\left(-\frac{R^2}{15}\right). \quad (43)$$

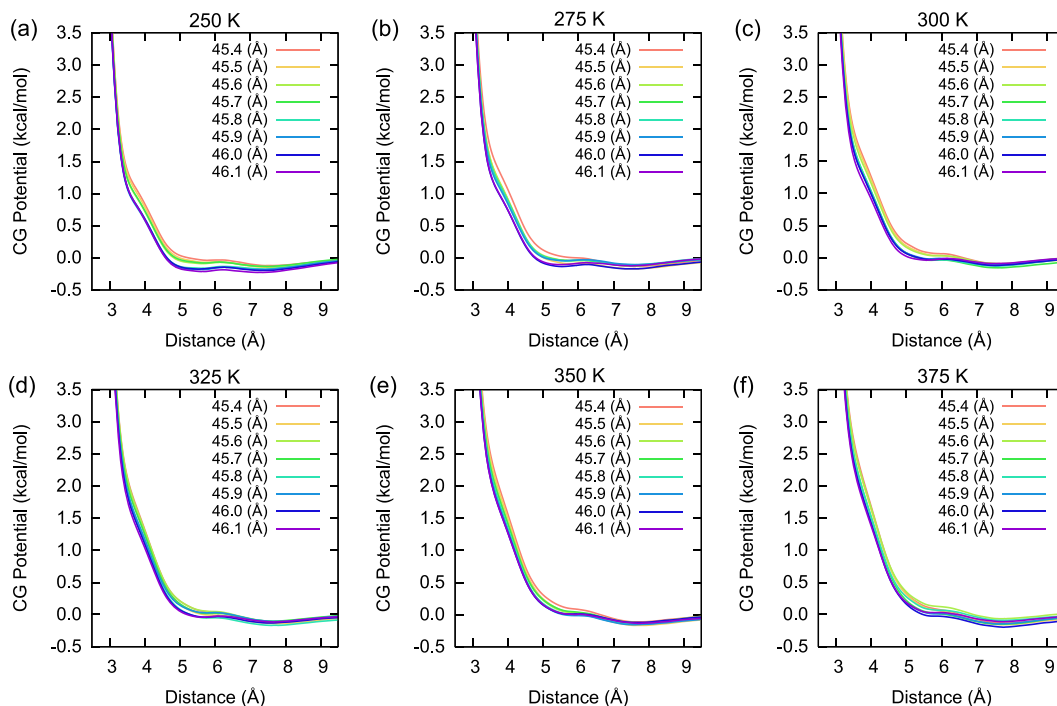


FIG. 7. Effective CG interactions of the acetonitrile system for various density conditions at temperatures of (a) 250 K, (b) 275 K, (c) 300 K, (d) 325 K, (e) 350 K, and (f) 375 K. At each temperature, we considered the different pressure conditions determined by the system box size ranging from 45.4 Å (red line) to 46.1 Å (violet line).

Following the initial parametrization using the fixed variables, we were able to capture the linear dependence of the hard sphere diameter for acetonitrile [Fig. 8(a)] as

$$d_{\text{MeCN}}(\rho, T) = 8.697 \times 10^{-4} \cdot T + 3.2611. \quad (44)$$

It should be mentioned that the hard-core repulsion coefficient e_{MeCN} was determined to be nearly 10, confirming our understanding that e_{MeCN} is a molecular-dependent coefficient, but it does not vary strictly with temperature or density for the same molecular system. Notably, this exponent value is smaller compared to that of methanol (15) or the LJ liquid (15).¹⁰¹ This lower e_{MeCN} value can be attributed to the single-site representation of the CG acetonitrile model, which significantly deviates from a spherical description due to the presence of a strong C≡N triple bond. This deviation results in a less repulsive exponent. As a potential avenue for further exploration, it would be valuable to establish a connection between the hard-core repulsive characteristics and molecular nature, e.g., the shape and topology of the molecule.

Next, we seek to determine the equation of state governing the strength of the collective solvation interactions, k_{MeCN} , as a function of temperature: $k_{\text{MeCN}}(\rho, T) = k_{\text{MeCN},\alpha} \cdot \rho + k_{\text{MeCN},\beta}(T)$. Figure 8(b) shows that the linearly varying behavior is exhibited in acetonitrile as

$$k_{\text{MeCN}}(\rho, T) = 28.245 \cdot \rho + k_{\text{MeCN},\beta}, \quad (45a)$$

$$k_{\text{MeCN},\beta} = 8.5692 \times 10^{-3} \cdot T - 22.56. \quad (45b)$$

Based on the parametrized d_{MeCN} and k_{MeCN} functionals, a_{MeCN} is described as

$$a_{\text{MeCN}}(\rho, T) = -4.078 \times 10^{-1} \cdot \rho + a_{\text{MeCN},\beta}, \quad (46a)$$

$$a_{\text{MeCN},\beta} = 3.6155 \times 10^{-3} \cdot T - 2.5537 \times 10^{-2}. \quad (46b)$$

Figure 9 illustrates the parametrization details for the remaining variables. In Fig. 9(a), g_{MeCN} still exhibits a linear dependence on density as

$$g_{\text{MeCN}}(\rho, T) = 9.4940 \cdot \rho + g_{\text{MeCN},\beta}, \quad (47a)$$

$$g_{\text{MeCN},\beta} = 1.9600 \times 10^{-3} T - 1.5830, \quad (47b)$$

but we noticed that f_{MeCN} , h_{MeCN} , and c_{MeCN} demonstrate temperature-independent behavior unlike g_{MeCN} . While we see some minor deviations from the linear trend, it is worth noting that the fluctuations of f_{MeCN} , h_{MeCN} , and c_{MeCN} are relatively smaller than those of the other variables: typically the variances are on the scale of 0.1 kcal/mol and 0.2 Å. These deviations may be due to statistical errors resulting from force-matching or the pairwise approximation. Thus, it is reasonable to fit the averaged data obtained from 275 to 375 K as a function of the bulk density, resulting in

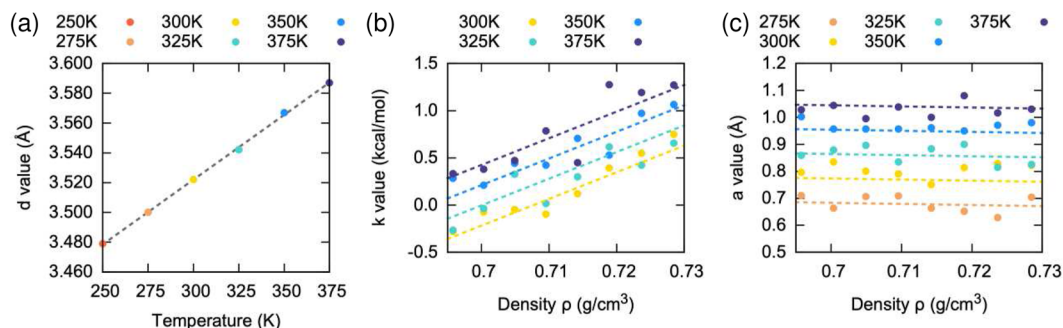


FIG. 8. Determination of the equations of state describing the parametrized d , k , and a variables for the Gaussian CG model. (a) Hard-core repulsive diameter at different temperatures from 250 K (red) to 375 K (navy). Interaction strengths at different temperatures and densities for (b) the collective solvation term and (c) the short-range repulsive Gaussian.

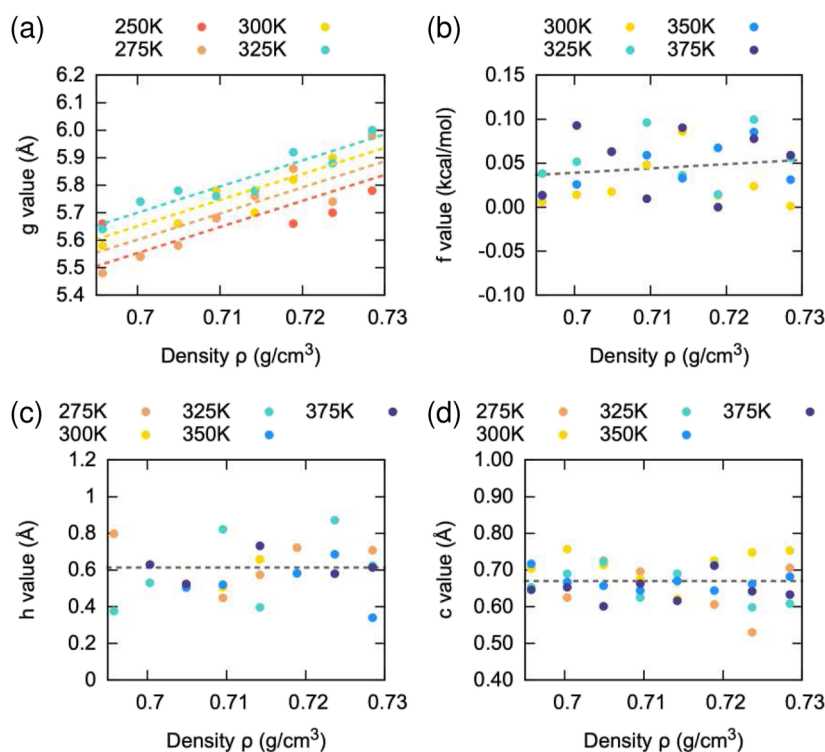


FIG. 9. Determination of the equations of state for the remaining variables (g , f , h , c) of the Gaussian CG acetone. (a) and (b) Local minima of the attractive long-range Gaussians as a function of temperature and density. (b) Interaction magnitudes for the long-range attractive Gaussian interaction as a function of density. (c) and (d) Average behavior of interaction widths of the short-range and long-range Gaussians for different state points. Coloring schemes are the same as for the previous figures.

$$f_{\text{MeCN}}(\rho) = 4.8701 \times 10^{-1} \cdot \rho - 3.0179 \times 10^{-1}, \quad (48a)$$

$$h_{\text{MeCN}} = 0.6131, \quad (48b)$$

$$c_{\text{MeCN}} = 0.6697. \quad (48c)$$

Altogether, we arrive at the Gaussian parametrized model for the acetonitrile system:

$$U_{\text{MeCN}}(R) = a_{\text{MeCN}}(\rho, T) \exp\left(-\frac{(R - 3.8)^2}{0.6697^2}\right) + \left(\frac{d_{\text{MeCN}}(\rho, T)}{R}\right)^{10} - f_{\text{MeCN}}(\rho) \exp\left(-\frac{(R - g_{\text{MeCN}}(\rho, T))^2}{0.6131^2}\right) + k_{\text{MeCN}}(\rho, T) \exp\left(-\frac{R^2}{15}\right). \quad (49)$$

E. Transferability: Thermodynamic state points

1. Methanol

Using Eq. (42), Fig. 10 compares the MS-CG interactions with the simplified CG equations of state based on the parametrized data for bulk methanol at different temperature and density conditions. Figure 10 shows one specific bulk density per temperature for clarity, illustrating the most drastic changes along the system conditions, i.e., 41.7 Å at 300 K, 41.5 Å at 325 K, 41.3 Å at 350 K, 41.1 Å at 375 K, 41.0 Å at 400 K, and 40.9 Å at 425 K. As mentioned earlier, bottom-up CG interactions undergo a continuous yet ambiguous change when varying the system conditions. The simplified Gaussian representation can faithfully modulate the PMF in response to these system variables. The double-well shape at 300 K and 41.7 Å transforms into a shouldered well at 375 K and 41.1 Å and finally becomes a double-shouldered shape at 425 K and 40.9 Å. We note that our models ignore interactions at long ranges ($R > 8$ Å) by design, yet the deviations among interactions within this region remain relatively minor. In turn, our parametrized model shows the capability of capturing the local interactions up to the second coordination shell. We acknowledge that these differences beyond the third coordination shell could potentially affect pressures and compressibilities at the CG level. However, it is worth noting that this issue could be ameliorated by introducing additional Gaussian basis functions beyond the third coordination shell. To comprehensively address this issue, we plan to conduct a systematic investigation into the optimal num-

ber of basis sets and their impact on the thermodynamic properties of CG models in a forthcoming study.

To assess the performance of the simplified equations of state, we performed additional CG simulations using the simplified interactions depicted in Fig. 10 and computed the RDFs obtained from the Gaussian representation, as shown in Fig. 11 (complete RDF values are provided in the supplementary material). As expected, the Gaussian representation accurately captures the structural changes in the liquid phase with high transferability, consistent with the agreement observed in the CG PMFs. While some deviations in the RDFs exhibit numerical instability during the parametrization, e.g., the second attractive Gaussian at higher temperatures, these deviations become relatively smaller as temperature increases. To summarize, our generalized description of the CG interaction in terms of system variables can correctly capture the locations and magnitudes of structural correlations.

2. Acetonitrile

Similarly, we generated the Gaussian representation for CG acetonitrile models using the parametrized equations of state, Eqs. (44)–(48), spanning from 250 to 375 K (see Fig. 12). In this case, we chose the following sets of the system conditions: 45.9 Å at 250 K, 45.8 Å at 275 K, 45.7 Å at 300 K, 45.6 Å at 325 K, 45.5 Å at 350 K, and 45.4 Å at 375 K. The Gaussian model reproduces these trends up to the second coordination shell with great precision.

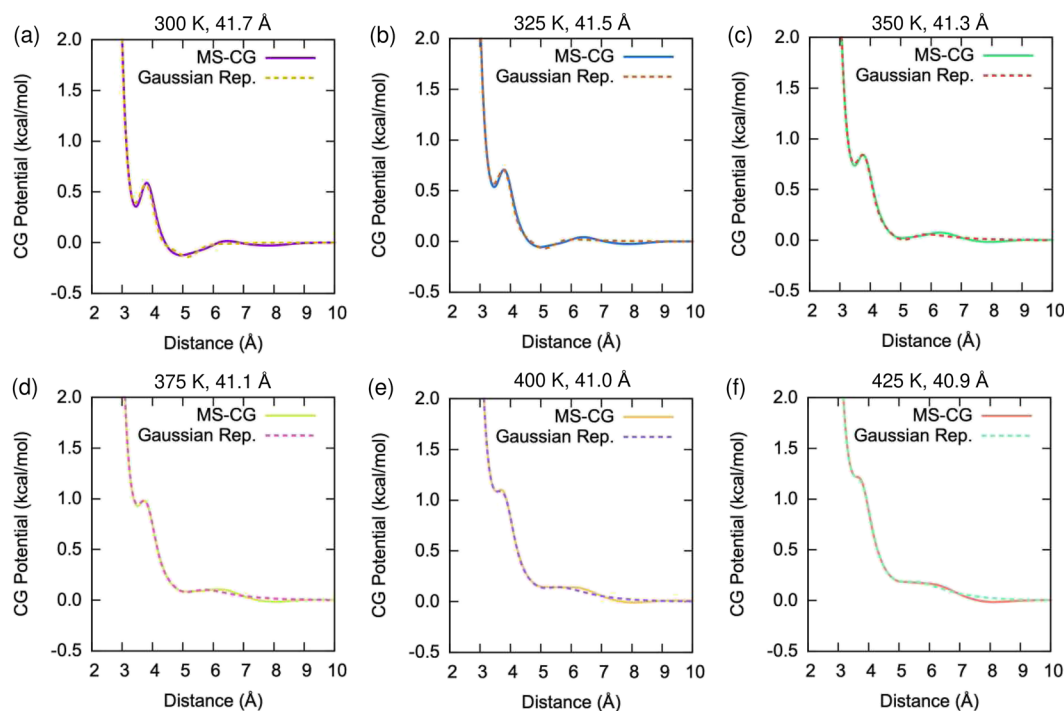


FIG. 10. Transferability analysis of the Gaussian representation by comparing the effective CG methanol potentials (dashed lines) with the reference MS-CG potentials (solid lines) over various thermodynamic state points: (a) 300 K and 41.7 Å, (b) 325 K and 41.5 Å, (c) 350 K and 41.3 Å, (d) 375 K and 41.1 Å, (e) 400 K and 41.0 Å, and (f) 425 K and 40.9 Å.

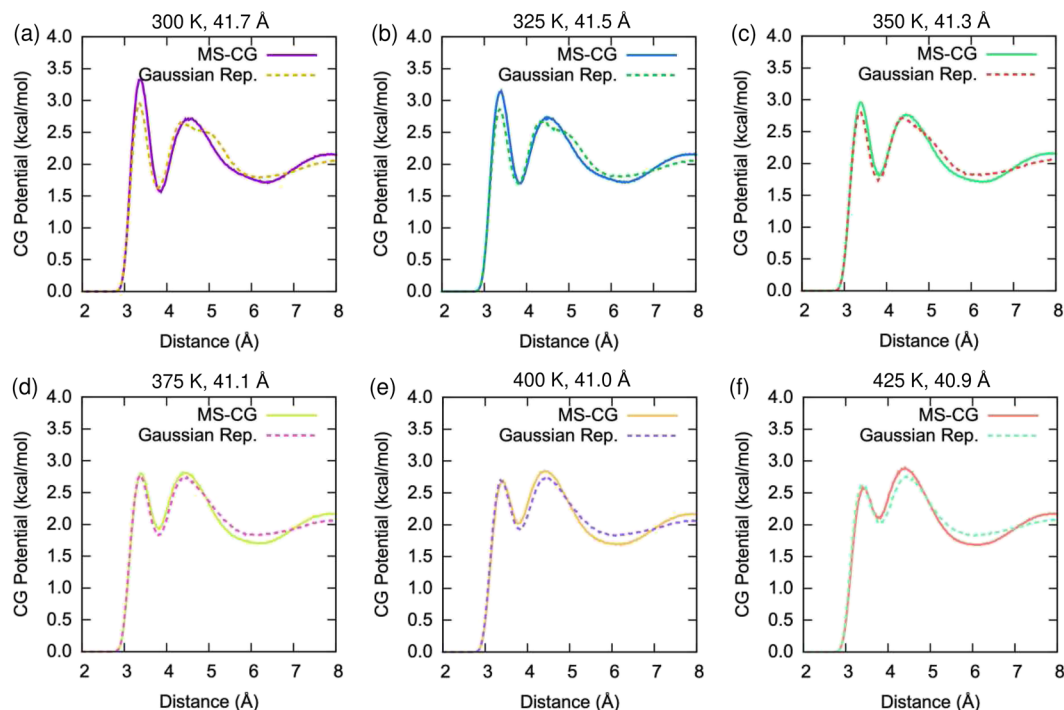


FIG. 11. Radial distribution function $g(R)$ of methanol pairs obtained from the CG simulations using the interactions plotted in Fig. 10. Each panel corresponds to the same models and conditions as in Fig. 10: (a) 300 K and 41.7 Å, (b) 325 K and 41.5 Å, (c) 350 K and 41.3 Å, (d) 375 K and 41.1 Å, (e) 400 K and 41.0 Å, and (f) 425 K and 40.9 Å.

Given that interactions at the third coordination shell near 7–10 Å could potentially lead to non-negligible errors, we verified the accuracy of the Gaussian model by computing the RDF at these conditions. Figure 13 indicates that a third Gaussian is not necessary. Nevertheless, one possible approach to improve this description for future study would be to include an additional attractive Gaussian to construct a more accurate Gaussian CG model. Considering the numerical instability of parametrizing such weak interactions, another direction would be to employ our two-Gaussian model as an initial parameter set and then infer the third Gaussian parameter.

To quantitatively evaluate the accuracy of the Gaussian models, we computed the following error metric $\mathcal{L}$ that has been widely applied in the previous CG modeling works^{29,170}

$$\Delta(g\{\mathcal{L}\}) := \frac{\int_0^{r_{\text{cut}}} |g^{\mathcal{L}}(\mathbf{r}) - g^{\text{target}}(\mathbf{r})| d\mathbf{r}}{\int_0^{r_{\text{cut}}} g^{\text{target}}(\mathbf{r}) d\mathbf{r}}. \quad (50)$$

By definition, Eq. (50) provides a measure of how close the model $\mathcal{L}$ is to the target in terms of the RDFs. Since we only introduced the Gaussians up to the second coordination shell, we chose $g + 1.96 \cdot h$ as the r_{cut} value to obtain a 95% confidence interval. Table I(a) lists the complete error metrics computed among the 60 state points studied in Figs. 1 and 12.

Table I confirms the transferability of our simplified models, with an average error of around 4% for methanol and 7% for acetone-trile. However, upon comparing the performance of the Gaussian

representation to the reference MS-CG, more pronounced differences in RDFs (Figs. 11 and 13) are observed compared to those in the pair interaction profiles (Figs. 10 and 12). Even though it is known that the RDF is often relatively insensitive to the pair potential in practice,^{171,172} in our case, this deviation might be attributed to the approximative nature of our Gaussian representation with the pair forces. Specifically, the truncated interaction profile beyond the third coordination shell and numerical issues in parametrization lead to distinct pair profiles when compared to the force-matched reference. Furthermore, the relatively significant errors observed for certain system conditions suggest a numerical challenge in accurately partitioning the CG interaction into its sub-interactions. To address this, a more robust numerical parametrization scheme can be designed with this condition as an initial target.¹⁷³ Progress in this work is currently underway.

3. Transferability to non-parametrized state points

Having established the accuracy of the Gaussian representation, we now investigate the transferability beyond the parametrized dataset. Even though we have observed gradual changes in the parameters during the optimization process, it remained unclear whether the parametrized models could accurately describe systems at unknown thermodynamic conditions. In order to provide a comprehensive analysis, we report the transferability outside of the dataset by extrapolating the previously parametrized Gaussian models for methanol. We performed two different extrapolations: (1) to

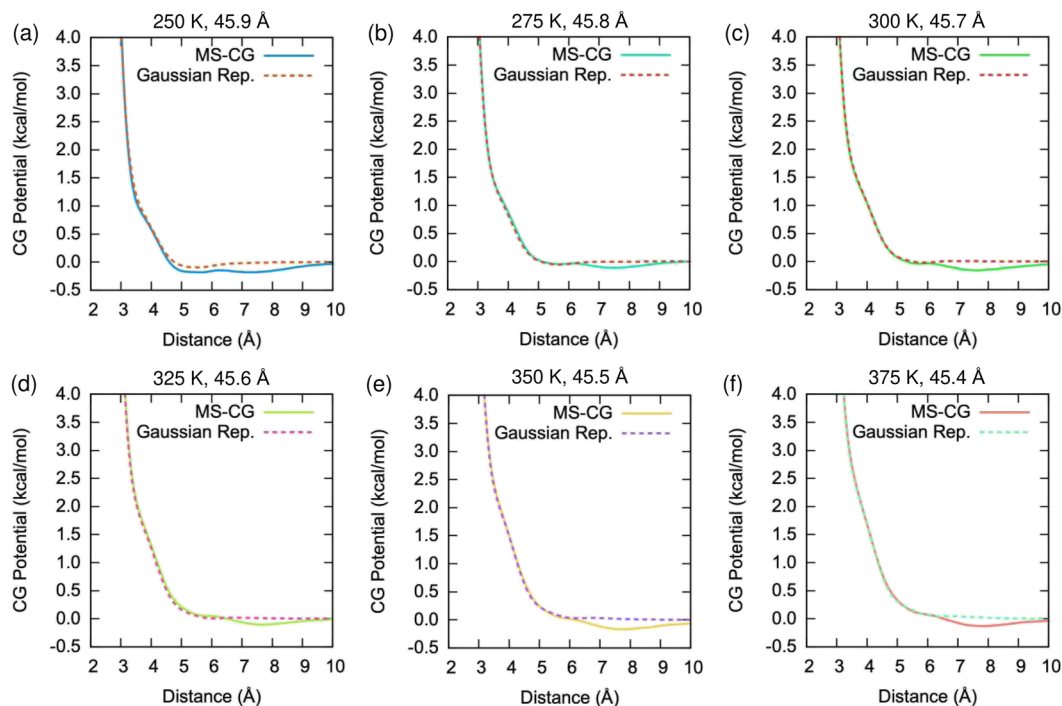


FIG. 12. Transferability analysis of the Gaussian representation by comparing the effective CG acetonitrile potentials (dashed lines) with the reference MS-CG potentials (solid lines) over various thermodynamic state points: (a) 250 K and 45.9 Å, (b) 275 K and 45.8 Å, (c) 300 K and 45.7 Å, (d) 325 K and 45.6 Å, (e) 350 K and 45.5 Å, and (f) 375 K and 45.4 Å.

lower temperatures ranging from 225 to 275 K, and (2) to lower densities (larger box lengths) ranging from 41.8 to 42.4 Å.

Notably, Fig. 14 shows an excellent recapitulation of the pairwise CG PMF at the non-parametrized state points. Previously, temperature transferability has been achieved mainly under fixed volume^{26,161,174} with limited pressure transferability,^{29,43,175} since the CG PMFs become the Helmholtz free energy $\Delta A(R)$, and then the entropic contributions $-T\Delta S(R)$ can be adjusted accordingly. Our framework greatly enhances the extensibility of the CG model to different temperatures and pressures without such strict limitations. Hence, we believe that this approach paves the way for the design of transferable CG models under any system conditions. Yet, further studies should aim to investigate the variation of the parameters and equations of state across different molecular systems as well as to develop a generalized design principle for any CG system. Such an effort can facilitate the construction of a database of bottom-up CG interactions with improved transferability.

F. System transferability: Combining rules for Gaussian representation

1. Remarks on system transferability

In the previous sections, we have demonstrated the transferability of the Gaussian-based CG model and its corresponding equations of state to different state points. Yet, another direction

of transferability is extending the applicability of the given model to different system compositions. We refer to this type of transferability as “system transferability.” Unlike state point transferability, achieving system transferability is much more challenging due to the need for a careful design of the combined interactions between more than two different CG sites, also known as “cross-interactions.” Because effective CG interactions are expressed as the many-body PMF, it is likely impossible to precisely determine the combined interaction via a bottom-up combining rule. However, in recent years, certain methodologies have been partially successful in this regard.^{26,30,46,47,176} In recent work, we aimed to derive the combined interactions between methanol and chloroform in mixed solution based on the interaction profiles from the liquids in their bulk phases.²⁶ We were able to approximate the combined CG interactions by capturing necessary structural correlations, but this attempt was only made possible by projecting the effective CG energy functional into the Gay-Berne potential form^{177,178} and then combining them using rod-sphere energetics.^{179,180} This approach, thus, is not extensible to any systems with different interaction forms other than the Gay-Berne interaction, resulting in limited applicability. The alternative theory that we recently developed¹⁰⁷ was to introduce the appropriate collective variables to decompose CG interactions into a combination of different conformational basins, e.g., using ultra-coarse-graining theory^{29,45,47,106,108} or multi-configurational coarse-graining,¹⁸¹ and predicted the cross-interaction akin to quantum mechanical approaches treating coupled states.¹⁸² However, imple-

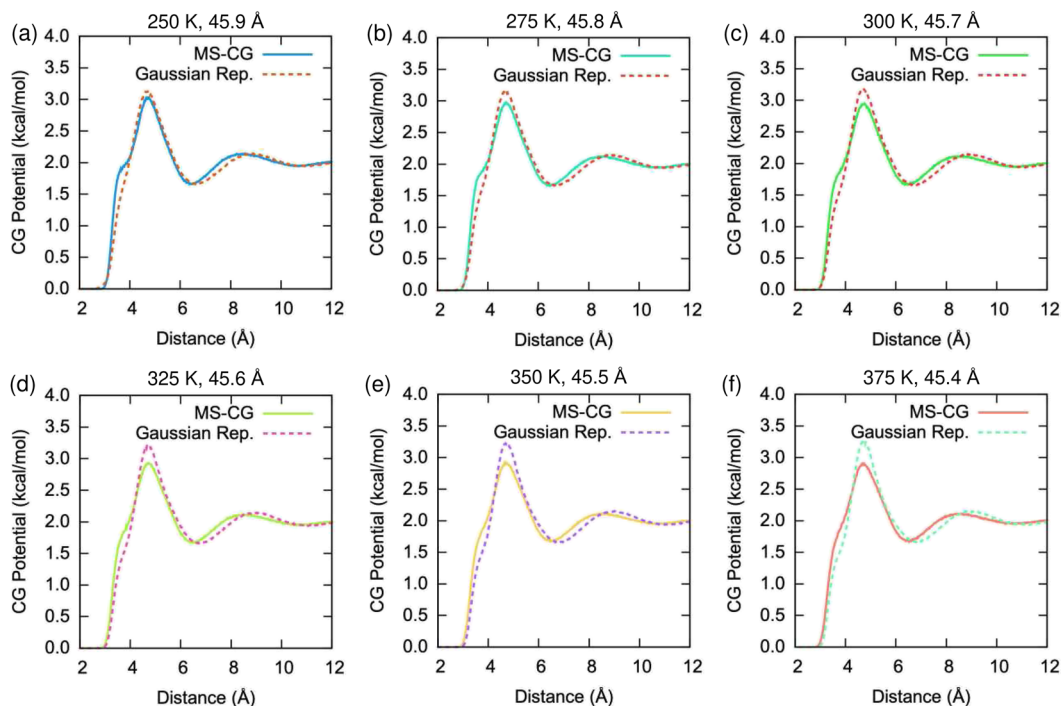


FIG. 13. Radial distribution function $g(R)$ of acetonitrile pairs obtained from the CG simulations using the interactions plotted in Fig. 12. Each panel corresponds to the same models and conditions as in Fig. 12: (a) 250 K and 45.9 Å, (b) 275 K and 45.8 Å, (c) 300 K and 45.7 Å, (d) 325 K and 45.6 Å, (e) 350 K and 45.5 Å, and (f) 375 K and 45.4 Å.

menting this approach in complex systems can be challenging since it requires an additional collective variable to distinguish between different configurations.¹⁰

On the other hand, by decomposing the CG interactions of the bulk phases in terms of the Gaussian basis functions, it is conceivable that one could combine each sub-interaction consisting of the overall CG interactions. Since the Gaussian function is described by its mean and variance, combining the Gaussian interactions can be more intuitive than devising a complex combining rule for the high-dimensional PMFs from first principles. Once we mix the interactions between two different Gaussian sub-interactions, we can construct the overall cross-interaction by combining these mixed sub-interactions.

With this in mind, we propose here a novel combining rule based on the Gaussian sub-interactions in this section. We first validate our approach by applying it to the LJ system, for which the combining rules are well-known by Lorentz and Berthelot.^{183,184} After comparing our approach to the mixed interaction from the Lorentz–Berthelot rule, we extend it to the mixed CG liquid system: methanol-acetonitrile mixture.

2. LJ system

In order to test the fidelity of the proposed combining rule, we consider the following two LJ systems, where the particles interact via 6–12 LJ interactions with the interaction parameter of $\sigma_1 = 4.0$ Å, $\epsilon_1 = 0.20$ kcal/mol for the first system (“weaker” LJ system) and $\sigma_2 = 4.0$ Å, $\epsilon_2 = 0.36$ kcal/mol for the second system (“stronger” LJ

system). From the standard Lorentz–Berthelot combining rule,^{183,184} their mixed interaction parameter is readily obtained as

$$\sigma_{1-2} = \left(\frac{\sigma_1 + \sigma_2}{2} \right) = 4.0 \text{ Å}, \quad (51a)$$

$$\epsilon_{1-2} = \sqrt{\epsilon_1 \epsilon_2} = 0.2683 \text{ kcal/mol}. \quad (51b)$$

We now demonstrate that the Gaussian parametrization strategy can be utilized to derive the conventional mixed interaction of simple LJ systems [Eq. (51)] in a similar way. Since the single LJ interaction has one attractive well, we only consider the attractive Gaussian with the hard sphere-like repulsion term:

$$U_{LJ}(R) = \left(\frac{d_{LJ}}{R} \right)^{e_{LJ}} - f_{LJ} \exp \left(- \frac{(R - g_{LJ})^2}{h_{LJ}^2} \right). \quad (52)$$

While, in principle, it is possible to determine the d_{LJ} , e_{LJ} , f_{LJ} , g_{LJ} , and h_{LJ} parameters directly using the same procedures as for the CG liquid systems, in order to ensure numerical stability of the parametrization process, we first estimate the effective hard sphere diameter term d_{LJ} using classical perturbation theory. This additional step helps eliminate any extra degeneracies in the parameter space. Following our recent work on mapping CG systems to effective hard spheres,^{63–65} we estimate the Barker–Henderson diameter σ_{BH} directly from the interaction profile^{58,59} using

$$\sigma_{BH} = \int_0^{R_0} [1 - \exp(-\beta U(R))] dR. \quad (53)$$

TABLE I. Error metric percentage $100 \times \Delta(g\{\mathcal{L}\})$ from Eq. (50) using the RDF at various temperatures for the CG liquid models: (a) Methanol and (b) Acetonitrile.

(a) Methanol									
Temp	Box length								
	40.9 Å	41.0 Å	41.1 Å	41.2 Å	41.3 Å	41.4 Å	41.5 Å	41.6 Å	41.7 Å
300 K	4.38	4.51	4.67	4.84	5.12	5.33	5.44	5.52	5.52
325 K	4.22	4.22	4.25	4.35	4.55	4.68	4.80	4.82	4.82
350 K	4.41	4.23	4.15	4.09	4.03	4.14	4.19	4.20	4.20
375 K	4.57	4.38	4.20	4.07	3.90	3.89	3.73	3.69	3.69
400 K	4.72	4.59	4.35	4.14	3.95	3.77	3.62	3.46	3.46
425 K	4.82	4.56	4.33	4.09	3.93	3.73	3.47	3.33	3.33

(b) Acetonitrile									
Temp	Box length								
	45.4 Å	45.5 Å	45.6 Å	45.7 Å	45.8 Å	45.9 Å	46.0 Å	46.1 Å	
250 K	8.20	7.92	7.52	7.14	6.67	6.21	5.73	5.36	
275 K	8.06	7.78	7.35	7.02	6.58	6.19	5.81	5.48	
300 K	8.07	7.74	7.36	7.00	6.64	6.20	5.86	5.51	
325 K	8.15	7.82	7.47	7.07	6.68	6.27	5.89	5.54	
350 K	8.35	8.08	7.72	7.30	6.88	6.51	6.12	5.76	
375 K	8.47	8.19	7.83	7.47	7.02	6.68	6.23	5.83	

In Ref. 64, we demonstrated that σ_{BH} could effectively capture the repulsive region in relatively soft molecules, e.g., water. At 300 K, Eq. (53) gives 3.741 Å for the weaker LJ system and 3.811 Å for the stronger LJ system. The values obtained were used to parametrize Eq. (52), resulting in the following Gaussian representation for LJ interactions in units of kcal/mol as

$$U_{\text{LJ}}^{1-1}(R) = \left(\frac{3.741}{R(\text{\AA})} \right)^{14.06} - 0.3856 \exp \left(- \frac{(R - 3.529)^2}{3.091(\text{\AA}^2)} \right), \quad (54a)$$

$$U_{\text{LJ}}^{2-2}(R) = \left(\frac{3.811}{R(\text{\AA})} \right)^{16.17} - 0.4282 \exp \left(- \frac{(R - 4.462)^2}{1.281(\text{\AA}^2)} \right). \quad (54b)$$

To obtain the combined LJ 1–2 interaction $U_{\text{LJ}}^{1-2}(R)$ by mixing U_{LJ}^{1-1} and U_{LJ}^{2-2} , we need to consider the hard sphere-like repulsion term as well as the Gaussian interaction. For the hard sphere-like contribution, it is reasonable to approximate d_{LJ}^{1-2} as the arithmetic mean of d_{LJ}^{1-1} and d_{LJ}^{2-2} , as done previously for constructing a hard sphere mixture system.¹⁸⁵ For the combined Gaussian interactions, we utilized the combining rule suggested by the Gaussian core model for designing binary mixtures,⁸³ where the strength follows the geometric average and the width follows the arithmetic average. Since the intensity of the Gaussian functions is in units of interaction strength, the magnitude of the combined sub-interaction between Gaussian sub-interactions $-f_{\text{LJ}}^{1-1} \exp \left(- \frac{(R - g_{\text{LJ}}^{1-1})^2}{(h_{\text{LJ}}^{1-1})^2} \right)$ and $-f_{\text{LJ}}^{2-2} \exp \left(- \frac{(R - g_{\text{LJ}}^{2-2})^2}{(h_{\text{LJ}}^{2-2})^2} \right)$ will also follow

the geometric average, and the mean of combined interaction can be regarded as their geometric average of

$$f_{\text{LJ}}^{1-2} := \sqrt{f_{\text{LJ}}^{1-1} f_{\text{LJ}}^{2-2}}. \quad (55)$$

Similarly, the characteristic lengths appearing in Gaussians g_{LJ}^{1-2} and h_{LJ}^{1-2} are also designed to be consistent with the Gaussian core model as

$$g_{\text{LJ}}^{1-2} := \frac{g_{\text{LJ}}^{1-1} + g_{\text{LJ}}^{2-2}}{2}, \quad (56a)$$

$$(h_{\text{LJ}}^{1-2})^2 = \frac{(h_{\text{LJ}}^{1-1})^2 + (h_{\text{LJ}}^{2-2})^2}{2}. \quad (56b)$$

Finally, we arrive at the combined interaction of the form:

$$U_{\text{LJ}}^{1-2}(R) = \left(\frac{3.776}{R(\text{\AA})} \right)^{15.12} - 0.4063 \exp \left(- \frac{(R - 4.000)^2}{2.186(\text{\AA}^2)} \right). \quad (57)$$

Remarkably, the final Eq. (57) is almost identical to the reference potential given by the Lorentz–Berthelot combining rule, as shown in Fig. 15(b). We further corroborated the fidelity of the new combining rule by performing LJ simulations. The LJ system consisted of 216 particles with 108 weaker and 108 stronger particles, and we used a cubic simulation box ($26.25 \times 26.25 \times 26.25 \text{ \AA}^3$) to ensure a liquid phase. The initial configuration was randomized at a lattice with a spacing of 4.25 Å, and consequently we performed the LJ simulation for 2.5 ns. The final structural correlation from the cross-interaction is illustrated in Fig. 15(c) by computing the

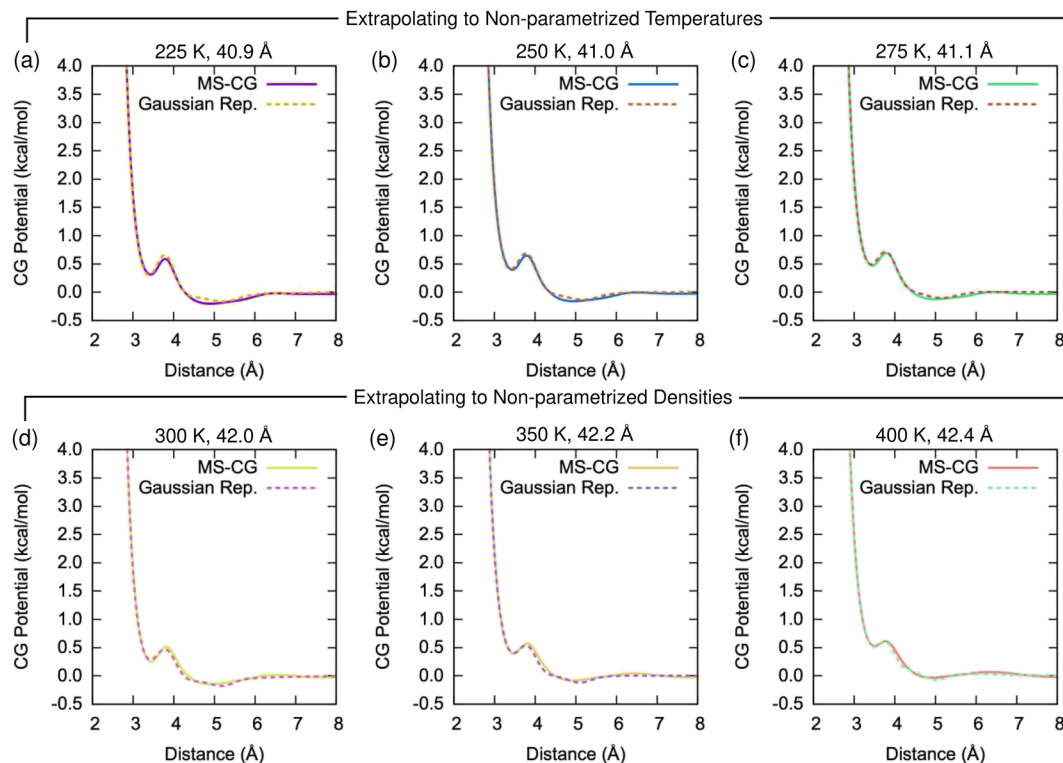


FIG. 14. Validating the transferability of the Gaussian model by extrapolating to thermodynamic state points beyond the parametrized conditions. Two types of extrapolation were considered: non-parametrized lower temperature conditions for (a) 225 K, (b) 250 K, (c) 275 K and lower density conditions for (d) 42.0 Å, (e) 42.2 Å, (f) 42.4 Å.

RDF between the weaker and stronger particles, showing excellent agreement with the well-known Lorentz–Berthelot combined model.^{183,184}

Despite the close agreement, the RDF shown in Fig. 15(c) suggests the parametrized interaction form based on Eq. (52) slightly underestimates the effective hard sphere diameter for the target LJ system. This observation aligns with the computed d_{LJ} values for 1–1, 2–2, and 1–2 pairs, which are 3.741, 3.811, and 3.776 Å, respectively, all smaller than the reference $\sigma = 4.0$ Å. Generally, the relative magnitude of the effective hard sphere diameter obtained through hard sphere mapping, when compared to the reference LJ liquid, varies across different reduced temperatures and densities. The temperature and density dependence of the effective hard sphere diameter for LJ potential parameters was extensively investigated by Ben-Amotz and Herschbach.¹⁸⁶ In Ref. 186, they reported the hard sphere diameters for the LJ potential as a function of reduced temperatures at $\rho^* = 0.7$. Our system condition, $\rho_{LJ}^* = \rho\sigma^3 = 0.76$, is proximate to the reported reduced density. Since our system is a mixture with different ϵ values, estimating the reduced temperature is not straightforward, but it should lie between $T_1^* = k_B T / \epsilon_1 = 2.98$ and $T_2^* = k_B T / \epsilon_2 = 1.66$. From Fig. 7 in Ref. 186, the Barker–Henderson approach, based on Eq. (53), is expected to yield a hard sphere diameter smaller than the reference LJ value, i.e., $\sigma^* = \sigma / \sigma_{LJ} < 1$. This also explains why our method slightly underestimates the effective hard sphere diameters for the mixture system.

While this conventional treatment utilizing Eq. (52) reasonably estimates the size of LJ particles, we would like to highlight a systematic method that could be considered for the future enhancement of our approach. One potential direction is to incorporate the entire contribution of the repulsive forces from the LJ interaction, which is not completely addressed by the conventional Barker–Henderson treatment. Specifically, the hybrid Barker–Henderson (hBH) approximation¹⁸⁷ employs a decomposition scheme akin to the Weeks–Chandler–Andersen approach,^{60–62} enabling the decomposition of the interaction $U(R)$ as

$$U^*(R) = \begin{cases} U(R) + \epsilon, & (R < \sigma_m) \\ 0, & (R > \sigma_m) \end{cases} \quad (58)$$

where $\sigma_m = 2^{1/6}\sigma$ represents the distance at which $U(R)$ attains its minimum value. Note that Eq. (58) divides $U^*(R)$ in a manner akin to the Weeks–Chandler–Andersen theory^{60–62} at distances larger than σ , and then the final effective hard sphere diameter is estimated by integrating the modified Barker–Henderson integrand up to σ_m , i.e.,

$$\sigma_{hBH} = \int_0^{\sigma_m} [1 - \exp(-\beta U^*(R))] dR. \quad (59)$$

Given that the integration range in Eq. (59), spanning from $R = 0$ to $R = \sigma_m$, is larger than that of the conventional

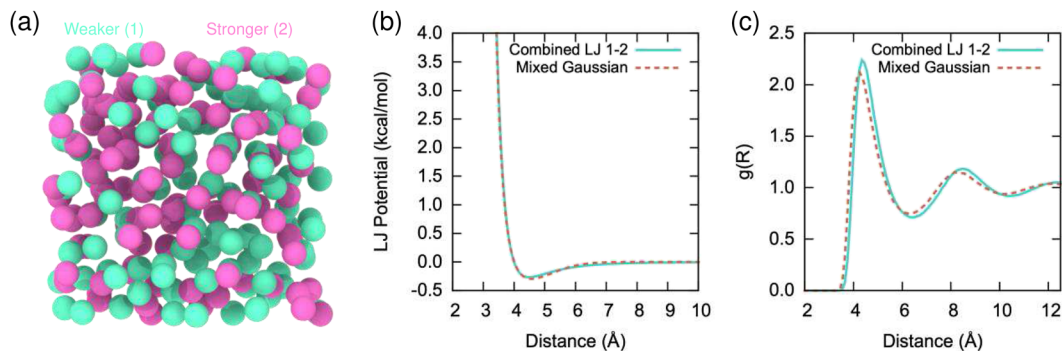


FIG. 15. Generalized combining rule for the Gaussian representation of the LJ interactions. (a) Final snapshot (configuration) of the LJ mixture composed of weaker (type 1, teal) and stronger (type 2, pink) LJ particles. (b) LJ interactions of the combined weaker-stronger cross-interaction interpolated using the reference Lorentz–Berthelot rule (teal solid line) and the proposed combining rule (red dashed line). (c) Structural cross-correlations using the RDF for weaker-stronger pairs from simulating interactions shown in panel (b) to the system in panel (a).

Barker–Henderson approach, Eqs. (58) and (59) are expected to correct the underestimated σ_{BH} ¹⁸⁷ observed in Fig. 15(c) and further improve the combining rule presented in this section.¹⁸⁸

3. Methanol-acetonitrile mixture

Motivated by the success of the new combining rule for accurately predicting the behavior of LJ liquid mixtures, we then aimed to extend its application to more complex CG interactions in liquid mixtures. A heterogeneous liquid mixture was prepared by placing 500 methanol and 500 acetonitrile molecules in a cubic box. This initial configuration was generated using the Packmol¹⁵⁵ and VMD¹⁵⁶ programs, following the same settings used for the bulk simulations. Energy minimization was performed on the initial configuration to eliminate artificial stress on the system. Next, the system was annealed to the target temperature of 300 K for 0.1 ns under the constant NVT dynamics using the Nosé–Hoover thermostat^{157,158} with $\tau_{\text{NVT}} = 0.1$ ps. The system volume was then equilibrated at 1 atm using Andersen barostat¹⁸⁹ with $\tau_{\text{NPT}} = 0.1$ ps for 1 ns, giving an equilibrated system box length of 43.739 Å. Finally, the FG trajectory for the CG parametrization was generated under constant NVT dynamics for 2.5 ns using the same thermostat setting. The final configuration at the FG level is shown in Fig. 16(a), along with its corresponding mapped CG system in Fig. 16(b).

In order to transfer the bulk methanol and acetonitrile interactions to the mixed phase, we started by constructing the self-interactions for methanol–methanol and acetonitrile–acetonitrile. The bulk density for each species was carefully adjusted by considering an excess volume term. It is known that for liquids, the excess volume in a mixture can either be larger (repulsively mixing) or smaller (attractively mixing) depending on the nature of their cross-interactions. In our case, the mixed volume term was estimated using the naïve bulk volume per molecule (or partial volume), V_u , of each species. For methanol, the naïve bulk volume per molecule was calculated as

$$V_u^{\text{MeOH}} = \frac{V_{\text{bulk}}^{\text{MeOH}}}{N_{\text{bulk}}} = 69.748 \text{ Å}^3/\text{molecule}. \quad (60)$$

Similarly, we obtain $V_u^{\text{MeCN}} = 96.058 \text{ Å}^3/\text{molecule}$. The mixing volume in this methanol/acetonitrile mixture system can then be readily calculated as

$$V_{\text{ex}} = V_{\text{mixture}} - (N^{\text{MeOH}} V_u^{\text{MeOH}} + N^{\text{MeCN}} V_u^{\text{MeCN}}) = 771.499 \text{ Å}^3. \quad (61)$$

The non-zero value of V_{ex} indicates that changes in the effective volume of liquid molecules due to mixing must be considered to accurately account for the densities of each species in the mixture. As our mixture has an equal number of species, i.e., $\chi_{\text{MeOH}} = \chi_{\text{MeCN}} = 0.5$, we assume that this excess volume effect contributes equally to each liquid. By using the corrected volume, we arrive at the overall densities

$$\rho^{\text{MeOH}} = \frac{m^{\text{MeOH}}}{N^{\text{MeOH}} V_u^{\text{MeOH}} + \chi_{\text{MeOH}} V_{\text{ex}}} = 0.7383 \text{ g} \cdot \text{cm}^{-3}, \quad (62a)$$

$$\rho^{\text{MeCN}} = \frac{m^{\text{MeCN}}}{N^{\text{MeCN}} V_u^{\text{MeCN}} + \chi_{\text{MeCN}} V_{\text{ex}}} = 0.6930 \text{ g} \cdot \text{cm}^{-3}, \quad (62b)$$

for methanol and acetonitrile, respectively. Plugging in χ_{MeOH} and χ_{MeCN} as 0.5 into Eqs. (62a) and (62b) at a temperature of 300 K results in the reduced interaction forms shown in Figs. 17(a)–17(c). We note that if we use the naïve bulk volume for each species $N^{\text{MeOH}} V_u^{\text{MeOH}}$ and $N^{\text{MeCN}} V_u^{\text{MeCN}}$, we obtain a CG PMF that is significantly more deviated, underscoring the importance of considering the effect of excess volume.

In Fig. 17, we assumed that each liquid in the mixture would behave identically to its bulk phase when borrowing the Gaussian representation from bulk liquids. This assumption may be strong since there is a cooperative correlation between the two molecules in the mixture; however, our hypothesis yields fairly reasonable CG interactions compared to the reference CG interactions in the mixture from MS-CG. The self-interactions shown in Figs. 17(a) and 17(b) agree with the reference MS-CG potentials and capture their shape accordingly. When comparing with the MS-CG potential as a reference, we would like to emphasize that the MS-CG interactions used here to parametrize the Gaussian representation

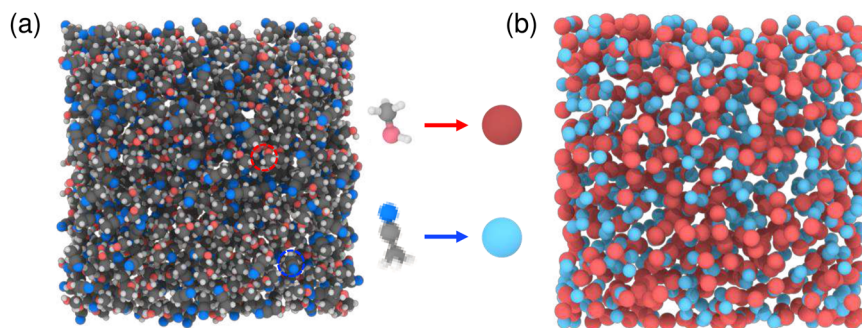


FIG. 16. Two-component CG liquid mixtures constructed by combining the CG methanol (red bead) and the CG acetonitrile (sky blue bead). (a) The final FG configuration of the generated liquid mixtures after the constant NVT simulation. (b) CG trajectories from panel (a) by manually mapping each molecule to its center of mass.

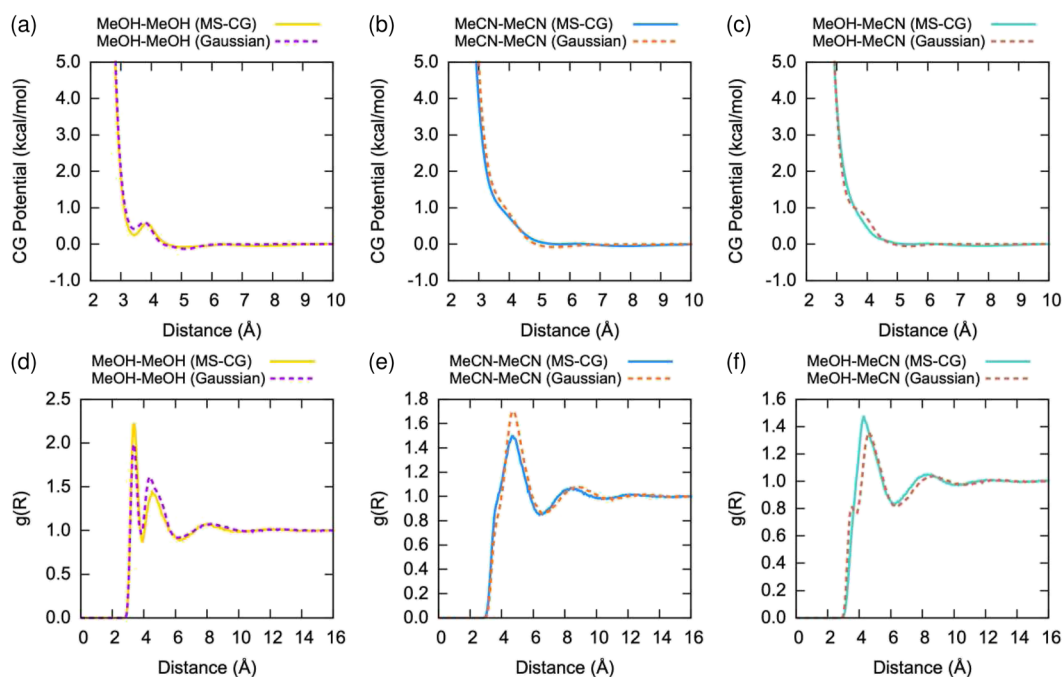


FIG. 17. Assessment of the system transferability for CG models in liquid mixtures using the developed combining rule [Eqs. (62)–(64)]. Panels (a)–(c) compare the effective CG interactions of the CG pairs, and panels (d)–(f) juxtapose the RDFs computed from the corresponding CG simulations. (a) and (d): Methanol–methanol pairs from the reference MS-CG (yellow solid) and the Gaussian (purple dashed) models. (b) and (e): Acetonitrile–acetonitrile pairs from the reference MS-CG (blue solid) and the Gaussian (red dashed). (c) and (f): Methanol–acetonitrile cross-pairs from the reference MS-CG (green solid) and the Gaussian (red dashed).

accurately reflect the structures of the liquid mixture, as shown in Appendix B, where the CG simulations precisely recapitulate the atomistic structural correlation from the atomistic MD simulation.

Designing the cross-interaction between methanol and acetonitrile is significantly more challenging than the self-interactions because there is no rigorous way to infer it from self-interactions. Nonetheless, we attempt to derive the cross-interaction by extending our argument from the simple LJ system. Since both liquids have the same number of Gaussian basis functions due to their similar coordination profiles, we can envision that the mixed interactions can be constructed by combining the Gaussians at the

same coordination shells. Since each sub-interaction can be separated, we assume that there is no correlation between the Gaussians located at different coordination shells, similar to counting the local cross-density function.^{104,190} Then, the combined Gaussian between $\epsilon_i^{1-1} N(\mu_i^{1-1}, (\sigma_i^{1-1})^2)$ and $\epsilon_i^{2-2} N(\mu_i^{2-2}, (\sigma_i^{2-2})^2)$ at the same $\{\bar{R}_i\}_i$ interval can be defined as

$$\epsilon_i^{1-2} N(\mu_i^{1-2}, (\sigma_i^{1-2})^2) \doteq \sqrt{\epsilon_i^{1-1} \epsilon_i^{2-2}} N\left(\frac{\mu_i^{1-1} + \mu_i^{2-2}}{2}, \frac{(\sigma_i^{1-1})^2 + (\sigma_i^{2-2})^2}{2}\right). \quad (63)$$

For the collective solvation interactions from higher-order correlations, we use the arithmetic average of the interaction strength, k , since the magnitude of the mean-field collective interactions should correspond to their mesoscopic densities.

Combined together, our description of the reduced representation of the cross-interaction is summarized as

$$U^{1-2}(R) = \left[\frac{(d_{1-1} + d_{2-2})/2}{R} \right]^{\frac{\epsilon_{1-1} + \epsilon_{2-2}}{2}} + \sum_i \sqrt{\epsilon_i^{1-1} \epsilon_i^{2-2}} N \left(\frac{\mu_i^{1-1} + \mu_i^{2-2}}{2}, \frac{(\sigma_i^{1-1})^2 + (\sigma_i^{2-2})^2}{2} \right) + \frac{k_{1-1} + k_{2-2}}{2} \exp \left(-\frac{R^2}{l^2} \right). \quad (64)$$

Figure 17(c) compares the suggested cross-interactions, using Eq. (64), with the reference MS-CG from the mixture. Despite a slight mismatch at 3–4 Å, our Gaussian combining approach is able to reproduce the general shape of the interaction profile. Notably, the short-range and long-range behaviors at $R < 3.5$ Å and $R > 6$ Å are in excellent agreement, indicating the high fidelity of our proposed combining rule.

To evaluate the performance of the combined model, we computed the RDF of the mixture and compared it with the MS-CG models that were directly parametrized from the mixture. The structural correlation displayed by the cross-interaction, as illustrated in Fig. 17(e), is more similar to the acetonitrile-acetonitrile self-interaction than the methanol-methanol. This particular feature is challenging to predict when designing the cross-interaction solely as an average of self-interactions. However, our proposed combined interactions can reasonably capture the overall cross-correlation profile, exhibiting good agreement, reproducing the local minima and maxima, and accurately describing the long-range behavior. There is room for systematic improvement of our combining rule by incorporating additional Gaussian basis sets for bulk interactions and employing an improved scheme to determine the hard-core distance.

IV. CONCLUDING REMARKS

Although bottom-up coarse-grained (CG) models can reproduce structural correlations with much less computational cost, their interactions are high-dimensional quantities that can only be determined from a specific fine-grained (FG) system. Therefore, accurately estimating CG interactions under unknown conditions is often not feasible. Additionally, unlike FG simulations or other top-down models, constructing cross-interactions between different CG entities presents a major challenge as there is no simple analytical functional form underlying the bottom-up CG interactions. This paper proposes a novel parametrization method for understanding the complex nature of the many-body CG interactions.

Our theory, inspired by classical liquid state perturbation theory, allows us to decouple the hard sphere repulsion term at short distances and perturbative interactions at long-range distances. In order to describe the long-range interactions, we introduced the Gaussian function as a basis, since each localized interaction gives rise to local densities that can be effectively separated from each other. Figure 18 schematically depicts this concept. The final

form of the simplified interaction consists of several Gaussian sub-interactions, along with the hard-core repulsion (inverse power law), and the collective solvation interaction terms that account for the mesoscopic repulsive interactions, which are derived from combining the Yvon-Born-Green equation with the Ornstein-Zernike integral equation. However, even when using two Gaussian functions with additional interactions, this functional representation requires ten parameters to be determined.

The goal of the proposed parametrization scheme for Gaussian basis functions is twofold, since each sub-interaction is related to the structural profile and density at a localized distance interval. The first goal is to determine the parameter values for each state point of the system, while the next goal is to infer a simple set of equations of state that can provide a more generalized description of how these interactions will change under different thermodynamic state points. To achieve this, we designed a sequential parametrization scheme that determines the largely varying parameter first by separating each interaction based on the CG PMF profile obtained by force-matching. We then confirmed that these parameters follow a simple linear relationship with temperature or density. Using the systematically parametrized equations of state, we successfully interpolated the interactions within the parametrized state points and extrapolated them outside the dataset. For both cases, we accurately reproduced the CG PMFs and the corresponding pair correlations in good agreement with the reference, supporting the high transferability of these reduced models.

We expanded our transferability analysis by predicting the cross-interaction between two CG sites in a multi-component mixture. Drawing on previous studies on combining rules for hard spheres and the Gaussian core models,^{83,185} we developed a generalized combining rule for two different Gaussian-represented CG models. Because of the Gaussian nature of basis sets, the proposed combining rule separately mixes each sub-interaction and then recombines them. By applying the proposed combining rule to a simple LJ and then a methanol-acetonitrile mixture, we were able

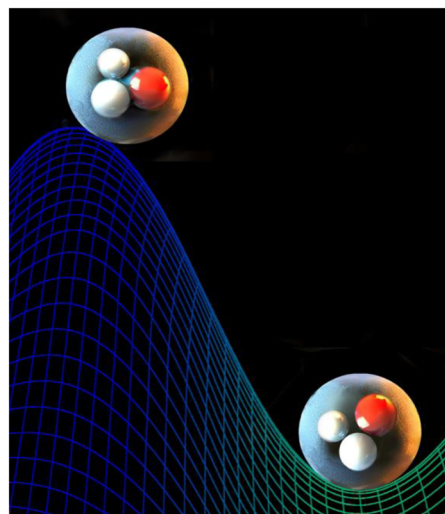


FIG. 18. Schematic illustrating the use of local Gaussian basis sets to describe CG liquid interactions.

to approximate the CG PMFs in different chemical environments. This achievement addresses a significant bottleneck in the field of bottom-up CG modeling.

In conclusion, our study provides a promising approach to enhance the accuracy and transferability of bottom-up CG models. We note that the presented parametrization protocol serves as an initial step in this direction to thoroughly address these challenges. As algorithmic development improves in being able to parametrize these variables, it is expected that further progress can be achieved in the overall performance of the simplified model. One important direction is to understand the physical nature of the equations of state for CG interactions. While we employed simple polynomial and exponential functional forms derived from physical principles, further studies in this direction should focus on elucidating the microscopic origins of these functional forms and deriving them from the bottom-up methodology. Addressing these areas would resolve any remaining uncertainties regarding the transferability problem. Another interesting direction is to examine how the devised combining rule would work for different CG systems with distinct coordination profiles. We chose methanol and acetonitrile as prototypical systems, where these molecules share a similar molecular structure. The remaining question is how to combine Gaussian basis sets that range over different distances, which will be of considerable interest for complex systems such as biomolecules or material systems. Investigating these avenues of research will provide deeper insight into the design of bottom-up CG models and pave the way for more accurate and efficient simulations of complex systems at varying spatiotemporal scales.

SUPPLEMENTARY MATERIAL

See the supplementary material for a complete description of the interpolated Gaussian CG interactions of methanol and acetonitrile, including the corresponding RDFs in comparison with those obtained by the MS-CG references for the range of thermodynamic state points analyzed in this study.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (NSF Grant No. CHE-2102677). Simulations were performed using computing resources provided by the University of Chicago Research Computing Center (RCC). J.J. acknowledges insightful discussions with Dr. Yining Han, Dr. Alexander J. Pak, and Dr. Ziwei He for feedback on the manuscript. We also thank Joohyun Jin for designing the graphical abstract. J.J. acknowledges an Arnold O. Beckman Postdoctoral Fellowship from the Arnold and Mabel Beckman Foundation.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Jaehyeok Jin: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Val-

idation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Jisung Hwang:** Conceptualization (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Writing – review & editing (supporting). **Gregory A. Voth:** Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this work are available from the corresponding author upon request.

APPENDIX A: DERIVATION OF EQ. (15)

Starting from Eq. (5), we can express the local density ρ_I as

$$\rho_I(R) = \frac{4\pi\rho}{V_R} \left\{ \sum_i \int_{\bar{R}_{i-1}}^{\bar{R}_i} r'^2 g(r') dr' \right\}. \quad (\text{A1})$$

Equation (A1) can be further expressed using the reversible work theorem [Eq. (6)], resulting in

$$\rho_I(R) = \frac{4\pi\rho}{V_R} \left\{ \sum_i \int_{\bar{R}_{i-1}}^{\bar{R}_i} r'^2 \exp[-\beta U_2(r')] dr' \right\}. \quad (\text{A2})$$

Now, considering the perturbative treatment of the PPMF, we can decompose $U_2(r')$ into $U'_2(r')$ and $U_{\text{HS}}(r')$. Additionally, it is reasonable to assume that $U_{\text{HS}}(r') = 0$ in the region between $\bar{R}_{i-1}$ and $\bar{R}_i$ due to the short-ranged and highly repulsive characteristics of $U_{\text{HS}}(r')$. Therefore, we arrive at

$$\rho_I(R) = \frac{4\pi\rho}{V_R} \left\{ \sum_i \int_{\bar{R}_{i-1}}^{\bar{R}_i} r'^2 \exp[-\beta U'_2(r')] dr' \right\}. \quad (\text{A3})$$

By introducing Gaussian basis sets, $U'_2(r')$ can be expressed as a sum of Gaussian sub-interactions [Eq. (14)], yielding the following expression for the local density

$$\rho_I(R) = \frac{4\pi\rho}{V_R} \left\{ \sum_i \int_{\bar{R}_{i-1}}^{\bar{R}_i} r'^2 \exp \left[-\beta \sum_i \epsilon_i N(\mu_i, \sigma_i^2) \right] dr' \right\}, \quad (\text{A4})$$

which corresponds to Eq. (15).

APPENDIX B: METHANOL-ACETONITRILE MIXTURE

Figure 19 compares the RDFs from the atomistic simulation with that from the CG simulations using the MS-CG potentials displayed in Figs. 17(a)–17(c). Remarkably, the radial distributions obtained from these distinct atomistic and CG simulations exhibit close agreement, underscoring the accuracy of the MS-CG model as a faithful representation of the multi-component liquid mixture. This reaffirms the suitability of these MS-CG interactions as a reliable reference for the parametrization of the Gaussian representation.

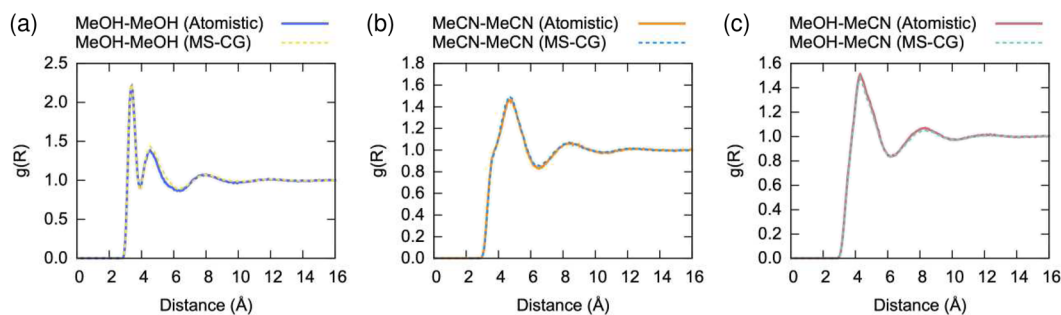


FIG. 19. Intermolecular RDF $g(R)$ acquired from the methanol-acetonitrile mixture: (a) Methanol–methanol pair. (b) Acetonitrile–acetonitrile pair. (c) Methanol–acetonitrile pair. Note that the atomistic RDFs are represented by solid lines, whereas the MS-CG RDFs are indicated by dashed lines.

REFERENCES

- ¹F. Müller-Plathe, *ChemPhysChem* **3**(9), 754–769 (2002).
- ²H. A. Scheraga, M. Khalili, and A. Liwo, *Annu. Rev. Phys. Chem.* **58**, 57–83 (2007).
- ³G. A. Voth, *Coarse-Graining of Condensed Phase and Biomolecular Systems* (CRC Press, 2008).
- ⁴C. Peter and K. Kremer, *Soft Matter* **5**(22), 4357–4366 (2009).
- ⁵T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, *Phys. Chem. Chem. Phys.* **11**(12), 1869–1892 (2009).
- ⁶S. Riniker and W. F. van Gunsteren, *J. Chem. Phys.* **134**(8), 084110 (2011).
- ⁷W. G. Noid, *J. Chem. Phys.* **139**(9), 090901 (2013).
- ⁸M. G. Saunders and G. A. Voth, *Annu. Rev. Biophys.* **42**, 73–93 (2013).
- ⁹A. J. Pak and G. A. Voth, *Curr. Opin. Struct. Biol.* **52**, 119–126 (2018).
- ¹⁰J. Jin, A. J. Pak, A. E. Durumeric, T. D. Loose, and G. A. Voth, *J. Chem. Theory Comput.* **18**(10), 5759–5791 (2022).
- ¹¹W. Noid, *J. Phys. Chem. B* **127**(19), 4174–4207 (2023).
- ¹²J. M. Grime and G. A. Voth, *Biophys. J.* **103**(8), 1774–1783 (2012).
- ¹³J. M. Grime, J. F. Dama, B. K. Ganser-Pornillos, C. L. Woodward, G. J. Jensen, M. Yeager, and G. A. Voth, *Nat. Commun.* **7**(1), 11568 (2016).
- ¹⁴A. J. Pak, J. M. Grime, P. Sengupta, A. K. Chen, A. E. Durumeric, A. Srivastava, M. Yeager, J. A. Briggs, J. Lippincott-Schwartz, and G. A. Voth, *Proc. Natl. Acad. Sci. U. S. A.* **114**(47), E10056–E10065 (2017).
- ¹⁵A. J. Pak, J. M. Grime, A. Yu, and G. A. Voth, *J. Am. Chem. Soc.* **141**(26), 10214–10224 (2019).
- ¹⁶A. Yu, K. A. Skorupka, A. J. Pak, B. K. Ganser-Pornillos, O. Pornillos, and G. A. Voth, *Nat. Commun.* **11**(1), 1307 (2020).
- ¹⁷A. Yu, E. M. Lee, J. Jin, and G. A. Voth, *Sci. Adv.* **6**(38), eabc6465 (2020).
- ¹⁸M. Gupta, A. J. Pak, and G. A. Voth, *Sci. Adv.* **9**(1), eadd7434 (2023).
- ¹⁹W. G. Noid, *Methods Mol. Biol.* **924**, 487–531 (2013).
- ²⁰M. S. Shell, in *Advances in Chemical Physics*, edited by S. A. Rice and A. R. Kinner (John Wiley & Sons, Inc., 2016), Vol. 161, pp. 395–441.
- ²¹J. W. Wagner, J. F. Dama, A. E. Durumeric, and G. A. Voth, *J. Chem. Phys.* **145**(4), 044108 (2016).
- ²²N. J. H. Dunn, T. T. Foley, and W. G. Noid, *Acc. Chem. Res.* **49**(12), 2832–2840 (2016).
- ²³T. D. Potter, J. Tasche, and M. R. Wilson, *Phys. Chem. Chem. Phys.* **21**(4), 1912–1927 (2019).
- ²⁴V. Krishna, W. G. Noid, and G. A. Voth, *J. Chem. Phys.* **131**(2), 024103 (2009).
- ²⁵L. Lu and G. A. Voth, *J. Chem. Phys.* **134**(22), 224107 (2011).
- ²⁶J. Jin, A. J. Pak, and G. A. Voth, *J. Phys. Chem. Lett.* **10**(16), 4549–4557 (2019).
- ²⁷N. J. H. Dunn and W. G. Noid, *J. Chem. Phys.* **143**(24), 243148 (2015).
- ²⁸M. R. DeLyser and W. G. Noid, *J. Chem. Phys.* **147**(13), 134111 (2017).
- ²⁹J. Jin, A. Yu, and G. A. Voth, *J. Chem. Theory Comput.* **16**(11), 6823–6842 (2020).
- ³⁰K. Shen, N. Sherck, M. Nguyen, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, G. H. Fredrickson, and M. S. Shell, *J. Chem. Phys.* **153**(15), 154116 (2020).
- ³¹S. Izvekov and G. A. Voth, *J. Phys. Chem. B* **109**(7), 2469–2473 (2005).
- ³²W. G. Noid, J.-W. Chu, G. S. Ayton, and G. A. Voth, *J. Phys. Chem. B* **111**(16), 4116–4127 (2007).
- ³³W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, *J. Chem. Phys.* **128**(24), 244114 (2008).
- ³⁴T. T. Foley, M. S. Shell, and W. G. Noid, *J. Chem. Phys.* **143**(24), 243104 (2015).
- ³⁵P. A. Golubkov, J. C. Wu, and P. Ren, *Phys. Chem. Chem. Phys.* **10**(15), 2050–2057 (2008).
- ³⁶P. Carbone, H. A. K. Varzaneh, X. Chen, and F. Müller-Plathe, *J. Chem. Phys.* **128**(6), 064904 (2008).
- ³⁷J. W. Mullinax and W. G. Noid, *J. Chem. Phys.* **131**(10), 104110 (2009).
- ³⁸E. Brini and N. Van der Vegt, *J. Chem. Phys.* **137**(15), 154113 (2012).
- ³⁹C. Dalgicdir, O. Sensoy, C. Peter, and M. Sayar, *J. Chem. Phys.* **139**(23), 234115 (2013).
- ⁴⁰P. Kar, S. M. Gopal, Y.-M. Cheng, A. Predeus, and M. Feig, *J. Chem. Theory Comput.* **9**(8), 3769–3788 (2013).
- ⁴¹P. Kar and M. Feig, *Advances in Protein Chemistry and Structural Biology* (Elsevier, 2014), Vol. 96, pp. 143–180.
- ⁴²D. D. Hsu, W. Xia, S. G. Arturo, and S. Ketten, *Macromolecules* **48**(9), 3057–3068 (2015).
- ⁴³V. Agrawal, P. Peralta, Y. Li, and J. Oswald, *J. Chem. Phys.* **145**(10), 104903 (2016).
- ⁴⁴N. J. H. Dunn and W. G. Noid, *J. Chem. Phys.* **144**(20), 204124 (2016).
- ⁴⁵J. F. Dama, J. Jin, and G. A. Voth, *J. Chem. Theory Comput.* **13**(3), 1010–1022 (2017).
- ⁴⁶T. Sanyal and M. S. Shell, *J. Phys. Chem. B* **122**(21), 5678–5693 (2018).
- ⁴⁷J. Jin and G. A. Voth, *J. Chem. Theory Comput.* **14**(4), 2180–2197 (2018).
- ⁴⁸D. Rosenberger, T. Sanyal, M. S. Shell, and N. F. van der Vegt, *J. Chem. Theory Comput.* **15**(5), 2881–2895 (2019).
- ⁴⁹K. H. Kanekal, J. F. Rudzinski, and T. Bereau, *J. Chem. Phys.* **157**(10), 104102 (2022).
- ⁵⁰E. Pretti and M. S. Shell, *J. Chem. Phys.* **155**(9), 094102 (2021).
- ⁵¹S. Izvekov and G. A. Voth, *J. Chem. Phys.* **123**(13), 134105 (2005).
- ⁵²W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, *J. Chem. Phys.* **128**(24), 244115 (2008).
- ⁵³L. Lu, S. Izvekov, A. Das, H. C. Andersen, and G. A. Voth, *J. Chem. Theory Comput.* **6**(3), 954–965 (2010).
- ⁵⁴M. S. Shell, *J. Chem. Phys.* **129**(14), 144108 (2008).
- ⁵⁵A. Chaimovich and M. S. Shell, *Phys. Rev. E* **81**(6), 060104 (2010).
- ⁵⁶A. Chaimovich and M. S. Shell, *J. Chem. Phys.* **134**(9), 094112 (2011).
- ⁵⁷R. W. Zwanzig, *J. Chem. Phys.* **22**(8), 1420–1426 (1954).
- ⁵⁸J. A. Barker and D. Henderson, *J. Chem. Phys.* **47**(8), 2856–2861 (1967).
- ⁵⁹J. A. Barker and D. Henderson, *J. Chem. Phys.* **47**(11), 4714–4721 (1967).

- ⁶⁰J. D. Weeks, D. Chandler, and H. C. Andersen, *J. Chem. Phys.* **54**(12), 5237–5247 (1971).
- ⁶¹H. C. Andersen, J. D. Weeks, and D. Chandler, *Phys. Rev. A* **4**(4), 1597 (1971).
- ⁶²J. D. Weeks, D. Chandler, and H. C. Andersen, *J. Chem. Phys.* **55**(11), 5422–5423 (1971).
- ⁶³J. Jin, K. S. Schweizer, and G. A. Voth, *J. Chem. Phys.* **158**(3), 034103 (2023).
- ⁶⁴J. Jin, K. S. Schweizer, and G. A. Voth, *J. Chem. Phys.* **158**(3), 034104 (2023).
- ⁶⁵J. Jin, E. K. Lee, and G. A. Voth, *J. Chem. Phys.* **159**(3), 164102 (2023).
- ⁶⁶Y. Rosenfeld, *Phys. Rev. Lett.* **63**(9), 980 (1989).
- ⁶⁷K. S. Schweizer, *J. Chem. Phys.* **91**(9), 5802–5821 (1989).
- ⁶⁸E. Kierlik and M. Rosinberg, *Phys. Rev. A* **42**(6), 3382 (1990).
- ⁶⁹J. F. Lutsko, *J. Chem. Phys.* **127**(5), 054701 (2007).
- ⁷⁰I. Pagonabarraga and D. Frenkel, *J. Chem. Phys.* **115**(11), 5015–5026 (2001).
- ⁷¹P. Warren, *Phys. Rev. E* **68**(6), 066702 (2003).
- ⁷²G. V. Efimov and G. Ganbold, *Phys. Status Solidi B* **168**, 165 (1991).
- ⁷³G. V. Efimov and E. A. Nogovitsin, *Physica A* **234**(1–2), 506–522 (1996).
- ⁷⁴S. A. Baeurle, R. Martoňák, and M. Parrinello, *J. Chem. Phys.* **117**(7), 3027–3039 (2002).
- ⁷⁵S. A. Baeurle, *Phys. Rev. Lett.* **89**(8), 080602 (2002).
- ⁷⁶S. A. Baeurle, G. V. Efimov, and E. A. Nogovitsin, *J. Chem. Phys.* **124**(22), 224110 (2006).
- ⁷⁷F. H. Stillinger, *J. Chem. Phys.* **65**(10), 3968–3974 (1976).
- ⁷⁸H. Löwen, *J. Phys.: Condens. Matter* **14**(46), 11897 (2002).
- ⁷⁹G. Rickayzen and D. Heyes, *J. Chem. Phys.* **138**(8), 084509 (2013).
- ⁸⁰F. H. Stillinger and T. A. Weber, *J. Chem. Phys.* **68**(8), 3837–3844 (1978).
- ⁸¹A. Louis, P. Bolhuis, and J. Hansen, *Phys. Rev. E* **62**(6), 7961 (2000).
- ⁸²A. Lang, C. Likos, M. Watzlawek, and H. Löwen, *J. Phys.: Condens. Matter* **12**(24), 5087 (2000).
- ⁸³A. Archer and R. Evans, *Phys. Rev. E* **64**(4), 041501 (2001).
- ⁸⁴S. Prestipino, F. Saija, and P. V. Giaquinta, *Phys. Rev. E* **71**(5), 050102 (2005).
- ⁸⁵B. J. Berne and P. Pechukas, *J. Chem. Phys.* **56**(8), 4213 (1972).
- ⁸⁶G. Milano, S. Goudeau, and F. Müller-Plathe, *J. Polym. Sci., Part B: Polym. Phys.* **43**(8), 871–885 (2005).
- ⁸⁷N. Basdevant, D. Borgis, and T. Ha-Duong, *J. Phys. Chem. B* **111**(31), 9390–9399 (2007).
- ⁸⁸S. John and G. Csányi, *J. Phys. Chem. B* **121**(48), 10934–10949 (2017).
- ⁸⁹A. B. de Oliveira, G. Franzese, P. A. Netz, and M. C. Barbosa, *J. Chem. Phys.* **128**(6), 064901 (2008).
- ⁹⁰A. Chaimovich and M. S. Shell, *Phys. Chem. Chem. Phys.* **11**(12), 1901–1915 (2009).
- ⁹¹M. E. Johnson, T. Head-Gordon, and A. A. Louis, *J. Chem. Phys.* **126**(14), 144509 (2007).
- ⁹²N. Sherck, K. Shen, M. Nguyen, B. Yoo, S. Kohler, J. C. Speros, K. T. Delaney, M. S. Shell, and G. H. Fredrickson, *ACS Macro Lett.* **10**(5), 576–583 (2021).
- ⁹³N. Basdevant, D. Borgis, and T. Ha-Duong, *J. Chem. Theory Comput.* **9**(1), 803–813 (2013).
- ⁹⁴R. D. Hills, *Protein Dynamics*, Methods in Molecular Biology Vol. 1084 (Humana Press, 2014), pp. 123–140.
- ⁹⁵J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Elsevier, 1990).
- ⁹⁶J. V. Sengers, R. Kayser, C. Peters, and H. White, *Equations of State for Fluids and Fluid Mixtures* (Elsevier, 2000).
- ⁹⁷Á. Mulero, *Theory and Simulation of Hard-Sphere Fluids and Related Systems* (Springer, 2008).
- ⁹⁸A. Santos, *A Concise Course on the Theory of Classical Liquids* (Springer, Berlin, 2016).
- ⁹⁹N. P. Bailey, U. R. Pedersen, N. Gnan, T. B. Schröder, and J. C. Dyre, *J. Chem. Phys.* **129**(18), 184507 (2008).
- ¹⁰⁰N. P. Bailey, U. R. Pedersen, N. Gnan, T. B. Schröder, and J. C. Dyre, *J. Chem. Phys.* **129**(18), 184508 (2008).
- ¹⁰¹U. R. Pedersen, T. B. Schröder, and J. C. Dyre, *Phys. Rev. Lett.* **105**(15), 157801 (2010).
- ¹⁰²R. Evans, *Fundamentals of Inhomogeneous Fluids* (Marcel Dekker, New York, 1992), Vol. 1, pp. 85–176.
- ¹⁰³J. W. Wagner, T. Dannenhoffer-Lafage, J. Jin, and G. A. Voth, *J. Chem. Phys.* **147**(4), 044113 (2017).
- ¹⁰⁴M. R. DeLyser and W. G. Noid, *J. Chem. Phys.* **151**(22), 224106 (2019).
- ¹⁰⁵M. R. DeLyser and W. G. Noid, *J. Chem. Phys.* **156**(3), 034106 (2022).
- ¹⁰⁶J. Jin, Y. Han, and G. A. Voth, *J. Chem. Theory Comput.* **14**(12), 6159–6174 (2018).
- ¹⁰⁷J. Jin and G. A. Voth, *J. Phys. Chem. Lett.* **14**(6), 1354–1362 (2023).
- ¹⁰⁸J. F. Dama, J. Jin, and G. A. Voth, *J. Chem. Theory Comput.* **14**(4), 2288 (2018).
- ¹⁰⁹T. Sanyal and M. S. Shell, *J. Chem. Phys.* **145**(3), 034109 (2016).
- ¹¹⁰D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, Oxford, UK, 1987).
- ¹¹¹T. L. Hill, *An Introduction to Statistical Thermodynamics* (Courier Corporation, 1986).
- ¹¹²T. Hamelryck, M. Borg, M. Paluszewski, J. Paulsen, J. Frellsen, C. Andreetta, W. Boomsma, S. Bottaro, and J. Ferkinghoff-Borg, *PLoS One* **5**(11), e13714 (2010).
- ¹¹³E. Kalligiannaki, V. Harmandaris, M. A. Katsoulakis, and P. Plecháč, *J. Chem. Phys.* **143**(8), 084105 (2015).
- ¹¹⁴N. N. Bogoliubov, *Problems of the Dynamical Theory in Statistical Physics* (Moscow-Leningrad, 1946).
- ¹¹⁵J. Yvon, *La théorie statistique des fluides et l'équation d'état* (Hermann and cie, 1935).
- ¹¹⁶J. G. Kirkwood, *J. Chem. Phys.* **14**(3), 180–201 (1946).
- ¹¹⁷M. Born and H. Green, *Proc. R. Soc. London, Ser. A* **188**(1012), 10–18 (1946).
- ¹¹⁸N. Bogoliubov, *Zh. Eksp. Teor. Fiz.* **16**(8), 691–702 (1946).
- ¹¹⁹J. G. Kirkwood, F. P. Buff, and M. S. Green, *J. Chem. Phys.* **17**(10), 988–994 (1949).
- ¹²⁰J. W. Mullinax and W. G. Noid, *Phys. Rev. Lett.* **103**(19), 198104 (2009).
- ¹²¹J. W. Mullinax and W. G. Noid, *J. Phys. Chem. C* **114**(12), 5661–5674 (2009).
- ¹²²J. W. Mullinax and W. G. Noid, *J. Chem. Phys.* **133**(12), 124107 (2010).
- ¹²³J. F. Rudzinski and W. G. Noid, *J. Phys. Chem. B* **118**(28), 8295–8312 (2014).
- ¹²⁴J. F. Rudzinski and W. G. Noid, *Eur. Phys. J.: Spec. Top.* **224**(12), 2193–2216 (2015).
- ¹²⁵A. Das and H. C. Andersen, *J. Chem. Phys.* **136**(19), 194114 (2012).
- ¹²⁶D. Hankins, J. Moskowitz, and F. Stillinger, *J. Chem. Phys.* **53**(12), 4544–4554 (1970).
- ¹²⁷J. Jin, Y. Han, A. J. Pak, and G. A. Voth, *J. Chem. Phys.* **154**(4), 044104 (2021).
- ¹²⁸J. Jin, A. J. Pak, Y. Han, and G. A. Voth, *J. Chem. Phys.* **154**(4), 044105 (2021).
- ¹²⁹S. Eliezer, *Fundamentals of Equations of State* (Allied Publishers, 2005).
- ¹³⁰L. Ornstein and F. Zernike, *Proc. K. Ned. Akad. Wet.* **19**(2), 1312–1315 (1914).
- ¹³¹T. Morita, *Prog. Theor. Phys.* **20**(6), 920–938 (1958).
- ¹³²J. A. Barker and D. Henderson, *Rev. Mod. Phys.* **48**(4), 587 (1976).
- ¹³³A. Vompe and G. Martynov, *J. Chem. Phys.* **100**(7), 5249–5258 (1994).
- ¹³⁴R. J. Speedy, *J. Chem. Soc., Faraday Trans. 2* **76**, 693–703 (1980).
- ¹³⁵J. Henderson, *Mol. Phys.* **48**(2), 389–400 (1983).
- ¹³⁶M. Llano-Restrepo and W. G. Chapman, *J. Chem. Phys.* **97**(3), 2046–2054 (1992).
- ¹³⁷M. Llano-Restrepo and W. G. Chapman, *J. Chem. Phys.* **100**(7), 5139–5148 (1994).
- ¹³⁸W. G. Chapman, *J. Chem. Phys.* **93**(6), 4299–4304 (1990).
- ¹³⁹E. Meeron and A. J. Siegert, *J. Chem. Phys.* **48**(7), 3139–3155 (1968).
- ¹⁴⁰E. Grundke and D. Henderson, *Mol. Phys.* **24**(2), 269–281 (1972).
- ¹⁴¹C. Likos, M. Watzlawek, and H. Löwen, *Phys. Rev. E* **58**(3), 3135 (1998).
- ¹⁴²P. Español, *Phys. Rev. E* **52**(2), 1734 (1995).
- ¹⁴³P. Español and P. Warren, *Europhys. Lett.* **30**(4), 191 (1995).
- ¹⁴⁴R. D. Groot and P. B. Warren, *J. Chem. Phys.* **107**(11), 4423–4435 (1997).
- ¹⁴⁵Y. Han, J. Jin, and G. A. Voth, *J. Chem. Phys.* **154**(8), 084122 (2021).
- ¹⁴⁶L. Lu and G. A. Voth, *J. Phys. Chem. B* **113**(5), 1501–1510 (2009).
- ¹⁴⁷See <https://github.com/uchicago-voth/MSCG-release> for Multiscale Coarse-Graining Force-Matching (MSCGFM) version 1.7 (2018).
- ¹⁴⁸Y. Peng, A. J. Pak, A. E. P. Durumeric, P. G. Sahrman, S. Mani, J. Jin, T. D. Loose, J. Beiter, and G. A. Voth, *J. Phys. Chem. B* **127**(40), 8537–8550 (2023).
- ¹⁴⁹T. Dannenhoffer-Lafage, J. W. Wagner, A. E. P. Durumeric, and G. A. Voth, *J. Chem. Phys.* **151**(13), 134115 (2019).

- ¹⁵⁰W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **118**(45), 11225–11236 (1996).
- ¹⁵¹G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, *J. Phys. Chem. B* **105**(28), 6474–6487 (2001).
- ¹⁵²S. Plimpton, *J. Comput. Phys.* **117**, 1–19 (1995).
- ¹⁵³W. M. Brown, P. Wang, S. J. Plimpton, and A. N. Tharrington, *Comput. Phys. Commun.* **182**(4), 898–911 (2011).
- ¹⁵⁴W. M. Brown, A. Kohlmeyer, S. J. Plimpton, and A. N. Tharrington, *Comput. Phys. Commun.* **183**(3), 449–459 (2012).
- ¹⁵⁵L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, *J. Comput. Chem.* **30**(13), 2157–2164 (2009).
- ¹⁵⁶W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**(1), 33–38 (1996).
- ¹⁵⁷S. Nosé, *J. Chem. Phys.* **81**(1), 511–519 (1984).
- ¹⁵⁸W. G. Hoover, *Phys. Rev. A* **31**(3), 1695 (1985).
- ¹⁵⁹S. Izvekov, P. W. Chung, and B. M. Rice, *J. Chem. Phys.* **133**(6), 064109 (2010).
- ¹⁶⁰S. Izvekov, P. W. Chung, and B. M. Rice, *J. Chem. Phys.* **135**(4), 044112 (2011).
- ¹⁶¹K. M. Lebold and W. G. Noid, *J. Chem. Phys.* **150**(1), 014104 (2019).
- ¹⁶²M. Dinpajoo and M. G. Guenza, *J. Phys. Chem. B* **122**(13), 3426–3440 (2017).
- ¹⁶³MATLAB, version 9.7.0 (R2019b), The MathWorks Inc., Natick, MA (2019).
- ¹⁶⁴J. J. Moré and D. C. Sorensen, *SIAM J. Sci. Stat. Comput.* **4**(3), 553–572 (1983).
- ¹⁶⁵E. Wilhelm, *J. Chem. Phys.* **58**(9), 3558–3560 (1973).
- ¹⁶⁶T. B. Schröder, N. P. Bailey, U. R. Pedersen, N. Gnan, and J. C. Dyre, *J. Chem. Phys.* **131**(23), 234503 (2009).
- ¹⁶⁷N. Gnan, T. B. Schröder, U. R. Pedersen, N. P. Bailey, and J. C. Dyre, *J. Chem. Phys.* **131**(23), 234504 (2009).
- ¹⁶⁸T. B. Schröder, N. Gnan, U. R. Pedersen, N. P. Bailey, and J. C. Dyre, *J. Chem. Phys.* **134**(16), 164505 (2011).
- ¹⁶⁹W. Kob and H. C. Andersen, *Phys. Rev. E* **51**(5), 4626 (1995).
- ¹⁷⁰C.-C. Fu, P. M. Kulkarni, M. S. Shell, and L. G. Leal, *J. Chem. Phys.* **137**(16), 164106 (2012).
- ¹⁷¹H. Wang, F. H. Stillinger, and S. Torquato, *J. Chem. Phys.* **153**(12) (2020).
- ¹⁷²R. Potestio, *JUnQ* **3**, 13–15 (2013).
- ¹⁷³O. Banerjee, L. E. Ghaoui, A. d’Aspremont, and G. Natsoulis, paper presented at the Proceedings of the 23rd International Conference on Machine Learning, 2006.
- ¹⁷⁴A. Das, L. Lu, H. C. Andersen, and G. A. Voth, *J. Chem. Phys.* **136**(19), 194115 (2012).
- ¹⁷⁵A. Das and H. C. Andersen, *J. Chem. Phys.* **132**(16), 164106 (2010).
- ¹⁷⁶H.-J. Qian, P. Carbone, X. Chen, H. A. Karimi-Varzaneh, C. C. Liew, and F. Müller-Plathe, *Macromolecules* **41**(24), 9919–9929 (2008).
- ¹⁷⁷J. Gay and B. Berne, *J. Chem. Phys.* **74**(6), 3316–3319 (1981).
- ¹⁷⁸P. A. Golubkov and P. Ren, *J. Chem. Phys.* **125**(6), 064103 (2006).
- ¹⁷⁹D. J. Cleaver, C. M. Care, M. P. Allen, and M. P. Neal, *Phys. Rev. E* **54**(1), 559 (1996).
- ¹⁸⁰D. Antypov and D. J. Cleaver, *J. Chem. Phys.* **120**(21), 10307–10316 (2004).
- ¹⁸¹M. E. Sharp, F. X. Vázquez, J. W. Wagner, T. Dannenhöffer-Lafage, and G. A. Voth, *J. Chem. Theory Comput.* **15**(5), 3306–3315 (2019).
- ¹⁸²R. A. Marcus, *Annu. Rev. Phys. Chem.* **15**(1), 155–196 (1964).
- ¹⁸³H. Lorentz, *Ann. Phys.* **248**, 127–136 (1881).
- ¹⁸⁴D. Berthelot, *Hebd. Séances Acad. Sci.* **126**, 1703 (1898).
- ¹⁸⁵D. Dimitrelis and J. M. Prausnitz, *Fluid Phase Equilib.* **31**(1), 1–21 (1986).
- ¹⁸⁶D. Ben-Amotz and D. R. Herschbach, *J. Phys. Chem.* **94**(3), 1038–1047 (1990).
- ¹⁸⁷I. Nezbeda and K. Aim, *Fluid Phase Equilib.* **17**(1), 1–18 (1984).
- ¹⁸⁸J. Kolafa and I. Nezbeda, *Fluid Phase Equilib.* **100**, 1–34 (1994).
- ¹⁸⁹H. C. Andersen, *J. Chem. Phys.* **72**(4), 2384–2393 (1980).
- ¹⁹⁰P. Vanya and J. A. Elliott, *Phys. Rev. E* **102**(1), 013312 (2020).