Integral equation theory based direct and accelerated systematic coarse-graining approaches

S. Y. Mashayak,[1,2] Linling Miao,[2,3] and N. R. Aluru[1,2]
[1]*Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[2]*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[3]*Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

(Received 22 December 2017; accepted 16 May 2018; published online 5 June 2018)

Coarse-grained (CG) molecular dynamics (MD) simulations have become popular for investigating systems on multiple length and time scales ranging from atomistic to mesoscales. In CGMD, several atoms are mapped onto a single CG bead and the effective interactions between CG beads are determined. Iterative coarse-graining methods, such as iterative Boltzmann inversion (IBI), are computationally expensive and can have convergence issues. In this paper, we present a direct and computationally efficient theoretical procedure for coarse-graining based on the Ornstein-Zernike (OZ) and hypernetted chain (HNC) integral equation theory. We demonstrate the OZ-HNC-based CG method by coarse-graining a bulk water system, a water-methanol mixture system, and an electrolyte system. We show that the accuracy of the CG potentials obtained from the OZ-HNC-based coarse-graining is comparable to iterative systematic coarse-graining methods. Furthermore, we show that the CG potentials from OZ-HNC can be used to reduce the number of iterations and hence the computational cost of the iterative systematic coarse-graining approaches, like IBI and relative entropy minimization. *Published by AIP Publishing.* https://doi.org/10.1063/1.5020321

I. INTRODUCTION

In recent years, coarse-grained (CG) simulations have become an important tool for investigating systems on larger time and length scales.[1–6] In coarse-grained simulations, several atoms are grouped into a single CG site and effective interaction potentials between CG sites are derived. Such coarse-graining leads to a significant speed-up in the molecular dynamics (MD) simulations due to (i) a reduced number of degrees of freedom and hence the number of interactions to compute, (ii) smoother interaction potentials among CG particles that enable larger time steps, and most importantly, (iii) a speedup in the intrinsic dynamics of the system which leads to faster diffusion and thus shorter equilibration times.[7] Determining accurate interaction potentials between CG sites is the most challenging part of the coarse-grained simulations.

Systematic coarse-graining offers a bottom-up approach to optimize CG models by systematically linking a low-resolution CG system to a reference high resolution all-atom (AA) system. In systematic coarse-graining, there are three main steps to construct a CG model. First, high quality AA MD simulations of the target system are performed to use as a reference. Then in the second step, a mapping scheme is defined to decide which atoms are collected into one CG bead and where the center of the bead is placed in relation to the atomistic description. Finally, in the third step, effective interactions between CG beads are optimized. In this work, we focus on the methods to determine accurate CG interactions.

There exist several systematic coarse-graining techniques of varying complexities and accuracies for optimizing CG interactions. These range from structure-based methods, such as the Boltzmann inversion, iterative Boltzmann inversion (IBI),[8] inverse Monte Carlo (IMC),[9] and relative entropy,[10] to force-based methods like force matching (FM).[11,12] Structure-based methods target (multi-body) distribution functions and use the relation between the distribution functions and the many-body potential of mean force (PMF)[8,12] to derive effective CG interactions. The force-matching method employs a variational principle to optimize the CG potential such that the many-body PMF field obtained from the reference fully atomistic simulations is reproduced as closely as possible. Moreover, Rudzinski and Noid[13] have shown that the FM method is related to the generalized-Yvon-Born-Green (g-YBG) integral equation (IE) framework. The g-YBG equation provides a relation between a given potential and the equilibrium two- and three-body correlation functions. It has been shown that, in the context of CG modeling, the g-YBG equation provides a framework to determine a variationally optimal CG potential to approximate the many-body PMF from a given equilibrium correlation.[13] Therefore, the g-YBG equation framework allows us to do "force-matching" without forces. Detailed descriptions about the various aspects and methods for determining CG potentials can be found in Refs. 1, 3, and 12–15.

Most of the coarse-graining approaches such as iterative Boltzmann inversion (IBI),[8] inverse Monte Carlo (IMC),[9] and relative entropy[10] are numerical and iterative methods. In these iterative methods, a trial function is provided as a starting CG

Algorithm 1. Workflow of an iterative coarse-graining method.

1: Initialize the CG potential u^0_{CG} and set the relaxation factor, χ, and the tolerance, *tol*
2: **repeat**
3: Perform canonical ensemble sampling of the CG system using the CG potential u^k_{CG}
4: Compute the ensemble averages and distribution functions from CG sampling, such as $g_{\mathrm{CG}}(r)$, $\left\langle \frac{\partial^2 U_{\mathrm{CG}}}{\partial\lambda_i \partial\lambda_j} \right\rangle_{\mathrm{CG}}$, etc.
5: Determine u^{k+1}_{CG} [Eq. (1) for IBI and Eq. (2) for the relative entropy]
6: **until** $|u^{k+1}_{\mathrm{CG}} - u^k_{\mathrm{CG}}| < tol$

potential which is then updated and optimized iteratively until a given error metric is minimized. For example, in IBI, the CG potential is refined iteratively as

$$u^{k+1}_{\mathrm{CG}}(r) = u^k_{\mathrm{CG}}(r) + k_{\mathrm{B}}T \ln \frac{g^k_{\mathrm{CG}}(r)}{g_{\mathrm{tgt}}(r)}, \tag{1}$$

where k is the iteration index, $u_{\mathrm{CG}}(r)$ is the CG pair potential, k_{B} is the Boltzmann constant, T is the temperature, $g^k_{\mathrm{CG}}(r)$ is the radial distribution function (RDF) determined using the CG potential $u^k_{\mathrm{CG}}(r)$ at the iteration step k, and $g_{\mathrm{tgt}}(r)$ is the target RDF. In relative entropy-based coarse-graining, the CG potential parameters, $\boldsymbol{\lambda}$, are refined iteratively as[14,16]

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \chi \mathbf{H}^{-1} \cdot \nabla_{\lambda} S_{\mathrm{rel}}, \tag{2}$$

where S_{rel} is the relative entropy, $\nabla_{\lambda} S_{\mathrm{rel}}$ is the vector of the first derivatives of S_{rel} with respect to $\boldsymbol{\lambda}$, $\mathbf{H}$ is the Hessian matrix of S_{rel}, and $\chi \in (0...1)$ is the relaxation parameter that can be adjusted to ensure convergence.

Therefore, in iterative coarse-graining methods, each iterative step requires canonical sampling of the coarse-grained system to estimate distribution functions and ensemble averages. Algorithm 1 shows a typical workflow of the iterative coarse-graining methods. In the iterative workflow, the canonical sampling, Step 3, and ensemble averages, Step 4, are the most computationally expensive parts of the iterative coarse-graining methods. The canonical sampling at each iteration can be done using either molecular dynamics, stochastic dynamics (SD), or Monte Carlo (MC) techniques. To ensure the convergence of the iterative coarse-graining schemes, statistically reliable CG ensemble configurations are required. For example, to ensure the convergence of IBI [Eq. (1)], the RDFs, $g_{\mathrm{CG}}(r)$, obtained from the CG ensemble average should converge to unity for the large distances, i.e., $r \to \infty$. To obtain statistically reliable CG ensemble averages, such as $g_{\mathrm{CG}}(r)$, iterative coarse-graining methods need to use a sufficiently large CG system and perform long CG simulations at each iteration step. The cost associated with the coarse-grained simulation and analysis of the CG ensemble configurations increases with the size of the system, i.e., number of particles, and the length of the simulation, i.e., number of time steps. Therefore, sampling of the coarse-grained systems and analysis of the CG ensemble configurations at each iteration step are computationally expensive and dominate the overall cost of the iterative coarse-graining approaches. Moreover, special care must be taken, such as smoothing of the potential updates, to ensure the stability of the numerical iterative schemes.[15] Computational cost and convergence issues of the iterative coarse-graining methods become particularly prohibitive for the systems involving macromolecules and mixtures of several molecules that consist of several different CG beads. For example, for dilute solutions, such as aqueous electrolytes and biomolecular systems which have low solute concentrations in aqueous environment, the solute-solute RDFs converge very slowly and therefore the size of the CG system, the length of the CG simulations per iteration, and the total number of iterations to reach convergence become prohibitively expensive.[1,17] Furthermore, for the systems involving multiple CG beads, due to a mutual dependence of the several CG potentials, iterative methods, like IBI, have a convergence problem.[18] On the other hand, FM and g-YBG methods are non-iterative methods and, hence, computationally less demanding. But, the pair CG potentials obtained from the FM and g-YBG methods do not ensure the accuracy of the target radial distribution functions.[13]

In this work, we present a direct (non-iterative) and computationally efficient theoretical procedure for coarse-graining. The theoretical procedure is based on the Ornstein-Zernike (OZ) and hypernetted chain (HNC) closure in the integral equation (IE) theory. HNC closure provides a direct approach to estimate an effective pair potential from a target distribution function without performing atomistic simulations. Therefore, OZ-HNC-based coarse-graining (OZ-HNC CG) is a computationally efficient approach to obtain CG potentials. In Sec. II, we provide theoretical and numerical details about the OZ-HNC-based coarse-graining procedure. In Sec. III, we demonstrate the OZ-HNC-based coarse-graining procedure by determining CG potentials for the bulk SPC/E (extended simple point charge model) water, water-methanol mixture, and potassium chloride solution systems.

II. METHOD

The goal for the coarse-grained model is to reproduce the target radial distribution function (RDF) of the reference system. The radial distribution function, $g(r)$, where $r = |\boldsymbol{r} - \boldsymbol{r}_0|$, is a conditional probability density, relative to the ideal gas limit, of finding a particle at a distance r away from a given reference particle at $\boldsymbol{r}_0$.[19] In the integral equation theory, for a uniform and isotropic system, the relation between an effective pair potential, $u(r)$, and RDF, $g(r)$, is defined as[20]

$$g(r) = \exp(-\beta u(r) + h(r) - c(r) + b(r)), \tag{3}$$

where $\beta = \frac{1}{k_{\mathrm{B}}T}$, $h(r)$ is the total correlation function, $c(r)$ is the direct correlation function, and $b(r)$ is the bridge function.

The total correlation function, $h(r)$, is the difference between the radial distribution function and its random value of unity,

$$h(r) = g(r) - 1. \tag{4}$$

The total correlation, $h(r)$, can be defined as a sum of the direct correlation part, $c(r)$, and the indirect correlations due to the propagation of the interactions via surrounding particles. This relation between $h(r)$ and $c(r)$ is given by the Ornstein-Zernike (OZ) equation[21] as

$$h(r) = c(r) + \rho \int c(|\boldsymbol{r} - \boldsymbol{r}'|)h(r')d\boldsymbol{r}', \tag{5}$$

where ρ is the uniform density of the bulk system.

The exact expression for the bridge function, $b(r)$, in Eq. (3) is not known. There exist several approximations, known as "closure approximations," to solve Eq. (3), such as the Percus-Yevick (PY) closure, mean spherical approximation (MSA) closure, and hyper-netted chain (HNC) approximation.[20] Here, we use the HNC closure approximation which ignores the bridge function and the expression for $g(r)$ is given by

$$g(r) = \exp(-\beta u(r) + h(r) - c(r)). \tag{6}$$

The HNC closure approximation maps three-body and higher-order many-body contributions into an effective pair potential. Therefore, it provides a direct, i.e., non-iterative, and analytical route to estimate an effective CG potential from a pair distribution. Given a target $g(r)$ and the thermodynamic state, i.e., ρ and T, of the reference system, the effective CG potential between coarse-grained particles can be obtained by inverting the HNC closure, Eq. (6), as

$$u(r) = k_B T(h(r) - c(r) - \ln(g(r))). \tag{7}$$

We note that the g-YBG equation also provides an alternative framework to directly relate the equilibrium correlations to the effective CG potentials. While the OZ-HNC-based approach uses the OZ equation together with the HNC-closure to relate $g(r)$ and $u(r)$, the g-YBG equation uses a target 3-body correlation function to directly obtain $u(r)$. However, determining the pair CG potentials directly from the target 3-body correlations is numerically challenging and may require simplifications which do not ensure the accuracy of the pair correlations[13]

The OZ equation, Eq. (5), and HNC closure, Eq. (7), are in the form of a single fluid component. For a multicomponent fluid mixture, the indirect correlation term in the OZ equation consists of a sum over all the fluid components which can be written in the Fourier space as

$$\tilde{h}_{ij}(k) = \tilde{c}_{ij}(k) + \sum_l \rho_l \tilde{h}_{il}(k)\tilde{c}_{lj}(k), \tag{8}$$

where $\tilde{h}_{ij}(k)$ and $\tilde{c}_{ij}(k)$ are the Fourier transform of the total correlation function and direct correlation function between species i and j, respectively. Equation (8) can be represented in the matrix form[22] as

$$\tilde{\mathbf{H}} = \tilde{\mathbf{C}} + \tilde{\mathbf{C}} \cdot \mathbf{R} \cdot \tilde{\mathbf{H}}, \tag{9}$$

where $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{C}}$ are the matrices of total and direct correlation functions such that $\tilde{\mathbf{H}}(ij)$ and $\tilde{\mathbf{C}}(ij)$ correspond to the total and direct correlation between species i and j. Therefore, $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{C}}$ are symmetric matrices. In Eq. (9), $\mathbf{R}$ is a diagonal matrix with $R_{ii} = \rho_i$. To solve for direct correlation functions, we use the reference RDFs of all component pairs as an input and the solution of $\tilde{\mathbf{C}}$ is given as

$$\tilde{\mathbf{C}} = (\mathbf{I} + \tilde{\mathbf{H}} \cdot \mathbf{R})^{-1}\tilde{\mathbf{H}}, \tag{10}$$

where $\mathbf{I}$ is an identity matrix.

The numerical procedure to obtain CG potentials from Eq. (7) is as follows. For a target RDF, $g(r)$, which can be obtained from an all-atom simulation or experimental data, $h(r)$ is determined from Eq. (4). Then, the OZ equation [Eq. (5)] is solved to obtain $c(r)$. We follow the numerical procedure described in Ref. 23 to solve the OZ equation. Once $c(r)$ is determined, the effective CG potential between CG sites is determined from Eq. (7). Therefore, in the OZ-HNC coarse-graining method, the CG potentials are obtained directly from the target distribution, and it does not require iteratively simulating the CG systems and analysis of the CG ensemble configurations. We note that, for the short separation distances where $g(r) \to 0$, $u(r)$ values need to be extrapolated.

III. RESULTS

To illustrate functionality of the OZ-HNC coarse-graining method, we coarse-grain three systems: an SPC/E bulk water system, a system consisting of a methanol-water mixture, and a potassium chloride electrolyte system. For all the systems, we first obtain the reference RDFs by performing the all-atom molecular dynamic (AAMD) simulations. Then, we use the reference RDFs in the OZ-HNC-based CG method, i.e., Eq. (7), to directly obtain the CG potentials. Then the OZ-HNC-based CG potentials are used to perform coarse-grained molecular dynamic (CGMD) simulations. We, then, compute the RDFs from the CGMD simulations and compare them with the reference RDFs from AAMD. In addition to RDF, we also evaluate the accuracy of the thermodynamic properties, such as pressure and chemical potential, from CGMD. Furthermore, we also compare the CG potentials from the OZ-HNC-based coarse-graining with the CG potentials from relative entropy optimization, which is an iterative coarse-graining method. All the AAMD and CGMD simulations are performed in the canonical ensemble with the Nosé-Hoover thermostat using GROMACS.[24]

A. Bulk water

For coarse-graining the bulk water, we consider bulk SPC/E water at a thermodynamic state of 300 K temperature and 1 bar pressure. In the coarse-grained model, we represent one water molecule by one CG bead positioned at its center of mass (COM) such that the CG beads solely interact via an isotropic two-body potential. The target RDF is determined from the reference all-atom NVT ensemble consisting of 2000 water molecules in a cubic box of volume $3.912 \times 3.912 \times 3.912$ nm^3 which is obtained from an NPT ensemble simulation of bulk water at $T = 300$ K and $P = 1$ bar. Then the target

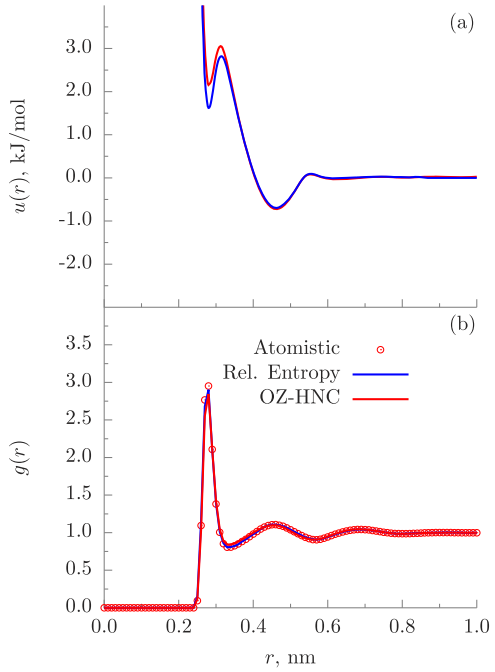

FIG. 1. Water-water CG potentials (a) and RDFs (b) for the bulk SPC/E water system. CG potential from the OZ-HNC CG procedure is close to the CG potential obtained from relative entropy minimization and can accurately predict the water-water COM RDF.

water-water RDF from AA simulations is used in OZ-HNC-based coarse-graining, i.e., Eq. (7), to obtain the water-water coarse-grained interaction.

Figure 1 shows the CG potential determined from HNC for bulk water. Figure 1 shows the water-water RDFs from the CGMD simulation along with the reference RDF from the AA simulation. We observe that the water-water RDF from the OZ-HNC-based CG potential compares well with the reference RDF from AA simulations. To have a better illustration of the differences between RDFs, especially at large distance r, we also compute the Kirkwood-Buff Integrals (KBIs)[25] as defined in the following equation:

$$G = 4\pi \int_0^\infty (g(r') - 1) r'^2 dr'. \tag{11}$$

Figure 2 compares the running KBIs, $G(r) = 4\pi \int_0^r (g(r') - 1)\, r'^2 dr'$, computed from the RDF of the reference AA simulation and the RDF from the OZ-HNC-based CG

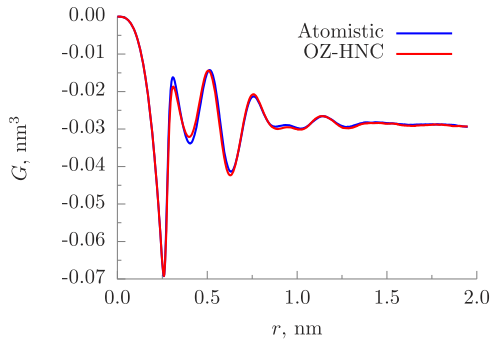

FIG. 2. Kirkwood-Buff integrals computed from the water-water RDFs for the bulk SPC/E water system. CG potentials from the OZ-HNC CG approach predict the KBI of reference atomistic simulations accurately.

approach. The KBIs approach to a plateau within half the simulation box size, which indicates that the system size we used here is large enough to have RDF converge to unity. We observe that the KBI from the OZ-HNC-based CG potential compares well with the reference KBI from AA simulations. Furthermore, the isothermal compressibility, κ_T, is related to the water-water KBI as

$$\kappa_T = \frac{1 + \rho G}{\rho k_B T}. \tag{12}$$

We find that the isothermal compressibilities, κ_T, from the KBI of the OZ-HNC-based CG potential (3.70×10^{-10} m^2/N) and the KBI of AA simulations (3.95×10^{-10} m^2/N) also compare well.

In Fig. 1, we also compare the OZ-HNC-based water-water CG potential to that from the relative entropy minimization obtained in Ref. 16. It can be seen that the CG potential from OZ-HNC is very similar to the optimized CG potential from relative entropy minimization. The small differences between the CG potentials from the OZ-HNC and relative entropy minimization could be attributed to the bridge function approximation in the HNC closure. Most importantly, as described in Sec. II, OZ-HNC-based coarse-graining is a direct approach and the cost of solving Eq. (7) is negligible, whereas relative entropy-based CG of bulk water (as described in Ref. 16) takes about 34 min to run one iteration on an Intel Xeon Processor E5-1630 with one thread. For each iteration, the water-water RDF is averaged from a 200 ps CGMD simulation, and the water-water CG potential is represented by a cubic B-spline function from 0 to the cutoff distance 0.9 nm with 48 spline knots.

B. Water-methanol mixture

Iterative coarse-graining methods, such as IBI, can have convergence problems for multicomponent systems.[1,18] IBI ignores the cross correlation terms, and hence, its convergence can be slower and may require relaxation [i.e., multiplying factor $\chi \in (0\ldots1)$] of the update function. On the other hand, IMC is more rigorous than IBI and accounts for the cross correlations, but it is computationally costly and requires large configurational sampling to achieve good accuracy. See Ref. 15 for detailed comparisons of IBI and IMC methods.

The OZ-HNC-based CG approach is rigorously derived from the statistical mechanics. As described in Sec. II, the HNC closure along with the OZ equation maps three-body and higher-order many-body contributions into an effective pair potential. Therefore, cross correlation terms are explicitly considered in determining CG potentials from OZ-HNC.

Here, to demonstrate robustness of the OZ-HNC-based coarse-graining for multicomponent systems, we consider an equimolar water-methanol mixture. In the CG model for the water-methanol mixtures, water and methanol molecules are represented by CG beads positioned at their COMs and the interactions between CG beads are modeled via isotropic two-body potentials. The reference water-water,

methanol-methanol, and methanol-water RDFs are generated from the AAMD simulations in an NVT ensemble of 2000 SPC/E water and 2000 methanol molecules described by the OPLS-AA force field[26] in a cubic box of length 5.727 nm at 300 K temperature. First, we use water-water, methanol-methanol, and methanol-water target RDFs from AAMD in the multicomponent version of the OZ equation, Eq. (10), to obtain the direct correlations between three different pairs. Then, we solve Eq. (7) for all three pairs to obtain corresponding CG potentials.

Figure 3 shows the CG potentials for the water-water, methanol-methanol, and methanol-water pairs determined from OZ-HNC. We then use the OZ-HNC-based CG potentials to perform molecular simulations of the CG water-methanol mixture system. Figure 3 shows that the RDFs from the CG simulations compare well with the reference RDFs from AA simulations. In Fig. 3, we also compare the OZ-HNC-based water-water, methanol-methanol, and methanol-water CG potentials to those from the relative entropy minimization obtained in Ref. 16. It can be seen that the CG potentials from OZ-HNC are very close to the optimized CG potentials obtained from the relative entropy minimization. Moreover, the computational cost of relative entropy-based CG for this mixture system is around 103 min for one iteration on Intel Xeon Processor E5-1630 with a single thread. For each iteration, the RDFs are averaged from a 500 ps CGMD simulation. The water-water CG potential is represented by a cubic B-spline function from 0 to 0.9 nm with 48 spline knots, while the water-methanol and methanol-methanol CG potentials are from 0 to 1.2 nm with 63 spline knots.

C. Potassium chloride solution

Coarse-graining of ionic systems can be very costly due to the long-range nature of the electrostatic Coulomb interactions between charged particles. Furthermore, for dilute electrolyte systems, though the number of ions is small, the CG ensemble sampling requires a large number of solvent molecules and long molecular simulations. For example, simulations of 10 mM ion concentration require thousands of water molecules per ion, and due to low ion counts, long simulations are required to achieve reliable statistics. Therefore, iterative coarse-graining techniques, which perform CG sampling at each iteration step, can be prohibitively expensive for electrolyte systems.

Here, we demonstrate a computationally efficient OZ-HNC-based coarse-graining method for electrolyte systems. We consider a bulk potassium chloride solution system at 300 K temperature and 1 M concentration. The reference all-atom simulation consists of 40 Cl^- ions and 40 K^+ ions together with 2180 solvent SPC/E water molecules in a cubic box of length 4.024 nm. The ion-ion and water-ion interactions are modeled by the force field of Joung and Cheatham.[27] In the coarse-grained electrolyte system, we coarse-grain each water molecule into a single bead positioned at its COM and use the same representation of ions as in AAMD. Therefore, for the CG electrolyte system, we need to coarse-grain the ion-water and water-water interactions.

An ion-water interaction has two parts: (i) the long-range ion-water electrostatic interaction and (ii) the short-range ion-water van der Waals (vdW) interaction. The ion-water

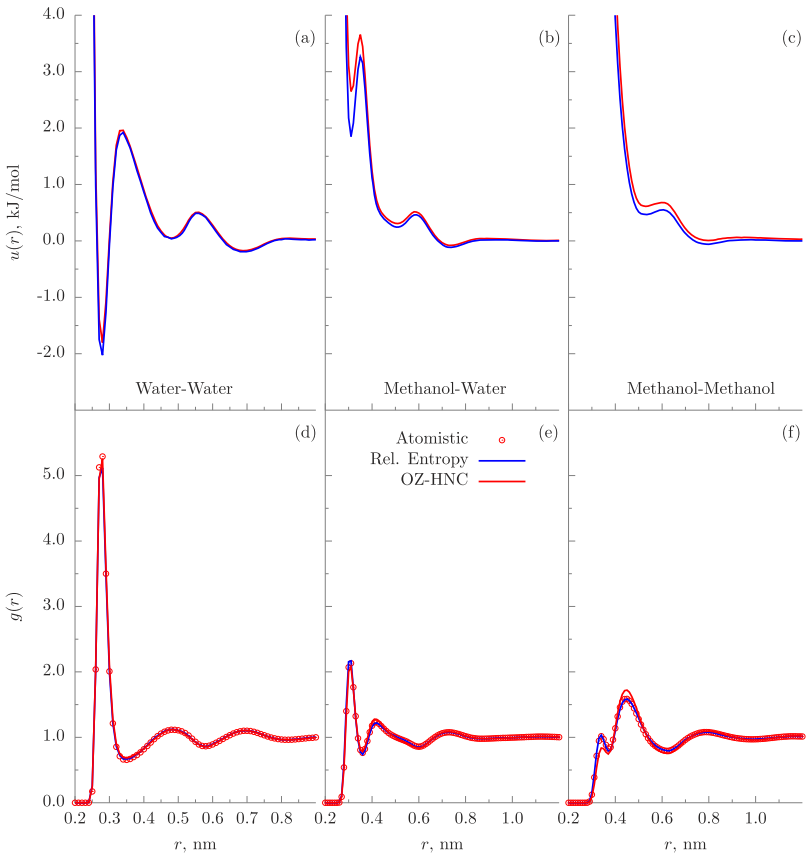

FIG. 3. CG potentials [(a)–(c)] and RDFs [(d)–(f)] of the methanol-water mixture system. CG potentials from the OZ-HNC CG procedure predict water-water and methanol-water COM RDFs accurately. However small errors are observed in the methanol-methanol COM RDFs using the OZ-HNC CG procedure.

electrostatic interaction causes screening of the ion-ion electrostatic interactions. The effects of the ion-water electrostatic interactions can be accounted via the dielectric permittivity of water which scales down (i.e., screens) the ion-ion electrostatic interactions. Therefore, here, we only determine the coarse-grained ion-water short-range vdW interactions. The ion-ion vdW interactions are the same as the Lennard-Jones (LJ) interactions in the atomistic force field. The charges on the ions are also kept the same as in the atomistic force field. The long-range ion-ion electrostatics are screened using the relative permittivity of 78.5. Similar to the methanol-water system, first, we use target RDFs from AAMD in the multicomponent version of the OZ equation, Eq. (10), to obtain the direct correlations between various ion and water pairs. Then, we solve Eq. (7) for water-water, K^+-water, and Cl^--water pairs to obtain the corresponding CG potentials.

Figure 4 shows the water-water, K^+-water, and Cl^--water CG interactions obtained from the OZ-HNC CG approach. We then use the OZ-HNC-based CG potentials to perform the CGMD simulations of the bulk KCl system. From Fig. 4, we observe that ion-water and water-water RDFs from the OZ-HNC CG potential-based CG-MD compare well with the reference AA-MD RDFs. Furthermore, Fig. 4 also shows that ion-water and water-water CG potentials from OZ-HNC are close to the CG potentials from the relative entropy minimization. For relative entropy-based CG of the ionic system, to get good averaged ion-water RDFs, the CGMD simulation in each iteration is run for 1 ns. We use cubic B-spline functions from 0 to 0.9 nm with 48 spline knots for all pair CG potentials. It takes approximately 90 min per iteration.

D. Thermodynamic representability problem

The above results show that the CG potentials from the OZ-HNC CG approach are structurally consistent with the reference AAMD systems. However, CG models usually suffer from representability and transferability issues; i.e., they cannot simultaneously reproduce all of the properties of interest and they may not be applicable for thermodynamic states far away from the one at which the system was parametrized.[28,29] CG models, by definition, lose some information from the underlying reference system. For instance, when replacing a water molecule by a single sphere, one loses information about the molecule's orientation. Here, we evaluate the ability of the OZ-HNC-based CG potentials to resolve other thermodynamic properties.

To check the thermodynamic consistency of the OZ-HNC-based CG potentials, we compute the pressure and excess chemical potential of the CG bulk water system. We follow two approaches to compute the pressure and chemical potential using CG potentials: theoretical approach and CGMD simulations. The theoretical approach is based on the statistical mechanics relation between the pair potential, correlation functions, and thermodynamic properties.[19] The pressure, P, of a system can be estimated from the virial equation as

$$\frac{\beta P}{\rho} = 1 - \frac{2\pi}{3}\sum_{i,j}\rho x_i x_j \int_0^\infty dr r^3 \frac{\partial(\beta u_{ij})}{\partial r} g_{ij}(r), \qquad (13)$$

where x_i is the composition of species i, u_{ij} is the pair potential between i and j fluid species, and $g_{ij}(r)$ is the RDF

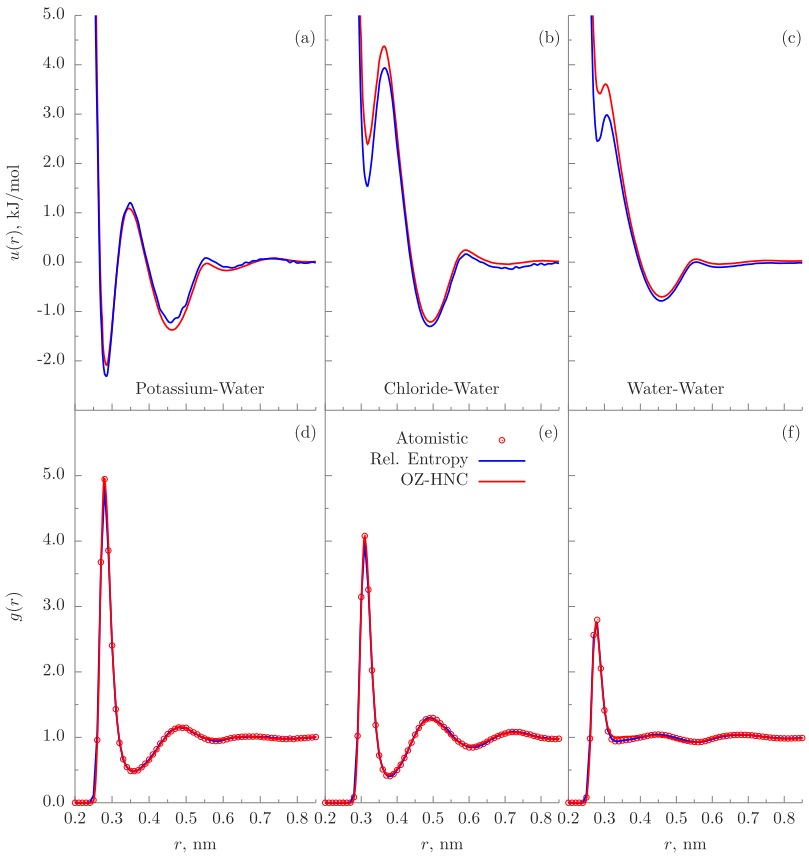

FIG. 4. CG potentials [(a)–(c)] and RDFs [(d)–(f)] of the potassium chloride electrolyte. The ion-water and water-water CG potentials from the OZ-HNC CG procedure are close to the CG potentials obtained from relative entropy minimization and can accurately predict the RDFs in the reference AA simulations.

between species i and j. In the HNC approximation, the excess chemical potential μ_i^{ex} of a given species i can be calculated from[30]

$$\mu_i^{ex} = 4\pi\beta^{-1}\sum_j \rho x_j \int_0^\infty dr r^2 [\frac{1}{2}h_{ij}(r)e_{ij}(r) - c_{ij}(r)], \quad (14)$$

where e_{ij} is the indirect correlation function given by $e_{ij} = h_{ij} - c_{ij}$. To compute the pressure and chemical potential of CG bulk water from theory [Eqs. (13) and (14)], we use the water-water CG potential from OZ-HNC (see Fig. 1). To compute the pressure and chemical potential from MD simulations, we use GROMACS analysis routines, in which the excess chemical potential is computed using the Bennett Acceptance Ratio (BAR) method[31] for AAMD and the test particle insertion method[32] for CGMD.

Table I shows the comparison of the pressure and chemical potentials from the reference AAMD, OZ-HNC-CG-based CGMD, OZ-HNC-CG-based theory, and relative entropy-CG-based CGMD. We observe that the pressure values from all the CG-based approaches are approximately close (within 10%) to each other; however, they differ significantly (more than 10 000 bars) from the reference AAMD pressure. Similarly, the excess chemical potential from different CG-based approaches are approximately similar (within 20%) but differ significantly (more than 50 kJ/mol) from the reference AAMD chemical potential. Therefore, similar to the relative entropy minimization, the CG potential for bulk water from OZ-HNC-based coarse-graining suffers from thermodynamic consistency. We find similar thermodynamic consistency issues for the OZ-HNC-based CG potentials for the methanol-water and KCl electrolyte systems. Dinpajooh and Guenza[33] argue that the integral equation based CG potentials are both structurally and thermodynamically consistent for the melts of polymer chains. However, our results show that, for the molecular liquids such as water and methanol, whose properties are dictated by highly directional intermolecular interactions, strong hydrogen bonding, and orientational ordering, the spherically symmetric isotropic CG pair potentials obtained from OZ-HNC-CG fail to predict the thermodynamic properties, such as pressure and chemical potential, accurately.

E. Accelerated HNC-based iterative coarse-graining

As discussed above, the RDFs from OZ-HNC-based CG potentials compare well with the reference AAMD RDFs and the CG potentials from HNC are very close to the CG potentials from the relative entropy optimization. However, HNC is an approximate theory, and it ignores the effects of the bridge function on the CG potential. Here, we propose an approach to combine OZ-HNC with the iterative coarse-graining approaches, such as IBI and relative entropy. The hybrid OZ-HNC plus iterative coarse-graining offers two main advantages: (i) it addresses the errors in the OZ-HNC-CG due to the bridge function approximation, and (ii) it can reduce the number of iterations (and hence the computational cost) required to converge iterative CG methods. Usually the inverse Boltzmann potential, $u_{\text{init}}(r) = -k_B T \ln\left(g_{\text{tgt}}(r)\right)$, is used as an initial guess for the iterative coarse-graining approach. In the hybrid OZ-HNC-CG-based approach, the CG potential from OZ-HNC is used as an initial guess for the iterative CG method. Since, the CG potential from OZ-HNC is already very close to the converged CG potential from iterative CG, starting with the OZ-HNC CG potential is expected to reduce the number of iterations required for convergence.

TABLE I. Comparison of thermodynamic properties of the bulk water system.

	AAMD	RE-CGMD	OZ-HNC-CGMD	OZ-HNC-CG theoretical
P (bar)	1.144	10218	11002	10680
μ^{ex} (kJ/mol)	−30.63	26.30	32.18	33.11

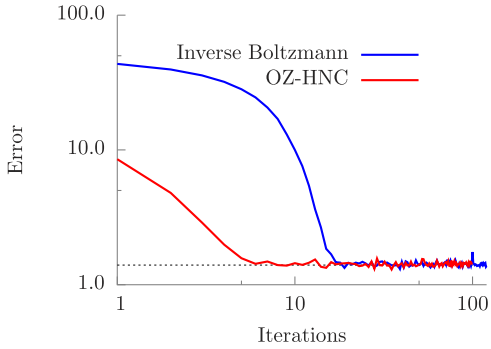

FIG. 5. Error in the water-water RDF of a bulk water system between an inverse Boltzmann initial guess and an OZ-HNC CG potential initial guess of relative entropy minimization. The number of iterations needed for the inverse Boltzmann initial guess is around 20 iterations, and only about 5 iterations are needed for the OZ-HNC CG potential initial guess.

Figure 5 compares the convergence of the relative entropy minimization of CG bulk water with the inverse Boltzmann initial guess and OZ-HNC-CG initial guess. The convergence is measured as a function of the error between the target RDF and CGMD RDF, defined as $|g_{AA}(r) - g_{CG}(r)|$, and the number of iteration steps. For both the initial guesses, we use the same number of B-spline knots to represent the CG potential and the relaxation factor of $\chi = 0.5$. Although the rate of convergence is not affected by the initial guess, the use of OZ-HNC-CG potential significantly decreases the error in the initial step; hence, much fewer iterations are needed for convergence.

IV. CONCLUSION

In this work, we demonstrated the OZ-HNC-based direct and computationally efficient coarse-graining approach by coarse-graining water, methanol-water mixture, and KCl electrolyte bulk systems. We showed that the RDFs from the CGMD with OZ-HNC-based CG potentials compare well with the target RDF from AAMD. Also, the CG potentials from OZ-HNC are very close to the iterative systematic coarse-graining approaches like relative entropy minimization. We also found that, although OZ-HNC-based CG potentials are structurally consistent, they fail to accurately predict the pressure and chemical potentials. Furthermore, we proposed and

demonstrated a hybrid OZ-HNC plus iterative coarse-graining approach which not only addresses the errors in the OZ-HNC-CG due to the bridge function approximation but also accelerates the iterative coarse-graining approaches such as relative entropy minimization.

ACKNOWLEDGMENTS

This work was supported by NSF under Grant Nos. 1420882, 1506619, 1708852, 1720701, 1720633, and 1545907. S.Y.M. and L.M. wish to thank M. H. Motevaselian for constructive feedback on the manuscript.

[1]C. Peter and K. Kremer, Soft Matter **5**, 4357 (2009).
[2]M. Praprotnik, L. D. Site, and K. Kremer, Annu. Rev. Phys. Chem. **59**, 545 (2008).
[3]F. Müller-Plathe, ChemPhysChem **3**, 754 (2002).
[4]G. A. Voth, *Coarse-Graining of Condensed Phase and Biomolecular Systems* (CRC Press, 2008), ISBN: 978-1-4200-5956-4.
[5]W. G. Noid, J. Chem. Phys. **139**, 090901(2013).
[6]C. Hyeon and D. Thirumalai, Nat. Commun. **2**, 487 (2011).
[7]D. Fritz, K. Koschke, V. A. Harmandaris, N. F. A. van der Vegt, and K. Kremer, Phys. Chem. Chem. Phys. **13**, 10412 (2011).
[8]D. Reith, M. Pütz, and F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003).
[9]A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).
[10]M. S. Shell, J. Chem. Phys. **129**, 144108 (2008).
[11]S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005).
[12]J. F. Rudzinski and W. G. Noid, J. Chem. Phys. **135**, 214101 (2011).
[13]J. F. Rudzinski and W. G. Noid, Eur. Phys. J. Spec. Top. **224**, 2193 (2015).
[14]A. Chaimovich and M. S. Shell, J. Chem. Phys. **134**, 094112 (2011).
[15]V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, J. Chem. Theory Comput. **5**, 3211 (2009).
[16]S. Y. Mashayak, M. N. Jochum, K. Koschke, N. R. Aluru, V. Rühle, and C. Junghans, PLoS ONE **10**, e0131754 (2015).
[17]A. Villa, N. F. A. van der Vegt, and C. Peter, Phys. Chem. Chem. Phys. **11**, 2068 (2009).
[18]C. Peter, L. D. Site, and K. Kremer, Soft Matter **4**, 859 (2008).
[19]J.-P. Hansen and I. R. McDonald, in *Theory of Simple Liquids*, 4th ed., edited by J.-P. Hansen and I. R. McDonald (Academic Press, Oxford, 2013), pp. 13–59, ISBN: 978-0-12-387032-2.
[20]J.-P. Hansen and I. R. McDonald, in *Theory of Simple Liquids*, 4th ed., edited by J.-P. Hansen and I. R. McDonald (Academic Press, Oxford, 2013), pp. 105–147, ISBN: 978-0-12-387032-2.
[21]J.-P. Hansen and I. R. McDonald, in *Theory of Simple Liquids*, 4th ed., edited by J.-P. Hansen and I. R. McDonald (Academic Press, Oxford, 2013), pp. 61–104, ISBN: 978-0-12-387032-2.
[22]L. Vrbka, M. Lund, I. Kalcher, J. Dzubiella, R. R. Netz, and W. Kunz, J. Chem. Phys. **131**, 154109 (2009).
[23]P. B. Warren, A. Vlasov, L. Anton, and A. J. Masters, J. Chem. Phys. **138**, 204907 (2013).
[24]S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel *et al.*, Bioinformatics **29**, 845 (2013).
[25]P. Ganguly and N. F. van der Vegt, J. Chem. Theory Comput. **9**, 1347 (2013).
[26]W. L. Jorgensen and J. Tirado-Rives, Proc. Natl. Acad. Sci. U. S. A. **102**, 6665 (2005).
[27]I. S. Joung and T. E. Cheatham, J. Phys. Chem. B **112**, 9020 (2008).
[28]A. A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).
[29]H. Wang, C. Junghans, and K. Kremer, Eur. Phys. J. E **28**, 221 (2009).
[30]J.-P. Hansen, G. Torrie, and P. Vieillefosse, Phys. Rev. A **16**, 2153 (1977).
[31]C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[32]B. Widom, J. Chem. Phys. **39**, 2808 (1963).
[33]M. Dinpajooh and M. G. Guenza, Polymer **117**, 282 (2017).