

Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations

M. Heidari,¹ K. Kremer,¹ R. Potestio,^{1,2,3} and R. Cortes-Huerta^{1,*}

¹*Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*

²*Physics Department, University of Trento, via Sommarive, 14 I-38123 Trento, Italy*

³*INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, I-38123 Trento, Italy*

(Dated: June 19, 2018)

We present an efficient method to obtain bulk isothermal compressibilities (κ_T) and Kirkwood-Buff (KB) integrals of single- and multicomponent liquids using fluctuations of the number of molecules obtained from small-sized molecular dynamics simulations. We write finite-size versions of the Ornstein-Zernike and the Kirkwood-Buff integral equations and include there finite size effects related to the statistical ensemble and the finite integration volumes required in computer simulations. Consequently, we obtain analytical expressions connecting κ_T and the KB integrals in the thermodynamic limit with density fluctuations in the simulated system. We validate the method by calculating various thermodynamic quantities, including the chemical potentials of SPC/E water as a function of the density, and of aqueous urea solutions as a function of the mole fraction. The reported results are in excellent agreement with calculations obtained by using the best computational methods available thus validating the method as a tool to compute the chemical potentials of dense molecular liquids and mixtures. Furthermore, the present method identifies conditions in which computer simulations can be effectively considered in the thermodynamic limit.

I. INTRODUCTION

Statistical mechanics establishes the connection between macroscopic thermodynamic quantities and the microscopic interactions and components of a physical system. In particular, integral equations relate the local structure of a fluid with density fluctuations in the grand canonical ensemble that, in the thermodynamic limit (TL), can be identified with equilibrium thermodynamic quantities such as the compressibility and the derivatives of the chemical potential [1, 2]. In computer simulations this relation is routinely employed in spite of the fact that the systems under consideration are constrained to the canonical ensemble and usually far away from TL conditions. Hence, various finite size effects need to be identified [3–5] and their impact should be evaluated for an appropriate interpretation of the simulation results [6–10].

Building on our previous work [11], in this study we explicitly write finite size versions the Ornstein-Zernike (OZ) [12] and the Kirkwood-Buff (KB) [13] integral equations of the theory of liquids. By including in these equations size effects related to i) the statistical ensemble used in computer simulations and ii) the finite integration volumes required when the system is not in the TL, we derive analytical expressions to obtain isothermal compressibilities and Kirkwood-Buff integrals (KBIs) in the thermodynamic limit from relatively small-sized simulations. The accuracy of the method presented here allows one to derive other thermodynamic quantities of interest such as the chemical potential of rather complex molecular liquids, in particular SPC/E water and aqueous urea

solutions, that well reproduce available simulation data for these systems.

The method presented here is inspired by the block analysis method introduced originally by Binder [14–20]. For the OZ integral equation we have generalised the study of Velasco *et al.* on off-lattice fluids [3, 4, 21–23] to molecular fluids. Moreover, we have extended the applicability of this approach to multicomponent liquids via the KBIs. The paper is organised as follows: In Section II we discuss the finite size version of the OZ integral equation and the evaluation of the bulk isothermal compressibility of SPC/E water which allows one to compute the chemical potential as a function of the density. Section III extends these results to multicomponent systems, where we study urea–water liquid mixtures and obtain the KBIs in the TL and subsequently the chemical potential of urea as a function of mole fraction. In Section IV we summarise our results.

II. SIMPLE LIQUIDS

Let us consider a molecular liquid of average density ρ at temperature T in equilibrium with a reservoir of particles, i.e. an *open* system. The fluctuations of the number of molecules are related to the local structure of the liquid *via* the Ornstein-Zernike integral equation [1, 12]

$$\frac{\Delta^2(N)}{\langle N \rangle} = 1 + \frac{\rho}{V} \int_V \int_V d\mathbf{r}_1 d\mathbf{r}_2 [g^o(\mathbf{r}_1, \mathbf{r}_2) - 1], \quad (1)$$

where $\Delta^2(N)/\langle N \rangle$ are the fluctuations of the number of particles, $\Delta^2(N) = \langle N^2 \rangle - \langle N \rangle^2$ and $g^o(\mathbf{r}_1, \mathbf{r}_2)$ is the pair correlation function of the open system and $\mathbf{r}_1, \mathbf{r}_2$ the position vectors of a pair of fluid particles. To solve the integral in Eq. (1) one assumes that the fluid is homogeneous,

* corteshu@mpip-mainz.mpg.de

isotropic and that the system is in the thermodynamic limit (TL), i.e. $V \rightarrow \infty$, $\langle N \rangle \rightarrow \infty$ with $\rho = \langle N \rangle / V = \text{constant}$. An infinite, homogeneous and isotropic system is translationally invariant, therefore we rewrite Eq. (1) as [1]:

$$\chi_T^\infty = \frac{\Delta^2(N)}{\langle N \rangle} = 1 + 4\pi\rho \int_0^\infty dr r^2 (g^\circ(r) - 1), \quad (2)$$

with $\chi_T^\infty = \rho k_B T \kappa_T$, κ_T being the isothermal compressibility of the bulk system. We have replaced $g^\circ(\mathbf{r}_1, \mathbf{r}_2)$ with $g^\circ(r)$ the radial distribution function (RDF) of the open system, with $r = |\mathbf{r}_2 - \mathbf{r}_1|$. Eq. (2) is only valid in the TL. However, it is routinely used in computer simulations where closed systems with finite volume and fixed number of particles are usually considered. The modified version of Eq. (2) used in simulations reads:

$$\chi_T^R = 1 + 4\pi\rho \int_0^R dr r^2 (g^c(r) - 1), \quad (3)$$

where $g^c(r)$ is the RDF of the *closed* system and the infinity upper limit in the integral on the r.h.s. of Eq. (2) is replaced by a radius $\zeta \ll R < L$ with ζ the correlation length of the system and L the size of the simulation box. These implicit assumptions introduce two different finite size effects: *ensemble* effects appear from assuming $g^\circ(r) = g^c(r)$, which only holds in the TL. If $R \gg \zeta$ then it is true that $g^\circ(r > R) = 1$. By contrast even small fluctuations of $g^c(r > R)$ might contribute significantly to the integral on the r.h.s. of Eq. (3). *Boundary* effects appear because in computer simulations the volume V in Eq. (1) is finite and therefore the step leading to Eq. (2) has to be handled with care. As a matter of fact, the errors introduced by using Eq. (3) yield significant deviations that become apparent when comparing standard molecular dynamics results with semi-grand canonical simulations [24]. An alternative Ornstein-Zernicke integral equation for finite size systems was proposed a long time ago [14, 16, 22, 25]. Let us consider a closed system with fixed number of particles N_0 and volume V_0 , with periodic boundary conditions (PBCs), where fluctuations of the number of particles are computed for subdomains of volume $V \leq V_0$. Therefore, we define [4]

$$\begin{aligned} \chi_T(V, V_0) &= \frac{\Delta^2(N; V, V_0)}{\langle N \rangle_{V, V_0}} \\ &= 1 + \frac{\rho}{V} \int_V \int_V d\mathbf{r}_1 d\mathbf{r}_2 [g^c(r_{12}) - 1], \end{aligned} \quad (4)$$

where $g^c(r_{12})$, $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$, is the pair correlation function of the closed system with total number of particles N_0 , and $\Delta^2(N; V, V_0) = \langle N^2 \rangle_{V, V_0} - \langle N \rangle_{V, V_0}^2$. The fluctuations of the number of particles thus depend on both subdomain and simulation box volumes. The question now is: is it possible to compute χ_T^∞ using Eq. (4) [26]?

To answer this question, we consider first the relation between the RDFs of an open system and its closed counterpart. The original idea is to write the grand canonical RDF as a sum of RDFs, in the canonical ensemble, weighted by the probability that the system has a number of particles N when in contact with a reservoir of particles [21, 27]. For a single component fluid of density ρ at temperature T with fixed number of particles N_0 and volume V_0 , its RDF can be written in terms of an expansion around N_0 as [21, 23, 25–27]:

$$\begin{aligned} g^c(r) &= g^\circ(r) - \frac{\chi_T^\infty}{N_0} \left\{ 1 + \frac{1}{2} \frac{\partial^2}{\partial \rho^2} [\rho^2 (g^\circ(r) - 1)] \right\} \\ &\quad + O(1/N_0^2). \end{aligned} \quad (5)$$

In the following we neglect the $O(1/N_0^2)$ terms in Eq. (5) because it has been demonstrated that, for sufficiently large systems, their contribution can be safely ignored [25]. By inserting Eq. (5) into the integral on the right hand side (r.h.s.) of Eq. (4) we obtain

$$\begin{aligned} \frac{\rho}{V} \int_V \int_V d\mathbf{r}_1 d\mathbf{r}_2 (g^c(r_{12}) - 1) &= \\ I_{V, V} - \frac{V}{V_0} \chi_T^\infty \left[1 + \frac{1}{V} \frac{1}{2} \frac{\partial^2}{\partial \rho^2} (\rho I_{V, V}) \right], \end{aligned} \quad (6)$$

where

$$I_{V, V} = \frac{\rho}{V} \int_V \int_V d\mathbf{r}_1 d\mathbf{r}_2 (g^\circ(r_{12}) - 1), \quad (7)$$

and we use that $\rho = N_0/V_0$.

At this point we include explicitly the second finite size effect, i.e. the fact that the volume V is finite. For this we rewrite $I_{V, V}$ as [17]

$$I_{V, V_0 - V} = I_{V, V_0} - I_{V, V},$$

with

$$\begin{aligned} I_{V, V_0} &= \frac{\rho}{V} \int_V \int_{V_0} d\mathbf{r}_1 d\mathbf{r}_2 (g^\circ(r_{12}) - 1), \\ I_{V, V_0 - V} &= \frac{\rho}{V} \int_V \int_{V_0 - V} d\mathbf{r}_1 d\mathbf{r}_2 (g^\circ(r_{12}) - 1). \end{aligned}$$

To solve these equations, we recall that both integrals, $I_{V, V}$ and I_{V, V_0} , are equal when \mathbf{r}_1 and \mathbf{r}_2 are both within the volume V . Moreover, the integrand is zero for $r_{12} > \zeta$ and therefore it does not contribute to the integral. However, for values of r_{12} close to the boundary of the subdomain V , and in particular when \mathbf{r}_1 lies inside V and \mathbf{r}_2 outside with $r_{12} < \zeta$, there are contributions missing in $I_{V, V}$ that are present in I_{V, V_0} . Thus, the difference $I_{V, V_0 - V} = I_{V, V_0} - I_{V, V}$ must be proportional to the surface-to-volume ratio S/V of the subdomain [17], i. e.

$$I_{V, V_0 - V} = \alpha' \rho \frac{S}{V} = -\frac{\rho \alpha}{V^{1/3}}, \quad (8)$$

with α' , $\alpha > 0$ proportionality constants that at this point we assume to be intensive. S is the surface area of the subdomain with volume V [17].

To compute I_{V, V_0} , we require that $V_\zeta < V < V_0$ with $V_\zeta = 4\pi\zeta^3/3$. Since we assumed PBCs, the system is translationally invariant. Hence, upon applying the transformation $\mathbf{r}_{12} \rightarrow \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ we obtain

$$I_{V, V_0} = \rho \int_{V_0} d\mathbf{r} (g^o(r) - 1) = \chi_T^\infty - 1, \quad (9)$$

where we assume that $g^o(r > \zeta) = 1$ thus ignoring fluctuations of the RDF beyond the volume V . By combining these two results we obtain

$$I_{V, V} = \chi_T^\infty - 1 + \frac{\rho \alpha}{V^{1/3}}, \quad (10)$$

and by including this result in Eq. (6) we obtain

$$\begin{aligned} \frac{\rho}{V} \int_V \int_V d\mathbf{r}_1 d\mathbf{r}_2 (g^c(r_{12}) - 1) &= \chi_T^\infty - 1 + \frac{\rho \alpha}{V^{1/3}} \\ - \frac{V}{V_0} \chi_T^\infty \left[1 + \frac{1}{V} \frac{1}{2} \frac{\partial^2}{\partial \rho^2} \left(\rho \left\{ \chi_T^\infty - 1 + \frac{\rho \alpha}{V^{1/3}} \right\} \right) \right]. \end{aligned} \quad (11)$$

The expression inside the square bracket on the r.h.s. is approximated to 1 since the term with the second derivative with respect to ρ is proportional to $V^{-4/3} \approx 0$. Replacing in Eq. (4) we obtain

$$\chi_T(\lambda) = \chi_T^\infty (1 - \lambda^3) + \frac{\rho \alpha}{V^{1/3}}, \quad (12)$$

with $\lambda \equiv (V/V_0)^{1/3}$ and $V_\zeta < V < V_0$. This expression is consistent with the two obvious limiting cases. In the thermodynamic limit, $V, V_0 \rightarrow \infty$, the correct value is recovered, i.e. $\chi_T(\lambda = 0) = \chi_T^\infty$. In the asymptotic limit $\lambda = 1$ we obtain $\chi_T(\lambda = 1) = 0$, that is, the fluctuations of the number of particles for a closed system are, by definition, equal to zero. The two size effects considered above can be easily identified in the previous expression. First, the term proportional to λ^3 takes into account the difference in ensemble, i.e., $g^c(r) \neq g^o(r)$. Second, the term proportional to $V^{-1/3}$ accounts for the fact that the integration domain V is finite. We emphasise here that this last result can be equally obtained by neglecting the derivative with respect to the density of Eq. (5), i.e. by directly using the expression:

$$g^c(r) = g^o(r) - \frac{\chi_T^\infty}{N_0}, \quad (13)$$

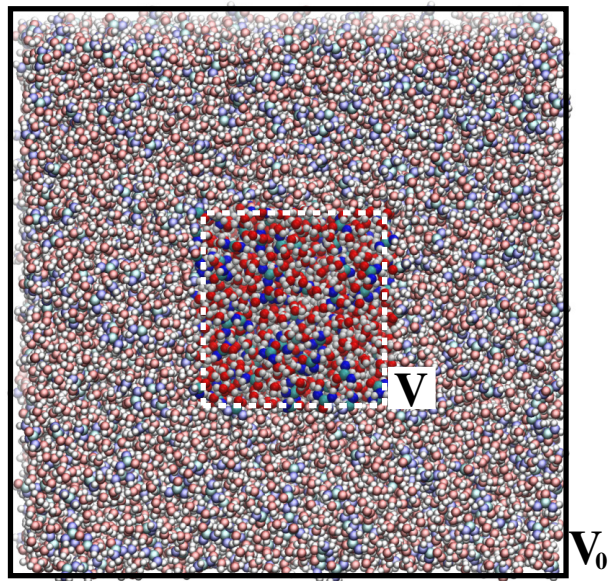


FIG. 1. Snapshot of a urea–water mixture simulation to illustrate the procedure used to obtain $\chi_T(\lambda)$. For a system with fixed volume V_0 , fluctuations of the number of particles are computed for various subdomains of volume $V \leq V_0$. The figure has been rendered with VMD [28].

valid for large values of r and with an asymptotic correction independent of r .

Finally, for practical purposes [22], we multiply Eq. (12) by λ and obtain:

$$\lambda \chi_T(\lambda) = \chi_T^\infty (1 - \lambda^3) \lambda + \frac{\rho \alpha}{V_0^{1/3}}. \quad (14)$$

To make use of Eq. (14), one runs a standard NVT simulation and selects a system with volume $V_0 \gg V_\zeta$. Then one computes systematically the fluctuations of the number of particles given by the l.h.s. of Eq. (4) for subvolumes V such that $V < V_0$, as schematically illustrated in Figure 1. Finally, one plots the results as a function of λ and extracts χ_T^∞ from a simple linear regression in the range $(V_\zeta/V_0)^{1/3} < \lambda \ll 1$. We recall here that extrapolations of thermodynamic properties from finite size computer simulations of single-component systems have been obtained in the past. Examples include the calculation of the compressibility of the Ising lattice gas [14] and hard disk fluids [22, 29], and the investigation of the gas-liquid transition in two-dimensional Lennard-Jones fluids [16, 18], as well as the calculation of the elastic constants of model solids [20].

We have used this strategy to calculate the isothermal compressibility of SPC/E [30] water. Molecular dynamics simulations have been carried out with GROMACS 4.5.1 [31] for 8000, 16000, and 32000 SPC/E [30] water molecules. To equilibrate the system, we started with a simulation box of size such that the initial density was ≈ 26 waters/nm³ (≈ 776 kg/m³). Then, optimisation

of the system was achieved by performing steepest descent minimisation (50000 steps are sufficient). Finally, an equilibration run of 3.5 ns has been carried out in the NPT ensemble at 1 bar. During this time the displacement of a randomly chosen water molecule is comparable to half the linear size of the simulation box.

Following this minimisation step we proceeded by alternating 3.5 ns (time step = 1 fs) constant pressure (NPT) at P=1 bar and constant volume (NVT) simulations at T=300 K. For NPT simulations we used a Berendsen barostat [32] whereas for NVT simulations temperature was enforced by a velocity rescaling thermostat [33]. We continued with this protocol until we verified that in the NPT ensemble the density is 33.5 waters/nm³ (1000 kg/m³) and that in the NVT simulation pressure never deviates significantly from the 1 bar value. For all the cases considered in the present study, the last NVT trajectory obtained after this sequence of NPT–NVT equilibration runs was used for the subsequent analysis in terms of the fluctuations of the number of molecules.

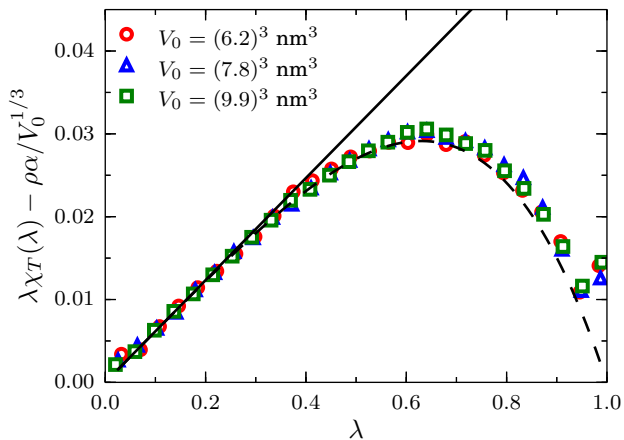


FIG. 2. Fluctuations of the number of water molecules as a function of λ for systems of 8000 (red circles), 16000 (blue triangles) and 32000 (green squares) corresponding to the volumes indicated in the legends. Using Eq. (14) and the linear fit to the data for 8000 water molecules in the range $\lambda < 0.3$, we obtain the curve represented by the dashed black line and the solid black line (neglecting the λ^3 term in Eq. (14)).

We evaluated $\lambda\chi_T(\lambda)$ as a function of λ for a system of 8000 water molecules. These data suggest a linear regime for $\lambda < 0.3$, consistent with Eq. (14), where we used a linear fit to the data to obtain χ_T^∞ and α . We report $\chi_T^\infty = 0.062$ (3 % error calculated from a simple linear regression), and by multiplying this value by the molar mass of water and dividing by $\rho k_B T$, we obtain $4.48 \times 10^{-5} \text{ bar}^{-1}$ in excellent agreement with the widely accepted value for the isothermal compressibility of the SPC/E model ($4.50 \times 10^{-5} \text{ bar}^{-1}$) [34]. Fig. 2 shows $\lambda\chi_T(\lambda) - \rho\alpha/V_0^{1/3}$ as a function of λ for systems of 8000 (red circles), 16000 (blue triangles) and 32000 (green squares) water molecules. We replaced χ_T^∞ and α

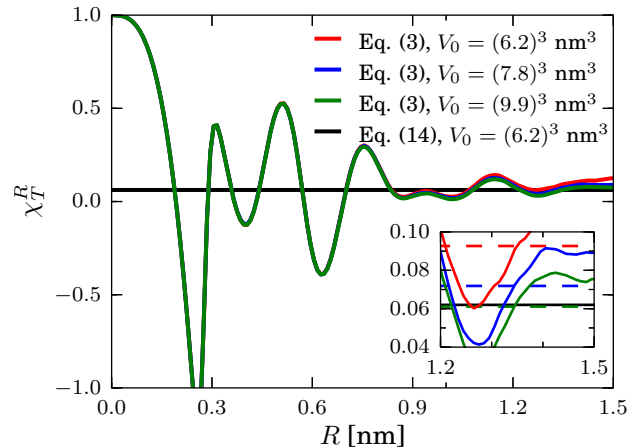


FIG. 3. Evaluation of Eq. (3) for $R = 1.5 \text{ nm}$ using the RDFs obtained from simulation boxes of side 6.2 (red line), 7.8 (blue line) and 9.9 nm (green line). The horizontal black line marks the asymptotic limit, $\chi_T^\infty = 0.062$, obtained from the linear fit on Fig. 2 and Eq. (14). The panel shows the interval $1.2 < R < 1.5 \text{ nm}$. χ_T^R for large values of R can be obtained by averaging over such an interval. These results correspond to the coloured horizontal lines. For $V_0 = (6.2)^3$ and $(9.9)^3 \text{ nm}^3$ we obtain the limits 0.093 ± 0.004 , 0.072 ± 0.003 and 0.061 ± 0.003 , respectively.

back in Eq. (14) to obtain the black curve that superposes on the simulation data for the full interval $0 < \lambda < 1$. The reader should keep in mind that for very small values of λ , when the linear size of the subdomain is comparable to the size of the molecules, the statistical analysis is questionable. The three sets of data follow the same curve, indicating that α is indeed an intensive quantity but also suggesting that it is possible to accurately estimate χ_T^∞ using the results of a relatively small simulation. By neglecting the term proportional to λ^3 in Eq. (14), we obtained the straight black line in Fig. 2 that indicates that deviations from linearity indeed become important for $\lambda > 0.3$. This limit $\lambda \ll 1$, where a linear behaviour is apparent, is rather interesting. In this regime the volume $V \ll V_0$, that is, the system is effectively in the grand canonical ensemble. If $V_0 \gg V_\zeta$ we obtain $\chi_T(\lambda \ll 1) = \chi_T^\infty$. Namely, provided that $V_\zeta < V < V_0$ in the interval $\lambda < 0.3$, both V and V_0 are large enough such that the system seems to be in the TL. A way to challenge this argument involves the RDF: if the system is in the TL then $g^c(r) = g^o(r)$ and Eq. (3) gives χ_T^R for $R > \zeta$. We choose $R = 1.5 \text{ nm}$ as a reasonable value since the O-O RDF of the SPC/E model goes to 1 at approximately 1 nm. This choice gives $V^{1/3} = 1.5 \times (4\pi/3)^{1/3} \text{ nm}$, and with the sizes $V_0^{1/3} \approx 6.2, 7.8$ and 9.9 nm we obtain $\lambda \approx 0.39, 0.31$ and 0.24 , respectively. For these systems we compute $g^c(r)$ and evaluate χ_T^R from Eq. (3). We obtain χ_T^R from the integral by averaging in the interval $1.2 < R < 1.5 \text{ nm}$. As expected, upon increasing the size of the simulation box (with $\lambda = 0.39, 0.31$ and 0.24

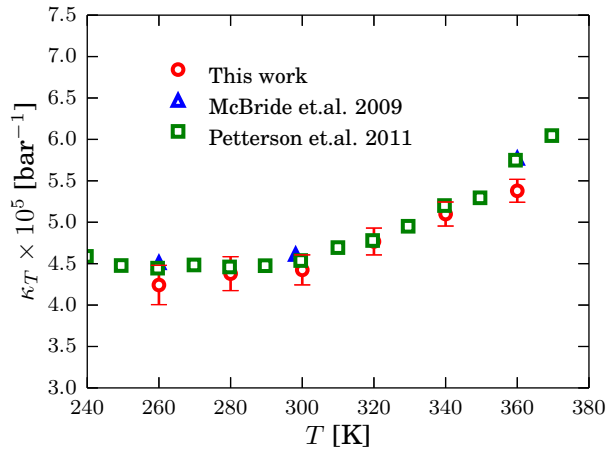


FIG. 4. Isothermal compressibility for SPC/E water as a function of temperature at fixed pressure (1 bar) using the method described in the text (red circles) and direct volume fluctuations: blue triangles [35] and green squares [36, 37].

the integral takes the values 0.093 ± 0.004 , 0.072 ± 0.003 , 0.061 ± 0.003) the integral converges to the value of 0.062 computed with the present method (See Fig. 3).

We conclude this part by saying that this method also allows us to identify simulation conditions in which the TL can be reached: in the case of SPC/E water, the volume at which the RDF is computed should be maximum of the order of one percent of the total volume of the simulation box.

To test the accuracy of the method, we have computed the isothermal compressibility of SPC/E water in the temperature interval 260 - 360 K. As shown in Fig. 4 there is reasonable agreement with previous calculations using directly volume fluctuations [35–37]. More important, the efficient and accurate calculation of isothermal compressibility for dense liquids opens the possibility to compute other thermodynamic quantities of great interest. For example, the isothermal compressibility is related to the chemical potential μ via

$$\kappa_T = \frac{1}{\rho^2} \left. \frac{\partial \rho}{\partial \mu} \right|_T, \quad (15)$$

which can be rearranged in terms of μ as:

$$\delta\mu = \int_{\rho_0}^{\rho} \frac{d\rho'}{\rho'^2 \kappa_T} \quad (16)$$

with $\delta\mu = \mu - \mu_0$ and μ_0 the chemical potential of the system at the reference density ρ_0 . This expression suggests that, upon an accurate calculation of the compressibility for different densities at fixed temperature, it is possible to obtain the chemical potential of the system, shifted by a constant μ_0 . We have computed the compressibility of SPC/E water in the density interval 0.99 - 1.05 g/cm³

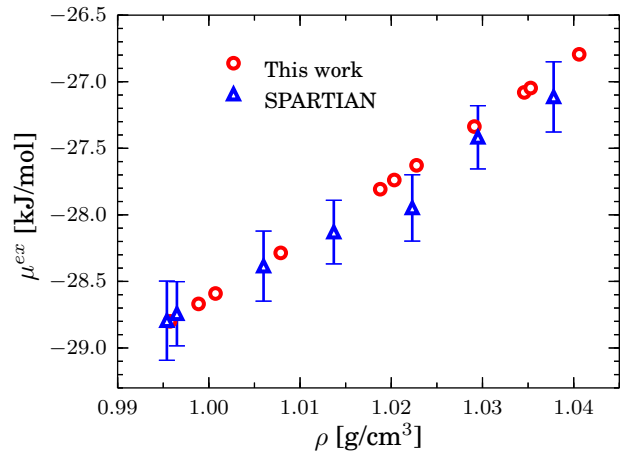


FIG. 5. Excess chemical potential as a function of density for the SPC/E water model at 300 K. The blue triangles are the data obtained with the SPARTIAN method [38] and the red circles correspond to the points computed with the method outlined in the text. For the latter the error bars are comparable to the point size.

and used the expression (16) to obtain $\delta\mu$. To estimate the shift and to prove that the method outlined here gives correct results, we have computed the excess chemical potential of SPC/E water in the same density interval using a method recently developed by us which can be interpreted as spatially resolved thermodynamic integration (SPARTIAN) [38]. To obtain the excess chemical potential we subtract the density-dependent part of the chemical potential of the ideal gas from $\delta\mu$, i.e.

$$\delta\mu^{\text{ex}} = \delta\mu - k_B T \ln \rho. \quad (17)$$

Results are presented in Fig. 5, and upon displacing the point corresponding to 0.995 g/cm³ we find that both data sets overlap remarkably well within error bars. In this way we have shown that with the strategy presented here it is possible to compute efficiently the chemical potential of dense liquids. When used in combination with another method that allows one to calculate the chemical potential at a reference state point, it also brings results consistent with state-of-the-art chemical potential calculations. Since in this context the chemical potential is obtained via the simple evaluation of fluctuations of the number of particles, we provide here an alternative strategy useful, in particular, in cases where the high density of the liquid becomes problematic for methods based on particle insertion or thermodynamic integration.

III. MULTICOMPONENT SYSTEMS

From the previous analysis, the generalisation to multi-component systems is straightforward [11]. For a mul-

ticomponent fluid of species i, j in equilibrium at temperature T , analogous to the Ornstein-Zernike integral equation, the Kirkwood-Buff integral (KB) is defined as [13]:

$$\begin{aligned} G_{ij}^o &= V \left(\frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\langle N_i \rangle} \right) \\ &= \frac{1}{V} \int_V \int_V d\mathbf{r}_1 d\mathbf{r}_2 [g_{ij}^o(\mathbf{r}_1, \mathbf{r}_2) - 1], \end{aligned} \quad (18)$$

with δ_{ij} the Kronecker delta. As before, the superscript (o) indicates that this definition holds for an open system, i.e. a system described by the grand canonical ensemble. $g_{ij}^o(\mathbf{r}_1, \mathbf{r}_2)$ is the multi-component pair correlation function of the open system. KB integrals connect the local structure of a multicomponent system to density fluctuations calculated in the grand canonical ensemble, that, in the thermodynamic limit, are related to equilibrium quantities such as the isothermal compressibility and derivatives of the chemical potential [2]. Recently, and to illustrate the robustness of this framework, KB integrals have been used to investigate the thermodynamics of Lennard-Jones binary mixtures [39], solvation of biomolecules [40], diffusion in multicomponent liquids [41, 42], complex phenomena in molecular mixtures [43] and denaturation [44] and self-assembly [45] of proteins. As in the single component case, by assuming that the fluid is homogeneous and isotropic we obtain the KB integrals in the TL:

$$G_{ij}^\infty = 4\pi \int_0^\infty dr r^2 (g_{ij}^o(r) - 1), \quad (19)$$

where $g_{ij}^o(r)$ is the multicomponent RDF of the infinite system. However, in computer simulations the expression

$$G_{ij}^R = 4\pi \int_0^R dr r^2 (g_{ij}^c(r) - 1), \quad (20)$$

is widely used. Here, $g_{ij}^c(r)$ is the multicomponent RDF of the finite system and $R \gg \zeta$ is a cutoff distance.

We explicitly include both finite size effects: ensemble and boundary effects, into an expression, analogous to Eq. (18), defined for a finite system. Let us consider systems with a total fixed number of particles N_0 and volume V_0 with PBCs. We define [4]

$$\begin{aligned} G_{ij}(V, V_0) &= V \left(\frac{\langle N_i N_j \rangle' - \langle N_i \rangle' \langle N_j \rangle'}{\langle N_i \rangle' \langle N_j \rangle'} - \frac{\delta_{ij}}{\langle N_i \rangle'} \right) \\ &= \frac{1}{V} \int_V \int_V d\mathbf{r}_1 d\mathbf{r}_2 [g_{ij}^c(r_{12}) - 1], \end{aligned} \quad (21)$$

with $g_{ij}^c(r_{12})$, $r_{12} = \mathbf{r}_2 - \mathbf{r}_1$, the pair correlation function of the closed system, and $\langle N_i \rangle' \equiv \langle N_i \rangle_{V, V_0}$ the average number of i -particles that depends on V and V_0 . Thus,

we define a quantity $G_{ij}(V, V_0)$ that can be evaluated by computing fluctuations of the number of particles in finite subdomains of volume V inside a simulation box of volume V_0 .

We have proposed an expression, valid for large separations r , relating the open and closed RDFs of the form [11]:

$$g_{ij}^c(r) = g_{ij}^o(r) - \frac{1}{V_0} \left(\frac{\delta_{ij}}{\rho_i} + G_{ij}^\infty \right), \quad (22)$$

based on the asymptotic limit $g_{ij}^c(r \rightarrow \infty) = 1 - (\delta_{ij}/\rho_i + G_{ij}^\infty)/V_0$ discussed in Ref. [2]. As expected, when the total volume $V_0 \rightarrow \infty$ we recover $g_{ij}^c(r) = g_{ij}^o(r)$. We include Eq. (22) in the integral on the r.h.s. of Eq. (21) and obtain:

$$\begin{aligned} G_{ij}(V, V_0) &= \frac{1}{V} \int_V \int_V d\mathbf{r}_1 d\mathbf{r}_2 [g_{ij}^o(r_{12}) - 1] \\ &\quad - \frac{V}{V_0} \left(\frac{\delta_{ij}}{\rho_i} + G_{ij}^\infty \right). \end{aligned} \quad (23)$$

In analogy to the single-component case, in Eq. (22) we neglect $O(1/V_0)$ contributions that include derivatives of $g_{ij}^o(r)$ with respect to ρ_i , as well as $O(1/V_0^2)$ terms, because their contribution to ensemble size effects becomes negligible upon integration.

Second, for the boundary effects, by explicitly taking into account the finite domain of integration V , we obtain [4, 11]:

$$\begin{aligned} G_{ij}(V, V_0) &= \frac{1}{V} \int_V \int_{V_0} d\mathbf{r}_1 d\mathbf{r}_2 [g_{ij}^o(r_{12}) - 1] \\ &\quad - \frac{V}{V_0} \left(\frac{\delta_{ij}}{\rho_i} + G_{ij}^\infty \right) + \frac{\alpha_{ij}}{V^{1/3}}, \end{aligned} \quad (24)$$

where α_{ij} is a constant that depends only on intensive thermodynamic system properties such as density and temperature. Finally, we rewrite Eq. (24) as [11]

$$G_{ij}(\lambda) = G_{ij}^\infty (1 - \lambda^3) - \lambda^3 \frac{\delta_{ij}}{\rho_i} + \frac{\alpha_{ij}}{V_0^{1/3} \lambda}, \quad (25)$$

with $\lambda = (V/V_0)^{1/3}$. And like in the single-component case, we can use Eq. (25) to extract the Kirkwood-Buff integrals in the thermodynamic limit, G_{ij}^∞ . In the case of multicomponent fluids, the explicit inclusion of finite size effects to obtain G_{ij}^∞ has been attempted in the literature by either using arguments from the thermodynamics of small systems [46] where only the $1/V^{1/3}$ term appears [47] or by introducing empirical tail corrections to the RDF and evaluating Eq. (20) with a modified kernel [48, 49]. Here we identify the relevant finite size effects at play and use them in a simple and physically sound framework that enables an accurate and efficient calculation of G_{ij}^∞

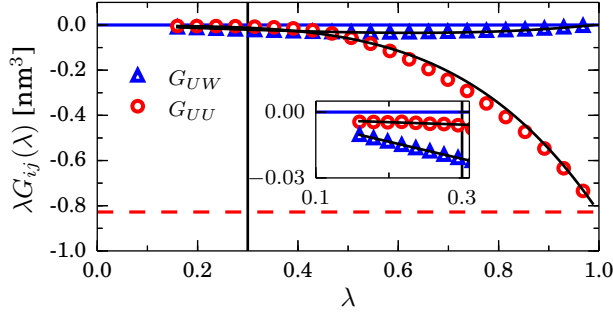


FIG. 6. Simulation results for a molar concentration of 2.00 M showing λG_{UU} (red circles) and λG_{UW} (blue triangles). The horizontal lines indicate the asymptotic limits for $G_{ij}(\lambda \rightarrow 1)$: $G_{UU} \rightarrow 1/\rho_U = 1/1.208 = 0.828$ and $G_{UW} \rightarrow 0$. The linear regime ($\lambda < 0.3$ as seen in the inset), with slope equal to G_{ij}^∞ , is indicated by the vertical black line. The black lines superimposing on the data for the whole interval $0 < \lambda < 1$ were obtained by fitting Eq. (26) in the linear regime.

[11]. Recently, the accuracy of this method has been tested by computing several thermodynamic properties of Lennard Jones mixtures showing good agreement with theoretical predictions [39].

We rewrite Eq. (25) as

$$\lambda G_{ij}(\lambda) = \lambda G_{ij}^\infty (1 - \lambda^3) - \lambda^4 \frac{\delta_{ij}}{\rho_i} + \frac{\alpha_{ij}}{V_0^{1/3}}. \quad (26)$$

To validate Eq. (26), we have performed simulations of aqueous urea solution [50, 51] using the Kirkwood-Buff derived force field [52] and SPC/E water [30] in GRO-MACS 4.5.1 [31] with a relatively small size of the simulation box ($L \sim 8$ nm). In addition to the trajectories from Ref. 51 we have considered four more molar concentrations for a total of seven molar concentrations: 2.00, 3.06, 3.90, 5.07, 6.03, 7.10 and 8.03. To equilibrate the system, we have alternated 3.5 ns (time step = 1 fs) constant pressure (NPT) at $P=1$ bar and constant volume (NVT) simulations at $T=300$ K. For NPT simulations we used a Berendsen barostat [32] whereas for NVT simulations temperature was enforced by a velocity rescaling thermostat [33]. We continued with this protocol until we verified that in the NVT simulation pressure never deviates significantly from the 1 bar value. The last NVT trajectory obtained after this NPT–NVT equilibration sequence was used to compute fluctuations of the number of molecules.

Fig. 6 shows fluctuations of the number of molecules of species $i, j = \text{urea (U), water (W)}$ as a function of λ for a molar concentration of 2.00 M. Blue triangles correspond to $\lambda G_{UW}(\lambda)$ that goes to zero when $\lambda \rightarrow 1$, as the horizontal blue solid line indicates. Red circles represent $\lambda G_{UU}(\lambda)$ that in the limit $\lambda \rightarrow 1$ goes to $1/\rho_U = 1/1.208 = 0.828$ as indicated by the horizontal red dashed line. As in the case of single component systems, here it is also clear that there is a linear region

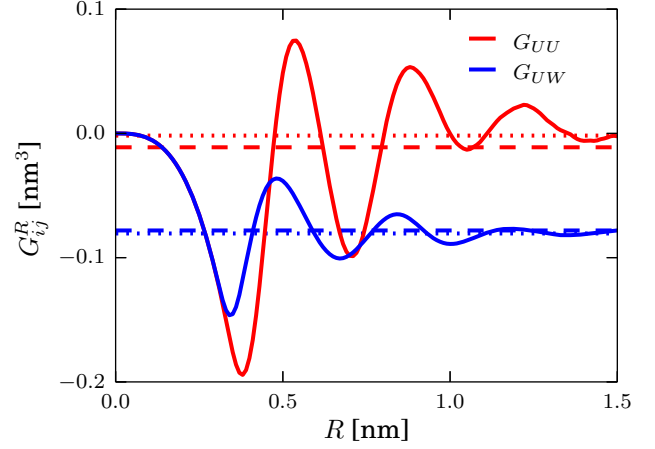


FIG. 7. Results for G_{ij}^R obtained from Eq. (20) for urea water mixtures at $C_U=2.00$ M. The dotted horizontal lines indicate the average of the curve obtained between 1.3 and 1.5 nm. The dashed horizontal lines correspond to G_{ij}^∞ as obtained from Eq. (26).

in the interval $0 < \lambda < 0.3$, indicated by the vertical solid black line (See inset in Fig. 6). From a simple linear regression in this region G_{UW}^∞ , G_{UU}^∞ as well as the values of α_{ij} are obtained. By inserting these values in Eq. (26) we obtain curves, in black solid lines, that overlap the simulation data points in all cases.

Since we can compute $g^c(r)$ directly from our trajectories, we compare the fluctuations results with the evaluation of the integral Eq. (20). Results for G_{UU}^R , G_{UW}^R with $R = 1.5$ nm are presented in Fig. 7 with solid red and blue curves, respectively. To obtain a value for the Kirkwood-Buff integrals we have averaged these curves in the interval 1.3 – 1.5 nm as indicated by the dotted horizontal lines. In addition, the dashed lines correspond to the values G_{UU}^∞ and G_{UW}^∞ obtained from the linear fit of Eq. (26). Differences in G_{UU}^∞ and the average from the curve are significant. The reason behind such differences lies in the fact that the concentration of urea is low, $C_U = 2.00M$, and a larger simulation box is needed to correctly compute the corresponding RDF. Conversely, the differences for G_{UW} are negligible because the size of the system is large enough to accurately obtain the urea – water RDF.

In our previous paper we have used these tools to investigate solvation thermodynamics [11]. In urea–water mixtures at pressure P , temperature T and number density ρ_U the derivative of the urea activity coefficient can be written as

$$\gamma_{UU} = 1 + \left(\frac{\partial \ln \gamma_U}{\partial \ln \rho_U} \right)_{P,T} = \frac{1}{1 + \rho_U(G_{UU} - G_{UW})}, \quad (27)$$

with γ_U the activity coefficient and $k_B T \ln \gamma_U$ the chemical potential of urea. We have computed γ_{UU} as a function of molar concentration and report the results in Fig.

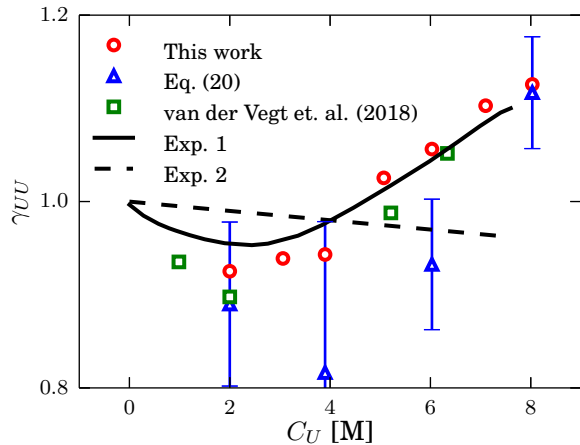


FIG. 8. Derivative of the activity coefficient γ_{UU} as a function of different urea molar concentrations C_U . Experimental data (Exp. 1 [52, 53] solid black line and Exp. 2 [56] dashed black line) are also presented for comparison. Concerning other simulation methods, the blue triangles have been obtained by using Eq. (20) [50, 51]. Recently, the activity coefficient has been computed by using a modified version of Eq. (20) with empirical tail corrections to the RDF. The green squares correspond to one data set extracted from Ref. [54] with the smallest error bars and that best reproduces the experimental results. For the latter, as well as for our data, the error bars are comparable to the symbol size.

8. These results are in good agreement with the experimental measurements [53] used to parameterise the force field [52]. Recently, a modified version of Eq. (20) with empirical tail corrections to the RDF have been used to compute γ_{UU} [54]. To compare with our method, we extract [55], among the various examples provided [54], the data set with the smallest error bars and that best reproduces the experimental results. The results presented in Fig. 8 suggest that a simple calculation of the fluctuations of the number of particles used in combination with eq. (26) produce results consistently in better agreement with Exp. 1 [53].

More important, we can also use Kirkwood-Buff integrals to compute the chemical potential of urea in urea–water mixtures. We use the expression

$$\frac{1}{k_B T} \left(\frac{\partial \mu_U}{\partial \rho_U} \right)_{P,T} = \frac{1}{\rho_U} + \frac{G_{UW} - G_{UU}}{1 + \rho_U (G_{UU} - G_{UW})}, \quad (28)$$

and as in the case of single component fluids, we integrate this equation to obtain a shifted chemical potential:

$$\delta \mu_U = k_B T \int_{\rho_{U,0}}^{\rho_U} d\rho'_U \left[\frac{1}{\rho'_U} + \frac{G_{UW} - G_{UU}}{1 + \rho'_U (G_{UU} - G_{UW})} \right], \quad (29)$$

where we have dropped the subindices P, T to lighten the notation. Using a combination of different methods,

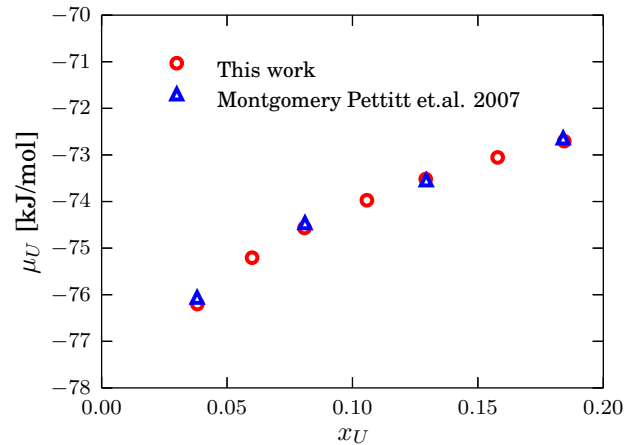


FIG. 9. Chemical potential of urea in aqueous solution as a function of mole fraction x_U . The results obtained with the method outlined in the text (red circles) are compared with the data from Ref. [57] (blue triangles).

the chemical potential of urea in urea–water mixtures as a function of mole fraction has been computed in Ref. [57]. In Fig. 9 we report these results in the interval 0,0.20 with blue triangles. We have obtained the shifted chemical potential of urea as a function of mole fraction using Eq. (29) and plot the results using the data for the mole fraction $x_U = 0.13$ ($C_U = 6.03$) as a reference state point (red circles). The two data points are in excellent agreement. This suggests that the method presented here constitutes an efficient and accurate tool to compute the chemical potential of liquid mixtures for a wide range of concentrations.

IV. CONCLUSIONS

During the last decades, we have witnessed an exponentially increasing computational capacity that allowed us to simulate enormous systems. In some cases, like the SPC/E example with $V_0 = (9.9 \text{ nm})^3$ discussed above, the system could practically be considered in the thermodynamic limit. Usually, this situation constitutes the exception rather than the rule, and the computation of bulk thermodynamic quantities from small-sized simulation results remains a challenging task. This limitation did not prevent pioneer computational scientists to investigate these properties for various complex systems. On the contrary, researchers have cleverly used the finite size effects present in their systems to extrapolate interesting quantities in the thermodynamic limit.

Inspired by these works, we used finite size variants of integral equations in the theory of liquids, namely for the Ornstein-Zernike and Kirkwood-Buff integral equations. By explicitly including *ensemble* and *boundary* size effects, we have derived analytical expressions that allow one to accurately calculate isothermal compressibilities

and Kirkwood-Buff integrals in the thermodynamic limit. In particular, the accurate estimation of these quantities for various density/concentration conditions allows one to determine, upon integration, the chemical potential for complex liquids and mixtures.

This protocol to compute chemical potentials simply relies on running a standard computer simulation. Provided that the linear dimension of the simulation box is larger than the correlation length of the system, the obtained trajectories can be easily used to compute the fluctuations of the number of molecules for different subdomain sizes. This can be carried out during the postprocessing and without any external sophisticated computational strategies, in contrast with the standard methods available in the literature.

The efficiency and accuracy of the method has been demonstrated by the calculation of shifted chemical potentials of SPC/E water as a function of the density, and aqueous urea solution as a function of the mole fraction. The reported results exhibit trends in excel-

lent agreement with state-of-the-art calculations. The reference chemical potential can be calculated using an alternative, well-established, method for a single density/concentration value. Thus, the simple estimation of density fluctuations, as described here, provides a powerful tool to determine the chemical potential of complex systems for a wide range of density/concentration conditions.

ACKNOWLEDGMENTS

We thank Debashish Mukherji and Roberto Menichetti for many stimulating discussions, and also thank the latter and Dominic Spiller for a critical reading of the manuscript. M.H., K.K. and R.P. gratefully acknowledge funding from SFB-TRR146 of the German Research Foundation (DFG). R.C.-H. gratefully acknowledges the Alexander von Humboldt Foundation for financial support.

-
- [1] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, 2006).
- [2] A. Ben-Naim, *Molecular Theory of Solutions* (Oxford University Press, 2006).
- [3] F. L. Román, J. A. White, A. González, and S. Velasco, *J. Chem. Phys.* **110**, 9821 (1999).
- [4] F. Román, J. White, A. González, and S. Velasco, "Theory and simulation of hard-sphere fluids and related systems," (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008) Chap. Ensemble Effects in Small Systems, pp. 343–381.
- [5] D. Frenkel, *The European Physical Journal Plus* **128**, 10 (2013).
- [6] J. I. Siepmann, I. R. McDonald, and D. Frenkel, *Journal of Physics: Condensed Matter* **4**, 679 (1992).
- [7] T. Zykova-Timan, C. Valeriani, E. Sanz, D. Frenkel, and E. Tosatti, *Phys. Rev. Lett.* **100**, 036103 (2008).
- [8] M. Vögele and G. Hummer, *The Journal of Physical Chemistry B* **120**, 8722 (2016), pMID: 27385207, <https://doi.org/10.1021/acs.jpcc.6b05102>.
- [9] P. Simonnin, B. Noetinger, C. Nieto-Draghi, V. Marry, and B. Rotenberg, *Journal of Chemical Theory and Computation* **13**, 2881 (2017), pMID: 28535342, <https://doi.org/10.1021/acs.jctc.7b00342>.
- [10] L. Belloni and J. Puibasset, *The Journal of Chemical Physics* **147**, 224110 (2017), <https://doi.org/10.1063/1.5005912>.
- [11] R. Cortes-Huerto, K. Kremer, and R. Potesio, *J. Chem. Phys.* **145**, 141103 (2016), <https://doi.org/10.1063/1.4964779>.
- [12] L. S. Ornstein and F. Zernike, *Proc. Akad. Sci. (Amsterdam)* **17**, 793 (1914).
- [13] J. G. Kirkwood and F. P. Buff, *J. Chem. Phys.* **19**, 774 (1951).
- [14] K. Binder, *Z. Phys. B* **43**, 119 (1981).
- [15] K. Kaski, K. Binder, and J. D. Gunton, *Phys. Rev. B* **29**, 3996 (1984).
- [16] M. Rovere, D. W. Hermann, and K. Binder, *EPL* **6**, 585 (1988).
- [17] M. Rovere, D. W. Heermann, and K. Binder, *J. Phys.: Condens. Matter* **2**, 7009 (1990).
- [18] M. Rovere, P. Nielaba, and K. Binder, *Z. Phys. B* **90**, 215 (1993).
- [19] H. Weber, D. Marx, and K. Binder, *Phys. Rev. B* **51**, 14636 (1995).
- [20] S. Sengupta, P. Nielaba, M. Rao, and K. Binder, *Phys. Rev. E* **61**, 1072 (2000).
- [21] F. L. Román, J. A. White, and S. Velasco, *J. Chem. Phys.* **107**, 4635 (1997).
- [22] F. L. Román, J. A. White, and S. Velasco, *EPL* **42**, 371 (1998).
- [23] F. L. Román, A. González, J. A. White, and S. Velasco, *Am. J. Phys.* **67**, 1149 (1999).
- [24] D. Mukherji and K. Kremer, *Macromolecules* **46**, 9158 (2013).
- [25] J. J. Salacuse, A. R. Denton, and P. A. Egelstaff, *Phys. Rev. E* **53**, 2382 (1996).
- [26] D. Villamaina and E. Trizac, *Eur. J. Phys.* **35**, 035011 (2014).
- [27] J. L. Lebowitz and J. K. Percus, *Phys. Rev.* **122**, 1675 (1961).
- [28] W. Humphrey, A. Dalke, and K. Schulten, *Journal of Molecular Graphics* **14**, 33 (1996).
- [29] J. Salacuse, *Physica A* **387**, 3073 (2008).
- [30] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.* **91**, 6269 (1987); L. X. Dang and B. M. Pettitt, *ibid.* **91**, 3349 (1987); Y. Wu, H. L. Tepper, and G. A. Voth, *J. Chem. Phys.* **124**, 024503 (2006).
- [31] 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, B. Hess, and E. Lindahl, *Bioinformatics* **29**, 845 (2013).
- [32] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* **81**,

- 3684 (1984).
- [33] G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- [34] P. G. Debenedetti, M. A. Ricci, and F. Bruni, eds., *Water: Fundamentals as the Basis for Understanding the Environment and Promoting Technology* (IOS Press, 2015).
- [35] H. L. Pi, J. L. Aragoes, C. Vega, E. G. Noya, J. L. Abascal, M. A. Gonzalez, and C. McBride, *Molecular Physics* **107**, 365 (2009), <http://dx.doi.org/10.1080/00268970902784926>.
- [36] L. G. Pettersson and A. Nilsson, *Journal of Non-Crystalline Solids* **407**, 399 (2015), 7th IDMRCs: Relaxation in Complex Systems.
- [37] K. T. Wikfeldt, C. Huang, A. Nilsson, and L. G. M. Pettersson, *The Journal of Chemical Physics* **134**, 214506 (2011), <http://dx.doi.org/10.1063/1.3594545>.
- [38] M. Heidari, K. Kremer, R. Cortes-Huerto, and R. Potesio, *ArXiv e-prints* (2018), arXiv:1802.08045 [cond-mat.soft].
- [39] A. A. Galata, S. D. Anogiannakis, and D. N. Theodorou, *Fluid Phase Equilibria* (2017), <https://doi.org/10.1016/j.fluid.2017.11.003>.
- [40] V. Pierce, M. Kang, M. Aburi, S. Weerasinghe, and P. E. Smith, *Cell Biochem. Biophys.* **50**, 1 (2008).
- [41] S. Kjelstrup, S. K. Schnell, T. J. H. Vlugt, J.-M. Simon, A. Bardow, D. Bedeaux, and T. Trinh, *Adv. Nat. Sci.: Nanosci. Nanotechnol.* **5**, 023002 (2014).
- [42] X. Liu, S. K. Schnell, J. M. Simon, P. Krüeger, D. Bedeaux, S. Kjelstrup, A. Bardow, and T. J. H. Vlugt, *Int. J. Thermophys.* **34**, 1169 (2013).
- [43] M. Pozar, J.-B. Segulier, J. Guerche, R. Mazighi, L. Zoranic, M. Mijakovic, B. Kezic-Lovrinevic, F. Sokolic, and A. Perera, *Phys. Chem. Chem. Phys.* **17**, 9885 (2015).
- [44] A. Ben-Naim, *The Journal of Chemical Physics* **137**, 235102 (2012), <https://doi.org/10.1063/1.4772463>.
- [45] A. Ben-Naim, *J. Chem. Phys.* **138**, 224906 (2013).
- [46] T. L. Hill, *Thermodynamics of Small Systems* (Dover, 1963).
- [47] S. K. Schnell, X. Liu, J. Simon, A. Bardow, D. Bedeaux, T. J. H. Vlugt, and S. Kjelstrup, *J. Phys. Chem. B* **115**, 10911 (2011).
- [48] P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, T. J. H. Vlugt, and J.-M. Simon, *J. Phys. Chem. Lett.* **4**, 235 (2013).
- [49] P. Ganguly and N. F. A. van der Vegt, *J. Chem. Theory Comput.* **9**, 1347 (2013).
- [50] D. Mukherji, N. F. A. van der Vegt, and K. Kremer, *J. Chem. Theory Comput.* **8**, 3536 (2012).
- [51] T. E. de Oliveira, P. A. Netz, K. Kremer, C. Junghans, and D. Mukherji, *J. Chem. Phys.* **144**, 174106 (2016).
- [52] S. Weerasinghe and P. E. Smith, *J. Phys. Chem. B* **107**, 3891 (2003).
- [53] R. H. Stokes, *Aust. J. Chem.* **20**, 2087 (1967).
- [54] J. Milzetti, D. Nayar, and N. F. A. van der Vegt, *The Journal of Physical Chemistry B* **0**, null (0), PMID: 29342355, <https://doi.org/10.1021/acs.jpcc.7b11831>.
- [55] A. Rohatgi, “Webplotdigitizer 4.0,” <https://automeris.io/WebPlotDigitizer> (2017).
- [56] O. Miyakami, A. Saito, T. Matsuo, and K. Nakamura, *Biosci. Biotechnol. Biochem.* **61**, 466469 (1997).
- [57] H. Kokubo, J. Rösger, D. W. Bolen, and B. M. Pettitt, *Biophys. J.* **93**, 3392 (2007).