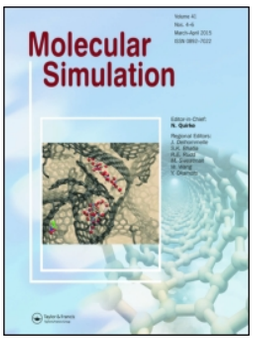

Molecular Simulation

ISSN: 0892-7022 (Print) 1029-0435 (Online) Journal homepage: https://www.tandfonline.com/loi/gmos20

Finite-size effects of Kirkwood–Buff integrals from molecular simulations

N. Dawass, P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, J. M. Simon & T. J. H. Vlugt

To cite this article: N. Dawass, P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, J. M. Simon & T. J. H. Vlugt (2018) Finite-size effects of Kirkwood–Buff integrals from molecular simulations, Molecular Simulation, 44:7, 599-612, DOI: 10.1080/08927022.2017.1416114

To link to this article: https://doi.org/10.1080/08927022.2017.1416114

Published online: 22 Dec 2017.

MOLECULAR SIMULATION, 2018
VOL. 44, NO. 7, 599–612
https://doi.org/10.1080/08927022.2017.1416114

Taylor & Francis
Taylor & Francis Group

Finite-size effects of Kirkwood–Buff integrals from molecular simulations

N. Dawass[a], P. Krüger[b,c], S. K. Schnell[d], D. Bedeaux[e], S. Kjelstrup[a,e], J. M. Simon[f] and T. J. H. Vlugt[a]

[a]Process & Energy Laboratory, Delft University of Technology, Delft, The Netherlands; [b]Graduate School of Science and Engineering, Molecular Chirality Research Center, Chiba University, Chiba, Japan; [c]ICB, UMR 6303 CNRS – Université de Bourgogne, Dijon, France; [d]Department of Materials Science and Engineering, Norwegian University of Science and Technology, NTNU, Trondheim, Norway; [e]PoreLab, Department of Chemistry, Norwegian University of Science and Technology, NTNU, Trondheim, Norway; [f]ICB, UMR 6303 CNRS – Université de Bourgogne, Dijon, France

ABSTRACT

The modelling of thermodynamic properties of liquids from local density fluctuations is relevant to many chemical and biological processes. The Kirkwood–Buff (KB) theory connects the microscopic structure of isotropic liquids with macroscopic properties such as partial derivatives of activity coefficients, partial molar volumes and compressibilities. Originally, KB integrals were formulated for open and infinite systems which are difficult to access with standard Molecular Dynamics (MD) simulations. Recently, KB integrals for finite and open systems were formulated (J Phys Chem Lett. 2013;4:235). From the scaling of KB integrals for finite subvolumes, embedded in larger reservoirs, with the inverse of the size of these subvolumes, estimates for KB integrals in the thermodynamic limit are obtained. Two system size effects are observed in MD simulations: (1) effects due to the size of the simulation box and the size of the finite subvolume embedded in the simulation box, and (2) effects due to computing radial distribution functions (RDF) from a closed and finite system. In this study, we investigate the two effects in detail by computing KB integrals using the following methods: (1) Monte Carlo simulations of finite subvolumes of a liquid with an analytic RDF and (2) MD simulations of a WCA mixture for various simulation box sizes, but at the same thermodynamic state. We investigate the effect of the size of the simulation box and quantify the differences compared to KB integrals computed in the thermodynamic limit. We demonstrate that calculations of KB integrals should not be extended beyond half the size of the simulation box. For finite-size effects related to the RDF, we find that the Van der Vegt correction (J Chem Theory Comput. 2013;9:1347) yields the most accurate results.

ARTICLE HISTORY
Received 12 September 2017
Accepted 5 December 2017

KEYWORDS
Kirkwood–Buff integrals; finite-size effects; molecular dynamics; thermodynamics

1. Introduction

The prediction of thermodynamic properties of fluids from local density fluctuations is relevant to many chemical and biological processes [1–4]. The Kirkwood–Buff (KB) theory provides a sound framework to connect molecular structure and density fluctuations in multicomponent liquid systems to macroscopic properties [5–9]. As a result, properties like the compressibility, partial molar enthalpies, partial molar volumes and derivatives of activity coefficients can be obtained from integrating radial distribution functions (RDFs) [10–14]. To compute these properties, accurate estimates of KB integrals are required [8,10]. Essentially, the theory by Kirkwood and Buff relates fluctuations in the number of molecules of an open system to integrals of RDFs. It is important to note that the KB theory is only applicable to isotropic systems [8]. The derivation of KB integrals is available in the original paper of Kirkwood and Buff [5]. For an infinitely large and open three-dimensional system, the KB integral for species α and β is defined as [5]:

$$G_{\alpha\beta}^{\infty} \equiv \int_0^{\infty} \mathrm{d}r 4\pi r^2 \left[g_{\alpha\beta}^{\infty}(r) - 1 \right], \tag{1}$$

where r is the radial distance and $g_{\alpha\beta}^{\infty}(r)$ is the RDF of species α and β, where the effects of the orientation are already integrated out [8,15]. Species α and β can be the same. To compute KB integrals from molecular simulation, RDFs of infinite and open systems have to be integrated. However, typical simulation techniques do not provide access to infinite systems as required by the KB theory. Moreover, it is generally difficult to simulate high-density fluids in open ensembles [16–20]. Several studies have simply truncated the integral of Equation (1) to finite distances when computing KB integrals from molecular simulations [21–25]. This approximation does not provide the actual integrals defined by Kirkwood and Buff but it does give an indication of its order of magnitude [8]. In other studies, the RDF from finite systems were extended to the thermodynamic limit, using mathematically involved methods [1,26]. For instance, Wedberag et al. [1] applied the Verlet extension [27] to extend $g(r)$ to large distances. These methods are difficult to apply to complex systems where internal degrees of freedom and molecular interactions should be considered.

Recently, we defined an expression for KB integrals for finite subvolumes, V, embedded in a larger reservoir [28]. The subvolume can exchange mass and energy with the surrounding and is grand canonical, which is the ensemble adopted by the theory of Kirkwood and Buff [5]. In molecular simulations, the simulation box, with volume L^3, acts as such a reservoir surrounding the subvolume (see Figure 1(a)). The size of the subvolume, R, is

CONTACT T. J. H. Vlugt t.j.h.vlugt@tudelft.nl

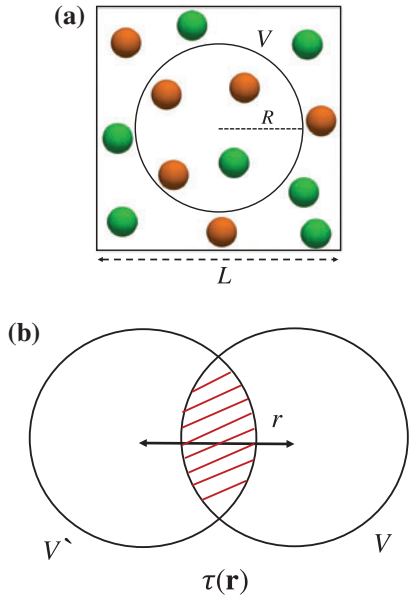

Figure 1. (Colour online) (a) Density fluctuations of two molecule types inside a spherical subvolume V, with radius R. This spherical subvolume is embedded inside a cubic simulation box with length L, and volume L^3. The finite simulation box acts as a reservoir surrounding the embedded subvolume V. $G^V_{\alpha\beta}$ is obtained using Equation (2) or Equation (3). The scaling of $G^V_{\alpha\beta}$ with the inverse size of the subvolume, $1/R$, is extrapolated to the thermodynamic limit to find $G^\infty_{\alpha\beta}$ (Equation (1)). (b) Schematic representation of the definition of the function $\tau(\mathbf{r})$ (Equation (11)), the overlap volume between two spheres separated by a distance r.

gradually increased and the KB integrals are computed for each subvolume size. The scaling of the KB integrals of finite subvolumes with $1/R$ is used to obtain integrals at the thermodynamic limit, $G^\infty_{\alpha\beta}$. As in the original KB theory, the starting point for the derivation of the finite-volume KB integrals is molecule fluctuations in a finite and open subvolume, V,

$$G^V_{\alpha\beta} \equiv V\frac{\langle N_\alpha N_\beta\rangle - \langle N_\alpha\rangle\langle N_\beta\rangle}{\langle N_\alpha\rangle\langle N_\beta\rangle} - \frac{V\delta_{\alpha\beta}}{\langle N_\beta\rangle}, \tag{2}$$

where N_α and N_β are the number of molecules of type α and β, inside the subvolume V and the brackets $\langle\cdots\rangle$ denote the ensemble average in an open system. $\delta_{\alpha\beta}$ is the Kronecker delta (equal to 1 when $\alpha = \beta$ and is zero otherwise).

Starting from Equation (2), the KB integrals for finite subvolumes in terms of the RDF of the system are derived in this work. The final working expression to compute the KB integrals from molecular simulations equals [28]

$$G^V_{\alpha\beta} = \int \mathrm{d}r\left[g^\infty_{\alpha\beta}(r) - 1\right]w(r) \tag{3}$$

where $g^\infty_{\alpha\beta}(r)$ is the RDF in the thermodynamic limit and $w(r)$ is a weighting function, which depends parametrically on the dimensionality and shape of the subvolume V. For a spherical subvolume, $w(r)$ is known analytically [28] and depends only on the dimensionless distance $r/2R$, where R is the sphere radius. For volumes of arbitrary shape, $w(r)$ may be computed numerically. Note that the integral in Equation (3) has a finite upper bound, because $w(r)$ is zero for r larger than the maximum distance in the subvolume V, which is $2R$ in the case of spheres. Equation (3) can be used to compute the integrals from RDFs obtained from molecular simulations of closed and finite simulation boxes, as in the case of the NVT ensemble.

When computing KB integrals from molecular simulations of closed, finite systems one must be aware of two system-size effects that originate from: (1) computing finite-volume KB integrals from subvolumes embedded in closed and finite simulation boxes [29] and (2) using RDFs obtained from finite and closed simulation boxes, in contrast to open, infinite systems as required by the KB theory. For the second effect, RDFs from open systems, $g^\infty(r)$, can be estimated from $g(r)$ obtained from molecular simulations of closed and finite systems. Previously, Krüger et al. [28] used a correlation based on expanding the difference between $g(r)$ and $g^\infty(r)$ in a Taylor series in the inverse of the number of molecules. Ganguly and van der Vegt [30] suggested a correction based on the excess or depletion of molecules within a distance, r, around a central molecule. Cortes-Huerto et al. introduced a correlation to compute the KB integrals that includes a correction to the RDF finite-size effects [31].

The objective of this work is to investigate finite-size effects associated with the computation of KB integrals from molecular simulations (e.g. MD) of finite and closed systems. For the two finite-size effects studied, we present the inaccuracies as a function of the simulation box size. Also, we investigate subvolume sizes and provide the distances up to which RDFs need to be calculated. The paper is organized as follows. In Section 2, a detailed theoretical derivation of the finite-volume KB integrals (Equation (3)) is presented, which was omitted from the our original paper due to size restrictions [28]. The finite-size effects are studied using two sets of simulations. Monte Carlo (MC) sampling of KB integrals from subvolumes embedded in larger simulation box. The KB integrals in the thermodynamic limit, and the inaccuracies associated with these computations are investigated for various simulation box sizes. Furthermore, we demonstrate the finite-size effects related to the RDF by performing Molecular Dynamics (MD) simulation of a Weeks–Chandler–Andersen [32] (WCA) mixture. Other than varying the simulation box size used to obtain KB integrals, we also test various RDF correction methods to estimate RDFs of infinite systems, $g^\infty(r)$. Sections 3.1 and 3.2 provide details regarding the simulation methods and specifications of the systems studied. In Section 4, we present two main sets of results: (1) the appropriate system sizes from which KB integrals could be extrapolated and the magnitude of the inaccuracies made when simulating finite subvolumes and (2) the numerical inaccuracies resulting from computing the RDFs in a closed and finite simulation boxes, and the effect of applying the discussed correction methods of the RDFs. Our findings are summarized in Section 5.

2. Theory

In this section, we present the theoretical derivation of Equation (3) which allows for the computation of KB integrals for finite subvolumes. Similar to the original Kirkwood–Buff theory [5] for infinitely large and open systems, Krüger et al. [28] derived an expression that relates local density fluctuations inside the subvolume with the integral of the RDF of the system. This was achieved by first considering the average densities and the fluctuations in density of the open subvolume, V, embedded in a large reservoir (see Figure 1(a)). The open subvolume, V is grand canonical. The system is characterized by the following variables: temperature (T), volume of the subvolume (V), and chemical potentials (μ_α and μ_β). In this ensemble, we consider the average number of molecules, $\langle N_\alpha \rangle$, and the average number of α and β pairs, $\langle N_\alpha N_\beta \rangle$, expressed as integrals of the one molecule density ($\rho_\alpha^{(1)}(\mathbf{r}_1)$) and the two molecule density ($\rho_{\alpha\beta}^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$),

$$\int_V \rho_\alpha^{(1)}(\mathbf{r}_1)\mathrm{d}\mathbf{r}_1 = \langle N_\alpha \rangle \tag{4}$$

$$\int_V \int_V \rho_{\alpha\beta}^{(2)}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 = \langle N_\alpha N_\beta \rangle - \delta_{\alpha\beta}\langle N_\alpha \rangle. \tag{5}$$

Integration of the local densities over the subvolume V yields the average number of molecules in the grand canonical ensemble [8]. Subsequently, the density fluctuations in the subvolume V are expressed as

$$\begin{aligned}\int_V \mathrm{d}\mathbf{r}_1 \int_V \mathrm{d}\mathbf{r}_2 [\rho_{\alpha\beta}^{(2)}(\mathbf{r_1}, \mathbf{r_2}) - \rho_\alpha^{(1)}(\mathbf{r}_1)\rho_\beta^{(1)}(\mathbf{r}_2)] \\ = \langle N_\alpha N_\beta \rangle - \langle N_\alpha \rangle \langle N_\beta \rangle - \delta_{\alpha\beta}\langle N_\alpha \rangle.\end{aligned} \tag{6}$$

For fluid systems, $\rho_\alpha^{(1)}(\mathbf{r}_1)$ and $\rho_{\alpha\beta}^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$, due to translational and rotational invariance of the molecules, can be replaced by c_α, and $c_\alpha c_\beta g_{\alpha\beta}(r_{12})$, respectively. Here, c_α is the macroscopic number density given by $c_\alpha = \langle N_\alpha \rangle / V$. The function $g_{\alpha\beta}(r_{12})$ is the RDF and $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$. For a finite multicomponent fluid, the integral, $G_{\alpha\beta}^V$, is defined by simply dividing Equation (6) by $c_\alpha c_\beta V$:

$$\begin{aligned}G_{\alpha\beta}^V &\equiv \frac{1}{V}\int_V \int_V \left[g_{\alpha\beta}(r_{12}) - 1\right] \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \\ &\equiv V\frac{\langle N_\alpha N_\beta \rangle - \langle N_\alpha \rangle \langle N_\beta \rangle}{\langle N_\alpha \rangle \langle N_\beta \rangle} - \frac{V\delta_{\alpha\beta}}{\langle N_\beta \rangle}.\end{aligned} \tag{7}$$

Here, the R.H.S (Equation (2)) and the L.H.S. of the equation are both equal to Equation (3) discussed in Section 1. Equation (3) is expressed in terms of integral of the RDF of the system, that could be obtained from molecular simulation. In the limit $V \to \infty$ and homogeneous conditions, the double integrals on the L.H.S. of Equation (7) can be reduced to a single integral by applying the transformation: $\mathbf{r}_2 \to \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, which yields the original expression for the KB integral for infinitely large systems (Equation (1)). However, for a finite subvolume, V, applying this transformation is not possible since the domain of integration over $\mathbf{r}$ depends on $\mathbf{r}_1$. In this case, the double volume integrals in Equation (7) are reduced to a single radial integral by rewriting the L.H.S. of Equation (7) as

$$G_{\alpha\beta}^V = \int \mathrm{d}r w(r) h_{\alpha\beta}(r), \tag{8}$$

where $h_{\alpha\beta}(r) = g_{\alpha\beta}(r) - 1$, and $w(r)$ is a purely geometric function characteristic of the volume V defined as

$$w(r) \equiv \frac{1}{V}\int_V \mathrm{d}\mathbf{r}_1 \int_V \mathrm{d}\mathbf{r}_2 \delta(r - |\mathbf{r}_1 - \mathbf{r}_2|) \tag{9}$$

Once the function $w(r)$ is known, the two-dimensional integral of Equation (7) reduces to the one-dimensional integral of Equation (8), and the expression for KB integrals for finite subvolumes is obtained. For the calculation of $w(r)$, we first rewrite the L.H.S. of Equation (7) as

$$G_{\alpha\beta}^V = \int_V \mathrm{d}\mathbf{r}\tau(\mathbf{r})h_{\alpha\beta}(\mathbf{r}), \tag{10}$$

where the integral is over all space and

$$\tau(\mathbf{r}) \equiv \int_V \mathrm{d}\mathbf{r}_1 \int_V \mathrm{d}\mathbf{r}_2 \delta(\mathbf{r} - \mathbf{r}_1 + \mathbf{r}_2). \tag{11}$$

The function $\tau(\mathbf{r})$ has a simple geometrical interpretation: it is the overlap between the subvolume V and the same subvolume V shifted by $\mathbf{r}$. This may be seen by making the variable substitution $\mathbf{r}_2' = \mathbf{r}_2 + \mathbf{r}$ which yields $\tau(\mathbf{r}) \equiv \int_V \mathrm{d}\mathbf{r}_1 \int_{V'} \mathrm{d}\mathbf{r}_2' \delta(\mathbf{r}_2' - \mathbf{r}_1)$, where V' is the subvolume V shifted by $\mathbf{r}$ (see Figure 1(b)). The function $w(r)$ is obtained from $\tau(\mathbf{r})$ by integrating over 4π solid angle (Ω) and dividing by V. We have

$$w(r) = \frac{1}{V}\int \mathrm{d}\mathbf{r}'\tau(\mathbf{r}')\delta(r - |\mathbf{r}'|) = \frac{r^{D-1}}{V}\int \mathrm{d}\Omega\, \tau(\mathbf{r}) \tag{12}$$

where D is the dimensionality of space. In the following, we consider for V hyperspheres of radius R, where by symmetry, the overlap volume does not depend on Ω, so $\tau(\mathbf{r}) = \tau(r)$. The volume of a hypersphere is $V = R^D \int \mathrm{d}\Omega / D$ which, together with Equation (12), yields

$$w(r) = \tau(r)\, D\, r^{D-1}/R^D. \tag{13}$$

The overlap volumes $\tau(r)$ of hyperspheres in D=1–3 dimensions (i.e. segment, circle and sphere) are well known [33]. From these, the corresponding functions $w(r)$ are obtained using Equation (13) and listed in Table 1. Using $w(r)$ and the L.H.S. of Equation (7), we arrive at the final expression for KB integrals for finite subvolumes (Equation (3)).

Table 1. Geometrical functions $w(x)$, $x = r/(2R)$, for D-dimensional hyperspheres of radius R. The expressions are valid for $x < 1$. For $x > 1$ we have $w = 0$ in all cases. Note that in our first paper [28], a factor of 2 was omitted in the expression of $w(x)$ for $D = 1$ because we took R as the segment length. Here, we take $2R$ which ensures consistency between all dimensions.

	$w(x)$
1D	$2(1-x)$
2D	$4r(\arccos(x) - x\sqrt{1-x^2})$
3D	$4\pi r^2(1 - 3x/2 + x^3/2)$

3. Methodology

3.1. Subvolumes finite-size effects

Computing KB integrals using molecular simulations of finite simulation boxes results in finite-size effects. These effects impact the accuracy of the computed integrals from finite subvolumes, and hence the integrals at the thermodynamic limit. In this section, we show the system, and simulation method, used to study the finite-size effects. Also, we relate the surface area of the subvolume embedded in the simulation box to the finite-size effects of the subvolume.

To deal solely with effects originating from the finite size of the system, and not RDF related effects, we study a fluid that is described by an analytic RDF. The following model describes a liquid with molecules of diameter σ [27,34]:

$$h(r) = g(r) - 1 = \begin{cases} \frac{3/2}{r/\sigma} \exp\left[\frac{1-r/\sigma}{\chi}\right] \cos\left[2\pi\left(\frac{r}{\sigma} - \frac{21}{20}\right)\right] & \frac{r}{\sigma} \geq \frac{19}{20}, \\ -1, & \frac{r}{\sigma} < \frac{19}{20} \end{cases} \quad (14)$$

where χ is the length scale at which the fluctuations of $h(r)$ decay. As we will consider a pure component fluid, for simplicity the indices α and β are dropped in this section. Also, the parameter σ is set to unity throughout this work, so we use L for the size of the box instead of L/σ. We show how finite KB integrals (Equation (3)) scale with the inverse of the size of the spherical subvolume, V (the subvolume can have any other shape but in this study, we choose to only consider spherical subvolumes).

The finite-size effects of the subvolume emerge from pairs of molecules, where one molecule is inside the subvolume, V, and the other is outside V. The contribution of these molecular pairs to the KB integrals decreases as the size of the subvolume gets smaller. This relation could be shown using the following function [28],

$$F \equiv \int_V \mathbf{dr}_1 \int_{L^3-V} \mathbf{dr}_2 h(r_{12}) = \int_V \mathbf{dr}_1 \int_{L^3} \mathbf{dr}_2 h(r_{12}) - \int_V \mathbf{dr}_1 \int_V \mathbf{dr}_2 h(r_{12}), \quad (15)$$

where non-zero contributions to F originate from molecule pairs where one molecule is inside V and the other one outside V. Assuming a finite correlation length ζ for a layer surrounding the subvolume, we have $h(r_{12}) \approx 0$ for $r_{12} > \zeta$, (see Figure 2(a)). The volume of this layer, and thus F, increases linearly with the surface area, A, of the spherical subvolume (for radius of the sphere larger than ζ). In the case of an infinitely large system, the homogeneous conditions allow for the substitution $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ in the integral over L^3, resulting in:

$$F = \int_V \mathbf{dr}_1 \int_{L^3} \mathbf{dr} h(r) - \int_V \mathbf{dr}_1 \int_V \mathbf{dr}_2 h(r_{12}) \approx VG^\infty - VG^V, \quad (16)$$

where G^∞ is the KB integral for an infinite volume. As F scales linearly with the surface area A, the difference $(G^\infty - G^V)$ scales as A/V, i.e. inversely with the linear dimension of the subvolume.

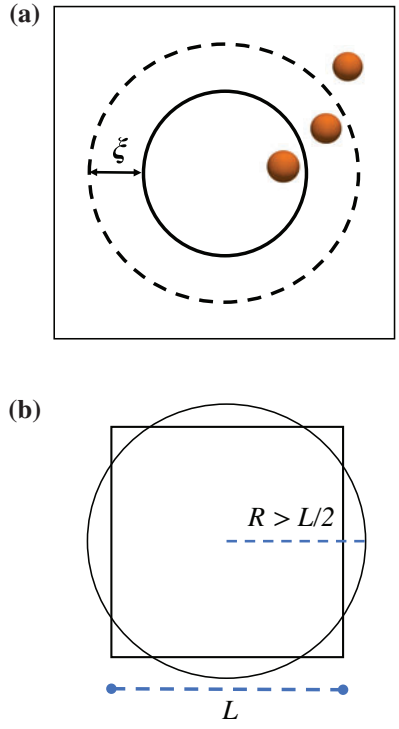

Figure 2. (Colour online) Schematic representation of finite-size effects in Section 3.1. (a) related to the function F in Equation (15). The only contributions to F are from molecule pairs with one molecule (1) inside the subvolume V and the other molecule (2) in a surrounding layer of thickness ζ, where ζ is the correlation length of the fluctuations of the RDF (Equation(14)). Molecules outside this layer ($r_{12} > \zeta$) do not contribute to F. (b) Schematic representation of the effect of extending R beyond half the simulation box length, $L/2$. The volume of the sphere no longer equals $\frac{4}{3}\pi R^3$, and the sphere caps falling outside the box has to be subtracted. Equation (17) and (18) provide the area and volume of the truncated spherical subvolume.

To quantify the inaccuracies resulting from these subvolume finite- size effects, the values of $G^V_{\alpha\beta}$ are computed from simulations of different sizes of the subvolume, R, for a specific simulation box size L. The KB integrals, $G^\infty_{\alpha\beta}$, are obtained from extrapolating the subvolumes KB integrals to infinite subvolume size ($R \to \infty$). The KB integrals are compared to integrals obtained from the direct numerical integration of Equation (3), with the RDF at each distance computed analytically from Equation (14). To quantify the inaccuracies in computed KB integrals due to finite-size effects, we evaluate the differences between numerical integration of Equation (3) and MC sampling of the double integral of Equation (7). Furthermore, we examine the distances up to which the computations of the KB integral are performed (i.e the appropriate subvolume sizes). In molecular simulation, RDFs are typically computed up to half the length of the simulation box and as a result the computed KB integrals are limited to this range. However, in simulations, one can in principle extend r up to $\frac{\sqrt{3}}{2}$ of the box length [35]. In this work, the RDF is extended up to $\frac{\sqrt{2}}{2}$ of the box length (the range $\frac{\sqrt{2}}{2} < L < \frac{\sqrt{3}}{2}$ involves complex calculations that will not be considered further). In the results section, we show how this extension affects the computations of KB integrals. When the radius of the subvolume, R, is larger than half the box size,

the surface area and volume of the subvolume are computed by disregarding the sphere caps falling outside the simulation box (see Figure 2(b)). Using $R - L/2$ for the height of the cap falling outside the simulation box, we obtain the following equations for the truncated surface area, $A_{\text{trunc:}}$, and truncated volume, $V_{\text{trunc:}}$ [36],

$$A_{\text{trunc:}} = 4\pi R\left[R - 3\left(R - \frac{L}{2}\right)\right], \qquad \frac{1}{2} < \frac{R}{L} < \frac{\sqrt{2}}{2} \tag{17}$$

$$V_{\text{trunc:}} = 2\pi\left[\frac{2}{3}R^3 - \left(R - \frac{L}{2}\right)^2\left(2R + \frac{L}{2}\right)\right], \frac{1}{2} < \frac{R}{L} < \frac{\sqrt{2}}{2} \tag{18}$$

To compute the KB integrals of the system using the RDF of Equation (14), MC sampling is performed to numerically evaluate Equation (7). For simplicity, a brute-force random sampling algorithm is used instead of importance sampling MC. The simulations follow the following algorithm:

(1) Specify the size of the cubic simulation box, L, the size of the spherical subvolume, with radius R, placed in the centre of the box, and the RDF parameters $\emptyset$ and σ. In this study, we fix $\sigma = 1$, and test different values of χ ($\chi = 1, 2,$ and 4).
(2) Select two random points inside the simulation box, p_1 and p_2, accept the points if they both fall inside the sphere otherwise, generate new points.
(3) Select a random point, p_3, and accept if the point is outside the sphere. Otherwise, choose a new point p_3.
(4) Calculate the distance between points p_1 and p_2 (r_{12}), and points p_1 and p_3 (r_{13}). Find $h.r_{12}/$ and $h.r_{13}/$ using Equation (14).
(5) Performing steps 2 and 3 results in one cycle. After a sufficient number of cycles (in this study we perform 10^{11} cycles), compute the average of the integrals $\int_V \int_V$ and $\int_V \int_{L^3-V}$ (from Equation (7) and Equation (15)) over the RDF.
(6) Finally, calculate $G^V_{\alpha\beta}$ (Equation (7)) and F (Equation (15)) using the sampled averages of the integrals $\int_V \int_V$ and $\int_V \int_{L^3-V}$.

3.2. Finite-size effects of the RDF

While our formulation enables the calculation of KB integrals from finite subvolumes, the used RDF has to be of that of an open and infinite system, $g^\infty(r)$. Using $g(r)$ from molecular simulation of a closed system results in a systematic error that affects the accuracy of the KB integrals, thus, the $g(r)$ has to be corrected. There are several approaches available to estimate $g^\infty(r)$ from RDFs computed in closed systems. In this section, we present the RDF correction methods that we consider in this work. To study the RDF correction methods, MD simulations of WCA molecules are performed. The system conditions and the simulation details are provided in this section.

3.2.1. Van der Vegt correction

When computing KB integrals using finite systems, Ganguly and van der Vegt [30] observe a drifting asymptote due to the asymptotic behaviour of RDFs in finite systems. Specifically, this asymptotic behaviour of the RDF does not converge to one. These authors proposed that the RDF could be corrected using a correlation that takes into account the excess, or depletion of the bulk density of molecule-type α around molecule of type β at a distance r due to the finite-size of the system (simulation box). The bulk density of molecules β is compensated by computing the excess or depletion of number of molecules of species β inside the considered subvolume, V. The subvolume is formed by taking a distance r from the central molecule of type α. The correlation takes into account the depletion of molecule-type β around a molecule of type α, $\Delta N_{\alpha\beta}(r)$, and provides the corrected RDF, $g^{\text{vdV}}_{\alpha\beta}(r)$:

$$g^{\text{vdV}}_{\alpha\beta}(r) = g_{\alpha\beta}(r)\frac{N_\beta\left(1 - \frac{V}{L^3}\right)}{N_\beta\left(1 - \frac{V}{L^3}\right) - \Delta N_{\alpha\beta}(r) - \delta_{\alpha\beta}}, \tag{19}$$

where $g_{\alpha\beta}(r)$ is the RDF obtained from a simulation of a closed system. For infinitely large and open systems, $g_{\alpha\beta}(r)$ and $g^{\text{vdV}}_{\alpha\beta}(r)$ are equal. N_β is the number of molecules of type β, L^3 is the volume of the closed simulation box (which is assumed to be cubic), and $\delta_{\alpha\beta}$ is the Kronecker delta. V is the subvolume that surrounds a molecule of type α, with radius r. This volume is calculated according to whether r extends to outside the simulation box or not [36],

$$V(r) = \begin{cases} \frac{4}{3}\pi r^3, & \frac{r}{L} < \frac{1}{2} \\ \frac{4}{3}\pi r^3\left(\frac{3-4R/L}{2R/L}\right), & \frac{1}{2} < \frac{r}{L} < \frac{\sqrt{2}}{2}. \end{cases} \tag{20}$$

The excess or depletion of molecule-type β around a molecule of type α, $\Delta N_{\alpha\beta}(r)$, can be calculated from, depending on whether r extends to outside the simulation box or not [36]:

$$\Delta N_{\alpha\beta}(r) = \begin{cases} \int_0^r \mathrm{d}r' 4\pi r'^2 \rho_\beta\left[g_{\alpha\beta}(r') - 1\right], & \frac{r}{L} < \frac{1}{2} \\ \int_0^r \mathrm{d}r' 4\pi r'^2 (1 - 3 + \frac{3L}{2r'})\rho_\beta\left[g_{\alpha\beta}(r') - 1\right], & \frac{1}{2} < \frac{r}{L} < \frac{\sqrt{2}}{2} \end{cases} \tag{21}$$

3.2.2. Inverse-N finite-size correction

A simple method to correct for the finite-size effect seen in radial distribution functions was presented by Krüger et al. [28], where the difference between $g(r)$ and $g^\infty(r)$ is expanded in a Taylor series in $1/N$ [37]:

$$g^{N_1}_{\alpha\beta}(r) = g^\infty_{\alpha\beta}(r) + \frac{c(r)}{N_1} + \mathcal{O}\left(\frac{1}{N_1^2}, r\right), \tag{22}$$

here, $g^{N_1}_{\alpha\beta}(r)$ is the RDF for a closed system with N_1 molecules, $g^\infty_{\alpha\beta}(r)$ is the RDF corrected for the finite-size effect, and $c(r)$ is a function that describes deviation from a open system. The function $c(r)$ is usually not known, but can be estimated using two systems with different sizes in the same thermodynamic

state (same density/pressure, temperature, and composition). The term $\mathcal{O}(\frac{1}{N_1^2}, r)$ is the error associated with the correlation in Equation (22), which is a function of the number of molecules used as well as r. From Equation (22), the corrected $g_{\alpha\beta}^{\infty}(r)$ can be expressed as

$$g_{\alpha\beta}^{\infty}(r) = \frac{N_1 g_{\alpha\beta}^{N_1}(r) - N_2 g_{\alpha\beta}^{N_2}(r)}{N_1 - N_2}, \tag{23}$$

where the subscripts 1 and 2 refer to two systems with different number of molecules, but with same density/pressure, temperature and composition. This method of correcting for the finite-size effect is straightforward, but requires two different set of simulations, with different box sizes and number of molecules. The box sizes should not be too different and the resulting $g_{\alpha\beta}^{\infty}(r)$ can only be extended to the size of the smallest system of the two systems. Another shortcoming of this method is related to the numerical accuracy arising from subtracting two numbers of the same magnitude, both in numerator and denominator. The resulting numerical instabilities are increased when using two system sizes that are very close to each other [38].

3.2.3. Cortes-Huerto et al. correction

Another RDF correction is proposed by Cortes-Huerto et al. [31]. These authors define KB integrals from finite systems in terms of number of molecules fluctuations as in the work of Krüger et al. [28]. However, their study consider the fluctuations inside a cubic subvolume. The use of a different subvolume geometry should not affect the values of the KB integrals at the thermodynamic limit, $G_{\alpha\beta}^{\infty}$ [29]. The KB integrals are also defined in terms of integrals over the RDF of the system (Equation (7)). To compute KB integrals using finite volumes, these authors modify the L.H.S. of Equation (7) to include RDF finite effects and finite subvolumes effects. For the RDF finite effects, a correction based on a relation from the work of Ben-Naim [8] is used, that strictly only applies when $r \to \infty$,

$$g_{\alpha\beta}(r) = g_{\alpha\beta}^{\infty}(r) - \frac{1}{L^3}\left(\frac{\delta_{\alpha\beta}}{\rho_\alpha} + G_{\alpha\beta}^{\infty}\right). \tag{24}$$

Using this relation means that the difference between $g_{\alpha\beta}(r)$ and $g_{\alpha\beta}^{\infty}(r)$ in Equation (24) is assumed to be independent of r (for all values of r). In the results section, we compare the validation of all RDF corrections over the whole range of r. When including the RDF correction (Equation (24)) in the L.H.S. of Equation (7), the following expression for the finite integrals is obtained [31],

$$G_{\alpha\beta}(V, L^3) = \frac{1}{V}\int_V \int_V \left[g_{\alpha\beta}^{\infty}(r_{12}) - 1\right] \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 - \frac{V}{L^3}\left(\frac{\delta_{\alpha\beta}}{\rho_\alpha} + G_{\alpha\beta}^{\infty}\right). \tag{25}$$

The effect of the finite size of the subvolume, V, is accounted for by considering the boundary effects considered in the function F (Equation (16)). The double integral in Equation (25), $\int_V \int_V$, is expanded to account for the other integration domains, $\int_V \int_{L^3}$ and $\int_V \int_{L^3-V}$. As explained earlier in Section 3.1, particles in a layer outside V in the volume $L^3 - V$ contribute to the double integral $\int_V \int_V$. This contribution scales with the surface area, S, of the subvolume. Considering the finite subvolume effect and using $S/V \propto 1/V^{1/3}$,

$$G_{\alpha\beta}(V, L^3) = \frac{1}{V}\int_V \int_{L^3} \left[g_{\alpha\beta}^{\infty}(r_{12}) - 1\right] \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 - \frac{V}{L^3}\left(\frac{\delta_{\alpha\beta}}{\rho_\alpha} + G_{\alpha\beta}^{\infty}\right) + \frac{C_{\alpha\beta}}{V^{1/3}}, \tag{26}$$

where $C_{\alpha\beta}$ is a constant that is unique for each thermodynamic state of the system. Cortes-Huerto et al. [31] restrict the volume V between V_ζ and L^3, where $V_\zeta = 4\pi\zeta^3/3$. As a result of the values of r being always larger than ζ, the value of $g_{\alpha\beta}^{\infty}(r_{12})$ is set to one. Additionally, it is assumed that the system is transitionally invariant and the transformation $\mathbf{r}_2 \to \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ applies which transforms the integrals in Equation (26) to the ones in the original KB integrals expression (Equation (3)). Applying these assumptions, the expression for KB integrals for finite subvolumes according to Cortes-Huerto et al. [31] is,

$$G_{\alpha\beta}(V, L^3) = G_{\alpha\beta}^{\infty}\left(1 - \frac{V}{L^3}\right) - \frac{V}{L^3}\frac{\delta_{\alpha\beta}}{\rho_\alpha} + \frac{C_{\alpha\beta}}{V^{1/3}}. \tag{27}$$

In this work, Equation (27) is used to find $G_{\alpha\beta}^{\infty}$ from simulations of finite subvolumes. The computed KB integrals are compared to those computed by other correction methods. Details of the system studied and of the MD simulations are provided in the following section.

3.2.4. Simulations details

To study effects resulting from computing RDFs from simulation of finite and closed systems, we examine a binary mixture interacting using the WCA potential [32] where the Lennard-Jones (LJ) potential is shifted and truncated at $2^{1/6}\sigma$. The WCA mixture is simulated in the *NVT* ensemble using the Nose–Hoover thermostat [39]. The MD package LAMMPS [40] is used to perform the simulations, with 1 million initialization timesteps and 5 million integration timesteps for each run. A timestep of 0.001 in LJ reduced units, which are the units used for other variables, are used. All simulations were performed for the same system properties with: $\sigma_{11} = \sigma_{12} = \sigma_{22} = 1.0$, $\epsilon_{11} = \epsilon_{22} = 1.0$, and $\epsilon_{12} = 0.1$. Additionally, the same thermodynamic condition was maintained for all system sizes. A mixture composition of $x_1 = 0.75$ was used along with the following reduced values: $T^* = 1.8$ and $\rho^* = 0.7$. The box length, L, and number of molecules, N, were varied to obtain $g_{11}(r), g_{12}(r), g_{22}(r)$ as a function of system size. For all box sizes, $g_{\alpha\beta}(r)$ were computed up to distances of $(\sqrt{2}/2)L$. The RDFs are used to calculate finite subvolumes KB integrals (Equation (3)) which are then extrapolated to the thermodynamic limit to obtain G_{11}^{∞}, G_{12}^{∞}, and G_{22}^{∞}, respectively. To evaluate the KB integrals in the thermodynamic limit for the system studied, simulations of a large system are performed. Specifically, G_{11}^{∞}, G_{12}^{∞}, and G_{22}^{∞} are evaluated from averaging the results of five simulations for a box with $L = 80$. Each simulation was initialized with a different configuration, thus allowing for computing the standard deviation for each variable.

4. Results and discussion

4.1. Subvolumes finite-size effects

KB integrals (G^{∞}) are computed for the liquid with the analytic RDF model in Equation (14), using the MC algorithm discussed in Section 3.1. The KB integrals for finite subvolumes, G^{V} (in this section, we drop the indices $\alpha\beta$ since a pure fluid is studied) are computed using simulation boxes with the following lengths, $L = 7.5, 10, 15, 20, 40$ and 50. For each box size, spherical subvolumes with R up to $(\sqrt{2}/2)L$ are used. The KB integrals for finite subvolumes, G^{V} scales with the inverse of the sphere size, $1/R$. For each box size, the linear part of the G^{V} scaling is extrapolated up to $1/R \to 0$, to find G^{∞}. Figure 3(a) shows the scaling behaviour for the case of simulating the system when $\chi = 2$. The regime where the scaling is linear depends on the size of the simulated box. Larger simulation boxes provide longer linear regimes. The accuracy of the computations of the KB integrals at the thermodynamic limit, G^{∞}, depends on the size of the simulation box. The computed KB integrals from the MC simulations are compared to KB integrals in the thermodynamic limit, $G^{\infty,\text{num}}$, obtained by numerically integrating Equation (3) up to very large distances. The differences between the numerically integrated KB integrals, $G^{\infty,\text{num}}$, and KB integrals from simulations are computed using

$$\text{Difference\%} = \frac{|G^{\infty,\text{num}} - G^{\infty}|}{|G^{\infty,\text{num}}|} * 100\%. \qquad (28)$$

In Table 2, the differences (%) are listed when using the system sizes shown in Figure 3 and for three χ parameters, $\chi = 1$, 2, and 4. For these parameters, the values of $G^{\infty,\text{num}}/\sigma^3$ are -1.785, -2.041 and -2.172 respectively. The values of G^{∞}/σ^3 were obtained by extrapolating the linear part of the lines in Figure 3, which extend until $R = L/2$ (indicated by a dot for each line). In general, for all fluctuation length parameters, χ, the difference decreases with the system size. For boxes with $L = 7.5$ or 10, the difference is equal to or larger than 1%, but the deviation decreases by approximately 75 and 90% when increasing L to 15σ and 20σ, respectively. Finally, obtaining the KB integrals from boxes with $L = 40\sigma$ and $L = 50\sigma$ results in marginal differences. The subvolumes finite-size effect is shown more clearly when plotting A/V instead of $1/R$ as shown in Figure 3(b). This is due to the fact that the linear scaling of G^{V} with $1/R$ is correct up to $R = L/2$ ($A/V \sim 1/R$). When R is larger than $L/2$, parts of the sphere fall outside the simulation box, and for these distances, the ratio $A_{\text{trunc:}}/V_{\text{trunc:}}$ (Equation (17) and (18)) is used instead of A/V.

Table 2. Differences (%), calculated using Equation (28), between KB integrals obtained from direct numerical integration of Equation (3), $G^{\infty,\text{num}}/\sigma^3$, and integrals computed using MC simulations of various simulation box sizes, G^{∞}/σ^3. The system used is the fluid described by the analytic RDF of Equation (14). The values of the KB integrals from numerical integration, $G^{\infty,\text{num}}/\sigma^3$, are -1.785, -2.041 and -2.172 for $\chi = 1$, $\chi = 2$, and $\chi = 4$, respectively. The values of G^{V}/σ^3 obtained from MC simulations of finite simulation boxes are extrapolated to the thermodynamic limit to obtain G^{∞}/σ^3.

Box length(L)	$\chi = 1$	$\chi = 2$	$\chi = 4$
7.5	1.979	2.262	2.408
10	1.052	1.042	1.000
15	0.323	0.228	0.140
20	0.154	0.105	0.069
40	0.005	0.036	0.007
50	0.003	0.017	0.055

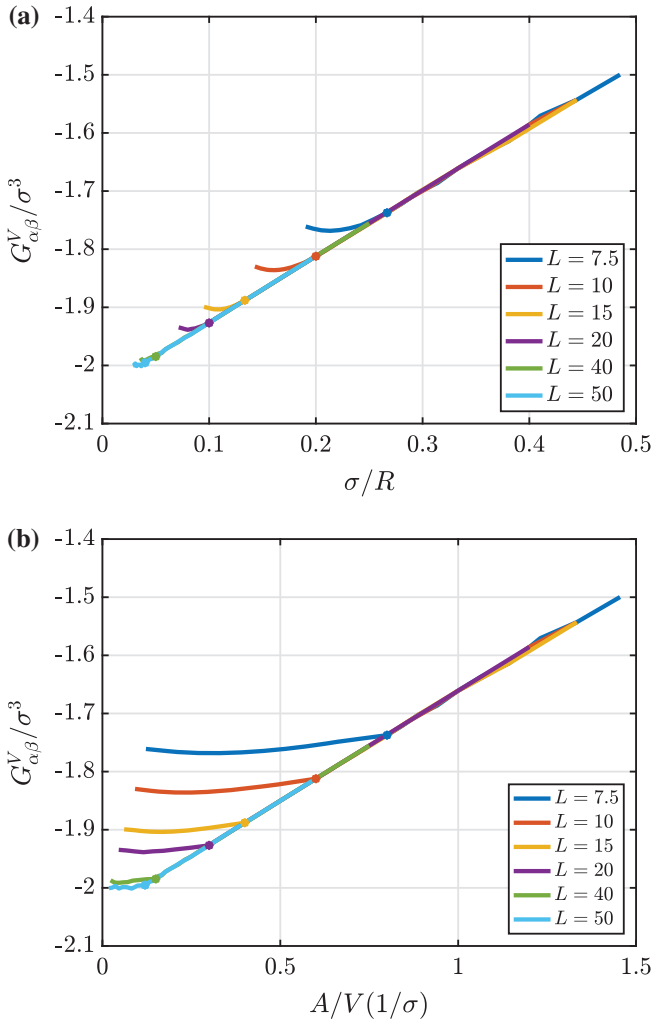

Figure 3. (Colour online) KB integrals from finite subvolumes of the fluid described with the RDF of Equation (14), with $\chi = 2$. G^{V}/σ^3 scales with: (a) the inverse of the sphere size $1/R$, and (b) the ratio of the area of the sphere to its volume A/V. The integrals are computed using MC simulations for different lengths of the simulation box, L, and different radii, R, of the spherical subvolume. The dots show the points where $R = L/2$ for each box size.

Another important observation made from Figure 3(b) is related to the size of subvolume used to compute KB integrals. For each box size, the dot in the G^{V} line indicates the point where the radius of the spherical subvolume equals $L/2$. As shown in this figure, increasing R beyond this value will not extend the linear regime. This finding is manifested when looking at the correlation between G^{V} and the ratio between the area and volume of the sphere (A/V) which is presented in Figure 3(b). As shown by Equation (16), G^{V} scales with A/V, but this scaling does not continue when R is larger than half the simulation box size. When the size of the subvolume extends beyond the simulation box the number of molecules surrounding the embedded subvolume decreases greatly, and the subvolume cannot be considered as grand canonical. Thus, extending the computations of KB integrals beyond $L/2$ is not necessary and does not improve the accuracy of the computed

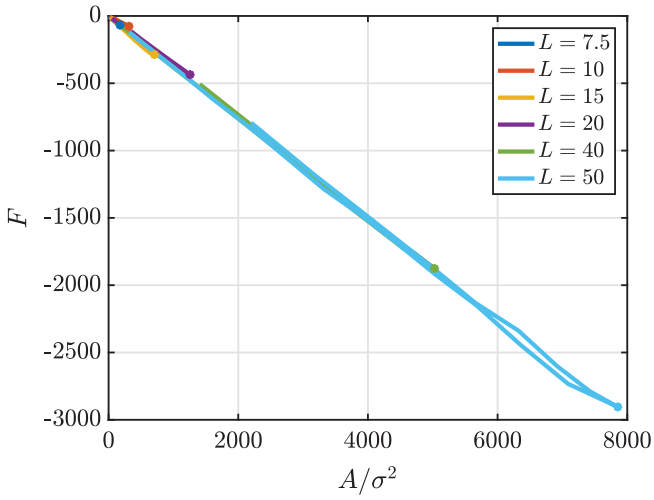

Figure 4. (Colour online) The function F (Equation (16)) of the fluid described by the RDF of Equation (14), with $\chi = 2$. The function scales linearly with the surface area, A/σ^2, for all the box sizes studied. The dots show the points where $R = L/2$.

KB integrals. In fact, when extrapolating G^{∞} this range should be avoided and only the linear part of the function G^V should be used. In Section 4.2.1, the best range used for extrapolating the scaling of $1/R$ and G^V, to properly obtain KB integrals, is further discussed. Finally, we examine the scaling of the function F (Equation (16)) with the surface area of the spherical subvolume. Figure 4 shows how the function F scales linearly with the surface area. The integrals over L^3 in Equation (16) become larger as the number of molecules around surface area of the subvolume increase.

For all simulation box sizes, the values of the function decrease when R is larger than $L/2$ due to the decrease in the number of molecules at the boundary of the sphere V. In any case, regardless of the shape of the subvolume, whether a fully embedded sphere or a truncated sphere, the function F scales linearly with the surface area.

4.2. Finite-size effects of the RDF

Here, we study the RDF-related effects discussed in Section 3.2. RDFs obtained from closed, and finite simulations have to be corrected. Figure 5 shows the enhanced scaling of $G^V_{\alpha\beta}$ with $1/R$ when applying an RDF correction. The correction methods for $g(r)$ are applied to KB integrals computed from closed, finite molecular dynamics of the WCA mixture described in Section 4.2. In the case of obtaining the RDFs from MD simulations, the extrapolation of $G^{\infty}_{\alpha\beta}$ is not straightforward. In the following section, we show how to identify the linear regime.

4.2.1. Identifying the linear regime

We consider the RDF computed for the binary WCA mixture, while applying the van der Vegt correction, since we find that it gives the most accurate KB integrals out of the three methods studied in this work (details are provided in Section 4.2.2). The MD simulation details and system conditions are provided in Section 3.2.4. The study is performed to identify the extrapolation range from the scaling of σ/R with $G^V_{\alpha\beta}$ from simulation boxes with $L = 10$, $L = 20$ and $L = 40$. To find the best linear

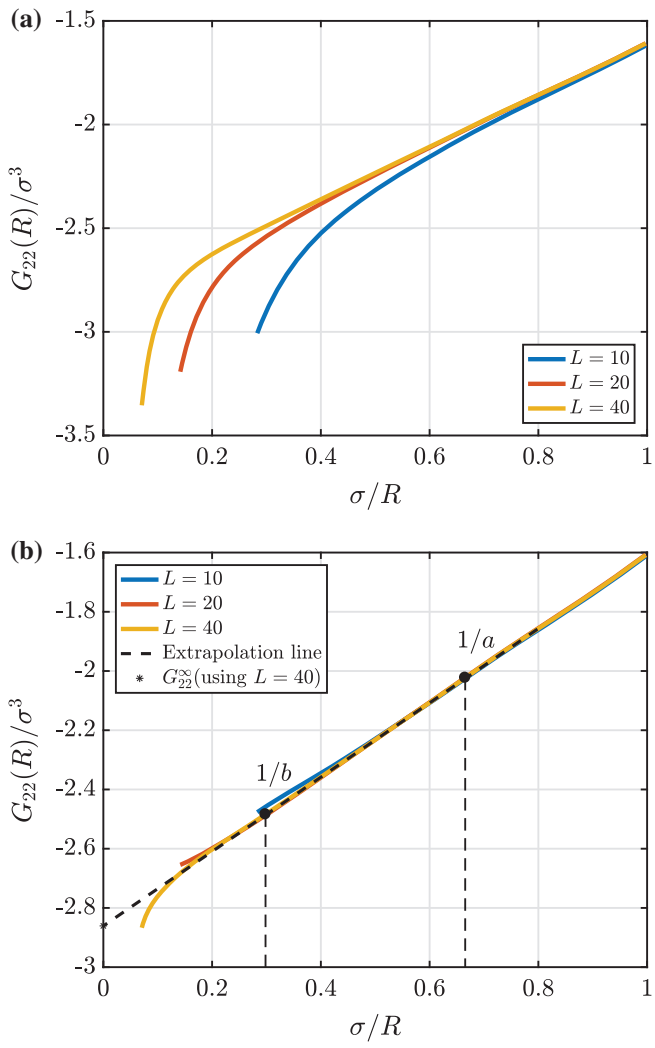

Figure 5. (Colour online) KB integrals for subvolumes from MD simulations of a binary WCA mixture. The system conditions, for all system sizes are $x_1 = 0.75$, $T^* = 1.8$ and $\rho^* = 0.7$. G^V_{22}/σ^3 is obtained from integrating $g_{22}(r)$ (Equation (3)) at each subvolume size, R. The KB integrals are computed for simulation boxes with $L = 10$, 20 and 40. In (a), the RDF is not corrected while in (b) the van der Vegt correction is applied and g^{vdV}_{22} is used (Equation (19)). In (b), the range used for extrapolation for the case of $L = 40$ is shown as well as the extrapolation line from which G^{∞}_{22}/σ^3 is obtained.

range, the start of the $1/R$ range, $1/a$, and the end, $1/b$, are varied (see Figure 5(b)). For each extrapolation range, the square of the correlation coefficient (denoted as c^2 in this work) is computed to assess the linearity of the selected range. Also, the difference between the extrapolated KB integrals, $G^{\infty}_{\alpha\beta}$, and the integrals computed from a large system ($L = 80$) are computed, using Equation (28). To relate the tested ranges to the dimensions of the system, a and b are related to the molecular diameter σ. Namely, a is set based on how many molecular diameters should be discarded at the beginning of the distances at which the RDF is computed. The starting point for the $1/R$ extrapolation could be varied as follows,

$$\frac{1}{a} = \frac{1}{x\sigma}. \tag{29}$$

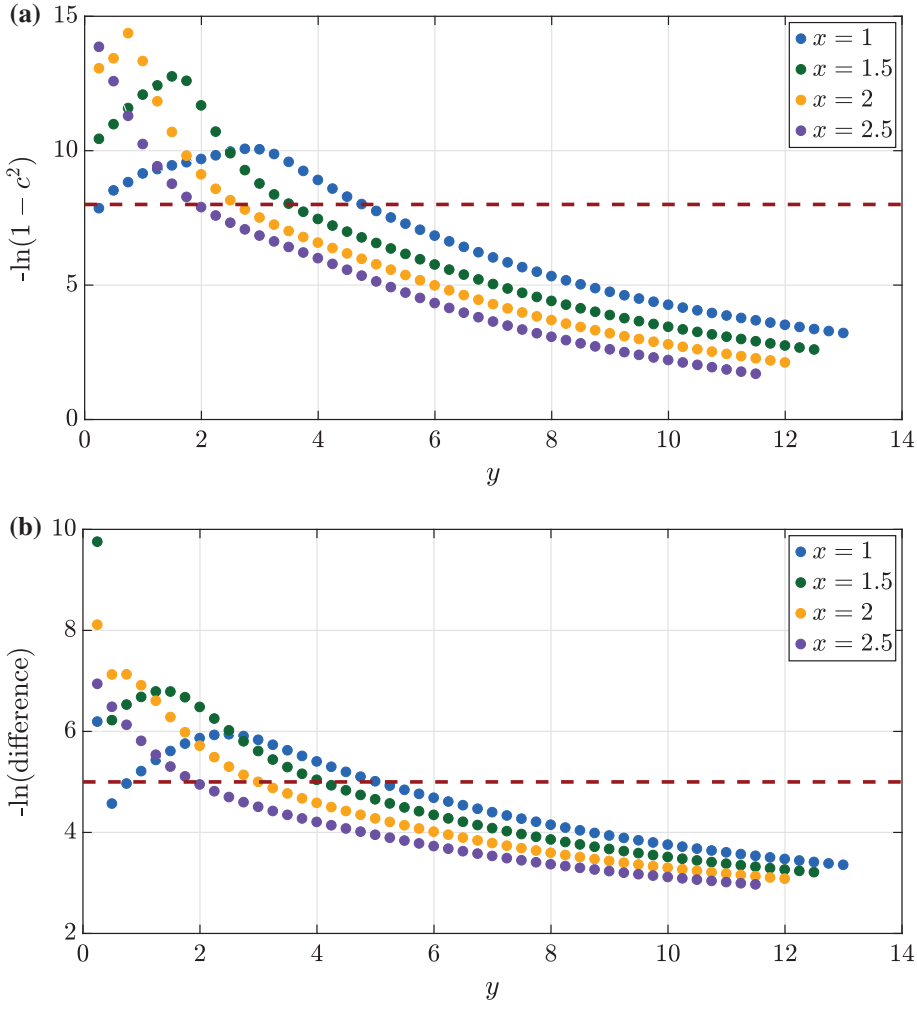

Figure 6. (Colour online) The natural logarithm of (a) the linearity $1 - c^2$ and (b) the differences, as a result of varying the linear range used to extrapolate to G_{22}^{∞}. The differences between KB integrals from MD simulations of large systems and KB integrals, G_{22}^{∞}, from different extrapolation ranges. The KB integrals are computed from MD simulations of the binary WCA mixture (details in Section 3.2.4). The size of the simulation box is set to $L = 40$. Points with different colours correspond to different starting point of the extrapolation range, x, and as a result a (Equation (29)). The variable y correspond to how far should the range be extended (Equation (30)).

Once x, and subsequently a, is set the end point of the extrapolation range could be again related to σ,

$$\frac{1}{b} = \frac{1}{a + y\sigma} \tag{30}$$

In this section, we demonstrate how to find the best extrapolation range for the case of the values of G_{22}^{V}. The results for the effect of varying x and y on the quality of the linear fit (c^2) and the accuracy in KB integrals computations (difference%) are shown in Figures 6–8 for the box sizes $L = 40$, $L = 20$ and $L = 10$, respectively. In these figures, the values of $-\ln(1 - c^2)$ and $-$ln.difference/ are plotted as functions of x and y. Also, in each plot a dashed line is added to indicate values of c^2 and difference% that we consider acceptable. x and y combinations that fall under the line are considered to give poor estimations of the KB integrals. Specifically, the following threshold values are considered, 8 for $-\ln(1 - c^2)$ and 5 for $-$ln.difference/ corresponding to $c^2 = 0.9997$ and difference = 0.7%.

Using a box size with $L = 40$ provides sufficient linear regime that results in very low values of the difference % between KB integrals computed and KB integrals from very large systems. Figure 6 shows that choosing lower x and y is favourable. The same observations are made when studying $L = 20$ (Figure 7). Using a large value of y could result in an extrapolating range that includes the diverging part of the G^V and $1/R$ scaling. Figures 6 and 7 show that, for a given x, the values of $-\ln(1-c^2)$ and $-$ln.difference/ start decreasing after a specific y value. In general, y should not be larger than 4σ. Finally, for $L = 10$, Figure 8 indicates that there is a range that is sufficiently linear. However, very few possibilities of x and y combinations provide low differences, i.e. the linear regime extends to a value that deviates from the integral at the thermodynamic limit, G^{∞}.

Besides examining the linear range for each system size individually, it is possible to investigate the effect of the system size by comparing Figures 6–8. Mainly, larger box sizes provide a longer linear regime and smaller differences between KB integrals computed from these boxes and integrals from very large boxes. When performing the study for the other integrals as well, G_{11}^{V} and G_{12}^{V}, a few general findings could be observed. First, it is recommended to choose x larger than 1, to avoid any fluctuations at small distances. This corresponds to discarding

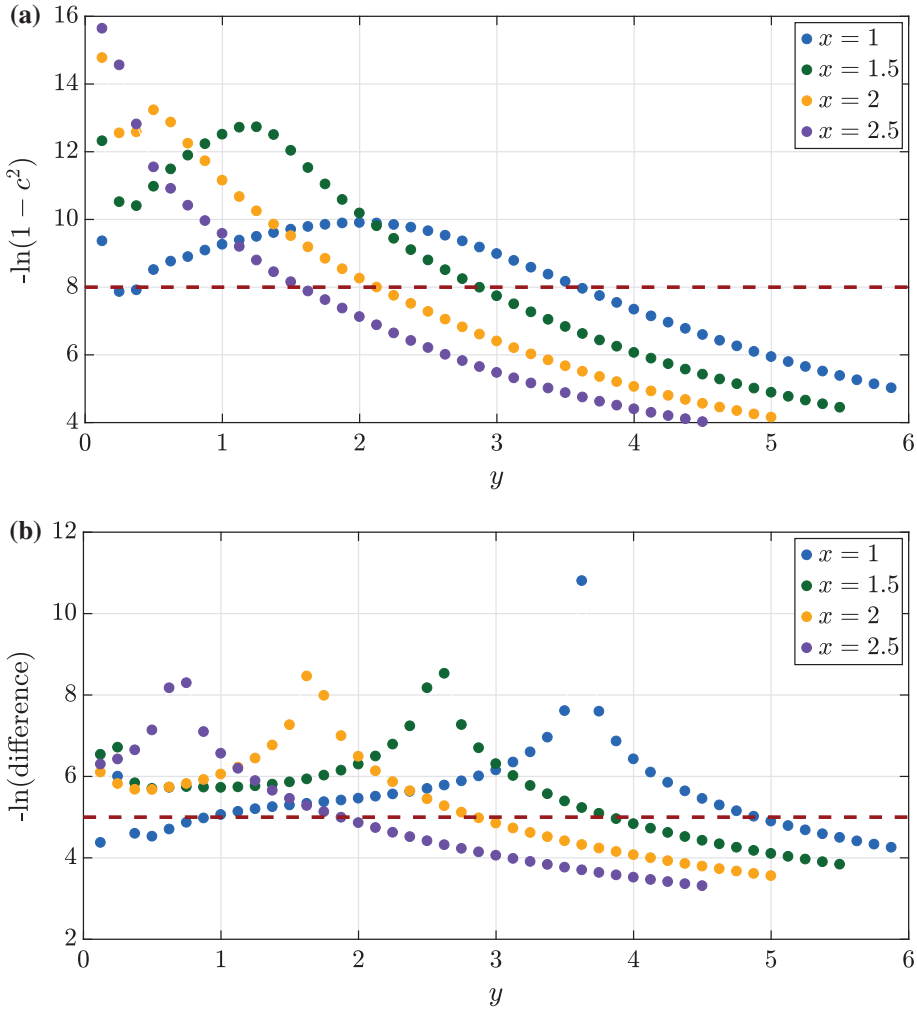

Figure 7. (Colour online) The natural logarithm of (a) the linearity $1 - c^2$ and (b) the differences, as a result of varying the linear range used to extrapolate to G_{22}^{∞}. The differences between KB integrals from MD simulations of large systems and KB integrals, G_{22}^{∞}, from different extrapolation ranges. The KB integrals are computed from MD simulations of the binary WCA mixture (details in Section 3.2.4). The size of the simulation box is set to $L = 20$. Points with different colours correspond to different starting point of the extrapolation range, x, and as a result a (Equation (29)). The variable y correspond to how far should the range be extended (Equation (30)).

at least 1σ from the G^V and $1/V$ scaling line. As for the end of the extrapolation range, y, taking short distances ensures that the extrapolation range does not extend to the end of the line, where the values of $G_{\alpha\beta}^V$ are diverging. We find that not exceeding $b = a + 4\sigma$ ensures that the linear fit is acceptable. This applies to boxes with L larger than 20. Finally, while it is important to use linear ranges with high values of the correlation coefficient, c^2 that does not always result in correct estimations of the KB integrals at the thermodynamic limit, especially for small simulation boxes. In the following section, inaccuracies in KB integrals resulting from finite sizes related to RDFs are discussed.

4.2.2. Comparing correction methods

Using RDFs computed from finite and closed simulation boxes lead to a systematic error in KB integrals computations. These RDF finite-size effects should be corrected. In Figure 5, we already showed the scaling of G_{22}^V with σ/R (a) when using RDFs that are not corrected and (b) when applying the van der Vegt correction [30]. For all the system sizes used, implementing the correction enhances the linear scaling of the finite-volume KB integrals. In addition to the van der Vegt correction, we considered the $1/N$ correlation method [28], and the method proposed by Cortes-Huerto et al. [31] We compare between the correction methods by considering the differences between the corrected RDF obtained from small systems and the RDF computed using a large system ($L = 80$). Figure 9 shows the quantity $\ln|g(r) - g^{80}(r)|$ for all distances, where the RDFs are computed for the WCA binary mixture using $L = 10, 20$ and 40. Generally, the differences decrease with larger box sizes, and the differences are larger for small r. Also, for all system sizes, the van der Vegt correction and the correction of Cortes-Huerto et al. result in lower deviations from the finite RDF than in the case of the $1/N$ correlation.

The integrals G_{11}^{∞}, G_{12}^{∞} and G_{22}^{∞} are computed for the binary WCA mixture using three different simulation boxes, with $L = 10$, $L = 20$ and $L = 40$. In the case of the $1/N$ correlation, the values of $G_{\alpha\beta}^{\infty}$ are obtained using two simulation box sizes, that are not very different. For instance, to compute KB integrals from a simulation box with $L = 10$, RDFs from simulations of boxes with $L = 10$ and $L = 11$, at the same density and temperature, are required. The obtained integrals from each

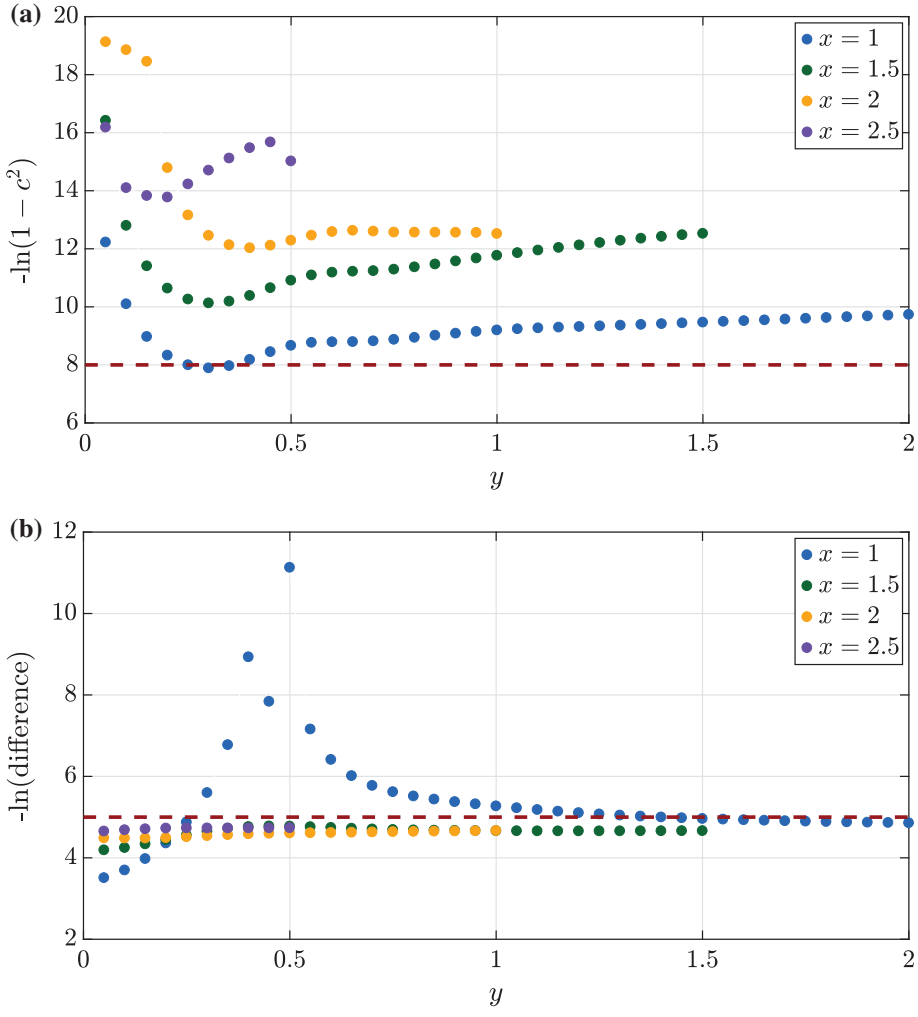

Figure 8. (Colour online) The natural logarithm of (a) the linearity $1 - c^2$ and (b) the differences, as a result of varying the linear range used to extrapolate to G_{22}^{∞}. The differences between KB integrals from MD simulations of large systems and KB integrals, G_{22}^{∞}, from different extrapolation ranges. The KB integrals are computed from MD simulations of the binary WCA mixture (details in Section 3.2.4). The size of the simulation box is set to $L = 10$. Points with different colours correspond to different starting point of the extrapolation range, x, and as a result a (Equation (29)). The variable y correspond to how far should the range be extended (Equation (30)).

Table 3. Differences (%) between $G_{\alpha\beta}^{\infty}/\sigma^3$ computed using MD simulations of a large system ($L/\sigma = 80$) and the integrals from simulations of small boxes, $L = 10$, 20 and 40. The MD results are for a binary WCA mixture ($x_1 = 0.75$) at $T^* = 1.8$ and $\rho^* = 0.7$ (reduced units). Each column includes differences as a result of implementing a different RDF correction method. For the $1/N$ correlation, two MD simulations were used, $L = 10, 11$, $L = 20, 21$ and $L = 40, 41$. Our best estimates for $G_{\alpha\beta}^{\infty}/\sigma^3$ are $G_{11}^{\infty}/\sigma^3 = -1.603$, $G_{12}^{\infty}/\sigma^3 = -0.5446$, $G_{22}^{\infty}/\sigma^3 = -2.870$.

Box length(L)	Without correction	van der Vegt [30]	$1/N$ correlation [41]	Cortes-Huerto method [31]
(a) G_{11}^{∞}/σ^3				
10	5.6	0.3	0.29	3.7
20	1.2	0.2	0.9	3.3
40	0.01	0.2	1.8	2.2
(b) G_{12}^{∞}/σ^3				
10	31	3	4	4
20	7	0.1	0.3	3
40	0.2	0.7	17	2
(c) G_{22}^{∞}/σ^3				
10	29	0.9	2	4
20	8	0.6	3	3
40	0.4	0.6	13	2

correction method are then compared to integrals computed using a system with $L = 80$. The differences (%) between KB integrals computed using $L = 80$ and KB integrals G_{11}^{∞}, G_{12}^{∞} and G_{22}^{∞} from finite simulation boxes ($L = 10, 20 and 40$) are presented in Table 3. For each KB integral, $G_{\alpha\beta}^{\infty}$, the differences are computed when the RDF is not corrected and when the

Table 4. Differences (%) between G_f^{∞}/σ^3 ($G_f = G_{11} + G_{22} - G_{12}/2$) computed using MD simulations of a large system ($L = 80$) and the integrals from simulations of small boxes, $L = 10, 20$*and*40. The MD results are for a binary WCA mixture ($x_1 = 0.75$) at $T^* = 1.8$ and $\rho^* = 0.7$ (reduced units). Each column includes differences as a result of implementing a different RDF correction method. For the $1/N$ correlation, two MD simulations were used, $L = 10, 11$, $L = 20, 21$ and $L = 40, 41$. Our estimate for G_f^{∞}/σ^3 equals -3.380 ± 0.005.

Box length(L)	Without correction	van der Vegt [30]	$1/N$ correlation [41]	Cortes-Huerto method [31]
10	37	1.8	2.4	5.2
20	9.4	0.3	2.5	3.9
40	0.4	0.8	17.4	2.8

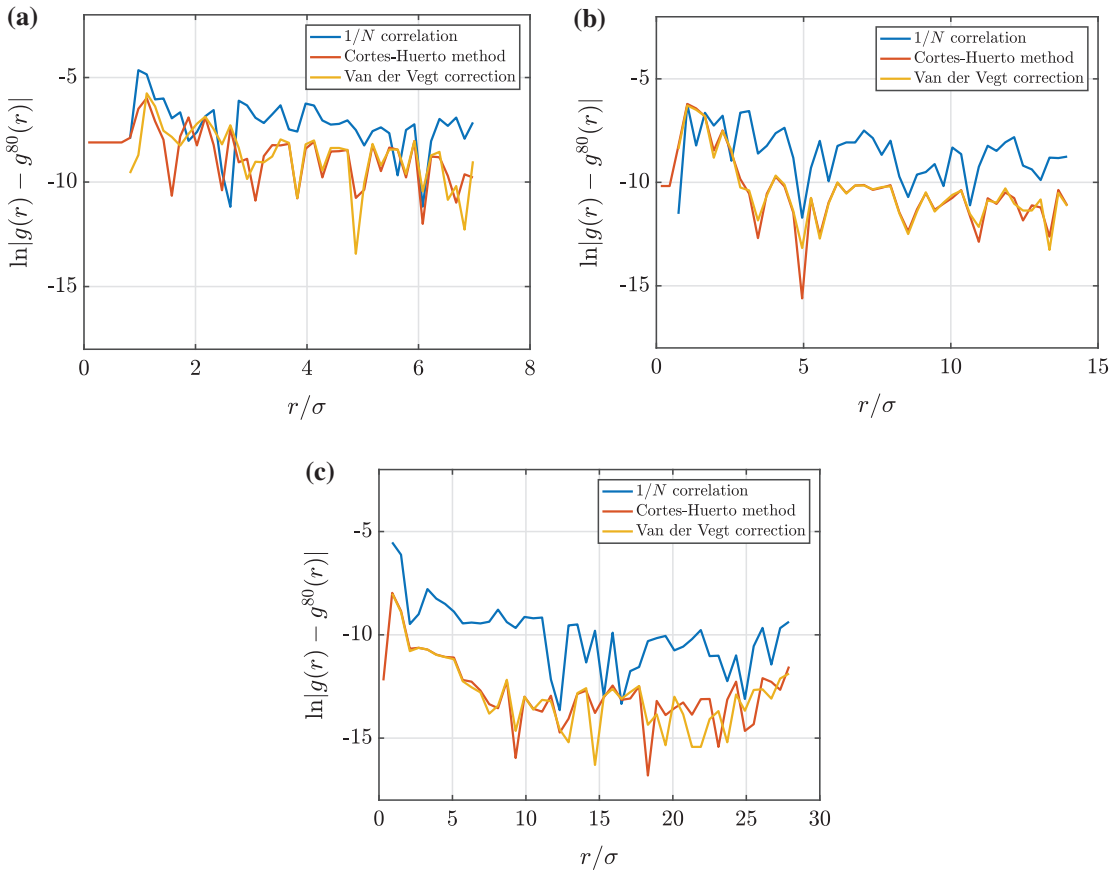

Figure 9. (Colour online) The natural logarithm of the difference between RDFs from MD simulation of systems with finite sizes, $g(r)$, and RDF from simulation of system with very large size, $g^{80}(r)$. The RDFs are computed for the binary WCA mixture with: $x_1 = 0.75$, $T^* = 1.8$ and $\rho^* = 0.7$. The differences are computed for three different simulation box sizes (a)$L = 10$, (b)$L = 20$ and (c)$L = 40$. For each system size, the difference between $g(r)$ and $g^{80}(r)$ is computed at each r. The different colours represent different RDF correction methods (Section 3.2).

correction methods discussed in this work are applied. Overall, the differences decrease with larger simulation boxes, with the exception of the $1/N$ correlation. Additionally, for box sizes up to $L = 20$, the van der Vegt method results in the lowest differences. For large boxes, Table 3 shows that a RDF correction is not needed and in the case of $L = 40$ the differences resulting from not using a corrections and the differences resulting from the correction methods are not comparable. This means that one does not have to correct the RDF for very large systems. Other than the individual KB integrals, we also consider the quantity, $G^f = G_{11} + G_{22} - 2G_{12}$, which is useful when computing the thermodynamic factor, Γ, from the KB integrals [10]. For $L = 80$, we find that $G^f/\sigma^3 = -3.38 \pm 0.005$, deviations from this value are presented in Table 4 for the case of not using a RDF correction as well as when using the three RDF correction methods. In the table, the differences (%) are computed using Equation (28). Using the RDF without a correction result in considerable differences in the values of G^f, especially for the smallest simulation box with $L = 10$. Out of the three methods, the van der Vegt correction [30] leads to the lowest differences in the KB integrals results in the thermodynamic limit. In Figure 9, we show that the method of Cortes-Huerto et al. [31] is similar to the van der Vegt correction [30] when estimating the function $g_{\alpha\beta}^{\infty}(r)$. Still, when computing the KB integrals the van der Vegt correction provides lower differences than the method of Cortes-Huerto et al. [31], which assumes that the RDF is independent of r. Also, the van der Vegt correction is fairly simple to implement, and the corrections are applied to one simulation for each size, unlike the $1/N$ correlation, where for each size two simulations are required. Another shortcoming of the latter method is its numerical inaccuracy when correcting the RDF, resulting in inconsistency with regard to predicting the KB integrals. Furthermore, we apply the correction method proposed by Cortes-Huerto et al. [31] to the same WCA mixture.

The method enhances the computation of the integrals from finite systems, but the differences(%) in the KB computations are not lowered as in the case of the van der Vegt correction.

5. Conclusions

In this work, we studied the finite-size effects related to the computation of KB integrals from molecular simulations of finite subvolumes. We presented the uncertainties in KB integrals due to the following: (1) effects due to the finite size of the subvolume, and simulation box, used to compute the KB integrals and (2) effects related to computing RDFs from molecular simulations of closed systems, in contrast to open systems as defined in the KB theory. We showed that uncertainties in the computations of the KB integrals decrease when increasing the size of the simulation box, and hence the embedded subvolume. We varied the system size and find that simulation boxes with lengths larger than 15σ are sufficient to reduce errors in computed KB integrals to below 0.1%. We vary the size of the subvolume, or the distance at which the RDF are computed. We find that a larger distance does not always ensure higher accuracy. In fact, for a given simulation box size, the radius of the spherical subvolume should not be extended beyond half the length of the simulation box. When using an analytic RDF model for the computations of the KB integrals, it is relatively straightforward to identify the linear regime in the scaling of finite subvolumes integrals with the inverse size of the subvolume. However, using RDFs computed from MD simulation of WCA molecules did not necessarily result in identifiable linear regime. In this work, we presented some guidelines for extrapolating the scaling of finite subvolumes KB integrals to the thermodynamic limit. In some cases small simulation boxes provided a sufficient linear regime; however, RDF finite effects caused the resulting KB integrals to deviate from these obtained from very large systems. Uncertainties arising from using RDFs of closed systems were evaluated for multiple simulation box sizes, as well as for various RDF correction methods. We demonstrate that using a RDF correction can significantly enhance the convergence of the KB integrals and eventually the accuracy of the computations of the integrals. We compare between the RDF correction methods and find that the van der Vegt correction of Ref. [30] achieves the lowest error and is the easiest to apply.

Acknowledgements

TJHV acknowledges NWO-CW for a VICI grant. SK and DB are grateful to the Research Council of Norway for the Centers of Excellence funding scheme, project number 262644 PoreLab.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was sponsored by NWO Exacte Wetenschappen (Physical Sciences)or the use of supercomputer facilities with financial support from the Nederlandse Organisatie voor weten schappelijk Onderzoek (Netherlands Organisation for Scientific research, NWO); NWO-CW for a VICI grant; Research Council of Norway for the Centers of Excellence funding scheme [project number 262644 PoreLab].

ORCID

N. Dawass http://orcid.org/0000-0001-5234-7127
P. Krüger http://orcid.org/0000-0002-1247-9886
S. Kjelstrup http://orcid.org/0000-0003-1235-5709
T. J. H. Vlugt http://orcid.org/0000-0003-3059-8712

References

[1] Wedberg R, O'Connell JP, Peters GH, et al. Accurate Kirkwood-Buff integrals from molecular simulations. Mol Simul. 2010;36:1243–1252.
[2] Shulgin I, Ruckenstein E. Kirkwood-Buff integrals in aqueous alcohol systems: comparison between thermodynamic calculations and X-ray scattering experiments. J Phys Chem B. 1999;103:2496–2503.
[3] de Pablo JJ, Escobedo FA. Molecular simulations in chemical engineering: present and future. AIChE J. 2002;48:2716–2721.
[4] Schnell SK, Englebienne P, Simon JM, et al. How to apply the Kirkwood-Buff theory to individual species in salt solutions. Chem Phys Lett. 2013;582:154–157.
[5] Kirkwood JG, Buff FP. The statistical mechanical theory of solutions. I. J Chem Phys. 1951;19:774–777.
[6] Ben-Naim A. Inversion of the Kirkwood-Buff theory of solutions: application to the water-ethanol system. J Chem Phys. 1977;67:4884–4890.
[7] Hall DG. Kirkwood-Buff theory of solutions. An alternative derivation of part of it and some applications. Trans Faraday Soc. 1970;67:2516–2524.
[8] Ben-Naim A. Molecular theory of solutions. Oxford: Oxford University Press; 2006.
[9] Nishikawa K, Ishizawa K, Kodera Y, et al. Determination of the energy spectrum of the primary beam in energy-dispersive X-ray diffractometry. Jpn J Appl Phys. 1986;25:1431–1434.
[10] Liu X, Schnell SK, Simon JM, et al. Diffusion coefficients from molecular dynamics simulations in binary and ternary mixtures. Int J Thermophys. 2013;34:1169–1196.
[11] Schnell SK, Skorpa R, Bedeaux D, et al. Partial molar enthalpies and reaction enthalpies from equilibrium molecular dynamics simulation. J Chem Phys. 2014;141:144501.
[12] Matteoli E, Lepori L. Solute-solute interactions in water. II. An analysis through the Kirkwood-Buff integrals for 14 organic solutes. J Chem Phys. 1984;80:2856–2863.
[13] O'Connell J. Thermodynamic properties of solutions based on correlation functions. Mol Phys. 1971;20:27–33.
[14] Patil KJ. Application of Kirkwood-Buff theory of liquid mixtures to water-butanol system. J Solution Chem. 1981;10:315–320.
[15] Koga K, Widom B. Thermodynamic functions as correlation-function integrals. J Chem Phys. 2013;138:114504.
[16] Baus M, Rull LF, Ryckaert JP. Observation, prediction and simulation of phase transitions in complex fluids. Vol. 460. Dordrecht: Springer Science & Business Media; 2012.
[17] Vlugt TJH, Martin M, Smit B, et al. Improving the efficiency of the configurational-bias Monte Carlo algorithm. Mol Phys. 1998;94:727–733.
[18] Dubbeldam D, Torres-Knoop A, Walton KS. On the inner workings of Monte Carlo codes. Mol Simul. 2013;39:1253–1292.
[19] Torres-Knoop A, Balaji SP, Vlugt TJH, et al. A comparison of advanced Monte Carlo methods for open systems: CFCMC vs CBMC. J Chem Theory Comput. 2014;10:942–952.
[20] Torres-Knoop A, Poursaeidesfahani A, Vlugt TJH, et al. Behavior of the enthalpy of adsorption in nanoporous materials close to saturation conditions. J Chem Theory Comput. 2017;13:3326–3339.
[21] Weerasinghe S, Smith PE. A Kirkwood-Buff derived force field for mixtures of urea and water. J Phys Chem B. 2003;107:3891–3898.
[22] Weerasinghe S, Smith PE. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. J Chem Phys. 2004;121:2180–2186.
[23] Smith PE. Cosolvent interactions with biomolecules: relating computer simulation data to experimental thermodynamic data. J Phys Chem B. 2004;108:18716–18724.

[24] Kang M, Smith PE. Preferential interaction parameters in biological systems by Kirkwood-Buff theory and computer simulation. Fluid Phase Equilib. 2007;256:14–19.
[25] Gee MB, Cox NR, Jiao Y, et al. A Kirkwood-Buff derived force field for aqueous alkali halides. J Chem Theory Comput. 2011;7:1369–1380.
[26] Nichols JW, Moore SG, Wheeler DR. Improved implementation of Kirkwood-Buff solution theory in periodic molecular simulations. Phys Rev E. 2009 Nov;80:051203.
[27] Verlet L. Computer "experiments" on classical fluids. II. Equilibrium correlation functions. Phys Rev. 1968;165:201–214.
[28] Krüger P, Schnell SK, Bedeaux D, et al. Kirkwood-Buff integrals for finite volumes. J Phys Chem Lett. 2013;4:235–238.
[29] Strøm BA, Simon JM, Schnell SK, et al. Size and shape effects on the thermodynamic properties of nanoscale volumes of water. Phys Chem Chem Phys. 2017;19:9016–9027.
[30] Ganguly P, van der Vegt NFA. Convergence of sampling Kirkwood-Buff integrals of aqueous solutions with molecular dynamics simulations. J Chem Theory Comput. 2013;9:1347–1355.
[31] Cortes-Huerto R, Kremer K, Potestio R. Communication: Kirkwood-Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. J Chem Phys. 2016;145:141103.
[32] Weeks JD, Chandler D, Andersen HC. Role of repulsive forces in determining the equilibrium structure of simple liquids. J Chem Phys. 1971;54:5237–5247.
[33] Weisstein EW. CRC concise encyclopedia of mathematics. 2nd ed. Boca Raton: Chapman and Hall; 2003.
[34] Kirkwood JG, Boggs EM. The radial distribution function in liquids. J Chem Phys. 1942;10:394–402.
[35] Deserno M. How to calculate a three-dimensional $g(r)$ under periodic boundary conditions. 2004. Available from: www.cmu.edu/biolphys/deserno/pdf/gr_periodic.pdf.
[36] Weisstein EM. Wolfram mathword. 2017. Available from: mathworld.wolfram.com/SphericalCap.
[37] Salacuse JJ, Denton AR, Egelstaff PA. Finite-size effects in molecular dynamics simulations: static structure factor and compressibility. i. Theoretical method. Phys Rev E. 1996 Mar;53:2382–2389.
[38] Hoffman JD, Frankel S. Numerical methods for engineers and scientists. New York: CRC Press; 2001.
[39] Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. Vol. 1. San Diego: Academic Press; 2001.
[40] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys. 1995;117:1–19.
[41] Schnell SK, Vlugt TJH, Simon JM, et al. Thermodynamics of small systems embedded in a reservoir: a detailed analysis of finite size effects. Mol Phys. 2012;110:1069–1079.