Improved implementation of Kirkwood-Buff solution theory in periodic molecular simulations

Joseph W. Nichols, Stan G. Moore, and Dean R. Wheeler*
Department of Chemical Engineering, Brigham Young University, Provo, Utah 84602, USA
(Received 26 August 2009; published 23 November 2009)

Kirkwood-Buff (KB) solution theory is a means to obtain certain thermodynamic derivatives from knowledge of molecular distributions. In actual practice the required integrals over radial distribution functions suffer inaccuracies due to finite-distance truncation effects and their use in closed systems. In this work we discuss how best to minimize these inaccuracies under traditional KB theory. In addition we implement a method for calculating KB quantities in molecular simulations with periodic boundary conditions and particularly within the canonical ensemble. The method is based on a finite-Fourier-series expansion of molecular concentration fluctuations and leads to more reliable results for a given computational effort. The procedure is validated and compared to the original method for a nonideal liquid mixture of Lennard-Jones particles intended to imitate a real system, carbon tetrafluoride, and methane.

DOI: 10.1103/PhysRevE.80.051203 PACS number(s): 61.20.Ja, 82.60.Lf, 62.10.+s, 65.20.De

I. INTRODUCTION

There are multiple ways to get free-energy-related properties from statistical mechanical computations. For instance, one can use thermodynamic integration, Widom's method, and free-energy perturbation methods in general [1]. Kirkwood-Buff (KB) solution theory [2] is another route to free-energy-related properties, in particular partial molar volumes and composition derivatives of chemical potential. Isothermal compressibility also can be calculated by the method.

The KB method is related to the local-composition concept used in continuum thermodynamics to correlate and predict activity coefficients. The basic idea is that all the information about macroscopic free-energy-related properties is embedded in the way that molecules of the same and different species associate, as manifested by composition and density variations in a local (molecular) level. This concept is found in Debye-Hückel theory and its extensions [3] and in liquid-mixture activity-coefficient models by Wilson and by Prausnitz, among others [4,5]. Because fluid microstructure can be obtained both by statistical mechanics calculations and by scattering experiments, over the last several decades KB theory has provided a useful connection between these methods and macroscopic thermodynamic observables.

In this work we consider ways to improve the accuracy of quantities calculated using KB theory. We develop an alternative implementation of KB theory for use in molecular simulations with periodic boundary conditions and particularly with fixed numbers of molecules. We validate and test the method, comparing it to the original KB implementation for a nonideal liquid mixture of Lennard-Jones (LJ) particles. For this system the method more accurately predicts the proper thermodynamic derivatives for a given simulation cell size and computational effort, meaning it is a more numerically efficient method.

II. KIRKWOOD-BUFF SOLUTION THEORY

The basis of Kirkwood-Buff solution theory is to connect thermodynamic free-energy properties to integrals over radial distribution functions $g_{ij}(r)$ in a grand canonical (μVT) ensemble. The details are available in the original 1951 paper by Kirkwood and Buff [2] as well as a later review by Newman [6] and a compilation volume edited by Matteoli and Mansoori [7]. We summarize here the general results for a macrohomogeneous mixture of n species.

Note that equations in this paper are given on a molecular basis rather than a molar basis; to effect the latter, one must simply substitute the gas constant R_g for Boltzmann's constant k_B, as well as take N_i and N to be numbers of moles rather than numbers of molecules. We use subscripts i, j, and k as species indices ranging from 1 to n, unless otherwise specified.

A. Pairwise distributions

To start we define a symmetric dimensionless matrix $\mathbf{A}$ according to

$$(\mathbf{A}^{-1})_{ij} = \langle N\rangle^{-1}[\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle], \tag{1}$$

where N is the fluctuating total number of molecules and N_i is the corresponding number for species i. Angle brackets $\langle\ldots\rangle$ have the traditional meaning of a statistical mechanical ensemble average. Note that our definition of matrix $\mathbf{A}$ intentionally differs by a factor of $\langle N\rangle$ from that of the original KB paper, in order to make elements of $\mathbf{A}$ intensive properties (i.e., not explicitly dependent on system size). Kirkwood and Buff related the above fluctuation expression to pairwise distributions, which we state in matrix form

$$\mathbf{A}^{-1} = \mathbf{Y} + \rho\mathbf{YGY}, \tag{2}$$

where $\rho = \langle N\rangle / V$ is average total concentration. Diagonal matrix $\mathbf{Y}$ is defined by $Y_{ij} = \delta_{ij} y_i$, where δ_{ij} is the Kronecker delta and $y_i = \langle N_i\rangle / \langle N\rangle$ is average species mole fraction. In Eq. (2) the first term on the right reflects the correlation of a molecule with itself. $\mathbf{G}$ is the symmetric matrix of so-called Kirkwood-Buff integrals, which express correlations between neighboring molecules or equivalently express the local composition deviations about molecules

*dean_wheeler@byu.edu

$$G_{ij} = \int_0^\infty [g_{ij}(r) - 1] 4\pi r^2 dr. \quad (3)$$

For an ideal (Lewis) mixture only the diagonal elements of matrix **G** are independent, with the off-diagonal elements given by $G_{ij} = \frac{1}{2}(G_{ii} + G_{jj})$ [8]. For an ideal-gas mixture all $G_{ij} = 0$. For finite-size simulations, the upper limit of integration in Eq. (3) cannot be infinity and must instead be some finite radius r_c—a matter discussed more fully below.

As an aside, we note that Eq. (2) is related to structure factors obtainable from scattering experiments. Various structure factors have been defined in the literature [9,10], and each is basically a spherically symmetric Fourier transform of some type of pair correlation function $(g(r) - 1)$. We define here a dimensionless partial structure factor S_{ij} as

$$S_{ij}(q) = y_i \delta_{ij} + y_i y_j \rho \int_0^\infty \frac{\sin(qr)}{qr} [g_{ij}(r) - 1] 4\pi r^2 dr, \quad (4)$$

where q is the magnitude of a reciprocal lattice vector or momentum change vector. In particular, $q = 4\pi\lambda^{-1} \sin(\frac{1}{2}\theta)$, where λ the radiation wavelength and θ is the scattering angle. In the limit $q \to 0$, Eq. (4) reduces to Eq. (2), namely, $S_{ij}(0) = (\mathbf{A}^{-1})_{ij}$. However, one cannot measure the $q = 0$ values directly but must instead extrapolate experimental results to $q = 0$. Furthermore, determination of S_{ij} from x-ray diffraction experiments is not straightforward, given that one obtains linear combinations of elements of matrix **S** and a deconvolution must be made. Through isotopic substitution, neutron diffraction better enables direct measurement of partial structure factors.

B. Thermodynamic derivatives

Kirkwood and Buff [2] showed that **A** is a matrix of derivatives of chemical potential

$$A_{ij} = \frac{N}{k_B T} \left(\frac{\partial \mu_i}{\partial N_j} \right)_{T,V,N_{k \neq j}}. \quad (5)$$

Recall that species chemical potential can be expressed as $\mu_i = \mu_i^0 + k_B T \ln(\gamma_i y_i)$, where γ_i is a species activity coefficient and μ_i^0 is a pure-component chemical potential that depends on temperature T and pressure P.

Equation (5) is a foothold for deriving other thermodynamic properties. For instance, we can generate the partial volumes $\bar{V}_i = (\partial V / \partial N_i)_{T,P,N_{k \neq i}}$ and isothermal compressibility $\kappa_T = -(\partial \ln V / \partial P)_{T,N_i}$ in terms of elements of **A**. The column vector of species partial volumes is given by

$$\bar{\mathbf{V}} = \frac{1}{\rho} \frac{\mathbf{A}\mathbf{y}}{\mathbf{y}^{\mathrm{T}} \mathbf{A} \mathbf{y}}, \quad (6)$$

where **y** is the column vector of species mole fractions. Superscript T indicates a matrix transpose so that matrix operation $\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{y}$ results in a scalar quantity. By definition the thermodynamic constraint $\mathbf{y}^{\mathrm{T}}\bar{\mathbf{V}} = \rho^{-1}$ is satisfied. It turns out that simulated partial volumes are generally the most reliable of the properties obtained from KB theory; it appears that systematic biases in the elements of **A** are largely canceled in the ratio that produces $\bar{\mathbf{V}}$.

The isothermal compressibility is given by

$$\kappa_T = (\mathbf{y}^{\mathrm{T}} \mathbf{A} \mathbf{y})^{-1} \kappa_T^{ig}, \quad (7)$$

where the ideal gas compressibility is $\kappa_T^{ig} = (\rho k_B T)^{-1}$. Note that for a single-component system one can use the above relations to derive

$$\kappa_T = (1 + \rho G_{11}) \kappa_T^{ig}, \quad (8)$$

which is known as the Ornstein-Zernicke relation or compressibility equation. For low-temperature liquids $\kappa_T \ll \kappa_T^{ig}$; therefore, in this case KB integral G_{11} is negative and κ_T is the small difference between two nearly equal quantities. One can generalize this to state that for liquids relatively small errors in KB integrals can lead to relatively large errors in calculated κ_T [11].

Furthermore, we wish to calculate a quantity known as the activity-coefficient correction matrix, which expresses the composition dependence of chemical potentials in terms of mole fractions, at constant T and P.

$$Q_{ij} = \frac{y_i}{k_B T} \left(\frac{\partial \mu_i}{\partial y_j} \right)_{T,P,y_{k \neq j,n}} = \delta_{ij} - \delta_{in} + y_i \left(\frac{\partial \ln \gamma_i}{\partial y_j} \right)_{T,P,y_{k \neq j,n}}. \quad (9)$$

Q may be related to matrix **A** from a combination of thermodynamics principles and matrix algebra:

$$\mathbf{Q} = \mathbf{Y} \left(\mathbf{A} - \frac{\mathbf{A}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{A}}{\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{y}} \right) \mathbf{I}^0, \quad (10)$$

where $I_{ij}^0 = \delta_{ij} - \delta_{in}$. Note that for the partial derivatives of Eq. (9), y_n is allowed to vary at the same time as y_j in order to maintain the constraint $\Sigma_i y_i = 1$. This means that species n is specially selected as a reference species in the formulation above. Matrix **Q** as defined here has dimensions $n \times n$ but contains only $\frac{1}{2}n(n-1)$ independent parameters in general. It is constrained by the following relations: $Q_{ij}/y_i - Q_{ji}/y_j + (Q_{ni} - Q_{nj})/y_n = 0$, $Q_{in} = 0$, and $\Sigma_i Q_{ij} = 0$. The first relation follows from the symmetry of matrix **A** and the last relation follows from the Gibbs-Duhem principle. For an ideal mixture $\mathbf{Q} = \mathbf{I}^0$.

Matrix **Q** contains the second derivatives of Gibbs free energy with respect to mole fractions and has multiple uses. It can be used in diffusive constitutive equations to determine parameters for activity-coefficient models and to evaluate phase stability at a composition point. For example, phase stability in a binary mixture requires that $Q_{11} > 0$. An application of interest to the authors is to model multicomponent diffusion with a driving force derived from a chemical potential gradient as used in the Stefan-Maxwell constitutive equation that has been extended for use in nonideal solutions [12–14]. To do this we need to relate the chemical potential gradient of each species i to $n-1$ independent species composition gradients. This requires use of an $(n-1) \times (n-1)$ submatrix of **Q** [15]. Such an activity-coefficient correction has been calculated previously for mutual diffusion in binary mixtures using KB theory [11,16,17].

Matrices $\mathbf{A}$, $\mathbf{S}(0)$, and $\mathbf{G}$ are symmetric and each in general contains $\frac{1}{2}n(n+1)$ independent parameters. This information is encapsulated in the set of matrix $\mathbf{Q}$, vector $\overline{\mathbf{V}}$, and scalar κ_T. If one has reliable values of these latter quantities in hand, the above equations can be inverted to obtain the former in order to check consistency or develop theoretical relationships. The general inversion equation to obtain KB integrals is

$$\rho\mathbf{G} = -\mathbf{Y}^{-1} + \mathbf{Y}^{-1}(\mathbf{I}^0 + \mathbf{y}\mathbf{y}^{\mathrm{T}}) \times \left[\mathbf{Q} + \rho\mathbf{Y}\overline{\mathbf{V}}\overline{\mathbf{V}}(\rho\overline{\mathbf{V}}^{\mathrm{T}}\mathbf{I}^0 + \mathbf{y}^{\mathrm{T}})\frac{\kappa_T^{ig}}{\kappa_T}\right]^{-1}. \quad (11)$$

An expression equivalent to this for binary mixtures has been published previously [8].

III. APPLICATION OF KB THEORY TO A PERIODIC SYSTEM

The fluctuation formula for $\mathbf{A}^{-1}$ formally requires an open system. When KB solution theory is implemented in a more computationally convenient ensemble with fixed molecule numbers, such as canonical or *NVT*, the hope is that the theory still can be used to obtain reasonably accurate results. If the KB integrals [Eq. (3)] are performed out to maximum spherical radius $r_c=0.5L$, they are sampling approximately half of the volume of a cubic unit cell of edge length L. Because the sampling volume occupies only a portion of the full system, the sampling volume can exchange mass with its surroundings and be considered an open system under constraint, one which approximates a grand canonical system to order N^{-1} [18]. Unfortunately, Kirkwood-Buff integrals may not have converged to "macroscopic" values when truncated at $0.5L$ for typical simulation sizes. Possible remedies include an extrapolation to $r_c\to\infty$ through an empirical fit of radial distribution functions [19] or use of integral equation theories [20]; however, one must be cautious because results from KB theory are quite sensitive to small errors in g_{ij} at large r values. As we illustrate below, oscillations in the KB integrals can be smoothed, but this does not remedy the systemic biases due to a too-small system. It appears the only fail-safe way to manage the truncation error of explicitly simulated Kirkwood Buff integrals is brute force: increase L (and r_c), resimulate, and repeat until apparent N-independence is achieved.

We propose here an alternative formulation that accounts for the periodic boundaries of the simulation unit cell as commonly employed in molecular simulations. As before, a fluctuation expression is applied to a portion of the total volume within a closed simulation (such as *NVT*). Rather than the sampling volume for Eq. (1) being a sphere centered on a single moving molecule, the sampling volume is instead composed of one or more rectangular-slab-like regions that are stationary with respect to the simulation cell. Debenedetti performed a similar division of a closed simulation cell into two or more open subcells to study fluctuation phenomena [21].

A. Structure factor from simulations

In order to count the particles in the designated sampling volume, we propose an expression for particle concentration that is consistent with the periodic boundaries. The fluctuation of the species concentration field from its equilibrium or average value can be described with a three-dimensional discrete Fourier series:

$$\delta\rho_i(\mathbf{r},t) = \rho\sum_{\mathbf{q}\neq 0}\psi_i(\mathbf{q},t)e^{\mathbf{i}\mathbf{q}\cdot\mathbf{r}}, \quad (12)$$

where $\mathbf{r}$ is a position vector, t is time, and ψ_i is a dimensionless Fourier coefficient or single-molecule structure factor. Vector $\mathbf{q}$ is a reciprocal lattice vector of the unit cell and $\mathbf{i}$ in the exponent is the imaginary number. For a cubic cell of length L we get the set of discrete wave or reciprocal lattice vectors

$$\mathbf{q} = \frac{2\pi}{L}\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}, \quad (13)$$

where m_1, m_2, and m_3 are independent integers. Note that $q=|\mathbf{q}|$. For simulations there is an upper bound q_c on the discrete q values one cares to use. In this work we set $q_c=2\pi L^{-1}\sqrt{17}$. After excluding $\mathbf{q}=0$, this makes 760 unique lattice vector pairs $(\pm\mathbf{q})$ within that cutoff.

Fourier mass coefficient ψ_i is calculated in a simulation from instantaneous molecular positions:

$$\psi_i(\mathbf{q},t) = \nu_i(q)\frac{1}{N}\sum_{a\in i}e^{-\mathbf{i}\mathbf{q}\cdot\mathbf{r}_a(t)}, \quad (14)$$

where $\mathbf{r}_a(t)$ is the center-of-mass position of molecule a of species i and the sum is over all molecules of species i. We note that fluctuations about equilibrium should average to zero, $\langle\delta\rho_i\rangle=\langle\psi_i\rangle=0$, except in the case of stable multiphase systems.

As structured molecules are not Dirac-delta points of mass, Eq. (14) includes function ν_i to effect a spatial "smearing out" of the molecule

$$\nu_i(q) = e^{-q^2r_{g,i}^2/6}. \quad (15)$$

In this formula, each molecule of species i is assumed to have an isotropic spatial distribution of mass with $r_{g,i}$ being the radius of gyration. Equation (15) is consistent with "Guinier's law" that is used in interpreting scattering experiments and is considered reasonably accurate even for nonspherical molecules when $qr_{g,i}$ is of order 1 or less [22]. Because function ν_i does not depend on the dynamics, if desired it can be omitted from Eq. (14) and reintroduced into the results after the conclusion of the simulation. In the present work we use Lennard-Jones particles to represent structured molecules and take $\nu_i(q)=1$ for simplicity.

B. Sampling volumes

Our objective is to choose a sampling volume that is commensurate with the periodic concentration fluctuations inside the cell—in fact we choose a separate sampling volume for

each lattice vector $\mathbf{q} \neq 0$. We then calculate the instantaneous particle-number fluctuation, $\delta N_i = N_i - \langle N_i \rangle$, corresponding to each sampling volume. This can be accomplished by integrating the concentration deviation $\delta\rho_i$ across the cell using a normalized weighting function Ω to select the desired regions:

$$\begin{aligned}\delta N_i(\mathbf{q},t) &= \int_V \delta\rho_i(\mathbf{r},t)\Omega(\mathbf{q}\cdot\mathbf{r})d\mathbf{r} \\ &= \rho \sum_{\mathbf{q}'\neq 0} \psi_i(\mathbf{q}',t) \int_V e^{\mathbf{i}\mathbf{q}'\cdot\mathbf{r}}\Omega(\mathbf{q}\cdot\mathbf{r})d\mathbf{r} \\ &= N\psi_i(\mathbf{q},t), \end{aligned} \tag{16}$$

where

$$\Omega(s) = (1 - \mathbf{i} + e^{-\mathbf{i}s}). \tag{17}$$

The weighting function is complex or in other words the sampling volume is complex, serving to maximize the computational efficiency of the method by incorporating information on both phase and amplitude of fluctuations. Because Ω returns nondiscrete values, the sampling volumes have "fuzzy" boundaries, leading to a continuous change in occupancy as particles move relative to the boundaries. This is in contrast to the method of Debenedetti [21], in which fluctuations result in strictly integer occupation numbers for each sampling volume and time. The weighting function given by Eq. (17) was chosen specifically to isolate fluctuations for a single lattice vector, that is, to generate $\delta N_i(\mathbf{q},t)$ that depends only on a single corresponding Fourier coefficient $\psi_i(\mathbf{q},t)$. In fact, the particular geometry of the sampling volume is immaterial as long as it fulfills this role.

The piecewise sampling volumes for each lattice vector effectively constitute separate fluctuation experiments, each yielding a set of δN_i that goes into Eq. (1). For convenience we reuse the notation $\mathbf{S} = \mathbf{A}^{-1}$ to give

$$S_{ij}(\mathbf{q}) = \frac{1}{N}\langle \delta N_i(\mathbf{q},t)\cdot\delta N_j(-\mathbf{q},t)\rangle = N\langle \psi_i(\mathbf{q},t)\cdot\psi_j(-\mathbf{q},t)\rangle. \tag{18}$$

Note that $\psi_j(-\mathbf{q},t)$ is the complex conjugate of $\psi_j(\mathbf{q},t)$. Any imaginary quantities inside the angle brackets average to zero. Thus, as expected, matrix $\mathbf{S}$ is real and symmetric. One also wants to be sure that $\langle \psi_i \rangle \approx 0$ as expected for homogeneous single-phase systems. One way to gauge this is to compare S_{ij} above to

$$S^0_{ij}(\mathbf{q}) = N\langle \psi_i(\mathbf{q},t)\rangle\langle \psi_j(-\mathbf{q},t)\rangle \tag{19}$$

calculated from the same trajectory. Expressions for $S_{ij}(q)$ in terms of molecular positions have previously been noted [23], but we are not aware of such being used previously to simulate Kirkwood-Buff quantities as is done in this work.

With $\mathbf{S} = \mathbf{A}^{-1}$ in hand, we can obtain a self-consistent set of thermodynamic properties for any lattice vector of our choice, excluding $\mathbf{q} = 0$. The calculated properties in general have a dependence on $\mathbf{q}$; although for isotropic materials the results should depend only on magnitude $q = |\mathbf{q}|$. Therefore, the results for equivalent discrete vectors (same q value) are averaged following the ensemble averaging in Eq. (18). As with Eq. (4), in order to obtain properties in the thermodynamic limit (to compare to macroscopic experiments), one must then extrapolate the results to $q = 0$. Obviously the accuracy of this extrapolation depends on simulation uncertainties and how rapidly a particular property varies as $q \to 0$. The extrapolation procedure used in this work is described below, accompanying the simulation results.

For an infinite-size system, the definitions of $\mathbf{S}$ in Eqs. (4) and (18) are equivalent. However, for finite systems, the pathology observed for KB integrals [Eq. (3)] is also observed for the integral in Eq. (4). That is, truncation of the integrals at maximum available radial distances can lead to large errors. The advantage of Eq. (18) relative to Eq. (4) is that the reformulation is based on a discrete Cartesian-based Fourier transform, rather than a continuous, spherically symmetric Fourier transform. The sampling volumes of the method do not truncate intermolecular correlations at a maximum radial distance; no assumption is made that $g_{ij} = 1$ at large intermolecular separations, an assumption generally violated for closed systems [2]. In summary, we have a means of calculating matrix $\mathbf{A}$ that is consistent with periodic boundary conditions and is immune to long-range truncation effects.

C. Extending the cut-off radius of KB integrals

To better understand and possibly ameliorate the limitations of KB theory as it is normally implemented in periodic simulations, we examined the effect of the cutoff or truncation radius r_c on the KB integrals and resulting thermodynamic properties. Changing r_c is in effect changing the fraction of the unit cell that is included in the sampling volume.

What happens if one extends r_c beyond $0.5L$ in a periodic cubic system? It leads to biased results if the integration is done in a purely spherical fashion because then one is beginning to double-count interactions between some pairs of molecule images and to incorporate into G_{ij} anomalous long-range order resulting from the periodicity of the system. For instance, at $r_c = L$ one would observe a molecule interacting with its own image, leading to a crystal-like spike in g_{ii}. Instead, for $r_c > 0.5L$ we extend the KB integrals "into the corners" of the unit cell to include additional unique molecule pairs without double counting. In effect this makes the integration volume a sphere circumscribed or bounded by a co-centered cube of length L. The running KB integrals for a periodic cubic system become

$$G_{ij}(r_c) = \int_0^{r_c} [g_{ij}(r) - 1]\frac{dv_s}{dr}dr. \tag{20}$$

Function $v_s(r)$, which is plotted in Fig. 1, provides the bounded volume out to radius r:

$$v_s(r) = L^3 \begin{cases} \dfrac{\pi}{6}x^3 & \text{if } 0 < x \leq 1 \\ -\dfrac{\pi}{3}x^3 + \dfrac{3\pi}{4}x^2 - \dfrac{\pi}{4} & \text{if } 1 < x \leq \sqrt{2} \\ p(x) & \text{if } \sqrt{2} < x \leq \sqrt{3}, \end{cases} \tag{21}$$

where $x = 2r/L$ and

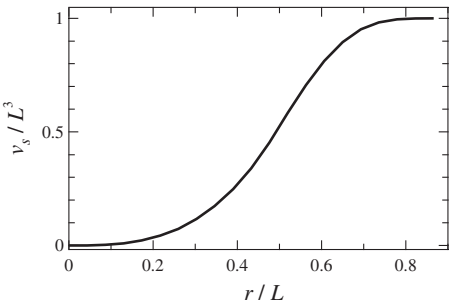

FIG. 1. Function $v_s(r)$ giving the volume of a sphere of radius r bounded by a co-centered cube of length L.

$$p(x) = \sqrt{x^2-2} + (1-3x^2)\left[\operatorname{atan}(\sqrt{x^2-2}) - \frac{\pi}{4}\right] + x^3\left[\operatorname{atan}(x\sqrt{x^2-2}) - \frac{\pi}{3}\right]. \tag{22}$$

One can calculate directly G_{ij} from recorded pair distances during a simulation without relying on numerical integration:

$$G_{ij}(r_c) = v_{ij}(r_c) - v_s(r_c), \tag{23}$$

where

$$v_{ij}(r_c) = \frac{V}{N_i(N_j - \delta_{ij})}\left\langle \sum_{a\in i}\sum_{\substack{b\in j \\ b\neq a}} \Phi(r_c - r_{ab}) \right\rangle, \tag{24}$$

where Φ is the Heaviside step function and r_{ab} is the nearest-image distance between molecules a and b. The double sum is over all molecules a of species i and molecules b of species j in the system, excluding $b=a$ in the case $i=j$. Function v_{ij} is implemented using histograms (arrays of bins) in terms of discrete r_c values. Furthermore, one can obtain the radial distribution function $g_{ij}(r)$ by combining Eqs. (20) and (23) and taking the derivative with respect to r_c. The result is $g_{ij}(r) = dv_{ij}/dv_s|_r$, where the derivative is implemented as a finite difference using the histograms created for v_{ij}.

Running KB integrals basically exhibit a pattern of damped oscillations. Thus, another step to improve the results is to employ a smoothing filter so that the results are less sensitive to the exact choice of integral cutoff r_c. Here we use a Gaussian kernel to effect the smoothing of some property f calculated from running KB integrals, such as an activity-coefficient correction, partial volume, or compressibility. The smoothed value is

$$f^*(r_c) = \frac{\int_0^{r_m} \exp\left[-\frac{1}{2}s^{-2}(r-r_c)^2\right] f(r)dr}{\int_0^{r_m} \exp\left[-\frac{1}{2}s^{-2}(r-r_c)^2\right] dr}, \tag{25}$$

where s is a smoothing-distance parameter, taken here to be 0.3 nm, and $r_m = \sqrt{0.75}L$ is the maximum distance available to the KB integrals.

TABLE I. Lennard-Jones parameters for binary mixture of CF_4 (species 1) and CH_4 (species 2).

	Species pair		
	11	22	12
σ(nm)	0.4150	0.3728	0.3939
ϵ/k_B(K)	175.0	149.0	142.1

IV. SIMULATION RESULTS AND DISCUSSION

A. Example LJ binary mixture

In order to test the results of the proposed methods, we selected and simulated a series of liquid state points for a nonideal binary Lennard-Jones mixture. The molecular-dynamics method was used. The equations of motion included an integral-control (Nosé-Hoover) thermostat and were integrated using a fourth-order Gear predictor-corrector scheme [24]. The radial cutoff for intermolecular forces was set equal to $0.3L$ for the simulations with $N=1200$ and to $0.2L$ for the single simulation with $N=4000$—in each case sufficient to include the three nearest shells of neighbors about each molecule. The standard long-range correction to pressure was used to account for interactions beyond the force cutoff. The size of the time step was selected for each simulation to generate a root-mean-square displacement of molecules of 1.5 fm per time step. A Verlet neighbor list was used to speed up computations [24]. ψ_i samples were taken at the same time as each update to the Verlet neighbor list, or approximately every 45 time steps.

Table I gives the LJ parameters for the binary system. Following Schoen *et al.* [25], the parameters for the two components were chosen to imitate carbon tetrafluoride (species 1) and methane (species 2), two molecules that are roughly spherical. However, we took the methane parameters from Vrabec and Fischer [26] because they better reproduce liquid density than the methane parameters used by Schoen *et al.* As for the cross interactions, we used the Lorentz combining rule, $\sigma_{12} = \frac{1}{2}(\sigma_{11} + \sigma_{22})$, but used a modified Berthelot combining rule, $\epsilon_{12} = (1-k_{12})\sqrt{\epsilon_{11}\epsilon_{22}}$, where $k_{12}=0.12$. This value was determined to approximate the experimental upper solution critical point of $T_{\text{crit}}=94.5$ K and $y_{1,\text{crit}}=0.43$ [25].

We assessed the presence of the liquid-liquid critical point by performing preliminary *NPT* simulations at $y_1=0.4$ and two temperatures that bracket the critical point, 90 and 100 K. We compared the averages of S_{ij}^0 [see Eq. (19)] at the two points to ensure significant demixing only at the lower temperature. A sharp increase in $|S_{ij}^0|$ as $q \to 0$ indicates stationary large-wavelength variations in composition, which in turn indicate an incipient or developed phase split. Schoen *et al.* likewise used a trial-and-error procedure around the critical point to arrive at what would be the equivalent of $k_{12}=0.30$ in the ϵ_{12} combining rule. However, this larger k_{12} value caused demixing over a wide composition range in our simulations at 120 K. The system size used by Schoen *et al.* was nearly five times smaller than that used in this work and therefore may have suppressed the phase separation we observed.

TABLE II. Parameters and results for *NVT*-simulated LJ binary systems, where $T=120$ K.

	Mole fraction y_1								
	0	0.1	0.2	0.4	0.4	0.6	0.8	0.9	1
$N/1000$	1.2	1.2	1.2	1.2	4	1.2	1.2	1.2	1.2
ρ^{-1}(cm^3/mol)	39.0	40.6	42.1	44.6	44.6	46.9	48.8	49.6	50.6
$\langle P\rangle$(MPa)	0.90	0.75	1.07	0.93	0.87	0.84	1.12	1.05	0.91
time (ns)	3.47	3.62	3.80	4.24	2.12	4.87	5.91	6.76	8.14
L(nm)	4.27	4.32	4.38	4.46	6.67	4.54	4.60	4.62	4.65

The simulated state points and a summary of results are presented in Table II. In order to imitate a conventional mixing experiment conducted at constant temperature and pressure, a series of preliminary *NPT* simulations was performed in order to determine appropriate cell volumes to use. Then *NVT* simulations of length 10^6 time steps (half that for the one larger simulation) following equilibration were performed to obtain the results in Table II.

In order to facilitate comparison with other work based on the LJ model, Table III lists the pure-component state points and results from Table II in commonly used dimensionless form. Both of these liquid state points are close to the LJ vapor-liquid coexistence curve. In fact, pure component 1 (representing CF_4) is essentially at the LJ triple point [27]. Careful equilibration (avoiding large temperature swings, for instance) was necessary to generate reliable starting configurations for our simulations at 120 K due to the system's proximity to both liquid-liquid and vapor-liquid two-phase regions.

B. Reference properties

In order to compare the results of the two KB methods and ascertain reliability, we used two correlations for mixture and pure-component properties. The first was the LJ-fluid equation of state (LJ EOS) by Johnson *et al.* along with van der Waals one-fluid mixing rules [27]. They showed that this method reproduces simulated vapor-liquid equilibrium results for LJ-fluid mixtures quite well, and our results confirm the accuracy of the method. The second correlation method we employed was the well-known UNIFAC activity-coefficient model (in this case indistinguishable from UNIQUAC [5]) as another way to get reference values of Q_{11}. Size and energy parameters for CH_4 and CF_4 species were obtained from Gmehling and co-workers [28,29] and are given in Table IV. It should be noted that these parameters overpredict the liquid-liquid critical temperature by about 9 K.

TABLE III. Dimensionless state points for *NVT*-simulated LJ pure-component systems.

	$i=1$	$i=2$
$T^+=k_BT/\epsilon_i$	0.686	0.805
$\rho^+=N\sigma_i^3/V$	0.851	0.799
$P^+=P\sigma_i^3/\epsilon_i$	0.027	0.023

C. Overall results

Figure 2 provides a summary and comparison of the results for the traditional implementation of KB theory ("old KB") and for the Fourier-based modification ("new KB"), both having been calculated simultaneously for each of the simulations in Table II. It is important to note that all the old KB points in Fig. 2 are "best case" results based on smoothing of properties to reduce oscillations and choice of an optimal cutoff as discussed below. The volumes presented here and in the following figures are given on a molar basis rather than a molecular basis for ease in interpretation. Figure 2 also gives reference curves of the relevant thermodynamic properties in order to assess accuracy of the methods. The LJ EOS (plus mixing rules) matches the simulated constant-pressure mixture volumes (labeled "*NPT*") well, differing by no more than 0.3% from any point. This gives confidence that this reference method is reasonably accurate for volumetric properties, namely, $\bar{V}_i$ and κ_T. Furthermore, the agreement between the LJ EOS and UNIFAC for Q_{11} is quite good, lending additional confidence in the accuracy of the LJ EOS for all three properties simulated.

For the system and properties studied here, the Fourier modification to the KB method produces results in greater agreement with the LJ EOS than does the traditional KB method. As shown in Fig. 2(a), both KB methods are generally reliable for partial molar volumes. There are a couple noticeable deviation trends, however. First, the modified KB method predicts partial molar volumes that depart further from the mixture-volume curve than does the LJ EOS. On the other hand, the traditional KB method predicts partial volumes that are closer to the mixture-volume curve, with significant deviations for dilute components. The origins of both of these biases are discussed below. In both cases the larger simulation size generated little change in the results.

As shown in Fig. 2(b), the traditional KB method—even with smoothing—produces significant errors in κ_T, consis-

TABLE IV. UNIFAC size and energy parameters for CF_4 (1) and CH_4 (2) [28,29].

$R_1=1.780$	$Q_1=1.820$	$a_{12}=142.06$ K
$R_2=1.129$	$Q_2=1.124$	$a_{21}=72.986$ K

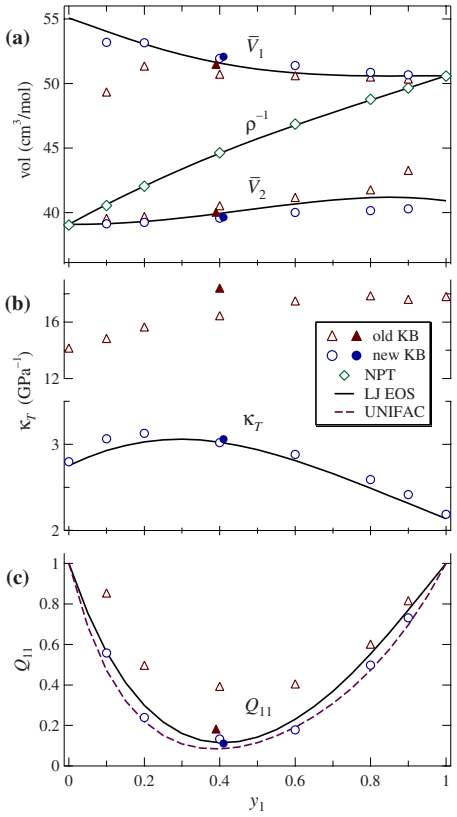

FIG. 2. (Color online) Comparison of calculated results upon changing composition for (a) partial molar volumes $\bar{V}_i$ and mixture volume ρ^{-1}, (b) isothermal compressibility κ_T, and (c) activity coefficient correction Q_{11}. Open and closed symbols indicate results from $N=1200$ and $N=4000$ simulations, respectively; closed symbols have been offset horizontally for clarity. Lines indicate reference correlations.

tently overpredicting by nearly an order of magnitude. The modified KB method shows remarkably good agreement with the LJ EOS, with a slight overprediction bias. Notably, the method correctly predicts the maximum in κ_T with changes in composition for this nonideal mixture. Figure 2(c) similarly shows the inherent bias toward mixture ideality in predicted Q_{11} values for the traditional KB method. The modified KB method, however, reproduces the mixture correlation curves quite well.

D. Analysis of the traditional KB method

We can generalize the old KB results in Fig. 2: the method always biases properties, to a greater or lesser degree, toward the result for an ideal-gas mixture. An ideal gas mixture produces $\bar{V}_i=\rho^{-1}$, $\kappa_T=\kappa_T^{ig}$, and $Q_{11}=1$. To better explain the biases in both methods, we present in Figs. 3 and 4 additional detail at one composition point ($y_1=0.4$) on the results and procedure for the respective methods. Simulations where $N=1200$ and $N=4000$ are included in order to assess system-size effects.

Figure 3(a) gives the unsmoothed elements of matrix $\mathbf{S}=\mathbf{A}^{-1}$, which come from running KB integrals. Each S_{ij} value is proportional to the corresponding G_{ij} value, and so these curves contain the same information on fluid structure found in KB integrals. The other thermodynamic properties, shown in parts (b) through (d), are in turn derived from the S_{ij} curves. Circles in Fig. 3 indicate the distance cutoff frequently employed by practitioners of traditional KB theory. This cutoff, while certainly convenient, does not necessarily lead to the most reliable results in a closed periodic system. Recall that beyond this point ($r_c>0.5L$) we are taking the KB integrals into the corners of the cube according to Eq. (23), and so the maximum evaluation distance is $r_m=\sqrt{0.75}L$ for each simulation. At this end point all unique pairs of molecules have been included exactly once in the KB integrals and $G_{ij}(r_m)=0$ for a closed system, which is a consequence of mass conservation. An ideal-gas mixture is defined by all $G_{ij}=0$. Thus, at the two radial limits of $r_c=0$ and $r_c=\sqrt{0.75}L$ in Fig. 3 we have ideal-gas-mixture behavior with $S_{ij}=\delta_{ij}y_i$. It is clear from the curves that either neglect of molecular interactions (on the left) or the closed-system constraint (on the right) increasingly bias the properties as an end point is approached, leading to the conclusion that the least amount of bias accrues if one chooses an intermediate distance for property evaluation. As already indicated the old KB results in Fig. 2 are based on smoothing of properties [Eq. (25)]; the cutoff was $r_c=0.45L$, a value intended to produce results least biased toward ideal-gas-mixture behavior.

Under what conditions would the traditional KB method produce results substantially free of bias or in agreement with the EOS for our LJ mixture? It is difficult to predict, but a rough estimate based on the results in Fig. 3 suggests a system size of at least $N=20\,000$ would be required for the traditional KB method to match results with the same quality as obtained by the new KB method for $N=1200$. Such an undertaking is beyond the scope of the present work.

E. Analysis of the Fourier KB method

Figure 4 provides an elaboration of the procedures to calculate the new KB results in Fig. 2. With this Fourier method we calculate and average a full set of S_{ij} and corresponding thermodynamic properties (the open symbols in Fig. 4) for a range of discrete q values. Then we extrapolate to the $q=0$ or thermodynamic limit, which in principle eliminates the biases of simulating in a closed finite system. However, extrapolation is always a hazardous affair—in addition to the inherent amplification of sample noise, there can be a systematic bias due to an incorrectly assumed functionality as $q\to 0$.

We attempted several different extrapolation procedures, including one based on Ornstein-Zernicke theory. By combining the Fourier transform of the multicomponent Ornstein-Zernicke equation [23] with Eq. (4), we obtain

$$\rho\mathbf{C}=\mathbf{Y}^{-1}-\mathbf{S}^{-1}, \tag{26}$$

where $\mathbf{C}(q)$ is a matrix of Fourier transforms of pairwise direct correlation functions $c_{ij}(r)$. With Eq. (26) we can obtain $\mathbf{C}(q)$ from simulated $\mathbf{S}(q)$ values. The postulated asymptotal form of the direct correlation function, consistent with commonly used closure relations for pure-component fluids, is $\lim_{r\to\infty} c(r)=-\phi(r)/(k_BT)$, where ϕ is the pairwise inter-

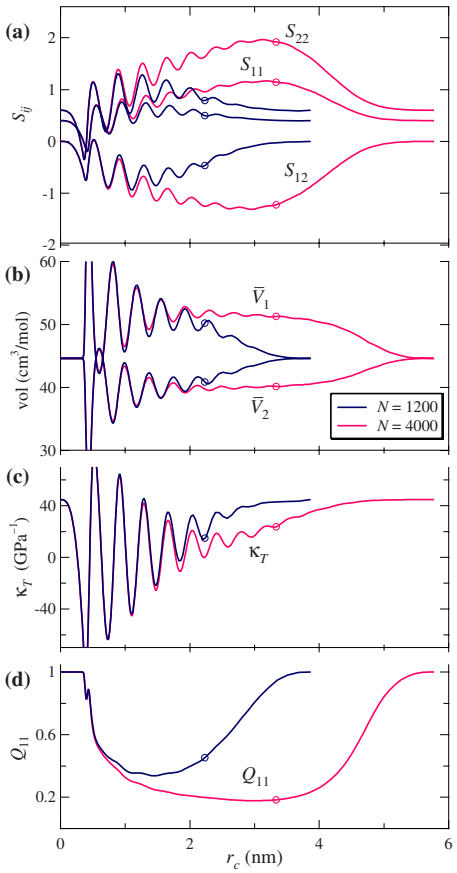

FIG. 3. (Color online) Properties derived from running KB integrals for two different system sizes at $y_1=0.4$. Circles indicate the oft-used $r_c=0.5L$ cutoff in traditional KB calculations.

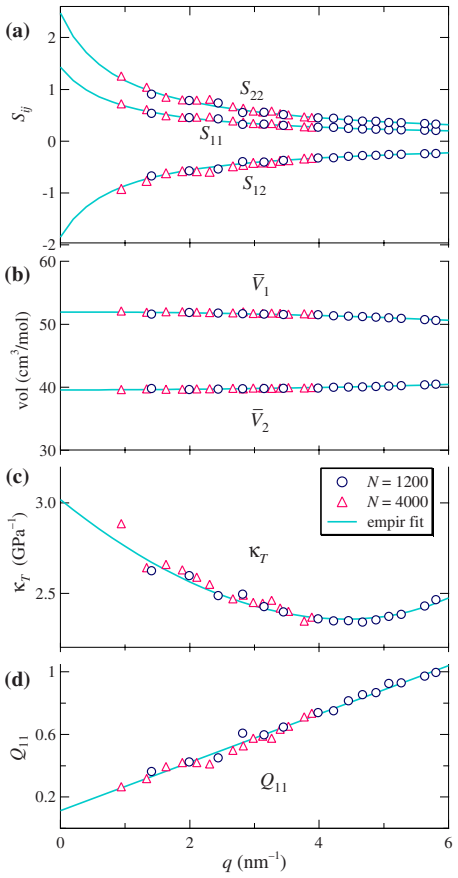

FIG. 4. (Color online) Properties derived from Fourier analysis for two different system sizes at $y_1=0.4$. The lines indicate empirical fits to the $N=1200$ results as explained in the text.

molecular potential [10]. If we assume that $c_{ij}(r)$ is finite at small r values and transitions to $c_{ij}(r) \propto r^{-6}$ at large r values (based on the long-range dispersion interaction), we arrive at

$$C_{ij}(q) = C_{ij}^A + C_{ij}^B q^2 + O(q^3), \tag{27}$$

where C_{ij}^A and C_{ij}^B are constants that depend on intermolecular potentials and the state point in a complicated way. We therefore truncated Eq. (27) at second order and took C_{ij}^A and C_{ij}^B to be adjustable fitting parameters (six independent values) to match Eq. (27) to simulated $\mathbf{C}(q)$ values. Unfortunately, this predicted asymptotal functionality did not fit the data well, as is shown in Fig. 5. One possible cause is the proximity of our simulated state point to a critical point; critical points are known to violate an assumption used to generate the asymptotal form of $c(r)$ [10]. More specifically, the $c_{ij}(r)$ may not approach their asymptotal forms rapidly enough for Eq. (27) to be accurate for the range of q values examined in this work.

In the end we arrived at the following empirical extrapolation procedure that fits the simulated data well and appears to produce relatively little bias for $q \to 0$ as shown in Figs. 2 and 4. Rather than extrapolate directly the $C_{ij}(q)$ or raw $S_{ij}(q)$ values, we instead calculated the three independent properties $\bar{V}_1$, κ_T, and Q_{11} from simulated $S_{ij}(q)$ and fit the three with separate functions of q: $\bar{V}_1(q)=V_1^A+V_1^B q^2$, $\kappa_T^{-1}(q)=K_T^A+K_T^B q+K_T^C q^3$, and $Q_{11}(q)=Q_{11}^A+Q_{11}^B q$. This means we had a total of seven adjustable fitting parameters $(V_1^A, V_1^B, K_T^A, K_T^B, K_T^C, Q_{11}^A, Q_{11}^B)$. The reason for correlating with the *inverse* of κ_T, also known as the bulk modulus, is that this function is bounded or analytic even at critical points, as are $\bar{V}_i$, Q_{ij}, and C_{ij}.

Although not necessary, for pedagogical purposes we took the fitted curves and, using the inversion procedure of Eq. (11), produced the equivalent fitted curves for $S_{ij}(q)$ shown in Fig. 4(a). All of the curves in Fig. 4 are based on the $N=1200$ simulation points. The data from the larger simulation are quite consistent with those for the smaller simulation and corresponding fitted curves (not shown) yield essentially the same $q=0$ intercepts, as is apparent from the points at $y_1=0.4$ in Fig. 2. Obviously this empirical extrapolation procedure is not perfect and hence some small biases relative to the LJ EOS remain.

For the sake of completeness, we show in Fig. 5 a comparison of simulated $\rho C_{ij}(q)$ values and the corresponding fitting curves using the empirical and asymptotic forms. In both cases the fits were made to the $N=1200$ simulation data only. The empirical fits (solid lines) were generated from the curves for $\bar{V}_1$, κ_T, and Q_{11}, as was done in Fig. 4(a). The asymptotic fits (broken lines) are least-squares fits of Eq. (27) to the simulated $\mathbf{C}$ values. The empirical fits are consistent with the $N=4000$ results and appear to provide a much

more reliable extrapolation to $q=0$ than do the asymptotic fits for a similar number of adjustable fitting parameters.

V. CONCLUSION

The original formulation of Kirkwood-Buff solution theory provides a molecule-centered picture of fluctuations. The Fourier method implemented here provides a simulation-cell- or laboratory-frame-based picture of fluctuations. Each gives distinct information about the fluid heterogeneities and in principle should lead to the same values of thermodynamic derivatives.

KB solution theory formally requires an open system, but the theory is nevertheless frequently applied to closed systems. The biases in the running KB integrals (and subsequent thermodynamic properties) at large r values are manifestations of the principal that $g_{ij}(r\to\infty)=1$ is violated for closed systems. One is assured that in the thermodynamic limit ($N\to\infty$) the statistical mechanical distinctions between open and closed systems are erased. This means that the original KB theory when incorrectly applied to a closed system nevertheless converges to the correct solution with increasing N. Yet, one wishes to adopt methods where this convergence is improved. This work demonstrates an implementation of KB theory that fills such a need. For *NVT* systems the method appears to be more accurate in calculating macroscopic thermodynamic derivatives than the old method, and it may have advantages for other ensembles (including open systems) as well although that was not examined here. In any case, neither the old nor new KB methods add much computational overhead, and so both can be included in a computer simulation code to enable comparisons to be made.

We showed how to extend KB integrals into the corners of the simulation cell in order to make maximal use of available information on radial pair correlations under the old KB method. When one changes the cut-off distance r_c, one changes the range over which pair correlations are included but also the degree to which the constraint of a closed system is imposed upon the KB integral values. It is difficult to decouple spatially these two effects unless the system size is quite large. Small-wavelength fluctuations in pair correlations can be moderated by a smoothing procedure. Correcting for the closed-system constraint is more problematic. Simply truncating Kirkwood Buff integrals at $r_c=0.5L$, as is commonly done, is not a rigorous solution to the closed-system problem but merely disguises or neglects it. One cannot avoid the fact that for any finite closed system the traditional KB method gives results that are biased toward ideal-gas-mixture values.

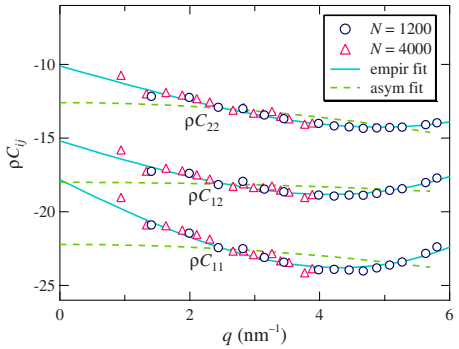

FIG. 5. (Color online) Fourier transform of pairwise direct correlation functions for two different system sizes at $y_1=0.4$. The lines indicate two different fits to the $N=1200$ results as explained in the text.

In this work we simulated a simple system representing the nonideal liquid mixture of CF_4 and CH_4, for which reference correlations of mixture properties are available. We did not intend to fully explore the phase diagram of this mixture nor to come up with the most realistic intermolecular potentials; however, the spherically symmetric LJ potentials used here appear to capture quite accurately volumetric and free-energy-related properties for this system.

Simulating near a critical point can be particularly challenging; even though we simulated ostensibly in a single-phase region about 20 K above the liquid-liquid critical point, we observed strong molecular fluctuations and occasional demixing. This necessitated long equilibration and simulation times for reliable single-phase results. The strong nonideality also made extrapolation of Fourier quantities to the $q=0$ wave vector particularly challenging.

We believe the adaptation or extension of KB theory given here will better enable simulation of thermodynamic derivatives for a wide range of molecules and mixtures. Although the Lennard-Jones fluid is often considered a relatively ideal system upon which to test theories, this is not necessarily the case with respect to KB solution theory. For instance, water and other small polar molecules simulated near their triple points have running KB integrals that dampen and approximately converge to their macroscopic asymptotal values within about 3 molecular diameter [30], compared to 6 or more diameter for LJ fluid.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, Grant No. CTS 0547610. Brigham Young University assisted in meeting the publication costs of this paper.

[1] D. Frenkel and B. Smith, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic Press, San Diego, 2002).
[2] J. G. Kirkwood and F. P. Buff, J. Chem. Phys. **19**, 774 (1951).
[3] J. E. Mayer, J. Chem. Phys. **18**, 1426 (1950).
[4] G. M. Wilson, J. Am. Chem. Soc. **86**, 127 (1964).
[5] J. Prausnitz, R. Lichtenthaler, and E. G. de Azevedo, *Molecular Thermodynamics of Fluid Phase Equilibria*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1986).
[6] K. E. Newman, Chem. Soc. Rev. **23**, 31 (1994).

[7] *Fluctuation Theory of Mixtures, Advances in Thermodynamics Series*, edited by E. Matteoli and G. A. Mansoori (Taylor & Francis, New York, 1990).
[8] A. Ben-Naim, J. Phys. Chem. B **111**, 2896 (2007).
[9] A. Perera, F. Sokolic, L. Almasy, P. Westh, and Y. Koga, J. Chem. Phys. **123**, 024503 (2005).
[10] N. March and M. Tosi, *Atomic Dynamics in Liquids* (Dover Publications, New York, 1976).
[11] R. Chitra and P. E. Smith, J. Chem. Phys. **114**, 426 (2001).
[12] C. Curtiss and R. Bird, Ind. Eng. Chem. Res. **38**, 2515 (1999).
[13] D. R. Wheeler and J. Newman, J. Phys. Chem. B **108**, 18353 (2004).
[14] D. R. Wheeler and J. Newman, J. Phys. Chem. B **108**, 18362 (2004).
[15] C. W. Monroe and J. Newman, Ind. Eng. Chem. Res. **45**, 5361 (2006).
[16] D. L. Jolly and R. J. Bearman, Mol. Phys. **41**, 137 (1980).
[17] M. Schoen and C. Hoheisel, Mol. Phys. **52**, 33 (1984).
[18] J. Lebowitz and J. Percus, Phys. Rev. **122**, 1675 (1961).
[19] E. Matteoli and G. A. Mansoori, J. Chem. Phys. **103**, 4672 (1995).
[20] D. McGuigan and P. Monson, Fluid Phase Equilib. **57**, 227 (1990).
[21] P. G. Debenedetti, J. Chem. Phys. **86**, 7126 (1987).
[22] G. Beaucage, J. Appl. Crystallogr. **28**, 717 (1995).
[23] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic Press, New York, 1986).
[24] M. Allen and D. Tildesley, *Computer Simulation of Liquids* (Oxford Science, 1989).
[25] M. Schoen, C. Hoheisel, and O. Beyer, Mol. Phys. **58**, 699 (1986).
[26] J. Vrabec and J. Fischer, Int. J. Thermophys. **17**, 889 (1996).
[27] J. K. Johnson, J. A. Zollweg, and K. E. Gubbins, Mol. Phys. **78**, 591 (1993).
[28] S. Horstmann, A. Jabłoniec, J. Krafczyk, K. Fischer, and J. Gmehling, Fluid Phase Equilib. **227**, 157 (2005).
[29] T. Holderbaum and J. Gmehling, Fluid Phase Equilib. **70**, 251 (1991).
[30] L. Zoranic, F. Sokolic, and A. Perara, J. Mol. Liq. **136**, 199 (2007).