

Hydrophobicity in a simple model of water: Entropy penalty as a sum of competing terms via full, angular expansion

Kevin A. T. Silverstein^{a)} and Ken A. Dill^{b)}

Department of Pharmaceutical Chemistry and Graduate Group in Biophysics, University of California, San Francisco, California 94143-1204

A. D. J. Haymet^{b),c)}

Department of Chemistry and Institute for Molecular Design, University of Houston, Houston, Texas 77204-5641

(Received 13 December 2000; accepted 24 January 2001)

The entropy penalty of solvation for nonpolar solutes dominates the hydrophobic effect at room temperature. We find that this entropy arises from a competition between a relatively localized “two-body” term, and a contribution arising from non-pairwise-decomposable three-body and higher-order terms. We use a full, angular dependent, expansion of solute–water correlation functions over the full range of fluid temperatures for a two-dimensional model of water. This water model has been shown to capture many of the basic anomalies of water and aqueous solutions of sparingly soluble nonpolar molecules, including the volume anomalies of water and the thermal anomalies of the hydrophobic effect. Our results show that for hot liquid water, the two-body approximation is sufficient to estimate the transfer entropy, but in cold liquid water, which is the main regime for biological hydrophobic interactions, the two-body assumption substantially overestimates the degree of ordering in water. © 2001 American Institute of Physics. [DOI: 10.1063/1.1355997]

I. SIGNATURES OF HYDROPHOBICITY

A signature of hydrophobic solvation at room temperature is the large negative entropy associated with inserting a nonpolar solute into water. This entropy penalty, which has been ascribed¹ to order induced by the solute in the surrounding solvent, leads to the low solubility of inert molecules in water at room temperature. What is the nature of this solvation entropy? We introduce the hypothesis² that the solvation entropy arises from the competition of (1) a two-body entropy, fairly local in nature, which highly orders the water surrounding a small, nonpolar solute, and (2) a three- and higher-body term, which cannot be decomposed into pairwise additive terms, which around room temperature and colder causes major disorder in the solvent.

Since the hydration thermodynamics, including the entropy, observed in real water is *also* observed for a small disk solute in a simple, two-dimensional model of water, recently studied by us,^{2–6} called the Mercedes–Benz (MB) model, we have exploited the reduced dimensionality of the simple model to test our hypothesis, and to calculate the magnitude and sign of certain terms in the full angular expansion of the entropy.

We find that (the ensemble invariant analog of) the simple “ $g \ln g$ ” term greatly *overestimates* the true entropy penalty at the temperatures of interest. The two-body term overestimates the order induced in the model water. There-

fore, higher-order terms must describe additional *disorder* in the liquid, and the sum of the terms yields the correct, observed, thermodynamic entropy penalty.

This paper is organized as follows. In Sec. II we summarize briefly the background for calculating entropy via multiparticle correlation functions, and the explicit formulas used in this work for both pure water and the solution. In Sec. III we describe the method by which a Fourier series expansion can be used to greatly speed up the convergence of the angular correlation functions, after briefly reviewing the model and computational procedures used. In Secs. IV and V, the results of our calculations for bulk water and a dilute nonpolar solution, respectively, are presented.⁷ Section VI summarizes our results, and in the appendixes we collect the mathematical details for (A) the angular expansion we employ, and (B) the entropy formula.

II. ENTROPY EXPANSIONS

A. The Kirkwood–Green expansion

For liquids with only pairwise-additive interactions between particles, many thermodynamic quantities, such as the average potential energy and the pressure, may be calculated directly from the pair correlation functions. However, for other important statistical quantities, such as the entropy and free energy, an analogous direct and exact calculation is not possible. Typically, such quantities have been calculated indirectly using the techniques of thermodynamic integration and histogram-reweighting methods, but these methods require significant computer resources.

A direct method for calculating the entropy (and hence the free energy) from the pair correlations alone has been explored for a variety of liquids.^{8–14} This method is based on

^{a)}Present address: Computational Biology Centers, Academic Health Center, University of Minnesota, Mayo Mail Code 43, 420 Delaware St SE, Minneapolis, Minnesota 55455.

^{b)}Authors to whom correspondence should be addressed.

^{c)}Electronic mail: haymet@uh.edu

truncation of an exact, but infinite-order, expansion for the entropy in terms of multiparticle correlation functions. Even for systems having only pairwise-additive energies, the expansion of the entropy in terms of multiparticle correlation functions does not truncate at the pair level, but instead has contributions from all orders.^{15,16} At present the calculation of multiparticle correlation functions is an extremely difficult task, and recent work has focused on manipulating the entropy expansion in order to obtain the best possible estimate of the entropy from the pair correlation functions alone.¹²

Building upon the work of Kirkwood,¹⁷ Green,¹⁵ Nettleton and Green,¹⁶ and Raveché,¹⁸ Hernando^{10,11,19} recast the entropy expansion into a particularly useful form, via a partial resummation of the higher-order terms. This expression has been generalized to a mixture of any number, μ , of components by Laird *et al.*,^{20,21} and applied to electrolytes systems by Laird and Haymet.²²

The resulting expansion may be written as the sum of four terms:

$$S[g^{(n)}]/Nk = s^{\text{ideal}} + s^{(2)} + s_{\text{ring}} + \sum_{i=3}^{\infty} s'_{\text{mix}}(i)[g_{\alpha\beta\dots}^{(m)}; m \leq i]. \quad (1)$$

The functions $g_{\alpha\beta\dots}^{(m)}(\mathbf{r}_\alpha, \mathbf{r}_\beta, \dots)$ are the m -particle (multi-component) correlation functions of the liquid. In the ensemble invariant form,⁸ the first term

$$s^{\text{ideal}} \equiv \frac{5}{2} - \sum_{\alpha=1}^{\mu} x_\alpha \ln \rho_\alpha \Lambda_\alpha^3 \quad (2)$$

is the ideal gas contribution, where α is the species label, Λ_α is the de Broglie thermal wavelength, ρ_α is the number density of component α , $\rho = \sum_{\alpha=1}^{\mu} \rho_\alpha$ is the total number density, and $x_\alpha = \rho_\alpha / \rho$ is the mole fraction. The second term,

$$s^{(2)} \equiv -\frac{\rho}{2} \sum_{\alpha=1}^{\mu} \sum_{\beta=1}^{\mu} x_\alpha x_\beta \int d\mathbf{r}_{\alpha\beta} \times [g_{\alpha\beta}^{(2)}(r_{\alpha\beta}) \ln g_{\alpha\beta}^{(2)}(r_{\alpha\beta}) - g_{\alpha\beta}^{(2)}(r_{\alpha\beta}) + 1], \quad (3)$$

results from the usual second-order truncation of the entropy expansion, and contains the familiar “ $g \ln g$ ” term. The next term is the so-called “ring” contribution,

$$s_{\text{ring}} \equiv \frac{1}{2(2\pi\rho)^3} \int d\mathbf{k} [\ln |\mathbf{I} + \tilde{\mathbf{H}}(\mathbf{k})| + \frac{1}{2} \text{Tr} \tilde{\mathbf{H}}^2(\mathbf{k}) - \text{Tr} \tilde{\mathbf{H}}(\mathbf{k})], \quad (4)$$

where \mathbf{I} is the $\mu \times \mu$ identity matrix and

$$\tilde{\mathbf{H}}(\mathbf{k})_{\alpha\beta} \equiv \rho_\alpha^{1/2} \tilde{h}_{\alpha\beta}(\mathbf{k}) \rho_\beta^{1/2}. \quad (5)$$

Hernando derived this term by extracting the leading contribution (“ring” diagram) from each individual order of the expansion hierarchy, and summing these contributions to infinite order.^{10,11} The function $\tilde{h}_{\alpha\beta}(\mathbf{k})$ is the Fourier transform of the total correlation function $h_{\alpha\beta}(r) \equiv g_{\alpha\beta}(r) - 1$. When pair correlation functions obtained from the hypernetted chain (HNC) approximation are inserted into this expansion, the sum of the first three terms is exactly equal to the well-known HNC entropy. Since the first three terms require

structural information only up to pair correlation functions, they form a potentially useful approximation to the total entropy. A subset of them is related to the HNC approximation.²¹ The fourth term contains the remaining terms of the expansion, which are more difficult to calculate. They are defined, for example, in Bush *et al.*²¹

Some applications of Eqs. (1)–(5) to aqueous systems initially used the canonical (i.e., “ $g \ln g$ ”) formulas.^{9,23} However, canonical formulas should not be used with correlation functions determined from simulations,^{24–26} due to their nonlocal nature⁸ (i.e., they would require unattainably large system sizes for reliable convergence). More recent publications use the correct formulas²⁷ but some²⁸ introduce additional approximations, such as orientation/translation decoupling. Application of these formulas to the problem of hydrophobicity was made by Silverstein.^{2,29}

B. Entropy expansion for pure water

To adequately describe the correlations among particles of a *molecular* fluid such as water, we need to go beyond the traditional centers-averaged n -body correlation functions. Water contains a high degree of angular ordering, and this information is lost in these orientationally averaged functions. Site–site correlation functions [e.g., $g_{\text{OO}}^{(2)}(r)$, $g_{\text{OH}}^{(2)}(r)$, $g_{\text{HH}}^{(2)}(r)$, $g_{\text{OOO}}^{(3)}(r, s, t)$, ...] have been used extensively in the literature. However, even these functions do not contain a full angular description. It is necessary to determine the distributions of the relative orientations between molecules in addition to their relative separation. The resulting full angular correlation function can be used to reconstruct all of the site–site functions; but the reverse is not possible. Different angular descriptions can be consistent with a single site–site correlation function.³⁰

The entropy of a molecular liquid can be expressed in terms of these and higher-order angular correlation functions.³¹ The exact expression for the general case of a mixture of molecular components is

$$\frac{S}{Nk} = s^{\text{ideal}} + s^{(2)} + s^{(3)} + \dots \quad (6)$$

for N total particles at temperature T and volume V , and k is the Boltzmann constant. The first term, which is analogous to the ideal gas term for atomic species in Eq. (2), but has additional angle-dependent variables, is derived in Appendix B:

$$s^{\text{ideal}} = 1 + \sum_{\alpha=1}^{\mu} x_\alpha \left[\frac{\nu_\alpha}{2} - \ln(\rho_\alpha \sigma_\alpha \Lambda_{t,\alpha}^d \Lambda_{r,\alpha}) \right], \quad (7)$$

where α is an index over μ particle species, x_α , ν_α , ρ_α , σ_α , $\Lambda_{t,\alpha}$, and $\Lambda_{r,\alpha}$ are the mole fraction, number of degrees of freedom, number density, symmetry number, translational thermal wavelength, and rotational wavelength, respectively, of species α , and d is the spatial dimensionality. Since the first term is an ideal-gas contribution, the two- and higher-body correlations account for the excess order in the liquid. The two-body term is of the form

$$s^{(2)} = -\frac{\rho}{2\Omega^2} \sum_{\alpha,\beta}^{\mu} x_{\alpha}x_{\beta} \int_V d\mathbf{r} d\omega_{\alpha} d\omega_{\beta} \times [g_{\alpha\beta}^{(2)}(\mathbf{r}, \omega_{\alpha}, \omega_{\beta}) \ln g_{\alpha\beta}^{(2)}(\mathbf{r}, \omega_{\alpha}, \omega_{\beta}) - g_{\alpha\beta}^{(2)}(\mathbf{r}, \omega_{\alpha}, \omega_{\beta}) + 1], \quad (8)$$

where $\rho = \sum_{\alpha=1}^{\mu} \rho_{\alpha}$ is the total number density, Ω is the total angular volume of integration, β is another molecular species index, \mathbf{r} is the center-of-mass interparticle separation, ω_{α} represents the Euler angles of species α , and $g_{\alpha\beta}^{(2)}(\mathbf{r}, \omega_{\alpha}, \omega_{\beta})$ is the two-body angular correlation function. Similarly, the three-body expression is

$$s^{(3)} = -\frac{\rho^2}{3!V\Omega^3} \sum_{\alpha,\beta,\gamma}^{\mu} x_{\alpha}x_{\beta}x_{\gamma} \int_V \int_V \int_V d\mathbf{r}_{\alpha} d\mathbf{r}_{\beta} d\mathbf{r}_{\gamma} d\omega_{\alpha} d\omega_{\beta} d\omega_{\gamma} \left[1 + g_{\alpha\beta\gamma}^{(3)}(\mathbf{r}_{\alpha\beta}, \mathbf{r}_{\beta\gamma}, \mathbf{r}_{\alpha\gamma}, \omega_{\alpha}, \omega_{\beta}, \omega_{\gamma}) \times \ln \left(\frac{g_{\alpha\beta\gamma}^{(3)}(\mathbf{r}_{\alpha\beta}, \mathbf{r}_{\beta\gamma}, \mathbf{r}_{\alpha\gamma}, \omega_{\alpha}, \omega_{\beta}, \omega_{\gamma})}{g_{\alpha\beta}^{(2)}(\mathbf{r}_{\alpha\beta}, \omega_{\alpha}, \omega_{\beta}) g_{\beta\gamma}^{(2)}(\mathbf{r}_{\beta\gamma}, \omega_{\beta}, \omega_{\gamma}) g_{\alpha\gamma}^{(2)}(\mathbf{r}_{\alpha\gamma}, \omega_{\alpha}, \omega_{\gamma})} \right) - g_{\alpha\beta\gamma}^{(3)}(\mathbf{r}_{\alpha\beta}, \mathbf{r}_{\beta\gamma}, \mathbf{r}_{\alpha\gamma}, \omega_{\alpha}, \omega_{\beta}, \omega_{\gamma}) + 3g_{\alpha\beta}^{(2)}(\mathbf{r}_{\alpha\beta}, \omega_{\alpha}, \omega_{\beta}) g_{\beta\gamma}^{(2)}(\mathbf{r}_{\beta\gamma}, \omega_{\beta}, \omega_{\gamma}) - 3g_{\alpha\beta}^{(2)}(\mathbf{r}_{\alpha\beta}, \omega_{\alpha}, \omega_{\beta}) \right], \quad (9)$$

where γ is a third index over molecular species. This term is identically zero in the well-known Kirkwood superposition approximation.

C. Entropy expansion for a solution

An expansion can also be written to describe the entropy of the solute-transfer process.²⁸ Using Eqs. (6)–(10) for a mixture of waters and solutes, the derivative with respect to solute number (at constant T , V , and N_W) is extrapolated to infinite dilution. After subtracting the ‘‘liberation entropy,’’ defined by Ben-Naim,³² we obtain the Ben-Naim standard state entropy:

$$\frac{\Delta S^*}{k} = \Delta s^{(2)} + \Delta s^{(3)} + \dots, \quad (10)$$

where the two-body term is

$$\Delta s^{(2)} = -\frac{\rho_W}{\Omega^2} \int_V d\mathbf{r}_{SW} d\omega_S d\omega_W [g_{SW}^{(2)} \ln g_{SW}^{(2)} - g_{SW}^{(2)} + 1] - \frac{\rho_W}{2\Omega^2} \int_V d\mathbf{r}_{WW} d\omega_{W_1} d\omega_{W_2} \times \left[\left(\frac{\partial g_{WW}^{(2)}}{\partial \rho_S} \right)_{T,V,N_W}^{\infty} \ln g_{WW}^{(2)} \right], \quad (11)$$

and the subscripts W and S denote water and solute, respectively. The partial derivative in the second term is taken at the limit of infinite dilution of the solute. In the case of an isotropic solute, the integration over the Euler angles of the solute is omitted, and we are left only with the angular degrees of freedom of the waters. From this expression, it is clear that only two terms contribute to the two-body transfer entropy: one for the orientational and translational restrictions of the waters about the solute, and a second term for the water–water reorganization induced by the presence of the solute. We neglect the contribution of the second term, because the water–water correlation function is essentially independent of solute concentration at the low concentrations

of interest to hydrophobicity studies, as shown in the following. This is the standard approximation in the literature,^{23,33,34} which may fail at the critical point of the pure solvent.³⁵ Further, the order *induced* in the surrounding fluid that contributes to this term at low and moderate temperatures would only accentuate the overestimate of the entropy penalty that we observe.

For the lowest temperature of this study (where the effects of system size are expected to be greatest), we have generated the full water–water correlation function in the presence of a solute of various dilutions. This was achieved by placing a single solute among 90, 120, 150, and 180 water molecules in the appropriate volume at a water density of 0.9. A two-dimensional ‘‘cut’’ of the water–water correlation function, shown³⁶ in Fig. 1, suggests that the water–water correlation function does not change with solute concentration. We checked many other cuts of the function, and all yielded the same conclusion. Hence, the partial derivative in the second term of Eq. (11) appears to be approximately

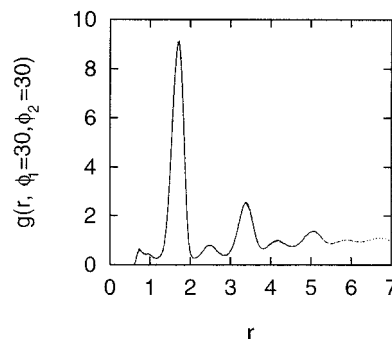


FIG. 1. A two-dimensional cut of the full, angular water–water correlation function for MB water in the presence of a single, fixed LJ solute. Four different simulations were run at a temperature just above the melting point, $T^*=0.16$, with a water density of 0.9, and contained 90 (solid), 120 (dashed), 150 (dotted), and 180 (dot-dashed) water molecules, respectively. The agreement shown here is typical of all of the other hypersurface cuts investigated. Definitions of the variables in the plot are given in Fig. 2.

zero throughout the range of the function. This agrees with the conclusions of Lazaridis.²⁷

The three-body term is comprised of two similar components:

$$\begin{aligned} \Delta s^{(3)} = & -\frac{\rho_w^2}{3!V\Omega^3} \int_V d\mathbf{r}_{w_1w_2} d\mathbf{r}_{sw_1} d\mathbf{r}_{sw_2} d\omega_s d\omega_{w_1} d\omega_{w_2} \\ & \times \left[g_{sww}^{(3)} \ln \left(\frac{g_{sww}^{(3)}}{g_{ww}^{(2)}g_{sw_1}^{(2)}g_{sw_2}^{(2)}} \right) - g_{sww}^{(3)} + 2g_{ww}^{(2)}g_{sw}^{(2)} + g_{sw_1}^{(2)}g_{sw_2}^{(2)} - 2g_{sw}^{(2)} - g_{ww}^{(2)} + 1 \right] \\ & - \frac{\rho_w^3}{3!V\Omega^3} \int_V d\mathbf{r}_{ww}^3 d\omega_w^3 \left\{ \left(\frac{\partial g_{www}^{(3)}}{\partial \rho_s} \right)_{T,V,N_w}^\infty \ln \left(\frac{g_{www}^{(3)}}{g_{w_1w_2}^{(2)}g_{w_2w_3}^{(2)}g_{w_1w_3}^{(2)}} \right) + 3 \left(\frac{\partial g_{ww}^{(2)}}{\partial \rho_s} \right)_{T,V,N_w}^\infty \left[2g_{ww}^{(2)} - 1 - \frac{g_{www}^{(3)}}{g_{ww}^{(2)}} \right] \right\}. \quad (12) \end{aligned}$$

This time, the first term describes the correlations of *pairs* of water about the solute, and the second term represents water reorganizations at the three-body level. As in Eq. (11), the superscripted infinity symbols indicate the partial derivatives are taken at the limit of infinite dilution of the solute. And as before, this term too is identically zero in the Kirkwood superposition approximation.

III. MODEL AND COMPUTATIONAL DETAILS

To study the multiple-body correlations of a molecular liquid such as water, it is necessary to track not only the intermolecular separations of particles, but their relative orientations as well. For two-body interactions of real water, this would amount to accumulating a correlation function that is a function of six variables in three dimensions (one center-of-mass separation variable, and five Euler angles). We have not had the computational resources for this task. Instead, we use a simplified, two-dimensional model of water, the MB model,^{3,37} which has recently been shown to have the anomalous thermodynamic trends of water and the hydrophobic effect.^{3,5} In $d=2$ there are only three variables (center-of-mass separation, and a single angular variable for each of the two molecules). This allows us to obtain well-converged correlation functions that can be integrated accurately. The convergence of these correlation functions can be improved further by accumulating the angular statistics as a Fourier series (see Sec. III B). Coefficients as a function of r are accumulated as ensemble averages in the simulation in a manner similar to that of Streett and Tildesley.³⁸

A. MB model

Here we review the properties of the MB model of water.^{3,5,29,37} In the model, water molecules are represented as disks in two dimensions, with three arms arranged as in the Mercedes-Benz logo. Molecules interact pairwise through a standard Lennard-Jones (LJ) term (two parameters), and an explicit hydrogen-bond interaction between arms of nearby molecules. There are no charges. The hydrogen bond has a defined optimal distance $r_{\text{HB}} = 1$ and optimal angle for each molecule's bonding arm ($\phi=0$) with the line that joins their centers (see Fig. 2). Deviations from this lowest-energy hydrogen bond are described by a Gaussian

expression with a single width parameter used to attenuate the interaction. The same width parameter is used for both separation and angle deviations.

In total, there are five parameters.³ The LJ well-depth parameter is one-tenth of ϵ_{HB} , the minimum value of the hydrogen bond function, and the LJ contact distance is 0.7 of r_{HB} . The width in the Gaussian $\sigma=0.085$ was chosen to be small enough that the direct h -bond contact is more favorable than a bifurcated h -bond. All energies and temperatures are reported in reduced units (i.e., normalized by $|\epsilon_{\text{HB}}|$). Likewise all distances are scaled by r_{HB} .

Monte Carlo simulations performed in this work were similar to those described earlier, but were carried out here in the canonical ensemble. Simulation boxes containing 60 molecules at two densities (the ice density, $\rho \approx 0.7698$ and $\rho = 0.9$) have been studied using standard periodic boundary conditions and the minimum image convention. Dilute non-polar solutions contained 60 water molecules and one LJ solute (with the same well depth and contact parameters as the water molecules) fixed at the center of the simulation box. Starting configurations at each temperature were selected at random.

The model has been shown by *NPT* Monte Carlo simulations³ to have the correct qualitative temperature dependence of a variety of pure water properties, including the existence of a density anomaly and related negative thermal expansion coefficient at low temperature, a minimum in the isothermal compressibility, a large anomalous heat capacity,

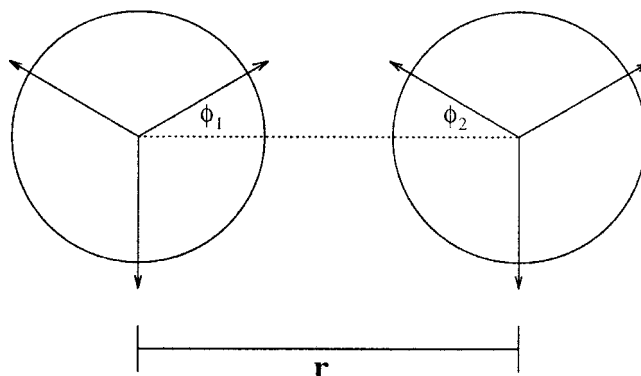


FIG. 2. Two representative MB "water" molecules, separated by a distance r . The angle, ϕ , that each molecule makes with the intermolecular axis is shown.

and spontaneous freezing to the $d=2$ model analog of ice, a low-density hexagonal crystal phase. Also, for the transfer of nonpolar solutes, the MB model predicts the correct temperature trends of the free energy, entropy, enthalpy, molar volume, and heat capacity.

B. Fourier expansion of the angular correlation function

In principle, the two-body angular correlation function for $d=2$, $g_{WW}^{(2)}(r, \phi_1, \phi_2)$, can be accumulated in a simple three-dimensional histogram. The histogram can then be accurately integrated numerically, provided that the bin size is sufficiently small. We have gathered such histograms using a bin size for r and ϕ of 0.02 and 2° , respectively. Reasonable convergence is achieved in 1×10^7 passes (1 pass = N Monte Carlo steps, and N is the number of molecules).

It is far more efficient, however, to expand the angular correlation function in a Fourier series. Streett and Tildesley have shown, in the analogous case for $d=3$ using spherical harmonics, that the coefficients of the series may be accumulated as a function of r as ensemble averages in the simulation.³⁸ Using this method for our system, excellent convergence of the correlation functions is attained in fewer than 5×10^5 passes, or about one to two orders of magnitude faster than the histogram method. Such an expansion would be crucial in the three-body case, since the three interparticle distances will already comprise a three-dimensional histogram.

In Appendix A, we show that the two-body water–water angular correlation function can be expanded as the Fourier series

$$g_{WW}^{(2)}(r, \phi_1, \phi_2) = \sum_{m_1} \sum_{m_2} c(m_1, m_2; r) e^{im_1\phi_1} e^{im_2\phi_2},$$

$$m_1, m_2 = 0, \pm 3, \pm 6, \dots, \quad (13)$$

where i is the imaginary unit, m_1 and m_2 are indices of the coefficients, and each of the coefficients themselves, $c(m_1, m_2; r)$ is a function of r . Only indices that are multiples of three contribute to the series due to the $2\pi/3$ symmetry of the MB water molecule.

The coefficients of this series may be determined as a product of angle- and r -dependent terms which are accumulated as ensemble averages in the simulation:

$$c(m_1, m_2; r) = g_{WW}^{(2)}(r) \langle e^{-im_1\phi_1} e^{-im_2\phi_2} \rangle_{r+\Delta r}. \quad (14)$$

The first term of the product is simply the centers-averaged pair correlation function. The second term is the ensemble average accumulated only over a shell about r of thickness Δr .

The expansion for the solute–water angular correlation function, $g_{SW}^{(2)}(r, \phi)$ is simpler, since the solute is isotropic. In this case, there is only one angular degree of freedom, ϕ , at the two-body level—that of the water molecule. Histograms of this function are slow to converge because there are statistically fewer water molecules around the solute than there are water–water pairs. Again, in this case, a corresponding Fourier series can greatly speed up convergence. Since this function is always symmetric about 0° , and there

is only one angle, a much simpler Fourier cosine series expansion (which derived in a manner similar to the water–water series in Appendix A) may be used in the solute–water case:

$$g_{SW}^{(2)}(r, \phi) = \sum_m c_m(r) \cos(m\phi), \quad m=0,3,6, \dots \quad (15)$$

In this case, the coefficients are

$$c_m(r) = \begin{cases} g_{SW}^{(2)}(r), & m=0 \\ 2g_{SW}^{(2)}(r) \langle \cos(m\phi) \rangle_{r+\Delta r}, & m=3,6,9, \dots, \end{cases} \quad (16)$$

where $g_{SW}^{(2)}(r)$ is the centers-averaged solute–water pair correlation function, and the ensemble average is defined as previously.

C. Calculation of the exact entropy

The exact excess (xs) entropy (i.e., the sum of all terms beyond the ideal gas term) has been computed for bulk water using thermodynamic integration by two independent pathways: (1) at constant density, from the given temperature to infinite temperature; and (2) at constant temperature, from zero density up to the desired density. In principle, only one pathway is necessary, but both were done to verify accuracy.

The entropy along the constant- ρ pathway is obtained from the thermodynamic relation

$$\frac{S^{xs}}{nk} = \int_{T_0}^{\infty} dT \left[\frac{C_v}{T} \right], \quad (17)$$

where T_0 is the temperature of interest, and C_v is the heat capacity, which is also a function of the temperature, T . This procedure yields the excess entropy for *any* temperature above T_0 at that fixed density. In practice, simulations can only be done at regular intervals up to some finite temperature. To overcome this obstacle, we performed more than a dozen simulations beyond a chosen temperature spanning six orders of magnitude in T , and fit the curve of C_v/T vs T to an expression of the form $C_v/T = (c/T)^\alpha$, where c is a constant. We found that this procedure fits all of the points, to within the error bars.

Along the constant- T pathway, the expression used by Hansen and Verlet³⁹ may be used to obtain the free energy:

$$\frac{F^{xs}}{nkT} = \int_0^{\rho_0} d\rho \left[\frac{\beta p / \rho - 1}{\rho} \right]. \quad (18)$$

Here, ρ_0 is the density of interest, p is the calculated pressure at each density ρ along the pathway, and $\beta p / \rho$ is the compressibility factor with $\beta \equiv 1/kT$. In practice, the compressibility factor is plotted as a function of density, and fit to a fifth or sixth order polynomial, which is forced to have a zeroth-order term of 1. In this manner, the above-mentioned integral can easily be performed analytically. The excess entropy can then be obtained from the relationship, $S^{xs}/nk = \beta U/n - \beta F^{xs}/n$, where U is the system energy at the given density and temperature.

For the case of the nonpolar solution, exact transfer entropies were computed using a fluctuation formula³ derived

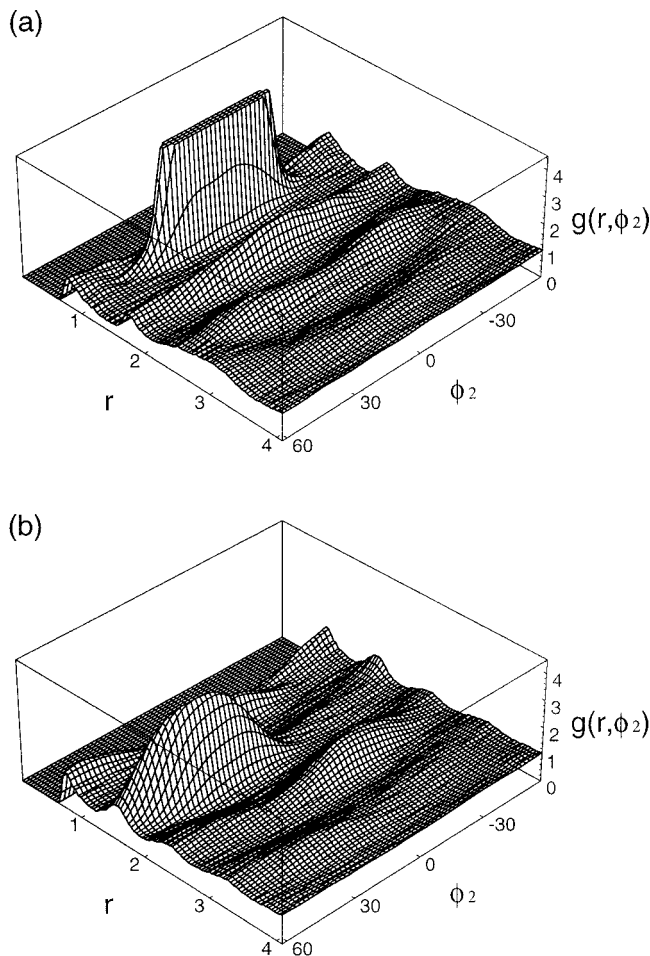


FIG. 3. Two surface cuts of the full water–water angular correlation curve $g_{WW}(r, \phi_1, \phi_2)$, where (a) $\phi_1 = 0^\circ$ and (b) $\phi_1 = 30^\circ$ for the phase point $T^* = 0.20$ and $\rho = 0.9$.

from the Widom method. This is the same approach used by Guillot and Guissani^{40,41} and us³ in earlier works.

IV. RESULTS: ENTROPY OF BULK MODEL WATER

The full two-body angular correlation function itself still comprises a four-dimensional plot, and hence cannot be visualized in its entirety. Thus we display “cuts” across the four-dimensional hypersurface. Two such cuts of the pure water distribution curve, $g_{WW}(r, \phi_1, \phi_2)$, are shown in Fig. 3. Figure 3(a) shows the distributions of a second water molecule, when the first one is fixed to point one of its arms

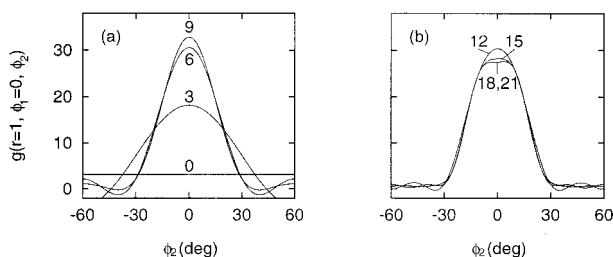


FIG. 4. Convergence of the Fourier series estimate for $g_{WW}^{(2)}(r=1.0, \phi_1=0^\circ, \phi_2)$: (a) inclusion of all coefficients up to $m_1, m_2 = 0$ through ± 9 as indicated by the labels; (b) coefficients up to $m_1, m_2 = \pm 12$ through ± 21 .

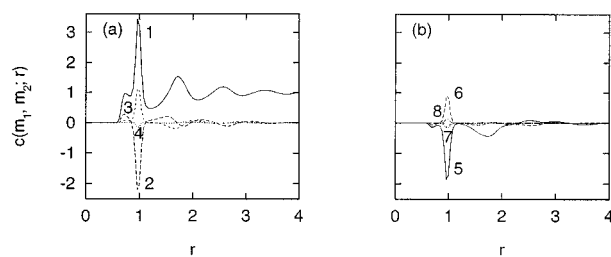


FIG. 5. The real part of the coefficients (m_1, m_2) of the Fourier series estimate of $g_{WW}^{(2)}(r, \phi_1, \phi_2)$, as a function of r : (a) 1: (0,0), 2: (0,3), 3: (0,6), 4: (0,9); (b) 5: (3,3), 6: (3,6), 7: (3,9), 8: (3,12).

directly at the other along the intermolecular axis (i.e., $\phi_1 = 0^\circ$). It is clear that the second water molecule has a preference to be one H-bonding unit away, pointing its arm toward the first molecule. The large peak has been cut away in order to better visualize the undulations marking the subsequent shells of water. In Fig. 3(b), the first molecule is fixed at an angle of 30° relative to the intermolecular axis of a given pair. This is the angle for an intervening molecule to form a bridge with a second-shell water. In an ideal arrangement, such a second-shell molecule would also have an angle of 30° relative to the fixed molecule, as shown by the peak maximum.

The Fourier series used to expand the correlation function converges after relatively few terms. In Fig. 4 we show the progression of a part of the correlation function, $g_{WW}^{(2)}(r=1.0, \phi_1=0^\circ, \phi_2)$, as higher order terms are added to the expansion. The zeroth-order term in Eq. (13) (i.e., $m_1, m_2 = 0$) has no angular dependence (see Fig. 4). The series rapidly converges.

The minimal contribution of higher order terms is more directly apparent in Fig. 5. Here, we show the r dependence of the real part of the coefficients themselves. The zeroth-order coefficient in Eq. (14) is the water–water pair correlation. Higher order coefficients make diminishing contributions (both in magnitude and radial extent).

The agreement is excellent between the direct histogram and series determinations of $g_{WW}^{(2)}(r, \phi_1, \phi_2)$. Because of the expense of accumulating histograms, correlation functions at all other temperatures were obtained only from the Fourier series method. Two separate two-dimensional cuts of the function are plotted in Fig. 6, comparing the two methods. Perhaps the most comprehensive measure of agreement is

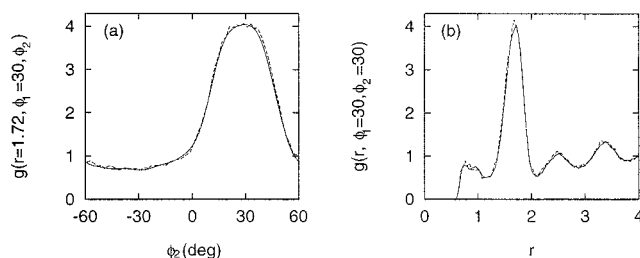


FIG. 6. Comparison of the Fourier series (solid line) and histogram (dashed line) estimates for two cuts of $g_{WW}^{(2)}(r, \phi_1, \phi_2)$: (a) as a function of ϕ_2 with r and ϕ_1 fixed at 1.72 and 30° , respectively; (b) as a function of r with ϕ_1 and ϕ_2 both fixed at 30° .

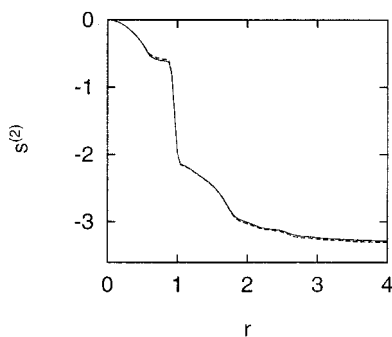


FIG. 7. Comparison of the Fourier series (solid line) and histogram (dashed line) estimates for the two-body entropy integral.

the two-body entropy expansion of Eq. (8). This integral is sensitive to all deviations from 1 in the correlation function. In Fig. 7 we show the integral as a function of the radius for both the histogram- and Fourier-derived functions. From Fig. 7, we see that (i) the integral converges at large r , and (ii) the histogram and series are in excellent agreement over the whole range of the correlation function.

The exact values for the (total) excess entropy, computed separately using a constant- ρ and constant- T thermodynamic integration pathway were in excellent agreement with each other. For the phase point, $T^*=0.20$, $\rho=0.7698$, the two pathways yielded values of -2.63 and -2.64 , respectively. Similarly, for the phase point $T^*=0.20$, $\rho=0.9$, the two values obtained were -2.78 and -2.82 .

In Fig. 8 we compare the two-body entropy with the exact excess values for two densities over a range of temperatures. It is clear that at high temperatures, the two-body term accounts for essentially all of the excess entropy (deviations are due probably to the finite size of the bins used in the numerical integration procedure). The two-body term is likely to account for the entropy also in the supercritical regime, and around the critical point, where the hydrogen bond network disappears.³⁵ But at lower temperatures, the two-body approximation leads to a significant *overestimate* of the true model entropy. (The lowest temperatures shown are just above the freezing point of MB water at these densities.) This discrepancy at low temperature must be made up by three- and higher-body terms. Note that the temperature dependence predicted by the two-body expression deviates significantly from the actual value.

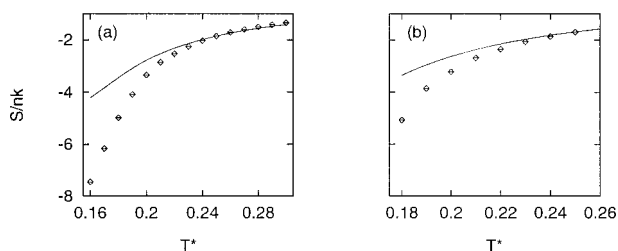


FIG. 8. Comparison of the two-body entropy term (diamonds) and the exact excess entropy (solid line, computed by thermodynamic integration) for the pure liquid. Trends are shown as a function of reduced temperature for two densities: (a) $\rho=0.9$, (b) $\rho=0.7698$.

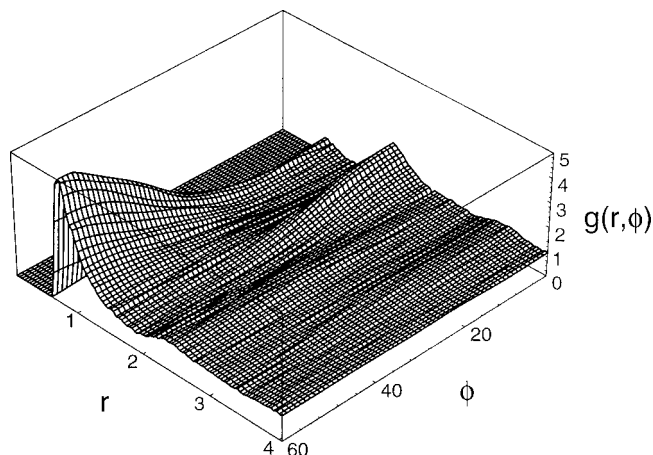


FIG. 9. The full solute-water angular correlation curve $g_{SW}(r, \phi)$ for the phase point $T^*=0.20$ and $\rho=0.9$. While the first shell water favors the orientation that straddles the nonpolar solute, the second shell waters point hydrogen bonds radially inward to coordinate with the first-shell waters.

V. RESULTS: ENTROPY FOR A NONPOLAR SOLUTE MOLECULE IN WATER

Figure 9 shows the full angular solute-water correlation curve, $g_{SW}(r, \phi)$, for a solute with the same size and LJ radius as the water, but no angle-dependent contribution to its interactions with other molecules. The first shell molecules prefer to straddle the solute (i.e., form a 60° angle), as we noted earlier.³ Similarly, the peak at zero angle and a distance of 1.8 shows that second-shell waters prefer to point their arms toward the first-shell molecules (and hence toward the solute).

The Fourier cosine series used to expand this function requires only three terms for convergence. In Fig. 10, we show the r dependence of the coefficients and the entropy

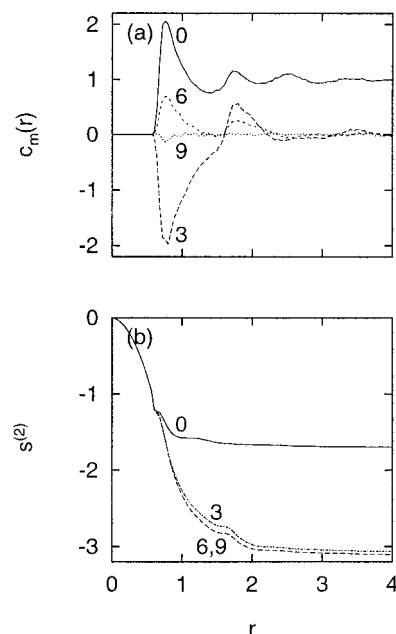


FIG. 10. Convergence of the Fourier series expansion for the solution: (a) the coefficients ($m=0-9$) as a function of r ; (b) the corresponding two-body entropy integral computed up to the order indicated by the labels.

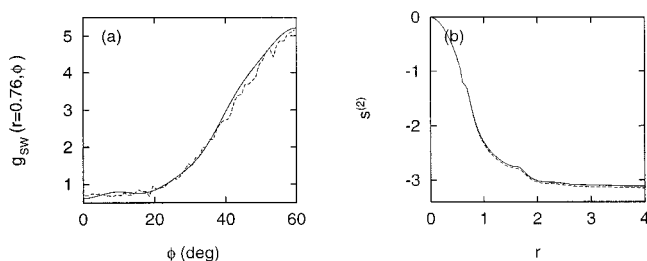


FIG. 11. Comparison of the Fourier series (solid line) and histogram (dashed line) estimates for $g_{SW}^{(2)}(r, \phi)$: (a) a cut of the correlation function as a function of ϕ with r fixed at 0.76; (b) the solute-water two-body entropy integral.

integral, which is computed from the correlation function truncated at each order. Since the zeroth-order term of Eq. (15) contains no angular information, the corresponding entropy integral corresponds simply to that which would be obtained by integrating only the centers-averaged pair correlation function. Figure 10(b) shows that the angular contributions to the entropy are significant, approximately 50% of the final value. In Fig. 11 we compare the histogram and series estimates of the correlation function directly, and by comparing the full entropy integral. Again, the agreement is excellent, with the Fourier series yielding better converged values.

Figure 12 shows the main result of this paper. We compare the two-body entropy term with the exact transfer entropy computed by the Widom fluctuation formula. Just as in pure water, two-body entropy is an excellent predictor of the exact entropy at high temperatures (but this is not true at low temperatures). Hence the difference between the two curves is the three- and higher-body entropy from the above-mentioned expansion. Thus at room temperature for this model, the entropy penalty of transfer is a result of both a highly ordering two-body term and disordering three- and higher-body terms.

Test of approximations to the entropy. To make contact with the literature, we have tested a number of cruder approximations to the solute entropy. If all of the angular information is omitted, the approximation to the two-body

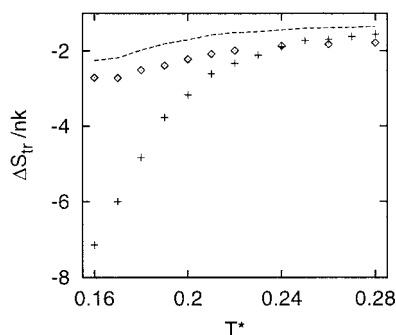


FIG. 12. Comparison of the two-body transfer entropy (+) and the corresponding exact value (\diamond), computed using the Widom particle insertion method) at $\rho=0.9$. The two-body entropy term which would arise from the proper integration of the *angle-averaged* pair correlation function, $g_{SW}(r)$, is also shown (dotted line). Trends are shown as a function of reduced temperature for the transfer of a simple inert Lennard-Jones solute with the same size as the solvent.

term is fortuitously in much better agreement with the exact transfer entropy. This is shown in Fig. 12, where we have plotted the two-body entropy integral computed from the angular-averaged pair correlation function, $g_{SW}(r)$, alongside the other two curves. This observation shows that the surprising success of theories which have omitted all angular information in predicting the thermodynamics of hydrophobic transfers⁴²⁻⁴⁶ arises from a cancellation of terms. It implies that, throughout the temperature range, the entropy loss attributed to the two-body angular terms almost exactly cancels the net disorder of the three- and higher-body terms. Further investigations of these multibody terms will be necessary to explain fully this remarkable observation.

In addition, we have tested two families of approximation for the pair term, which attempt to decouple translational and orientational components of the pair entropy of hydrophobic hydration. In view of the above-mentioned results, such an approximate decoupling would be of practical value at the temperatures of interest if equally accurate and simple approximations were found for the higher order terms, but unfortunately we know of none. In the approximation for the pair term, if it is assumed that (1) orientational correlations (relative to the solute) only occur in the first shell of waters, and (2) within a shell, water orientations have a homogeneous distribution, then the transfer entropy arising from solute-water correlations can be formally separated into the following components:^{23,34}

$$\Delta s^{(2)} \approx \Delta s_{\text{approx}}^{\text{tr}} + \Delta s_{\text{approx}}^{\text{or}} \quad (19)$$

and

$$\Delta s_{\text{approx}}^{\text{tr}} = -\rho_W \int_0^\infty d\mathbf{r} [g_{SW}(\mathbf{r}) \ln g_{SW}(\mathbf{r}) - g_{SW}(\mathbf{r}) + 1]. \quad (20)$$

The solute-water correlation function displayed here is merely a function of the center of mass positions of the two respective particles, and,

$$\Delta s_{\text{approx}}^{\text{or}} = -\sum_{\text{sh}_i} \frac{N_{W, \text{sh}_i}}{\Omega} \int d\omega_W [p_{\text{sh}_i}(\omega_W) \ln p_{\text{sh}_i}(\omega_W)], \quad (21)$$

where N_{W, sh_i} and $p_{\text{sh}_i}(\omega_W)$ are simply the coordination number and the orientational probability distributions, respectively, of the waters in the i th shell. The orientational probabilities are all normalized such that $\int d\omega_W p(\omega_W) = \Omega$.

Figure 13 shows that at all temperatures the ‘‘single shell’’ approximation²³ does not recover the calculated entropy, but a new, extended version of this approximation (in which the orientations of waters in three shells, rather than simply the first one are included) reproduces the full two-body angular entropy. More importantly, it is apparent in Fig. 13(b) that the orientational degrees of freedom of the water are far more temperature sensitive than the corresponding translational distributions.

VI. CONCLUSIONS

We have used a multiparticle expansion of the entropy to test our hypothesis that the entropy of solvation is the sum of two competing terms, explicitly by calculating two-body

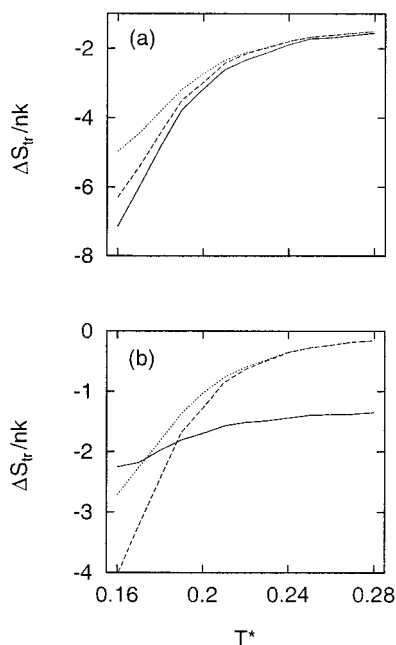


FIG. 13. Translational and orientational contributions to the entropy of transfer at $\rho=0.9$: (a) Comparison of $\Delta S_{tr}^{(2)}$ (solid line) with $\Delta S_{tr}^{tr} + \Delta S_{tr}^{or}$ as described in text (dotted line: orientations in the first shell included; dashed line: orientations in the first three shells included); (b) temperature dependence of $\Delta S_{tr}^{tr, approx}$ (solid line) and $\Delta S_{tr}^{or, approx}$, first shell (dotted line), first three shells (dashed line).

contribution for MB water, and a dilute hydrophobic solution. To do this, we calculate full angle-dependent pair correlation functions.

We have determined that the two-body term overestimates the exact entropy at low temperatures for both the pure liquid and for simple nonpolar solute transfers. At high temperatures, the agreement is remarkably good. Surprisingly, if all angular information is discarded, the corresponding two-body term is in reasonable agreement with the exact transfer entropy values at all temperatures studied.

We have tested candidate approximations²³ in which the translational and orientation components of the transfer entropy for simple solutes are assumed to be independent. When the orientations of only a single shell are included, their approximation underestimates the two-body entropy term at low temperature. This underestimate, combined with the overestimate of the full two-body term, with respect to the total excess entropy, leads to a fortuitous cancellation between two unrelated errors as suggested by Smith *et al.*²⁴ However, if the shell method is extended to multiple shells, then the agreement of the decoupling approximation with the full two-body term is reasonably good. Hence, under these conditions, the approximation appears to be a valid means to avoid calculating the full angular correlation function at the two-body level. In order to be of practical value, similar approximations of comparable accuracy will need to be made for the other terms that contribute to the full entropy.

ACKNOWLEDGMENTS

K.A.T.S. gratefully acknowledges support under a U.S. National Science Foundation Graduate Research Fellowship

and a UCSF Regent's Fellowship. Some of the calculations were performed in Australia using SydCom, the USyd/UTS Distributed Computing Facility funded by an ARC infrastructure grant, and supported by the Australian Research Council (ARC) (Grant No. A29530010). A.D.J.H. thanks the Welch Foundation for partial support of this research, and K.A.D. was supported by NIH Grant No. GM34993. In addition, K.A.T.S. would like to thank Karen Tang for many pertinent and lengthy discussions.

APPENDIX A: FOURIER SERIES EXPANSION OF THE ANGULAR CORRELATION FUNCTION

In this Appendix, we derive the formulas used in this work to express the water–water angular correlation function in terms of a Fourier series expansion. Additionally, some details that simplify the implementation are also discussed. We accumulate the coefficients of this series as ensemble averages in a manner similar to that used by Streett and Tildesley³⁸ for the analogous case using spherical harmonics in three dimensions.

1. Derivation

We begin by assuming that the water–water angular pair correlation function, $g_{WW}^{(2)}(r, \phi_1, \phi_2)$, (which is a smooth, continuous function of ϕ_1 and ϕ_2 at any given r) can be expressed as a product of Fourier series of the form:

$$g_{WW}^{(2)}(r, \phi_1, \phi_2) = \sum_{m_1, m_2 = -\infty}^{+\infty} c(m_1, m_2; r) e^{im_1 \phi_1} e^{im_2 \phi_2}. \quad (A1)$$

The complex coefficients of this series, $c(m_1, m_2; r)$ are a function of r , and have two integral indices, m_1 and m_2 . The angles, ϕ_1 and ϕ_2 have the range $[0, 2\pi)$ and correspond⁴⁷ to those in Fig. 2.

Now we have the task of obtaining expressions for the coefficients. We begin by multiplying both sides by the complex conjugate and integrating over all angular space to obtain

$$\begin{aligned} \int_0^{2\pi} \int_0^{2\pi} d\phi_1 d\phi_2 g_{WW}^{(2)}(r, \phi_1, \phi_2) e^{-im'_1 \phi_1} e^{-im'_2 \phi_2} \\ = \sum_{m_1, m_2 = -\infty}^{+\infty} c(m_1, m_2; r) \\ \times \int_0^{2\pi} d\phi_1 e^{-im'_1 \phi_1} e^{im_1 \phi_1} \int_0^{2\pi} d\phi_2 e^{-im'_2 \phi_2} e^{im_2 \phi_2}, \end{aligned} \quad (A2)$$

where m'_1 and m'_2 are arbitrary coefficients. Exchanging the right- and left-hand sides of the equation, and applying the orthogonality property

$$\int_0^{2\pi} d\phi (e^{im\phi})^* e^{in\phi} = 2\pi \delta_{m,n} \quad (A3)$$

we are left with

$$c(m_1, m_2; r) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} d\phi_1 d\phi_2 g_{WW}^{(2)}(r, \phi_1, \phi_2) e^{-im_1\phi_1} e^{-im_2\phi_2} \quad (\text{A4})$$

after the sums over m_1 and m_2 have been performed, and the dummy prime variables have been relabeled. But Eq. (A4) is not very practical, since the determination of the coefficients requires the computation of the angular correlation function itself—the very quantity we are trying to approximate!

So we use the definition, following Streett and Tildesley, for the ensemble average of some quantity, $X(r, \phi_1, \phi_2)$, in a given shell of infinitesimal thickness about r :

$$\langle X \rangle_{r+\Delta r} = \frac{\int_0^{2\pi} \int_0^{2\pi} d\phi_1 d\phi_2 X(r, \phi_1, \phi_2) f^{(2)}(r, \phi_1, \phi_2)}{\int_0^{2\pi} \int_0^{2\pi} d\phi_1 d\phi_2 f^{(2)}(r, \phi_1, \phi_2)}. \quad (\text{A5})$$

Here

$$f^{(2)}(r, \phi_1, \phi_2) \equiv \frac{\rho^2 \sigma^2}{\Omega^2} g_{WW}^{(2)}(r, \phi_1, \phi_2) \quad (\text{A6})$$

is the two-particle distribution function, and ρ , σ , and $\Omega \equiv \int_0^{2\pi} d\phi = 2\pi$ refer to the water density, symmetry number (3 for MB water), and total volume of integration, respectively. In particular, since Eq. (A4) gives

$$c(0,0;r) = g_{WW}^{(2)}(r) = \frac{1}{\Omega^2} \int_0^{2\pi} \int_0^{2\pi} d\phi_1 d\phi_2 g_{WW}^{(2)}(r, \phi_1, \phi_2), \quad (\text{A7})$$

it follows from Eqs. (A5) and (A6) that

$$\int_0^{2\pi} \int_0^{2\pi} d\phi_1 d\phi_2 X(r, \phi_1, \phi_2) g_{WW}^{(2)}(r, \phi_1, \phi_2) = 4\pi^2 g_{WW}^{(2)}(r) \langle X \rangle_{r+\Delta r}. \quad (\text{A8})$$

Finally, using Eq. (A4) to interpret the product of Fourier terms as the quantity X , we obtain a useful equation for the coefficients:

$$c(m_1, m_2; r) = g_{WW}^{(2)}(r) \langle e^{-im_1\phi_1} e^{-im_2\phi_2} \rangle_{r+\Delta r}. \quad (\text{A9})$$

2. Simplifications for implementation

The Fourier series expansion derived in the previous section may be simplified by exploiting additional properties of the angular correlation function. The first of these simplifications uses the $2\pi/3$ symmetry of the MB molecule to obtain a condition on the indices m_1 and m_2 which eliminates $2/3$ of the needed coefficients. The second makes use of the fact that the correlation function is always real, and results in a further 50% reduction in the number of coefficients that must be accumulated.

Due to the symmetry of the MB water molecule, the following condition holds:

$$g_{WW}^{(2)}(r, \phi_1, \phi_2) = g_{WW}^{(2)}\left(r, \phi_1 + \frac{2\pi}{3}, \phi_2\right). \quad (\text{A10})$$

This implies that, after expanding the right-hand side in the Fourier series of Eq. (A1),

$$g_{WW}^{(2)}(r, \phi_1, \phi_2) = \sum_{m_1, m_2 = -\infty}^{+\infty} c(m_1, m_2; r) e^{im_1\phi_1} e^{im_2\phi_2} e^{im_1 2\pi/3}. \quad (\text{A11})$$

Since this expansion must be equivalent to the original expression of Eq. (A1), the coefficients must be zero whenever the last term is not 1 (i.e., whenever m_1 is not a multiple of 3). Hence the coefficients are only nonzero for $m_1 = 0, \pm 3, \pm 6, \dots$. Similarly, the condition $g_{WW}^{(2)}(r, \phi_1, \phi_2) = g_{WW}^{(2)}(r, \phi_1, \phi_2 + 2\pi/3)$ requires that $m_2 = 0, \pm 3, \pm 6, \dots$ as well.

The requirement that the correlation function be real can be expressed as

$$g_{WW}^{(2)}(r, \phi_1, \phi_2) = g_{WW}^{(2)*}(r, \phi_1, \phi_2), \quad (\text{A12})$$

where the asterisk denotes the complex conjugate. Substituting the Fourier expansion for both sides, we have

$$\sum_{m_1, m_2 = -\infty}^{+\infty} c(m_1, m_2; r) e^{im_1\phi_1} e^{im_2\phi_2} = \sum_{m_1, m_2 = -\infty}^{+\infty} c^*(m_1, m_2; r) e^{-im_1\phi_1} e^{-im_2\phi_2}. \quad (\text{A13})$$

After multiplying both sides by the complex conjugate of an arbitrary rank, integrating over all angular space, using the orthogonality property of Eq. (A3), and performing the sums over m_1 and m_2 , we obtain

$$c(m_1, m_2; r) = c^*(-m_1, -m_2; r). \quad (\text{A14})$$

Equation (A14) allows us to reexpress the series expansion the form:

$$g_{WW}^{(2)}(r, \phi_1, \phi_2) = \sum_{m_1, m_2 = 0}^{+\infty} [c(m_1, m_2; r) e^{im_1\phi_1} e^{im_2\phi_2} + c^*(m_1, m_2; r) e^{-im_1\phi_1} e^{-im_2\phi_2} + c(m_1, -m_2; r) e^{im_1\phi_1} e^{-im_2\phi_2} + c^*(m_1, -m_2; r) e^{-im_1\phi_1} e^{im_2\phi_2}]. \quad (\text{A15})$$

The advantage of this form, which we use in practice, is that only the positive–positive and positive–negative coefficients need to be determined from simulations. The remaining coefficients can then simply be obtained by taking the appropriate complex conjugates, hence reducing the direct computational requirements by half.

APPENDIX B: DERIVATION OF THE IDEAL-GAS ENTROPY TERM

We derive here a simple, but rigorous, expression for the first term in the entropy expansion. This is done for the three-dimensional pure molecular fluid, but can be generalized straightforwardly to a mixture, in d dimensions. The nomenclature used is based upon that of Gray and Gubbins.³⁰

For a fluid of N rigid molecules in a volume V at temperature T , the classical expression for the canonical partition function is

$$Q_N = \frac{Z_N}{h^{\nu N} N! \sigma^N}, \quad (B1)$$

$$Z_N = \int_v d\mathbf{r}^N d\mathbf{p}^N d\omega^N d\mathbf{J}^N \exp(-\beta H_N), \quad (B2)$$

where h is Planck's constant, ν is the number of degrees of freedom (five for linear molecules, six for nonlinear molecules), σ is the symmetry number, $\beta = 1/kT$ is the inverse temperature with k denoting the Boltzmann constant, H_N is the Hamiltonian, \mathbf{r}^N represents the Cartesian coordinates of the N molecules, and \mathbf{p}^N , ω^N , and \mathbf{J}^N are the linear momenta, Euler angles, and angular momenta of those N molecules, respectively.

The probability, $P_N^{(n)}$, of observing a given *specific* subset of n of the total N particles in a given arrangement of positions, orientations, and momenta is

$$P_N^{(n)} = \frac{1}{Z} \int_v d\mathbf{r}^{N-n} d\mathbf{p}^{N-n} d\omega^{N-n} d\mathbf{J}^{N-n} \exp(-\beta H_N). \quad (B3)$$

It is often more useful, however, to know the probability of observing any n *indistinguishable* particles with the given arrangement in phase space, $(\mathbf{r}^n, \mathbf{p}^n, \omega^n, \mathbf{J}^n)$. This is the n -particle distribution function:

$$f_N^{(n)} = \frac{N!}{(N-n)!} \sigma^n P_N^{(n)}. \quad (B4)$$

The one-particle distribution function is thus, from Eqs. (B3) and (B4):

$$f_N^{(1)} = \frac{N\sigma}{Z} \int_v d\mathbf{r}^{N-1} d\mathbf{p}^{N-1} d\omega^{N-1} d\mathbf{J}^{N-1} \exp(-\beta H_N). \quad (B5)$$

Noting that

$$H_N = H_{N-1} + \frac{\mathbf{p}_1^2}{2m} + \sum_{\alpha=X,Y,Z} \frac{\mathbf{J}_{1\alpha}^2}{2I_\alpha} + \psi, \quad (B6)$$

where m is the molecular mass, I_α is the moment of inertia in the $\alpha = X, Y, Z$ dimension, and ψ is the potential of interaction between one molecule and the remaining $N-1$ molecules, we can obtain an explicit formula for the one particle distribution function. After substituting Eq. (B6) into Eq. (B5), separating the integrals, factoring the momentum integrals into products, and dividing out like terms, we obtain an expression for the one-particle distribution function that is only dependent on one particle's momentum:

$$f_N^{(1)}(\mathbf{p}_1, \mathbf{J}_1) = \frac{\rho\sigma}{\Omega} \left(\frac{\beta}{2\pi m}\right)^{3/2} \left[\prod_{\alpha=X,Y,Z} \left(\frac{\beta}{2\pi I_\alpha}\right)^{1/2} \right] \times \exp\left(-\frac{\beta\mathbf{p}_1^2}{2m}\right) \exp\left(-\beta \sum_{\alpha=X,Y,Z} \frac{\mathbf{J}_{1\alpha}^2}{2I_\alpha}\right), \quad (B7)$$

where $\rho = N/V$ is the number density and Ω is the total angular volume of integration.

As Lazaridis and Paulaitis have shown,²³ the canonical expression for the one-particle term in the entropy expansion is

$$S_N^{(1)} = -\frac{kN\Omega}{\rho\sigma} \int_v d\mathbf{p} d\mathbf{J} f_N^{(1)}(\mathbf{p}, \mathbf{J}) \ln[h^\nu f_N^{(1)}(\mathbf{p}, \mathbf{J})]. \quad (B8)$$

Substituting the expression of Eq. (B7) into this equation, and assigning the constants $a = \beta/2m$, $b_\alpha = \beta/2I_\alpha$, $c_1 = -Nk(a/\pi)^{3/2} \prod_{\alpha=X,Y,Z} (b_\alpha/\pi)^{1/2}$, and $c_2 = -c_1 h^\nu \rho \sigma / (Nk\Omega)$ for mathematical simplicity, we have:

$$S_N^{(1)} = c_1 \int_v d\mathbf{p} d\mathbf{J} \left\{ \exp(-a\mathbf{p}^2) \exp\left(-\sum_{\alpha=X,Y,Z} b_\alpha \mathbf{J}_\alpha^2\right) \times \left[\ln c_2 - a\mathbf{p}^2 - \sum_{\alpha=X,Y,Z} b_\alpha \mathbf{J}_\alpha^2 \right] \right\}. \quad (B9)$$

If the linear and angular momenta are factorizable,³⁰ then we obtain, after separating the integrals, and changing to spherical coordinates:

$$S_N^{(1)} = 4\pi c_1 \left\{ \ln c_2 \int_0^{+\infty} dp [p^2 \exp(-ap^2)] \int_{-\infty}^{+\infty} d\mathbf{J} \exp\left(-\sum_{\alpha=X,Y,Z} b_\alpha \mathbf{J}_\alpha^2\right) - a \int_0^{+\infty} dp [p^4 \exp(-ap^2)] \int_{-\infty}^{+\infty} d\mathbf{J} \times \exp\left(-\sum_{\alpha=X,Y,Z} b_\alpha \mathbf{J}_\alpha^2\right) - \int_0^{+\infty} dp [p^2 \exp(-ap^2)] \sum_{\alpha=X,Y,Z} b_\alpha \int_{-\infty}^{+\infty} d\mathbf{J} \left[\mathbf{J}_\alpha^2 \exp\left(-\sum_{\alpha'=X,Y,Z} b_{\alpha'} \mathbf{J}_{\alpha'}^2\right) \right] \right\}, \quad (B10)$$

which, after evaluating the linear momentum integrals from standard tables, and regrouping, we are left with

$$S_N^{(1)} = \left(\frac{\pi}{a}\right)^{3/2} c_1 \left\{ (\ln c_2 - 3/2) \times \int_{-\infty}^{+\infty} d\mathbf{J} \exp\left(-\sum_{\alpha=X,Y,Z} b_\alpha \mathbf{J}_\alpha^2\right) - \sum_{\alpha=X,Y,Z} b_\alpha \times \int_{-\infty}^{+\infty} d\mathbf{J} \left[\mathbf{J}_\alpha^2 \exp\left(-\sum_{\alpha'=X,Y,Z} b_{\alpha'} \mathbf{J}_{\alpha'}^2\right) \right] \right\}. \quad (B11)$$

The two integrals in Eq. (B11) are

$$\int_{-\infty}^{+\infty} d\mathbf{J} \exp\left(-\sum_{\alpha=X,Y,Z} b_\alpha \mathbf{J}_\alpha^2\right) = \prod_{\alpha=X,Y,Z} \left(\frac{\pi}{b_\alpha}\right)^{1/2} \quad (B12)$$

and

$$\sum_{\alpha=X,Y,Z} b_\alpha \int_{-\infty}^{+\infty} d\mathbf{J} \left[\mathbf{J}_\alpha^2 \exp\left(-\sum_{\alpha'=X,Y,Z} b_{\alpha'} \mathbf{J}_{\alpha'}^2\right) \right] = \frac{3}{2} \left[\prod_{\alpha=X,Y,Z} \left(\frac{\pi}{b_\alpha}\right)^{1/2} \right], \quad (B13)$$

respectively. After substituting these relationships back into

Eq. (B11), replacing the constants c_1 , c_2 , a , and b_α as defined earlier, and collecting and canceling terms, we obtain

$$S_N^{(1)} = Nk \left[\frac{\nu}{2} - \ln(\rho\sigma\Lambda_t^3\Lambda_r) \right], \quad (\text{B14})$$

where the translational and rotational thermal wavelengths are

$$\Lambda_t = \left(\frac{\beta h^2}{2\pi m} \right)^{1/2} \quad (\text{B15})$$

and

$$\Lambda_r = \begin{cases} \pi^{-1/2} \left(\frac{\beta h^2}{8\pi^2 I_X} \right)^{1/2} \left(\frac{\beta h^2}{8\pi^2 I_Y} \right)^{1/2} \left(\frac{\beta h^2}{8\pi^2 I_Z} \right)^{1/2} \\ \text{(nonlinear)} \\ \frac{\beta h^2}{8\pi^2 I} \quad \text{(linear)} \end{cases}, \quad (\text{B16})$$

respectively. Baranyai and Evans⁸ have shown that the configurational contribution to the entropy of the ideal gas in the *canonical* ensemble is exactly 1. Hence, this value must be added to $S_N^{(1)}$ of Eq. (B14), to at last obtain the ideal-gas entropy:

$$s^{\text{ideal}} = 1 + Nk \left[\frac{\nu}{2} - \ln(\rho\sigma\Lambda_t^3\Lambda_r) \right]. \quad (\text{B17})$$

¹H. S. Frank and M. W. Evans, *J. Chem. Phys.* **13**, 507 (1945).

²A. D. J. Haymet, K. A. T. Silverstein, and K. A. Dill, in *International Symposium on Molecular Thermodynamics and Molecular Simulation*, Hosen University, Tokyo, Japan, 1997, pp. 143–152.

³K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill, *J. Am. Chem. Soc.* **120**, 3166 (1998).

⁴K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill, *J. Chem. Phys.* **111**, 8000 (1999).

⁵K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill, *J. Am. Chem. Soc.* **122**, 8037 (2000).

⁶A. D. J. Haymet, K. A. T. Silverstein, and K. A. Dill, *Faraday Discuss.* **103**, 117 (1996).

⁷Some of these results were presented in preliminary form at a conference (Ref. 2).

⁸A. Baranyai and D. J. Evans, *Phys. Rev. A* **40**, 3817 (1989).

⁹D. C. Wallace, *J. Chem. Phys.* **87**, 2282 (1987).

¹⁰J. A. Hernando, *Mol. Phys.* **69**, 319 (1990).

¹¹J. A. Hernando, *Mol. Phys.* **69**, 327 (1990).

¹²B. B. Laird and A. D. J. Haymet, *Phys. Rev. A* **45**, 5680 (1992).

¹³B. B. Laird and A. D. J. Haymet, *J. Chem. Phys.* **97**, 2153 (1992).

¹⁴L. Lee, *J. Chem. Phys.* **97**, 8606 (1992).

¹⁵H. S. Green, *The Molecular Theory of Fluids* (North-Holland, Amsterdam, 1952).

¹⁶R. E. Nettleton and M. S. Green, *J. Chem. Phys.* **29**, 1365 (1958).

¹⁷J. G. Kirkwood, *J. Chem. Phys.* **10**, 394 (1942).

¹⁸H. J. Raveché, *J. Chem. Phys.* **55**, 2242 (1971).

¹⁹J. A. Hernando and L. Blum, *Phys. Rev. E* **62**, 6577 (2000).

²⁰B. B. Laird, J. Wang, and A. D. J. Haymet, *Phys. Rev. E* **47**, 2491 (1993).

²¹M. R. Bush, M. J. Booth, A. D. J. Haymet, and A. G. Schlijper, *Mol. Phys.* **95**, 601 (1998).

²²B. B. Laird and A. D. J. Haymet, *J. Chem. Phys.* **100**, 3775 (1994).

²³T. Lazaridis and M. E. Paulaitis, *J. Phys. Chem.* **96**, 3847 (1992).

²⁴D. E. Smith, B. B. Laird, and A. D. J. Haymet, *J. Phys. Chem.* **97**, 5788 (1993).

²⁵J. W. Arthur and A. D. J. Haymet, *J. Chem. Phys.* **109**, 7991 (1998).

²⁶J. W. Arthur and A. D. J. Haymet, *J. Chem. Phys.* **110**, 5873 (1999).

²⁷T. Lazaridis, *J. Phys. Chem. B* **104**, 4964 (2000).

²⁸H. S. Ashbaugh and M. E. Paulaitis, *J. Phys. Chem.* **100**, 1900 (1996).

²⁹K. A. T. Silverstein, Ph.D. thesis, University of California, San Francisco, 1997.

³⁰C. G. Gray and K. E. Gubbins, *Theory of Molecular Fluids* (Clarendon, Oxford, 1984).

³¹An expansion which includes the orientational degrees of freedom needed for a molecular fluid was presented by Lazaridis and Paulaitis (Ref. 23) in the *canonical* ensemble. As described earlier, however, the canonical formulas are nonlocal. To obtain the correct, ensemble-invariant formulas, we have applied the procedure of Baranyai and Evans (Ref. 8).

³²A. Ben-Naim, *Hydrophobic Interactions* (Plenum, New York, 1980).

³³T. Lazaridis and M. E. Paulaitis, *J. Phys. Chem.* **97**, 5789 (1993).

³⁴T. Lazaridis and M. E. Paulaitis, *J. Phys. Chem.* **98**, 635 (1994).

³⁵This observation was brought to our attention by an anonymous referee of our manuscript.

³⁶The corresponding correlation function for 60 water molecules has its two primary peaks dampened (not shown to avoid confusion) at this lowest temperature, due to interference through the box boundaries. Though our study here has been carried out with 60 water molecules, we emphasize that no corresponding box size effects have been observed at intermediate temperatures, and that increasing the system size would merely accentuate our two-body entropy overestimate by an additional 3% at the lowest temperature. Thus we have not redone the remainder of the calculations with a larger system size.

³⁷A. Ben-Naim, *J. Chem. Phys.* **54**, 3682 (1971).

³⁸W. B. Streett and D. J. Tildesley, *Proc. R. Soc. London, Ser. A* **348**, 485 (1976).

³⁹J.-P. Hansen and L. Verlet, *Phys. Rev.* **184**, 151 (1969).

⁴⁰B. Guillot, Y. Guissani, and S. Bratos, *J. Chem. Phys.* **95**, 3643 (1991).

⁴¹B. Guillot and Y. Guissani, *J. Chem. Phys.* **99**, 8075 (1993).

⁴²L. R. Pratt and D. Chandler, *J. Chem. Phys.* **67**, 3683 (1977).

⁴³T. Head-Gordon, *Chem. Phys. Lett.* **227**, 215 (1994).

⁴⁴T. Head-Gordon, *J. Am. Chem. Soc.* **117**, 501 (1995).

⁴⁵G. Hummer, S. Garde, A. E. Garcia, A. Pohorille, and L. R. Pratt, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8951 (1996).

⁴⁶B. J. Berne, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8800 (1996).

⁴⁷Actually, when this work was performed, ϕ_1 and ϕ_2 were both taken to be the angles relative to a fixed unit vector from molecule 1 to molecule 2 (i.e., ϕ_1 was the same, but ϕ_2' was related to ϕ_2 in Fig. 2 by $\phi_2' = \pi - \phi_2$). Then, the resulting curves were properly transformed to the angles of Fig. 2 which are more intuitive given the model potential. In retrospect, the original, intuitive angles could have been used in the expansion, yielding the same results with less trouble. However, we make this note because the coefficients themselves, plotted in Fig. 5, would have been different.