JCTC Journal of Chemical Theory and Computation

Thermodynamic Properties of Liquid Water: An Application of a Nonparametric Approach to Computing the Entropy of a Neat Fluid

Lingle Wang, Robert Abel, Richard A. Friesner, and B. J. Berne*

Department of Chemistry, Columbia University, New York, New York 10027

Received February 13, 2009

Abstract: Because of its fundamental importance to molecular biology, great interest has continued to persist in developing novel techniques to efficiently characterize the thermodynamic and structural features of liquid water. A particularly fruitful approach, first applied to liquid water by Lazaridis and Karplus, is to use molecular dynamics or Monte Carlo simulations to collect the required statistics to integrate the inhomogeneous solvation theory equations for the solvation enthalpy and entropy. We here suggest several technical improvements to this approach, which may facilitate faster convergence and greater accuracy. In particular, we devise a nonparametric *k*th nearest-neighbors (NN)-based approach to estimate the water−water correlation entropy, and we suggest an alternative factorization of the water−water correlation function that appears to more robustly describe the correlation entropy of the neat fluid. It appears that the NN method offers several advantages over the more common histogram-based approaches, including much faster convergence for a given amount of simulation data; an intuitive error bound that may be readily formulated without resorting to block averaging or bootstrapping; and the absence of empirically tuned parameters, which may bias the results in an uncontrolled fashion.

1. Introduction

Water is unique among liquids for its biological significance. It plays an active role in the formation of the structures of proteins, lipid bilayers, and nucleic acids in vivo, both through direct hydrogen-bonding interactions with these biomolecules, and also through indirect interactions, where the unique hydrogen-bonded structure of liquid water is known to drive hydrophobic assembly.[1] It has been suggested that a robust characterization of the thermodynamic properties and structure of water solvating the active site of a protein is essential to rationalize the various binding affinities of small molecules that will displace that solvent to bind to the protein active site.[2,3]

As such, great interest has continued to persist in developing novel techniques to efficiently characterize the thermodynamic and structural features of liquid water in different environments. A particularly fruitful approach, first applied to liquid water by Lazaridis and Karplus,[4−6] used molecular dynamics or Monte Carlo simulations to collect the required statistics to integrate the inhomogeneous solvation theory (IST) equations for the solvation enthalpy and entropy. In this theory, the solvation enthalpy is determined from an analysis of the change in the solute−solvent and solvent−solvent interaction energy terms, and the solvation entropy is computed from an expansion of the entropy in terms of increasingly higher order solute−solvent correlation functions.[4] This approach has been used to characterize the thermodynamics and structure of neat water,[6] hydration of small hydrophobes,[4] and the hydration of the active sites of proteins.[7,8] Recently, it has also been extended to allow for the rapid computation of the relative binding affinities of a set of congeneric ligands with a given protein, via a semiempirical displaced-solvent functional.[2]

Because of the increasing interest in applying this technique to water[9−12] in various environments, we have chosen to reexamine the factorization and correlation function integration scheme originally suggested by Lazaridis and Karplus[6] for bulk water and later adopted by

* Corresponding author e-mail: bb8@columbia.edu.

10.1021/ct900078k CCC: $40.75
Published on Web 05/18/2009

others.[13] We have found that several technical improvements in this scheme are possible, which may facilitate faster convergence and greater accuracy than the more typical expressions. In this Article, we (1) devise a nonparametric *k*th nearest-neighbors (NN)[14]-based approach to estimate the water−water correlation entropy, in lieu of the more common histogram-based approaches, and (2) suggest an alternative factorization for the water−water correlation function that appears to more robustly describe the water−water correlation entropy of the neat fluid. To our knowledge, this is the first application of the NN method to compute the entropy of a neat fluid. It appears that the NN method offers several advantages over the more common histogram-based approaches, including (1) much faster convergence for a given amount of simulation data, especially when the correlation function is highly structured; (2) an intuitive error bound may be readily formulated without resorting to block averaging or bootstrapping techniques, which may be problematic to apply to estimators of the entropy; and (3) the absence of empirically tuned parameters, such as the histogram bin width, which may bias the results in an unpredictable fashion. Our alternative factorization of the water−water correlation function explicitly includes correlations between the water−dipole-vector−intermolecular-axis angle with the angle of rotation of the water molecule about its dipole vector. This contribution, although neglected by others,[6] has been found in our work to increase the agreement of results obtained by the entropy expansion with those obtained by less approximate methods, such as free energy perturbation theory. We also extensively compare the solvation entropies obtained from the truncated entropy expansion to those obtained from a finite difference analysis of free energy perturbation theory results. This comparison allows us to characterize the errors in both precision and accuracy associated with the NN method of integrating the entropy expansion presented here.

Our primary interest in developing this technique was to later adapt the method to study the solvation of solutes; thus, we were interested in determining realistic estimates of the convergence of the technique when the isotropic symmetry of the fluid was not present. As such, when extracting the solvent configurations to compute the pair correlation function (PCF), we chose to use only the configurations of a distinguished solvent molecule with the rest of the system, instead of collecting statistics from all pairs of solvent molecules. Such a protocol allows for an interrogation of the relative convergence properties of the various methods that might be obscured by the additional statistics offered by taking advantage of the symmetry of the system.

2. Methods

2.1. The Entropy Expression of a Neat Fluid. First derived by Green,[15] and later by Raveché[16] and Wallace,[17] the entropy of a fluid can be expressed as a sum of integrals over multiparticle correlation functions. For a molecular fluid,[5] the expression is

$$\begin{aligned} s = s^{\mathrm{id}} + s_{\mathrm{e}} = s^{\mathrm{id}} - \frac{1}{2!}k\frac{\rho}{\Omega^2} \\ \int [g^{(2)} \ln(g^{(2)}) - g^{(2)} + 1]\, \mathrm{d}\boldsymbol{r}\, \mathrm{d}\omega^2 \\ -\frac{1}{3!}k\frac{\rho}{\Omega^2}\int [g^{(3)} \ln(\delta g^{(3)}) - g^{(3)} + \\ 3g^{(2)}g^{(2)} - 3g^{(2)} + 1]\, \mathrm{d}\boldsymbol{r}_1\, \mathrm{d}\boldsymbol{r}_2\, \mathrm{d}\omega^3 - \ldots \end{aligned} \tag{1}$$

where, s^{id} is the entropy of an ideal gas with the same density and temperature as the fluid, s_{e} is the excess entropy of the fluid over that of the ideal gas, k is Boltzmann's constant, ρ is the number density, ω denotes the orientational variables of one molecule, Ω is the total volume of the orientational space (for a nonlinear molecule like water, Ω is $8\pi^2$), $g^{(2)}$ is the pair correlation function, $g^{(3)}$ is the triplet correlation function, and $g^{(3)}$ is the deviation of $g^{(3)}$ from the superposition approximation. In practice, it is very difficult or even impossible to converge the three-particle and higher order correlation terms. However, it has been established that, for most fluids, the largest contribution to the excess entropy comes from the two-particle correlation term,[6] and the error induced by neglecting the higher order terms of the expansion may often be safely ignored.

Following the work of Lazaridis and Karplus,[6] we evaluate the two-particle excess entropy of liquid water by separating the two-particle term into translational and orientational components by factorization:

$$g(r, \omega^2) = g(r)g(\omega^2|r) \tag{2}$$

$$s_{\mathrm{e}}^{(2)} = s_{\mathrm{trans}}^{(2)} + s_{\mathrm{orient}}^{(2)} \tag{3}$$

$$s_{\mathrm{trans}}^{(2)} = -\frac{1}{2}k\rho\int [g(r)\ln g(r) - g(r) + 1]\, \mathrm{d}\boldsymbol{r} \tag{4}$$

$$s_{\mathrm{orient}}^{(2)} = \frac{1}{2}k\rho\int g(r)S^{\mathrm{orient}}(r)\, \mathrm{d}\boldsymbol{r} \tag{5}$$

$$S^{\mathrm{orient}} = -\frac{1}{\Omega^2}\int J(\omega^2)g(\omega^2|r)\ln g(\omega^2|r)\, \mathrm{d}\omega^2 \tag{6}$$

where r is the oxygen−oxygen distance of two water molecules, ω^2 are the angles that define the relative orientation of the two water molecules, $J(\omega^2)$ is the Jacobian of the angular variables, $g(r,\omega^2)$ is the pair correlation function, and $g(\omega^2|r)$ is the conditional-angular pair correlation function in the typical Bayesian notion. (Note that $g(r,\omega^2)$ is identical to $g^{(2)}$ as it appears in eq 1.) We denote the relative orientation of the two water molecules by the five angles[6] $[\theta_1,\theta_2,\phi,\chi_1,\chi_2]$, where θ_1,θ_2 are the angles between the intermolecular axis and the dipole vector of each molecule, ϕ describes the relative dihedral rotation of the dipole vector around the intermolecular axis, and χ_1,χ_2 describe the rotation of each molecule around its dipole vector. In the following discussion, we denote the entropy defined by formula 6 the orientational Shannon entropy,[18] and denote the entropy defined by formula 5 the orientational excess entropy.

In line with prior work,[6] we calculated the orientational Shannon entropy as defined by formula 6 for three different ranges of r: $(0 < r \leq 2.7)$, $(2.7 < r \leq 3.3)$, and $(3.3 < r \leq 5.6)$, which correspond to the various peaks and troughs in

the radial distribution function. In this way, the orientational excess entropy is related to Shannon entropy by:

$$s_{\text{orient}} = \frac{1}{2}N_i k S^{\text{orient}} \quad i = 1, 2, 3 \tag{7}$$

where N_i is the average number of water molecules in the ith shell.

2.2. Factorization of the Orientational Pair Correlation Function Using Generalized Kirkwood Superposition Approximation. The orientational pair correlation function (PCF) of water is a function of five angles, which is very difficult to converge from currently accessible molecular dynamics simulation time scales. The idea of factorization is to approximate the higher dimensional probability density function by the product of its lower dimensional marginal probability density functions. The generalized Kirkwood superposition approximation (GKSA)[19−21] allows an m-dimensional distribution to be estimated using corresponding $m - 1$-dimensional distributions:

$$\rho(x_1, x_2, ..., x_m) = \begin{cases} \dfrac{\prod_{c_{m-1}^m} \rho_{m-1} \cdots \prod_{c_2^n} \rho_2}{\prod_{c_{m-2}^m} \rho_{m-2} \cdots \prod_{c_1^n} \rho_1} & m \text{ is odd} \\ \dfrac{\prod_{c_{m-1}^m} \rho_{m-1} \cdots \prod_{c_1^n} \rho_1}{\prod_{c_{m-2}^m} \rho_{m-2} \cdots \prod_{c_2^n} \rho_2} & m \text{ is even} \end{cases} \tag{8}$$

where ρ_{m-k} represents a specific probability density function of $m - k$ dimensionality, and c_{m-k}^m indicates all possible combinations of $m - k$ groupings from the set of m total variables. Reiss[20] and Singer[21] have demonstrated that the GKSA is the optimal approximation of an n-particle distribution for $n \geq 3$ from a variational point of view, and it has been applied in numerous settings.[22,23]

From the results of our simulations, and as indicated by Lazaridis and Karplus,[6] the distribution has no structure along angle ϕ; that is, $g(\phi)$ is close to 1 over the range of ϕ and has no correlation with other angles. Thus, we approximated the five-dimensional PCF by:

$$g(\theta_1, \theta_2, \phi, \chi_1, \chi_2) = g(\theta_1, \theta_2, \chi_1, \chi_2) g(\phi) \tag{9}$$

Note that for any properly defined orientational PCF $g(x_1,x_2,...x_N)$,

$$\frac{1}{\Omega^{[x_1,x_2...x_n]}} \int J(x_1, x_2...x_n) g(x_1, x_2...x_n)\, dx_1\, dx_2...dx_n = 1 \tag{10}$$

where

$$\Omega^{[x_1,x_2...x_n]} = \int J(x_1, x_2...x_n)\, dx_1\, dx_2...dx_n \tag{11}$$

That is, $\Omega^{[x_1,x_2,...x_n]}$ is the integral of the Jacobian $J(x_1,x_2,...x_n)$ over angular variables $x_1,x_2,...x_n$. Therefore, $g(x_1,x_2,...x_n)$ is proportional to $\rho(x_1,x_2,...x_n)$ with proportional coefficient $\Omega^{[x_1,x_2,...x_n]}$. Via application of the GKSA (formula 8), it follows:

$$g(\theta_1, \theta_2, \chi_1, \chi_2) = \frac{g(\theta_1, \theta_2) g(\theta_1, \chi_1) g(\theta_1, \chi_2) g(\theta_2, \chi_1) g(\theta_2, \chi_2) g(\chi_1, \chi_2)}{g^2(\theta_1) g^2(\theta_2) g^2(\chi_1) g^2(\chi_2)} \tag{12}$$

Note that this factorization differs from that introduced by Karplus and Lazaridis[6] by the explicit inclusion of $g(\theta_1,\chi_1)$ and $g(\theta_2,\chi_2)$ terms. Taking this approximation of $g(x_1,x_2,...x_n)$ into the argument of the logarithm of formula 6, we find

$$S^{\text{orient}} = -\frac{1}{\Omega^2} \int J(\omega^2) g(\omega^2|r) \ln g(\omega^2|r)\, d\omega^2 \tag{13}$$

$$= -\sum_{C_2^4} \frac{1}{\Omega^{[x_1,x_2]}} \int J(x_1, x_2) g(x_1, x_2) \ln g(x_1, x_2)\, dx_1\, dx_2 + 2 \sum_{C_1^4} \frac{1}{\Omega^{[x]}} \int J(x) g(x) \ln g(x)\, dx \tag{14}$$

$$= \sum_{C_2^4} S^{[x_1,x_2]} - 2 \sum_{C_1^4} S^{[x]} \tag{15}$$

where x_1,x_2 is any combination of two variables from the $[\theta_1,\theta_2,\chi_1,\chi_2]$ set, x is any variable from the $[\theta_1,\theta_2,\chi_1,\chi_2]$ set, $J(x_1,x_2)$ is the Jacobian of the corresponding two variables, $J(x)$ is the Jacobian corresponding to variable x, $\Omega^{[x_1,x_2]}$ is the total accessible angular volume of variables x_1,x_2, $\Omega^{[x]}$ is the total accessible angular volume of variable x, $S^{[x_1,x_2]}$ is the Shannon entropy of angular variables x_1 and x_2, and $S^{[x]}$ is the Shannon entropy of angular variable x.

We note that an ambiguity seems to exist in the literature as to how to properly apply an approximation of the type suggested in eq 12 to eq 6. We have adopted here to apply the approximation only to the logarithm of eq 6 (as was done in the original derivation of eq 1), which allows result 15 to be interpreted through the language of information theory.[24] An alternate approach, which has been adopted by others, has been to apply approximation 12 to both occurrences of the PCF in eq 6, taking care to renormalize the factorization of the PCF introduced in eq 12 so that meaningful results will still be obtained. Interestingly, the results of these two approaches do not numerically agree, which may not be obvious from cursory inspection. We leave this proof as an exercise for the reader, which can be readily shown for instance from a correlated multidimensional Gaussian distribution.

2.3. The *k*'th Nearest-Neighbor Method. The NN method[14] gives an asymptotically unbiased estimate of an integral of the form:

$$I = -\int \rho(x_1, x_2, ..., x_s) \ln \rho(x_1, x_2, ..., x_s)\, dx_1\, dx_2...dx_s \tag{16}$$

where $\rho(x_1,x_2,...x_s)$ is the probability density function. Given a reasonable estimation of probability density function $f(x^i)$, the value of integral can be approximated as

$$I \approx -\frac{1}{n} \sum_{i=1}^{n} \ln f(x^i) \tag{17}$$

which follows from x^i being sampled from the true distribution $\rho(x^i)$. The NN method of nonparametrically estimating $f(x^i)$ at a point $x^i = (x_1^i,x_2^i,...x_s^i)$ is[25]

$$f(x^i) = \frac{k}{n}\frac{1}{V_s(R_{i,k})} \tag{18}$$

$$V_s(R_{i,k}) = \frac{\pi^{s/2}R_{i,k}^s}{\Gamma\left(\frac{1}{2}s + 1\right)} \tag{19}$$

where n is the number of data points in the sample, $V_s(R_{i,k})$ is the volume of an s-dimensional sphere with radius $R_{i,k}$, and $R_{i,k}$ is the Euclidean distance between the point x^i and its kth nearest neighbor in the sample. This approximation amounts to assuming that the distance between neighboring sampled points in configuration space will be small where the probability density function is large, and vice versa. So this integration may be estimated as

$$I \approx -\frac{1}{n}\sum_{i=1}^{n}\ln f(x^i) = \frac{1}{n}\sum_{i=1}^{n}\ln\frac{n\pi^{s/2}R_{i,k}^s}{k\Gamma\left(\frac{1}{2}s + 1\right)} \tag{20}$$

However, the estimate in eq 20 is systematically biased[14] and will deviate from the correct result in the limit of large n by $L_{k-1} - \ln k - \gamma$, where $L_j = \Sigma_{i=1}^{j}1/i$ and $\gamma = 0.5772...$ is Euler's constant. By subtracting the bias $L_{k-1} - \ln k - \gamma$, the modified unbiased estimate is formulated as

$$I \approx \frac{s}{n}\sum_{i=1}^{n}\ln R_{i,k} + \ln\frac{n\pi^{s/2}}{\Gamma\left(\frac{1}{2}s + 1\right)} - L_{k-1} + \gamma \tag{21}$$

Now our goal is to modify our expressions for the Shannon entropies into a form that is amenable to a kth NN evaluation of the integral. The expression of the two-dimensional orientational Shannon entropy has the form of

$$S^{[x_1,x_2]} = -\frac{1}{\Omega^{[x_1,x_2]}}\int J(x_1,x_2)g(x_1,x_2)\ln g(x_1,x_2)\,dx_1\,dx_2 \tag{22}$$

where $J(x_1,x_2)$ is the Jacobian associated with x_1 and x_2. Here, for χ_1 and χ_2 the Jacobian is 1, but for θ_1 and θ_2 the Jacobian is sin θ_1 and sin θ_2. However, by a change of variables from θ to $t = \pi/2(\cos\theta + 1)$, the Jacobian for t becomes 1, and the total angular volume is π for one-dimensional distribution and π^2 for two-dimensional distributions. Next, $g(x_1,x_2)$ is proportional to $\rho(x_1,x_2)$ in eq 16, with proportional coefficient π^2. Following the NN method, the statistically unbiased estimation of the one- and two-dimensional orientational Shannon entropies may now be approximated as

$$H_k^{[x]}(n) = \frac{1}{n}\sum_{i=1}^{n}\ln R_{i,k} + \ln\frac{n\pi^{1/2}}{\Gamma\left(\frac{1}{2} + 1\right)\Omega^{[x]}} - L_{k-1} + \gamma \tag{23}$$

$$H_k^{[x_1,x_2]}(n) = \frac{2}{n}\sum_{i=1}^{n}\ln R_{i,k} + \ln\frac{n\pi^{1}}{\Gamma\left(\frac{1}{2}\times 2 + 1\right)\Omega^{[x_1,x_2]}} - L_{k-1} + \gamma \tag{24}$$

where $H_k^{[x]}(n)$ is the kth NN estimate of the Shannon entropy of random variable x from a sampling of n data points, and $H_k^{[x_1,x_2]}(n)$ is the kth NN estimate of the joint Shannon entropy of random variables x_1,x_2 from a sampling of n data points. Thus, we are now equipped to apply the NN method of estimating the entropy to liquid state problems. We also note that to compute the NN distances, we made use of the ANN code,[26] which utilizes the $k-d$ tree algorithm[27] for obtaining the kth NN distances $R_{i,k}$ between sample points as necessary.

2.4. Error Analysis of the *k*th Nearest-Neighbor Method. It has been shown through an analysis of the limiting distribution[14] that the variance of the kth NN estimate of the entropy $H_k(n)$ is

$$\mathrm{var}[H_k(n)] = \frac{Q_k + \mathrm{var}[\ln f(x)]}{n} \tag{25}$$

where $f(x)$ is the probability density function and $Q_k = \sum_{j=k}^{\infty}1/j^2$. Formally, this result follows from using the Poisson approximation of the binomial distribution to characterize the fluctuations of $H_k(n)$ in the large n limit (please see ref 14 for details). Because $H_k(n)$ is asymptotically unbiased,[14] the asymptotic mean square error of the estimate is of the order given by eq 25. Typically, the true value $H(n)$ will be estimated by computing $H_k(n)$ for several values of k, typically 1−5. Because the analytical form of the variance is known, we may combine these estimates by a weighted averaging procedure, that is, $H(n) = \sum w_k H_k(n)$. For independent variables with the same average, the weight that minimizes the variance of the estimate of the average is a weight proportional to the inverse of the variance of the variable (see Appendix A for details), that is,

$$w_k = \frac{1/(Q_k + \mathrm{var}[\ln f(x)])}{\sum_{i=1}^{m}1/(Q_k + \mathrm{var}[\ln f(x)])}\text{for } k = 1, 2...m \tag{26}$$

where w_k is the ideal weight of $H_k(n)$ when averaging $H(n)$. Such calculations may also be readily extended to compute the standard deviation of such an estimate (Appendix A). Interestingly, two well-defined limits exist here: (1) if var[ln $f(x)$] is small, then the proper weighting will be

$$w_k = \frac{1/Q_k}{\sum_{k=1}^{m}1/Q_k}\text{for } k = 1, 2...m \tag{27}$$

and, (2) if var[ln $f(x)$] is large, then the proper weighting will be a flat function, which will lead to a simple arithmetic average. Therefore, the best possible estimate of $H(n)$ from m estimates of $H_k(n)$ will always be bound by these two limiting averages. Further, if these two limiting averages converge in the given sampling, it is highly probable the estimate of $H(n)$ is also converged. We also note here that an intuitive sense of which regime best fits the given data can be discerned by inspecting the relative noise in plots of the m $H_k(n)$ estimates as a function of n (where n is the amount of simulation time in this application). If the $H_1(n)$ estimate noticeably suffers greater fluctuations than the other estimates, then the var[ln $f(x)$] term must be small, because the Q_1 component is dominating relative variances of the estimates. However, if the m $H_k(n)$ estimates all appear

graphically to have fluctuations of a similar magnitude, then the var[ln $f(x)$] term must be large, and the simple arithmetic average is more appropriate. Such inspection of our data revealed var[ln $f(x)$] to be small. As such, the weighted average determined by application of eq 27 was taken in this work as our best possible estimate of $H(n)$.

2.5. Calculation of the Excess Energy, Enthalpy, and Free Energy. The excess molar energy of a fluid is simply

$$\Delta E = \frac{1}{2}\frac{\rho}{\Omega^2}\int g(r,\omega^2)u(r,\omega^2)\,\mathrm{d}r\,\mathrm{d}\omega^2 \qquad (28)$$

where $u(r,\omega^2)$ is the interaction energy between two molecules with distance r and orientation determined by ω^2. This quantity is straightforward to extract from the simulation, as it is merely one-half of the interaction energy between the water molecule of interest with the rest of the system. The molar excess enthalpy can be obtained by approximating the $\Delta(PV)$ term. For the liquid phase, the PV term may be safely neglected, and for the gas phase, we may use the ideal gas equation of state $PV = NkT$ to derive an excellent approximation to the PV term analytically. Combined with the excess entropy, we find the excess free energy of the fluid may be expressed as

$$\Delta G = \Delta E + \Delta(PV) - Ts_e \qquad (29)$$

as is typical.

2.6. The Finite-Difference Method of Entropy Calculation. To generate reference data to examine the accuracy of the kth NN method of evaluating the entropy expansion, we pursued a finite difference analysis of the solvation free energy, as computed from free energy perturbation theory (FEP). The finite-difference (FD) method of computing an entropy from FEP data proceeds by first noting that the entropy is the temperature derivative of the free energy, and then attempting to accurately estimate this slope,[28] that is

$$-\Delta S(T) = \left\langle\frac{\partial\Delta G}{\partial T}\right\rangle_P = \frac{\Delta G(T+\Delta T) - \Delta G(T-\Delta T)}{2\Delta T} \qquad (30)$$

This method relies on the assumption that the heat capacity of the system is independent of temperature in the range $[T - \Delta T, T + \Delta T]$.[29] This assumption appears to be valid near room temperature with ΔT even as large as 50 K.[28] Here, we use the Bennett acceptance ratio[30] method to calculate the excess free energy of liquid water at $T = 298 \pm 20$ K, and then use FD to calculate the excess entropy at $T = 298$ K. The datails of this method are included in the appendices. These data allow for independent validation of the NN approach and the approximations therein.

2.7. Details of the Simulation. Dynamics trajectories were generated using the Desmond molecular dynamics program.[31] A 25 Å cubic box of the TIP4P[32] water model was first equilibrated to 298 K and 1 atm with Nose−Hoover[33,34] temperature and Martyna−Tobias−Klein[35] pressure controls, followed by 30 ns NVT dynamics simulation with a Nose−Hoover[33,34] temperature control. To integrate the equations of motion of the system, the RESPA[36] integrator was used, where the integration step was 2 fs for the bonded and the nonbonded-near interactions and 6 fs for the nonbonded-far interactions. Configurations were collected every 1.002 ps. The cutoff distance was 9 Å for the van der Waals interaction, and the particle-mesh Ewald[37] method was used to model the electrostatic interactions. Similar simulations were performed for the SPC,[38] SPC/E,[39] TIP3P,[32] and TIP4P-Ew[40] water models.

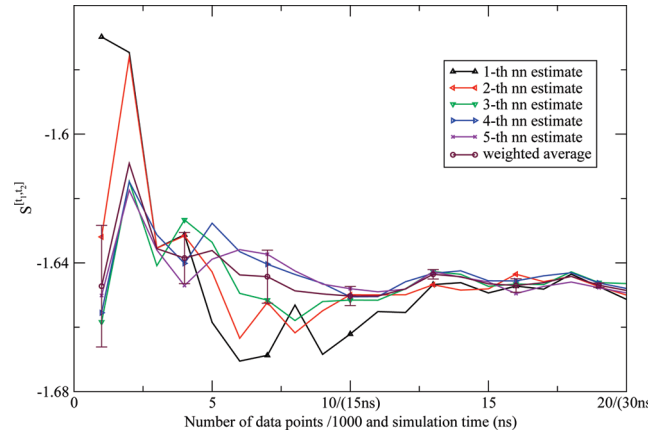

Figure 1. The first shell orientational Shannon entropy $S^{[t_1,t_2]}$ for the TIP3P model as a function of the number of data points (labeled on the horizontal axis in front of "/" in units of 1000) and the corresponding simulation time (labeled on the horizontal axis in parentheses) using the NN method. The weighted average estimate and the associated error bar were also depicted.

When extracting the solvent configurations to compute the PCF, we chose to only use the configurations of a distinguished solvent molecule with the rest of the system, instead of collecting statistics from all pairs of solvent molecules. Our primary interest in developing this technique was to later adapt the method to study the solvation of solutes; thus, we were interested in determining realistic estimates of the convergence of the technique when the isotropic symmetry of the fluid was not present. Such a protocol allows for an interrogation of the relative convergence properties of the various methods that might be obscured by the additional statistics offered by taking advantage of the symmetry of the system.

3. Results and Discussion

3.1. The Shannon Entropies. The NN estimates of the two-dimensional orientational Shannon entropies $S^{[r_1,r_2]}$ of the TIP3P water model for the three shells are given in Figures 1, 2, and 3. The results reported in these figures were generally representative of those results obtained for the other models. We see from the figures that the weighted average estimate of all of the Shannon entropies is converged over the course of the simulations. The results of all of the one- and two-dimensional orientational Shannon entropies for each of the three shells for all of the water models studied are given in Table 1. By application of formulas 4 and 7, we computed the translational excess entropies and orientational excess entropies for all of the water models studied. All of the final results are shown in Table 2. From the table, we

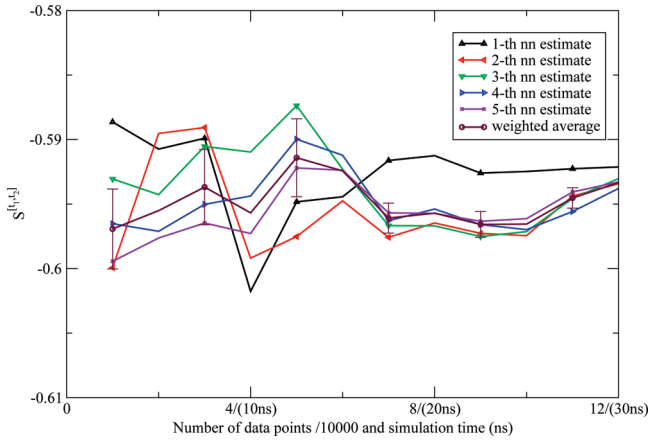

Figure 2. The second shell orientational Shannon entropy $S^{[t_1,t_2]}$ for the TIP3P model as a function of the number of data points (labeled on the horizontal axis in front of "/" in units of 10 000) and the corresponding simulation time (labeled on the horizontal axis in parentheses) using the NN method. The weighted average estimate and the associated error bar were also depicted.

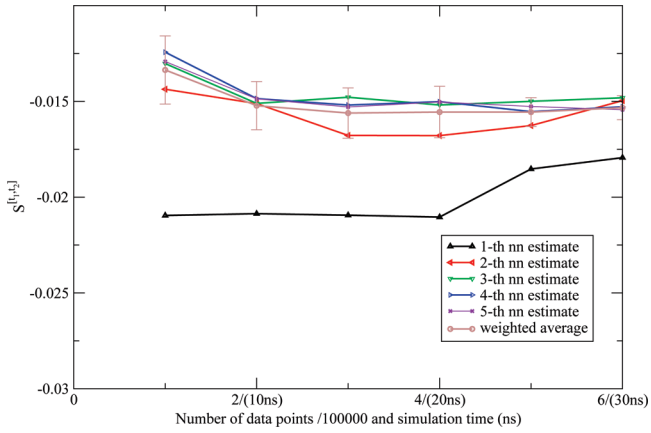

Figure 3. The third shell orientational Shannon entropy $S^{[t_1,t_2]}$ for the TIP3P model as a function of the number of data points (labeled on the horizontal axis in front of "/" in units of 100 000) and the corresponding simulation time (labeled on the horizontal axis in parentheses) using the NN method. The weighted average estimate and the associated error bar were also depicted.

see that for the TIP4P model the excess entropy result from the NN method, −13.67 eu, is very close to the experimental value, −14.1 eu. We also note excellent agreement between the excess entropies computed here and those derived from cell theory.[41] The agreement for the TIP3P and SPC models was slightly diminished as compared to the other models, for reasons that will be explained later.

3.2. Convergence Properties. We extensively compared the commonly employed histogram method to compute the orientational Shannon entropy to the NN method weighted average (Figures 4, 5, and 6). We see clearly that the NN method weighted average converges much faster than the histogram method for shells 1 and 2. For shell 3, both methods give similar results. This is easily understood: for the first and second shells, the water molecules are highly correlated, and the histogram results will have a strong dependency on the bin size used to do the integration; however, for the third shell, there is little correlation, so the histogram method has similar convergence properties as compared to the NN method.

Figures 7, 8, 9, 10, and 11 depict the total orientational excess entropies as a function of simulation time from the various histogram estimates and the NN weighted average estimate. For all of the models studied, the 10° histogram estimate (which is most commonly used currently[6,10]) gave results closest to the NN estimate. However, for a bin size of 20°, the entropy result is biased away from the correct result, and for bin sizes of 5° and 2.5°, much longer simulation time would be needed to converge the results. Because ideal bin size is problem specific, it cannot be deduced unless other reference data are already known. Thus, the absence of such a parametric bias in the NN method is a notable advantage of the technique.

3.3. Error Analysis. As described in the Methods, we calculated the variance associated with the weighted average of the NN estimates for each of the one- and two-dimensional Shannon entropies. Because the NN estimate is asymptotically unbiased, the error of the estimate is also given by the variance. We calculated the error on the basis of the weighted average, which assumes var ln $f(x)$ is 0. However, even in the extreme cases where var ln $f(x)$ goes to infinity and the five NN estimates contribute equally to the average, the variance of the arithmetic average only differs slightly from weighted average, and they are within the error bar of each other, strongly indicating the convergence of these calculations (Figures 12 and 13).

3.4. The Radial Dependence of Orientational Shannon Entropy. We calculated the orientational Shannon entropies in three radial regions, assuming the orientational distribution would be independent of r in each subregion. To validate this approximation, we calculated the orientational Shannon entropies at different intervals of r from 2.5 to 4.0 Å. Typical Shannon entropies $S^{[t_1,t_2]}$ at different values of r are shown in Figure 14.

We see from the figure that the Shannon entropy increases as the distance between the two water molecules r increases, and goes to zero when r is sufficiently large. Additionally, the change of the Shannon entropy with respect to r is smooth in the respective first and second hydration shells. Because of the slow variation of the orientational Shannon entropy with respect to r, the sum of the orientational excess entropy at each interval will differ from the sum of the orientational excess entropy of the three shells only by at most 0.5 eu, which is within statistical uncertainty of the calculation. Thus, this approximation was not a large source of error in these calculations.

3.5. Inclusion of $g(\theta_1,\chi_1)$ in the Factorization. The factorization of the PCF used here differs from the more common formulation[6] by the explicit inclusion of $g(\theta_1,\chi_1)$ and $g(\theta_2,\chi_2)$. The distribution functions $g(\theta_1)*g(\chi_1)$ and $g(\theta_1,\chi_1)$ for the TIP4P model are shown in Figures 15 and 16. Careful inspection of these figures suggests that $g(\theta_1,\chi_1)$ differs from $g(\theta_1)g(\chi_1)$ quantitatively, which is supported by the two-dimensional Shannon entropy $S^{[\theta_1,\chi_1]}$ differing significantly from the sum of $S^{[\theta_1]}$ and $S^{[\chi_1]}$. For example, for the TIP4P model, the first shell Shannon entropy of $S^{[\theta_1,\chi_1]}$

Table 1. Orientational Shannon Entropies of the Five Water Models[a]

	water models	$S^{[t_1,t_2]}$	$S^{[t_1,\chi_1]}$	$S^{[t_1,\chi_2]}$	$S^{[\chi_1,\chi_2]}$	$S^{[t_1]}$	$S^{[\chi_1]}$
	TIP4P	−1.33	−1.21	−1.15	−1.02	−0.34	−0.29
	SPC	−1.67	−1.28	−1.24	−0.89	−0.50	−0.27
shell 1	TIP3P	−1.65	−1.16	−1.14	−0.74	−0.47	−0.23
	SPC/E	−1.70	−1.32	−1.29	−0.94	−0.51	−0.29
	TIP4P-Ew	−1.44	−1.29	−1.23	−1.05	−0.39	−0.30
	TIP4P	−0.59	−0.44	−0.46	−0.38	−0.10	−0.10
	SPC	−0.69	−0.42	−0.46	−0.30	−0.11	−0.09
shell 2	TIP3P	−0.60	−0.29	−0.34	−0.18	−0.09	−0.06
	SPC/E	−0.71	−0.46	−0.50	−0.33	−0.13	−0.10
	TIP4P-Ew	−0.68	−0.51	−0.53	−0.38	−0.12	−0.12
	TIP4P	−0.010	−0.007	−0.002	−0.003	−0.001	−0.000
	SPC	−0.014	−0.007	−0.005	−0.001	−0.002	−0.000
shell 3	TIP3P	−0.015	−0.003	−0.003	−0.001	−0.002	−0.000
	SPC/E	−0.013	−0.007	−0.005	−0.003	−0.001	−0.000
	TIP4P-Ew	−0.012	−0.007	−0.004	−0.001	−0.001	−0.000

[a] $t = \pi/2(\cos(\theta) + 1)$; all of these entropies are unitless.

Table 2. Comparison of Entropy Results from the NN Method and Cell Theory[a]

	EXP	TIP4P	TIP3P	SPC	SPC/E	TIP4P-Ew
$s^{(2)}_{trans}$		−3.15(3.14[b])	−2.99	−2.99	−3.19	−3.33
$s^{(2)}_{orient}$		−10.52(9.10[b])	−8.58	−10.20	−11.53	−11.76
$s^{(2)}_{ex}$		−13.67(−12.2[b])	−11.57	−13.19	−14.72	−15.09
s_{ex}	−14.05[c]	−14.32[d]	−13.36[d]	−14.01[d]	−14.79[d]	−14.99[d]

[a] Entropies in cal/(mol·K) (eu). [b] Data from Lazaridis.[6] [c] Data from Wagner.[42] [d] Data from Henchman by cell theory.[41]

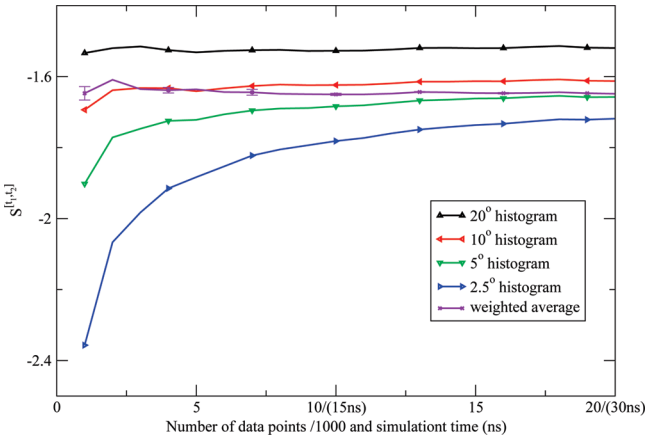

Figure 4. The first shell orientational Shannon entropy $S^{[t_1,t_2]}$ for the TIP3P model as a function of the number of data points (labeled on the horizontal axis in front of "/" in units of 1000) and the corresponding simulation time (labeled on the horizontal axis in parentheses) using histogram method. The weighted average of the NN estimates and the associated error bar were also depicted.

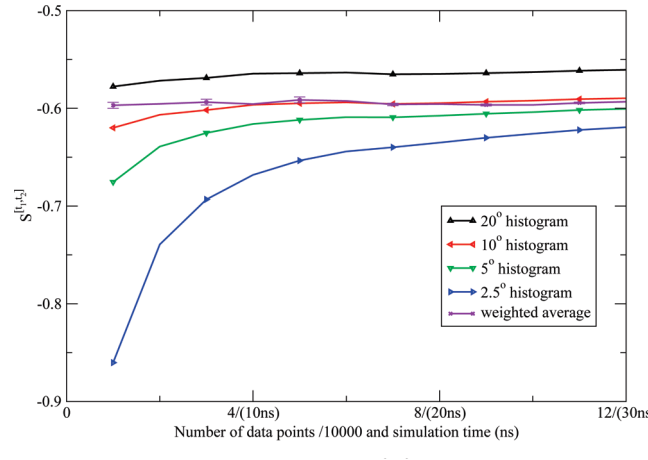

Figure 5. The second shell orientational Shannon entropy $S^{[t_1,t_2]}$ for the TIP3P model as a function of the number of data points (labeled on the horizontal axis in front of "/" in units of 10 000) and the corresponding simulation time (labeled on the horizontal axis in parentheses) using histogram method. The weighted average of the NN estimates and the associated error bar were also depicted.

is −1.21, while $S^{[\theta_1]}$ is −0.34 and $S^{[\chi_1]}$ is −0.29. This result indicated a non-negligible correlation between χ_1 and θ_1, which suggested that the explicit inclusion of $g(\theta_1,\chi_1)$ and $g(\theta_2,\chi_2)$ in our factorization would lead to greater quantitative precision. This also explains why our excess entropy result for the TIP4P model (−13.67 eu) is about 1.5 eu more negative than the previously reported value (−12.2 eu),[6] which is in better agreement with both the FD estimate of the entropy of the model and the experimental estimate of liquid water.

3.6. Comparison of Free Energy Results. From these simulations, we computed the excess molar energies and excess free energies of the various water models. The results of these calculations for all models studied are listed in Table 3 alongside the relevant literature values. The excess free energies we have obtained here show excellent agreement (within 0.5 kcal/mol uniformly) with the high precision FEP results obtained by Shirts et al.[43] Interestingly, the TIP4P model gives results closest to the experimental quantities.

The SPC/E, TIP4P, and TIP4P-Ew models all give free energy results somewhat closer to the Shirts[43] results than the other models. This may not be accidental. In our calculations, the higher order multiparticle correlation entropies were ignored. There is some literature precedence expecting these higher order contributions to the excess entropy to vanish at the temperature of solid−liquid phase transition.[44,45] Recently, Saija has shown that for the TIP4P

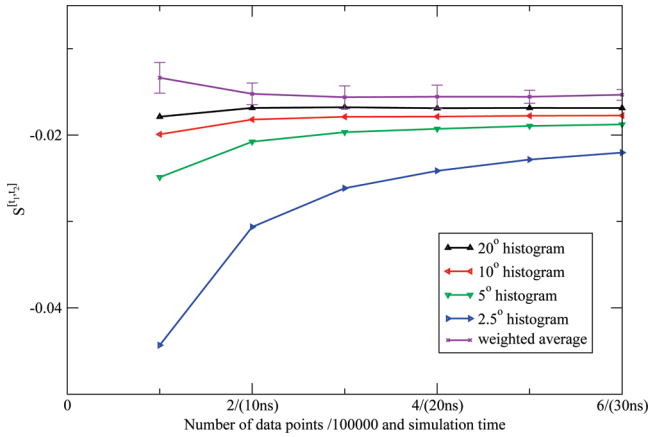

Figure 6. The third shell orientational Shannon entropy $S^{[t_1,t_2]}$ for the TIP3P model as a function of the number of data points (labeled on the horizontal axis in front of "/" in units of 100 000) and the corresponding simulation time (labeled on the horizontal axis in parentheses) using histogram method. The weighted average of the NN estimates and the associated error bar were also depicted.

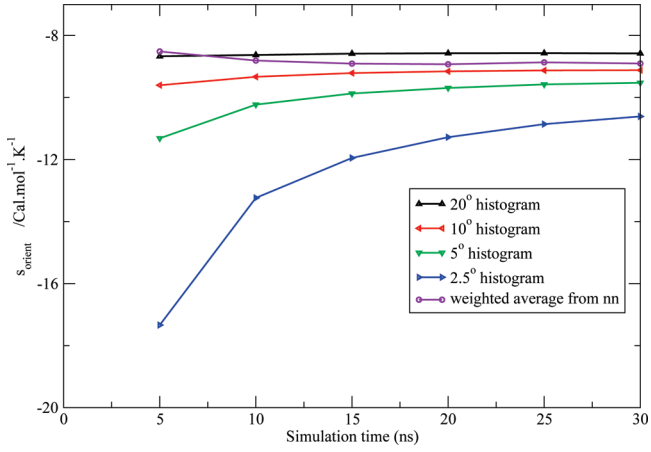

Figure 7. Total orientational excess entropy as a function of simulation time from the NN method and histogram method with different bin width for the TIP3P model.

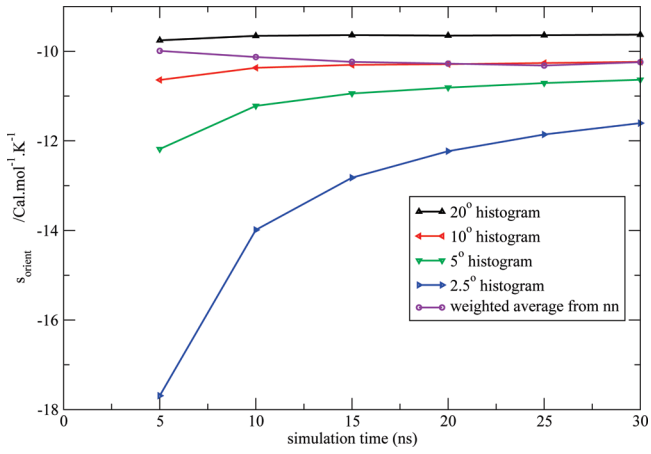

Figure 8. Total orientational excess entropy as a function of simulation time from the NN method and histogram method with different bin width for the SPC model.

model, the temperature of maximum density (TMD) coincides with the temperature where higher order contributions

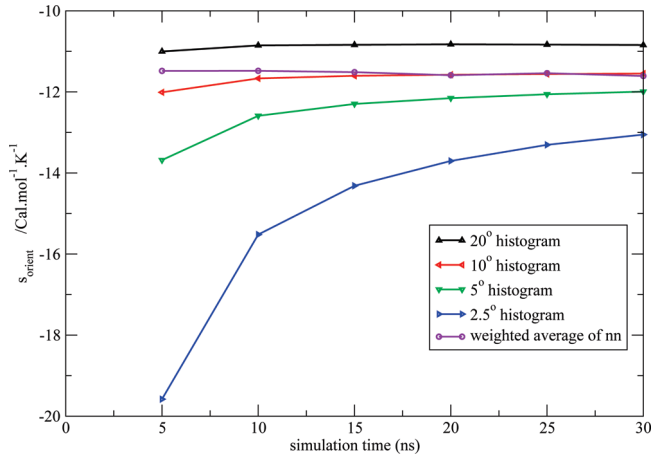

Figure 9. Total orientational excess entropy as a function of simulation time from the NN method and histogram method with different bin width for the SPC/E model.

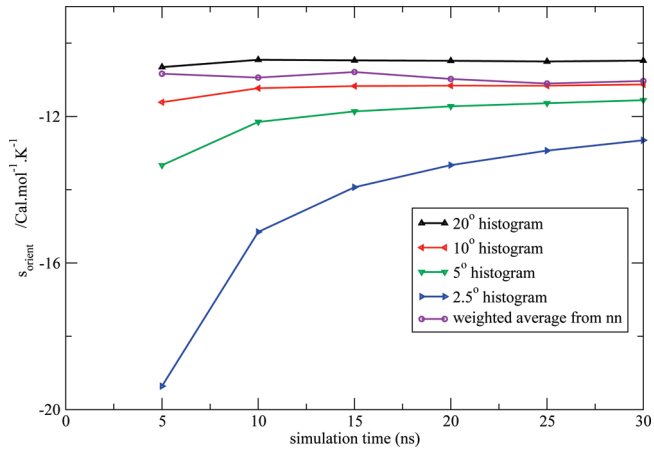

Figure 10. Total orientational excess entropy as a function of simulation time from the NN method and histogram method with different bin width for the TIP4P model.

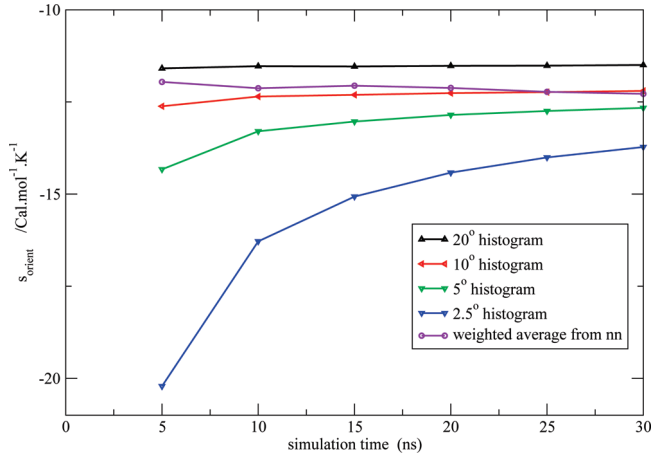

Figure 11. Total orientational excess entropy as a function of simulation time from the NN method and histogram method with different bin width for the TIP4P-Ew model.

to the entropy should vanish.[13] Studies of temperature dependence of the densities of the different water models studied here[46] have shown that the TMD of the TIP4P model occurred at 258 K, the TMD of the SCP/E model occurred at 235 K,[47] the TMD of the TIP4P-Ew model occurred at

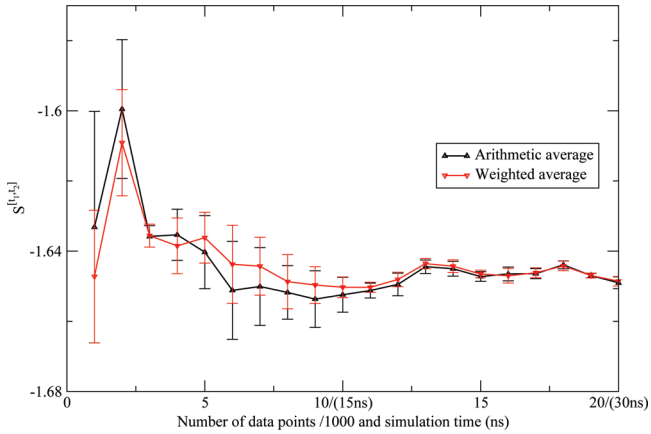

Figure 12. Comparison between the arithmetic average and the weighted average of the NN estimates for the first shell Shannon entropy $S^{[t_1,t_2]}$ for the TIP3P model. They are within the error bar of each other.

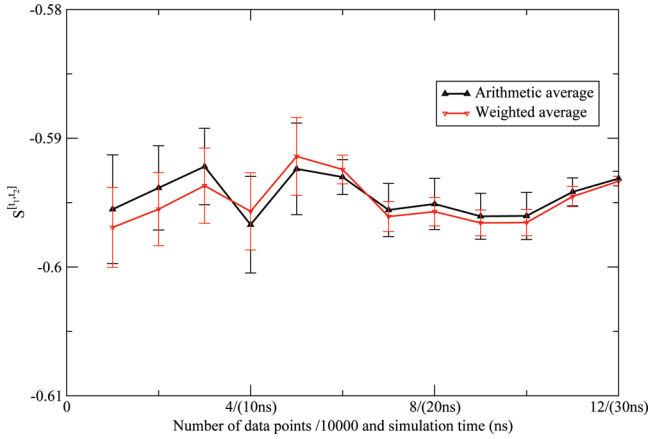

Figure 13. Comparison between the arithmetic average and the weighted average of the NN estimates for the second shell Shannon entropy $S^{[t_1,t_2]}$ for the TIP3P model. They are within the error bar of each other.

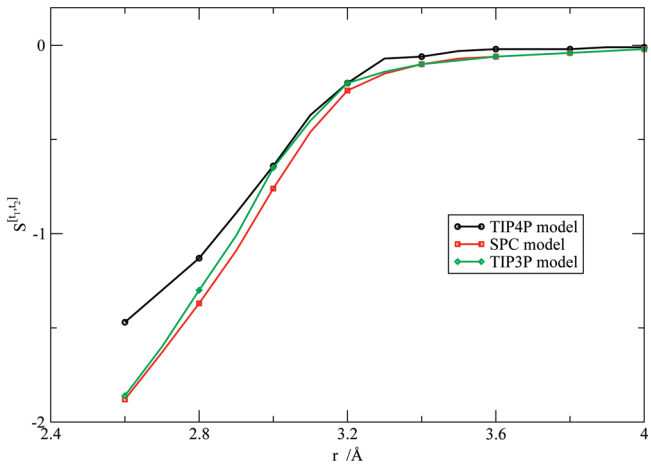

Figure 14. Orientational Shannon entropy $S^{[t_1,t_2]}$ as a function of r for the various water models.

272 K,[40] and the density of the SPC and TIP3P models increases monotonically as temperature decreases in the range [220,370].[46] This indicates, for the TIP3P and SPC models, multiparticle correlation entropy may contribute more to the total entropy than for the other models, which may be why

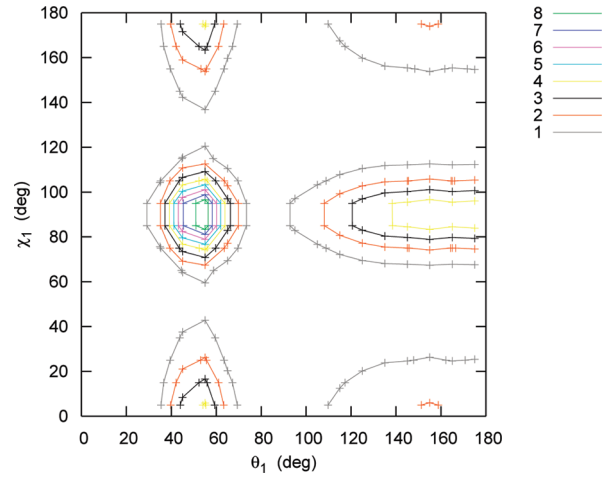

Figure 15. Products of one-dimensional marginal distribution function $g(\theta_1)^*g(\chi_1)$ for the TIP4P model in the first shell.

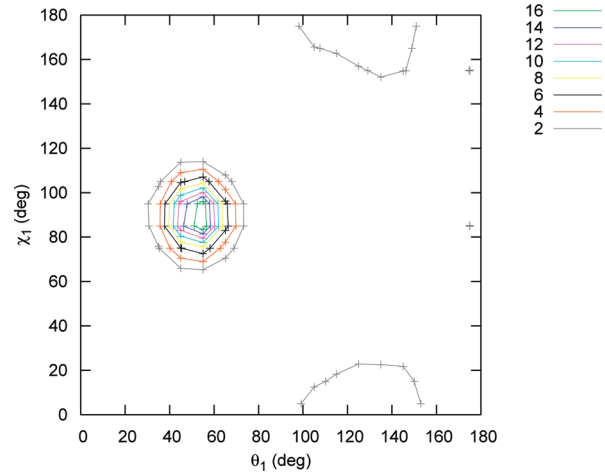

Figure 16. Two-dimensional marginal distribution function $g(\theta_1, \chi_1)$ for the TIP4P model in the first shell.

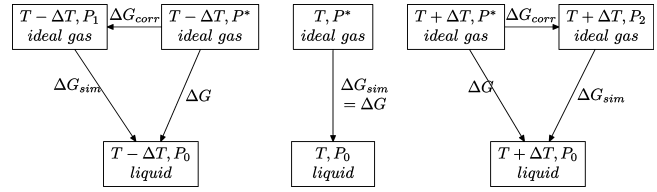

Figure 17. Thermodynamic cycle depicting the constant pressure corrections to ΔG_{sim} at temperatures $T \pm \Delta T$ when computing the slope of ΔG_{sim} with respect to T.

our quantitative accuracy for them is somewhat diminished. However, the molecular detail afforded by this technique in yielding both a value of the entropy and a physical interpretation of its meaning, in terms of the fluid structure implied by the shape of the pair correction function (PCF), gives it a comparative advantage over techniques such as FEP, which will generally only yield a value of the entropy without any additional molecular understanding of the system.

3.7. Entropy Calculation from FD Method. We calculated the excess free energy of water at temperature 298 ± 20 K with the Bennett acceptance ratio[30] method and obtained entropies at 298 K by the FD formula. The results

Table 3. Results for the Energy, Enthalpy, and Entropy of Liquid Water from Various Methods[a]

water models	TIP4P	TIP3P	SPC	SPC/E	TIP4P-Ew
excess energy	−9.85	−9.49	−9.90	−11.08	−10.91
excess enthalpy	−10.43	−10.07	−10.48	−11.66(−10.48[d])	−11.49(−10.45[e])
excess enthalpy[b]	−10.41	−10.09	−10.47	−11.69(−10.51[d])	−11.61(−10.57[e])
excess entropy from NN	−13.67	−11.57	−13.19	−14.72	−15.09
excess entropy[c]	−14.43	−13.39	−14.46	−15.57	−15.53
excess free energy from NN	−6.36	−6.63	−6.55	−7.27(−6.09[d])	−7.00(−5.96[e])
excess free energy[b]	−6.11	−6.10	−6.16	−7.05(−5.87[d])	−6.98(−5.94[e])
excess free energy from exp				−6.33	
excess enthalpy from exp				−10.52	

[a] Energies in kcal/mol, entropies in cal/(mol·K) (eu). [b] Results from Shirts.[43] [c] Results from Shirts[43] by subtracting enthalpy from free energy. [d] Include polarization correction.[39] [e] Include polarization correction.[40]

Table 4. Entropy Results from FD Method and Comparison with Other Methods[a]

water models	TIP4P	TIP3P	SPC	SPC/E	TIP4P-Ew
excess free energy at 278 K	−6.35[b]	−6.21(−6.24[d])	−6.36(−6.39[d])	−7.19(−7.23[d])	
excess free energy at 298 K	−6.03[b]	−5.95	−6.06	−6.89	
excess free energy at 318 K	−5.73[b]	−5.71(−5.69[d])	−5.80(−5.78[d])	−6.66(−6.62[d])	
excess entropy from FD	−15.2[b]	−13.8(±0.8[e])	−15.2(±0.8[e])	−15.3(±0.8[e])	
excess entropy from NN	−13.67	−11.57	−13.19	−14.72	−15.09
excess entropy from FEP[c]	−14.43	−13.39	−14.46	−15.57	−15.53

[a] Energies in kcal/mol, entropies in cal/(mol·K) (eu). [b] Results from Saija.[13] [c] Results from Shirts[43] by subtracting enthalpy from free energy. [d] Results in parentheses include constant pressure correction (Appendix B). [e] Indicates the error associated with the entropy.

are presented in Table 4. The excess entropies computed from the FD method are consistently larger in magnitude than those computed from the NN method, consistent with us neglecting the contributions from the higher order terms of the expansion.

As in the proceeding section, the NN and FD excess entropies of the SPC/E water are in very close agreement; however, the agreement of the NN and FD entropies of the SPC and TIP3P models is much poorer. We again expect the reason for this discrepancy to be due to the TMD of the SPC/E model being close to the range of temperatures treated in this study, while the TMDs of the SPC and TIP3P models fall well outside this range. Thus, the higher order terms of the entropy expansion are expected to make larger contributions to the excess entropies for the SPC and TIP3P models versus the contribution made to the excess entropy of the SPC/E water.

4. Conclusion

Our results indicate that the NN method of computing entropies in the liquid state offers several compelling advantages over the more common histogram approaches, including (1) much faster convergence for a given amount of simulation data; (2) an intuitive error bound for the uncertainty of the calculation without resorting to block averaging or bootstrapping techniques, which may be problematic to apply to estimators of the entropy; and (3) not relying on empirically tuned parameters, such as the histogram bin width, which may bias the results in an unpredictable fashion. We also found that inspection of the limiting behaviours of var ln $f(x)$ may be used to both analyze the convergence of the given calculation and develop the best possible estimate of the entropy given a set of calculated $H_k(n)$. Although we also found that a judicious choice of the histogram bin width may mitigate these advantages, such a choice is difficult to make without prior knowledge of the properties of the limiting distribution, which may not be available when new problems are investigated.

Our alternative factorization of the water−water correlation function, which explicitly included correlations between the angle formed by the water dipole vector and the intermolecular axis with the angle of rotation of the water molecule about its dipole vector, was found to increase the agreement of results obtained by the entropy expansion with those obtained by less approximate methods, such as FEP and the FD benchmark calculations. This result suggests that this contribution should not be ignored in future studies of the excess entropy of liquid water and other fluids.

Acknowledgment. This research was supported by the National Institutes of Health through a grant to R.A.F. (NIH-GM-40526), by the National Science Foundation through a grant to B.J.B. (NSF-CHE-1689) and an NSF Fellowship to R.A., and an allocation of computer time on TeraGrid resources provided by NCSA under NSF auspices.

Appendix A: Determination of Most Proper Weights

Given that $x_1, x_2,...x_n$ are independent variables with the same average u but different variance $v_1, v_2,...v_n$, we may define $\bar{x} = \sum_{i=1}^{n} w_i x_i$, with constraint $\sum_{i=1}^{n} w_i = 1$. We may find the weights w_i such that the variance of $\bar{x}$ is minimized:

$$\mathrm{var}[\bar{x}] = \sum_{i=1}^{n} (w_i)^2 v_i \tag{1}$$

Using Lagrange multipliers, we find:

$$w_i = \frac{\frac{1}{v_i}}{\sum_{i=1}^{n} \frac{1}{v_i}} \tag{2}$$

and

$$\mathrm{var}[\bar{x}] = \frac{1}{\sum_{i=1}^{n} \frac{1}{v_i}} \tag{3}$$

$$E[\sum_{i=1}^{n} w_i(x_i - \bar{x})^2] = E[\sum_{i=1}^{n} w_i((x_i - u) - (\bar{x} - u))^2] \tag{4}$$

$$= E[\sum_{i=1}^{n} w_i((x_i - u)^2 - 2(x_1 - u)(\bar{x} - u) + (\bar{x} - u)^2)] \tag{5}$$

$$= E[\sum_{i=1}^{n} w_i(x_i - u)^2] - 2E[\sum_{i=1}^{n} w_i(x_i - u)(\bar{x} - u)] + E[\sum_{i=1}^{n} w_i(\bar{x} - u)^2] \tag{6}$$

By application of eq 2 and $\sum_i{}^n{}_{=1} w_i = 1$, we find:

$$E[\sum_{i=1}^{n} w_i(x_i - \bar{x})^2] = \frac{n-1}{\sum_{i=1}^{n} \frac{1}{v_i}} \tag{7}$$

Thus, we can approximate the variance of the weighted average by the estimator:

$$V = \frac{1}{n-1} \sum_{i=1}^{n} w_i(x_i - \bar{x})^2 \tag{8}$$

Appendix B: Constant Pressure Correction to ΔG_{sim} for the FD Entropy

In the FEP simulations, we turned on/off the interaction between one distinguished water molecule with the rest of the system at constant temperature T and constant pressure P_0, over the series of several λ windows. The solvation free energy of the distinguished water molecule corresponds to the difference in the chemical potential μ between two phases: (1) the liquid phase and (2) the ideal gas phase with the same temperature and number density as the liquid.[48] For example,

$$\Delta G_{\mathrm{sim}}(T) = -kT \ln \frac{\tilde{\Delta}(\lambda = 1)}{\tilde{\Delta}(\lambda = 0)} = \mu_{\mathrm{l}}(N, P_0, T) - \mu_{\mathrm{g}}(N, P^*, T) \tag{9}$$

where P^* is the pressure of the ideal gas with the same temperature T and number density as the simulated liquid at pressure P_0, and $\tilde{\Delta}$ is the isobaric−isothermal partition function of the system specified by lambda. (For details, please see ref 48.)

The heat capacity of the ideal gas at constant pressure P^* is trivially constant with respect to temperature, and we may well approximate the heat capacity of liquid water to also be constant under constant pressure P_0 over the temperature range studied here. It then follows:

$$\Delta G(T) = \Delta H(T) - T\Delta S(T) \tag{10}$$

$$\Delta H(T \pm \Delta T) = \Delta H(T) \pm \Delta C_P \Delta T \tag{11}$$

$$\Delta S(T \pm \Delta T) = \Delta S(T) + \Delta C_P \ln \frac{T \pm \Delta T}{T} \tag{12}$$

$$\Delta S(T) \approx -\frac{\Delta G(T + \Delta T) - \Delta G(T - \Delta T)}{2\Delta T} \tag{13}$$

which are the typical equations of the finite difference method of computing the thermodynamic entropy. In these equations, all of the Δ quantities correspond to the difference of the thermodynamic quantities between the liquid phase at P_0 and the ideal gas phase at P^*.

In similar simulations run at pressure P_0 but temperatures $T \pm \Delta T$, we analogously find

$$\Delta G_{\mathrm{sim}}(T - \Delta T) = \mu_{\mathrm{l}}(N, P_0, T - \Delta T) - \mu_{\mathrm{g}}(N, P_1, T - \Delta T) \tag{14}$$

$$\Delta G_{\mathrm{sim}}(T + \Delta T) = \mu_{\mathrm{l}}(N, P_0, T + \Delta T) - \mu_{\mathrm{g}}(N, P_2, T + \Delta T) \tag{15}$$

where P_1 and P_2 correspond to the ideal gas pressure with the same temperature and number density as the simulated liquids. Note that the ΔG values obtained from simulation differ from those occurring in eq 13 because the reference gas-phase free energies differ, and thus we must explicitly correct for this difference in the reference state. By adding a correction term $\Delta G_{\mathrm{corr}}(T \pm \Delta T)$ to the simulated free energy, we were able to use eq 13 to calculate the entropy at temperature T, where:

$$\Delta G_{\mathrm{corr}}(T - \Delta T) = \mu_{\mathrm{g}}(N, P_1, T - \Delta T) - \mu_{\mathrm{g}}(N, P^*, T - \Delta T) = k(T - \Delta T) \ln \frac{P_1}{P^*} \tag{16}$$

$$\Delta G_{\mathrm{corr}}(T + \Delta T) = \mu_{\mathrm{g}}(N, P_2, T + \Delta T) - \mu_{\mathrm{g}}(N, P^*, T + \Delta T) = k(T + \Delta T) \ln \frac{P_2}{P^*} \tag{17}$$

and

$$\Delta S(T) = -\frac{\Delta G_{\mathrm{sim}}(T + \Delta T) + \Delta G_{\mathrm{corr}}(T + \Delta T) - \Delta G_{\mathrm{sim}}(T - \Delta T) - \Delta G_{\mathrm{corr}}(T - \Delta T)}{2\Delta T} \tag{18}$$

These corrections, although small in magnitude, were systematically of opposite sign at temperatures $T \pm \Delta T$ because the thermal expansion coefficient of liquid water differs from the thermal expansion coefficient of the ideal gas. As a result, failure to apply these corrections will lead to a non-negligible systematical bias in the FD-FEP entropy.

The thermodynamic cycle indicating the whole process, including correction terms, is depicted in Figure 17. Note that in the cycle depicted in Figure 17, we must compute the correction terms at temperatures $T \pm \Delta T$ to compute the slope of ΔG with respect to T, that is, the entropy associated

with the solvation free energy of transferring the water molecule from the gas phase to the liquid phase at temperature T.

References

(1) Berne, B. J.; Weeks, J. D.; Zhou, R. *Annu. Rev. Phys. Chem.* **2009**, *60*, 85–103.

(2) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.

(3) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 808–813.

(4) Lazaridis, T. *J. Phys. Chem. B* **1998**, *102*, 3531–3541.

(5) Lazaridis, T.; Paulattis, M. E. *J. Phys. Chem.* **1992**, *96*, 3847–3855.

(6) Lazaridis, T.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 4294–4316.

(7) Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2006**, *110*, 1464–1475.

(8) Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2005**, *109*, 662–670.

(9) Zielkiewicz, J. *J. Phys. Chem. B* **2008**, *112*, 7810–7815.

(10) Zielkiewicz, J. *J. Chem. Phys.* **2005**, *123*, 104501.

(11) Esposito, R.; Saija, F.; Saitta, A. M.; Giaquinta, P. V. *Phys. Rev. E* **2006**, *73*, 040502.

(12) Silverstein, K. A. T.; Dill, K. A.; Haymet, A. D. J. *J. Chem. Phys.* **2001**, *114*, 6303–6314.

(13) Saija, F.; Saitta, A. M.; Giaquinta, P. V. *J. Chem. Phys.* **2003**, *119*, 3587–3589.

(14) Singh, H.; Misra, N.; Hnizdo, V.; Fedorowicz, A.; Demchuk, E. *Am. J. Math. Manag. Sci.* **2003**, *23*, 301–322.

(15) Green, H. S. *Molecular Theory of Fluids*; North-Holland: Amsterdam, 1952; Chapter 3.

(16) Raveché, H. J. *J. Chem. Phys.* **1971**, *55*, 2242–2250.

(17) Wallace, D. C. *J. Chem. Phys.* **1987**, *87*, 2282–2284.

(18) Shannon, C. E. *Bell. Syst. Tech. J.* **1948**, *27*, 379–423.

(19) Fisher, I. Z.; Kopeliovich, B. L. *Dokl. Akad. Nauk SSSR* **1960**, *133*, 81–83.

(20) Reiss, H. *J. Stat. Phys.* **1972**, *6*, 39–47.

(21) Singer, A. *J. Chem. Phys.* **2004**, *121*, 3657–3666.

(22) Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 024107.

(23) Hnizdo, V.; Darian, E.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. *J. Comput. Chem.* **2007**, *28*, 655–668.

(24) Matsuda, H. *Phys. Rev. E* **2000**, *62*, 3096–3102.

(25) Loftsgaarden, D. O.; Quesenberry, C. P. *Ann. Math. Statist.* **1965**, *36*, 1049–1051.

(26) Arya, S.; Mount, D. M. Approximate nearest neighbor queries in fixed dimensions. *SODA '93: Proceedings of the fourth annual ACM-SIAM Symposium on Discrete Algorithms*; Philadelphia, PA, 1993; pp 271−280.

(27) Freidman, J. H.; Bentley, J. L.; Finkel, R. A. *ACM Trans. Math. Softw.* **1977**, *3*, 209–226.

(28) Smith, D. E.; Haymet, A. D. *J. Chem. Phys.* **1993**, *98*, 6445–6454.

(29) Wan, S. Z.; Stote, R. H.; Karplus, M. *J. Chem. Phys.* **2004**, *121*, 9539–9548.

(30) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(31) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable algorithms for molecular dynamics simulations on commodity clusters. *SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing*; New York, NY, 2006; p 84.

(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(33) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(34) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(35) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(36) Tuckerman M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

(37) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(38) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, 1981; pp 331−342.

(39) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(40) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(41) Henchman, R. H. *J. Chem. Phys.* **2007**, *126*, 064504.

(42) Wagner, W.; Pruβ, A. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387–478.

(43) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(44) Wallace, D. C. *Int. J. Quantum Chem.* **1994**, *52*, 425–435.

(45) Giaquinta, P. V.; Giunta, G. *Physica A* **1992**, *187*, 145–158.

(46) Jorgensen, W. L.; Jenson, C. *J. Comput. Chem.* **1998**, *19*, 1179–1186.

(47) Baez, L. A.; Clancy, P. *J. Chem. Phys.* **1994**, *101*, 9837–9840.

(48) Horn, H. W.; Swope, W. C.; Pitera, J. W. *J. Chem. Phys.* **2005**, *123*, 194504.

CT900078K