A Simple Nonadditive Model of Water

Gregory G. Wood[1]

[1]Department of Mathematics and Physics, California State University Channel Islands, Camarillo, CA, USA

Abstract—*Liquid water has rich thermodynamic behavior over a range of temperatures and pressures[1]. Models of water used in protein folding simulations must be fast and reflect the underlying hydrogen bond network accurately. Although greatly simplified, current models of water can account for more then 99% of the CPU time during numerical simulations[2]. Current models assume simple additivity of free energy which is incorrect over coupled degrees of freedom[3], [4] - and in the liquid state all of the water is coupled. A novel statistical mechanical model of water is presented encapsulating the essential nonadditive free energies without recourse to computationally expensive techniques.*

Keywords: model of water, thermodynamic properties, statistical mechanics, monte carlo simulation

1. Introduction

Liquid water is a system of great significance for the study of biomolecules[5], [6]. Although pure liquid water is a poor analog of the aqueous environments found in cells, in nature, it is nonetheless widely used in both experiment and numerical simulation. Even simple models of water are computationally expensive with orders of magnitude more CPU time spent on the water rather then the biomolecule itself, in typical simulations[2]. Typically one to a few layers of molecules are modeled around the biomolecule and beyond that bulk water is described with a continuum model. The water molecules are modeled as a set of points, with fixed distances between each, and interact with one another and with the biomolecule by a set of energy functions[7], [8]. Among other, simplified, models of water are treating water as two dimensional disks[9] and the Mercedes-Benz model[10]. Simple models, such as the Ising model in magnetism, assist fundamental understanding without employing complex or computationally expensive mathematics.

The first step in many models of liquid water for use with biomolecules is to model pure water and reproduce physical properties of liquid water[7], [8]. For the purposes of this work, the density and specific heat of liquid water over a broad range of temperatures and pressures is employed.

Instead of using continuous energy functions, a new model where the bond between adjacent molecules can be classified in one of a discrete set of states is presented. The character of each bond is given by an average length, energy and entropy. Further model parameters employed give rise to long range interactions: (A) a strain energy parameter which adds a

small energy term if two parallel bonds across a single unit cell are of dissimilar length and (B) an extra bond entropy term which adds extra entropy to atoms which more then one high-entropy bond adjoining them.

Specific heat at constant pressure, c_p, and density, ρ, are computed by standard statistical methods[11] from the partition function. An average value, $\overline{x}$, such as the average energy $\overline{E}$, required above, or the average bond length from which the density can be computed, is found via a sum over all states i as follows:

$$\overline{x} = \frac{\Sigma x_i \exp\left(-G_i/k_B T\right)}{\Sigma \exp\left(-G_i/k_B T\right)}, \qquad (1)$$

where k_B is Boltzmann's constant and T is the temperature. The Gibbs free energy, G_i, is detailed below.

To compute the specific heat at constant pressure, the average energy of the system is found via Eq. 1, above, and the appropriate derivitive is taken at fixed pressure as follows:

$$c_p = \left.\frac{\partial \overline{E}}{\partial T}\right|_p. \qquad (2)$$

However, the complete sum over all states of the system is daunting for even modest size systems: the number of states is 3^N. Most of these states are highly unfavorable. Further, most are very similar to a very large number of other states. For small systems, the exact results of sums over all states is compared with approximations of summing over all types of states, as detailed below and checked for consistency.

Having three bond lengths implies the oxygen to oxygen distance distribution would be a collection of delta functions. However, these should be regarded as the centroids of gaussian-like distributions of possible bond lengths. To illustrate this, using the Heisenberg uncertainty relation with the equipartition of energy, it is possible to estimate a lower limit on the size of the width of such gaussians, in angstroms, as a function of temperature. At room temperature, 300 Kelvin, this is about 0.05 angstrom, which is about six times smaller then the experimental width of the nearest neighbor peak of the oxygen-oxygen distribution function[12], [13]. The details of this calculation are reproduced in appendix A, below.

2. Contributions to the Free Energy

The Gibbs Free Energy of some state i of the system is given by, $G_i = E_i + PV_i - TS_i$, where E is the energy, P

the pressure, T the temperature, V the volume, and S the entropy. The energy, volume and entropy of the states are adjustable model parameters of each type of bond modeled, thus for three states nine parameters are required. However the zero of both energy and entropy are arbitrary, reducing the number of parameters to seven.

The three states are referred to as short, medium and long in this article, reflecting the rank ordering of their bond length parameters. The short bond has the lowest entropy and energy and the long bond has the highest. Thus the low temperature anomalies of water[14] are not studied currently within this model, since that would require one or more extra states which violate these rules. The goal of this work is to show that most of the temperature and pressure phase space can be modeled well with a limited model.

In addition to straightforward local statistical mechanics, in which only energy and entropy are required, long range interactions are invoked to model strain energy and the localization of bonds by neighbors. These are termed nonadditive interactions as they are extra terms on top of the regular summation of energy and entropies. The two terms lead to quite different effects.

The first is the strain energy for parallel mismatched bonds. A small energy penalty is added to the state energy E_i, for each mismatched pair of bonds.

To give all bonds a physical location, all bonds are placed in an idealized hexagonal lattice and neighboring bonds, and the six nearest parallel bonds, are identified for each. This exercise was carried out by hand for 128 atoms in a 4x4x8 lattice and used to construct computer code which identifies neighboring bonds and parallel bonds for lattices of arbitrary size. Periodic boundary conditions are employed.

The second term is the extra bond entropy. A small entropy term is added when two or more non-ground state bonds meet at an oxygen atom. Six parameters are employed to account for two, three or four of either the medium or long bonds meeting at an oxygen. What about mixed states? A long bond can be treated as a medium bond if it will increase the extra bond entropy. For example, if one short, one medium and two long bonds meet at an oxygen, the larger of the extra bond entropy due to two long or three medium bonds will be used. In future work, a single parameter controlling the strength of these interactions will be employed from which all possible combinations of bond entropies will be derived. There is no unique way to determine such a parameter and many possibilities are being considered. Further, due to the tiny magnitudes of some of these parameters, they may be dropped altogether and only four, or perhaps three or four, higher entropy bonds meeting at an oxygen will warrent consideration for extra bond entropy. However, this is of great interest when increasing the number of states, or generalizing this model in any way. For the moment, only rank ordering of the six parameters is enforced such that: the extra bond entropy must increase

as the number of non-ground state bonds increases and (b) the extra bond entropy is greater for an equal number of long bonds over medium bonds. The total number of free parameters is 14, although as will be seen below, some of these parameters are quite tiny - four orders of magnitude smaller then the bond entropies.

3. Scaling the System

Even for the 128 atom system described above (256 bonds) the number of terms in the partition function is $3^{256} \approx 10^{122}$. Since this would take longer then the age of the universe to compute on all CPUs in existence, some kind of approximation must be made. First a very small system was created, a six atom oxygen ring with 18 bonds for which the partition function could be computed exactly and compared with various methods of sampling. The method arrived at is to sample each combination of numbers of short, medium and long bonds some large number of times, weighing each triplet of numbers (e.g. the number of short, medium, long bonds) by the appropriate multinomial coefficient. This is checked to ensure the two results are consistent and then the sampling algorithm is scaled to the larger systems. All possible triplets of number of types of bonds are sampled, although for large enough systems, sampling over various fractions of each could replace the numbers (1% short bonds, 3% medium bonds and 96% long bonds would be a sampled state instead of, say, 2 short 4 medium and 250 long bonds, for example).

4. Results

Three different pressures are considered, 0.013 MPa, 400 MPa and 1000 MPa. The lowest and highest pressures available from a thorough set of experiments[1] were employed and the middle number was chosen such that the density of the 400 MPa data should fall about half way between the two extremes. With only the three states considered and the fourteen adjustable parameters, excellent agreement to either density or specific heat and adequate fits to the other physical property are possible. In the data below, the closer fit is to density, as seen in Fig. 2 below. The excellent agreement with density across temperature and pressure is superior to the results of the TIP-5p model[7]. The TIP-5p model, along with the SPC/E[8] model are the most commonly employed models for molecular dynamics[7].

The effect of introducing the extra bond entropy and the strain is to improve the goodness of fit (chi-squared) by a factor of 3.5 - despite many of the extra bond entropy parameters being very small, see Table 1 below.

5. Conclusions

A simple three state model of liquid water yields excellent agreement with experimental density measurements, and adequate agreement with experimental specific heat data
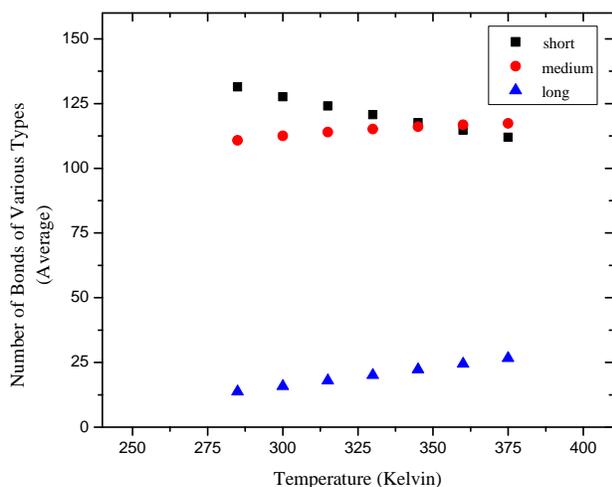
Fig. 1: Number of long, medium and short bonds at medium (400 MPa) pressure as a function of temperature. Note the crossover from a plurality of short to medium bonds at a temperature near 350 K. At higher pressure, the short bond state is more highly favored, with both the medium and long bonds suppressed. At low pressure, the longer bond is significantly more favored at the expense of the short bond, with about equal propensity for medium length bonds.
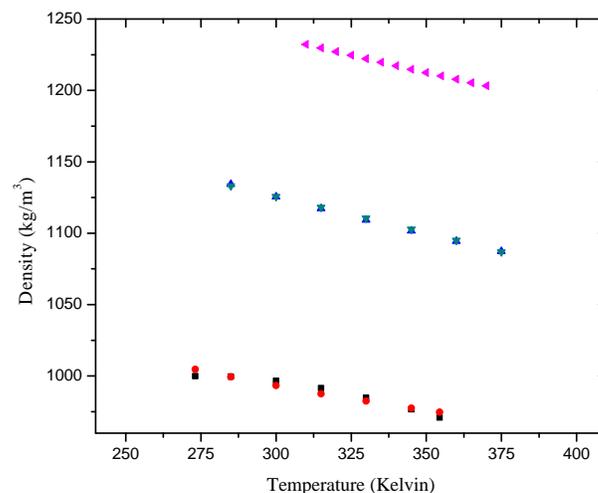

Fig. 2: Density in units of kilograms per cubic meter versus temperature in kelvin for liquid water at three pressures: low (0.013 MPa), medium (400 MPa) and high (1000 MPa). The higher pressure result in correspondingly higher densities. At each pressure, the density is nearly linear in temperature with a small negative slope. The calculated values are represented by squares (at low pressure), upright triangles (at medium pressure) and left pointing triangles (at high pressure). Agreement is within the size of the symbols at high and medium pressure and only slight differences exist at the lowest pressure, the largest of which is an 0.5% difference at the lowest temperature.

over a broad range of temperatures (273-373 kelvin) and pressures (0.013 MPa - 1000 MPa). The model employs two novel nonadditive terms: a (generally small) extra entropy term added at the intersection of multiple high-entropy bonds and a strain energy term for mismatched parallel bonds. Despite some terms being small, these nonadditive terms improve the goodness of fit by a factor of three and a half.

6. Future Work

A study of physical properties of small molecules in water shall be employed to determine similar parameters for a small set (as small as possible) between water and various atoms of biological relevance. Such a set of parameters would then be employed to model the water around and between proteins for the purposes of protein folding, drug design and docking. By replacing traditional means, either the run-time of simulations can be greatly reduced or the quantity of water modeled around proteins greatly increased.

7. Acknowledgements

The author would like to thank Brian W. Gilreath for assistance visualizing the structure of the tetragonal structure of ice and many useful discussions.

8. Appendix A

The Heisenberg uncertainty relation is that the product of the uncertainties in position and momentum must exceed half the rationalized Plank's constant. Applying this to the axis of the bond, for convenience labeled the x-axis, gives:

$$\Delta x \Delta p_x > \hbar/2. \tag{3}$$

The x-momentum is the product of mass times velocity (since any velocity here is far below the speed of light thus relativistic effects are negligible). All that is needed is a relation between momentum and position to give a lower bound on the uncertainty in position. There is a relation in energy, but to employ it, a specific form of potential (binding) energy is required. For this purpose, an approximation is introduced: that the bond acts like a single harmonic oscillator and thus the potential energy is given by $\frac{1}{2}kx^2$, where k is the spring constant and x is the displacement from equilibrium position. More complex relations such as the Lennard-Jones (6-12) potential can be considered, but these are well approximated by a harmonic oscillator when the energy is far below the dissociation
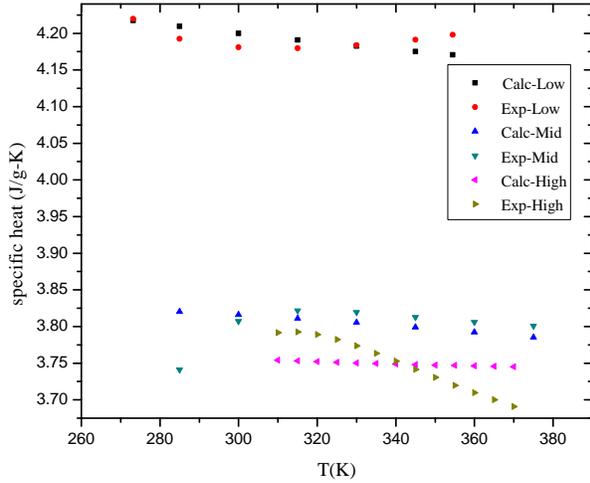
Fig. 3: Specific heat at constant pressure of liquid water in units of joules per gram-kelvin versus temperature in kelvin for three different pressures: low (0.013 MPa), medium (400 MPa) and high (1000 MPa). With only three states, the calculated data reflects the correct magnitudes in specific heat but is unable to account for the more abrupt changes at low temperature.

energy, which should be the case in question (liquid water at room temperature).

The equipartition of energy theorem requires each the degree of freedom to have the same energy as given below:

$$\frac{1}{2}k_B T = \frac{p_x^2}{2m} = \frac{1}{2}kx^2. \tag{4}$$

Combining these relations by setting $\Delta x = x$ and $\Delta p_x = p_x$, we find:

$$x > \frac{\hbar}{2(mk_B T)^{0.5}} \tag{5}$$

and

$$k^2 < k_B T/\hbar. \tag{6}$$

Plugging in known values at room temperature, the minimum value of x is about 0.05 angstroms, which is about six times smaller then the experimental width from the radial distribution function. It is worth noting that for a single, isolated harmonic oscillator, the wavefunctions can be found exactly along with the uncertainties in both position and momentum. In that case, equality in the Heisenberg relation holds, meaning the wavefunctions are the "tightest" possible. It is not surprising that a fluid cannot be modeled as a collection of isolated single harmonic oscillators.

Table 1: Model parameters for short, medium and long bonds of liquid water. All enthalpies, ΔH, have units of kcal/mol and all entropies are dimensionless "pure" entropies (to produce entropies in the proper units, these values need only be multiplied by Boltzmann's constant in the appropriate units). The bond lengths, $< x >$, are in angstroms and the extra bond entropies xbs, are also unitless. The energy, entropy of the short bond is set to zero without loss of generality[11]. The subscript of the extra bond entropy parameter denotes the number of such bonds meeting at a particular oxygen atom. The strain energy is 2.22×10^{-4} kcal/mol-K. Although fourteen free parameters are employed, three of the six extra bond entropy parameters are very small - three or four orders of magnitude less then the typical change in entropy from state to state of a single bond.

Parameter	long	medium	short
ΔH	3.25e-3	6.21e-4	0
ΔS	4.47e-3	2.89e-3	0
$< x >$	3.37	3.01	2.71
xbs_2	7.32e-7	6.83e-7	0
xbs_3	3.84e-5	5.36e-6	0
xbs_4	1.51e-4	1.83e-5	0

References

[1] Wagner and Pruβ, "The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use," *J. Chem. Ref. Data*, vol. 31, pp. 387–535, 2002.
[2] D. Qiu, P. S. Shenkin, F. P. Hollinger and W. C. Still, "The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii ", *J. Phys. Chem. A*, vol. 101, pp. 3005-3014, 1997.
[3] K. A. Dill "Additivity Principles in Biochemistry", *J. Biol. Chem.*, vol. 272, pp. 701–704, 1997.
[4] A. E. Mark, and W. F. van Gunsteren, "Decomposition of the Free Energy of a System in Terms of Specific Interactions." *J. Mol. Biol.*, vol. 240, pp. 167–176, 1994.
[5] K. A. Dill, T. M. Truskett, V. Vlachy, and B. Hrivar-Lee, "Modeling water, the hydrophobic effect, and ion solvation", *Annu. Rev. Biophys. Biomol. Struct.*, vol. 34, pp. 173–199, 2005.
[6] M. J. Tait, and F. Franks, "Water in Biological Systems", *Nature*, vol. 230, pp. 91–94, 1971.
[7] M. W. Mahoney, and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions", *J. Chem. Phys.*, vol. 112, pp. 8910–8922, May 2000.
[8] H. J. C. Berendsen, J. R. Grigera, and T. P. Staatsma, "The missing term in effective pair potentials", *J. Phys. Chem.*, vol. 91, pp. 6269–6271, 1987.
[9] T. M. Truskett, and K. A. Dill, "A Simple Statistical Mechanical Model of Water", *J. Phys. Chem.*, vol. 106, pp. 11829–11842, 2002.
[10] T. Urbic, T. Vlachy, and K. A. Dill, "Confined Water: A Mercedes-Benz Model Study", *J. Phys. Chem. B.*, vol. 110, pp. 4963–4970, Feb. 2006.
[11] F. Reif, *Fundamentals of Statistical and Thermal Physics*, 1st ed., New York, USA: McGraw Hill, 1965.
[12] A. K. Soper, and M. G. Philips, "A new determination of the structure of water at 25 C", *Chem. Phys.*, vol 107, pp. 47–60, Aug. 1986.
[13] A. K. Soper, "On the determination of the pair correlation function from liquid structure factor measurements", *Chem. Phys.*, vol. 107, pp. 61–74, Aug. 1986.
[14] C. H. Cho, S. Singh, and G. W. Robinson, "Water anomalies and the double-well Takahashi model", *Chem. Phys.*, vol. 232, pp. 329–341, June 1998.