1 Sanity check

The static structure factor is the Fourier transform of the radial distribution function $g_2(r)$.

$$S(\vec{q}) = 1 + \rho \int_V d\vec{r} e^{-i\vec{q}\cdot\vec{r}} g_2(r) \tag{1}$$

r is the distance between two water molecules and $\vec{r}$ is the vector connecting them. To compute the integral directly from MD simulation of water, I have to do the following modifications: for each frame k from the simulation, find all unique combination of pairs of molecules within distance R and sum up their contributions to the Fourier transform. R has to be chosen since the size of the simulation box is finite. Hence an estimate of $S(\vec{q})$ from on simulation frame is:

$$S(\vec{q}) \approx \frac{1}{N} \sum_{i,j,=1}^{N} e^{-i\vec{q}\cdot(\vec{r_i} - \vec{r_j})} \Delta_{ij} \tag{2}$$

where $\Delta_{ij} = 1$ if $\left|\vec{r_i} - \vec{r_j}\right| \leq R$ and is 0 otherwise.

Since liquid water is isotopic, intuition tells us that the pointing $(\vec{r_i} - \vec{r_j})$ shouldn't matter and only its length should. Debye's formula formalizes this and the above summation is equivalent to:

$$S(Q) \approx \frac{1}{N} \sum_{i,j,=1}^{N} \frac{\sin(Qr_{ij})}{Qr_{ij}} \Delta_{ij} \tag{3}$$

where $r_{ij} = \vec{r_i} - \vec{r_j}$ and Q is the magnitude of $\vec{q}$.

Averaging over many independent simulation frames gives an estimate from the simulation of $S(Q)$. There is also a finite-size effect introduced by the simulation box. See the two Salcuse papers in the notes for details. The finite-size effect is implemented in the code.

I have done some checking to make sure that the explicit Fourier transform makes sense with my simulation data. So far I have checked that it produces a small imaginary part, and also seems to be only dependent on the magnitude of $\vec{q}$ but not its pointing. Also, equations 2 and 3 give the same answer when I use them to compute $S(Q)$ from averaging over independent frames from the simulations. This makes me think that I have been doing the explicit Fourier transform correctly for the radial distribution.

2 Direct calculation of the correlator from simulation

The intensity from scattering from a pair of water molecules is:

$$I(\vec{q}, \omega) = \left| \sum_{i}^{2} f(q) e^{-i\vec{q}\cdot(R_\omega \vec{r_i})} \right|^2 \tag{4}$$

where R_ω is the rotation operator by angle ω and $f(q)$ is the atomic form factor for oxygen. I am ignoring hydrogen for now since it does not scatter x-ray strongly. If we just expand equation 4 explicitly, we have

$$I(\vec{q},\omega) = 2\left|f(q)\right|^2 (1 + \cos\vec{q} \cdot (R_\omega \vec{r_{12}})). \tag{5}$$

We want to compute, from the simulation data, the correlator between two pairs of water molecule, i.e. a single 4-point ensemble. The correlator is the following integral:

$$C(\vec{q_1},\vec{q_2}) = \int_\omega I(\vec{q_1},\omega) I(\vec{q_2},\omega) d\omega. \tag{6}$$

Assuming our water box simulation is an good approximation of the statistical behavior of real water, we can estimate $C(\vec{q_1},\vec{q_2})$ by summing over all possible 4-point ensemble (tetrahedrons) in the simulations. Specifically,

$$C(\vec{q_1},\vec{q_2}) = \frac{4\left|f(q)\right|^4}{N_{tthd}} \sum_{m=1}^{M} \sum_{i,j,k,l}^{N} (1 + \cos\vec{q_1} \cdot \vec{r_{ij,m}})(1 + \cos\vec{q_2} \cdot \vec{r_{kl,m}}) \tag{7}$$

where N_{tthd} id the total number of tetrahedrons in M simulations frames, N is the total number of water molecules in the simulation, the index m denotes the m-th frame in the simulation. We are averaging over M statistically independent frames. The summation over indices i, j, k, and l are over all unique 4-point ensembles.

The dependence on ω is removed going from equation 6 to 7 as the MD simulation takes care of sampling over all possible orientation of tetrahedrons. The assumption here is that the water model used in the MD simulation is a reasonable and we have enough statistically independent frames from the MD trajectory.

The atomic form factor for oxygen is approximately the sum of some gaussian functions. With $0 < q < 25\text{Å}^{-1}$, $f(q)$ is

$$f(q) = \sum_{i}^{4} a_i \exp\left(-b_i (\frac{q}{4\pi})^2\right) + c \tag{8}$$

and the a_i's and b_i's are constants summarized in the table below.

a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4	c
3.0485	13.2771	2.2868	5.7011	1.5463	0.3239	0.867	32.9089	0.2508

3 Method of computation

The computation is done in three steps: 1) MD simulation; 2) finding unique tetrahedrons formed by four water molecules; 3) computing C according to equation 7.

3.1 MD simulation

We use GROMACS version 5.0.2 to run the MD simulations. We start with a single water molecule, define a cubic simulation box that is 2.2 nm on each side, and then use the build-in solvation function in GROMACS to fill the box with water molecules. The automatic solvation results in 339 water molecules in the box. We use an AMBER force field (amber99sb-ildn) and a TIP3P model for water molecules. Periodic boundary condition is also applied. Specific simulation parameters are saved in nvt-pr-md_RUNNAME.mdp files in /home/shenglan/MD_simulations/water_box on zauber. Every 1 ps in simulation time, the coordinates of the water molecules are recorded. The trajectories ranges from 100 frames to 10000 frames (100 ps to 10 nm simulation times).

3.2 Defining tetrahedrons

From the MD simulations, we can build a set water tetrahedrons. The position of a water molecule is defined only by the oxygen atom since the hydrogen atoms scatter x-ray much more weakly. In every simulation frame, we can find the three nearest neighbors for every water molecules and form one tetrahedron this way. Using only the nearest neighbors, which probably contribute most to C than neighbors further away, we can reduce the the number of tetrahedrons we need to sample and thus computation time. There are approximately as many unique[1] nearest-neighbor tetrahedrons as water molecules per simulation frame. For instance, 1000 statistically independent simulation frames generate about 330k tetrahedrons.

If the simulation is true the actual physics of liquid water, this set of water tetrahedron is a representation of the true distribution of tetrahedron geometries that are possible for water. Its underlying statistics are therefore a good estimate of those of the true distribution. One way to check is to look at the radial distribution function, i.e. the distribution of distances between pairs of water molecules (oxygens); it has a well-known shpae. Figure 3.2 shows the radial function produced by looking at as simulation of 1000 frames and 339 water molecules in $2.2 \times 2.2 \times 2.2$ nm box. The shape looks reasonable with a peak at about 0.3 nm. It will be interesting to compare radial distribution function produced by systems of different sizes since my understanding is that the periodic boundary condition has a larger effect on smaller systems.

Look into the some statistics of the tetrahedrons (e.g. length of sides) and write about them here.

3.3 Computing the correlator

The sampled tetrahedrons coordinates provide r_{ij} and r_{kl} in equation 7. We then need to choose $\vec{q_1}$ and $\vec{q_2}$. Specifically, we can fix $\vec{q_1}$ and then define

[1] The combination of four water molecules that make up the vertices of the tetrahedron is unique. Some simulation frames yield fewer than 339 tetrahedrons because two molecules that are nearest neighbors of each other happen to share the two other nearest neighbors.

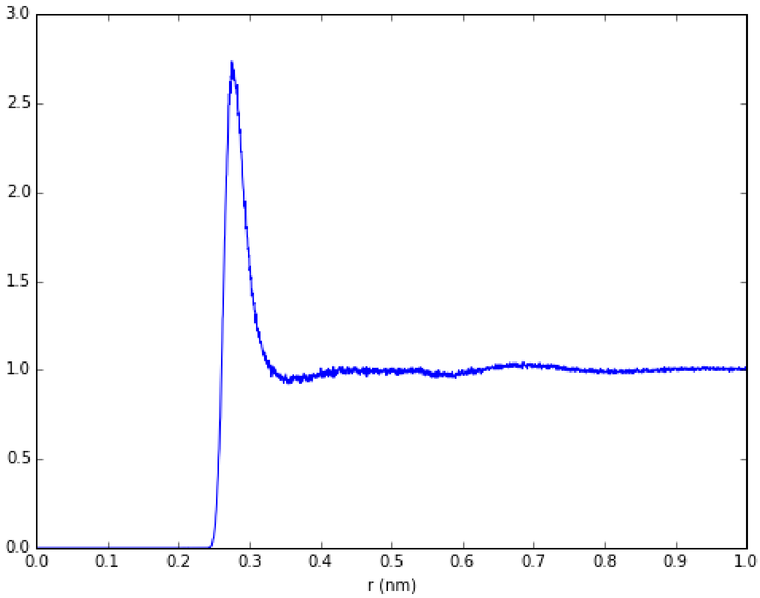

Figure 1: Radial distribution function of simulated water.

$\vec{q_2}(|q_2|, \phi)$, where ϕ is the angle between the two vectors on the detector plane. For autocorrelation, the magnitude of both vectors is $|q|$. For $\vec{q} = (q_x, q_y, q_z)$,

$$q_x = |q| \sqrt{1 - (\frac{|q|}{2q_{beam}})^2} \cos(\phi) \tag{9}$$

$$q_y = |q| \sqrt{1 - (\frac{|q|}{2q_{beam}})^2} \sin(\phi) \tag{10}$$

$$q_z = -\frac{|q|^2}{2q_{beam}} \tag{11}$$

In the case of the autocorrelator, we will fix ϕ for $\vec{q_1}$ at zero and generate $\vec{q_2}(\phi)$ $(\phi \in [0, \pi])$. In the case of different magnitudes of $\vec{q_1}$ and $\vec{q_2}, |q|$ in the equations above are replaced by $|q_1|$ and $|q_2|$.