1 Introduction

Theory

Issues related to supercritical water are based on industrial interest in using supercritical water as a medium to dissolve waste products. Solvation properties of supercritical water is contrary to those in ambient water, e.g. hydrophobic particles repulse each other and ions are attracted to each other. Our interest lies in how the underlying structure of water changes from ambient to supercritical conditions, especially its hydrogen bonding properties.

Liquid water is characterized by a tetrahedral network of hydrogen bonds. Neutron diffraction experiments show that under supercritical conditions the hydrogen bonding network is disrupted. In this homework you will use computer simulations to study this phenomenon. As a preliminary step you should read the experimental paper “Site-site pair correlation functions of water from 25 to 400°C: Revised analysis of new and old diffraction data”, A.K. Soper, F. Bruni, and M. A. Ricci, Journal of Chemical Physics, 106, Page 247-254, 1997.

You will perform simulations of a small box of water with a classical force field, calculate various $g(r)$’s and compare those results to experiment and other theoretical calculations.

2 Simulation Setup

Building the Water box

Open Maestro.

/opt/schrodinger/suite2015-2/maestro &

With an empty workspace, pull up the Build window using the Edit pull-down menu. In the Build menu, select Organic in the Fragments menu. The “H–H” should be highlighted in blue. If not, select “H–H”. Move the mouse to the center of the viewing window and left click. A line will appear representing a hydrogen molecule. This hydrogen molecule will be the center around which the water box will be placed. Close the Build window.

Locate the Applications menu at the top of the window. If you do not see the Applications menu, enable it by selecting: Tasks → Application View. In the Applications pull-down
menu, go to the Impact submenu and select Soak. A Soak window will appear. Use the default box dimension which are 18.65Å in the XYZ directions.

Open the Job Settings window by clicking the gear icon beside the Run button. A Soak window will appear. Ensure that the Host is localhost. Name the job “impact_soak_1”. Select Run. When finished, a bunch of water molecules will appear around our hydrogen molecule (zoom out or press the Recenter button).

IMPORTANT! BEFORE YOU PROCEED: You must delete the hydrogen molecule. Under the delete button on the left, select delete molecules and click on an atom of the hydrogen molecule (atoms are at the end of the line). Click on the delete button to turn it off. Click the recenter button and view the entire water box. Close the Soak window.

We can view the progress of our project with the Project Table by going to the Project drop-down menu and selecting Show Table (or typing CMD + T). Right now there is not a lot to see in the table. There should be one entry. To see additional information about the water box, click Show... → All from the row of icons at the top of the Project Table window. The main items you will need in a later step are the dimensions of the water box:

PDB~CRYST1~a  PDB~CRYST1~b  PDB~CRYST1~c

Currently they should all be 18.650. The box should roughly look like Fig 2.1. But you can change the representation of the water molecules by selecting under Workspace the option Molecular Representation. In the Molecular Representation window, you can set all to CPK, Tube, or Ball & Stick. The default is Wire. If you do change the representation, you must click the All button to apply the changes to the workspace.

Energy Minimization of the Water Box

Before performing any dynamics, we want to energy minimize the water box to take care of any overlaps that may occur in the soak process. Under Applications and the Impact submenu, select Minimization... to open the Impact Energy Minimization window. We want Use structures from: to be Workspace (included entry). There are three tabs: Potential, Constraints, and Minimization.

1. In the Potential tab, click on the Use periodic boundary conditions. Click on Settings... to the right to see the dimensions of the water box (18.650Å × 18.650Å × 18.650Å as shown in the Project Table).

2. Go to the Minimization tab. The defaults should be fine. You will need less than 100 cycles to relax any problems in the initial box.

3. Name the job impact mini.

4. Open the Job Settings window by clicking the gear icon beside the Run button. Make sure the Host is set to localhost. Click Run.

This will take about 10 seconds. Close Impact Energy Minimization window. When finished, notice that there is a new entry in the Project Table and that the box size has not changed.
3. LAUNCH SIMULATIONS

Figure 2.1: A completed water box

3 Launch Simulations

Thermalization of the Water Box to 298K

Minimization tends to “freeze” molecules. We need to warm up the water box. We will perform a 1ps MD simulation to thermalize the waters. Go to Applications → Impact → Dynamics....

1. In the Impact Dynamics window, Use structure from: Workspace.

2. Under the Potential tab, make sure Use periodic boundary conditions is clicked on.

3. Under the MD Parameters, the Stop overall motion is clicked on, but Record trajectory is clicked off. We will record the trajectory later, but not for the thermalization. Set Frequency of printing information: to 100 to save disc space.

4. The Dynamics tab contains the bulk of your settings. Set Number of MD steps to 1000 (with a time step of 0.001ps, 1000 steps will be 1ps total). The Ensemble type is Constant temperature (NVT). Set the Target temperature to 298.15 K and the Temperature relaxation time (ps): to 0.20. Set the Initial temperature (K) to 100 to reflect that we are starting with a cooled box of water. In 1000 steps, the system will heat from 100K to 298.15K.
5. Open the Job Settings window by clicking the gear icon beside the Run button. Run this job locally by setting the Host to localhost. Name the job “impact_dyn_therml” (we will refer back to this job later).

It will take under 1 minute.

**Production Run at 298K**

After completing the thermalization, you will see a new entry in the project table. Because we ran at constant volume, the box dimensions have not changed. Return to the Impact Dynamics window. Set Use structures from: to Workspace. Check to be sure in the Potential tab that periodic boundary conditions are still turned on and the box dimensions are still 18.650³.

Go to the MD Parameters tab and set frequency of printing information to 500 or 1000. Turn on Record trajectory. Change Frames written every to 1000 and turn off Sample velocities.

Go to the Dynamics tab. Set the number of MD steps to 100000 (which will give us 100ps of trajectory and 100 frames in the trajectory file). Use the NVT ensemble and set the target and initial temperature to 298.15. Set the temperature relaxation time to 0.2. Start. Name it “impact_dyn_298”. Submit this job locally by setting Host to localhost. It will take about 30 minutes. To monitor the progress of the job, you can select the button with a circle on it in the Impact Dynamics window next to the Job Settings gear icon.

Once the job is complete, there will be a new entry with a blue “T” button in the Aux column of the Project Table. Click on the T to open the Trajectory window. At the bottom, click on the Structure... button to open the Export Structure window. Export to file “impact_dyn_298_traj.maegz” with Structures: set to Selected frames. This file will be used to calculated the radial distribution function.

If running short on remaining class time, this is a good place to stop and resume from during the next class period.

To view the trajectory, return to the Trajectory window and press the Structure... button. Instead of exporting to a file, Export to the Project Table. The frames will be listed in the Project Table. In the second row of the project table are some round blue buttons to play through the frames in the Project Table. Press the right-pointing triangle to play forward through the frames.

**IMPORTANT:** Before leaving, write down the number of molecules in the simulation (that is listed on the row immediately below the viewing window under Mol:) and the box dimensions. These will be used in the radial distribution calculation.

**Thermalizing Supercritical Water**

We want a water box at 673K with a density of 0.66g/ml. We will start where the last simulation ended. In the Project Table, click on the last frame (Row 204 or so) to select that set of water molecules. When you click on that row, only that row should be highlighted in yellow. Go to the Impact Dynamics window (Applications → Impact → Dynamics...).

Set Use structures from: to Workspace (since the last structure of the trajectory is in the workspace) or to Project Table (selected entry) since you have selected an entry in the Project Table in the previous step.
3. LAUNCH SIMULATIONS

Make sure in Potential that periodic boundary conditions are selected and the box size is 18.650 cubed in the periodic boundary conditions Settings... window.

In MD Parameters, turn off record trajectory and set frequency of printing to 50 instead of 500.

In Dynamics,

1. Set number of MD steps to 5000.

2. Ensemble type should be constant pressure (NPT) so that we can change the volume of the box (and change the density).

3. The target temperature (K) is 673.

4. Set temperature relaxation time to 0.2.

5. The target pressure is something you will have to play with. Start with 800 atm (the pressure for supercritical water at this temperature).

6. Set the volume relaxation time to 0.2.

7. For our purposes, the solvent isothermal compressibility can be left at the default of $4.96 \times 10^{-5} \text{atm}^{-1}$.

8. The initial temperature is 298.15.

Run the simulation for 5000 steps (1.5 minutes to complete). Monitor the progress in the Monitor window. Look at the rows of numbers below every Step number; the last number in the second row is the density. Our target is 0.66g/ml. If the final density reported is within 5% of our target value, continue onto the next section.

Otherwise, we will rerun this short simulation again until the reported density is within 5% of the target. In order to restart the simulation, we first need to remove the output of the last one. In the Project Table, right click on the current (last) entry and select Delete and select yes when prompted to confirm. Next, we need to select a new starting point to thermalize the supercritical water. Select either the entry resulting from the 298K thermalization (‘impact_dyn_therm1’) or the last frame of the 298K production run. You have selected an entry properly when the box to the left of the entry name is shaded and the row is highlighted in yellow.

Lastly, vary the density. If your last density was too high, reduce the pressure this run; if the density was too low, increase the pressure. Increase or decrease the pressure in 100 to 200 atm increments. When you’re done, make sure to note the pressure that gives roughly the correct density.

Production Run of Supercritical Water

Starting with the thermalization run with the right density, we will run a 100ps MD trajectory for calculating the radial distribution function. Open the Impact Dynamics window (if not already open). Select Use structures from to Workspace. Go to the Potential tab. Make sure use periodic boundary conditions are turned on. In Settings..., the box lengths have to be set to those reported in the Project Table. Scroll across the last, highlighted row to the values under PDB CRYST1 a, PDB CRYST1 b, and PDB CRYST1 c (you may need to
select the Show...All button on the Project Table Window for these to display). These are the final bo... the Periodic Boundary Conditions window. If you do not do this, your simulation will crash.

Move to the MD Parameters tab. Set the frequency of printing to 500, turn on record trajectory, and have frames written every 1000 steps and turn off Sample velocities.

Move to the Dynamics tab. Set the number of MD steps to 100000. Set the ensemble type to Constant temperature (NVT). Set both the target and initial temperatures to 673K. The temperature relaxation time should be 0.2. Start the run, naming it “impact_dyn_673”. This will take roughly 30 minutes to complete.

Use the same procedure described above for the run at 298 to save the trajectory files into “impact_dyn_673_traj.maegz” and to view the final results.

IMPORTANT: Before leaving, write down the number of molecules in the simulation (that is listed on the row immediately below the viewing window under Mol:) and the box dimensions. These will be used in the radial distribution calculation.

4 Simulation Post-Processing

Setting Up $g(r)$ Calculation

Once you have finished the simulations in Maestro, we will do some analysis on the commandline. First exit Maestro. Next, there are several scripts and source data you must copy to your working directory from online:

```
curl -O https://ronlevygroup.cst.temple.edu/courses/2015_fall/chem5302/water_lab.tgz
tar -xzf water_lab.tgz
```

Additionally, for the scripts to function properly, enter the following commands into the terminal:

```
echo "SCHRODINGER=/opt/schrodinger/suites2015-2/" >> ~/.bashrc
source ~/.bashrc
```

The first command makes it so that you will always have the proper environment variables setup in all future terminals. The second line sets up the variables for the current terminal session.

The four scripts needed are:

- `rdf.py`
- `gr.script`
- `getpressure.awk`
- `awksd`

and some experimental data needed later is located in h2odata. To copy and ensure proper permissions on the scripts, in your working directory, type

```
cp -R source_files/{rdf.py,gr.script,getpressure.awk,awksd,h2odata} .
chmod 755 rdf.py gr.script getpressure.awk awksd
```
The `rdf.py` is a Python script that is executed by IMPACT to read in the trajectory files

- `impact_dyn_298_traj.maegz`
- `impact_dyn_673_traj.maegz`

and calculate the radial distribution function, \( g(r) \), and the number of nearest neighbors within a particular radius, \( \int_0^r dr' 4\pi r'^2 g_{OO}(r') \), based on a series of flags. `gr.script` calls `rdf.py` and gives details about how those flags should be set. Here is a brief description of the flags that are used in `gr.script`:

- **-a <ASL selection>**: this is the atom designation of the first set of atoms used in the \( g(r) \) calculation. Atoms are O for oxygen and H for hydrogen. This is pre-set in `gr.script`.

- **-b <ASL selection>**: this is the atom designation of the second set of atoms used in the \( g(r) \) calculation. If not specified, -b is automatically set to -a. This is pre-set in `gr.script`.

- **-d <dens>**: this is the number density of the second set of atoms (under the -b flag). The number density is the number of atoms divided by the volume of the box (in Å\(^3\)). If the number of molecules is 218, the number of O’s is 218 and the number of H’s is 436. You will need to calculate this and change it in `gr.script`.

- **-x, -y, -z**: these are the dimensions of the box you saved earlier. Remember that the box sizes are different for `impact_dyn_298_traj.maegz` and `impact_dyn_673_traj.maegz`.

Make the following changes in `gr.script`:

1. Change the O and H density for the 298K and 673K calculations.
2. Ensure the dimensions of the box are correct for 298K calculations.
3. Change the dimensions of the box for 673K calculations.

### Running `gr.script`

On your workstation, you will submit `gr.script` locally. At the command line enter `. /gr.script` then press enter. When the job is finished, you will have 6 output files: `g_r_*`. They contain the following information:

- **g_r_O-O_298.out**: \( r, g_{OO}(r) \) and \( \int_0^r dr' 4\pi r'^2 g_{OO}(r') \) at 298K
- **g_r_O-H_298.out**: \( r, g_{OH}(r) \) and \( \int_0^r dr' 4\pi r'^2 g_{OH}(r') \) at 298K
- **g_r_H-H_298.out**: \( r, g_{HH}(r) \) and \( \int_0^r dr' 4\pi r'^2 g_{HH}(r') \) at 298K
- **g_r_O-O_673.out**: \( r, g_{OO}(r) \) and \( \int_0^r dr' 4\pi r'^2 g_{OO}(r') \) at 673K
- **g_r_O-H_673.out**: \( r, g_{OH}(r) \) and \( \int_0^r dr' 4\pi r'^2 g_{OH}(r') \) at 673K
- **g_r_H-H_673.out**: \( r, g_{HH}(r) \) and \( \int_0^r dr' 4\pi r'^2 g_{HH}(r') \) at 673K

Make plots of \( g(r) \) and \( \int_0^r dr' 4\pi r'^2 g(r') \) versus \( r \) for O-O, O-H, and H-H at the two temperatures.
5 Analysis

- Use the output of IMPACT to estimate the average pressure and its statistical uncertainty at the two state points.

**Question 5.1:** How does the experimental pressure – 1 atm at ambient conditions and 800 atm at the supercritical conditions – compare to the average calculated pressures? How do the computed values fluctuate?

The scripts `getpressure.awk` and `awkstd` can be used to extract the pressure from the output files ('`impact_dyn_298.out`' and `impact_dyn_673.out`). At the command line, type:

```
./getpressure.awk impact_dyn_298.out > 298.pressure
./getpressure.awk impact_dyn_673.out > 673.pressure
```

The files `*.pressure` will have the instantaneous pressure. Use `awkstd` to get the average and standard deviation.

```
./awkstd 298.pressure
./awkstd 673.pressure
```

- There are experimental results at ambient conditions located in the `h2odata` directory.

**Question 5.2:** Compare in a plot calculated \(g_{OO}(r)\) and \(g_{OH}(r)\) with these experimental results. Discuss any differences in qualitative and quantitative terms.

**Question 5.3:** Repeat the analysis for your results at the supercritical conditions.

Read the README file to locate the appropriate files for your simulation condition.

- From the plots of the radial distribution functions, \(g(r)\), and the number of nearest neighbors, \(\int_0^\sigma dr' 4\rho\pi r'^2 g(r')\), obtain the number of neighbors for each atom as a function of distance at ambient conditions.

**Question 5.4:** Give an estimate of the number of nearest neighbor from the oxygen-oxygen radial distribution function by only integrating up to the first minimum of the distribution function. How does this compare with an ideal tetrahedral coordination of 4?

**Question 5.5:** Repeat the analysis for your results at the supercritical conditions.

- In simple liquids the second peak of \(g(r)\) is located at \(2\sigma\).
5. ANALYSIS

**Question 5.6:** If we assign an effective $\sigma$ of water to coincide with the first peak of $g_{OO}(r)$, where should the second maximum occur (in Å)? Comment on the observed difference between water and a simple liquid. How does the calculated second peak compare to the measured one?

**Question 5.7:** Repeat the analysis for your results at the supercritical conditions.

- **Question 5.8:** How many hydrogen are there within the first peak of $g_{OH}(r)$, as measured up to the first minimum?

An approximate number of hydrogen bonds a water molecule is engaged in with its nearest neighbors can be calculated by multiplying the number of hydrogens you just calculated by 2. (This takes care of the normalization, but ignores the finer points of defining hydrogen-bonds as discussed in the Soper et al. article.)

**Question 5.9:** How does the number of hydrogen bonds calculated from your simulation compare with the same number calculated by performing the same analysis on the experimental data?

**Question 5.10:** Repeat the analysis for your results at the supercritical conditions.

- **Question 5.11:** What can you conclude on the local structure of supercritical water compared to the local structure of liquid water? Is hydrogen bonding still important in supercritical water?

- The large qualitative changes observed in the experiment are not reproduced too well by the simulations. Either the experiments are wrong or the calculations are wrong.

**Question 5.12:** Comment on the changes from ambient to supercritical conditions. Discuss possible sources of error in the theoretical treatment of the system. Can you suggest a better set of simulations?

- **Optional.** In the data directory h2odata there are two further sets of simulation data at supercritical conditions. The nemo.*.data files contain data from a simulation using a classical potential but which is fully polarizable, whereas the sprik.*.data were produced using an ab initio simulation using wave-function propagation techniques. Compare the four data sets with respect to the $g_{OO}(r)$, $g_{OH}(r)$ and number of hydrogen bonds. Can you draw any conclusions?