

Condensed Matter Physics Final Project

Author: Brian Andrews

Due: 12:00 PM, Dec 22nd, 2023

This project was inspired by the "Making It Rain" project [1, 2]. You will be conducting dynamics simulations of many water molecules using so-called Molecular Dynamics (MD) software. You will be introduced to MD and will set up, run, and analyze simulations using Google Colab which provides a free graphical processing unit (GPU) to perform the simulation and calculations. You may find that simulating or modeling molecular systems is far from trivial and we, as a collective research field, are not as far along as we might like in terms of accuracy or computational efficiency!

The sections of this document correspond with sections in the Colab notebook. Follow through the sections and use the outputs of the notebook to answer the questions. The simulations and some analyses will take some time so it is highly recommended you run simulations while answering questions or have other work to do. Some questions will require you to write your own code outside of the Colab notebook. **The project requires a Google account with access to Google Drive. If you do not have one, please make one for the duration of this project.**

Colab Notebook Link:

<https://colab.research.google.com/drive/1RScebYm6gca-M03dB0Fi5QGv7N1ckGpH?usp=sharing>

Github Link: <https://github.com/andrewsb8/Water-Dynamics-MD>

Background

Why are dynamics simulations useful? Experiments performed on microscopic systems are limited in their scope. Only time-averaged, bulk properties can be directly measured but often no, or very little, information is obtained at the atomistic scale to explain the observable emergent behavior. The physics of the system at the microscopic level are usually described theoretically. However, the typical systems of interest are examples of many-body problems which cannot be treated analytically without extreme simplifications. Dynamics simulations, while are not without approximation, are powerful tools for identifying atomistic mechanisms which drive experimentally observable, emergent properties of bulk matter. The accuracy of our computational models is paramount and our simulations must be able to reproduce experimental results to verify our systems are accurately evolving over time. This project will focus on judging the accuracy of these simulations and understanding the difficulty of accurately simulating the dynamics many molecules in the condensed phase in the classical limit.

A Brief Introduction to MD In MD, each particle in the system evolves through the numerical integration of Newton's equations. After each time interval or time step, dt , the forces are calculated for the whole system. Then, the positions and velocities are updated through another time step. Many time steps are required when running these simulations.

$$x(t + dt) = x(t) + v(t)dt + \frac{1}{2}a(t)dt^2 \quad (1)$$

$$v(t + dt) = v(t) + a(t)dt \quad (2)$$

$$a(t) = F(t)/m \quad (3)$$

The time evolution of each particle depends on the forces acting on the particle. To accurately determine the forces, the potential must be defined. The most accurate potential would be defined using quantum

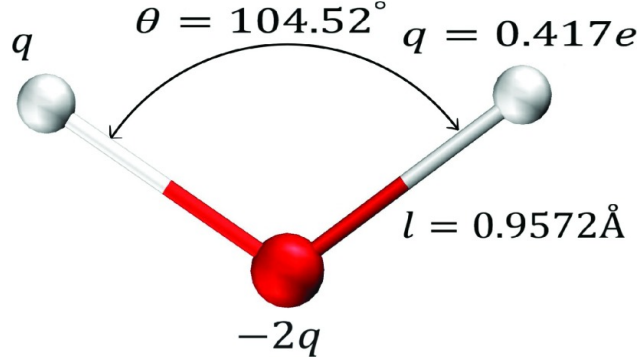


Figure 1: A water molecule with equilibrium bond lengths and bond angles as well as partial charges labeled. Taken from [3]

mechanics. In this case, we will treat the system classically to reduce computational cost and increase the number of atoms or molecules we can consider. In the case of water, we are dealing with a molecule which has covalent bonds with specific equilibrium properties and charge distribution (see Fig. 1).

The covalent bonds between atom pairs, bond angles between atom triplets, van der Waals interactions, and coulombic interactions all contribute to the potential energy of the system. To calculate over the total energy, a double sum must be conducted over all atoms i and each atom, j , with which atom i interacts. The potential for water molecules looks like the following:

$$U = \sum_{bonds} k_{bonds}(x - x_o)^2 + \sum_{angles} k_{angles}(\theta - \theta_o)^2 + \frac{1}{2} \sum_{ij} \epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) + \sum_{ij} \frac{1}{4\pi\epsilon_o} \frac{q_i q_j}{r_{ij}} \quad (4)$$

where x and θ define the length of the bond vector and angle at time t , x_o and θ_o are the equilibrium bond lengths and angles, r_{ij} are interparticle distances, and q_i and q_j are the atomic partial charges. The values for the parameters are determined typically through computational quantum mechanics experiments. The forces are then $\vec{F} = -\vec{\nabla}U$.

Periodic Boundary Conditions and the Minimum Image Convention Molecular dynamics takes advantage of periodic boundary conditions (PBC) such that water molecules can "exit" the boundary and "reenter" the cell on the opposite side. Using PBC allows us to simulate a smaller number of water molecules, lowering the computational cost, while still approximating a bulk water system. This is analogous to using a unit cell to describe an infinite crystal. If PBC was not used in simulations, the water molecules would "bounce" off the boundary of the cell and the cell sizes would have to be so large simulations would become computationally intractable. Fig. 2 can be used to understand how PBC works in 2D. The central grid of the nine, "O", is the unit cell containing six atoms. The squares around it are copies of the unit cell but are not explicitly represented in the simulation. In our 2D example, atom 4 is near the edge boundary and interacts with atoms inside the dashed box which includes 1_A , 2_B , and 6_H outside of the original unit cell. These atoms correspond to atoms 1, 2, and 6 in the central unit cell and are called images of those atoms. Therefore, interactions across the boundary are between atoms on the opposite side of the cell and this is known as the minimum image convention. By using PBC and the minimum image convention, a small unit cell of water molecules can be considered to be bulk water, or be considered within the thermodynamic limit.

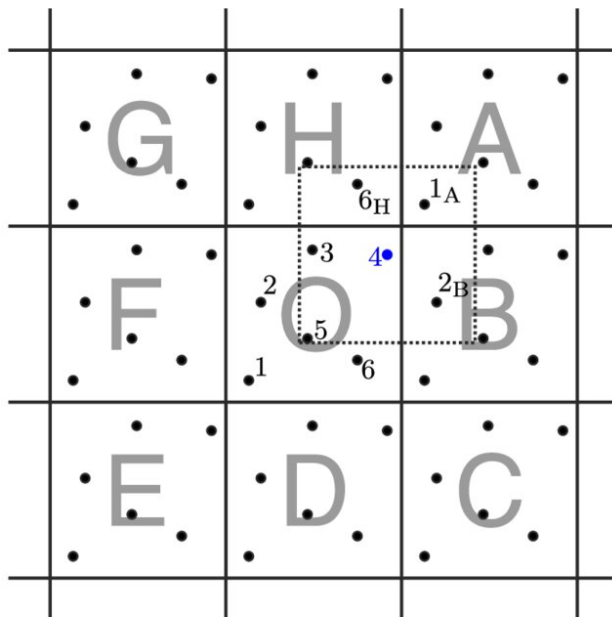


Figure 2: Illustrative schematic demonstrating the design and implementation of periodic boundary conditions for a two-dimensional system. Taken from [4].

Quick Discussion of Thermodynamics and Ensembles Our system has constant number of particles, N , and is in contact with a heat bath such that the system maintains constant average temperature. Therefore, our system is in an ensemble analogous to the canonical ensemble which you may remember from thermodynamics and statistical mechanics. However, instead of maintaining constant volume, these simulations will maintain constant average pressure to mimic experimental conditions. This means the system size can vary slightly as a function of time. This ensemble is called the NPT ensemble (isobaric-isothermal ensemble) because Number of particles, Pressure, and Temperature are held constant.

Why aren't we using quantum mechanical methods? Generally, computational methods of solving the Schrodinger equation are only feasible, even on large computing clusters, for systems on the order of ≈ 100 atoms which is not large enough to replicate bulk phenomenon. Even with 100 atoms, evolving the system dynamically is too computationally costly. Using a classical potential allows for simulating more atoms but also forces researchers to become creative to overcome problems, as will be seen in later sections of this project.

How will this project compare to published studies which use MD? Generally speaking, the systems studied in this project will be smaller than those used for research purposes to reduce the time to run the simulations and analyses. These simulations are also typically used to study larger molecules like lipids and proteins which would require larger system sizes than explored in this project.

0 Setting up the Python Environment for Molecular Dynamics Simulations

Open the Colab notebook in your browser. Click "Connect" under the "Share" button. You should see 'T4' (GPU model) and small RAM and Disk usage graphs. If you don't see 'T4', click on the dropdown next to "Connect", select 'Change runtime type', and select a T4. The cells in this section will install all of the necessary packages we need to perform the simulations and analysis and make directories in your Google Drive to organize the information. Links to the primary packages we will be using are at the top of the notebook for the curious. **You may have to execute these cells again if you lose internet**

connection, restart your computer, or exit the browser. Be prepared for this as it can be time consuming!

If you get disconnected after your initial setup, reconnect to a runtime, and execute the cells with these titles to get going again:

- Import Google Drive
- Install Conda Colab
- Install dependencies
- Load dependencies

If you finished a simulation prior to being disconnected, that data is still in your Google Drive and you do not have to repeat the simulation!! Also, if you finish all of the simulations and just have to run analyses, you can connect to a CPU runtime (instead of T4) to reduce risk of usage limits interrupting your work.

1 Preparing a Molecular Dynamics Simulation of Water

Any files output in this section will be in 'tip3p.sim' on your Google Drive. The cells in this section create a cubic box of randomly placed water molecules in a $3 \times 3 \times 3$ nm (27 nm^3) cubic box. The choice of a cubic cell is arbitrary and could have been chosen to be a number of potential 3D shapes. Then, the energy of the system is minimized using a gradient descent algorithm which modifies the relative positions of the molecules to lower the energy. Random velocities based on the Boltzmann distribution are assigned to the atoms. Other parameters, such as the temperature (300 K), pressure (1 bar), and time step (2 fs) are also defined.

1. Include a screenshot of your cell of water molecules in your submission. Report the number of water molecules in your simulation box from the output of the cell.

2. To explore the idea of how PBC can be useful, answer the following questions. If PBC is not used, do you think the shape of the unit cell affect the dynamics of the system (a cube vs a octahedron, for example)? Additionally, do you think the number of water molecules, as reported in question 1, would be sufficient to be considered bulk water in the absence of a periodic boundary? To gain intuition for the latter of the two, compare the volume of the cell defined in our simulation to that of 0.05 mL of water (0.05 cm^3 or the equivalent of about 1 drop of water).

3. (a) We choose a time step of 2 fs (2×10^{-15} s). To understand why, we can do a small computational kinematics experiment. Consider a ball that is being launched from the ground at an angle in 2D. Using the kinematics equations, you can plot y vs x for any time, t. For a ball where $m = 1 \text{ kg}$ and an initial velocity of 3 m/s launched at an 45 degree angle with respect to the horizontal, make a y vs x plot from $t = 0 \text{ s}$ to the time when the ball touches the ground and include enough data points such that the trajectory is clear. Now, you can do the same thing with Eqs. 1-3 for different values of $dt = 0.000001, 0.001, 0.1$, and 1. Plot all of these curves on the same plot and provide it, as well as the code used to generate the plot, in your submission. You can use whatever language you prefer to generate the plot (Mathematica, Python, Matlab, etc). What do you notice about your results as dt increases? (b) Based on the results of (a), why is it necessary choose a time step so small for the MD simulations? Think of the length of the box edges from the simulation.

4. Why do you think an energy minimization step is necessary? Remember that we initiated the system by randomly placing water molecules throughout the box and water molecules consist of fermions.

2 Running & Analyzing the Simulation of Water

Any files output in this section will be in 'tip3p_sim' on your Google Drive. Now that we have a minimized system, we can begin the production simulation. Execute cells in this section of the notebook. This will execute a simulation for 15,000,000 steps which corresponds to 30 ns ($2 \text{ fs} \times x \text{ steps}$). The output from the cell will include information about simulation time and real time to let you follow the progress of the simulation. Additionally, in your Google Drive, there will be a trajectory file (extension .dcd) and a log file (.csv), the latter of which will include information about the system as the simulation progresses such as instantaneous temperature and density.

5. The simulation you run will evolve the system through 30 ns with a time step of 2 fs. The third column of the output is labeled "Elapsed Time" which is the amount of "real" time (measured on our phone or wall clocks) it took to execute 30 ns of simulation time. Report the total elapsed time in seconds it took for your simulation to run in hours, minutes, and seconds. Determine then how long it would take to evolve the simulation through 1 second of simulation time. It will be a very large number and this is only for < 3000 atoms! This highlights a major issue with researching condensed phase systems with dynamics simulations, particularly in biology. It's very difficult to simulate complex systems on relevant time scales: 1 second of simulation time or longer for millions of atoms.

6. In the output, you will see a plot of temperature vs simulation time and the rolling average of temperature vs simulation time. Aside from the beginning of the simulation where the system is "warming up", you may notice fairly large fluctuations from the instantaneous temperature while the average is fairly constant at 300 K. We don't notice these types of fluctuations in real life. Explain why. Include a screenshot of the plot in your submission.

7. The next cell reports the average temperature and average kinetic energy of the simulation. Use the equipartition theorem, considering only kinetic energy and its associated degrees of freedom, to determine what you would expect the ratio of these two quantities to be. How close is the ratio of average temperature and KE output from simulation to the expected value?

8. The next cell calculates and plots the radial (or pair) distribution function, $g(r)$, for the oxygen atoms of water molecules. It also integrates over the first peak. Report the value of the integration and explain its meaning. Include a screenshot of the plot in your submission.

9. The final cell in this section calculates the average dielectric constant and average mass density. The latter is just the sum of the mass divided by the volume of the box averaged over the whole simulation. Remember that in our ensemble, the box size can fluctuate slightly. The former, ϵ , is calculated by calculating the net dipole moment $M = \sum_i q_i \vec{r}_i$ where q_i is the partial charge and \vec{r}_i is the position of particle i . The dielectric constant is then $\epsilon = 1 + \frac{\langle M^2 \rangle - \langle M \rangle^2}{3\epsilon_0 V k_B T}$. Averages are calculated over time using individual time steps or "snapshots" of the trajectory of the water molecules, similar to the Random Walks discussed in class and the final homework. Fill in values for TIP3P in Table 1 (or make your own table in your submission).

	Experiment	TIP3P	TIP4P/2005
Density [kg/m ³]	997		
Dielectric constant (ϵ)	78.4		

Table 1: Table for logging calculated properties of simulated water and facilitating comparison with experimental values.

3 Simulating a Different Representation of Water: TIP4P/2005

Any files output in this section will be in 'tip4p_sim' on your Google Drive. The water model in the previous section is called TIP3P because it is a three-point water model. However, it has been demonstrated that three-point water models may have inherent limitations in classical MD [3]. Therefore, researchers have tried adding what are known as virtual particles to water as shown in Fig. 3. Here, we will explore an example of a four-point water model, TIP4P/2005 [5], and compare its ability to reproduce the mass density and dielectric constant, ϵ , with the TIP3P water model. In the case of TIP4P/2005, the virtual particle "M" is massless, does not contribute to the van der Waals potential, but it carries the negative partial charge instead of the oxygen. This changes the dipole moment of the molecule, or the distribution of partial charges, of the molecule. Additionally, σ and ϵ , of the van der Waals potential, for the oxygen and hydrogens of TIP4P/2005 are different than for TIP3P.

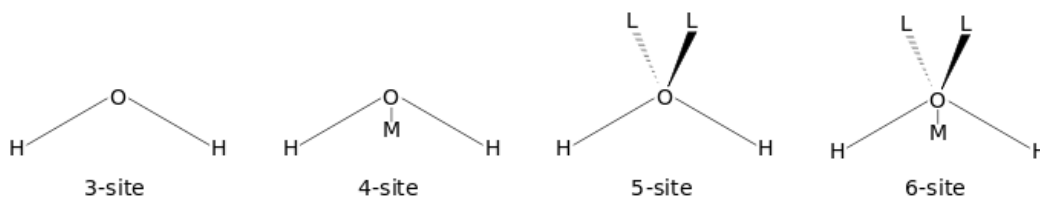


Figure 3: Schematics of different ways researchers have attempted to model water in classical simulations. In section 2, the three-point representation is used. In section 3, the four-point representation is used. Taken from Wikipedia.

10. Execute the cells in this section and report the number of water molecules in your box. Also report the results for average mass density and ϵ for the TIP4P/2005 water model in Table 1 (or to your table in your submission). Do you notice a potential relationship between the density and the dielectric constant?

11. Are the calculated values in Table 1 close to the experimental ones for either water model? Based on your results from simulating two representations of water, do you think the dielectric constant is an emergent property or an intrinsic property?

12. The general expression of Coulomb's law is $\frac{1}{4\pi\epsilon\epsilon_0} \frac{q_i q_j}{r_{ij}}$. In a vacuum, $\epsilon = 1$. Experimental measure of ϵ is 78.4 for water as shown in Table 1. Using a water model that cannot reproduce the correct dielectric constant can have consequences in simulations of solute molecules (e.g. polymers, proteins, lipids, small molecules for pharmaceuticals). You might think about a simulation where we are trying to investigate the strength of the coulombic interaction between two particles or molecules surrounded by water. If the dielectric constant of the water model in simulation was too high, what would the simulation predict about the coulombic interaction strength (or energy) relative to what you might expect from experiment? This has major consequences for simulating complex molecular phenomenon like protein folding, protein-ligand

binding, drug design, and many other applications.

4 Simulating a Phase Transition

Any files output in this section will be in 'tip4p_phase_transition' on your Google Drive. Generally, for reproducing experimentally observable characteristics of water, TIP4P/2005 is considered more accurate than TIP3P. Although you may have noticed neither produce the experimental dielectric constant very well. Here we will use TIP4P/2005 to computationally investigate the freezing phase transition of water to ice. Execute the cell in this section. The output will be two plots showing the volume of the box and the mass density as a function of temperature from 305 K to 260 K. To do this, 10 short (10 ns) simulations of the system will be conducted in 5 K intervals. In total, this simulation should take around 2 times as long as the first simulation in this project.

13. Our discussion of thermal expansion in Kittel Chapter 5 showed that $\langle x \rangle$, directly related to mass density and volume of the box, is linear or proportional to T . Describe the relationship between density and volume with temperature for water based on your plots. Why do you think water behaves this way? Include the plots in your submission.

14. Using the graphs, estimate the temperature in Kelvin where the phase transition happens, a.k.a the critical temperature, based on your data. Why did you choose this temperature? Look up and report the real transition temperature in K for comparison.

15. Identify an order parameter associated with this transition. What type of phase transition is this? Explain why you know the type of phase transition.

References

- [1] Pablo R. Arantes, Marcelo D. Poletto, Conrado Pedebos, and Rodrigo Ligabue-Braun. Making it rain: Cloud-based molecular simulations for everyone. *JOURNAL OF CHEMICAL INFORMATION AND MODELING*, 61(10):4852–4856, OCT 25 2021.
- [2] Pablo Arantes. making-it-rain. <https://github.com/pablo-arantes/making-it-rain>, 2021.
- [3] Saeed Izadi and Alexey V. Onufriev. Accuracy limit of rigid 3-point water models. *JOURNAL OF CHEMICAL PHYSICS*, 145(7), AUG 21 2016.
- [4] Franz Bamer, Firaz Ebrahim, Bernd Markert, and Benjamin Stamm. Molecular mechanics of disordered solids. *ARCHIVES OF COMPUTATIONAL METHODS IN ENGINEERING*, 30(3):2105–2180, APR 2023.
- [5] JLF Abascal and C Vega. A general purpose model for the condensed phases of water: Tip4p/2005. *JOURNAL OF CHEMICAL PHYSICS*, 123(23), DEC 15 2005.