
3.34 Molecular dynamics simulations on water

Advanced physical lab exercises: Section B
October 4, 2023



TECHNISCHE
UNIVERSITÄT
DARMSTADT

1 Introduction

In physics, a distinction is usually made between theory and experiment. However, this neglects a methodological area that has existed for a long time and is becoming increasingly important, which is different from both.

Simulations are not pure theory, since data and measured values are generated that cannot be found by analytical solutions. At the same time, they are not real experiments, since only effects can be observed that allow underlying models. These are simplified and do not fully reflect the real nature, which is why quantitative deviations often exist. On the other hand, simulations have the great advantage that they can be evaluated in more detail and experimental systems, or model systems, that are difficult to access can also be investigated. This allows to test theories in a different way.

In this experiment, you will investigate some properties of water using so-called molecular dynamics (MD) simulations and compare them with experimental data.

This explicitly means that you set up a simulation with pure water at room temperature (295 K) and then heat and cool it. This is accomplished with console inputs. A list of helpful commands is given at the end of the manual. Evaluations of the trajectories (coordinates of the particles as a function of time) are performed with a script.

2 Basics

2.1 MD simulations

The simulation method used here considers systems on the size scale of a few to a hundred nanometers [1, 2]. The motions of atoms due to their interaction with each other and under the influence of external fields are simulated. So, in general, the Schrödinger equation should be solved for the Hamiltonian operator \hat{H} of the whole system:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle \quad (1)$$

However, with today's computer power, this is only possible for very small systems and short times. In the Born-Oppenheimer approximation, which allows to consider the motion of electrons and atomic nuclei separately because of the strongly different masses, the system of equations is simplified. It is based on the fact that the electrons react fast enough to the displacement of the nuclei. From here on, the electrons are not considered explicitly and the atomic nuclei are assigned the corresponding charges, partial or total. If now simple potentials and the coordinates of the particles are used instead of Hamiltonian and wave functions, classical Newtonian equations of motion can be applied. Assuming that we have a system with N particles, then the interaction potential can be written as

$$V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N). \quad (2)$$

By derivation we obtain the forces acting on the atoms:

$$\vec{F}_i = -\vec{\nabla}_{\vec{r}_i} V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N). \quad (3)$$

Calculation of the forces already allows energy minimization of the system. Iterative shifting of the particles along their forces reduces the potential energy to a, mostly local, minimum. For dynamics Newtonian equations of motion are applied:

$$m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} = \vec{F}_i, i = 1, \dots, N \quad (4)$$

An analytical solution is not possible, so numerical means are resorted to in the simulation. Numerical integration for sufficiently small time steps then gives us the trajectory $\vec{r}_i(t)$ of each particle. In the case of the integrator used in the experiment (leap frog), this means iteratively updating the positions with the following equations:

$$\vec{v}_i \left(t + \frac{1}{2} \Delta t \right) = \vec{v}_i \left(t - \frac{1}{2} \Delta t \right) + \frac{\Delta t}{m_i} \cdot \vec{F}_i(t) \quad (5)$$

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \Delta t \cdot \vec{v}_i \left(t + \frac{1}{2} \Delta t \right) \quad (6)$$

Rounding errors and asymmetric potentials can easily cause drifts in the kinetic energy and thus in the internal energy. How these are handled is described in the next section.

The system consists of a box, preferably cubic, filled with particles. To avoid unrealistic effects at the edges of the box and to artificially increase the system size, so-called periodic boundary conditions are used. The simulation box is effectively repeated infinitely often in each direction. Particles can enter periodic images of the box and thus diffuse freely.

2.2 Temperature and pressure coupling

The simulations are to be carried out at a specific thermodynamic state (pressure, temperature and constant particle number). The corresponding thermodynamic ensemble is therefore called the NpT (isothermal-isobaric) ensemble, since particle number, pressure and temperature are kept constant. In order to achieve this, the velocities must be regularly adjusted to the respective temperature due to the previously mentioned drifts. Assuming that the distribution of velocities obeys the Maxwell-Boltzmann distribution at the corresponding temperature, the velocities can be rescaled if deviations have occurred. The algorithm used here, the velocity-rescaling thermostat [3], keeps the temperature constant via the following equation:

$$dE_{\text{kin}} = (\bar{E}_{\text{kin}} - E_{\text{kin}}) \cdot \frac{dt}{\tau} + 2 \cdot \sqrt{\frac{E_{\text{kin}} \bar{E}_{\text{kin}}}{N_{\text{df}}}} \cdot \frac{dW}{\sqrt{\tau}} \quad (7)$$

Here, \bar{E}_{kin} is the target energy, N_{df} is the number of degrees of freedom, τ is a time constant for approaching the target, and dW is stochastic Wiener noise. The first term on the right-hand side describes a simple exponential decay in the difference between current kinetic energy and the target value, while the second term provides sufficient fluctuations due to Wiener noise. Overall, therefore, temperature fluctuations are not suppressed, resulting in a correct thermodynamic ensemble.

Observables calculated from a simulation are only correct if the results for all microstates with their respective probabilities in thermodynamic equilibrium are determined and averaged for this purpose. Therefore, it is important that the algorithms used produce a correct ensemble. In order for the system to explore the phase space sufficiently long simulations must be performed. When the time average of the observables gives the same as the ensemble average, it is called ergodicity. In our simulations, we have a sample in time and must choose it long enough to get the correct values for our observables with small error. For the pressure coupling, the so-called Parrinello-Rahman barostat is used [4]. The pressure tensor is calculated via the positions, velocities and forces. It describes whether the particles push 'outward' or 'inward'. Accordingly, the pressure is positive or negative. To keep the pressure constant, the size of the simulation box is adjusted. The Parrinello-Rahman barostat modifies the equations of motion with an additional friction term. The box volume can now fluctuate and is on average according to the density belonging to the thermodynamic state. Alternatively, the very robust Berendsen barostat can be used. It provides an exponential approximation to the mean box volume but it does not produce the true thermodynamic ensemble.

2.3 Interactions of the particles

The previously described potential $V(\vec{r}_1 \dots \vec{r}_N)$ generally consists of bound and unbound interactions. Unbound interactions are:

- Coulomb potential
- van der Waals attraction
- Pauli principle repulsion

The latter are often grouped together in the Lennard-Jones (LJ) potential.

$$U_{\text{LJ}}(\vec{r}_i, \vec{r}_j) = 4 \cdot \epsilon_{ij} \cdot \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (8)$$

The Coulomb potential,

$$U_{\text{C}}(\vec{r}_i, \vec{r}_j) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \quad (9)$$

is with $\frac{1}{r}$ of longer range than the LJ potential. Both potentials are calculated explicitly for particle pairs ij up to a cutoff radius r_c . The long-range Coulomb interaction is augmented with a sum in reciprocal space, the particle-mesh Ewald summation, for longer distances and the periodic images. A faster analytical solution is used for the LJ-interaction. For bonds, as they always occur in molecules, harmonic potentials are usually used. For particles separated by two (or three) bonds, angular potentials (or torsional potentials) are used instead of unbound interactions. These potentials take better account of the actual interaction than the unbound interaction would describe, because their parameterization is usually determined from quantum mechanical simulation of the forces.

2.4 Water model

Water is a very important topic in MD simulations. It is ubiquitous in nature and technology and its ability to form hydrogen bonds shapes its properties. This interaction must therefore be well modeled in the simulation. However, hydrogen bonds in water also have a covalent character to a small extent and cannot be described only with Coulomb and Lennard-Jones interaction. This, and the general simplification of the molecule to atomic nuclei with partial charges, ensure that no model of water correctly reflects all its properties quantitatively. There are several models, which differ in the parameters for the potentials or the geometry. In addition, there are extended models, which try to better represent the charge distribution by the free electron pairs and therefore use more than three particles for a water molecule. In general, it is tried to fit the parameters to experimental data. Therefore, water models exist which, for example, fit the experimentally measured dynamics well, but whose melting temperature lies next to it.

The water model used here, TIP4P/2005 [5, 6], fits some properties of water very well, including its dynamics and density anomaly. However, its melting temperature, which Figure we do not examine in detail in this experiment, is about 20 K too low. It has a virtual particle for the negative charge (see Fig. 1) and a stiff geometry.

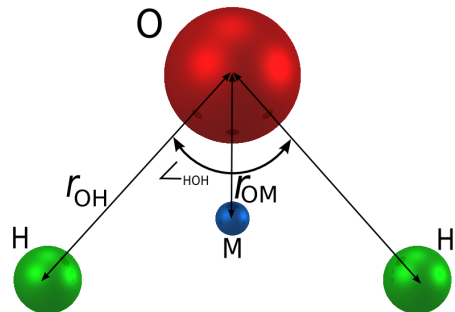


Figure 1: The water model used, TIP4P/2005 [5]. The individual parameters can be taken from Tab. 1.

Table 1: Parameters for TIP4P/2005 [5, 6]

r_{OH} / nm	$\angle\text{HOH} / ^\circ$	σ / nm	$\epsilon/k_B / \text{K}$	q_O / e	q_H / e	q_M / e	r_{OM} / nm
0.09572	104.52	0.31589	93.2	0	0.5564	$-2q_H$	0.01546

2.5 Water anomalies

Although H_2O is a very small and simple molecule, it shows the most anomalies of all molecules [7]. Some of these anomalies are thought to be responsible for making life possible. It has a particularly high melting and boiling temperature for its size and can be found on Earth in all three phases, solid, liquid, and gas. In the solid phase, water can assume over 10 different structures. It has a high heat capacity and heat of vaporization compared to other molecular liquids. It is also important for this experiment that water has a density anomaly, a maximum at 277 K, hat [8]. Some of these properties are attributed to the ability to form hydrogen bonds. The relatively high binding energy for an unbound interaction provides the energetic properties. The geometry, e.g., hydrogen bonds to four neighbors in a tetrahedral arrangement, provides structural properties and effects related to pressure and density. On cooling, the viscosity of water increases as usual. However, increased pressure causes a decrease in viscosity, unlike almost all other liquids. Decreased pressure instead increases viscosity.

2.6 Supercooled liquids

For many atoms, molecules and mixtures of such, crystalline arrangements with low entropy and enthalpy exist. Ordinary ice, for example, has a hexagonal structure. Below the melting temperature T_m these are thermodynamically stable with respect to the liquid because the Gibb's energy, $G = U + pV - TS = H - TS$, of the crystal is smaller. Although the entropy decreases due to the higher order but the enthalpy decreases more. The actual thermodynamic equilibrium is below T_m thus the crystal. The liquid is in a metastable state. However, crystallization is a process with finite probability and dynamics.

Crystallization starts with so-called crystallization nuclei. However, the contact surface of liquid and crystallization nucleus is unfavorable. Two contributions compete, a favoring volume term G_V of the crystal and a hindering surface term G_O of the contact area. Thus, the difference ΔG depends on the size of the crystal nucleus. Above a critical size, $\Delta G < 0$ and the nucleus is thermodynamically stable. In addition, G_V and G_O are temperature dependent. The deeper one cools below T_m the more favorable the crystal becomes and the smaller the critical size becomes. Thermodynamic

fluctuations allow the formation of crystallization nuclei. Small nuclei decay again while nuclei above the critical size grow. The energy barrier for the critical size $\Delta G_{\text{crit.}}$ determines the probability for the formation of a crystallization nucleus ($p \sim \exp(-\Delta G_{\text{crit.}}/T)$). However, the probability for a critical crystallization nucleus increases with size of the system and waiting time. After emergence of a nucleus, it must grow and the system must crystallize completely. It is therefore possible to supercool and study a liquid without it freezing immediately [9].

For water this is limited by the homogeneous nucleation temperature (≈ 235 K) below which the energy barrier for a nucleus disappears. Liquid water can then only be studied on very short time scales. Below this comes the so-called "no man's land", which is virtually impossible to study in experiments [10]. Simulations have a few advantages here. The very small system size makes it extremely rare for a critical nucleus of crystallization to occur. Also, the system is studied on very short time scales. One thus often has the possibility to supercool the liquid deeper without being hindered by crystallization.

2.6.1 Dynamics and the glass transition

The dynamics and viscosity of liquids is strongly temperature dependent and changes over many orders of magnitude. Usually one finds an Arrhenius or super-Arrhenius law, i.e. an exponential increase in reciprocal temperature. Therefore, it is suitable to use a so-called Arrhenius plot for larger temperature ranges (see Fig. 2). The majority of liquids follow a super-Arrhenius law. Glasses are an important example here. Their composition prevents crystallization very well, they are a so-called good glass former, and they therefore remain liquid up to the glass transition temperature T_g . This is the temperature above which the viscosity is so low, or the correlation time so long compared to the experimental time scale, that the system begins to behave like a solid [9]. This is true for all liquids as long as crystallization can be avoided, including water. A final model for this behavior is still lacking and is the subject of current research. Since simulations operate on much shorter time scales, the effective glass transition temperature is higher.

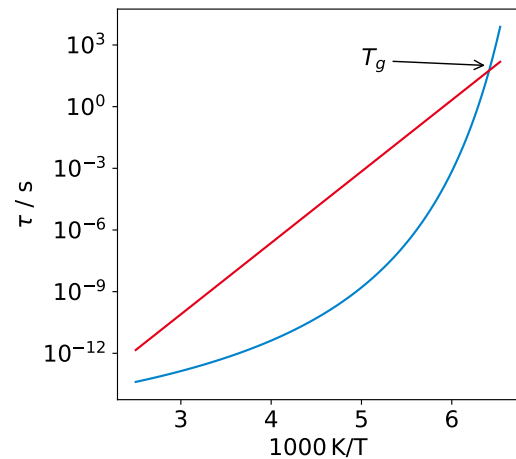


Figure 2: Exemplary temperature dependence of the correlation time τ in a so-called Arrhenius plot. This plot linearizes an Arrhenius (red) and distinguishes it well from a super-Arrhenius behavior (blue).

3 Observables

3.1 Structure

A simple observable is the density, which can be obtained directly from the mean value of the density in the NpT simulation. A quantity for the details of the packing of the particles on microscopic scale is the radial pair distribution function:

$$g(r) = \frac{1}{\rho \cdot N} \sum_i^N \sum_j^N \frac{\delta(r_{ij} - r)}{4\pi r^2}; r_{ij} = |\vec{r}_i - \vec{r}_j| \quad (10)$$

Here \vec{r}_i are the particle positions and ρ the mean particle density. With a discrete histogram, $g(r)$ can be calculated from the trajectory. For better statistics and ergodicity, the result is averaged over different time points t_0 . The cumulative particle number up to a certain distance can be determined by integration:

$$n(r) = 4\pi\rho \int_0^r g(r')r'^2 dr' \quad (11)$$

As indicated in Fig. 3 the pair distribution function can be used to distinguish which aggregate state the system is in. A crystal shows sharp peaks while a liquid shows only short range structures. Important is also the connection with the structure factor, which plays an important role in scattering experiments.

$$S(\vec{q}) = \frac{1}{N} \sum_{i,j} e^{-i\vec{q} \cdot (\vec{r}_j - \vec{r}_i)} \quad (12)$$

By Fourier transformation $g(r)$ and $S(q)$ can be converted into each other (q is the magnitude of the scattering vector). For isotropic systems simplified is valid:

$$S(q) = 4\pi\rho \cdot \int r^2 \frac{\sin(qr)}{qr} (g(r) - 1) dr \quad (13)$$

A problem in simulations is that $g(r)$ can only be determined for distances within the simulation box. Assuming that $g(r > R) = 1$, i.e. that there is no more correlation, the following calculation is very efficient for fluids:

$$S(q) = 1 - \frac{4\pi\rho}{q^2} \left(\frac{\sin(qR)}{q} - R \cdot \cos(qR) \right) + \sum_i \sum_{j, r_{ij} < R} \frac{\sin(qr_{ij})}{qr_{ij}}. \quad (14)$$

As long as the length scale of q ($\frac{2\pi}{q}$) is smaller than R , one gets reasonable results.

Tetrahedral order An essential property of water is locally largely tetrahedral arrangement of the molecules. One can well approximate the position of water molecules by the position of their oxygen atom. The arrangement of the nearest four neighbors should then correspond to a tetrahedron. To check this, we can directly calculate the tetrahedral order parameter from the trajectory. It is zero for completely random arrangement and one for a perfect tetrahedron, with thermal vibrations of the molecules leading to deviations. It is defined for a particle as the sum over its four nearest neighbors:

$$Q_i = 1 - \frac{3}{8} \sum_{j=1}^3 \sum_{k=j}^4 \left(\cos(\Phi_{jik}) + \frac{1}{3} \right)^2 \quad (15)$$

Here Φ_{jik} is the angle of the connecting lines from the central oxygen to two neighbors. For the tetrahedron angle $\cos(\theta_{\text{tetra}} \approx 109.47^\circ) = -\frac{1}{3}$. The tetrahedral order parameter thus determines something like a mean square deviation to the Tetrahedron.

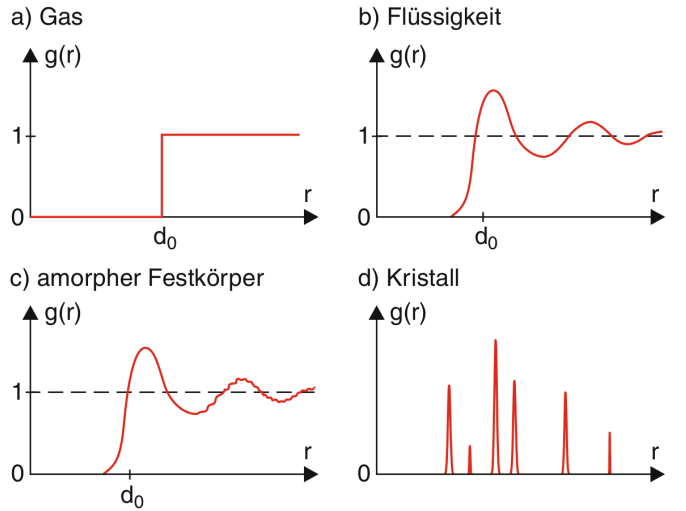


Figure 3: Exemplary pair distribution functions for different aggregate states (a: gas, b: liquid, c: amorphous solid, d: crystal) [11].

3.2 Dynamics

3.2.1 Mean displacement square

The mean-squared displacement (MSD) describes how far particles are from their original position after a certain time difference.

$$\langle \Delta \vec{r}^2 \rangle(t) = \left\langle \frac{1}{N} \sum_i^N (\vec{r}_i(t+t_0) - \vec{r}_i(t_0))^2 \right\rangle_{t_0} \quad (16)$$

Here, too, the average is taken over several temporal origins t_0 . One finds up to three regimes in simple systems. For very short times, $t < 1$ ps, the particles fly freely and one is in the ballistic regime, $\langle \Delta \vec{r}^2 \rangle(t) \sim t^2$. If they meet neighbors, they may be reflected: A plateau is formed. For long times, the diffuse particles according to a random-walk. It can be shown that then with the self-diffusion coefficient (D) the relation

$$\langle \Delta \vec{r}^2 \rangle(t) = 6Dt \quad (17)$$

applies. Determining viscosity from MD simulations involves a lot of effort (a lot of simulation time), which we cannot audit in this experiment. Fortunately, the Stokes-Einstein equation links

$$D = \frac{k_B T}{6\pi\eta r} \quad (18)$$

D with the viscosity η . Here T is the temperature, k_B the Boltzmann constant and r the radius of a particle assumed to be spherical. Thus the self-diffusion coefficient is a practical quantity for the dynamics. In addition, for long times, the MSD can be taken as a simple measure of ergodicity. If it is only very few particle radii large, then the system has not yet assumed enough different microstates. Furthermore, the value of D cannot be determined as long as the MSD is not yet proportional to time.

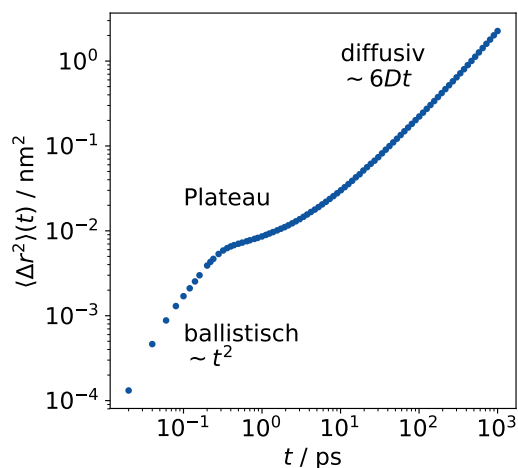


Figure 4: MSD for 250 K. Note the double logarithmic plot.

3.2.2 Autocorrelation of the dipole moment

To study the rotational dynamics of water molecules, we now correlate the dipole moment $\vec{\mu}_i$ of one and the same molecule at two different times, t_0 and $t_0 + t$. The correlation function evaluated for this purpose is called $F_1(t)$ and is the first Legendre polynomial for $\cos(\theta_i(t)) = \frac{\vec{\mu}_i(t_0+t) \cdot \vec{\mu}_i(t_0)}{|\vec{\mu}_i(t_0+t)| \cdot |\vec{\mu}_i(t_0)|}$, where θ the angle between the directions of the dipole moment to the considered times. Mathematically, the following averaging over different time origins t_0 and with unit vectors \vec{e}_μ aus:

$$F_1(t) = \left\langle \frac{1}{N} \sum_i^N \vec{e}_{\mu,i}(t+t_0) \cdot \vec{e}_{\mu,i}(t_0) \right\rangle_{t_0} \quad (19)$$

This function thus measures how the molecules reorient. It starts at one and, in simple cases, decays exponentially ($C(t) = A \cdot e^{-t/\tau}$) to zero. It can be measured approximately by dielectric spectroscopy.

The correlation time τ of the decay then describes how fast the local dynamics is. The shorter τ the faster the water molecules 'forget' their original orientation. In addition, the system can react faster to perturbations. A sheared fluid has internal stress. These pressure fluctuations can be compensated for by a system with a shorter τ or lower viscosity be dismantled more quickly.

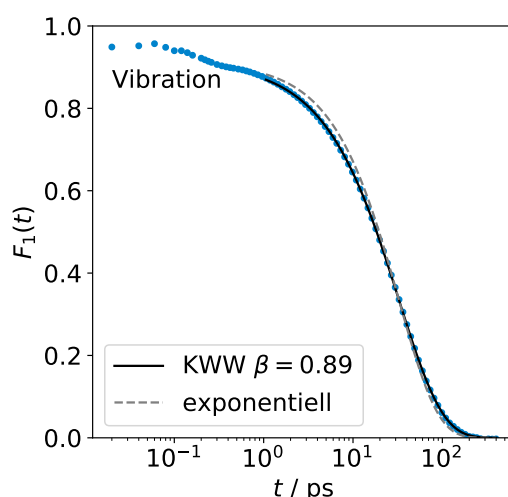


Figure 5: F_1 for 250 K. The data can be better described with a stretched-exponential decay.

4 Preparation

4.1 MD simulations

- Why Newtonian equations of motion?
- Why is it necessary to solve the equations of motion numerically?
- Which potentials serve as a basis (quantum mechanical? classical?)
- How do you deal with the limited system size?
- What problems can arise from discretization in time?

4.2 Thermodynamics and ergodicity

- In which order of magnitude is the particle number of our simulation? Is it a lot or a little?
- Which thermodynamic ensemble corresponds most closely to experiments? How do we achieve this?
- What does ergodicity mean? Why is it important? Why do we have to pay attention to it in simulations and usually not in experiments?

4.3 Measurement of structure and diffusion

- What does the pair distribution function mean and what do we expect for fluids?
- What do we expect for the temperature dependence of the pair distribution function?
- What does the mean square of displacement mean?

4.4 Autocorrelations

- What does one mean by correlations?
- How can we quantify the rotation of water molecules?

4.5 Water Anomalies

- What is the structure of a water molecule?
- What are hydrogen bonds and how many can a water molecule be involved in?
- What local structure does water form?
- What is the density anomaly of water?
- What is the temperature dependence of the viscosity or diffusion of ordinary liquids?
- What happens to the structure of water when pressure changes?
- What happens to the dynamics of water when pressure changes?
- How are these two effects related and why is water anomalous?

4.6 Supercooled liquids

- What is a supercooled liquid?
- Why can water be supercooled?
- How do dynamics change with temperature?
- Why does this limit the lowest temperature that can be studied with simulations? (e.g. in terms of ergodicity)

4.7 Linux use

Familiarize yourself with the concept of entering commands in consoles and read the help at the end of the manual.

5 Implementation

An experiment protocol is also useful for simulations in order to understand how the results were produced. Therefore, note down your steps; especially if you deviate from the instructions or you encounter problems and solve them.

5.1 Density equilibration at 295 K

Starting with the coordinates of a single TIP4P/2005 water molecule, we want to create a configuration with about 1000 molecules and then equilibrate the density at room temperature (295 K) and atmospheric pressure (1 bar).

1. Use the GROMACS tool 'gmx insert-molecules' to fill a box with water (important arguments can be -ci, -box, -nmol and -h to show help).
For safety, you can estimate the required box size by approximating the molecules as spheres and assuming a small packing density, e.g., 0.4. As diameter you can take the σ -parameter of the LJ-interaction from the '.top' file.
2. Optional: Visualize your result with vmd (call vmd with the '.gro' file).
3. Go to the prepared folder T295K_p1bar_NPT. Copy your '.gro' file and the '.mdp', '.top', 'run.sh' and 'eval.yaml' file into it and name it 'gro_water.gro'. Enter the correct number of molecules at the end of the '.top' file.
4. Make sure that in the '.mdp' file the thermodynamic parameters are correct. First, a short simulation with pressure coupling should be performed with the Berendsen Barostat.
5. Run the simulation with the run.sh script. If the simulation stops because of an error, it usually means that the random start configuration is very unfavorable. Fix this with the following steps for an energy minimization.
 - a) In the '.mdp' file set the integrator to 'steep'.
 - b) Use the run.sh script.
 - c) Copy the produced 'out/out.gro' as new start configuration.
 - d) In the '.mdp' file set the integrator to 'md'.
 - e) Optional: 'simclean.sh -f' to clean up the folder.
6. Use 'boxcut.sh' to create a medium density '.gro' file. You can use 'xmgrace box.xvg' to view the fluctuations of the box size. Strong drifts at the beginning of the simulation should be ignored with 'boxcut.sh' ('boxcut.sh -s 0.*'; default 0.1).

-
7. The obtained 'equilibrium.gro' file should be close to the expected density. Now use it to run the simulation again, this time with the Parrinello-Rahman barostat. Estimate a reasonable simulation length in number of integration steps. One integration step corresponds to 2 fs and the self-diffusion coefficient at 295 K is about $2 \cdot 10^{-3} \frac{\text{nm}^2}{\text{ps}}$. An MSD of more than 20 nm² should be achieved. In the '.mdp' file, change the values for 'nsteps' and 'Pcoupl' accordingly. Before simulation you can clean up the folder with 'simclean.sh -f' and get rid of old output.
 8. Use 'boxcut.sh' again.
 9. Optional: Visualize the trajectory with 'vmd'.
 10. Start the evaluation with 'mdeval'. The results will be stored in the subfolder 'eval' in readable files.

5.2 Successive cooling/heating

Now add the remaining NpT simulations: 230 K to 370 K in 10 K increments as well as a simulation at 275 K. For this, it is a good idea to always start with a configuration that is already close to the new temperature (Why?). Therefore you should always cool down step by step and take the configuration of the previous temperature. For 310 K and higher you can reduce the simulation length (halve 'nsteps' in the '.mdp' file). You can also do the evaluation of the simulations explained in the next section on the side to speed up the experiment.

1. Create new folder with appropriate name.
2. Adopt the top, mdp, eval.yaml and run.sh files.
3. Take over the 'equilibrium.gro' of the previous temperature as 'gro_water.gro'.
4. Adjust the temperature in the '.mdp' file
5. Run 'boxcut.sh' again after the simulation
6. Evaluate the simulation with 'mdeval' and view the MSD. Make sure that at least an MSD of 2 nm² has been reached and, if necessary, extend the simulation ('nsteps' in the '.mdp' file).

The last step can also be carried out subsequently. To save time, first perform all the temperatures once and then extend the lowest temperatures if necessary.

5.3 Simulations at elevated/negative pressure

We want to investigate the behavior of water under pressure changes. To do this, create two simulations at 250 K, each with a pressure of -500 and 500 bar. Name the folders accordingly and adjust the 'ref_p' parameter in the '.mdp' file. The evaluation is carried out as before.

5.4 Completion of the evaluation

The evaluation of the simulations is performed with 'mdeval' ('mdeval' in the simulation folder or 'mdeval -d SIMULATIONFOLDER'). The results are stored in the subfolder 'eval' in readable files. You can also run the command again for a simulation and already calculated quantities will not be recalculated. The results for the densities and mean tetrahedral order parameters can be collected using the makefile ('make all -kj 8' in the parent folder). With 'xmgrace' you can open the '.dat' files and check your work.

6 Evaluation and discussion

Many of the measured variables have small errors due to statistics. The systematic errors are much larger but unknown. Therefore, you can often omit errors. However, if it makes sense, for example after an estimation or a fit of the data to a function, you should specify errors.

6.1 Amorphous structure

- Compare your structure factor with the experimental result of X-ray scattering [12]. How well does the simulation fit?
- How does $g_{OO}(r)$ change with the temperature? Do we have a phase transition to the crystal?
- What is the local structure that you can determine from the pair distribution functions? For this, take your lowest temperature and show all $g(r)$ of this temperature in a graph.
 - The first peak in $g_{OO}(r)$ represents the nearest neighbor shell. Use the cumulative particle number of the neighbors $n_{OO}(r)$ at a distance between the first two peaks (Why?) in $g_{OO}(r)$, u to determine the coordination number of the water molecules.
 - To determine the angle α and β in Fig. 6 you need corresponding distances. Determine them with the peaks in the corresponding $g(r)$ or the geometry of the TIP4P/2005.
 - What stands out for α and β in combination with the coordination number? Which local structure does this correspond to?

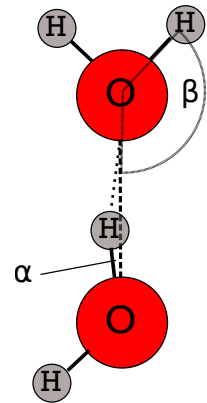


Figure 6: Example of a locale arrangement of water molecules with hydrogen bonding.

6.2 Density anomaly

- Compare your densities with the experimental data [8]. How well does the simulation fit and at what temperature does the water model have its density maximum?
- Discuss this effect along with the data for the mean tetrahedral order (a $\langle Q \rangle$ vs T Plot). Is the local structure that water forms on cooling more densely or more thinly packed?

6.3 Dynamics: diffusion and rotation

- Determine the self-diffusion coefficient at long times from the MSDs. Compare it with the literature data of Holz et al. [13]. Since the dynamics changes over several orders of magnitude, a logarithmic axis is useful.
- Determine the relaxation times from the F_1 autocorrelations. Why do the correlation functions have a drop at very short times? The final decay at long times can be described by a so-called Kohlrausch-Williams-Watts (KWW) function. The KWW is a stretched or compressed exponential decay.

$$C_{\text{KWW}}(t) = A \cdot e^{-(t/\tau)^\beta} \quad (20)$$

It is sufficient here to ignore data points for short times and find the fit parameter τ . But check that the amplitude does not become larger than 1 and that values between 0.5 and 1 are found for β .

- Create a so-called Arrhenius plot ($\log \tau$ vs $1000 \text{ K}/T$) with the relaxation times of the rotation. Add the diffusion by converting it to a time with a suitable jump length $a^2 = 6D\tau$ (e.g., some peak in $g(r)$). What do they notice about this?
- The energy barriers E underlying the motions, e.g. the breaking of a hydrogen bond, could be constant as already known from the studies. Then we expect an Arrhenius law for the temperature dependence:

$$\tau = \tau_0 \cdot e^{E/T} \quad (21)$$

However, super-Arrhenius behavior at low temperatures is usually found for liquids, see also Fig.2. This can be well described by the empirical Vogel-Fulcher-Tammann (VFT) equation

$$\tau = \tau_0 \cdot e^{\frac{B}{T-T_0}} \quad (22)$$

Which of the two temperature dependences better describes your data? Fit the relaxation times of the rotation to both functions and add the curves to the Arrhenius plot. Try to explain your result. For example, they may try it along with the change in structure, the tetrahedral order.

Note: The raw τ s change exponentially. Either logarithmize before the fit or adjust the weighting (e.g., set the error as proportional to τ).

6.4 Stokes-Einstein Relation

The Stokes-Einstein relation links the self-diffusion coefficient to the viscosity:

$$D = \frac{k_B T}{6\pi\eta a} \quad (23)$$

The Debye formula $\tau_D = \frac{4\pi\eta r^3}{k_B T}$ for T reorientation and experimental data show that $\tau \sim \frac{\eta}{T}$. Together with the Stokes-Einstein relation results in $\tau \cdot D = \text{const.}$. This is called a Stokes-Einstein breakdown if this relation is no longer satisfied, e.g. $\tau^\theta \cdot D = \text{const.}$ with $\theta \neq 1$ holds. Verify this relation in a graph.

6.5 Behavior during pressure changes

If you have already plotted the data points for the two additional pressures in the previous graphs, you can now still discuss them in the context of the usual expectations. Increasing the pressure actually means more dense packing of the particles and thus slower dynamics and more pronounced structures. What do you find for water in the supercooled region?

7 References

- [1] Michael Griebel. *Numerical simulation in molecular dynamics : numerics, algorithms, parallelization, applications*. Vol. 5. Texts in computational science and engineering. Berlin [u.a.], 2007. ISBN: 9783540680949.
- [2] Mark Abraham Berk Hess David van der Spoel and Erik Lindahl. *GROMACS Reference Manual Version 2018.8*. The GROMACS development teams at the Royal Institute of Technology and Uppsala University, Sweden. 2019. URL: <https://manual.gromacs.org/documentation/2018-current/manual-2018.8.pdf>.
- [3] Giovanni Bussi, Davide Donadio, and Michele Parrinello. “Canonical sampling through velocity rescaling”. In: *The Journal of Chemical Physics* 126.1 (2007), p. 014101.
- [4] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190.
- [5] SklogWiki: TIP4P/2005 model of water. URL: http://www.sklogwiki.org/SklogWiki/index.php/TIP4P/2005_model_of_water. (accessed: 05.04.2022).
- [6] J. L. F. Abascal and C. Vega. “A general purpose model for the condensed phases of water: TIP4P/2005”. In: *The Journal of Chemical Physics* 123.23 (2005), p. 234505.
- [7] Martin Chaplin. *Water Structure and Science: Anomalous properties of water*. URL: https://water.lsbu.ac.uk/water/water_anomalies.html. (accessed: 30.03.2022).
- [8] Francesco Mallamace et al. “The anomalous behavior of the density of water in the range $30\text{ K} < T < 373\text{ K}$ ”. In: *Proceedings of the National Academy of Sciences* 104.47 (2007), pp. 18387–18391.
- [9] Andrea Cavagna. “Supercooled liquids for pedestrians”. In: *Physics Reports* 476.4 (2009), pp. 51–124. ISSN: 0370-1573.
- [10] H. E. Stanley et al. “The puzzling behavior of water at very low temperature. Invited Lecture”. In: *Phys. Chem. Chem. Phys.* 2 (8 2000), pp. 1551–1558.
- [11] Wolfgang Demtröder. *Experimentalphysik 3 : Atome, Moleküle und Festkörper*. 5th ed. 2016. Springer-Lehrbuch. Berlin, Heidelberg, 2016. ISBN: 9783662490945.
- [12] Lawrie B. Skinner et al. “Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range”. In: *The Journal of Chemical Physics* 138.7 (2013), p. 074506.
- [13] Manfred Holz, Stefan R. Heil, and Antonio Sacco. “Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^1H NMR PFG measurements”. In: *Phys. Chem. Chem. Phys.* 2 (20 2000), pp. 4740–4742.

8 Helpful information about the use of Linux and the provided programs

8.1 Linux-tips

In this experiment, the use of a Linux-based operating system and the so-called console/terminal is very advantageous. Therefore, a minimalistic introduction is given here, so that the experiment can be carried out. A console can be opened via the start menu but also by right clicking into a folder and 'open in terminal'. In the latter case one is also already in the correct folder. For the console applies namely that paths can be specified absolute or relative to your current directory. Generally one can execute 'installed programs' simply by typing into the console. Additionally you can pass so called arguments after the program, further options or inputs to be used. This could look like this: 'gedit mdp_water.mdp'. 'gedit' is a simple text editor and 'mdp_water.mdp' is an argument, the file to be displayed. Since spaces separate arguments, you should not use spaces in folder and file names.

Here are some useful commands (CAPITAL LETTERS are to be replaced):

- cd DIRECTORY (the console directory changes to the one specified)
- ls DIRECTORY (lists all files; without argument from current directory)
- rm FILE (deletes the specified file; with -r directories can also be deleted)
- mkdir DIRECTORY (creates the DIRECTORY)
- mv FILE1 FILE2 (moves or renames files)
- cd .. (goes up one folder)
- gedit FILE1 FILE2 ... (text editor)
- vmd GRO XTC (visualizes molecular dynamics files)
- gmx (to call the programs of GROMACS)
- bash SCRIPT (can be used to run scripts)
- ./SCRIPT (alternative method to run scripts if they are marked as executable)
- boxcut.sh -d DIRECTORY (a script of the attempt to obtain a configuration of medium density)
- mdeval -d DIRECTORY (python package for evaluating simulations, needs the eval.yaml file in DIRECTORY)
- make all (Understands so-called makefiles, which are used here to collect the results)
- simclean.sh -d DIRECTORY (cleans simulation order, with -f also output is deleted!)
- xmgrace FILE1 FILE2 ... (a program for plotting data)

Additionally '*' can be used as a placeholder for characters. 'xmgrace eval/F1_*p1bar*' opens all files that match 'F1_*p1bar*' in the 'eval' sub-folder. Most **PRACTICAL** is the tabulator key. It completes text in the console to the next position where there are several valid possibilities. Quick double pressing makes these possibilities appear in the console!

8.2 Use of GROMACS

GROMACS is a very comprehensive and performant package for MD simulations. It is C-based, can solve several other simulations besides Newtonian equations of motion, and comes with several programs that can perform standard evaluations. It was actually designed for the simulation of proteins, but can also be used for arbitrary systems just because of its performance with unbound interactions. Important here is the use of GPUs, which enables a further acceleration by a factor of 4 or more.

For simple MD simulations a topology ('.top'-file), initial coordinates ('.gro'-file) and a file with simulation parameters ('.mdp'-file) are needed.

-
- '.gro' files consist of the number of atoms, the indexes, names and coordinates themselves and lastly the geometry of the simulation box. They can additionally contain velocities.
 - Topologies contain the atom types and their WW parameters. Furthermore, the molecules with their bonds are defined. Last, the content of the system, the number of different molecules, is given.
 - '.mdp' files contain type of simulation and length, parameters for the output, barostat and thermostat, parameters for the interactions and more to control the simulation.

The program grompp (listed as 'gmx grompp') uses these input files to create a single complete input file in tpr format. This is then the input for 'gmx mdrun', the simulation itself. Both calls are made in a prepared script that only needs to be executed.

8.3 Use of mdevaluate

Evaluations of the trajectories could also be performed with the GROMACS tools. To reduce the complexity and duration, the python package mdevaluate¹ is used. Evaluations are predefined and are specified in the 'eval.yaml' file. Here the results are stored in a folder 'eval' relative to the simulation folders. mdevaluate is called with 'mdeval -d SIMULATION FOLDER'. With 'mdeval -r -np 8' you can also evaluate all simulations at the same time.

The following files are relevant:

- 'F1_T*.dat' are the F_1 autocorrelations. (t / ps and $F_1(t)$)
- 'gr_*.dat' are the Ppair distribution functions corresponding between OW, HW and their combination. (r / nm and $g(r)$)
- 'Sq_*.dat' are the structure factors corresponding between OW, HW and their combination. (q / nm^{-1} und $S(q)$)
- 'msd_T*.dat' are the MSDs (t / ps and $\Delta r^2 / \text{nm}^2$)
- 'nr_T*.dat' sare the cumulative nearest neighbors (r / nm and $n(r)$)

8.4 Accumulation of the results with the help of the makefile

Some of the results have only one value per temperature. These are stored collectively with the makefile. In addition, missing boxcut.sh executions can be made up. A makefile with the command 'make' is used. By adding '-kj 8' errors are ignored and everything is parallelized. Only if the input has changed the output file will be updated. So it does not hurt to execute it again.

The following files, also in the 'eval' folder, are relevant:

- 'densities_p*bar.dat' for the densities in SI units.
- 'meanQ_p*bar.dat' for the mean tetrahedral order.

¹<https://github.com/mdevaluate>