Final Project: Simulating Water and Ethanol

2MMN40: Introduction to Molecular Modeling and Simulation
Last update: December 15, 2020

Contents

1 Introduction

This document describes the course project. By the end of the course you have written a fairly powerful Molecular Dynamics (MD) simulator in Python. In this project you will make use of your simulator to study a water and ethanol mixture. As you work on this project, keep in mind the theory that you have learned during the lectures, and try to make connections to it in your analysis and results.

1.1 Groups

For this project you will be working in groups of two.

2 Report Structure

The report should have the structure of a scientific paper, it should have an introduction explaining the problem. This should also include a short literature review which would include a summary of what has been done in a *scientific* context, experimentally or otherwise, to study ethanol in water. Outline some interesting phenomena — will you be able to study these phenomena with your code? Describe the structure of the molecules that you will study. What problems do you foresee in studying these things? Do you think that your code and force-field are enough to capture all of the Physics in this system? Why or why not?

You may use the headings in this document as a guideline of what sections to have in your report. You also have the freedom to write your report however you like, however, the questions posed in this document should be answered at a minimum. You may of course explore these topics further, or consider other topics you find interesting.

Treat this report like a scientific paper. That means that one should cite any resources used and include a properly formatted bibliography. Important scientific/mathematical statements made in your report should be either supported by a citation or by a result that you produce yourself. There is no such requirement as a *minimum* number of citations, but there should be a few that you are bound to collect during your work.

There is a hard upper limit of 15 pages for this report, not including the code and any appendices. There is no minimum page count.

3 Theory

Summarize the methods that you are using to study this system. Make connections to the theoretical concepts that you have learnt in the lectures.

Describe the physics of this system with some detail. Talk about some of the physical and chemical properties of the water/ethanol system and discuss what you think can and cannot be captured by classical MD. Why is this system interesting to study? Name some interesting physical or chemical phenomena that this system has. Out of the things you think can be studied with MD, do you think your code can simulate it?

3.1 Integrator

You have spent some time studying some different integrators, provide an analysis of the different integrators that you have used. Use H_2 as an example molecule. We encourage you to try an integrator that was not covered in the course. Discuss why velocity verlet is a good integrator to use for the simulation.

4 Simulation

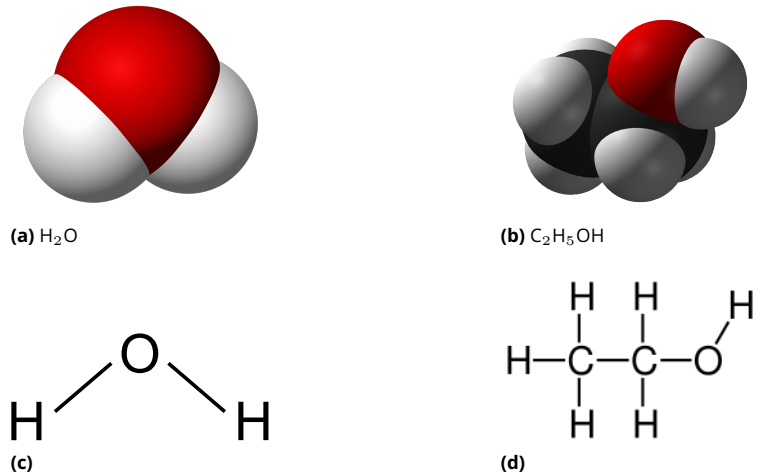


Figure 1. The molecules in this system.

Figure 1 shows the molecules that you will be simulating for this project. You will be simulating solutions of ethanol in water.

All force-field parameters are given in section 4.3.

4.1 Systems to Simulate

Simulate the following three cases with your NVT simulator:

- A system of pure water
- A system of pure ethanol
- A system of 14.3% ethanol in water (percents are in mole, the strange percentage is to be able to compare the results to literature, see the results section)

To simulate the system you will need to set a system size. We recommend a system size of $5\times5\times5$nm and a simulation duration of 1ns. This is, however, a large simulation (it is not strange if the simulation takes 40 to 50 hours, but this depends heavily on your implementation). Depending on your implementation you might need to scale down, for example to a box of $3\times3\times3$nm and a simulation time of 0.1ns (that should be around 40 to 50 times quicker, but if you have some inefficiencies in you code might still take 20-30 hours). There are two things to consider here, computational complexity/time and the physics. What goes wrong from a physical perspective if the simulation system is to small?

To get the number of molecules that should be present in the box, you should google for the density of water, ethanol and the mixture. From the density you can compute how many molecules you need of each type (this sounds hard, but is just a high school chemistry computation). Do the simulations at 298.15K. If you have time, try a range of temperatures in [250K, 350K].

You should follow the guidelines in this section in your simulations. However, please make some arguments with regards to why the numbers here are, or aren't reasonable. These arguments may be physical or computational (performance reasons).

4.2 Simulation Parameters

Timestep

Pick a good timestep, as a starting point you could take 2fs. Depending on your system you might need a smaller one, but smaller than 0.01fs should never be necessary. Realize that a smaller time step means a longer simulation. It might be worthwhile to experiment a bit with the timestep, to find the largest timestep you can take (without the system exploding) before you do a full simulation.

You can also arrive at the largest possible time step in a formal way. Find the fastest oscillation in your molecules, determine its period and say you want ten "points" within one oscillation then your time step will be the period divided by ten.

Lennard-Jones

For Lennard-Jones, use an appropriate cut-off length. Once again, a larger cut-off probably results in longer simulation times, but a small cut-off means you will lose important physics. Discuss in the report why and how you chose your cut-off. Do not calculate any Lennard-Jones between atoms within the same molecule (is this the right thing to do?).

Mixing Rules

The Lennard-Jones (LJ) potential assumes that we are considering interactions between identical atom types. They have to be tweaked slightly when considering the interaction between two different atom types. This amounts to a slight rescaling of the LJ constants. If σ_i and σ_j are the constants for two different atoms i and j (say Oxygen and Carbon), then the value corresponding to their interaction will be

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j). \tag{1}$$

And for the ϵ's,

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}. \tag{2}$$

Give some details with regards to why these mixing rules are needed. Can you think of a better way to consider the mixing?

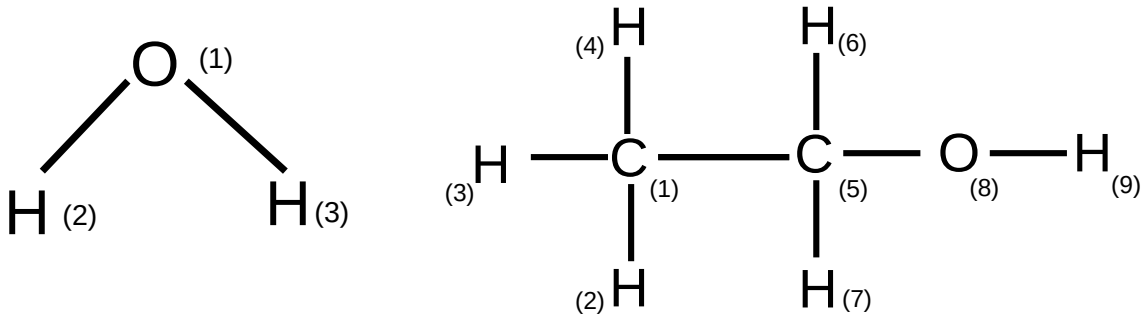
4.3 Force-Field Parameters

4.3.1 Water

The bonds and angles defined below use the indices in figure 2. The parameters below were taken from the TIP3P water model.

Species	Mass (amu)
H	1.0080
O	15.9994

Table 1. H_2O masses



(a) **(b)**

Figure 2. Labelled structure for Water and Ethanol. You may, naturally, change the indexing to whatever you like. The bonds, angles, and dihedrals below are defined with these indices in mind.

Species	σ (nm)	ϵ (kJ/mol)
O	0.315061	0.66386
H	0	0

Table 2. H_2O Lennard-Jones parameters.

ij	R_o (nm)	k_b (kJ mol^{-1} nm^{-2})
12, 13	0.09572	502416

Table 3. H_2O bonds.

ijk	θ_o (^{o})	k_θ (kJ mol^{-1} rad^{-2})
213	104.52	628.02

Table 4. H_2O Angles.

4.3.2 Ethanol

Refer to figure 2. These parameters were taken from the OPLS/AA force-field.

Species	Mass (amu)
C	12.0110
Rest	Table 5

Table 5. Ethanol masses

Species	σ (nm)	ϵ (kJ/mol)
H (bonded to C)	0.25	0.125520
H (bonded to O)	0.00	0.00
C	0.35	0.276144
O	0.312	0.711280

Table 6. Ethanol Lennard-Jones parameters.

ij	R_o (nm)	k_b (kJ mol^{-1} nm^{-2})
15	0.1529	224262.4
31,41,21,65,75	0.1090	284512.0
58	0.1410	267776.0
98	0.0945	462750.0

Table 7. Ethanol bonds.

ijk	θ_o (^{o})	k_θ (kJ mol^{-1} rad^{-2})
215, 315, 415	108.5	292.880
413, 412, 312, 657	107.8	276.144
157,156	110.7	313.800
158	109.5	414.400
589	108.5	460.240
658, 758	109.5	292.880

Table 8. Ethanol Angles.

4.3.3 Dihedrals

You will need to include dihedrals for the ethanol molecule. We will use a slightly more complicated potential named the Ryckaert-Bellman potentials. These are not that complicated actually because they're periodic, so taking their gradients is straight-forward.

$$V(\psi) = \frac{1}{2}\left(C_1(1+\cos(\psi)) + C_2(1-\cos(2\psi)) + C_3(1+\cos(3\psi)) + C_4(1-\cos(4\psi))\right)$$

where $\psi = \theta_{ijkl} - 180^o$.

$ijkl$	C_1 (kJ mol^{-1})	C_2 (kJ mol^{-1})	C_3 (kJ mol^{-1}	C_4 (kJ mol^{-1})
2156, 3156, 4156 2157, 3157, 4157	0.62760	1.88280	0.00000	-3.91622
2158, 3158, 4158	0.97905	2.93716	0.00000	-3.91622
1589	-0.44310	3.83255	0.72801	-4.11705
6589, 7589	0.94140	2.82420	0.00000	-3.76560

Table 9. Ethanol Dihedrals.

4.4 Thermostat

Use the simple velocity rescaling thermostat to set the temperature of your system. This is done by simply rescaling the momenta up and down towards the desired temperature by a factor λ

$$\vec{p}_s = \lambda\vec{p}$$

where $\lambda = \sqrt{(T/T_i)}$. T being the desired temperature, and T_i being the current temperature.

Discuss the efficacy of this thermostat. Do you think that it is suitable for your simulation? If you have time, you may implement one of the other thermostats covered in the lectures and compare.

5 Code

Attach your code to this report in a single file[1]. The code should be well commented and you should provide enough documentation such that it can be understood and run by us. It should also run flawlessly with a standard Anaconda3 installation. You may use libraries that are not included[2] in Anaconda3, but then these libraries should be clearly listed. Obviously, you may not use any libraries that do the MD computations for you...

In your report, make a flowchart showing how your code works. Make an assessment of the performance, and flexibility of your code. Outline the things you have learnt while writing it. What do you think is particularly impressive regarding your code? What is not so impressive? Would you change anything? Make some proposals that you think will improve your code[3].

6 Results

For each of your simulations measure:

1. Total Energy
2. Potential Energy
3. Kinetic Energy

[1] This means either one source file or a zip file if you have multiple source files

[2] We don't expect you to do this, but we allow it nevertheless.

[3] At the end of this course, we encourage you to publish your code on github and share it with the world. We think it will look particularly impressive on a CV.

4. Temperature
5. Some Radial Distribution Functions (Pair Correlation Functions) (these are 6 separate things)
   (a) O-O (water-water, ethanol-ethanol, and ethanol-water)
   (b) O-H (water-water, ethanol-ethanol, and ethanol-water)

Compare your results to other known results or results from other simulations in the literature. As a minimum, compare your simulation results to the *Molecular dynamics simulation of ethanol/water mixtures for structure and diffusion properties* paper. They use a different method to simulate the same system. It is interesting to see and discuss what the differences are. The pure water case can be compared against the *The Radial Distribution Functions of Water...* paper. You can find copies on Canvas.

7 Conclusion

Summarize your report and present some interesting conclusions. Propose some extensions to this project in the future.

Appendix A Tips and Hints

Some hints you can follow but you do not have to do:

- Running an MD simulation requires three steps. Setting up an initial structure, running an MD simulation and analyzing the data. Writing three programs, where each does one of these tasks, will make your code more readable and give you more safety as you can save intermediate results. It requires a bit more code for writing and reading input files though.
- Besides the geometry and the topology your program will require additional input options, timestep, temperature , number of steps etc.... You can write these parameters into your code or better make an option file which is read by your code. It separates your data more from your code and makes running multiple simulations easier.
- It is important to realize that your system starts in an artificial configuration. The initial positions of all the molecules are determined by you and not by the physics of the system. You will need to let your system run for a bit before your measurements of the energy and the RDFs make any physical sense. This might take a while depending on your initial configuration, to speed up the convergence towards this physical system, it is beneficial to introduce some randomness in your initial configuration. In particular for the ethanol-water mixture, if you define your initial positions on a grid, make sure the water and ethanol molecules are already mixed, don't put all the ethanol molecules on one side of the box and all the water molecules on the other, because then you will also have to wait for all the molecules to mix.

Appendix B Assessment Rubric for the Final Project 2MMN40

Criteria	Excellent (4)	Good (3)	Sufficient (2)	Insufficient (1)
Organization of written report (A: 10%)	Report is clear and logical. Reader can easily follow line of reasoning.	Report is generally clear. A few minor points may be confusing.	Reader can follow the report with effort. Organization not well thought out.	Report is very confusing and unclear. Reader cannot follow it.
Report style (B: 10%)	Style is appropriate and makes use of the field's terminology. Not too casual. Reader can clearly understand impact and purpose of the work.	Style is generally appropriate. May have some trouble in explaining results or purpose of the work.	Report is too informal or unprepared. Difficult to understand. Much of information lacks focus and clarity.	Report is consistently at an inappropriate level. Information is not well synthesized. Reader can't understand the point of the work.
Content: Depth (C: 30%)	Theory, methods, results, discussion and conclusions are clearly and coherently elucidated. Logical and persuasive agreement between data and conclusions. Impact and implications of results.	Description of project and results is generally clear. Some discussion of what results mean.	Some components of project description are minimal or missing. Little discussion of what results mean.	Description of the project and results is very difficult to follow. No discussion of meaning of results. Readers learn little.
Content: Accuracy (D: 20%)	Information given is consistently accurate. Facts and calculations are correct.	No significant errors are made. Readers recognize errors as result of oversight and offer advice to correct.	Enough errors made to be distracting, but some information is accurate.	Information is so inaccurate that reader cannot depend on the reported results and work.
Code quality (E: 10%)	Code runs and is well documented (i.e. there are sufficient but not too many comments). The code has a good structure (i.e. it can be read and understood easily without the need of an excessive number of comments)	Code runs and is well documented. The code is somewhat structured.	Code works partially. Some comments are present and the code is somewhat structured.	Code does not work.
Use of Figures and Tables (F: 10%)	Figures and tables are prepared in a professional manner. Font is large enough to be seen. Well organized. Main points stand out.	Figures contribute, but not all material is visually represented well. Font size is appropriate for reading.	Figures are poorly prepared or used inappropriately font is too small. Too much information is included.	No Figures are used, or they are so poorly prepared that they detract from the report.
Final synthesis and summary of results (G: 10%)	Main study questions are directly addressed and clarified in the discussion.	Generally responsive to main study questions. Work is put in a general framework of related work.	Report poorly addresses research questions. Very little connection to related work.	Not responsive to research questions. Little to no connection to prior work in the field, and how this work extends that.

Appendix C Determination of the Final Grade

Grade for the report

To compute the grade for the report we will use the following formula

$$\text{Report Grade} = 2.5 \cdot [0.1(A + B + E + F + G) + 0.3C + 0.2D],$$

capitalized letters refer to elements of the rubric, the points can be found between the parentheses in the rubric (e.g. Excellent (4), means you get 4 points).

Final grade

Your final grade will be determined after a small oral exam. We will ask you a few questions about the project, but it will not be a full defense. Your final grade will be the grade for the report, plus or minus a small adjustment based on your oral exam.