2AMM10 Assignment 2 :
Molecular Dynamics using Machine Learning

1 Introduction

Over the past years, the amount of CO_2 in the atmosphere has been increasing, and greenhouse effects have become more evident. A potential way to reduce the carbon levels in the atmosphere is through carbon capture, using nanoporous materials such as zeolites. Zeolites consist of multiple TO_4 tetrahedra, where T can be either aluminium or silicon. By combining multiple of these tetrahedra, a nanoporous material is formed. As a result, zeolites have a large surface area, on which molecules such as CO_2 and CH_4 can be adsorbed.

Dr. Menko Vladovski is world-renowned for his research on zeolites, and has recently come up with a new theoretical 2D zeolite called Vladolite (Figure 1), which he has not been able to synthesize yet. To still be able to study the behaviour of this potential zeolite, dr. Vladovski has developed a Molecular Dynamics simulation software, which models the physical interactions of various atoms being adsorbed inside his zeolite. However, since most of dr. Vladovski's research group has turned to computer simulations, he is only able to perform a limited number of simulations each day, greatly limiting his ability to study his potential solution for the climate crisis. Therefore, he asked you, as AI and machine learning experts, to develop a simulator based on Deep Learning, in the hopes that it can greatly speed up Vladolite simulations.

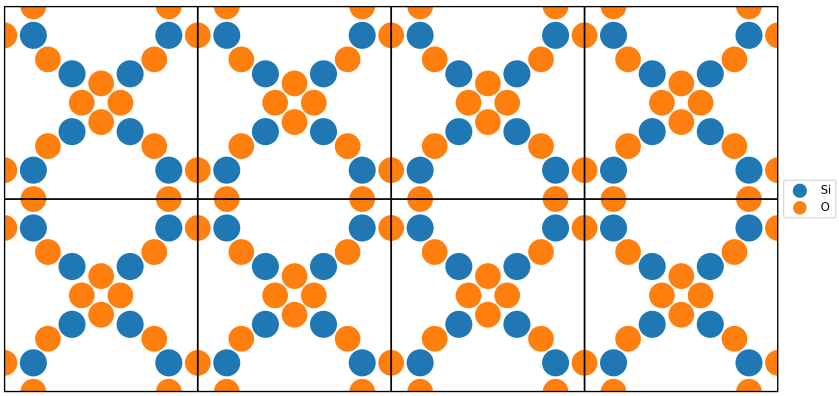

Figure 1: 4 by 2 unit cells of all-silica Vladolite

In particular, dr. Vladovski wants to carry out simulations where charged ions are inserted into the zeolite. He has provided you with the following information regarding the simulations:

- **Unit Cell** Like all other crystals, Vladolite is defined through its unit cell. The unit cell is the minimum set of atoms, which, when tiled (infinitely) in all directions, forms the crystal structure. By analyzing the behaviour of a crystal inside the unit cell, it is possible to extrapolate properties to the full crystal structure. In the case of Vladolite, the unit cell is cubic and has dimensions of 20Å by 20Å[1].

[1] Angstrom. $1Å = 10^{-10}m = 0.1nm$

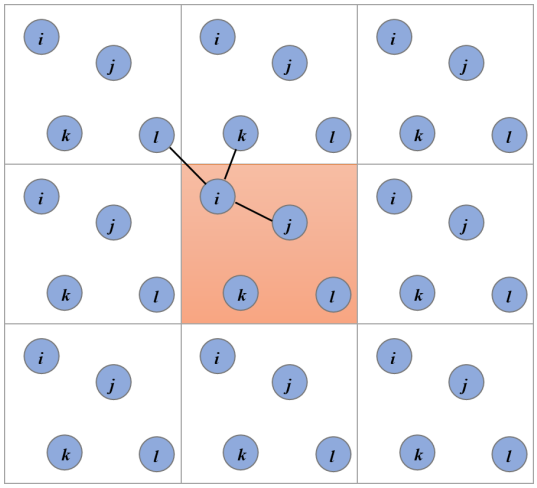

Figure 2: Example of minimum image convention. The unit cell with the orange background is the "original" unit cell, while the other cells are periodic images. Here the distances between i and j, k, l are calculated. As can be seen, to calculate the distance to l and k, the periodic boundaries need to be crossed. Image credit: https://commons.wikimedia.org/wiki/File:Minimum_Image_Convention.png

- **Periodic Boundary Conditions** When modelling the unit cell of the crystal, we assume that the system repeats itself periodically. When a particle crosses a boundary, it re-enters through the opposing boundary (similar to the game Pac-Man). When calculating geometric features, such as distances, we account for this effect using the *minimum image convention*. Under this rule, the distance is calculated by taking the *minimum* distance between atom i and all periodic *images* (in neighbouring unit cells) of atom j. An example of the minimum image convention can be found in Figure 2.
- **Energies and Forces** In molecular dynamics, the evolution of the system is guided by the energies at each time step. Here, atoms follow the gradient of the energy, the force, to move towards a low-energy state. Energies and forces are calculated between each pair of atoms in the system. Dr. Vladovski has pointed out to you that in the system we are studying, the energies and forces between two atoms become negligible (close to 0) when the distance between them is more than 10Å.

2 Assignment Description

It was already mentioned that we will study molecular dynamics in this assignment. More precisely, we study how a system of $n \in \{3, 4, 5\}$ atoms $p_1, p_2, \ldots, p_n$ moving in a 2D unit cell evolves over time. At time t, the system is described by the charge $q_i \in \{-1, 1\}$, the position $\mathbf{x}_i^t \in [0, 20)^2$ and velocity $\mathbf{v}_i^t \in \mathbb{R}^2$ of each atom p_i. The evolution of the system at time t is solely dependent on the charges, positions, and velocities of the atoms at time $t-1$. Furthermore, for any two objects, the force between them depends only on their relative locations and the charges of the objects. For all tasks, functions for loading and reading the data and a short explanation of the data structure are provided in the skeleton Jupyter notebook accompanying this assignment.

2.1 Task 1: Energy Prediction

Like many other physicists, dr. Vladovski is skeptical about the use of Deep Learning for speeding up Molecular Dynamics simulations, as he considers these models to be magical black boxes. As such, he first wants to investigate whether it is possible to model a set of atoms in a periodic, cubic, 20Å by 20Å box, **without** the crystal. Since Molecular Dynamics is governed by the energy of the system, being able to predict it is an important

step towards building a Deep Learning simulator. In this task, you are given a set of positions $\mathbf{x}_i$, velocities $\mathbf{v}_i$, and charges q_i of atoms p_i. Then, your goal is to predict the energy e of the system.

For this task, the training data consists of trajectories $\mathbf{x}_i^{0:T}$ and $\mathbf{v}_i^{0:T}$ and charges q_i of atoms p_i and the associated energies $\mathbf{e}^{0:T}$. The test set consist of datapoints containing the trajectories $\mathbf{x}_i^{0:T}$, velocities $\mathbf{v}_i^{0:T}$ and charges q_i, for which the associated energy $\mathbf{e}^{0:T}$ needs to be predicted.

Hints:

- It is important that your model architecture is aligned with the properties (symmetries, structure) of the data, to make your model as data-efficient as possible. Keep this in mind when coming up with a solution for this task.
- As a result of the periodic domain, some ways in which we typically calculate or represent geometric features might not work out of the box. Which geometric features is your model using? Do you need to modify anything to respect the periodic domain? Keep in mind that there are multiple possible solutions for the representation of the data, which can all have different geometric features that need to be adapted for periodicity.

2.2 Task 2: Molecular Dynamics with Atoms

After predicting energies, the next step towards building a simulator is modelling the dynamics. Specifically, your task is to build a model, that, given a set of initial positions $\mathbf{x}_i^0$, velocities $\mathbf{v}_i^0$ and charges q_i of atoms p_i, predicts a sequence of positions $\mathbf{x}_i^{1:T}$, up to terminal state of $T = 40$. The length of each time step is 0.5 seconds.

For this task, the training data consists of the same data as in the previous task. The test set consists of datapoints containing initial positions $\mathbf{x}_i^0$ and velocities $\mathbf{v}_i^0$ and charges q_i, as well as the energy $\mathbf{e}^0$, paired with the remainder of the trajectories $\mathbf{x}_i^{1:T}$ of the atoms p_i for evaluating the performance of your method.

Hints:

- While building a model, keep in mind how your design choices affect the ability of your model to handle symmetries. What will happen to the trajectory if the initial positions and velocities are rotated? What will happen if they are reflected horizontally?
- We only care about the performance of your model with respect to the predicted positions for its evaluation. However, depending on your approach, it might be useful to also have your model predict velocities and energies in order to infer positions later on.
- When performing the evaluation of your model, keep in mind that you are only allowed to give it as input $\mathbf{x}_i^0$, $\mathbf{v}_i^0$, $\mathbf{e}^0$. The rest of the trajectory needs to be generated by your model!

2.3 Task 3: Molecular Dynamics in a Crystal

Seeing that its possible to model trajectories of a few charged atoms, dr. Vladovski is convinced that this approach could also work to simulate charged atoms inside his zeolite. In reality, the zeolites dr. Vladovski often works with are flexible, meaning that the crystal atoms move around, and cause pores to open and close. However, simulating zeolites as a flexible crystal is very computationally expensive. In addition, very few simulation engines can perform this task accurately. Luckily, dr. Vladovski has found out that using simulations where the crystal is kept rigid (atoms belonging to the crystal are not allowed to move) still yield properties which match values from real-life experiments. Dr. Vladovski has already performed Molecular Dynamics for various instances of the Vladolite zeolite, which all have varying numbers of silicon and aluminium atoms.

Similarly to the previous task, given a set of initial positions $\mathbf{x}_i^0$, velocities $\mathbf{v}_i^0$ and charges q_i of atoms p_i, you have to predict a sequence of positions $\mathbf{x}_i^{1:T}$. However, in addition to the positions of atoms, you are also provided with the positions of the crystal atoms $\mathbf{z}_i$, as well as the charges of the crystal atoms c_i. Silicon atoms have a charge of 0.2, aluminium atoms have a charge of 0.15 and oxygen atoms have a charge of -0.1.

For this task, the training data consists of trajectories $\mathbf{x}_i^{0:T}$ and $\mathbf{v}_i^{0:T}$, charges q_i of atoms p_i and the associated energies $\mathbf{e}^{0:T}$ as well as the positions $\mathbf{z}_i$ and charges c_i of the crystal atoms. The test set consists of datapoints containing initial positions $\mathbf{x}_i^0$, velocities $\mathbf{v}_i^0$, charges q_i for atoms p_i, the initial energy $\mathbf{e}^0$, as well as the positions $\mathbf{z}_i$ and charges c_i of the crystal atoms. This is paired with the remainder of the trajectories $\mathbf{x}_i^{1:T}$ of the atoms p_i for evaluating the performance of your method.

Hints:

- You should be able to re-use (a large part of) your model from the previous task, since it is still only the charged atoms which move. How will you incorporate information about the crystal in your model?
- Keep in mind that for each trajectory, the associated crystal can be different (i.e. aluminium atoms can be replaced by silicon and vice-versa).

3 Deliverables

The submission of this assignment is done in ANS. There are two deliverables:

- Implementation of the solution
- Description and explanation of your approach

For the implementation, a skeleton Jupyter Notebook with functions that can load the provided data for each task are provided. The notebooks also contain short bits of self-explanatory code on how the data is formatted and how to load the data in python. Please submit your code in the provided skeleton Jupyter notebook. You need to upload the Jupyter Notebook code (.ipynb file extension) and a PDF version (.pdf file extension) with the execution output. The maximum upload size is 25MB. So, you may need to clean some image output from the code Notebook before uploading. If you are using Google colab, you can generate the PDF by going to "File → Print → Save as PDF". Please generate the PDF after running all cells of your solution (with possibly images removed if necessary for file size constraints).

The description of your approach is submitted as a digital group test in ANS. Rather than an unstructured report, the group test is formatted in a number of questions that you need to answer to describe your solution and understanding of the assignment. Please see ANS for the details.