

TITLE OF THE MASTER THESIS

by

Filip Henrik Larsen

THESIS

for the degree of

MASTER OF SCIENCE



Faculty of Mathematics and Natural Sciences
University of Oslo

May 2017

Abstract

This is an abstract text.

To someone

This is a dedication to my cat.

Acknowledgements

Flora Joelle Larsen
Anders Malthe-Sørenssen
Anders Håfreager
Henrik Sveinson
Kjetil

Contents

1	Introduction	1
2	Molecular dynamics	3
2.1	Time-integration	3
2.1.1	Velocity Verlet	4
2.2	Potentials	6
2.2.1	Lennard-Jones	6
2.2.2	Stillinger-Weber	7
2.2.3	Vashishta	7
2.3	Boundary conditions	8
2.3.1	Minimum image convention	9
2.4	Measuring physical quantities	11
2.5	Thermostats	11
2.5.1	Berendsen	12
2.5.2	Andersen	12
2.5.3	Nosé-Hoover	13
2.6	Efficiency improvements	13
2.6.1	Cut-off	13
2.6.2	Cell lists and neighbor lists	14
2.6.3	Parallelization	14
3	LAMMPS	17
3.1	Installation	17
3.1.1	Linux	17
3.1.2	Mac OS X with Homebrew	18
3.2	Input scripts	19
3.2.1	System configurations	19
3.2.2	Run-time commands	20
3.2.3	Output	22
3.3	Visualization	24
3.3.1	OVITO	24
4	Preparing a molecular dynamics simulation	25

Contents

4.1	Silica	25
4.1.1	Unit cell of β -cristobalite	26
4.2	Building a crystal	26
4.3	Verifications	28
4.3.1	Melting point	28
4.4	Shaping the silica	30
4.5	Moving the sphere towards the slab	32
5	This must be sorted in designated chapters	33
5.1	Radial distribution of normal force	33
5.1.1	Radial binning	33
6	Computing the normal force distribution	37
6.1	Creating a custom compute	37
6.1.1	Find a similar compute	38
6.1.2	Creating the class	38
6.2	Least squares regression	44
A	Source code	47
A.1	compute_group_group.h	48
A.2	compute_group_group_atom.h	49
A.3	compute_group_group_atom.cpp	50
B	Something	55
B.1	LAMMPS units	55

Chapter 1

Introduction

Why is the subject of this thesis of any interest?

What is our take on the problem?

What do we hope to accomplish?

How will this be of any contribution to anything?

How is the thesis laid out?

Chapter 2

Molecular dynamics

Molecular dynamics is a method that uses Newton's equations of motion to simulate the movement of interacting atoms. Each atom is treated as a point particle with a mass, and the interaction between them is described by a force field. Molecular dynamics aid our understanding of macroscopic phenomena by giving us insight in microscopic effects and behavior. It can be used if an experiment is too difficult, dangerous or expensive to do otherwise, for instance experiments at extreme temperatures or pressure. The areas of application are wide, and it has become a very popular field in material science, biochemistry and biophysics.

In this chapter we will present the basic principals of molecular dynamics, including essential algorithms, common methods and efficiency improvements.

2.1 Time-integration

The evolution of particles position and velocity is carried out through time integration. In molecular dynamics, atoms movement is governed by Newton's second law

$$\sum_{i=1}^N \mathbf{F}_i = m_i \mathbf{a}_i \quad (2.1)$$

Theoretically, the computation of particles movement is easy to assimilate. We know that acceleration is the derivative of velocity, which in turn is the derivative of position. Thus we should be able to retrieve the velocities and positions using time-integration of the acceleration.

$$\mathbf{v}_i(t) = \mathbf{v}_i(0) + \int_0^t \mathbf{a}_i(t) dt \quad (2.2)$$

$$\mathbf{r}_i(t) = \mathbf{r}_i(0) + \int_0^t \mathbf{v}_i(t) dt \quad (2.3)$$

When solving an integral numerically, we're forced to discretize the problem and use an approximation method. This is done by dividing the domain into steps

of size Δt and approximate the area of the function inside each such partition. Typically, the error of the approximation decreases when reducing the time step Δt . However, a smaller step length leads to higher computational expense, and may limit the size or time-frame we are able to compute. These are common considerations when doing computational research. One has to balance precision, size/time-frame, time and expense. It is therefore very important to choose a numerical scheme that has a satisfying level of precision as well as speed.

There are numerous numerical integration methods, but not all these have properties that are desirable for molecular dynamics simulation. The most fundamental features the scheme should have are high precision, computationally cheap, good energy conservation, reversibility and be deterministic. Conservation of energy is fundamental in physics, and in order to simulate the micro-canonical ensemble (NVE), the integration method cannot have an energy drift. The dynamics should be reversible, meaning one should be able to simulate the system in the opposite direction in time, and arrive at past states. It should be deterministic in the sense that by the information about a specific state $(r(t_0), v(t_0))$ one should be able to compute the state at any other time $(r(t), v(t))$, be it future or past. Molecular dynamics simulations are computationally expensive, and practically all the work lies in the time-integration loop. More precisely computing the sum of forces on each atom. The choice of integration scheme is therefore very important. As it turns out, probably the best scheme for the task is also one of the simplest.

2.1.1 Velocity Verlet

In molecular dynamics the most common scheme for time-integration is the *Velocity Verlet* algorithm, which is a specific type of Verlet integration. It is derived by Taylor expanding the position both one time step forward and one backwards as follows:

$$r(t + \Delta t) = r(t) + \frac{\Delta t}{1!}r'(t) + \frac{\Delta t^2}{2!}r''(t) + \frac{\Delta t^3}{3!}r'''(t) + \mathcal{O}(\Delta t^4) \quad (2.4)$$

$$r(t - \Delta t) = r(t) - \frac{\Delta t}{1!}r'(t) + \frac{\Delta t^2}{2!}r''(t) - \frac{\Delta t^3}{3!}r'''(t) + \mathcal{O}(\Delta t^4) \quad (2.5)$$

If we add these two together, the odd terms cancel. After rearranging, we are left with the very simple verlet algorithm.

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \Delta t^2 r''(t) + \mathcal{O}(\Delta t^4) \quad (2.6)$$

Or written in a compact notation where

$$\mathbf{r}^n = r(t) \quad \mathbf{r}^{n\pm 1} = r(t \pm \Delta t) \quad \mathbf{a}^n = \frac{\partial \mathbf{v}^n}{\partial t} = \frac{\partial^2 \mathbf{r}^n}{\partial t^2} \quad (2.7)$$

this can be expressed as

$$\mathbf{r}^{n+1} = 2\mathbf{r}^n - \mathbf{r}^{n-1} + \Delta t^2 \mathbf{a}^n . \quad (2.8)$$

Notice that we do not rely on information about the velocity in order to predict the next position. The algorithm in equation (2.8) is known as Strömer-Verlet. It's not self-starting, since you need to know the previous and current steps to predict the next. Thus, it needs to be initialized in a way, e.g. by using algorithm (2.10) for the first step. It has a high order truncation error, proportional to $\mathcal{O}(\Delta t^4)$, however it's vulnerable to round-off errors. This is because the last term is proportional to Δt^2 , and is potentially much smaller than the other terms. When trying to add it with the other terms, this may be a source of round-off error [2]. As a remedy to this problem, the Velocity Verlet algorithm incorporates a second order differential approximation of the velocity.

$$\mathbf{v}^n = \frac{\mathbf{r}^{n+1} - \mathbf{r}^{n-1}}{2\Delta t} + \mathcal{O}(\Delta t^2) \quad (2.9)$$

Resulting in

$$\mathbf{r}^{n+1} = \mathbf{r}^n + \Delta t \mathbf{v}^n + \frac{\Delta t^2}{2} \mathbf{a}^n \quad (2.10)$$

$$\mathbf{r}^{n+1} = \mathbf{r}^n + \Delta t \left(\mathbf{v}^n + \frac{\Delta t}{2} \mathbf{a}^n \right) \quad (2.11)$$

We may recognize the parenthesis as $\mathbf{v}^{n+1/2}$, leaving us with

$$\mathbf{r}^{n+1} = \mathbf{r}^n + \Delta t \mathbf{v}^{n+1/2} . \quad (2.12)$$

Next, we taylor expand the velocity and use a first order approximation of the acceleration.

$$\mathbf{v}^{n+1} = \mathbf{v}^n + \Delta t \mathbf{a}^n + \frac{\Delta t^2}{2} \dot{\mathbf{a}}^n + \mathcal{O}(\Delta t^3) \quad (2.13)$$

$$\mathbf{v}^{n+1} = \mathbf{v}^n + \Delta t \mathbf{a}^n + \frac{\Delta t^2}{2} \frac{\mathbf{a}^{n+1} - \mathbf{a}^n}{\Delta t} \quad (2.14)$$

$$\mathbf{v}^{n+1} = \mathbf{v}^n + \frac{\Delta t}{2} (\mathbf{a}^{n+1} + \mathbf{a}^n) \quad (2.15)$$

$$\mathbf{v}^{n+1} = \mathbf{v}^{n+1/2} + \frac{\Delta t}{2} \mathbf{a}^{n+1} \quad (2.16)$$

Equations (2.12) and (2.16) define the Velocity Verlet algorithms. The final Velocity Verlet algorithm is thus:

$$\mathbf{v}^{n+1/2} = \mathbf{v}^n + \frac{\Delta t}{2} \mathbf{a}^n \quad (2.17)$$

$$\mathbf{r}^{n+1} = \mathbf{r}^n + \Delta t \mathbf{v}^{n+1/2} \quad (2.18)$$

$$\mathbf{a}^{n+1} = -\nabla U(\mathbf{r}^{n+1})/m \quad (2.19)$$

$$\mathbf{v}^{n+1} = \mathbf{v}^{n+1/2} + \frac{\Delta t}{2} \mathbf{a}^{n+1} \quad (2.20)$$

The Velocity Verlet algorithms has the same truncation error as the original, but a lower round-off error. This algorithm has good energy conservation, and is symplectic.

SOME MORE

2.2 Potentials

2.2.1 Lennard-Jones

One of the simplest and most known potentials is the Lennard-Jones potential. It was first proposed in 1924 by John Edward Lennard-Jones. It takes the form

$$V(r) = 4\epsilon \left[\left(\frac{\theta}{r} \right)^{12} - \left(\frac{\theta}{r} \right)^6 \right], \quad (2.21)$$

where the first term represents Pauli repulsion due to overlapping electron orbitals, and the last represents the van der Waal force. The constant θ express the distance at which the inter-atomic potential is zero, while ϵ is the depth of the well. r is of course the inter-atomic distance. A plot of the potential is shown in figure 2.1. The inter-atomic distance at which the potential is at it's minimum can easily be shown to be $r = 2^{1/6}\theta$.

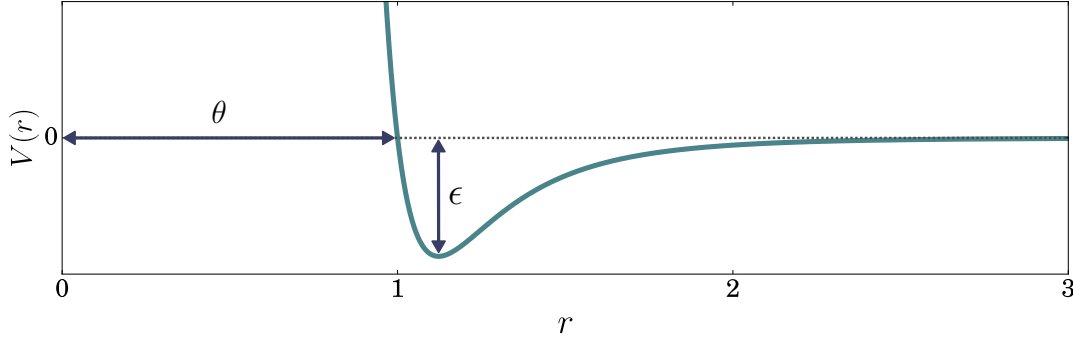


Figure 2.1: Lennard-Jones potential as function of inter-atomic distance, r , in units of θ .

Since the potential only account for the Pauli repulsion and van der Waal forces, its applicability is limited to systems consisting of neutral atoms or molecules, where there are no bonds present. It would suffice for studying noble gases, for instance, but that quickly becomes quite dull.

2.2.2 Stillinger-Weber

2.2.3 Vashishta

The Vashishta potential consists of two terms that represent two- and three-body interactions respectively. These incorporate physical effects such as steric repulsion, coulomb interaction, dipole interaction, Van der Waal interaction and energy related to covalent bonds. It can be expressed as

$$V = \sum_{i < j} V_{ij}^{(2)}(r_{ij}) + \sum_{i < j < k} V_{ijk}^{(3)}(\mathbf{r}_{ij}, \mathbf{r}_{ik}), \quad (2.22)$$

where \mathbf{r}_i represents the position of the i -th atom, $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, and $r_{ij} = |\mathbf{r}_{ij}|$ is the distance between atom i and atom j . The two-body term, denoted by superscript (2), sums over all atoms j within a cutoff range r_c from atom i , and is given as

$$V_{ij}^{(2)}(r) = \frac{H_{ij}}{r^{\eta_{ij}}} + \frac{Z_i Z_j}{r} e^{-r/r_{1s}} - \frac{D_{ij}}{2r^4} e^{-r/r_{4s}} - \frac{W_{ij}}{r^6}. \quad (2.23)$$

where H_{ij} and η_{ij} are the strength and exponent of the steric repulsion respectively, Z_i is the effective charge of atom i , r_{1s} and r_{4s} are screening lengths for the coulomb and dipole interaction respectively, D_{ij} is the strength of the dipole interaction, and W_{ij} is the strength of the Van der Waal interaction.

The three-body term of the potential, denoted by superscript (3), sums over all atoms j and k , that are covalently bonded to atom i , within a cutoff range

of r_0 , and is expressed as

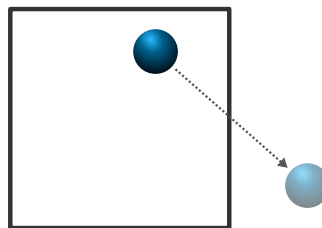
$$V_{ijk}^{(3)}(\mathbf{r}_{ij}, \mathbf{r}_{ik}) = B_{ijk} \exp\left(\frac{\xi}{r_{ij} - r_0} + \frac{\xi}{r_{ik} - r_0}\right) \frac{(\cos \theta_{ijk} - \cos \theta_0)^2}{1 + C_{ijk} (\cos \theta_{ijk} - \cos \theta_0)^2} \quad (2.24)$$

where B_{ijk} is the strength of the three-body term, ξ and C_{ijk} are constant parameters, and θ_{ijk} is the angle between \mathbf{r}_{ij} and \mathbf{r}_{ik} . The values of all these parameters depend on the system at hand. For instance, the strength of the steric repulsion, H_{ij} , for Si-Si interaction is different from that of Si-O interactions. For the case of silica, these metrics have been computed already, and when using the Vashishta potential in LAMMPS there is a file containing them. As for the Lennard-Jones potential, the two-body potential term is truncated and shifted to prevent unphysical jerks, and retain stability. This is done the same way as in equation (2.29).

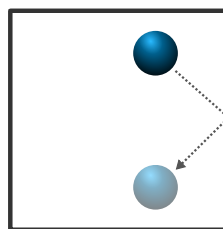
2.3 Boundary conditions

Boundary conditions is a crucial detail to decide upon when setting up a molecular dynamic experiment. The way we treat atoms at the boundary can have an immense effect on their behavior and how physically reasonable the results will be. There are several types of boundary conditions one may desire, and one may define custom conditions. A few examples are listed below.

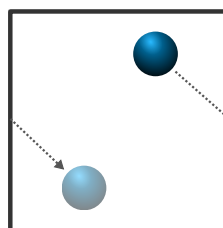
No boundary conditions. Particles are not subject to any special rules at any boundary. The system size may be regarded as infinite. This might be a reasonable choice when studying explosions for instance, or in experiments that is on a really short time scale.



Reflecting boundary conditions, which act as hard walls. In stead of passing through the boundary, the velocity component in the direction normal to the face of the boundary changes sign, thus causing the atoms to be confined within the simulation box.



Periodic boundary conditions, which allow atoms on separate sides of a boundary to interact through the boundary as if they were neighbors, and atoms crossing the boundary reappears on the other side of the simulation box. This is useful when studying bulk atoms of a material or materials that has a periodic structure.



The boundary conditions may include more details than only how to handle the trajectory of atoms intersecting the boundaries. For instance with periodic boundaries, atoms that are far apart are effectively close. These atoms should therefore interact, since their motion would seem artificial otherwise. Thus, the interpretation of atoms positions must in some cases be included in the boundary conditions. In order to interpret atoms positions in the case of periodic boundaries, one has to remember that an atom should not interact with multiple images of another atom. Which image, if not the original, must therefore be decided. Also, an atom should not interact with an image of itself. That could potentially cause strange correlations. The standard way of determining atoms positions is by using the *minimum image convention*.

2.3.1 Minimum image convention

The minimum image convention simply states that the periodic image of an atom closest to the reference atom is the one to consider as its position. Example cases of 1D and 2D systems are shown in figure 2.2 and figure 2.3 respectfully. In the

examples we look at systems consisting of two atoms and we try to determine the minimum image of the smaller, blue atom with respect to the larger, red, reference atom. The original system is distinguishable from the periodic copies by the background color; they have a slightly darker background. Also, we only bother showing the periodic copies of the atom of consideration, the blue one. The systems depicted are of size L and $L \times L$ in the 1D and 2D cases respectfully.

By studying figure 2.2 it becomes apparent that if the atom of consideration is closer to the reference atom than half the system length, the original image will be used to interpret its position. Otherwise, if it's farther apart than half the system length, a periodic image will be closer and therefor used as the atoms position. This principle is applicable at higher dimensions as well. We simply decompose the positions and check each dimension separately. For a system of arbitrary number of dimensions the minimum image convention is assured by using the progression of listing 2.1. Figure 2.3 illustrate how we decompose the positional separation of the atoms in the 2D case.

```

1  for (int k=0; k<numberOfDimension; k++){
2      delta[k] = atom[j].pos[k] - atom[i].pos[k]
3      if (delta[k] > L/2) {
4          delta[k] = delta[k] - L
5      }
6      else if (delta[k] < -L/2){
7          delta[k] = delta[k] + L
8      }
9  }
```

Listing 2.1: Loop to compute the position of the closest periodic image of atom j with respect to the reference atom i .

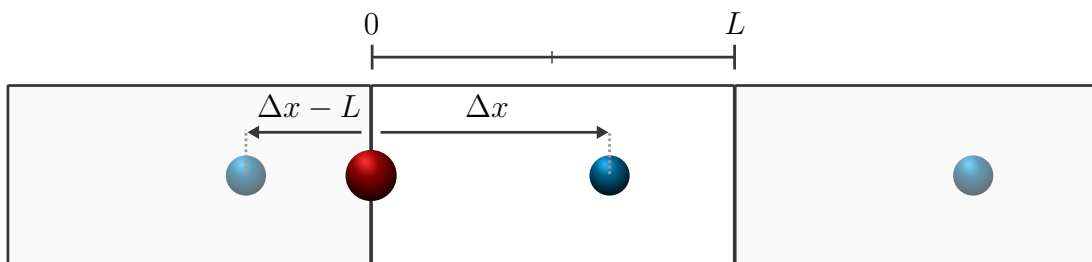


Figure 2.2: Minimum image convention in 1 dimension. If an atom is separated from the reference atom (red) by more than half the system length, $L/2$, the minimum image convention states that the distance between the two is the distance to the periodic copy, which is $|\Delta x - L|$. The white area indicates the system, while the gray areas are periodic replications of the system. Atoms in the gray areas are periodic images of the atom of consideration in the white area.

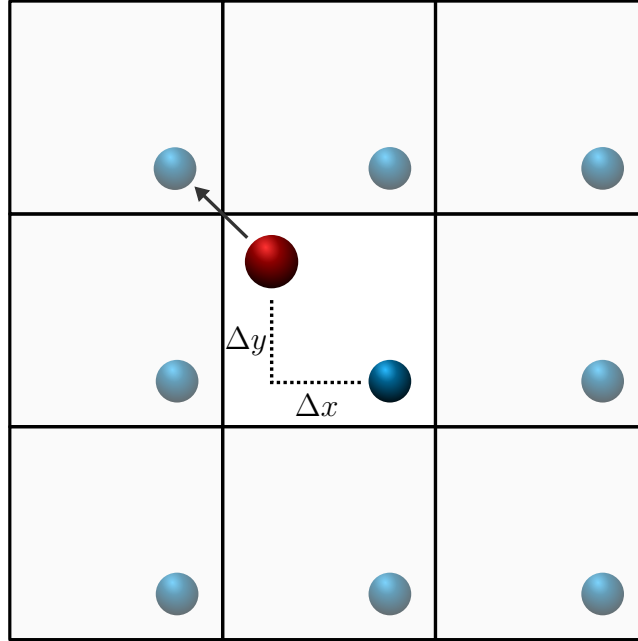


Figure 2.3: Minimum image convention in 2 dimensions. The position of an atom, with respect to the reference atom (red), is the position of the original or any of the replicated images that has the shortest distance to the reference atom. The minimum image is found by decomposing the atoms position, and carry out the The shortest distance in this case is marked with an arrow. Thus, when computing positional dependent quantities it's the position of the north-west replica that will be considered as the position of the blue atom in this case.

2.4 Measuring physical quantities

2.4.1 Energy

2.4.2 Temperature

2.4.3 Pressure

2.4.4 Something else?

2.5 Thermostats

From thermodynamics we have that the kinetic energy, or total thermal energy is

$$E_k = \frac{f}{2} N k_b T \quad (2.25)$$

, for a system of N atoms with f degrees of freedom and temperature T . k_b is the Boltzmann constant. If we assume newtonian motion, the total kinetic energy in a system is statistically

$$E_k = \frac{1}{2} \langle m_i \mathbf{v}_i^2 \rangle. \quad (2.26)$$

We can equate these to find the statistical temperature given as

$$T = \frac{\langle m_i \mathbf{v}_i^2 \rangle}{f N k_b}. \quad (2.27)$$

Thus, we can approximate the instantaneous temperature using the atoms velocity. This value will fluctuate about the real value, so time-averaging is often used when measuring.

Often, we want to control the temperature of our simulated system. This is done by fixing the temperature in equation (2.27) and manipulate the atoms velocities. There are several methods of manipulating the velocities, and these are called *thermostats*. There are many different types of thermostats, and they all have their strengths and weaknesses.

We will look at two quite basic and different thermostats, as well as one that is common in professional use.

2.5.1 Berendsen

The Berendsen thermostat rescales the velocity of all atoms in the system so that the temperature approaches that of the surrounding heat bath. It rescales the atoms velocity by multiplying them by a scaling factor γ , given as

$$\gamma = \sqrt{1 + \frac{\Delta t}{\tau} \left(\frac{T_{bath}}{T} - 1 \right)}, \quad (2.28)$$

where Δt is the time step length, and τ is a relaxation time, which tune the coupling to the heat bath. T_{bath} and T is the temperature of the heat bath and the current temperature of the system, respectfully. Typically the relaxation time is set to $\tau \approx 20\Delta t$. Using $\tau = \Delta t$ would result in the velocities rescaling so the temperature would be exactly the target temperature, T_{bath} , every time step. This thermostat suppress temperature fluctuations efficiently, rendering it well suitable for equilibration procedures. However, because it rescales the velocities, it produce dynamics that are inconsistent with the canonical ensemble.

2.5.2 Andersen

The Andersen thermostat simulates the coupling between the system and an imaginary heat bath as collisions between their atoms. Atoms that collide are assigned a new normally distributed velocity with standard deviation $\sqrt{3k_B T_{bath}/m}$

about the target temperature, T_{bath} . The procedure of the thermostat goes as follows: For each atom in the system, we generate a random uniformly distributed number in the interval $[0, 1]$. If this number is less than $\Delta t/\tau$, the atom is considered to be colliding, and is assigned a new velocity. Here, τ is regarded as a collision time, and its value should be similar to τ in the Berendsen thermostat ($20\Delta t$). The Andersen thermostat is useful when equilibrating systems, but because randomly chosen particles are assigned randomly distributed velocities, it disturbs the dynamics of e.g. lattice vibrations. Also, this disturbance will decorrelate the system, rendering this thermostat ill-equipped when measuring e.g. diffusion coefficient.

2.5.3 Nosé-Hoover

2.6 Efficiency improvements

The major part of the CPU time is spent in the force loop. At every time step we must recompute the force acting on each individual atom. When doing so, we should in theory include the contribution from all other atoms. Having a system consisting of N atoms would result in $N(N-1)/2 \propto N^2$ computations, if we apply Newton's third law. In this section we will look at the most fundamental efficiency improvements applied in molecular dynamics simulations.

2.6.1 Cut-off

Depending on the potential in use, the forces become negligible at certain distances. For instance if one uses the Lennard-Jones potential (introduced in section 2.2.1) the contributions are practically zero for atoms positioned at a distance $r \geq 2.5\sigma$. Therefore, during a simulation we choose to only account for the contributions from atoms closer than this *cut-off* distance, denoted r_c . For the Lennard-Jones potentials the cut-off range is usually set to 2.5θ . Past this range, the pair-potential is practically zero. It's not zero, though. Without any alterations, particles intersecting the cut-off limit will experience an unphysical jerk, which may render the simulation unstable. To counteract this effect, we raise the potential, ensuring it and its derivative to be zero at the cut-off. We use the following adjustment.

$$V(r) = \begin{cases} V(r) - V(r_c) - \left. \frac{\partial V(r)}{\partial r} \right|_{r=r_c} (r - r_c) & , r \leq r_c \\ 0 & , r > r_c \end{cases} \quad (2.29)$$

which implies the force

$$F(r) = \begin{cases} F(r) = F(r) - F(r_c) & , r \leq r_c \\ 0 & , r > r_c \end{cases} \quad (2.30)$$

In practice we need to keep track of which atoms are within the cut-off range of each atom. This is achieved using cell lists and neighbor lists.

2.6.2 Cell lists and neighbor lists

The main purpose of the cell list is to make the building of neighbor lists more efficient. We need to check which atoms are neighboring atoms, but obviously we do not need to check the entire domain, since the cut-off length is relatively small. Therefore, we partition the system into several cubes of size equal to the cut-off length. We store the atoms contained by each cell in a *cell list*. Finally, when we build the neighbor lists we check only the atoms within the neighboring cells and those in the same cell, 27 cells in total. In this sense, by neighboring cells we mean any cells that share a face, edge or vertex. This is illustrated in figure 5.4. The pay-off of using cell- and neighbor lists is tremendous. Since we now, for each atom, only include contributions from atoms within a constant volume of size $4\pi r_c^3/3$, the number of contributions will only depend on the density, which is an intensive¹ property. Thus, the number of computations is reduced to $\mathcal{O}(N)$, which is an immense relief in computational expense!

2.6.3 Parallelization

It might be misleading to refer to parallelization as an efficiency improvement, when on the contrary it most likely increases the CPU time usage. However, the real time consumed may be greatly decreased. It is intuitive that partitioning the work and processing these simultaneously will decrease the time as compared to processing it serially. The speedup is defined as

$$S = \frac{T_s}{T_p}, \quad (2.31)$$

where T_s is the time used when executing the program on a single processor, and T_p the time used when running on p processors simultaneously. The time spent running a parallel implementation of a code using p processors is seldom trivially $T_p = T_s/p$. This is due to the fact that there is a certain amount of time used on *overhead*. This includes interprocess communications, idling and excess

¹Physical property of a system that does not depend on the system size.

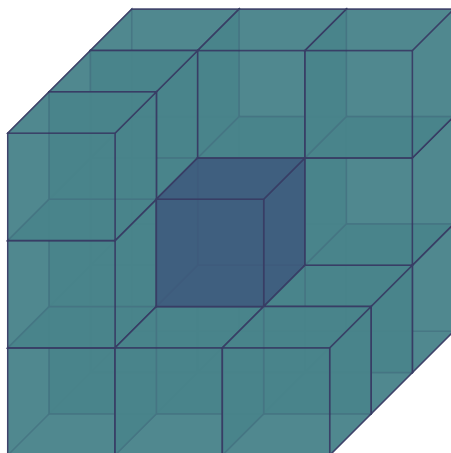


Figure 2.4: Illustrative figure of cells of concern when building the neighbor lists. The lighter cubes are neighboring cells; the darker cube is the cell containing the reference atom. The 7 cells in front of the dark cell are removed from the figure, but are also included.

computations. In molecular dynamics simulations there is communication between processors when building the cell- and neighbor lists, and when computing thermodynamical properties such as energy, pressure, temperature, etc..

During this project we have mainly been using the local supercomputer at the department of physics at the University of Oslo. It provides users with the possibility to run up to 256 processes at once. Though, before doing so, it is considered as good practice to check the speedup obtained by using several numbers of cores. In order to compute the speedup we initialized a system containing $15 \times 15 \times 15$ unit cells of beta-cristobalite and saved it as a restart file. We then remotely ran the input script shown in Listings 2.2 from the supercomputer using 1, 2, 4, 8, 16, 32 and 64 processors in the same fashion as shown in Listing 2.3. The resulting speedup of using the respective number of processors is plotted in figure 2.5.

```

1  include "system.in.init"
2  read_restart ${filename}
3  include "system.in.settings"
4
5  variable N equal 1
6  variable T equal 293
7
8  neighbor 0.3 bin
9  neigh_modify delay 10
10
11 timestep 0.002
12
13 dump myDump slabUpperGroup atom 1 dumpFiles/surface_*.dump
14
```

```

15  fix nvt all nvt temp ${T} ${T} 1.0
16  run ${N}

```

Listing 2.2: LAMMPS input script executed using several numbers of processors, and timed separately.

```

1  mpirun -n 8 lmp_mpi -in speedup.in -var filename
    speedup.restart

```

Listing 2.3: Command used to execute the input script speedup.in on 8 parallel processors and set the filename variable to speedup.restart.

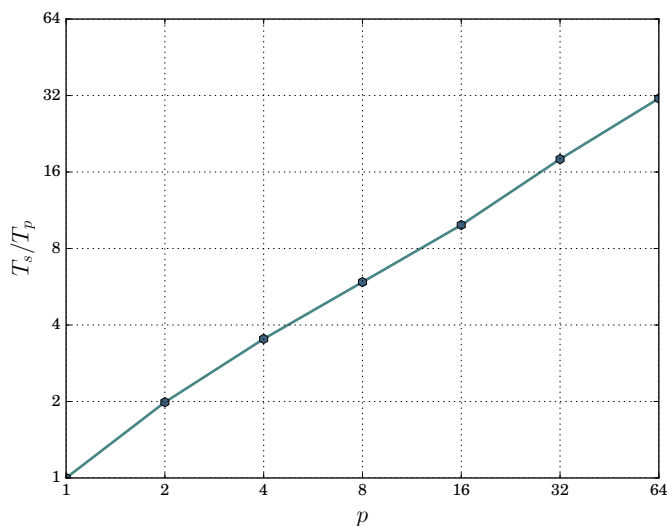


Figure 2.5: Speedup as a function of number of processors.

The result indicate that the speedup is in fact not simply $S = p$. Anyhow, we clearly see the advantage of using 64 processors in parallel as opposed to 1. We can finish a job that would have taken an hour in two minutes! Also, we clearly see that the initial claim holds; this is not more efficient when regarding CPU time. In fact this result suggest that using 1 processor is twice as energy efficient as using 64.

Chapter 3

LAMMPS

LAMMPS is an acronym for *Large-scale Atomic/Molecular Massively Parallel Simulator*. It is a classical molecular dynamics simulation code designed to run efficiently on parallel computers. Its development began in the mid 1990s at Sandia National Laboratories, with funding from the U.S. Department of Energy. It was a cooperative project between two DOE labs and three private companies. The development is still ongoing and contributions are revised thoroughly. Today LAMMPS is an open-source code with extensive and user friendly documentation. This is one of the main reasons why we have chosen to use LAMMPS as opposed to other molecular dynamics software.

In this chapter we will be presented with a guide to install LAMMPS, a brief explanation of commands and syntax, and an example of a valid input script.

3.1 Installation

Installing LAMMPS is a fairly simple procedure if only the basic settings are needed.

3.1.1 Linux

Users with a Unix based OS may download the lammps distribution as a tarball from LAMMPS' download page¹ and then unpack it from the command line.

```
1 gunzip filename.tar.gz
2 tar xzvf filename.tar
```

The user should then change directory into `/path/to/lammps/src/`, and execute the following commands in order to list available packages.

```
1 make package-status
```

¹<http://lammps.sandia.gov/download.html>

Installing specific packages is accomplished as shown below.

```
1 make yes-molecule yes-manybody yes-python yes-rigid
```

The above example installs the packages *molecule*, *manybody*, *python* and *rigid*. Next, the user can build LAMMPS using either of the lines below. Assuming the user has MPI installed, line 2 makes the resulting executable compatible with parallelization in MPI.

```
1 make serial
2 make mpi
```

At this point there should be an executable in the `/path/to/lammps/src/` directory named `lmp_serial` or `lmp_mpi`, depending on the previous choice. These are now ready to run. To use it one has to point to this file from the command line at every run. It may be practical to set up a symlink as shown below.

```
1 sudo ln -s /path/to/lammps/src/lmp_mpi
   /usr/local/bin/lmp_mpi
```

Then the executable will be available as `lmp_serial` or `lmp_mpi` from any directory.

3.1.2 Mac OS X with Homebrew

Mac users can follow the procedure described above, however they may also install even easier using *Homebrew*².

```
1 brew tap homebrew/science
2 brew install lammps           # serial version
3 brew install lammps --with-mpi # mpi support
```

Where the user obviously should choose either line 2 or line 3, depending on whether the user wants MPI comparability. This will install an executable named "lammps", a python module named "lammps", and resources with standard packages. This is basically it. LAMMPS is now ready to run, however, not all packages are installed. The location of the resources and available packages can be found using the following command.

```
1 brew info lammps
```

Specific packages are available as options, and may be installed using the following syntax.

```
1 brew install lammps --enable-manybody
```

²<http://brew.sh/>

In the example shown we installed the package `manybody`.

3.2 Input scripts

The executable made in the previous section can be used to read so-called input scripts. The input scripts contain LAMMPS commands to configure the simulation. This includes all settings and actions. This is naturally partitioned into two sections: *system configurations* and *run-time commands*. The scripts are executed on 8 CPU's in parallel as

```
1 mpirun -n 8 lmp_mpi -in in.system
```

Despite the widely common convention of letting the file type be determined by the letters following the punctuation mark, it's common in the LAMMPS manuals to name the input scripts with the name last. i.e. `in.name`. Luckily, both conventions are meretricious since the executable accepts any name/type for the input script.

3.2.1 System configurations

The first part of every input script contains information of the system properties. This includes: size, number of dimensions, boundary conditions, potentials, unit convention, information on the containing atoms, time step length, neighbor list update frequency and possibly a lot more. Listing 3.1 show the configurations that has been used for the entire duration of this project. Below there is a description of each command and its purpose.

```
1 units          metal
2 boundary       p p p
3 atom_style     atomic
4 read_data      "system.data"
5 pair_style     vashishta
6 neighbor       0.3 bin
7 neigh_modify   delay 10
8 timestep       0.002
```

Listing 3.1: Typical system configurations applied in this project.

`units` determines what unit convention should be used for the simulation. It might not seem like a difficult decision, but if you import data from a file for instance, it is crucial that LAMMPS interpret the data correctly. In this project we have chosen *metal* as the unit convention. The units are therefore as shown in table B.1.

boundary sets the style of boundaries for the simulation box in each dimension. There are four options: **p**, **f**, **s** and **m**. In the example above, we have chosen to use periodic boundaries through all the faces of the simulation box. It's possible to set different conditions in separate dimensions, e.g. **boundary p f p** sets fixed boundaries on the faces normal to the y-direction. One can even set different conditions on the two faces in the same dimension by using two letters, e.g. **boundary p fs p**, where the first letter indicates the boundary to be used on the *low face* and the last on the *high face*.

atom_style tells LAMMPS the structure of atom related data stored in a data file. This includes information about particle types, positions, charges, mass, bonds, angles, and potentially more, depending on the **atom_style**.

read_data reads a data file containing information as described above and additional information about the system, such as its size and shape. This is one of three ways to distribute initial atom positions. Another is the **read_restart** command, which is used extensively to load saved states from restart files. The last option is to use **create atoms**, which distributes atoms in a predefined way, i.e. in a lattice or a random collection. To use this last option, one has to first create a simulation box using the **create box** command.

pair_style determines which potential to use in the simulation.

neighbor sets the additional range of the neighbor list cutoff, and the method of constructing the neighbor lists. The consequence of having a larger range for the neighbor lists is that there are more particles to check for force contribution, however, we don't have to update the neighbor lists as often.

neigh_modify determines how often to update the neighbor lists.

timestep self-explanatory.

3.2.2 Run-time commands

Variables

LAMMPS offer the ability to store values in variables. These can be constants or dependent on other variables, like the time step for instance. Declaration of variables is done using the following syntax.

```

1  variable A equal 1000
2  variable B equal step/${A}
3  variable C equal ${B}
```

```
4 variable D equal v_B
```

Listing 3.2: Declaration of variables.

All of the above are valid variables. The `step` variable is a LAMMPS standard for current time step. We attend for `B` to be a linear function running from 0 to 1 over the course of `A` time steps. `C` evaluates `B` at the time step it is declared and returns that value whenever called upon. Thus, `C` will be a constant value. `D` will evaluate `B` whenever called upon, and return the current value of variable `B`. The variables `C` and `D` are of course superfluous, since we could only use `B` directly. The user can use math operations on variables, such as: addition, subtraction, multiplication, division, as well as square roots, logarithms, exponentials, and lots more.

Region & Group

In almost all simulations it is necessary to define regions and groups in order to give certain parts of the domain special properties. If the user wish to create a sphere of atoms, the typical process is to start from a block of atoms, define a spherical region in the block and remove exterior atoms. The regions specified can take the shapes: block, cone, cylinder, plane, prism or sphere, and regions can be combined. An example of removing atoms within a combined region is shown in listing 4.3. This example also show how to assign a group to atoms within a region. Each atom can be part of up to 32 different groups at a time. The user can do tons of actions on groups. Most computes and fixes act on specific groups. There is only one standard group in LAMMPS; the group `all`.

Computes

A compute defines a computation that will be performed on a group of atoms. This does not affect the dynamics in any way. The returned values are instantaneous. That is, they are computed from information about atoms on the current time step. The returned values can be global, local or per-atom quantities. These can be retrieved by an output commands using the following syntax:

```
1 c_computeID
2 c_computeID[1]
3 c_computeID[*][2]
```

The first alternative would return all values (scalars, vectors, arrays) computed. The second would return only the first element of a vector, or row of an array, and the last one would return the second element of all rows of an array. Note that LAMMPS counts from 1, which also deviate from other established standards. In section 3.2.3 we will see this in several example, where we refer to the `compute com` command.

```
1 compute comID groupID com
```

This computes the position of the center of mass of all atoms within the specified group, and stores the result in a 3-element vector.

Fix

A fix is an operation that's applied to the system during time iteration. Examples include time evolution of atoms, i.e. updating their positions and velocities, using thermostats, applying external force on atoms, averaging values, etc. There are numerous fixes and computes in LAMMPS, and you can easily add new ones if you know the structure of the framework. Some Fixes also compute and store values that can be retrieved by output commands. For instance the `fix addforce` command

```
1 fix addforceID groupID addforce fx fy fz
```

adds a given force to all atoms within the specified group. This `fix` stores the total force acting on the group before the force was added. This is stored as a 3-element vector. Its values can be obtained in the same manner as for computes:

```
1 f_fixID
2 f_fixID [1]
3 f_fixID [*] [2]
```

3.2.3 Output

In order to benefit from our simulations, we need to be able to extract the data of interest somehow. Here we will present the four most basic types of output.

Thermodynamic output

prints computed values to the screen and logfile every N time steps. It is activated by the command

```
1 thermo N
```

where N should be replaced by the desired frequency of the output. This can be a variable. The default quantities given in the output are: time step, temperature, pairwise energy, molecular energy, total energy and pressure. The user can specify what values should be included by using the `thermo_style` command with the syntax of the following example.

```
1 thermo_style custom step c_comID v_Fz f_addforceID [3]
```


This line specifies that when the thermodynamic output is printed, it should contain the time step, all quantities returned by the compute `comID`, the current value of the variable `Fz` and the third element of the vector retrieved by the fix `addforceID`. The naming of the compute- and fix names can be whatever desired, even numbers. I just find it convenient to name them by what they represent.

Dump

Dump files store data about the state of atoms in a specified group with the given frequency. Below we illustrate two quite different dump commands.

```
1 dump dumpID1 all atom 100 name.dump
2 dump dumpID2 groupID custom 50 name*.dump id x y vx fx
```

Line 1 creates a file named `name.dump`. Every 100 time steps it writes data to this file in the *atom* format, i.e. `id type xs ys zs`, where `xs`, `ys` and `zs` are scaled coordinates relative to the size of the simulation box. Line 2 creates a file for every time step, where the asterisk in the name is replaced by the current time step. Also we define the format of the output to be unscaled x- and y-coordinates, as well as velocity and force in the x-direction.

Common visualization software can read most of the predefined formats in LAMMPS: *atom*, *xyz*, *cfg*, etc..

Fixes

Certain fixes can also write specified quantities to files. For instance the commands

```
1 fix timeAvgID all ave/time 100 5 1000 c_comID file name.avg
2 fix spatialAvgID all ave/chunk 100 10 1000 chunkID c_comID
  file vel.profile
```

do respectfully time- and spatial averaging of the values returned by the compute `comID`. The numbers indicate how many values should be sampled, number of time steps between each sample, and how often to write this average to file. Note that the last command refers to a compute `chunk`, which governs how to grid the system. There is also a `fix print` command that writes single-line output to screen or file with a prescribed frequency.

```
1 fix printID all print 100 "z-component of COM: c_comID[3]"
```

This acts practically like the `thermo` command.

Restart files

In many cases it is practical to save certain checkpoints in a simulation, in order to be able to start a new simulation from said point. The `restart` command does just that.

```
1 restart N name.*.restart
```

This command saves the state of the system every `N` time step to individual restart files, distinguishable by the number representing the time step of the save. By "state of the system" we mean positions and velocities of all atoms, information of what groups each atom belongs to, the size and shape of the simulation box, boundary conditions, potential, etc.. When starting a new simulation, one of these "snapshots" can be loaded using the `read_restart` command as illustrated below.

```
1 read_restart name.timeStep.restart
```

3.3 Visualization

The LAMMPS package does not provide high-quality visualization software. However, the default formats of the dump command are well known to most of the ones out there. A couple of so-called high-quality visualization tools are listed below:

- VMD
- AtomEye
- OVITO
- ParaView
- PyMol

In this section we will discuss how to visualize an MD simulation and some of the features of one of the softwares listed, namely Ovito. Lastly we will be presented with a new concept that is currently under development here at the University of Oslo; a software named Atomify.

3.3.1 OVITO

Chapter 4

Preparing a molecular dynamics simulation

We wish to construct a system consisting of **two** elements made out of silica: a slab and a sphere cap. In order to do this we need to generate the spacial position coordinates (x,y,z) of every single atom. Considering that we are making a system consisting of about 10^5 atoms, this is obviously not done manually. We have chosen to use a tool named *Moltemplate*¹, which is included in the LAMMPS distribution.

The main idea is to manually enter the coordinates of only the atoms in a unit cell of the material one wish to generate, and then simply copy this unit cell wherever desired. The software will shift the coordinates of the copied unit cell by the displacement from the original image. In addition it will generate files containing data such as which atoms they share bonds with, if any, and angles between such bonds.

4.1 Silica

Silica is a chemical compound also known as Silicon dioxide, having the chemical formula SiO_2 . It has several polymorph structures, the most common being quartz, which is one of the most abundant minerals in the Earth's crust. Other polymorphs include cristobalite, tridymite, coesite and more. For our purpose it is insignificant which one we choose. Once the material is melted, it is indifferent which configuration we started from, as long as the density is correct. In this project we will build the constituents of the system from a type of cristobalite named β -cristobalite. This is mainly because it has a simple structure and a cubical unit cell.

¹<http://www.moltemplate.org/index.html>

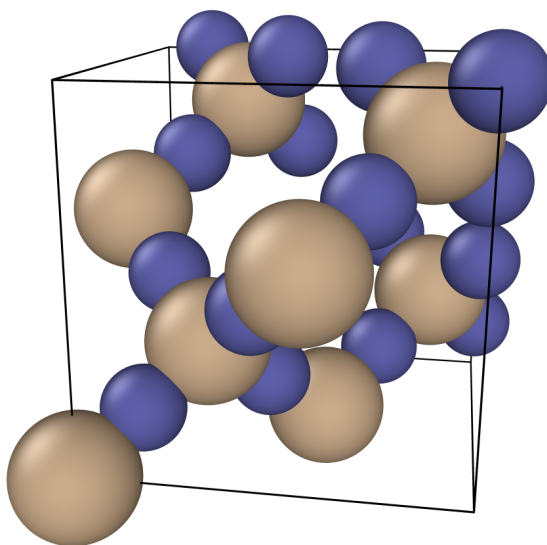


Figure 4.1: Unit cell of β -cristobalite. Tan and blue spheres represent silicon and oxygen atoms respectively. The unit cell is cubical, with edges of length 7.12\AA .

4.1.1 Unit cell of β -cristobalite

In order to construct the unit cell of a material, one should look up the coordinates of the atoms in a crystallography database. We have used the unit cell of β -cristobalite found at *Crystallography Open Database*². At this site one can download a `.cif`-file (Crystallographic Information Framework) containing information about the spatial positions of each atom, the length of the unit cell edges and angles between faces of the cell. In the case of β -cristobalite the unit cell is cubical with edges of length 7.12\AA . It contains 8 silicon atoms and 16 oxygen atoms. The density of the unit cell can easily be computed and is 2.2114 g/cm^3 . The format of the `.cif`-file is not easily readable. To extract the information we have used a tool named `cif2file`³, developed by Torbjörn Björkman.

4.2 Building a crystal

The coordinates gotten from the `.cif`-file can now be implemented into *moltem-plate* together with whatever bond and angle data required by the potential. In our simulations we will use the Vashishta potential, which does not require these.

²<http://www.crystallography.net/cod/1010944.html>

³<http://www.cif2cell.com-about.com/>

Moltemplate has its own structure and syntax. The first step to build up a larger material is, as mentioned, to create the unit cell. Data concerning the unit cell are placed in a `.lt`-file, which is readable by Moltemplate. Such a file is shown in Listing 4.1. For a more profound understanding of the structure and syntax of these files, the reader is advised to read the moltemplate manual⁴.

```

1  # file "beta-cristobalite.lt"
2
3  beta-cristobalite {
4    write("Data Atoms") {
5      $atom:Si1    @atom:Si    0.00    0.00    0.00
6      $atom:Si2    @atom:Si    0.00    3.56    3.56
7      $atom:Si3    @atom:Si    1.78    1.78    1.78
8      $atom:Si4    @atom:Si    3.56    0.00    3.56
9      $atom:Si5    @atom:Si    1.78    5.34    5.34
10     $atom:Si6    @atom:Si    5.34    5.34    1.78
11     $atom:Si7    @atom:Si    3.56    3.56    0.00
12     $atom:Si8    @atom:Si    5.34    1.78    5.34
13     $atom:O1     @atom:O     0.89    0.89    0.89
14     $atom:O2     @atom:O     6.23    4.45    2.67
15     $atom:O3     @atom:O     2.67    2.67    0.89
16     $atom:O4     @atom:O     4.45    0.89    4.45
17     $atom:O5     @atom:O     0.89    4.45    4.45
18     $atom:O6     @atom:O     4.45    4.45    0.89
19     $atom:O7     @atom:O     2.67    6.23    4.45
20     $atom:O8     @atom:O     2.67    0.89    2.67
21     $atom:O9     @atom:O     4.45    2.67    6.23
22     $atom:O10    @atom:O     6.23    2.67    4.45
23     $atom:O11    @atom:O     2.67    4.45    6.23
24     $atom:O12    @atom:O     0.89    6.23    6.23
25     $atom:O13    @atom:O     0.89    2.67    2.67
26     $atom:O14    @atom:O     4.45    6.23    2.67
27     $atom:O15    @atom:O     6.23    6.23    0.89
28     $atom:O16    @atom:O     6.23    0.89    6.23
29   }
30
31   write_once("Data Masses") {
32     @atom:Si 28.0855
33     @atom:O  15.9994
34   }
35 } # end definition of beta-cristobalite molecule type

```

Listing 4.1: Typical Moltemplate file containing unit cell data. The columns of the "Data Atoms" section hold, from left to right, information of atom ID, atom type, x-, y- and z-position. The "Data Masses" section stores the weight of silicon and oxygen atoms in atomic mass units.

We use the unit cell as building blocks, placing them concurrently until we

⁴http://www.moltemplate.org/doc/moltemplate_manual.pdf

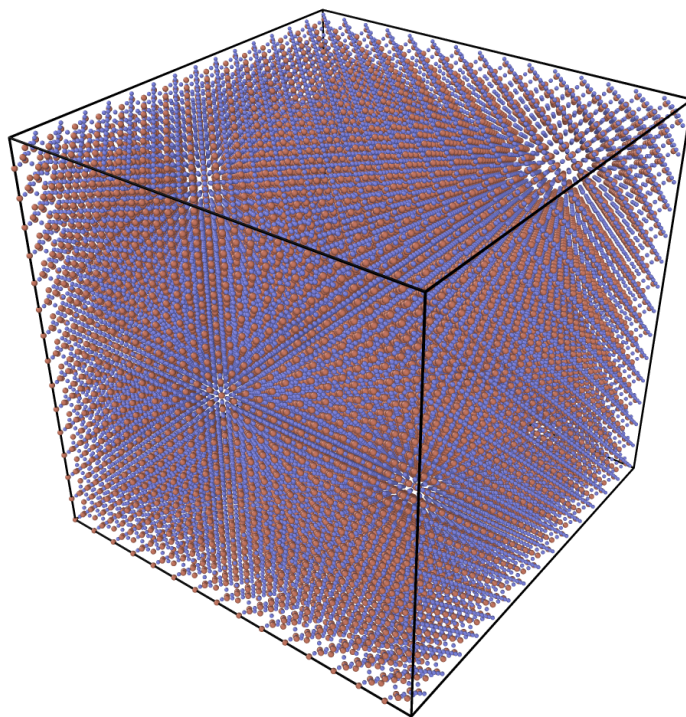


Figure 4.2: System built from $15 \times 15 \times 15$ unit cells of b-cristobalite.

have a crystal of the desired size. For our purpose, we generate a large cube of $15 \times 15 \times 15$ unit cells. This is done as follows.

4.3 Verifications

4.3.1 Melting point

Finding the melting point is a very important verification of our system. There are several factors that may affect the result. For instance, if the density of the system is too high, the melting point will be at a higher temperature than it normally would. Also, errors in the potential model may affect the temperature of the melting point. When increasing the temperature of the system that is in a solid state, we will eventually reach the melting point. At the phase transition from a solid state to a liquid the atoms of the silica will have energy great enough to break the interatomic bonds. They will break loose from their regular arrangement and move about much more freely. An approach to computing the melting point is therefore to systematically increase the temperature stepwise and sample the mean square displacement of the atoms at each designated temperature. The mean square displacement is the average of the square of the displacement every

atom has from its initial position. It can be expressed as:

$$\langle r^2(t) \rangle = \frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i(t) - \mathbf{r}_i(0))^2. \quad (4.1)$$

where $\mathbf{r}_i(t)$ is the position of atom i at time t and N is the total number of atoms. In practice the mean square displacement was computed using a standard *compute* in LAMMPS, namely the *msd compute*⁵. At every time step it stores a vector of 4 elements; the first 3 are the squared dx , dy and dz displacements averaged over the atoms of the specific group, while the 4th is the total mean square displacement for the specific group, i.e. $(dx^2 + dy^2 + dz^2)$. The $15 \times 15 \times 15$ system is sufficient for this test. The procedure is rather simple. For β -cristobalite we expect the melting point to be about 1986K⁶, so we start out at 1500K. After equilibration the following steps are repeated until we reach a target temperature.

- equilibrate for the current temperature
- compute the msd for N time steps
- increase the temperature by a given step size

```

1  label meltingLoop
2  variable i loop 11
3      fix nvtID all nvt temp ${T} ${T} 1.0
4      run ${N}
5      compute msdID all msd
6      fix msdDumpID all ave/time 1 1 3 c_msdID[4] file
          msd_N${N}_T${T}.txt
7      run ${N}
8      uncompute msdID
9      unfix nvtID
10     unfix msdDumpID
11     variable T equal ${T}+50
12 next i
13 jump SELF meltingLoop

```

Listing 4.2: Main blalbabla.

⁵http://lammps.sandia.gov/doc/compute_msd.html

⁶<https://en.wikipedia.org/wiki/Cristobalite>

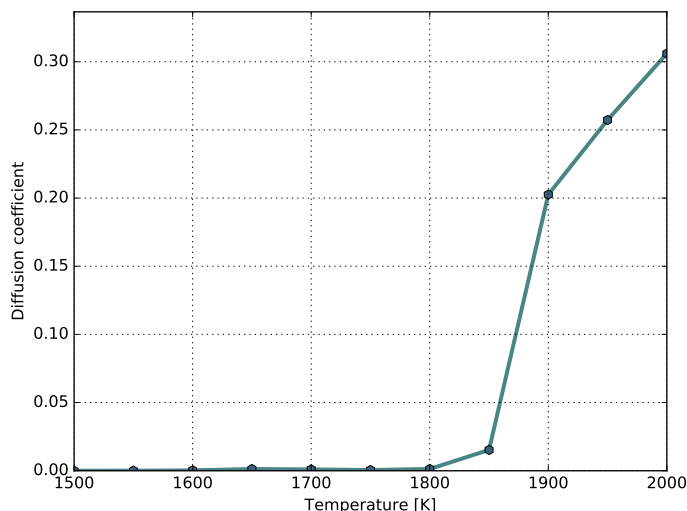


Figure 4.3: Mean square displacement as a function of temperature. The face transition occurs where we see a rapid change in the behavior.

This result infer a melting point somewhere in the range 1800K-1900K. Believe it or not, this is actually a pass on the test! In fact, anything deviating less than 20% from the experimental results is considered a pass, as stated by the developers of the potential. Unfortunately, this also show that there are some qualitative detail that is lost, and that when using this potential we should probably focus on the quantitative behavior.

4.4 Shaping the silica

The huge cube of silica can be carved however we like by defining regions from which we delete the containing atoms. In LAMMPS this is done using the `region`, `union`, `intersect` and `delete_atoms` commands. Our implementation is stated in Listing 4.3, which is very simple due to the way we are going to treat the boundary conditions. We start out by defining a spherical region labeled `sphereRegion`, described by the xyz-coordinates of its center and a radii. The atoms within this region are assigned to a group labeled `sphereGroup`. Next, we define a cuboid (block) region named `slabRegion`, described by the position of its faces in x-, y- and z-direction. The atoms within this region are assigned to a group, which we label `slabGroup`. We combine these two regions using the `union` command and label the region outside of these regions `outRegion`. Finally, we delete the atoms that are not in the sphere nor the slab; we delete the ones contained by `outRegion`.

```
1 region sphereRegion sphere 53.4 53.4 226.8 150
```



```

2  group  sphereGroup region sphereRegion
3
4  region slabRegion block 0 INF 0 INF 0 35.6
5  group  slabGroup region slabRegion
6
7  region outRegion union 2 sphereRegion slabRegion side out
8  delete_atoms region outRegion

```

Listing 4.3: Defining regions to keep or delete from a system of dimensions $106.8 \times 106.8 \times 106.8$ Å.

For the purpose of deleting atoms, the creation of groups was superfluous. However, at a later stage we will utilize them and this is an appropriate place for them to be assigned. The appliance of the script in Listing 4.3 on the system is shown in figure 4.2 is shown in figure 4.4, where our perspective is along the y-axis.

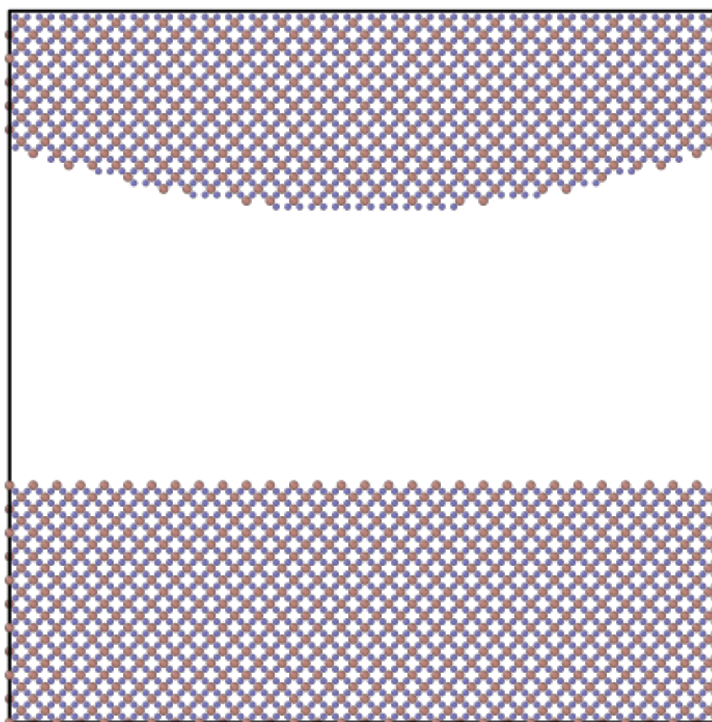


Figure 4.4: xz-perspective on a system built from $15 \times 15 \times 15$ unit cells of b-cristobalite, with certain regions carved out. This is a result from applying Listing 4.3 to the system shown in figure 4.2. The top shape is a sphere cap, while the bottom is a slab.

4.5 Moving the sphere towards the slab

We wish to push the sphere down onto the slab in order to create a deformation on the slab. There are probably a lot of smart ways to do this. The author has settled on the following strategy:

We apply periodic boundary conditions in all three dimensions. Secondly, we freeze the atoms at the bottom of the slab so that their positions are fixed. This will, due to the periodic boundary conditions, allow us to consider the sphere cap and the slab as if they are not connected to each other, but to independent blocks of silica glass. Then, for every N time steps we decrease the height of the system, while remapping the positions of the atoms. The remapping is a very important procedure. It ensures that we do not lose any atoms that otherwise would be lost when moving the z-boundary. A side-effect of the remapping is that the atoms in the system will be somewhat compressed in the z-direction. Though, if we do the compression slowly, this will not be of any concern.

```
1 fix freezeID groupID setforce 0 0 0
2 velocity groupID set 0 0 0
```

Listing 4.4: LAMMPS commands for hard coding the forces and velocities of atoms within a specific group. Effectively freezing them.

```
1 fix ID all deform 1 z delta 0 -${compressionLength} remap x
```

Listing 4.5: LAMMPS command for changing the size of the simulation box.

Chapter 5

This must be sorted in designated chapters

5.1 Radial distribution of normal force

In order to find a radial distribution of the normal force, F_N , we partition the system into a grid in the xy-plane. We then use the command

```
1 compute chunkID all chunk/atom bin/2d x 0 7.12 y 0 7.12
2 compute stressID all stress/atom NULL
3 fix fixChunkID all ave/chunk 1 1 10 chunkID
  c_stressID[3] file forcesInChunks.txt
```

to compute the stress of every chunk in the z-direction, σ_{zz} (sum of every individual atom stress in the chunk).

Line 1 establishes the grid, with bin width 7.12Å.

Line 2 creates a compute of the stress

Line 3 stores the sum of individual stresses in each chunk to the file `forcesInChunks.txt`.

This is done every 10 time steps in order to reduce correlation effects.

The data is stored from each time step can easily be averaged to produce a result as shown in figure X.

We can then find the radial distribution simply by binning this matrix in radial bins, and average the normal forces of the chunks within the bins.

5.1.1 Radial binning

Our system is partitioned into a grid. Each cell in the grid holds an averaged value of the normal force in that cell. The radial distribution of normal force should express the averaged value of the normal force at a given radial distance from the **center of the sphere**. A coarse method of doing this is to average the weights of the cells whose center is within the bin. This is illustrated in figure 5.1. In many applications where the bin size can be large compared to the length

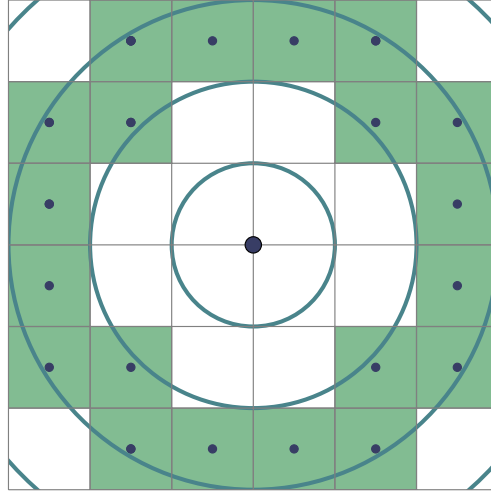


Figure 5.1: Coarse radial binning. The value appointed to the radial bin is the average of the weights of the cells whose center is within the bin. The cells with center within the third radial bin are colored, and their centers drawn.

of the cells, this method might suffice. However, the current radii of the contact area between the sphere and the slab is only about 10 unit cells, and therefore having a large bin width will result in very few data points.

A different, slightly more sophisticated approach is to compute the fraction of the area of each cell that actually is within the bin, and multiply this by the weight associated with the cell. We can sum all these contributions and average them by dividing by the area of the bin. This means that even cells that do not have their center within the bin might contribute to the resulting bin value. How much, however, will depend on the fraction of the cell that is within the bin. This may be regarded as a smoothing of our coarse force distribution, and will probably give a more correct result than the coarse binning method already described. An illustration of this binning method is shown in figure 5.2. The figure clearly shows that some cells have a larger area within the bin than others, and thus contribute more.

Computing the area of the cell within the bin should be described here!

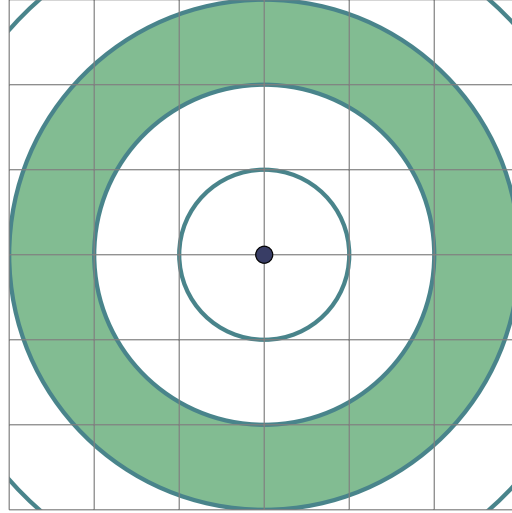


Figure 5.2: Radial binning based on weighted contributions of intersecting cells. The third radial bin is colored, and its value will be the average of the intersecting cells weight times the fraction of the cells area that intersects the bin.

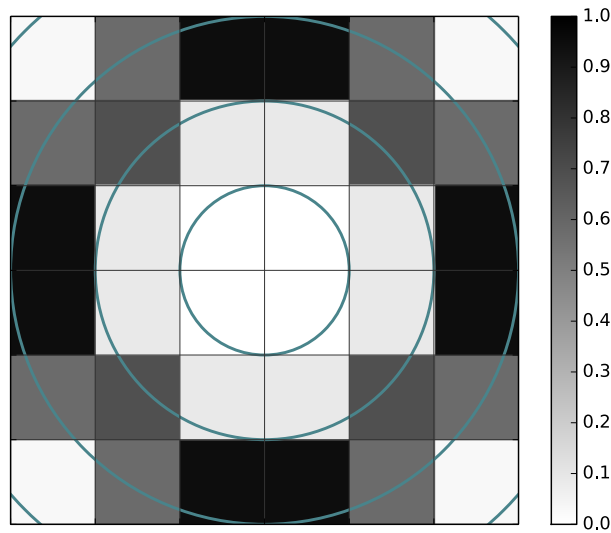
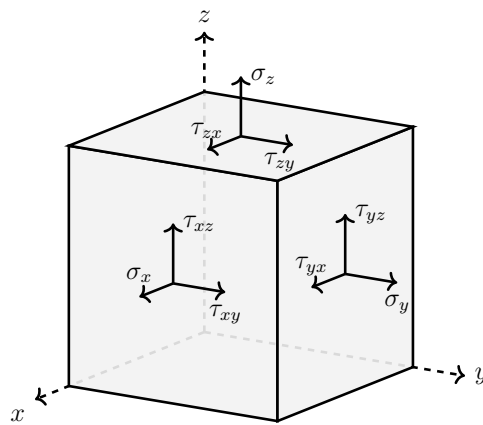


Figure 5.3: Fraction of cell's area that are within the third radial bin.

**Figure 5.4:** Stress

Chapter 6

Computing the normal force distribution

The normal force is defined as the force exerted on an object that is perpendicular to the contact surface. In this chapter we will make an attempt to find the distribution of the normal forces. This is not a trivial thing to compute in LAMMPS. As a matter of fact, to achieve this we have expanded the LAMMPS library by creating a custom compute class. The details of that procedure will be described.

Our strategy is simple, but not necessarily easy. First of, we divide the system into a grid. Secondly, we compute the average force exerted on one body from another within each cell. We approximate the slope of the contact surface within the cells using a least squares regression method. Finally, we project the average force of the atoms in a cell onto the normal vector of the cell.

6.1 Creating a custom compute

A *compute* is a LAMMPS command that defines a computation that will be performed on a group of atoms. The *computes* produce instantaneous values, using information about the atoms on the current time step.

In LAMMPS there are more than 100 computes and chances are, they already have what you're looking for. If not, one might treat the data from other computes in some way to get the desired information. However, if there are no compute command that does the desired task, it is possible to create an own custom class and thereby expanding the LAMMPS library.

In order to compute the normal forces acting on the sphere, we have written a custom compute class. The purpose of the class was to save the forces acting on atoms in one group from atoms of another group. In this section we will try to give brief instructions on how this was done.

[MAYBE AN ILLUSTRATIVE FIGURE HERE]

6.1.1 Find a similar compute

Obviously, before writing any code we should know what we want the compute to calculate and how this should be done. Before starting off with a blank sheet in the editor, one should definitely search for similar computes in LAMMPS. This can potentially save hours of hard work!

For instance there is a compute named *group/group*¹ which computes the total energy and force interaction between two groups of atoms. This is almost what we want, but we need to know the total force acting on all atoms from atoms of other groups. **It should also work with the Vashishta potential.**

Thus, there are minor modifications needed and because of the similarities we chose to make our compute a subclass of this one.

6.1.2 Creating the class

All computes in LAMMPS are subclasses of the class named *compute*. From this superclass they inherit a bunch of variables, functions and flags, which the user may decide to set. Functions are of course declared in the header file, while variables and flags are set in the source file. The source code of the *group/group* compute is shown in Appendix A.1. Since we will be making a subclass of it, we change the *private* property to *protected* so that we have access to all the variables and functions.

We start out by creating a header file and decide upon a name for our class. We have chosen the name *group/group/atom* since it is basically a per-atom version of the already existing compute *group/group*. A complete header-file is shown in Listing 6.2 and explained in detail below.

```

1  #ifdef COMPUTE_CLASS
2  ComputeStyle(group/group/atom, ComputeGroupGroupAtom)
3  #else
4
5  #ifndef LMP_COMPUTE_GROUP_GROUP_ATOM_H
6  #define LMP_COMPUTE_GROUP_GROUP_ATOM_H
7
8  #include "compute.h"
9  #include "compute_group_group.h"
10
11 namespace LAMMPS_NS {
12
13 class ComputeGroupGroupAtom : public ComputeGroupGroup {
14 public:
15     ComputeGroupGroupAtom(class LAMMPS *, int, char **);
16     ~ComputeGroupGroupAtom();
17     void compute_peratom() override;
18     int nmax;

```

¹http://lammps.sandia.gov/doc/compute_group_group.html


```

19     double **carray;
20
21     private:
22     void pair_contribution() override;
23 };
24 }
25 #endif
26 #endif

```

Listing 6.1: Header file of our new compute: `compute_group_group_atom.h`.

`ComputeStyle` defines the command to be used in the LAMMPS input script to be `group/group/atom`, and the name to be *ComputeGroupGroupAtom*. The name will be redundant to us.

`nmax` is the number of atoms which are subject to a non zero force from atoms of another group at the current time step; it may vary.

`carray` is a two dimensional array containing the force on atoms in one group induced by atoms of another group. Its dimension will necessarily be $nmax \times 3$.

`compute_peratom()` and `pair_contribution()` are functions which will be described below the corresponding source file.

```

1  #include <mpi.h>
2  #include <string.h>
3  #include "compute_group_group_atom.h"
4  #include "atom.h"
5  #include "update.h"
6  #include "force.h"
7  #include "pair.h"
8  #include "neighbor.h"
9  #include "neigh_request.h"
10 #include "neigh_list.h"
11 #include "group.h"
12 #include "kspace.h"
13 #include "error.h"
14 #include <math.h>
15 #include "comm.h"
16 #include "domain.h"
17 #include "math_const.h"
18 #include "memory.h"
19
20 using namespace LAMMPS_NS;
21 using namespace MathConst;
22
23 #define SMALL 0.00001
24
25 ComputeGroupGroupAtom::ComputeGroupGroupAtom(LAMMPS *lmp,
26     int narg, char **arg) :
27     ComputeGroupGroup(lmp, narg, arg),
28     carray(NULL),
29     nmax(0)

```

```

29 {
30     if (narg < 4) error->all(FLERR,"Illegal compute
        group/group command");
31
32     peratom_flag      = 1; // Indicating a peratom compute
33     size_peratom_cols = 4; // # of Columns per atom.
34     extarray          = 0; // 0/1 if global array is all
        intensive/extensive
35     scalar_flag       = 0;
36     vector_flag       = 0;
37 }
38
39
40 ComputeGroupGroupAtom::~~ComputeGroupGroupAtom()
41 {
42     memory->destroy(carray);
43 }
44
45
46 void ComputeGroupGroupAtom::compute_peratom()
47 {
48     // grow array if necessary
49     if (atom->nmax > nmax) {
50         memory->destroy(carray);
51         nmax = atom->nmax;
52         memory->create(carray, nmax, size_peratom_cols,
            "group/group/atom:carray");
53         array_atom = cararray;
54     }
55
56     if (pairflag) pair_contribution();
57     if (kspaceflag) kspace_contribution(); // This doesn't
        happen though. See compute_group_group.cpp
        constructor.
58 }
59
60
61 void ComputeGroupGroupAtom::pair_contribution()
62 {
63     int i,j,ii,jj,inum,jnum,itype,jtype;
64     double xtmp,ymtp,ztmp,dex,dely,dely;
65     double rsq,eng,fpair,factor_coul,factor_lj;
66     int *ilist,*jlist,*numneigh,**firstneigh;
67
68     double **x = atom->x;
69     int *type = atom->type;
70     int *mask = atom->mask;
71     int nlocal = atom->nlocal;
72     double *special_coul = force->special_coul;
73     double *special_lj = force->special_lj;
74     int newton_pair = force->newton_pair;

```

```

75     double *columns;
76
77     // invoke half neighbor list (will copy or build if
       necessary)
78
79     neighbor->build_one(list);
80
81     inum = list->inum;
82     ilist = list->ilist;
83     numneigh = list->numneigh;
84     firstneigh = list->firstneigh;
85
86     // loop over neighbors of my atoms
87     // skip if I,J are not in 2 groups
88
89
90     for (ii = 0; ii < inum; ii++) {
91         i = ilist[ii];
92
93         // skip if atom I is not in either group
94         if (!(mask[i] & groupbit || mask[i] & jgroupbit))
           continue;
95
96         xtmp = x[i][0];
97         ytmp = x[i][1];
98         ztmp = x[i][2];
99         itype = type[i];
100        jlist = firstneigh[i];
101        jnum = numneigh[i];
102
103        for (jj = 0; jj < jnum; jj++) {
104            j = jlist[jj];
105            factor_lj = special_lj[sbmask(j)];
106            factor_coul = special_coul[sbmask(j)];
107            j &= NEIGHMASK;
108
109            // skip if atom J is not in either group
110            if (!(mask[j] & groupbit || mask[j] &
              jgroupbit)) continue;
111
112            int ij_flag = 0;
113            int ji_flag = 0;
114            if (mask[i] & groupbit && mask[j] & jgroupbit)
              ij_flag = 1;
115            if (mask[j] & groupbit && mask[i] & jgroupbit)
              ji_flag = 1;
116
117            // skip if atoms I,J are only in the same group
118            if (!ij_flag && !ji_flag) continue;
119
120            delx = xtmp - x[j][0];

```

```

121         dely = ytmp - x[j][1];
122         delz = ztmp - x[j][2];
123         rsq = delx*delx + dely*dely + delz*delz;
124         jtype = type[j];
125
126         if (rsq < cutsq[itype][jtype]) {
127             eng = pair->single(i, j, itype, jtype,
128                               rsq, factor_coul, factor_lj, fpair);
129
130             // energy only computed once so tally full
131             // amount
132             // force tally is jgroup acting on igroup
133
134             if (newton_pair || j < nlocal) {
135                 array_atom[i][0] += eng;
136                 if (ij_flag) {
137                     array_atom[i][1] += delx*fpair;
138                     array_atom[i][2] += dely*fpair;
139                     array_atom[i][3] += delz*fpair;
140                 }
141                 if (ji_flag) {
142                     array_atom[j][1] -= delx*fpair;
143                     array_atom[j][2] -= dely*fpair;
144                     array_atom[j][3] -= delz*fpair;
145                 }
146
147                 // energy computed twice so tally half
148                 // amount
149                 // only tally force if I own igroup
150                 // atom
151                 }
152             else {
153                 array_atom[i][0] += 0.5*eng;
154                 if (ij_flag) {
155                     array_atom[i][1] += delx*fpair;
156                     array_atom[i][2] += dely*fpair;
157                     array_atom[i][3] += delz*fpair;
158                 }
159             }
160         }
161     }
162 }

```

Listing 6.2: Source file of compute: compute_group_group_atom.cpp.

In the constructor we set specific flags that LAMMPS uses to interpret what structure our data should have, and how to store them. We set the `peratom_flag` to be `True`, which indicates that we desire to store some data for each atom. `size_peratom_cols` defines the number of data values to store for each atom. Also, we set the `scalar_flag` and `vector_flag` to `False`, since we do not wish

to return a vector or scalar value.

Following the constructor is the destructor on line 40. Its only task is to free the memory occupied by the array once it is no longer needed.

`compute_peratom()` will resize the array to the number of atoms of concern, `nmax`. It does this using LAMMPS internal functions, which we will not care to describe here. Finally it calls upon functions

I ENDED HERE
LAST TIME!

Note that we should only compute the force for one of the groups. factor 2...

6.2 Least squares regression

The method of least squares aims to find parameters which minimize the sum of the squared residuals, where residuals are the difference between observed values and the approximated value. We will use this method to approximate the slope of the surface of the substrate. This will be done by partitioning the system in a grid and do a plane approximation on each cell of the grid. In other words, we seek the coefficients in the plane equation

$$z = ax + by + c \quad (6.1)$$

that minimizes the sum of the squared residuals

$$S = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n (z_i - f(x_i, y_i, \boldsymbol{\beta}))^2, \quad (6.2)$$

where $f(x_i, y_i, \boldsymbol{\beta})$ is the right hand side of the plane equation and $\boldsymbol{\beta}$ is the set of coefficients. The minima has the property that the differential with respect to any coefficient is zero.

$$\frac{\partial S}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial r_i^2}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial r_i^2}{\partial r_i} \frac{\partial r_i}{\partial \beta_j} = -2 \sum_{i=1}^n r_i \frac{\partial f(x_i, y_i, \boldsymbol{\beta})}{\partial \beta_j} = 0, \quad \forall \beta_j \in \boldsymbol{\beta} \quad (6.3)$$

When approximating a plane we have three coefficients to account for: a , b and c . This leaves us with the following set of equations:

$$-2 \sum_{i=1}^n (z_i - ax_i - by_i - c) \frac{\partial}{\partial a} (ax_i + by_i + c) = 0 \quad (6.4)$$

$$-2 \sum_{i=1}^n (z_i - ax_i - by_i - c) \frac{\partial}{\partial b} (ax_i + by_i + c) = 0 \quad (6.5)$$

$$-2 \sum_{i=1}^n (z_i - ax_i - by_i - c) \frac{\partial}{\partial c} (ax_i + by_i + c) = 0, \quad (6.6)$$

which corresponds to

$$\sum_{i=1}^n z_i x_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i y_i + c \sum_{i=1}^n x_i \quad (6.7)$$

$$\sum_{i=1}^n z_i y_i = a \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n y_i^2 + c \sum_{i=1}^n y_i \quad (6.8)$$

$$\sum_{i=1}^n z_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n y_i + nc. \quad (6.9)$$

This can be expressed as a matrix equation.

$$\begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i \\ \sum x_i & \sum y_i & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix} \quad (6.10)$$

where we have omitted the indices to better readability. Solving this linear system retrieves the optimal coefficients in the sense of the least squares method. The normal vector of the plane will be $\mathbf{n} = [a, b, 1]$. This vector will be used to compute the size of the normal force. Since we know the average force on an atom in the chunk, and the normal vector from the approximated slope of the surface, we can compute the normal force simply as

$$\mathbf{F}_N = |\mathbf{F}| \cos \theta \frac{\mathbf{n}}{|\mathbf{n}|}. \quad (6.11)$$

The cosine of the angle between the two vectors is given as

$$\cos \theta = \frac{\mathbf{F} \cdot \mathbf{n}}{|\mathbf{F}| \cdot |\mathbf{n}|}, \quad (6.12)$$

meaning that the normal force may be expressed as

$$\mathbf{F}_N = (\mathbf{F} \cdot \mathbf{n}) \frac{\mathbf{n}}{|\mathbf{n}|^2}. \quad (6.13)$$

We will assume that the normal vector \mathbf{n} is always in the same general direction as the average force, though obviously it may just as well point in the opposite direction and still be a normal vector to the plane. Programatically this was done in python as shown in Listing 6.3.

```

1      def getAngle(self, v1, v2):
2          'Computes angle between a vector and a line
          parallel to another vector'
3
4          lv1 = np.linalg.norm(v1)
5          lv2 = np.linalg.norm(v2)
6
7          angle = np.arccos(np.dot(v1, v2)/(lv1*lv2))
8          angle = min(angle, abs(np.pi-angle))

```

Listing 6.3: Python function to compute the smallest angle between a vector and a line parallel to another vector.

Appendix A

Source code

A.1 compute_group_group.h

```

1  #ifdef COMPUTE_CLASS
2  ComputeStyle(group/group, ComputeGroupGroup)
3  #else
4
5  #ifndef LMP_COMPUTE_GROUP_GROUP_H
6  #define LMP_COMPUTE_GROUP_GROUP_H
7
8  #include "compute.h"
9
10 namespace LAMMPS_NS {
11
12 class ComputeGroupGroup : public Compute {
13 public:
14     ComputeGroupGroup(class LAMMPS *, int, char **);
15     ~ComputeGroupGroup();
16     void init();
17     void init_list(int, class NeighList *);
18     double compute_scalar();
19     void compute_vector();
20
21 protected: // private
22     char *group2;
23     int jgroup, jgroupbit, othergroupbit;
24     double **cutsq;
25     double e_self, e_correction;
26     int pairflag, kspaceflag, boundaryflag;
27     class Pair *pair;
28     class NeighList *list;
29     class KSpace *kspace;
30
31     virtual void pair_contribution();
32     void kspace_contribution();
33     void kspace_correction();
34 };
35
36 }
37
38 #endif
39 #endif

```

A.2 compute_group_group_atom.h

```
1  #ifndef COMPUTE_CLASS
2  ComputeStyle(group/group/atom,ComputeGroupGroupAtom)
3  #else
4
5  #ifndef LMP_COMPUTE_GROUP_GROUP_ATOM_H
6  #define LMP_COMPUTE_GROUP_GROUP_ATOM_H
7
8  #include "compute.h"
9  #include "compute_group_group.h"
10
11  namespace LAMMPS_NS {
12
13  class ComputeGroupGroupAtom : public ComputeGroupGroup {
14  public:
15      ComputeGroupGroupAtom(class LAMMPS *, int, char **);
16      ~ComputeGroupGroupAtom();
17      void compute_peratom() override;
18      int nmax;
19      double **carray;
20
21  private:
22      void pair_contribution() override;
23  };
24  }
25  #endif
26  #endif
```

A.3 compute_group_group_atom.cpp

```

1  #include <mpi.h>
2  #include <string.h>
3  #include "compute_group_group_atom.h"
4  #include "atom.h"
5  #include "update.h"
6  #include "force.h"
7  #include "pair.h"
8  #include "neighbor.h"
9  #include "neigh_request.h"
10 #include "neigh_list.h"
11 #include "group.h"
12 #include "kspace.h"
13 #include "error.h"
14 #include <math.h>
15 #include "comm.h"
16 #include "domain.h"
17 #include "math_const.h"
18 #include "memory.h"
19
20 using namespace LAMMPS_NS;
21 using namespace MathConst;
22
23 #define SMALL 0.00001
24
25 ComputeGroupGroupAtom::ComputeGroupGroupAtom(LAMMPS *lmp,
26   int nargs, char **arg) :
27   ComputeGroupGroup(lmp, nargs, arg),
28   carray(NULL),
29   nmax(0)
30 {
31   if (nargs < 4) error->all(FLERR,"Illegal compute
32     group/group command");
33
34   peratom_flag      = 1; // Indicating a peratom compute
35   size_peratom_cols = 4; // # of Columns per atom.
36   extarray          = 0; // 0/1 if global array is all
37     intensive/extensive
38   scalar_flag       = 0;
39   vector_flag        = 0;
40 }
41
42 ComputeGroupGroupAtom::~ComputeGroupGroupAtom()
43 {
44   memory->destroy(carray);
45 }

```

```

46 void ComputeGroupGroupAtom::compute_peratom()
47 {
48     // grow array if necessary
49     if (atom->nmax > nmax) {
50         memory->destroy(carray);
51         nmax = atom->nmax;
52         memory->create(carray, nmax, size_peratom_cols,
53             "group/group/atom:carray");
54         array_atom = cararray;
55     }
56
57     if (pairflag) pair_contribution();
58     if (kpaceflag) kspace_contribution(); // This doesn't
59     happen though. See compute_group_group.cpp
60     constructor.
61 }
62
63 void ComputeGroupGroupAtom::pair_contribution()
64 {
65     int i,j,ii,jj,inum,jnum,itype,jtype;
66     double xtmp,ymtp,ztmp,dely,delz;
67     double rsq,eng,fpair,factor_coul,factor_lj;
68     int *ilist,*jlist,*numneigh,**firstneigh;
69
70     double **x = atom->x;
71     int *type = atom->type;
72     int *mask = atom->mask;
73     int nlocal = atom->nlocal;
74     double *special_coul = force->special_coul;
75     double *special_lj = force->special_lj;
76     int newton_pair = force->newton_pair;
77     double *columns;
78
79     // invoke half neighbor list (will copy or build if
80     necessary)
81
82     neighbor->build_one(list);
83
84     inum = list->inum;
85     ilist = list->ilist;
86     numneigh = list->numneigh;
87     firstneigh = list->firstneigh;
88
89     // loop over neighbors of my atoms
90     // skip if I,J are not in 2 groups
91
92     for (ii = 0; ii < inum; ii++) {
93         i = ilist[ii];

```

```

93 // skip if atom I is not in either group
94 if (!(mask[i] & groupbit || mask[i] & jgroupbit))
95     continue;
96
97 xtmp = x[i][0];
98 ytmp = x[i][1];
99 ztmp = x[i][2];
100 itype = type[i];
101 jlist = firstneigh[i];
102 jnum = numneigh[i];
103
104 for (jj = 0; jj < jnum; jj++) {
105     j = jlist[jj];
106     factor_lj = special_lj[sbmask(j)];
107     factor_coul = special_coul[sbmask(j)];
108     j &= NEIGHMASK;
109
110     // skip if atom J is not in either group
111     if (!(mask[j] & groupbit || mask[j] &
112         jgroupbit)) continue;
113
114     int ij_flag = 0;
115     int ji_flag = 0;
116     if (mask[i] & groupbit && mask[j] & jgroupbit)
117         ij_flag = 1;
118     if (mask[j] & groupbit && mask[i] & jgroupbit)
119         ji_flag = 1;
120
121     // skip if atoms I,J are only in the same group
122     if (!ij_flag && !ji_flag) continue;
123
124     delx = xtmp - x[j][0];
125     dely = ytmp - x[j][1];
126     delz = ztmp - x[j][2];
127     rsq = delx*delx + dely*dely + delz*delz;
128     jtype = type[j];
129
130     if (rsq < cutsq[itype][jtype]) {
131         eng = pair->single(i, j, itype, jtype,
132             rsq, factor_coul, factor_lj, fpair);
133
134         // energy only computed once so tally full
135         // amount
136         // force tally is jgroup acting on igroup
137
138         if (newton_pair || j < nlocal) {
139             array_atom[i][0] += eng;
140             if (ij_flag) {
141                 array_atom[i][1] += delx*fpair;
142                 array_atom[i][2] += dely*fpair;
143                 array_atom[i][3] += delz*fpair;
144             }
145         }
146     }
147 }

```

```
138         }
139         if (ji_flag) {
140             array_atom[j][1] -= delx*fpair;
141             array_atom[j][2] -= dely*fpair;
142             array_atom[j][3] -= delz*fpair;
143         }
144
145         // energy computed twice so tally half
146         // amount
147         // only tally force if I own igroup
148         // atom
149         }
150         else {
151             array_atom[i][0] += 0.5*eng;
152             if (ij_flag) {
153                 array_atom[i][1] += delx*fpair;
154                 array_atom[i][2] += dely*fpair;
155                 array_atom[i][3] += delz*fpair;
156             }
157         }
158     }
159 }
```


Appendix B

Something

B.1 LAMMPS units

Table B.1: Unit convention of the *metal* unit style in LAMMPS.

Mass	grams/mole
Distance	Angstroms
Time	picoseconds
Energy	eV
Velocity	Angstroms/picosecond
Force	eV/Angstrom
Torque	eV
Temperature	Kelvin
Pressure	bars
Dynamic viscosity	Poise
Charge	multiple of electron charge
Dipole	charge · Angstroms
Electric field	volts / Angstrom
Density	gram / cm ³

Bibliography

- [1] Bin Luan and Mark O Robbins. The breakdown of continuum models for mechanical contacts. *Nature*, 435(7044):929–932, 2005.
- [2] Knut Mørken. Numerical Algorithms and Digital Representation, 2016.