Computer Lab 3a - Monte Carlo Simulations

During today's lab you will be going back and forth between the Mac and the cluster. You will use **ssh** to remotely log on to the cluster and do a variety of tasks, you will use **sftp** to fetch files from the cluster to the Mac, and you will do a variety of tasks directly on the Mac. Many people find it useful to have several terminal windows open at the same time. Keep one for your **ssh** session on the cluster, one for doing **sftp** "gets" and "puts" and one to do things directly on the Mac. Of course it is helpful if you go to the same subdirectory on the Mac in each of these windows before entering any commands.

Answer the questions highlighted in yellow and send the answers to **achin14@jhu.edu** either in the body of an email or as an attached file with your JHEDID in the name (e.g. *JHEDID_lab1b.txt* if made with vi or *JHEDID_lab1b.docx* if made with MSWord).

I. **Anatomy of a Monte Carlo simulation program**. You will first obtain the files to be used in this lab, copy them to your working directory on the cluster and look at the contents of the following files:

- *modules.py* - a file containing different subroutines called by the main program
- *mc_nvt.py* - the main program file
- *mc_params* - a file listing the parameters of the simulation
- *average.py* – a python script for calculating means and other statistics of data
- *rdf_calc.py* – a python script for calculating radial distribution functions
- *normal-python3* – a batch control file for submitting python jobs to the job queue on the cluster

1. Log on to a Mac, change to your home directory or flashdrive, make a new directory called *lab3a*, change to that directory. This will be your main Mac working directory today.

2. Open a second terminal window on the local machine and **`ssh`** to your account on the cluster (**`compbio2@kirin.kit.jhu.edu`**), then **`cd JHEDID`**.

3. To set up your UNIX shell environment on the cluster just enter the following,

```
tcsh
```

The terminal app window preferences only work on the Mac, not on kirin. Just give the command every time you ssh to kirin (the .tcshrc file is there but you can't see it).

4. Make a subdirectory on the cluster for today's lab (below your JHEDID directory) and **cd** there

```
mkdir lab3a
cd lab3a
```

5. Get the tar archive of files for this lab, unpack it in your *lab3a/* directory on the cluster and change directories to the new unpacked directory called *mc*.


```
cp /home/compbio2/Shared/mc.tar .
tar xvf mc.tar
cd mc
cp /home/compbio2/Shared/normal-python3 .
```

1. How do you find out more about the **tar** command?

6. *Stop while we discuss some highlights of these files in class*. You can use **more**, **less** or **view** to look at the contents of the files.

II. **Run equilibration simulation MC_NVT on Argon box**

1. Edit the *normal-python3* file. Change the following line so that the word `yourJHEDID` is really your actual JHED_ID. Leave the `#$ -N` as is.

   ```
   #$ -N yourJHEDID
   ```

2. Make a starting system of 144 Argon molecules called *init.pdb* using the program *lattice.py*. To do this just run the *normal-python3* batch control file with the following command,

   ```
   ./normal-python3
   ```

3. The script will only take a few minutes to run and you should see a new file, *init.pdb*, appear in your directory. Check that *init.pdb* looks like a PDB file in format. You will also see some new files with the names yourJHEDID.o***** and yourJHEDID.p*****. You should remove these with the following command,

   ```
   rm yourJHEDID.*
   ```

4. Now that you have a starting structure (*init.pdb*) you can run the Monte Carlo program to equilibrate the system. Edit the *normal-python3* file so the command under the label, `Run the job` is,

   ```
   $PYTHONBIN mc_nvt.py mc_params init.pdb
   ```

This command translates to: Use Python to run your code (*mc_nvt.py*) with input parameter file *mc_params*, and input system configuration file *init.pdb*.

3. Submit your job and check to see if your job is running,

   ```
   ./normal-python3
   qstat –u \*
   ```

You can re-enter the **qstat** command anytime to see if your job is still running. When your job has finished (it will take a few minutes) view the *yourJHEID.o***** file on the cluster using the **`less`** command. Look at the progression of maximum displacement distances and the progression of fractional

acceptances by the Metropolis criteria. Compare the relationship between fraction of attempts accepted and magnitude of displacement for each move. It appears that the displacement is increased until the acceptance rate is ~50%. 2. Why don't we set the algorithm to give us a larger percent acceptance (i.e. use smaller displacement)?
You should have several new configuration files, *ensemMC.pdb*, *finalMC.pdb* and *restart.pdb* and a file with the energy, pressure and density values, *epd.dat*.

III. **View equilibration simulation results with molecular graphics**

1. Fetch the *init.pdb, ensemMC.pdb, epd.dat* and *average.py* files back to the Mac. On the Mac from the directory called *lab3a/* use the following commands.

```
sftp compbio2@kirin.kit.jhu.edu
cd JHEDID/lab3a/mc
get init.pdb
etc.
bye
```

When you have the *init.pdb* file back on the Mac type these commands,

```
pymol init.pdb
```
init | S | spheres

3. What is the average inter-atomic distance?

Wizard | Measurement
(Click on two adjacent atom centers)

Make a note of the starting interatomic distance and quit PyMOL.
4. What do you expect to happen to this lattice during your simulation?

Now use PyMOL to view the ensemble created during your equilibration run. Color the spheres with a spectrum by residue number,

```
pymol ensemMC.pdb
```
ensemMC | S | spheres
ensemMC | C | spectrum | rainbow

set the frame rate to 15 FPS

Movie | Frame Rate | 15 FPS

and play the movie.

5. Why do many of the blue spheres suddenly appear on the other side of the system?

Stop the movie (■) and go to the last frame (▶|).

6. Are the colors well-mixed (an indication of equilibrium)?
7. What are some inter-atomic distances (in Å)?
(Hint: Use the Wizard | Measurement menu to measure several different adjacent atomic distances.)
8. What is the ideal (minimum energy) atomic distance for this system?
9. Why aren't all atomic distances at the minimum energy distance?

Now would be a good time to skip ahead to section V and start your production run. Them come back and do section IV.

IV. **Analyze equilibrium ensemble**

1. Plot the energy versus cycle number.

```
awk '{print $1,$2}' epd.dat > energy.dat
```

Edit *energy.dat* using **vi** to delete (**dd**) the first two lines (the column labels), then

```
xmgrace energy.dat &
```

10. Does the system appear to have reached equilibrium?
11. What should the plot look like at equilibrium?

2. Calculate block statistics from the trajectory.
Enter the following commands to obtain different blocks of data from the *epd.dat* file. 12. What does each command and argument in a line below do?

```
sed -n '1,200p' energy.dat | awk '{print $2}' > 1.dat
sed -n '201,400p' energy.dat | awk '{print $2}' > 2.dat
sed -n '401,600p' energy.dat | awk '{print $2}' > 3.dat
sed -n '601,800p' energy.dat | awk '{print $2}' > 4.dat
```

Then enter the following commands to calculate statistics on each successive block.

```
python3 average.py 1.dat
python3 average.py 2.dat
python3 average.py 3.dat
python3 average.py 4.dat
```

Inspect the means and standard deviations for the energy values in each block. 13. Does the system appear to have reached equilibrium?

Let's assume that the system is at equilibrium and go on to the next step.

V. **Restart simulation for production of Boltzmann ensemble.**

On the cluster edit the *mc_params* file to run for 6000 cycles.
Then edit the *normal-python3* file to use the *restart.pdb* file instead of *init.pdb*.
I.e.,

```
$PYTHONBIN mc_nvt.py mc_params restart.pdb
```


Then submit your job to the queue,

```
./normal-python3
```

Go back to section IV if you skipped ahead. When the production run finishes fetch the *ensemMC.pdb* and *epd.dat* files back to the Mac (Note: This will overwrite the existing files).
View the simulation ensemble in PyMOL.
Does the simulation appear stable?
Are the colors well-mixed?

Plot the energy versus cycle number for the production run.
14. Does the system appear to have maintained a stable energy value with small fluctuations?
Calculate and draw a linear regression line through the data:

- Click Data | Transformations | Regression
- Click Load: Function
- Enter the following values: Start load at: **1**, Stop load at: 5**999**, # of points: **100**
- Click Accept
- Expand and inspect the Console that appears to see the slope of the fitted line.

15. What would be the slope of the fitted line at equilibrium and how does your calculated slope compare?

If everything points to the system being stable and fully equilibrated, you should have a Boltzmann distribution of Argon molecules and this ensemble would be adequate for testing hypotheses about Argon under the conditions of the simulation.

Congratulations! You have just run a molecular simulation.

Save the file *ensemMC.pdb* from the production run because you will use it in the next lab.

Historical Note: In 1961 Richard Feynman took over the introductory course in physics at Caltech and his lecture notes were later collected and became the cultural classic, *The Feynman Lectures on Physics*. In the first lecture he said the following:

If, in some cataclysm, all scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis (or atomic fact, or whatever you wish to call it) that all things are made of atoms – little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another. In that one sentence you will see an enormous

amount of information about the world, if just a little imagination and thinking are applied.

Your simulation today was a demonstration of that concept.

