Molecular Dynamics tutorial lab

Eduardo M. Bringa (ebringa@yahoo.com) and Diego Tramontina (diego.tramontina@gmail.com)

I) Simple performance issues

Computational cost of MD simulations is usually measured by how much time (wall clock) does it take a given simulation, divided the number of atoms and the steps run in that simulation (TPAS). For simple pair potentials, TPAS is in the range ~5-50 μs/atom/step. For complex potentials, or long range potentials it can be 1-2 orders of magnitude longer. You need to make sure you run enough steps to get meaningful timing, i.e. at least 1,000 steps.

0) Warm-up

a) We will focus on the LAMMPS MD software (http://lammps.sandia.gov), and run the version installed in the local cluster. Copy everything needed to your local dir, including the required input file, potential and data files, etc.

b) Please bring output from the code from the cluster to your local machine for plotting/analysis.

1) Lennard-Jones (LJ) simulations (without MPI)

Every MD tutorial should start with Lennard-Jones (LJ) simulations ☺

$$V(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$

You should see the input scripts in dir "melt". This script creates a LJ solid, and then increases the temperature above the melting point. The script also allow to obtain the diffusivity of the liquid, by calculating the mean square displacement $<r_i^2(t)>$, where <> indicates average over all N particles, i=1,..,N.

There is one sample submission script, which you need to change according to which input and run mode you want to use.

a) Run the input script, modifying the submission script as needed and saving the logs. Use a single core. What is the TPAS? Note the force loop timing, neighbor timing, etc.

b) Now increment the number of steps from 1,000 to 10,000. How much does the "cost" of the simulation changes? Why?

c) Now increment the number of atoms in the simulation by a factor of 2^3 (box size changes a factor of 2 along each direction) going back to 1,000 steps. Does the "cost" of the simulation changes? Why? What happens if you simulate an elongated box, where only one of the dimensions is extended 2^3?

d) What happens if your neighbor “skin” is doubled? Can you comment on cases where you would need that?

e) Using your favorite plotting program -for instance, gnuplot (short tutorial at http://people.virginia.edu/~lz2n/mse627/Eduardo/)-, plot the total, kinetic and potential energies from the different logs. Are they different? Are they statistically different (calculate mean and standard deviation using for instance OpenOffice)?

f) Test how much the timing is affected by the disk I/O, un-commenting the “dump” commands in the script. Note that most benchmarks showing TPAS or equivalent do not include disk I/O, which will be needed in production runs. In this case you would be writing to a single file, but LAMMPS is capable to writing to one file/process to achieve better parallel efficiency. This starts mattering seriously for runs with +100 cores.

g) Calculate the diffusion coefficient, D, for three different temperatures. In 3D, for “normal” diffusion, D=MSD(t)/6t. How long do you need to simulate for reliable results? What are the values of the diffusion coefficient if the temperatures are below melting?

2) Parallelization: Lennard-Jones (LJ) simulations with MPI

Run the input script, modifying the submission script as needed and saving the logs. Use only 1000 steps.

a) Run in 2, 3, 4, 6, 8, 12, 16 cores if you run before in a single core. Do a table with the timing for the different cases. For a “strong” scaling case like this one, how good is the parallelization scaling?

b) Do a “weak” scaling test, increasing the system size by a factor of 2, 4, 8, 16, and running case (a). Is the scaling good? What is the maximum number of atoms which can be simulated due to memory constrains?

3) OPTIONAL: GPU Lennard-Jones (LJ) simulations

Run the input script, modifying the submission script as needed to run in the TESLA and VOLTA GPUs, saving the logs. Use only 1000 steps.

a) How does the timing compare with the timing for CPUs? How many CPU cores are needed to reach the performance of a single GPU?

b) Do a “weak” scaling test, increasing the system size by a factor of 2, 4, 8, 16, and running case (2.a). Is the scaling good? What is the maximum number of atoms which can be simulated due to memory constrains?