

Rapid Parameterisation of Small Molecules

Dr Benjamin Roberts
[\(ben.roberts@nesi.org.nz\)](mailto:ben.roberts@nesi.org.nz)



Outline

① Introduction

② Technical Overview

③ Case Studies

presentation

Eshan's Simulations

Nick's networks

Molecular Mechanics

- An approximation of a molecule's potential energy using a Newtonian model
- Add a time component to get molecular dynamics (MD)
- The system is represented in three parts: Coordinates, topology and force field
- Coordinates: Where all the particles are (arbitrary frame of reference)
- Topology: Which particles are connected to which other particles
- What is a force field?

Force fields



A force field is an approximate mathematical model of the potential energy surface of the system. It can be expressed as an equation in $3N - 6$ variables, where N is the number of particles.

Force fields

Functional form

- The form of the equations used to model the various types of interaction that contribute to the overall potential energy (bond stretching, angle bending, etc.)
- All interactions of the same type will use the same basic model (e.g., bonds commonly use a quadratic equation in the manner of Hooke's law)

Parameters

- The constant terms

Technical Overview

There are three structured approaches to run our jobs on a MPI machine

- ① MPMD
- ② batcher
- ③ custom (usually a form of the previous two but integrated within your workflow)

We will use three case studies to show the problems and the solutions adopted.

Case Studies

Presentation

We'll go through the work of three researchers showing various problems and solutions and we'll summarize their pros and cons.

- ① Ehsan's simulations (the flood)
- ② Nick's network (MPMD in testing)
- ③ Jing Wang: DNA matching (batcher)

Ehsan's Simulations

Ehsan is a PhD student in computer science. His problem is simulating wireless networks to improve the efficiency of the communication in mobile devices.

- first submission to the cluster was 40,000 simulations
- done 4 at a time
- first reaction: increase the number of simulations he can run (after all they are not big MPI jobs)
- Bad idea: the simulations were spread all over the cluster blocking big MPI jobs
- developed a custom solution to contain him to a set of nodes
- we later found out that it was equivalent to using MPMD

We now identify high-throughput project quickly and offer them better solutions as soon as possible.

Nick's Networks

- Nick Baker is a PhD student in biological sciences with a background in mathematics. He studies ecological networks (complex food chains).
- His work focuses on analysing networks to find "keystone" species that are essential to the network. The work is what we call embarrassingly parallel, he has to go through a high number of networks and perform his analyses on each and everyone of them.
- I spotted him on the cluster running four jobs at a time out of close to a hundred. There was definitely room for improvement. He became a test subject for MPMD

MPMP: Multiple Program Multiple Data

SMPD: Instances of a program process the data distributed amongst them.

MPMD: Programs, not necessarily identical, process distinct data.

Questions & Answers

