CMSC416: Introduction to Parallel Computing

Topic: Parallel CSE Applications
Date: April 30, 2024

Molecular Dynamics Overview

- Parallel Computing can be used to calculate trajectories of atoms and molecules by solving Newton's equations of motions
- A molecular system is a 3-dimensional simulation box with some number of atoms
- In a simulation, you have to find the forces on each number of atoms and figure out acceleration to find the new position of the atom.
- Whenever there is a bond between atoms you have to calculate the force of one atom on another, which include electrostatic forces which are related to the charge of the atom, and Van der Waal's forces, which are everything else not related to bonds or atom charges.
- You have to calculate all of this to find the net force of atoms, and divide by the mass of the atom to find the acceleration.
- There are thousands to millions of atoms in a simulation.
- A simulation step is 1 femtosecond, (a very small timestep), and at every timestep, you need to calculate the forces of all the atoms.
- You need a very large number of very small timesteps just to simulate the life of 1 millisecond of atoms

Sequential algorithm for atom simulation

- At every time step, calculate the forces on each atom (bonded and short-range forces at every step, and long-range non-bonded forces every few time steps).
- Calculate the accelerations, velocities, and new positions at every timestep and repeat.

Approaches for parallelization - Atom decomposition

- In atom decomposition, you don't worry about the position of atoms and simply distribute n atoms across p processes and each process calculates the forces on the atoms assigned to it
- There can be issues with this naive approach, such as load imbalance and lack of communication among atoms split across processes.
- For load imbalance issues, you can distribute a force matrix to processes, which tells you if you need to calculate forces between two atoms, but it doesn't solve the issue of lots of communication being needed if bonded atoms are split among processes.

Approaches for parallelization - Spatial decomposition

- Spatial decomposition assigns a region of the 3D simulation to each process which solves the issue of atoms bonded to each other being split up.
- Create smaller boxes and assign a box to each process, (typically 16 angstroms), and you only need to communicate with neighboring boxes.

- Within 16 angstroms, calculations will be done explicitly, beyond 16 angstroms, calculations will be done using approximations.
- If an atom moves into a different box, then it gets sent to the other process.

Approaches for parallelization - Hybrid Parallelization

- Hybrid parallelization combines force and space decomposition.
- The 3D space is divided into smaller boxes, which are then allocated to various processes.
- Instead of assigning a single process to manage both the atoms and their force calculations, distinct processes are designated specifically for force computations.
- Objects dedicated to calculating forces between atoms across different boxes are created. These objects are then independently assigned to processors, which are responsible for performing the actual force calculations.
- While this method introduces more communication, it enhances scalability, allowing it to handle larger simulations.
- Initially, atoms are assigned to processors through spatial decomposition. Calculations involving interactions between different boxes are assigned to independent processes, similar to force decomposition.

Neutral territory (NT) methods

- Desmond is a molecular dynamics simulator developed by D.E. Shaw, which parallelizes its simulations using the midpoint method.
- The simulation space is divided into smaller three-dimensional volumes known as boxes, each designated as a home box for the particles within it.
- Each home box is managed by a single process, responsible for maintaining all relevant particle data (mass, charge, position, etc.) for that box.
- Molecular interactions can extend beyond a single home box, necessitating communication between processes managing adjacent boxes.
- This is done through mechanisms that allow each process to access data on particles that are near the edges of their home boxes or in neighboring boxes.
- To ensure accurate force computations for interactions that span across home box boundaries, Desmond clones particle data from neighboring boxes.
- All relevant particle data has to be available for interactions whose computational midpoint lies between multiple boxes.

Parallelization of Particle Mesh Ewald (3D FFT)

- Direct force calculations get replaced by splitting electrostatic interactions into short-range and long-range components.
- Short-range forces are calculated directly in real space with an effective cutoff radius, while long-range forces are handled in Fourier space to manage interactions more efficiently.
- A 3D mesh or grid is created to represent the charge densities of atoms across the simulation box.

- This grid is used with Fast Fourier Transforms (FFT), which compute both the discrete Fourier transform (DFT) and its inverse.
- By using FFT on a mesh grid, PME significantly reduces computational complexity from O(N^2) to O(NlogN), making it feasible to simulate large molecular systems.
- This method allows for highly accurate calculations of electrostatic forces.
- In the 1D or slab decomposition method of parallelizing FFT, the three-dimensional data grid used is divided along one dimension into slabs that span the other two dimensions, with each assigned to a different process.
- The 2D or pencil decomposition technique divides the 3D grid along two dimensions into “pencils”, with each being is assigned to a different processor.

Epidemiology

- Computational and mathematical modeling of epidemics is important to assist governments in responding to outbreaks, but is challenging because of large populations and dense urbanizations.
- There is also increased local and global travel since people are traveling quickly every day, so all modes of transportation must be simulated.
- There is also an increasingly immuno-compromised population.
- One simulation method is agend based modeling, which model people and interactions between them to simulate epidemic diffusion
- Create a model of if two people interact with each other for let’s say 30 minutes, what’s the probability of one person infecting the other if one of them has the disease.