Overview of NAMD and Molecular Dynamics

Jim Phillips

Low-cost Linux Clusters for Biomolecular Simulations Using NAMD

Outline

- Overview of molecular dynamics
- Overview of NAMD
- NAMD parallel design
- NAMD config file format
- NAMD config file walkthrough

Molecular Dynamics

Energy function: $U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R})$

used to determine the force on each atom:

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R})$$

Newton's equation represents a set of N second order differential equations which are solved numerically at discrete time steps to determine the trajectory of each atom.

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \vec{F}_i(t)$$

Advantage of the Verlet Method:
requires only one force evaluation per timestep

Energy Functions

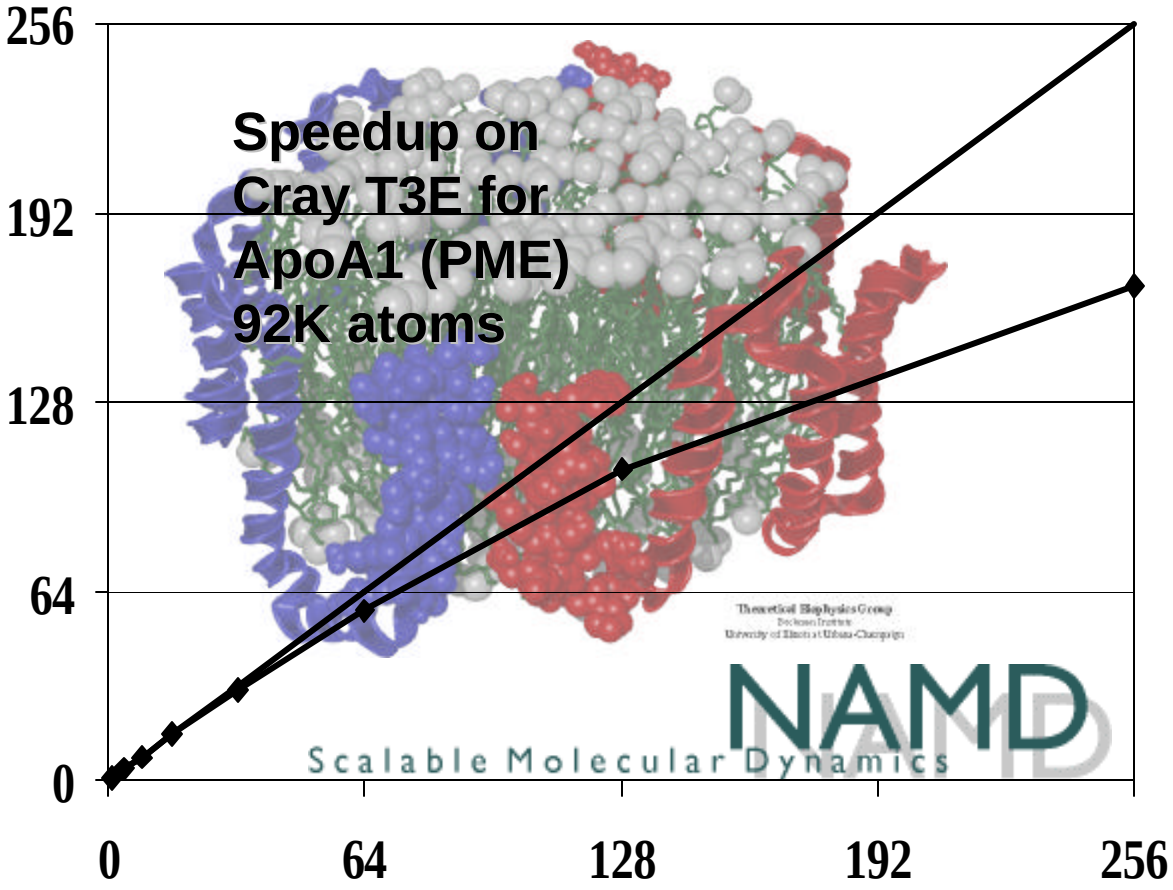
$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

Ubond = oscillations about the equilibrium bond length
Uangle = oscillations of 3 atoms about an equilibrium angle
Udihedral = torsional rotation of 4 atoms about a central bond
Unonbond = non-bonded energy terms (electrostatics and Lenard-Jones)

NAMD: Scalable Molecular Dynamics



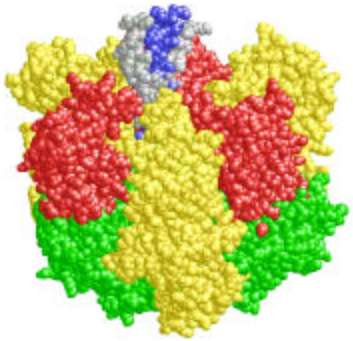
Runs on SP, Origin, T3E, clusters…and Windows.

Scalable to 1000's of CPUs.

Data file compatible with CHARMM and X-PLOR.

Fast full electrostatics and constant pressure ensembles.

Ready for 100,000 atoms and beyond…



ATPase

Freely available with C++ source code from http://www.ks.uiuc.edu/Research/namd/

Scalability: The Problem

- Speedup = (serial time) / (parallel time).
- Ideal (linear) speedup is number of CPUs.

- Amdahl's law says best speedup possible is (total work) / (non-parallelizable work).

- In MD, each timestep must be parallelized.

NAMD Parallel Design

- Atoms are “spatially decomposed” into cubes which are distributed among CPUs.
  - Generates 25-400 cubes to distribute.
- Interactions between pairs of cubes are also distributed among CPUs.
  - Generates 300-5000 interaction groups.
- Interactions are redistributed among CPUs based on measurement of their run time.

Active Load Balancing

- Initial distribution minimizes communication, but load estimate is poor.
- Calculations are redistributed from scratch based on initial load measurements.
- Additional measurements allow incremental refinement of the distribution, taking into account additional communication load.

Data Driven Execution

- Some communication is necessary, but we don’t want to sit idle waiting for it.
- Work is broken down into interaction sets which require different parts of the incoming data.
- Interactions are calculated as soon as required data is received, not in any particular order.

Configuration File Format

- NAMD is case insensitive (Tcl isn't).
- All options follow: <name> <value>
- Some options may appear several times.
- Some options contain scripts inside { }.
- Everything following '#' is a comment.
- Include other files with "source <file>".
- If you know Tcl, use print rather than puts.

```
# molecular system
structure   bpti.psf

# force field
parameters  param19.pro
exclude     scaled1-4
1-4scaling  0.4

# approximations
switching   on
switchdist  7.5
cutoff              8
pairlistdist        9.5
margin              0
stepspercycle       10
```

- The .psf file lists atom types, charges, masses, bonds, angles, etc., which do not change.
- The parameter file defines an energy function based on the types of the atoms. It also defines an exclusion policy which is given separately here.
- Nonbonded forces are cutoff by 8A with a smooth switching function starting at 7.5A.
- Atoms outside pairlistdist will stay outside cutoff for 10 steps.

```
# initial conditions
coordinates
      bpti_lang.coor
velocitities
      bpti_lang.vel

# integrator
rigidbonds  all
timestep    2.0

#protocol
firsttimestep     50000
numsteps    75000
```

- Coordinates and velocities in PDB formatted files start the simulation. (Velocities can also be randomly generated based on a given temperature.)
- All bonds involving hydrogen atoms will be held rigid. (Other options are none and water.)
- This allows a 2fs timestep.
- Timesteps will be numbered starting at 50,000.
- Stop on reaching timestep 75,000 (run for 25,000 steps).

```
# output (log)
outputenergies    10
outputtiming      50

#output (files)
binaryoutput      no
outputname  bpti_free
dcdfreq     500
```

- Energies and temperature will be printed out every 10 steps.
- Elapsed and estimated time will be printed out every 50 steps.

- Generate text PDB files (rather than binary files) of final state.
- Output files will be named bpti_free.coor, bpti_free.vel, bpti_free.dcd, etc.
- Coordinates will be written to DCD format trajectory file every 500 steps (1ps).

Conclusion

- Manuals, etc. online at:

  http://www.ks.uiuc.edu/Research/namd/
- NAMD email helpdesk:

  namd@ks.uiuc.edu