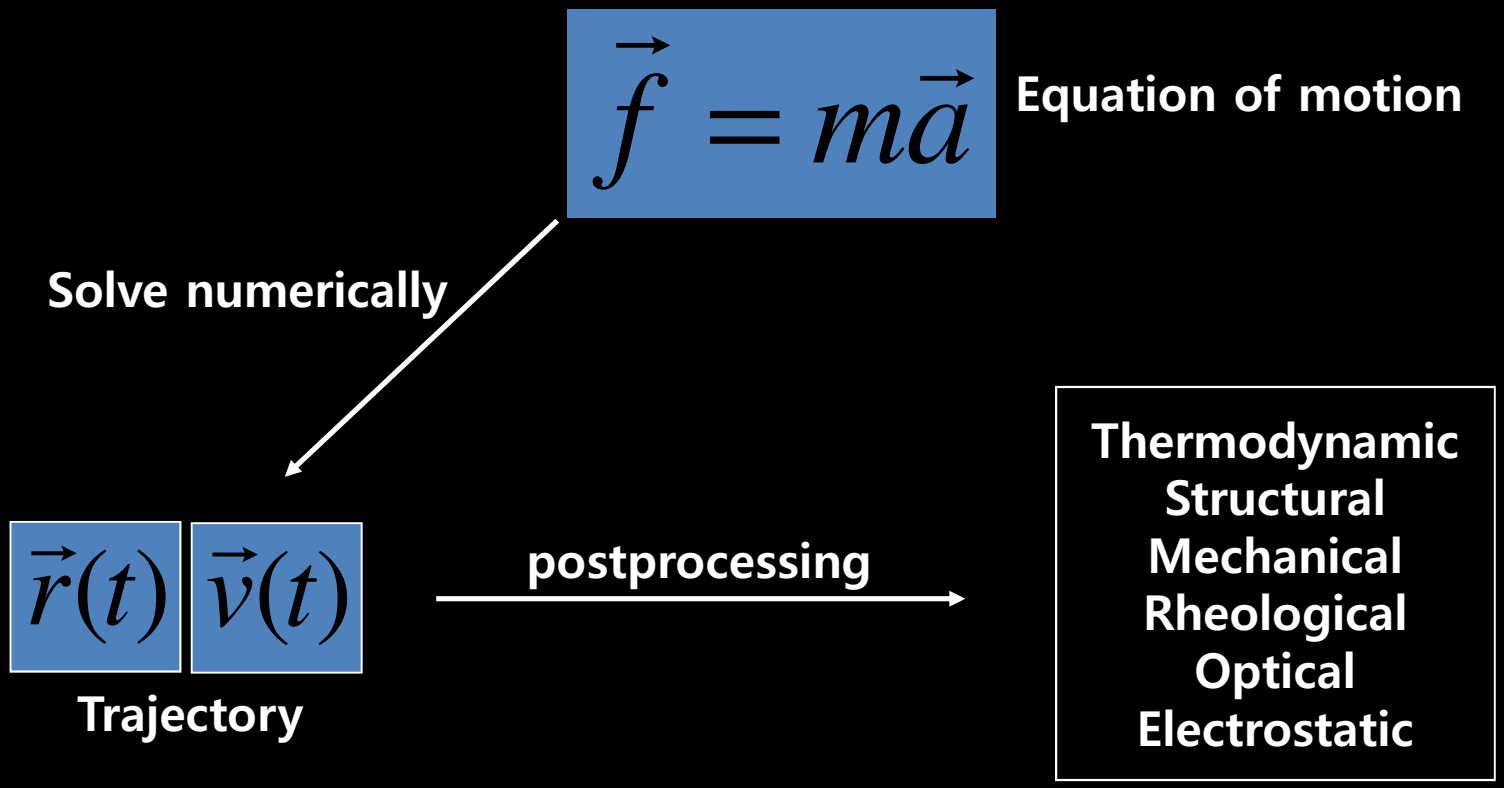
Developing a highly scalable molecular dynamics simulation program

Kwang Jin Oh

KISTI Supercomputing center

What is molecular dynamics simulation ?



Molecular dynamics simulation in NVE ensemble

$$\frac{d\vec{r}_i}{dt} = \vec{v}_i$$

$$\frac{d\vec{v}_i}{dt} = \frac{\vec{f}_i}{m_i}$$

$$H' = K_f + V$$

$$\vec{v}(\frac{\Delta t}{2}) = \vec{v}(0) + \frac{1}{2}\frac{\Delta t}{m}\vec{F}(0)$$

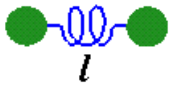
$$\vec{r}(\Delta t) = \vec{r}(0) + \Delta t\vec{v}(\frac{\Delta t}{2})$$

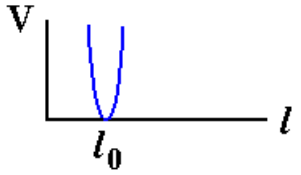
$$\vec{v}(\Delta t) = \vec{v}(\frac{\Delta t}{2}) + \frac{1}{2}\frac{\Delta t}{m}\vec{F}(\Delta t)$$

velocity Verlet algorithm

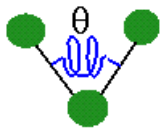
Empirical Potential Energy Function

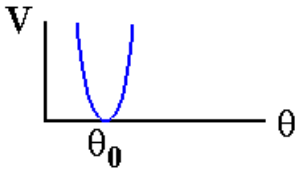
Bonds



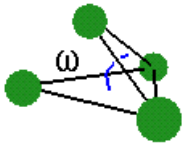


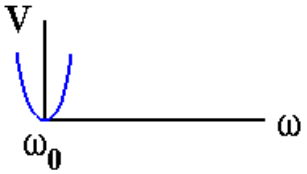
Angles



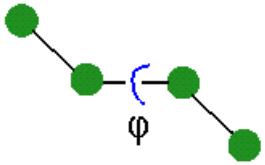


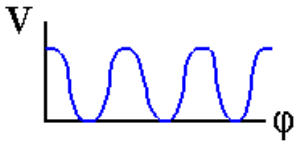
Improper Dihedrals



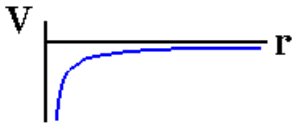


Torsions

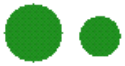


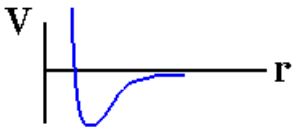


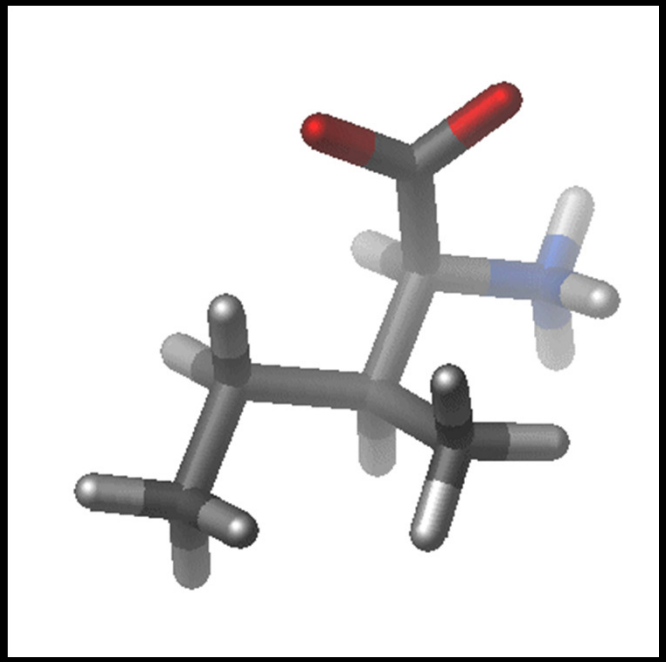
Electrostatics



van der Waals







Trajectory analysis

Temperature

$$< K >= \frac{3}{2} k_B T$$

Pressure

$$PV = Nk_B T + < W >$$

Hear capacity

$$< \delta K^2 >= k_B T^2 C_V$$

Radial distribution function

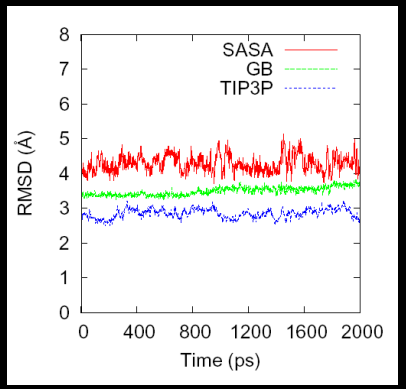
$$g(r) = \frac{V}{N^2} < \sum \sum \delta(\vec{r} - \vec{r}_{ij})$$

Diffusion constant

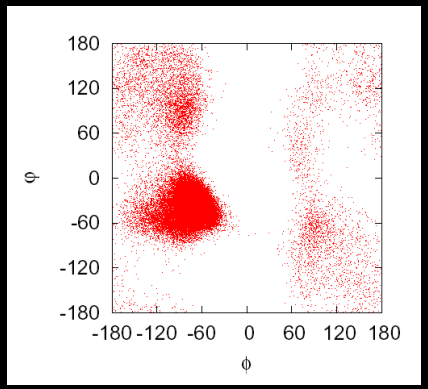
$$2Dt = < (\vec{r}(t) - \vec{r}(0))^2 >$$

Trajectory analysis

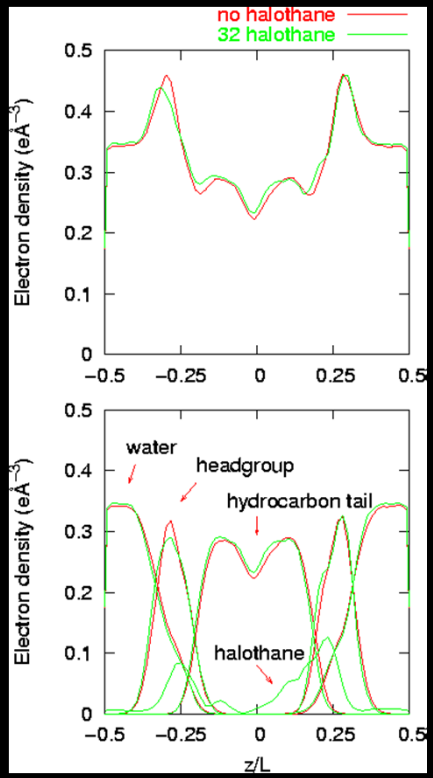
RMSD



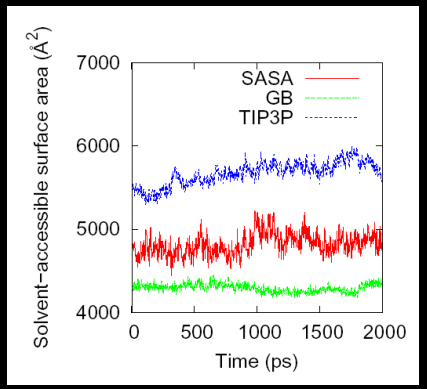
Ramachandran plot



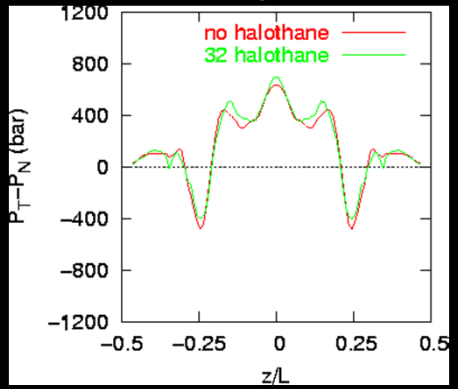
Density profile



SASA

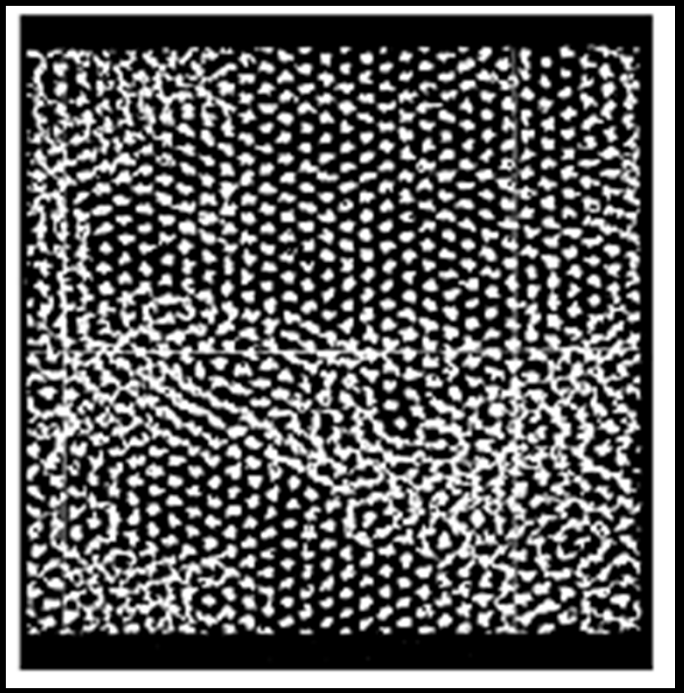


Pressure profile



Computing power

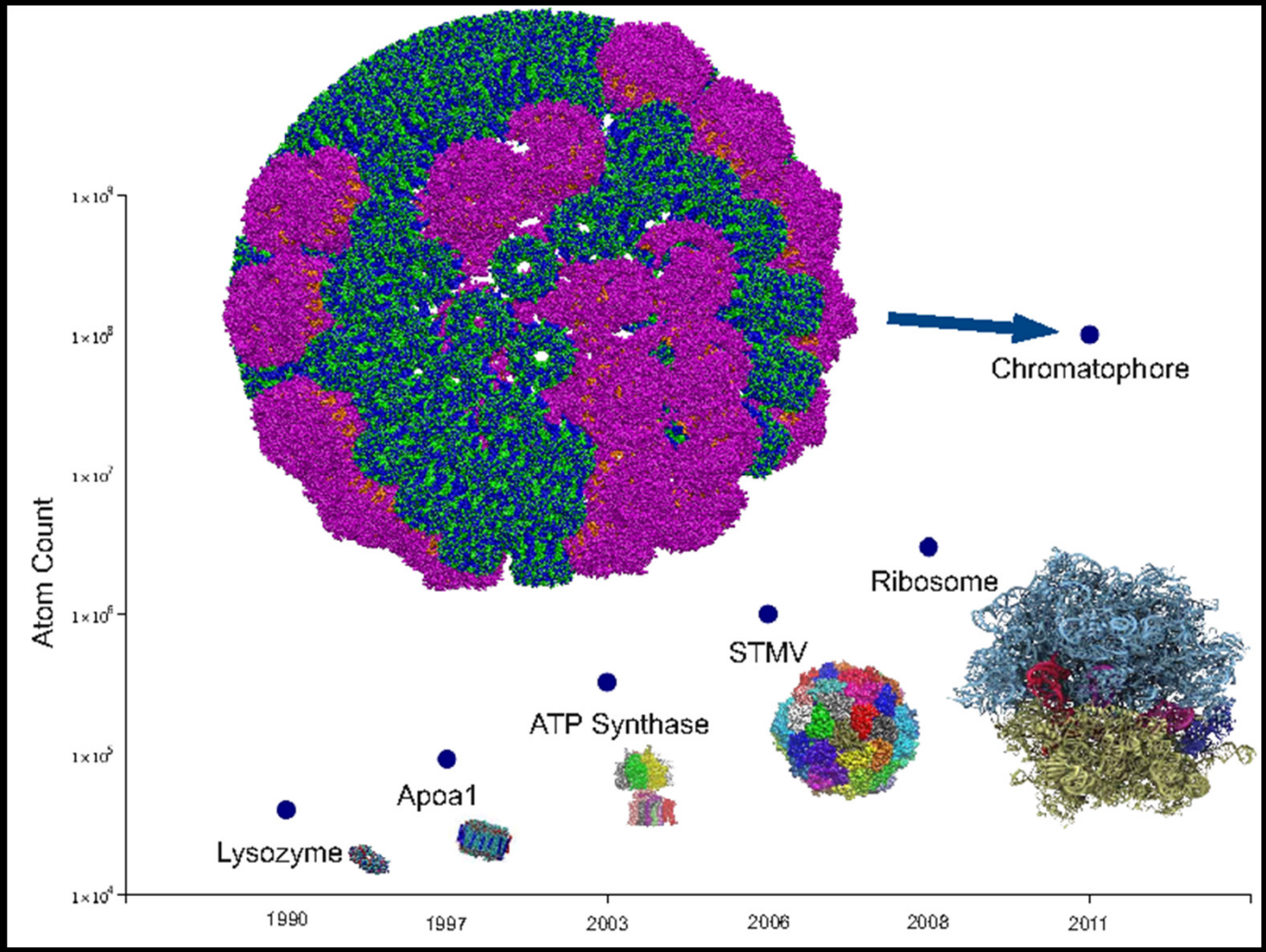
IBM 704 at Lawrence Livermore
(5 KFLOPS)



Trajectory near liquid-solid
Phase transition

MD of hard spheres by Alder and
Wainright: First MD

Computing power



What is problem?

Time scale issue -> Need large number of iterations

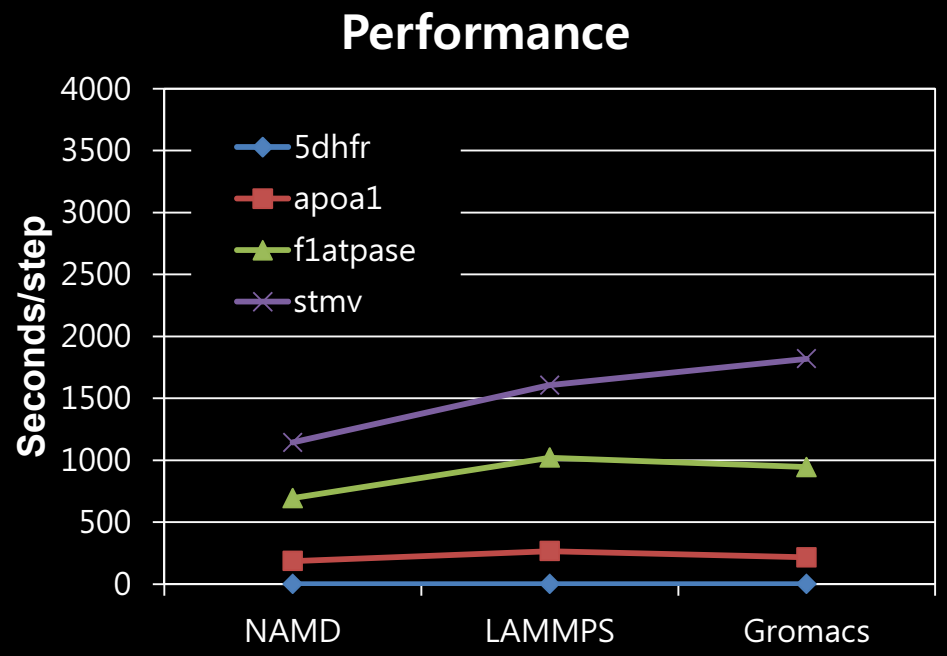
- Local Motions (0.01 to 5 Å, 10^{-15} to 10^{-1} s)
  - Atomic fluctuations
  - Sidechain Motions
  - Loop Motions
- Rigid Body Motions (1 to 10Å, 10^{-9} to 1s)
  - Helix Motions
  - Domain Motions (hinge bending)
  - Subunit motions
- Large-Scale Motions (> 5Å, 10^{-7} to 10^{4} s)
  - Helix coil transitions
  - Dissociation/Association
  - Folding and Unfolding

Computationally very intensive !!!

Length scale issue -> Need large number of particles

- About 10 billion atoms in μm^3 -> prohibitive

Simple example



5dhfr: 23558
apoa1: 92224
f1atpase: 327506
stmv: 1066628

1 ns simulation of **stmv** with
1 fs integration time step
-> 1M steps
-> ~10^9 seconds
-> ~10000 days

What we are trying to do?

- Faster force calculation algorithm (neighbor list, SPME, FMM, coarse-grained model, …)
- More efficient numerical integration algorithm (RESPA, …)
- **Higher scalable parallel algorithm (DD, load balancing, …)**

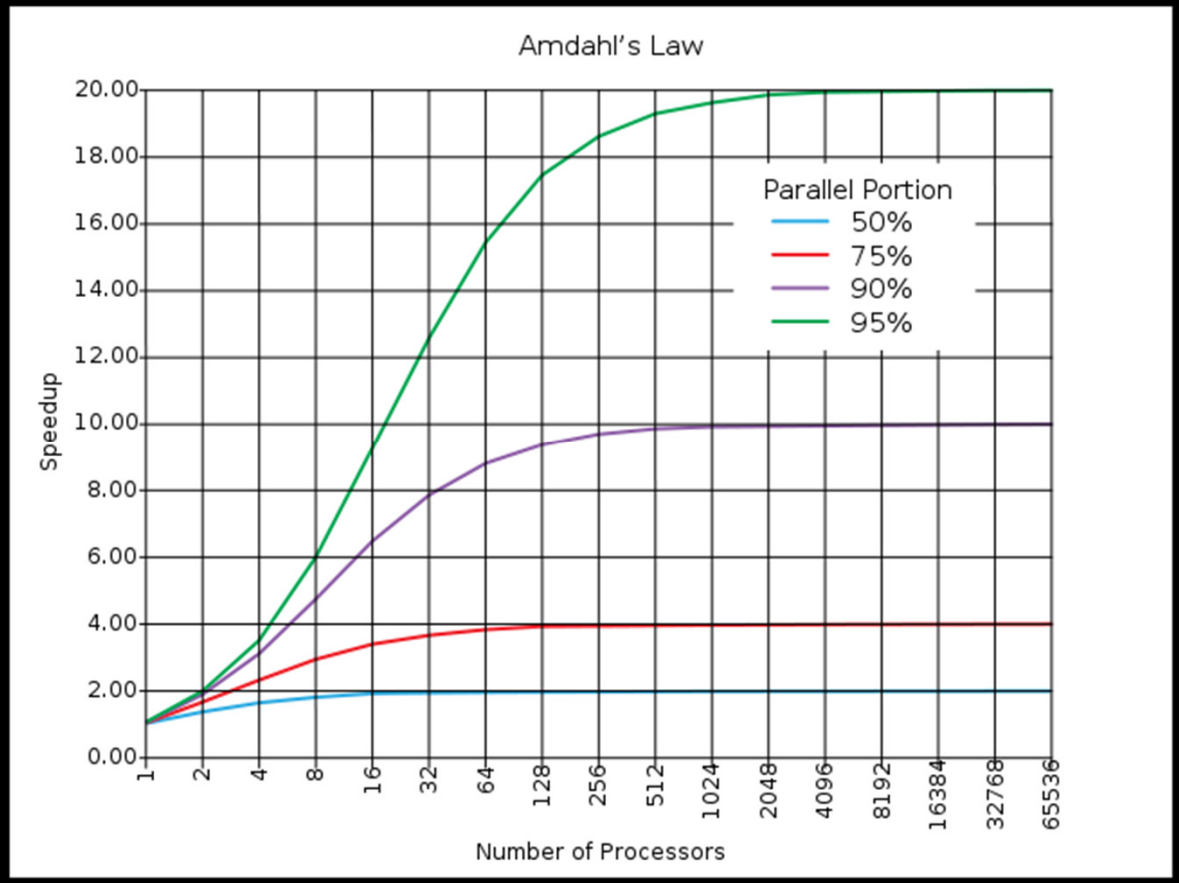
Challenges we face

- Utilizing cores of $O(10^5)$ with good scalability
  - N/P is small
  - computation/communication is low
- Utilizing emerging architectures
  - need new programming model
  - need an efficient parallelization scheme
- Optimizing performance to reduce the gap

Amdahl's law

$$S = \frac{1}{(1-p) + \frac{p}{N}}$$

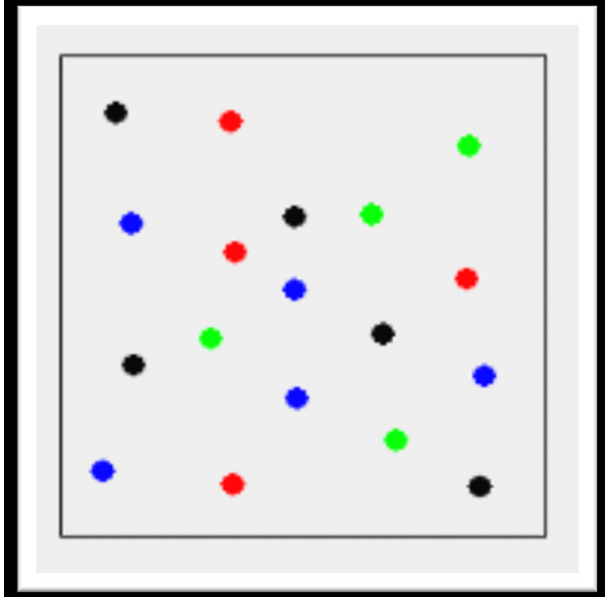
N: # of processors
p: parallel portion



We need better parallel algorithm !!!

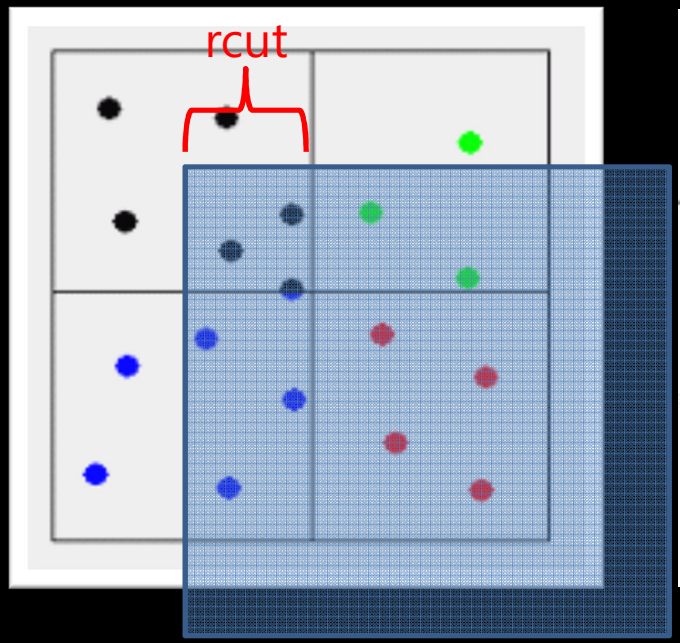
Parallelization schemes

Atom decomposition



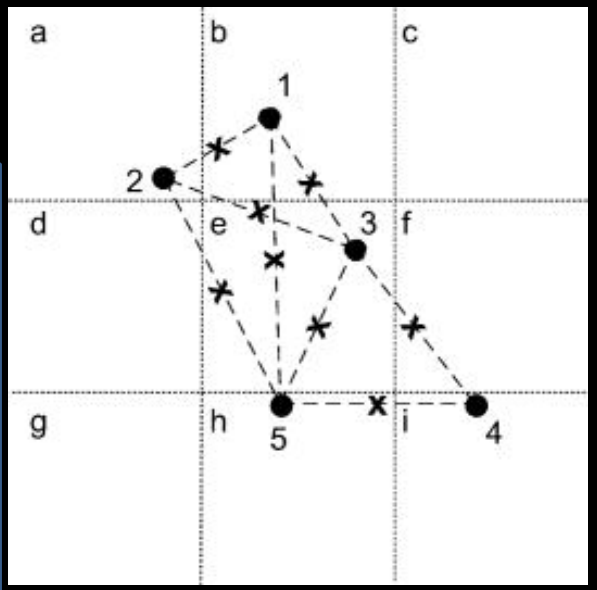
- All-to-all communication
- Poor scalability
- Better load balancing

Domain decomposition



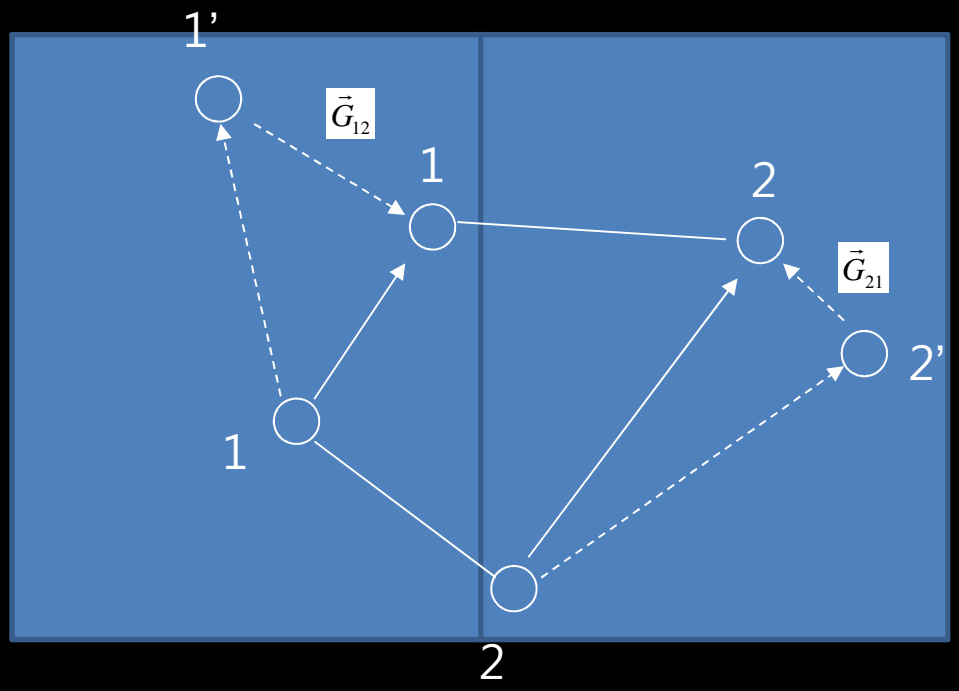
- Communication only with neighbor processors
- Better scalability
- Poor load balancing
- # of cores are limited by the domain size

Midpoint method



- 0.5*rcut

Bond constraints



SHAKE

$$\vec{G}_{ij} \approx \frac{\mu_{ij}(d_{ij}^2 - d_{ij}'^2)}{2\Delta t^2 \vec{d}_{ij}^0 \bullet \vec{d}'_{ij}} \vec{d}_{ij}^0$$

- Integrating vibrational motion of high frequency with large time step -> the trajectory blows up eventually
- To constrain bonds or angles like C-H, H-O-H, and so on

Smooth particle mesh Ewald

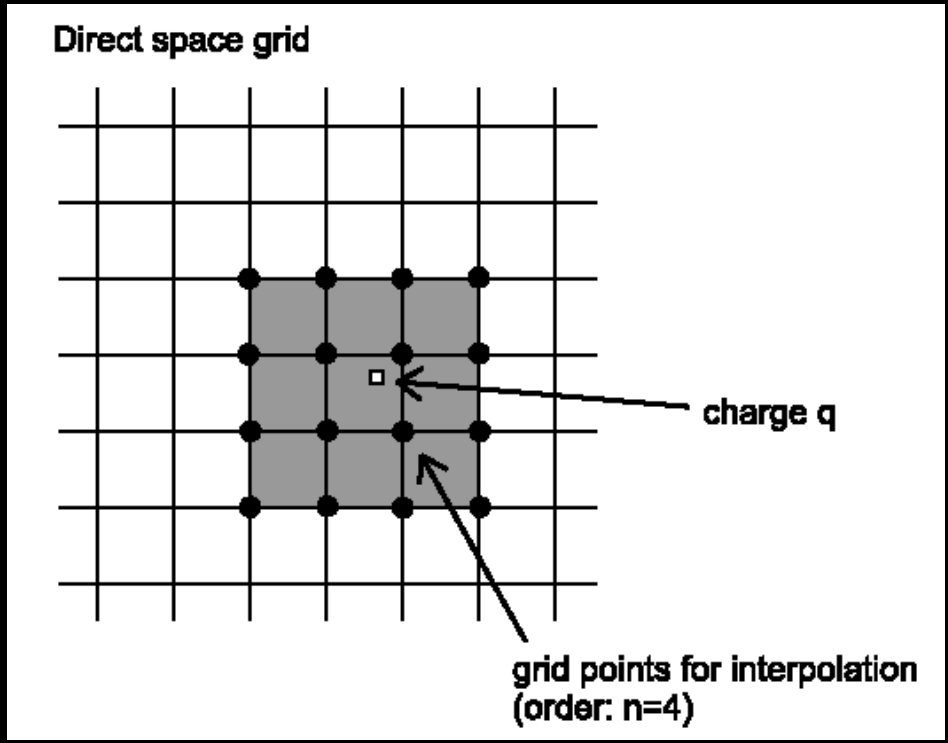
Ewald: $O(N^{3/2})$

SPME: $O(N \log N)$

$$V = \sum_i \sum_{j>i} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j erfc(\alpha r_{ij})}{r_{ij}}$$

$$+ \frac{1}{\varepsilon_0} \sum_{\vec{k} \neq 0} \frac{\exp(-k^2 / 4\alpha^2)}{k^2} \left| S(\vec{k}) \right|^2$$

$$- \frac{1}{4\pi\varepsilon_0} \sum_i q_i^2 \frac{\alpha}{\sqrt{\pi}}$$



$$S(\vec{k}) = \sum_i q_i \exp(i\vec{k} \bullet \vec{r}_i)$$

Ewald

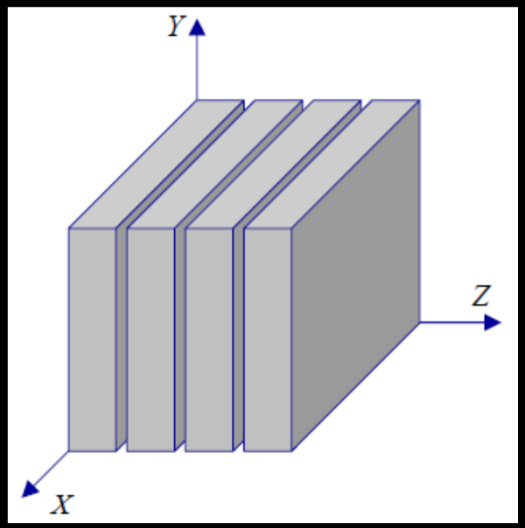
$$S(\vec{k}) \approx b_1(k_1) b_2(k_2) b_3(k_3) F(Q)(k_1, k_2, k_3)$$

SPME

volume->slab

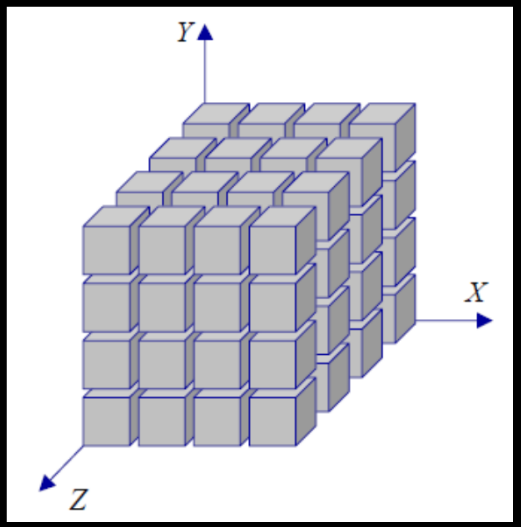
slab->volume

1. Get charges on grid points,
2. Block the charge array,
3. Backward FFT,
4. Calculate energy and all-reduce,
5. All-gather and Forward FFT,
6. Calculate forces.

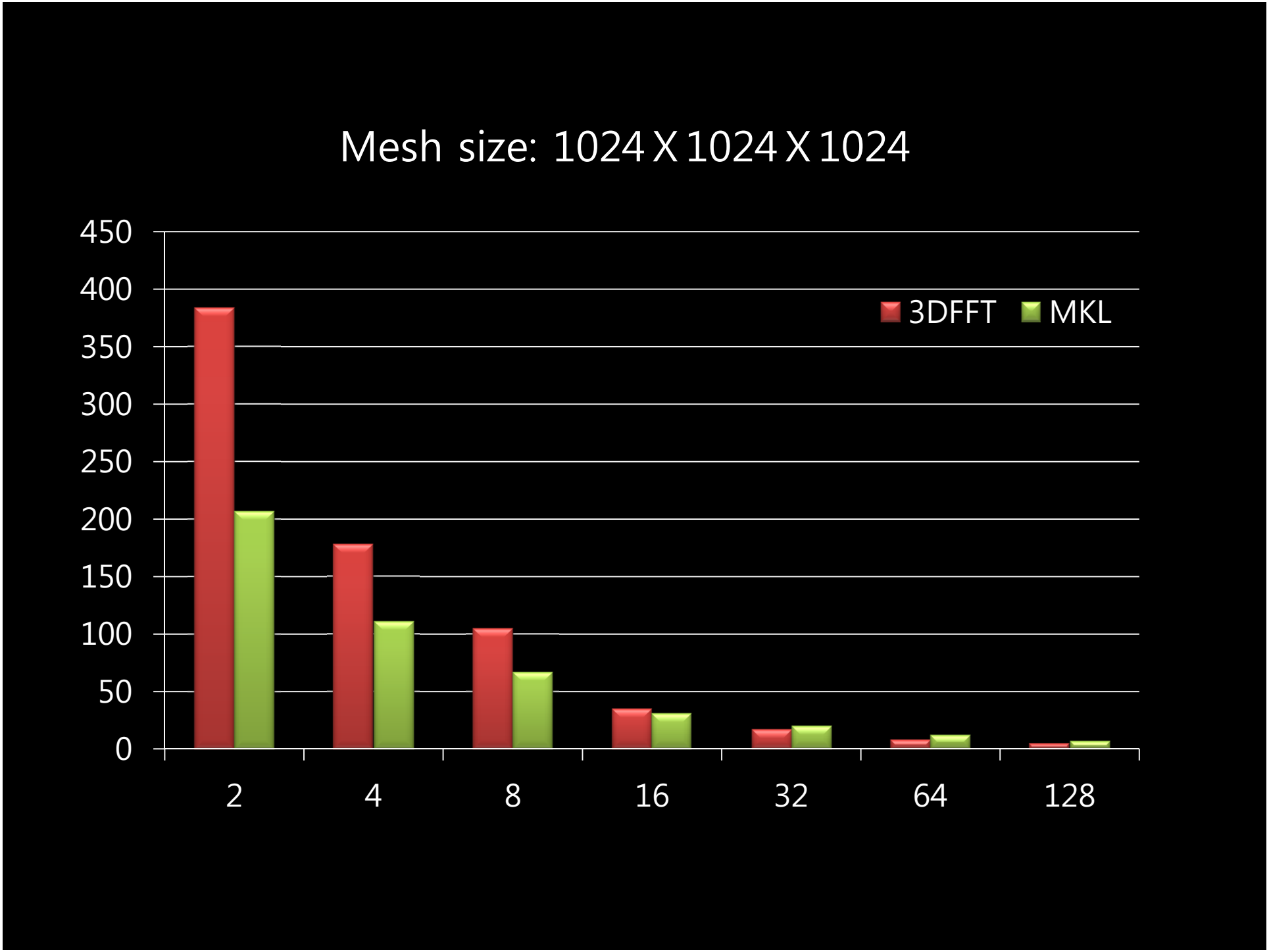


slab decomposition

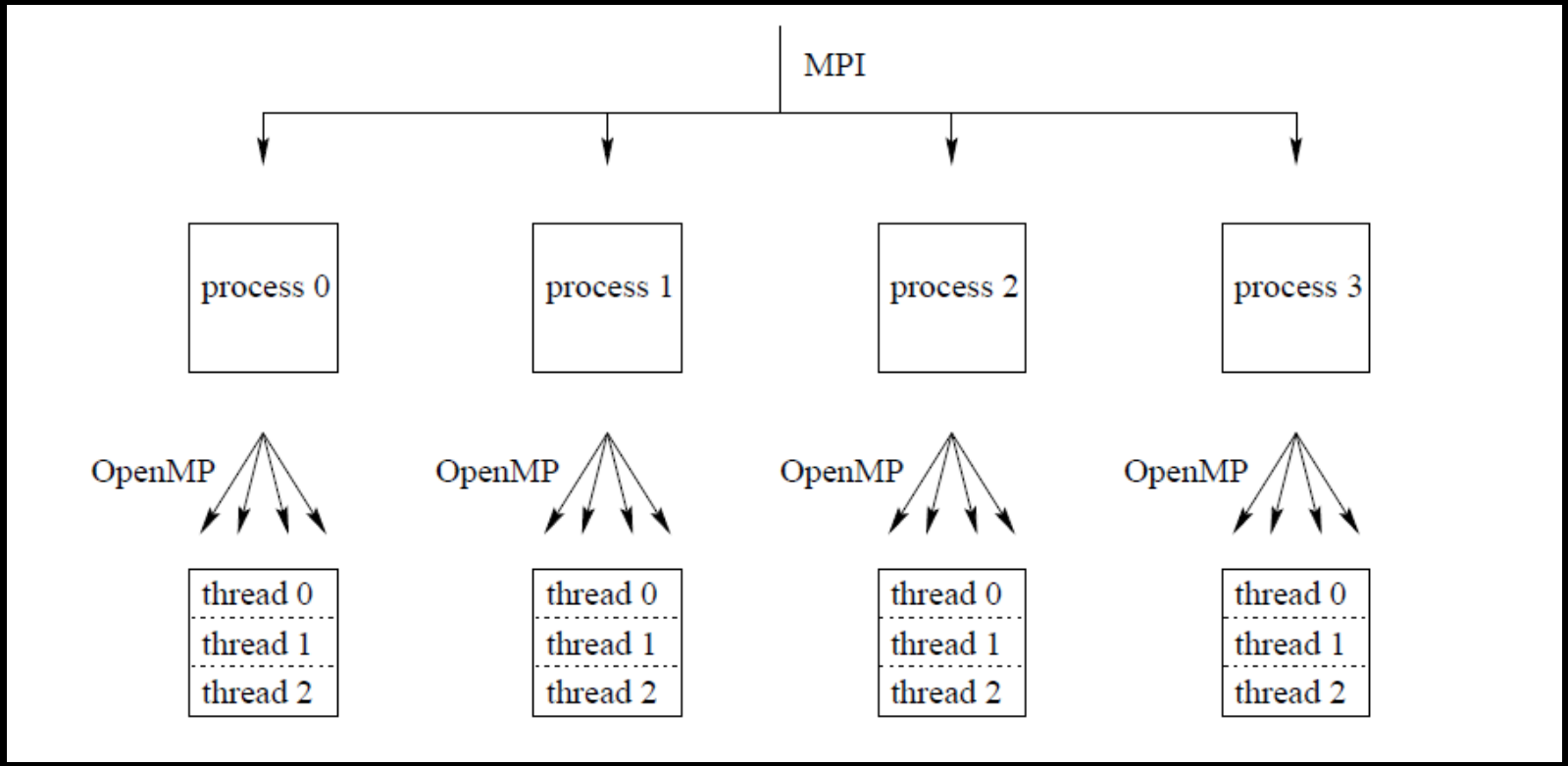
# of processors is limited by the dimension !!!



volumetric decomposition

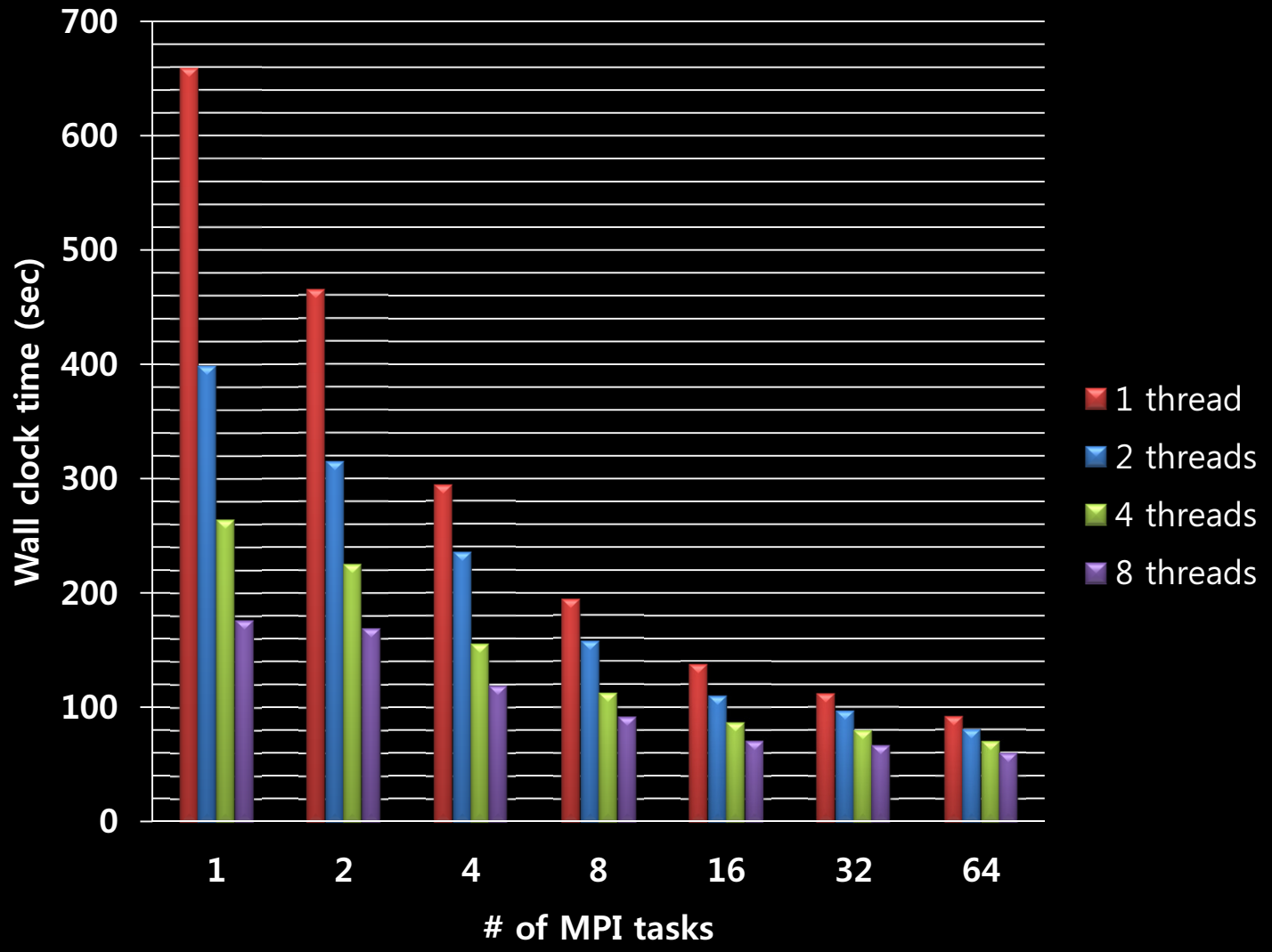

Mesh size: 1024 X 1024 X 1024
450
400
350
300
250
200
150
100
50
0
3DFFT
MKL
2
4
8
16
32
64
128

Hierarchical parallelization

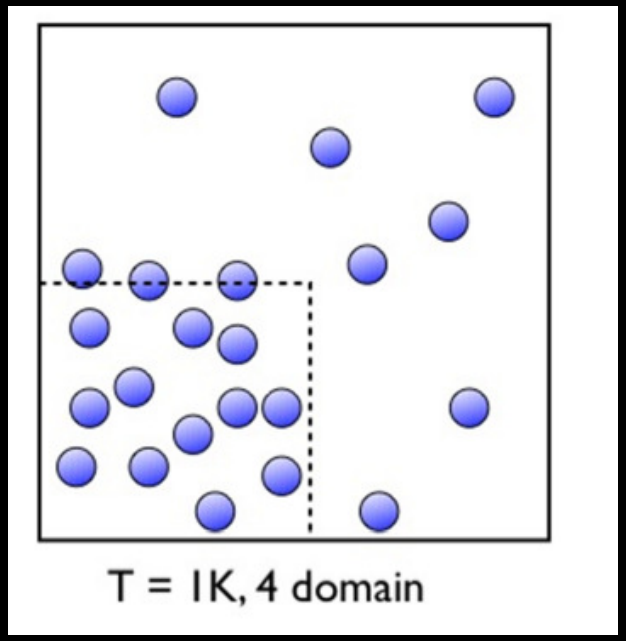


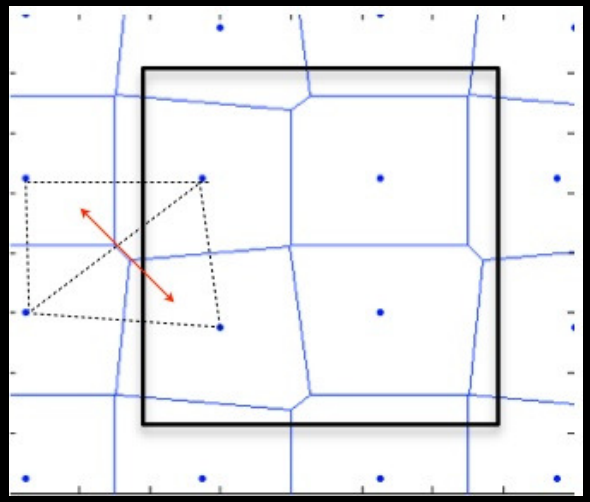
Parallel performance

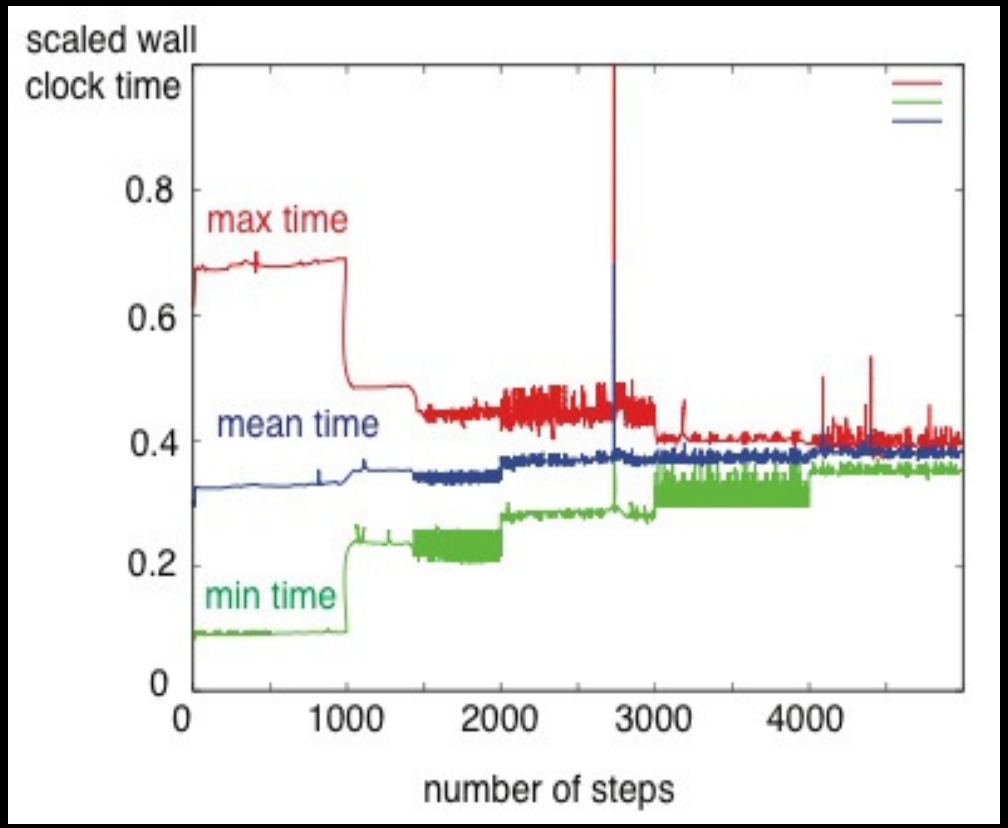
5DHFR (charmm22) + TIP3P
MDNVE 1000 steps on tachyon2
Box: ~64 A, rcut: 12 A



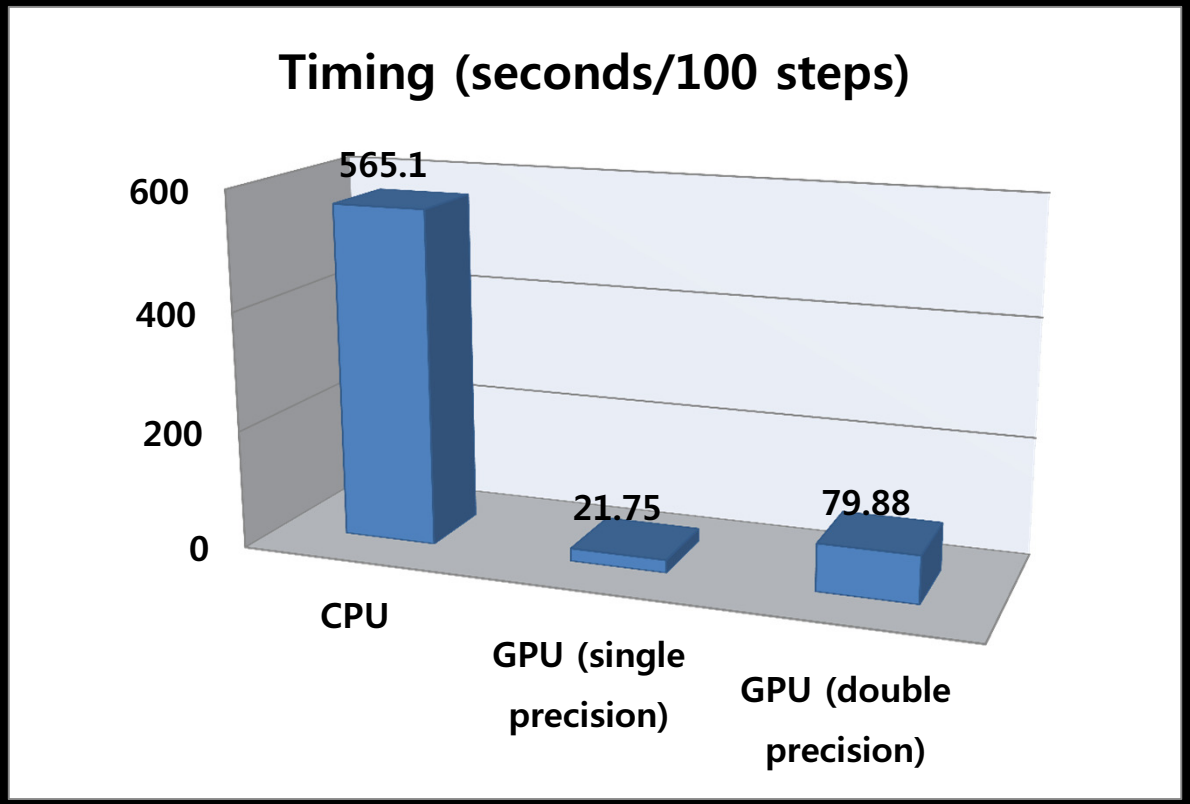
Load balancing: Voronoi diagram



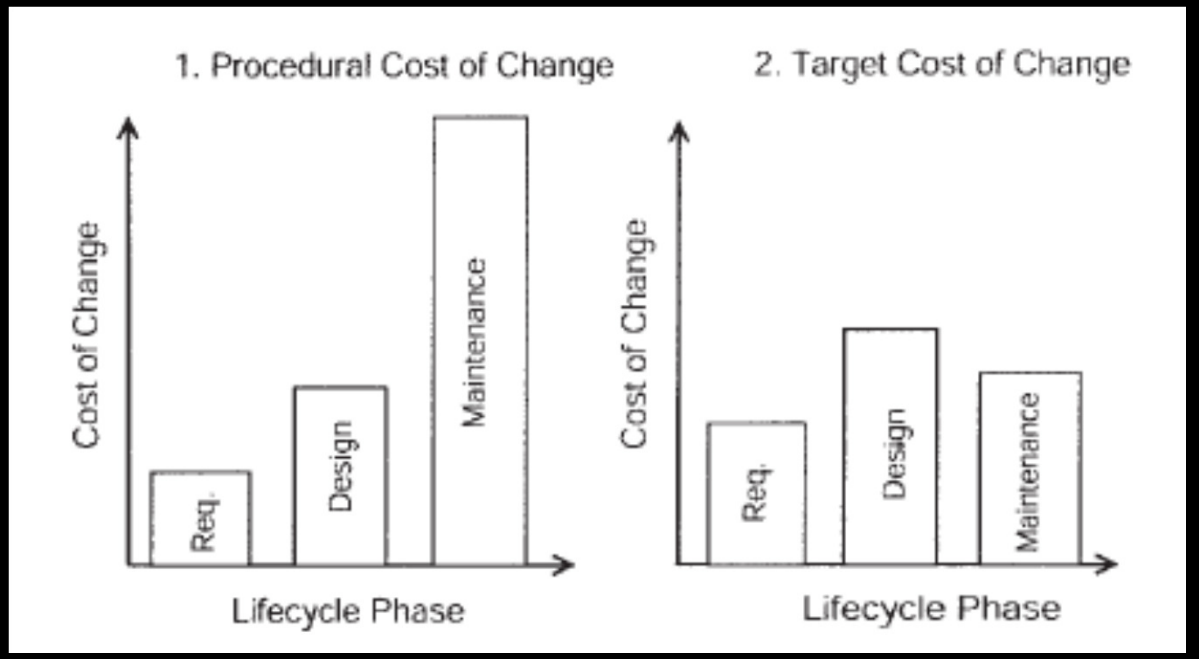




GPU computing



Object-oriented programming

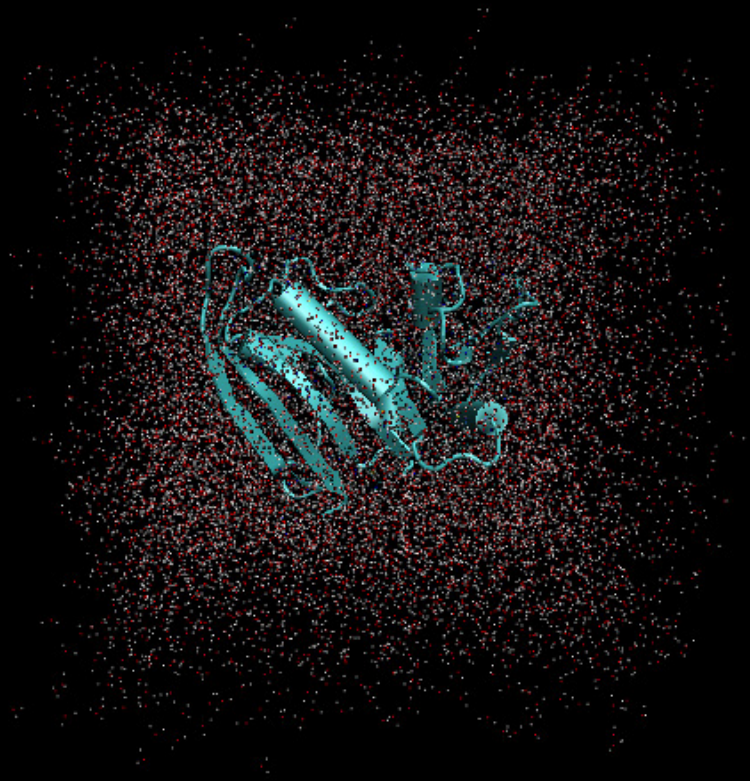


Abstraction, inheritance, polymorphism …

mm_par: A general purpose parallel MD code

- Object-oriented design using C++
- Hierarchical parallelization using MPI and OPENMP
- Domain decomposition based on atomic group
- Charmm, Amber, and easily extended to handle other forcefields
- NVE MD, NVT MD (global, molecule type, molecule, rigid group), NPT MD (flexible, isotropic, x-y-z, xy-z, z), LD, DPD
- Multiple time step using RESPA
- Replica exchange molecular dynamics
- Shake/Rattle for constraint dynamics
- Electrostatic force calculation using SPME
- Neighbor lists (Verlet neighbor list, cell-linked list, combined list)
- 3D FFT using volumetric decomposition
- Implicit solvent model (SASA, GB)
- Trajectory analysis

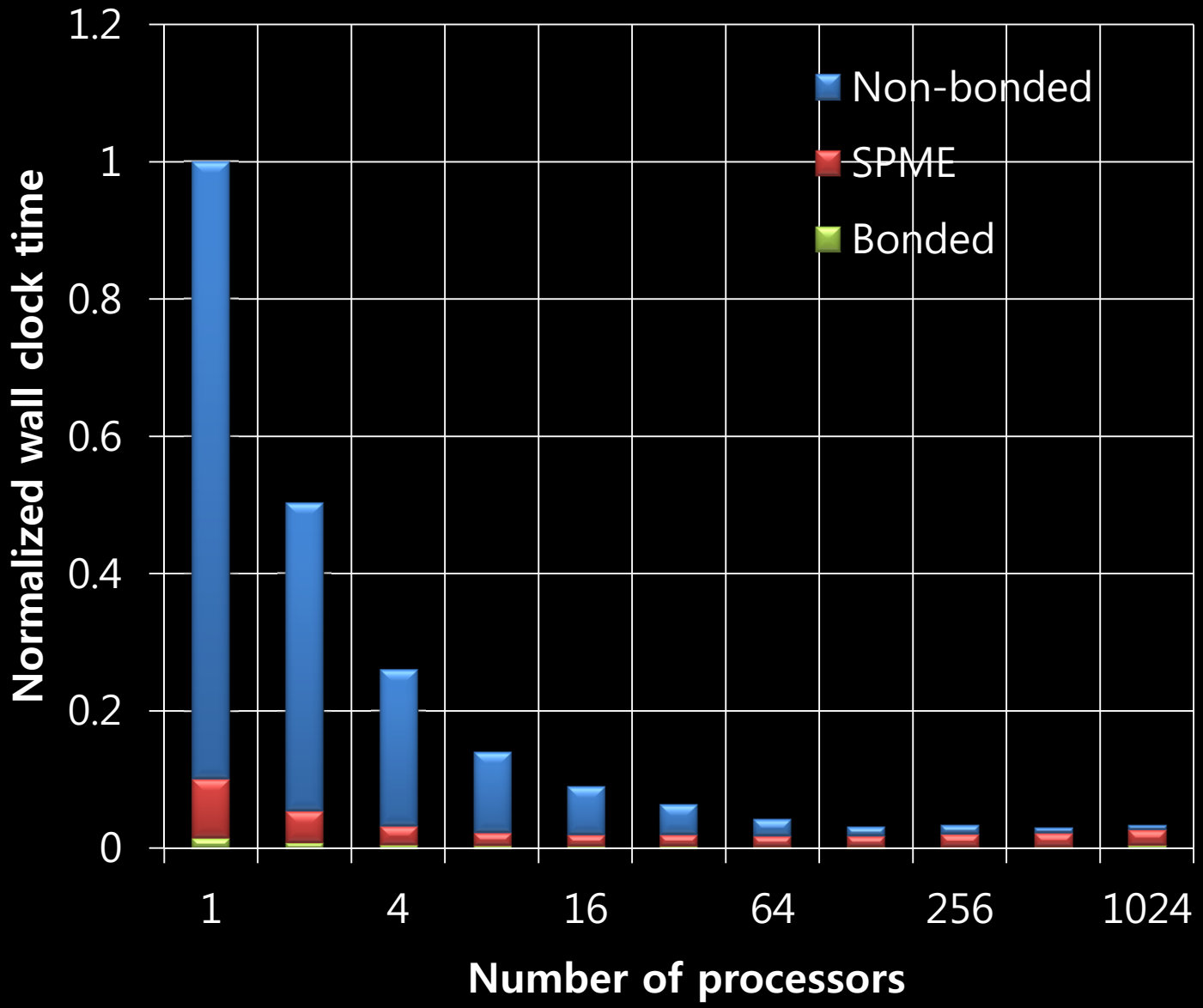
Future issues: strong scalability



5DHFR (CHARMM22)+7023 water (TIP3P)

	Bonded force	Non-bonded force (real space/ reciprocal space)	
Time (sec)	0.03	2.57	0.3

Future issues: strong scalability



Thank you !!!

koh@kisti.re.kr