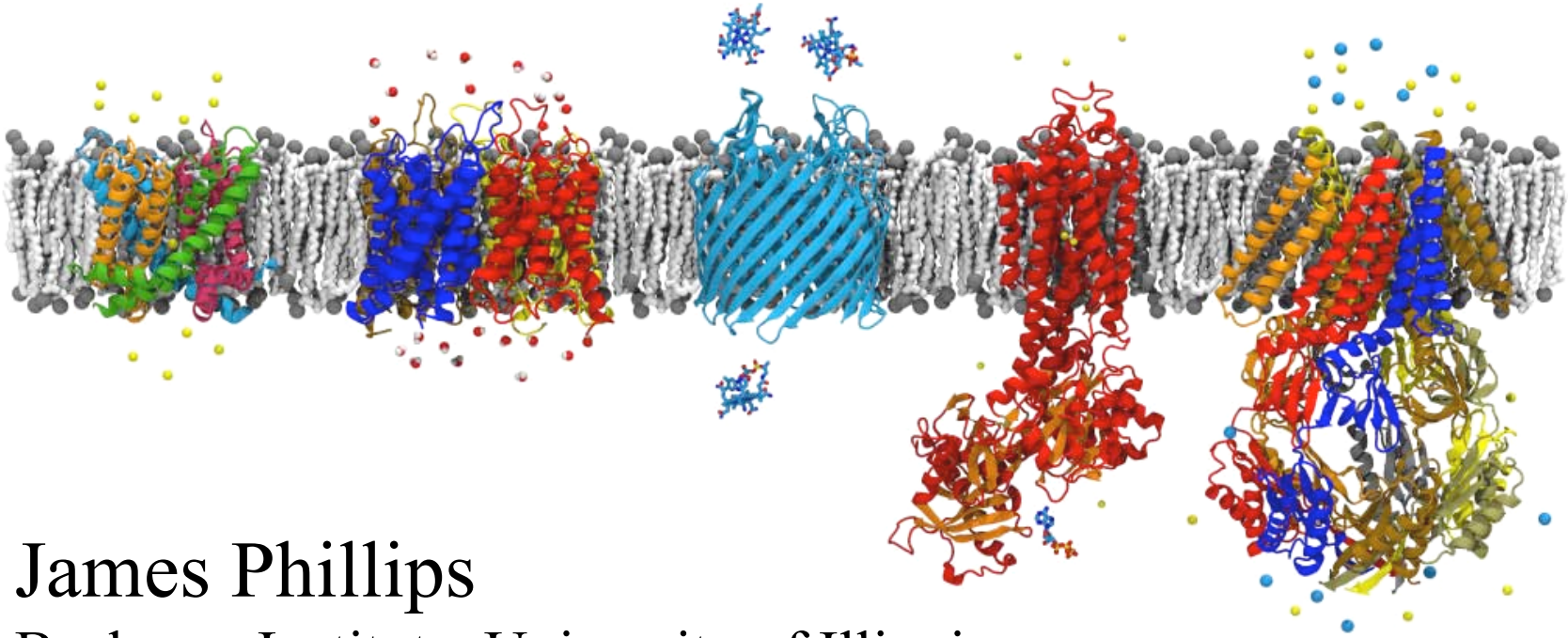
Petascale Molecular Dynamics with NAMD



James Phillips

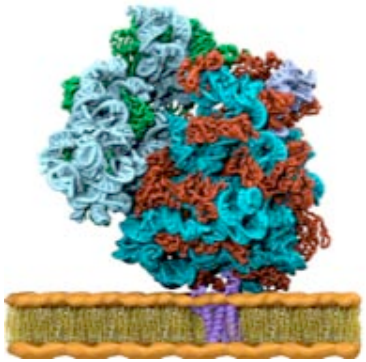
Beckman Institute, University of Illinois

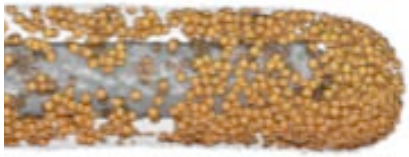
http://www.ks.uiuc.edu/Research/namd/

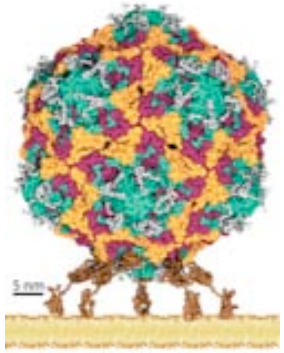
CScADS workshop on Leadership Computing Platforms,
Extreme-scale Applications, and Performance Strategies
July 23-26, 2012

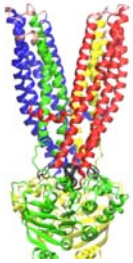

NIH BTRC for Macromolecular Modeling and Bioinformatics
1990-2017
Beckman Institute
University of Illinois at
Urbana-Champaign

Collaborative Driving Projects

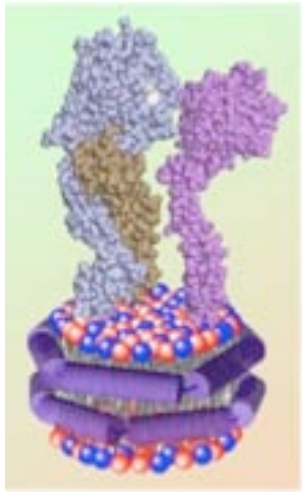


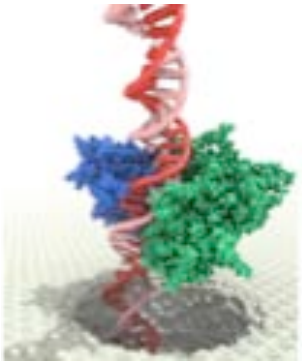


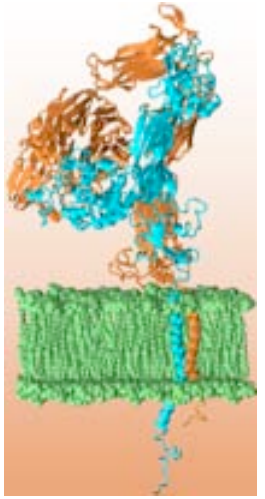




1. Ribosome	R. Beckmann (U. Munich) J. Frank (Columbia U.) T. Ha(UIUC) K. Fredrick (Ohio state U.) R. Gonzalez (Columbia U.)
2. Blood Coagulation Factors	J. Morrissey (UIUC) S. Sligar (UIUC) C. Rienstra (UIUC) G. Gilbert (Harvard)
3. Whole Cell Behavior	W. Baumeister (MPI Biochem.) J. Xiao (Johns Hopkins U.) C.N. Hunter (U. Sheffield) N. Price (U. Washington)
4. Biosensors	R. Bashir (UIUC) J. Gundlach (U. Washington) G. Timp (U. Notre Dame) M. Wanunu (Northeastern U.) L. Liu (UIUC)
5. Viral Infection Process	J. Hogle (Harvard U.) P. Ortoleva (Indiana U.) A. Gronenborn (U. Pittsburgh)
6. Integrin	T. Ha (UIUC) T. Springer (Harvard U.)
7. Membrane Transporters	H. Mchaourab (Vanderbilt U.) R. Nakamoto (U. Virginia) D.-N. Wang (New York U.) H. Weinstein (Cornell U.)

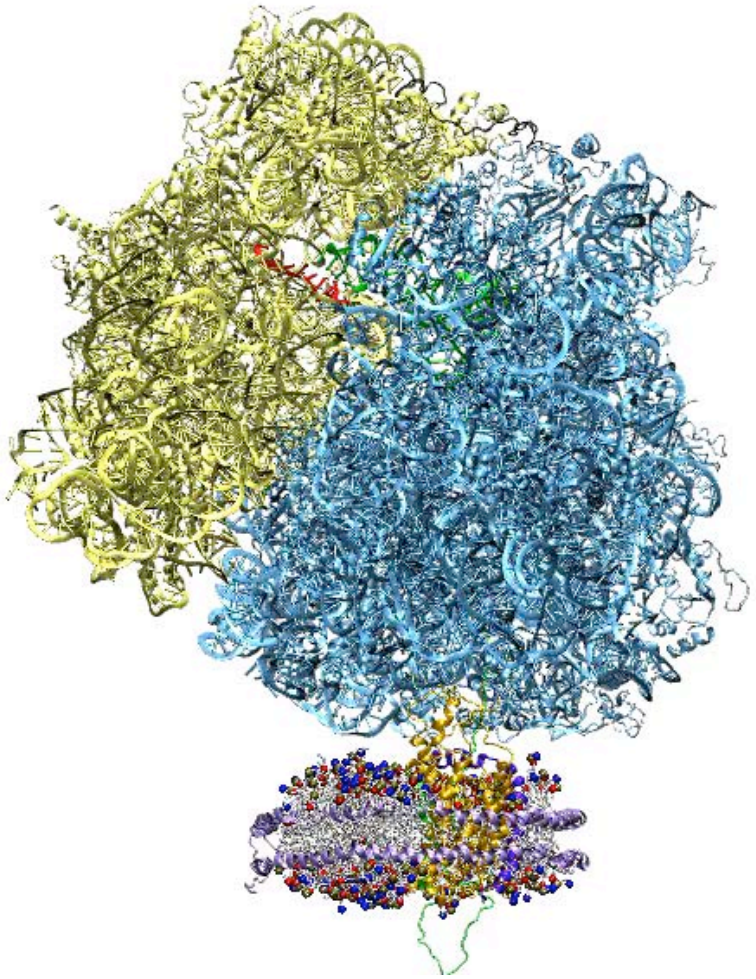






Ribosome Driving Project

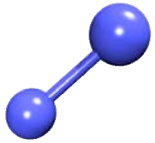
Target of over 50%
of antibiotics

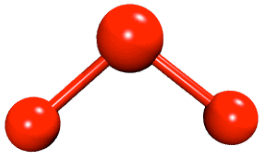


Many related diseases. e.g. Alzheimer's disease due to dysfunctional ribosome (J. Neuroscience 2005, 25:9171-9175)

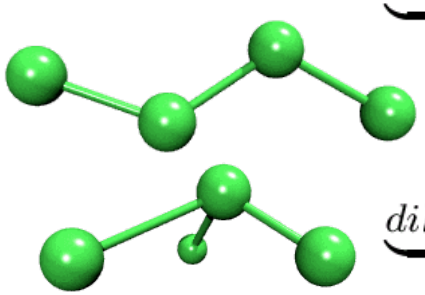
Localization failure of nascent chain lead to neurodegenerative disease (Mol. Bio. of the Cell 2005, 16:279-291)

Molecular Mechanics Force Field





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



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

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

Classical 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:

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

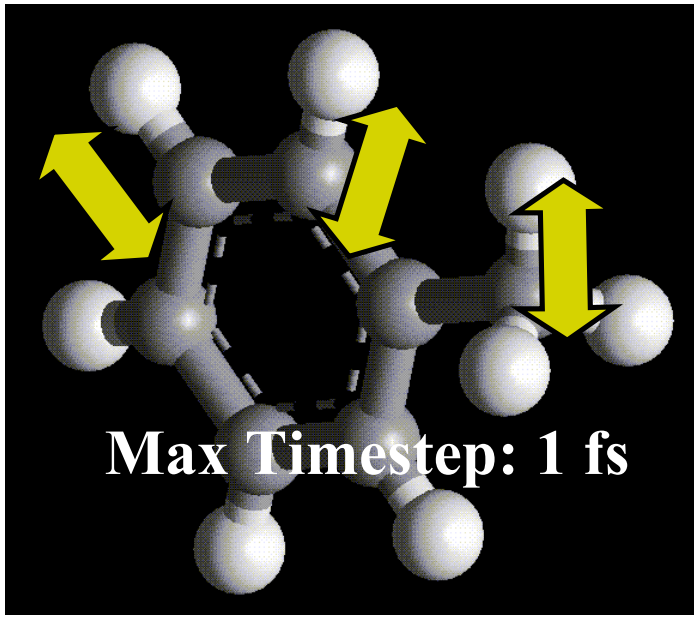
Newton's equation represents a set of N second order differential equations which are solved numerically via the Verlet integrator at discrete time steps to determine the trajectory of each atom.

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

Small terms added to control temperature and pressure.

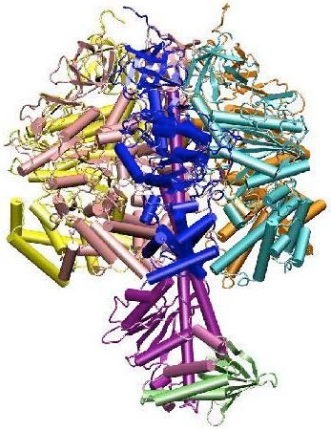
Biomolecular Time Scales

Motion	Time Scale (sec)
Bond stretching	10^{-14} to 10^{-13}
Elastic vibrations	10^{-12} to 10^{-11}
Rotations of surface sidechains	10^{-11} to 10^{-10}
Hinge bending	10^{-11} to 10^{-7}
Rotation of buried side chains	10^{-4} to 1 sec
Allosteric transistions	10^{-5} to 1 sec
Local denaturations	10^{-5} to 10 sec



NAMD: Scalable Molecular Dynamics

2002 Gordon Bell Award



ATP synthase

PSC Lemieux

51,000 Users, 2900 Citations

Computational Biophysics Summer School

Blue Waters Target Application

Illinois Petascale Computing Facility

GPU Acceleration

NVIDIA Tesla

NCSA Lincoln

Parallel Programming Lab
University of Illinois at Urbana-Champaign

Siebel Center for Computer Science

http://charm.cs.illinois.edu/

National Center for Research Resources

NIH BTRC for Macromolecular Modeling and Bioinformatics
http://www.ks.uiuc.edu/

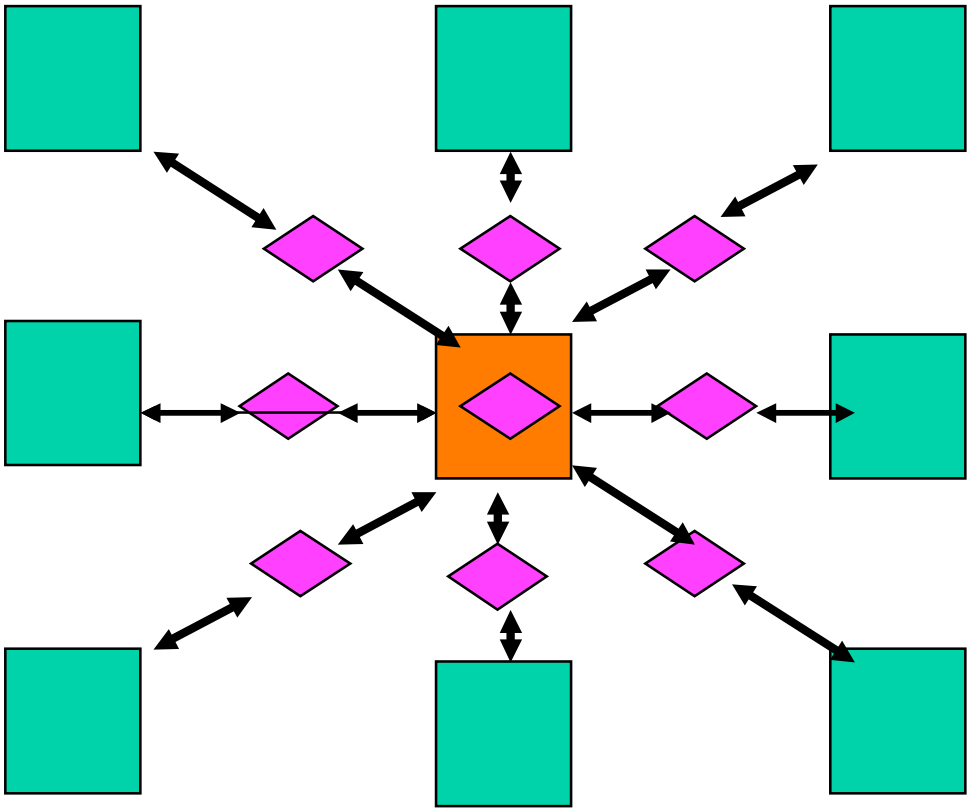
Beckman Institute, UIUC

Charm++ Used by NAMD

- Parallel C++ with *data driven* objects
- Asynchronous method invocation
- Prioritized scheduling of messages/execution
- Measurement-based load balancing
- Portable messaging layer
- Portable multithreading model
- *Projections* performance analysis tool

NAMD Hybrid Decomposition

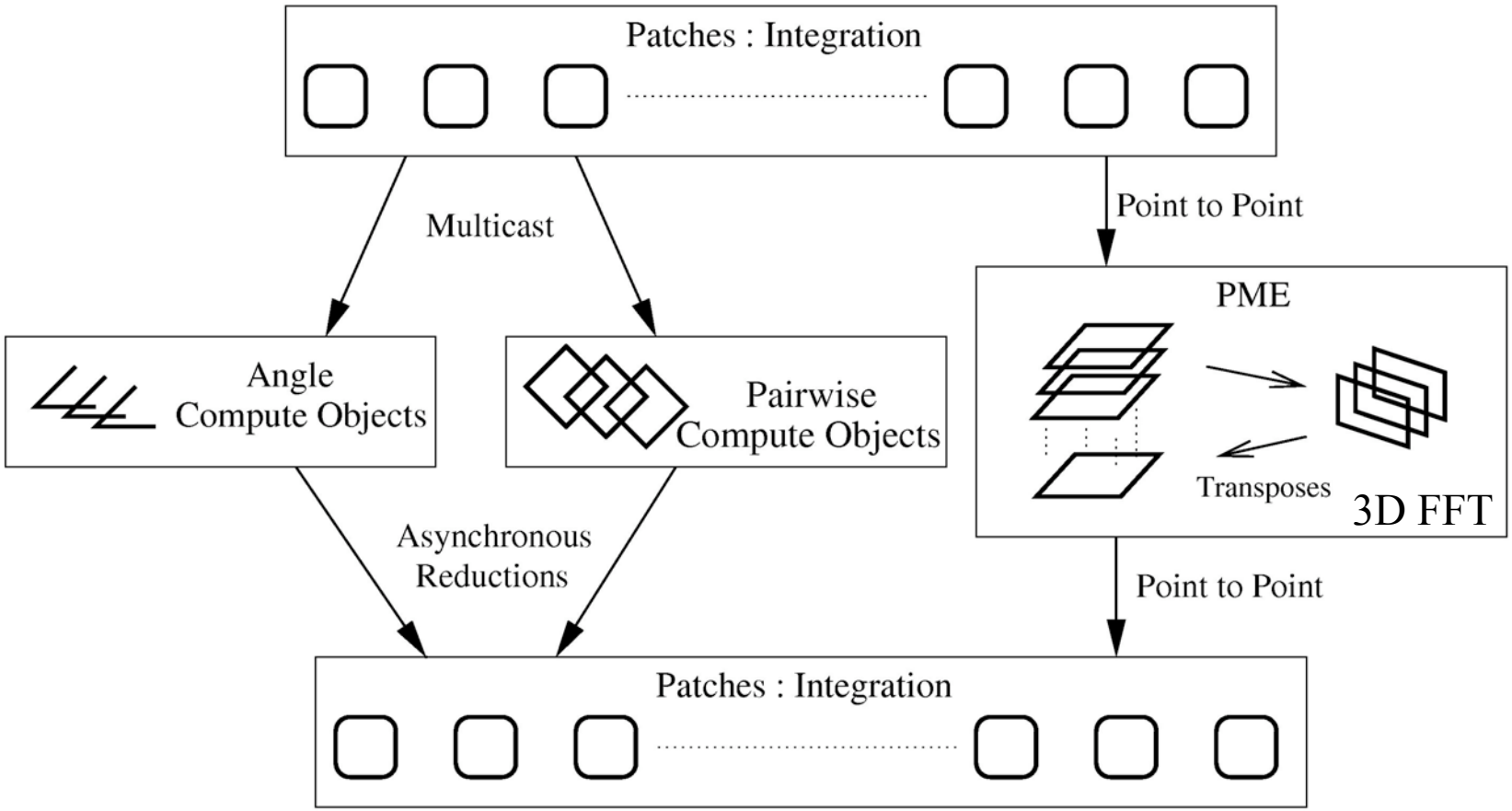
Kale *et al., J. Comp. Phys.* **151**:283-312, 1999.



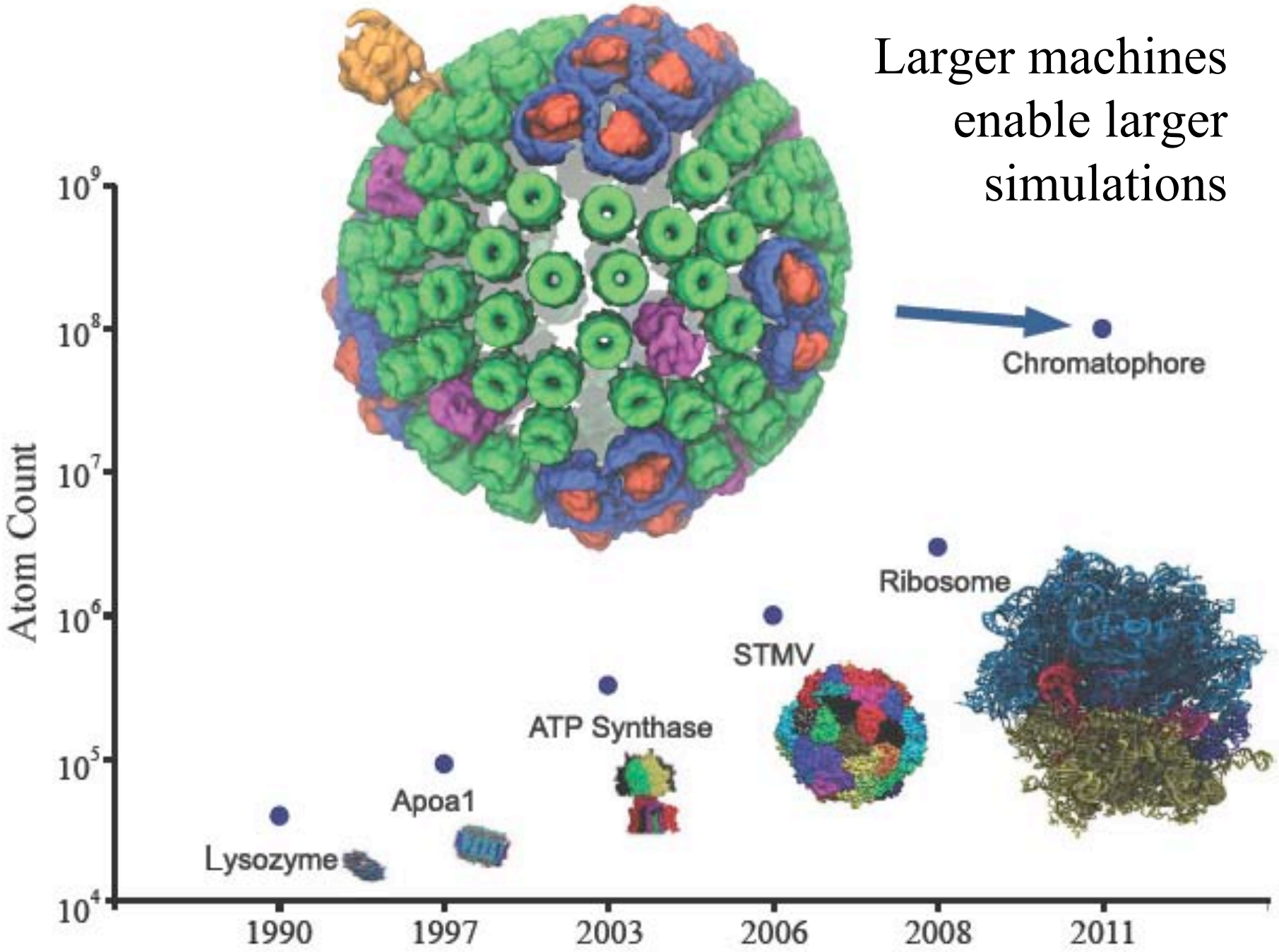
- Spatially decompose data and communication.
- Separate but related work decomposition.
- "Compute objects" facilitate iterative, measurement-based load balancing system.

NAMD Overlapping Execution

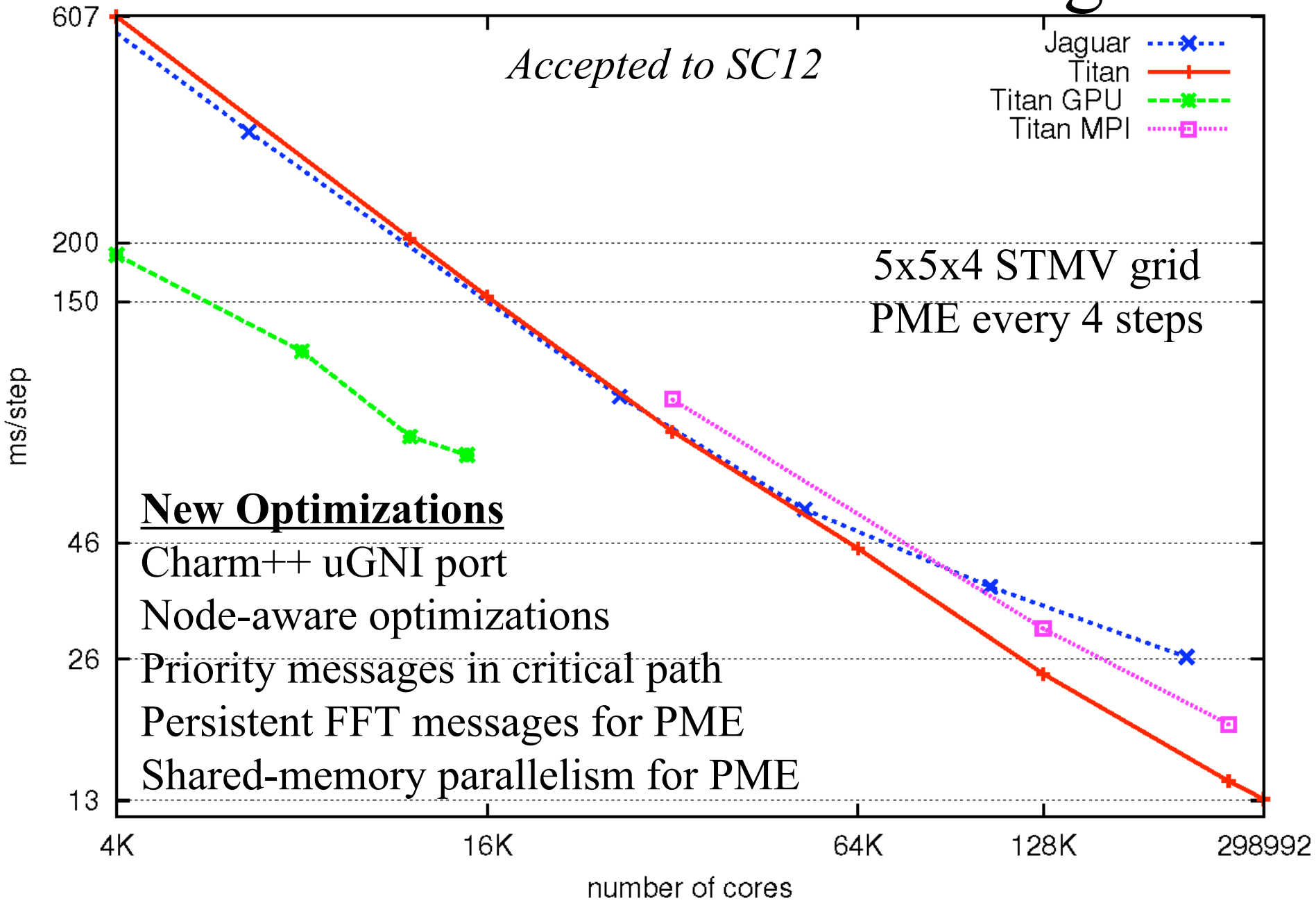
Phillips *et al., SC2002*.



Objects are assigned to processors and queued as data arrives.


Larger machines enable larger simulations
Chromatophore
Ribosome
STMV
ATP Synthase
Apoa1
Lysozyme
Atom Count
10^9
10^8
10^7
10^6
10^5
10^4
1990
1997
2003
2006
2008
2011

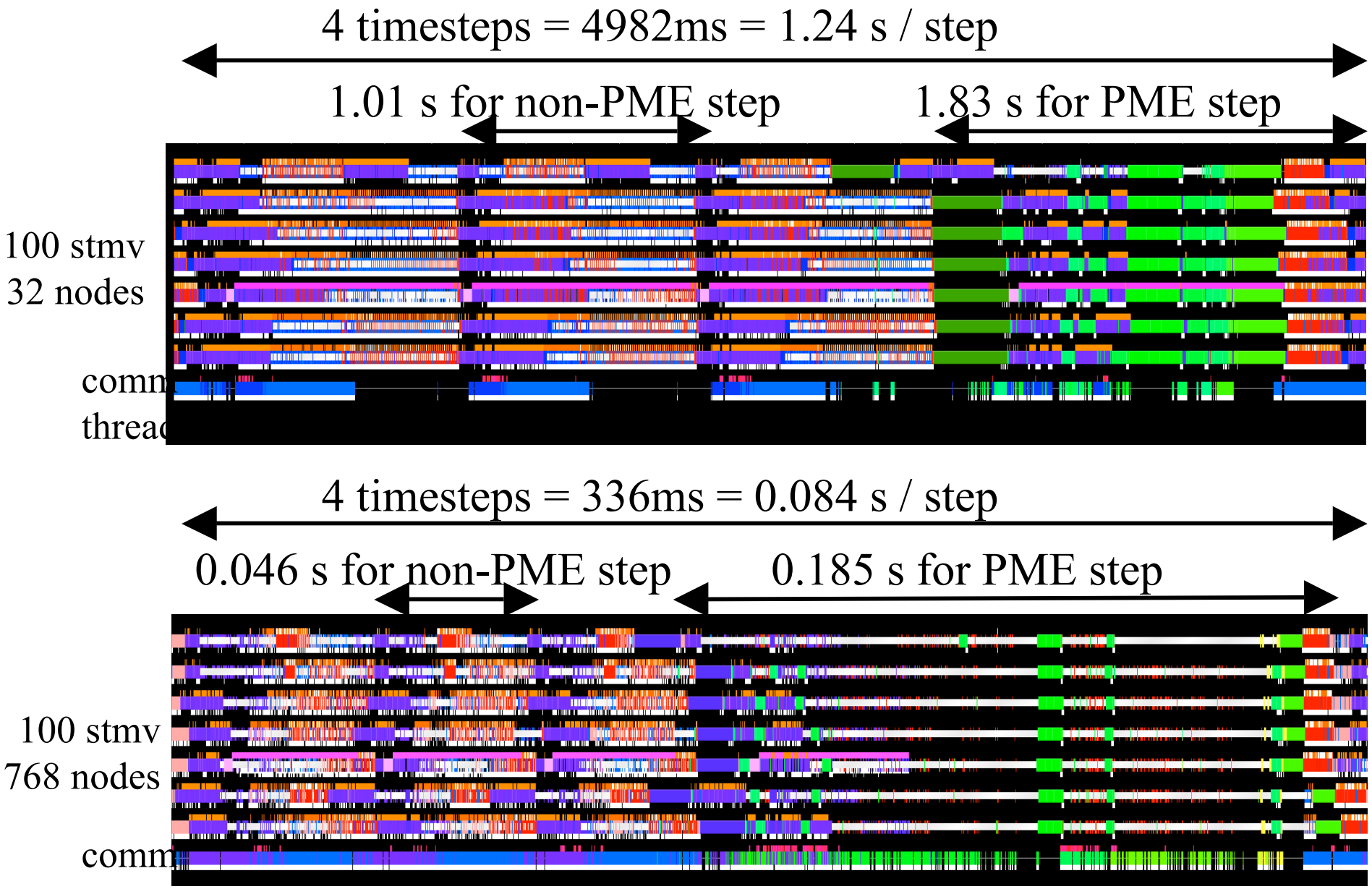
100M Atoms on Titan vs Jaguar



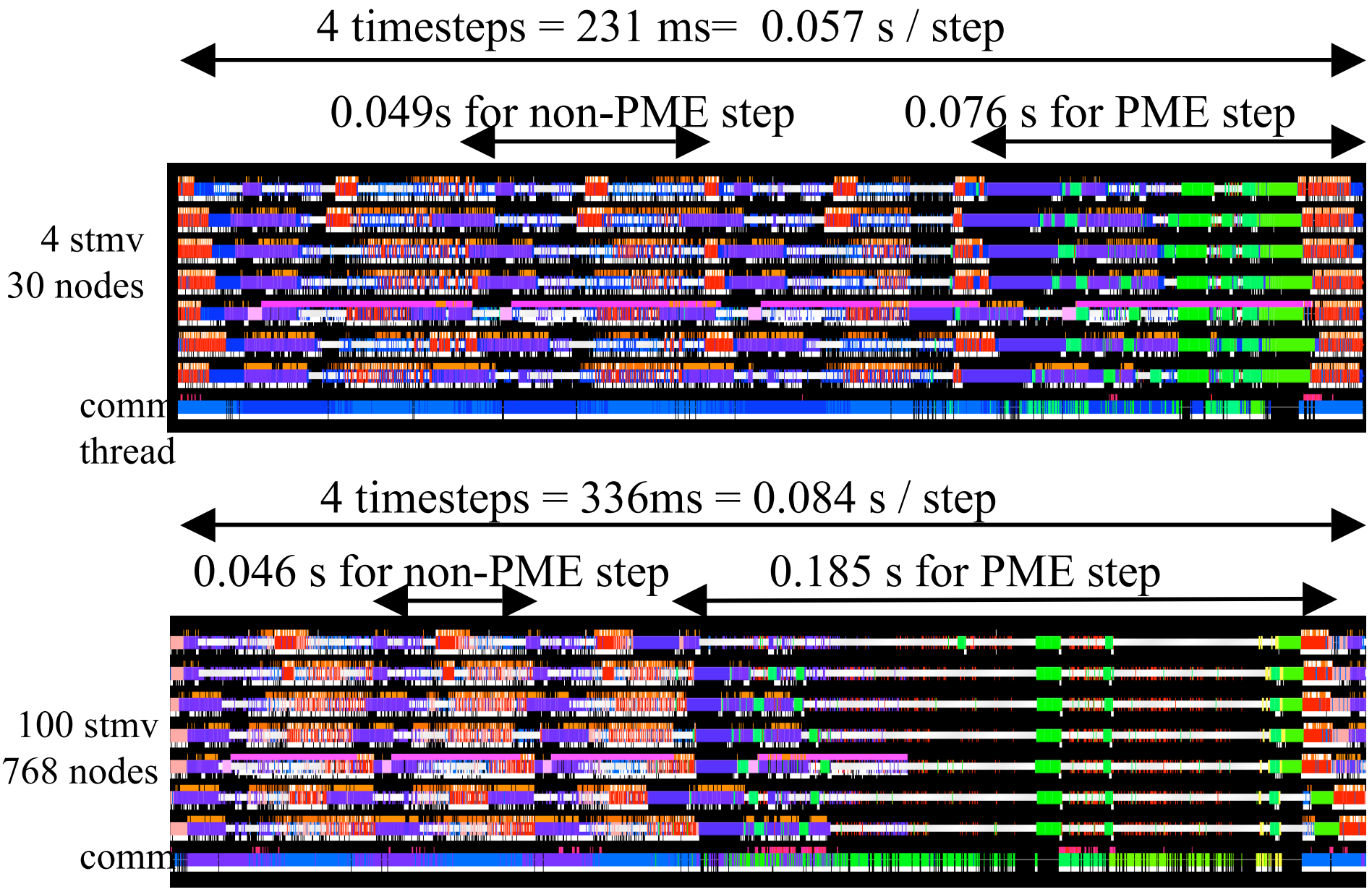
Challenges

- Variety of accelerator programming models
  - CUDA, OpenCL, OpenACC, Intel ???
- Vector instructions for atom-based math
  - Also variety of vector widths
- Tight coupling of threads on a single core
  - Intel MIC and Hyperthreads, IBM BG/Q
- Network latency and bisection bandwidth

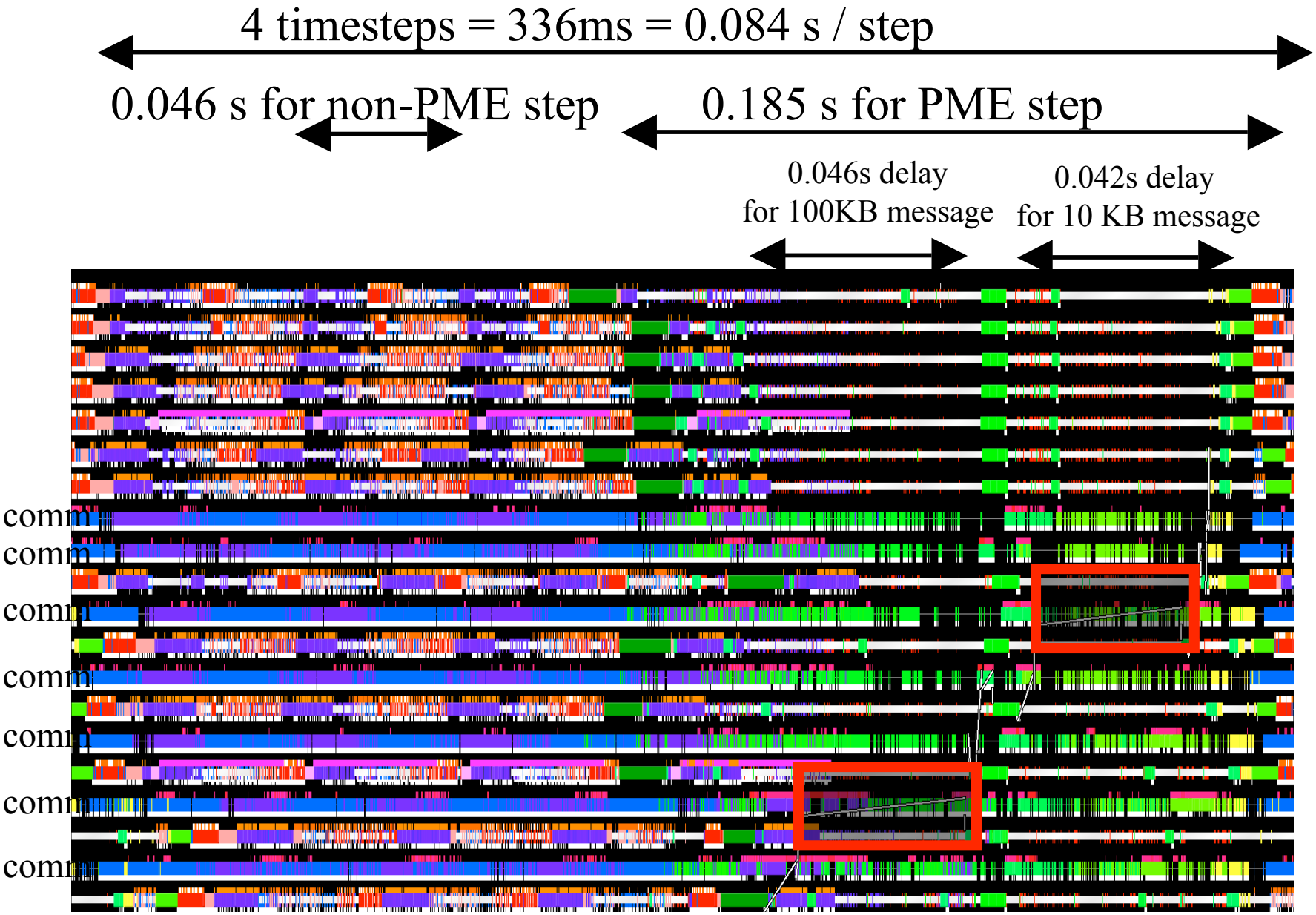
TitanDev Strong Scaling



TitanDev Weak Scaling



PME delays – tracing data needed for one ungrid calculation



NAMD I/O

- Modest data written incrementally during run
  - 4.8 GB restart, 1.2GB frame for 100M atoms
- Traditionally single I/O node, POSIX
- Distributed (<100) due to memory limits
- Challenge is Lustre performance for:
  - Multiple writers to single file
  - Simultaneous open by multiple nodes
- Dedicated I/O threads possible solution?

VMD – "Visual Molecular Dynamics"

- Visualization and analysis of molecular dynamics simulations, sequence data, volumetric data, quantum chemistry simulations, particle systems, …
- User extensible with scripting and plugins
- http://www.ks.uiuc.edu/Research/vmd/