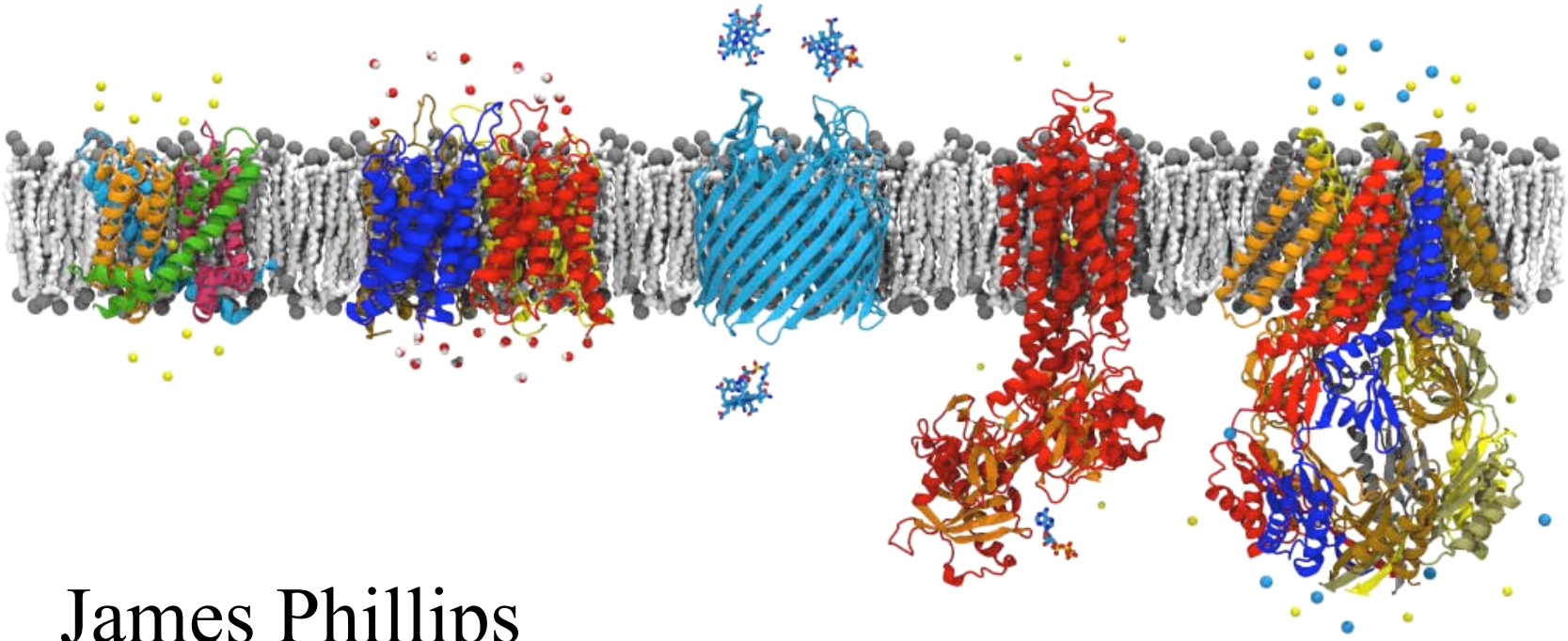
# Petascale Molecular Dynamics Simulations on GPU-Accelerated Supercomputers

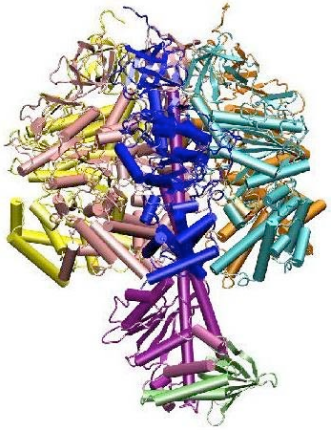


James Phillips
Beckman Institute, University of Illinois
http://www.ks.uiuc.edu/Research/namd/

**GTC 2012**

# 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

# NIH BTRC for Macromolecular Modeling and Bioinformatics

**1990-2017**

**Beckman Institute**
**University of Illinois at**
**Urbana-Champaign**

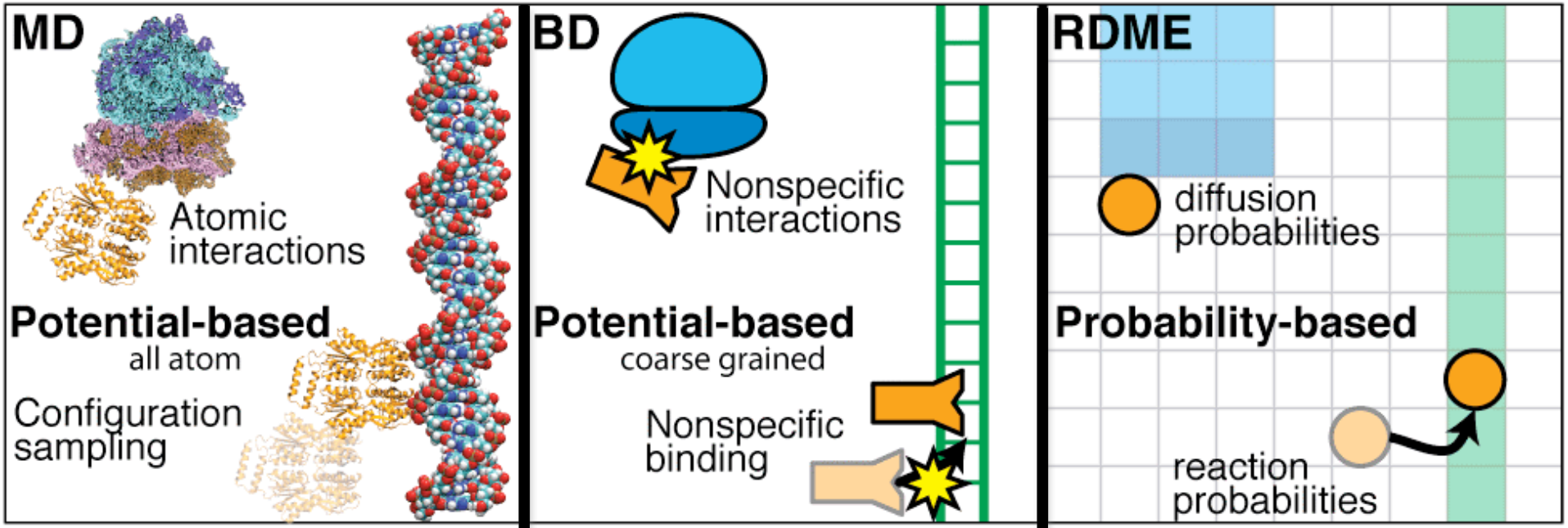
# Physics of *in vivo* Molecular Systems

***Biomolecular interactions span many orders of magnitude in space and time.***

***Center software provides multi-scale computational modeling.***

**femtoseconds** → **hours**

**Ångstrom** → **microns**




- NAMD — Scalable Molecular Dynamics
- MDFF — Molecular Dynamics Flexible Fitting
- HMMM — Highly Mobile Membrane Mimetic
- VMD — Visual Molecular Dynamics

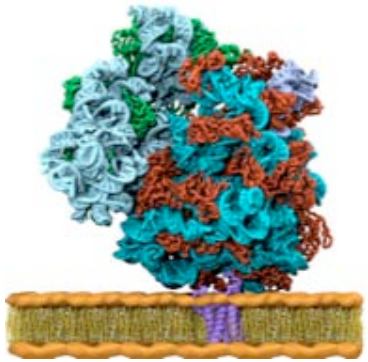
- BrownianMover — Brownian Dynamics
- VMD — Visual Molecular Dynamics
- NAMD — Scalable Molecular Dynamics

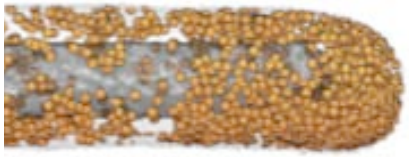
**BD:** Brownian Dynamics

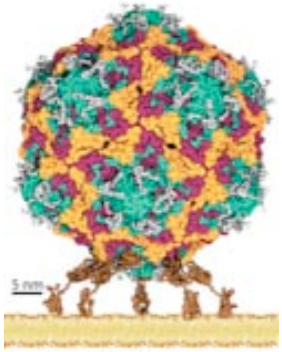
- LatticeMicrobes — Whole Cell Simulations
- VMD — Visual Molecular Dynamics

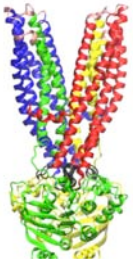
**RDME:** Reaction-diffusion master equation

# Collaborative Driving Projects

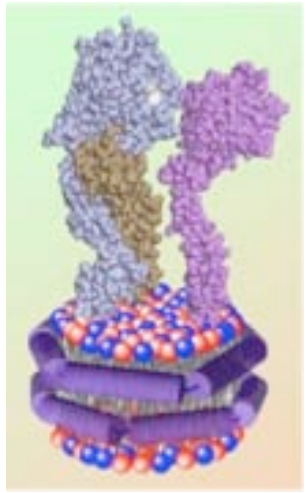


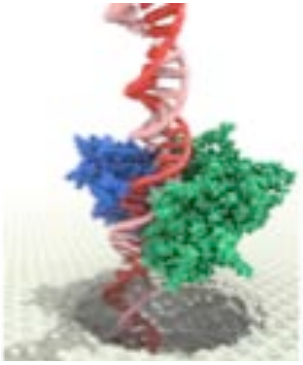


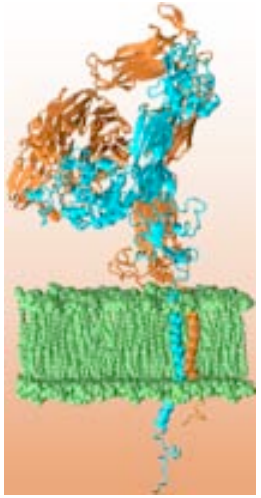




| | |
|---|---|
| 1. Ribosome | R. Beckmann (U. Munich)<br>J. Frank (Columbia U.)<br>T. Ha(UIUC)<br>K. Fredrick (Ohio state U.)<br>R. Gonzalez (Columbia U.) |
| 2. Blood Coagulation Factors | J. Morrissey (UIUC)<br>S. Sligar (UIUC)<br>C. Rienstra (UIUC)<br>G. Gilbert (Harvard) |
| 3. Whole Cell Behavior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

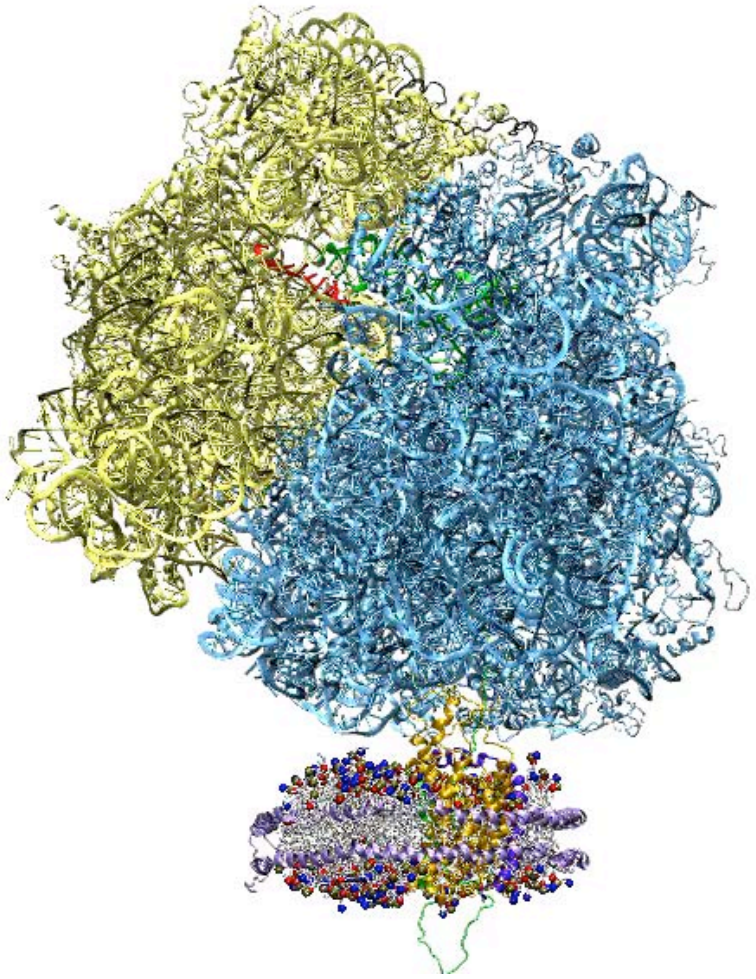






# Ribosome Driving Project

Target of over 50%
of antibiotics



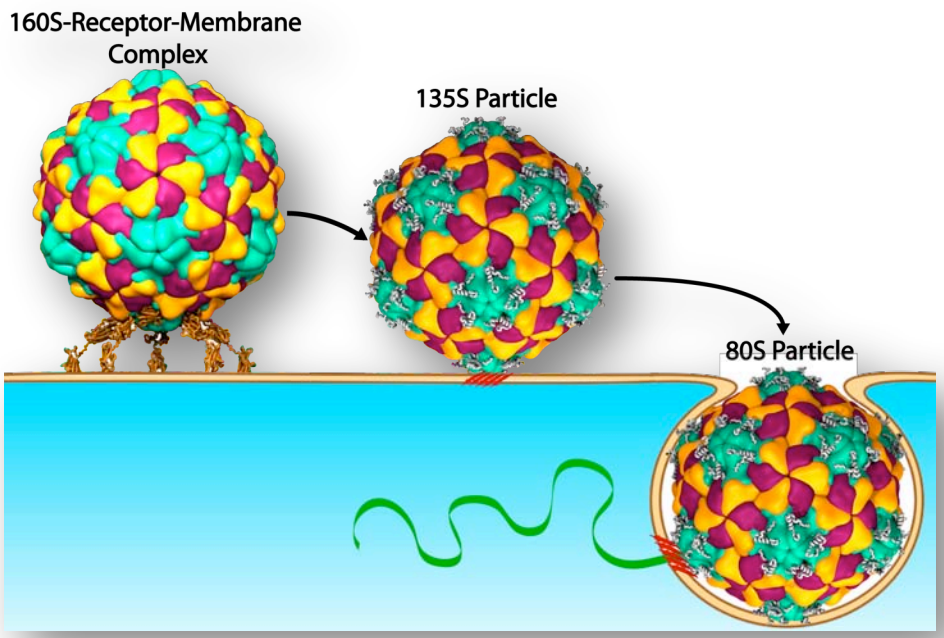
Many related diseases. e.g. Alzheimer's disease due to dysfunctional ribosome (J. Neuroscience 2005, 25:9171-9175)

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

# Viral Infection Driving Projects

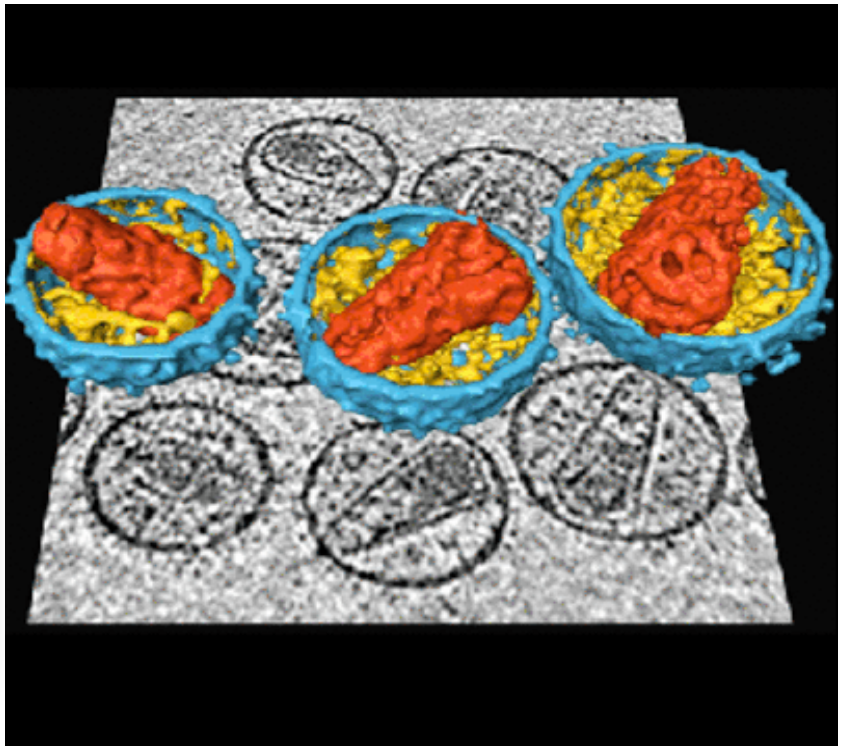
## Poliovirus

Poliovirus is a model system for understanding how non-enveloped viruses bind to and enter a host cell.




## Human Immunodeficiency Virus 1

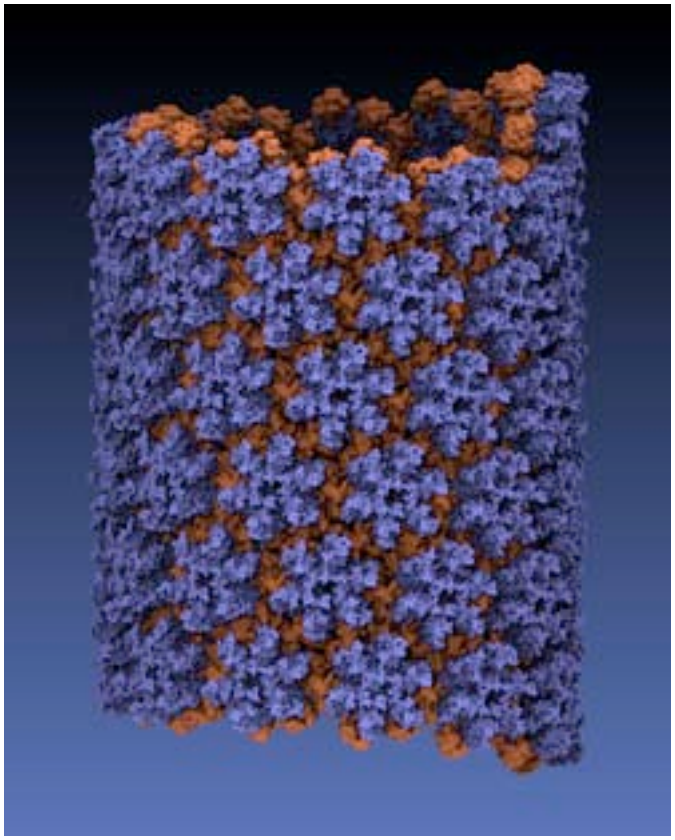
Knowledge of HIV capsid atomic structure may reveal disassembly mechanism and guide novel therapies.



Briggs et al. ***Structure*** (2006) 14:15-20.

# Blue Waters Early Science Project

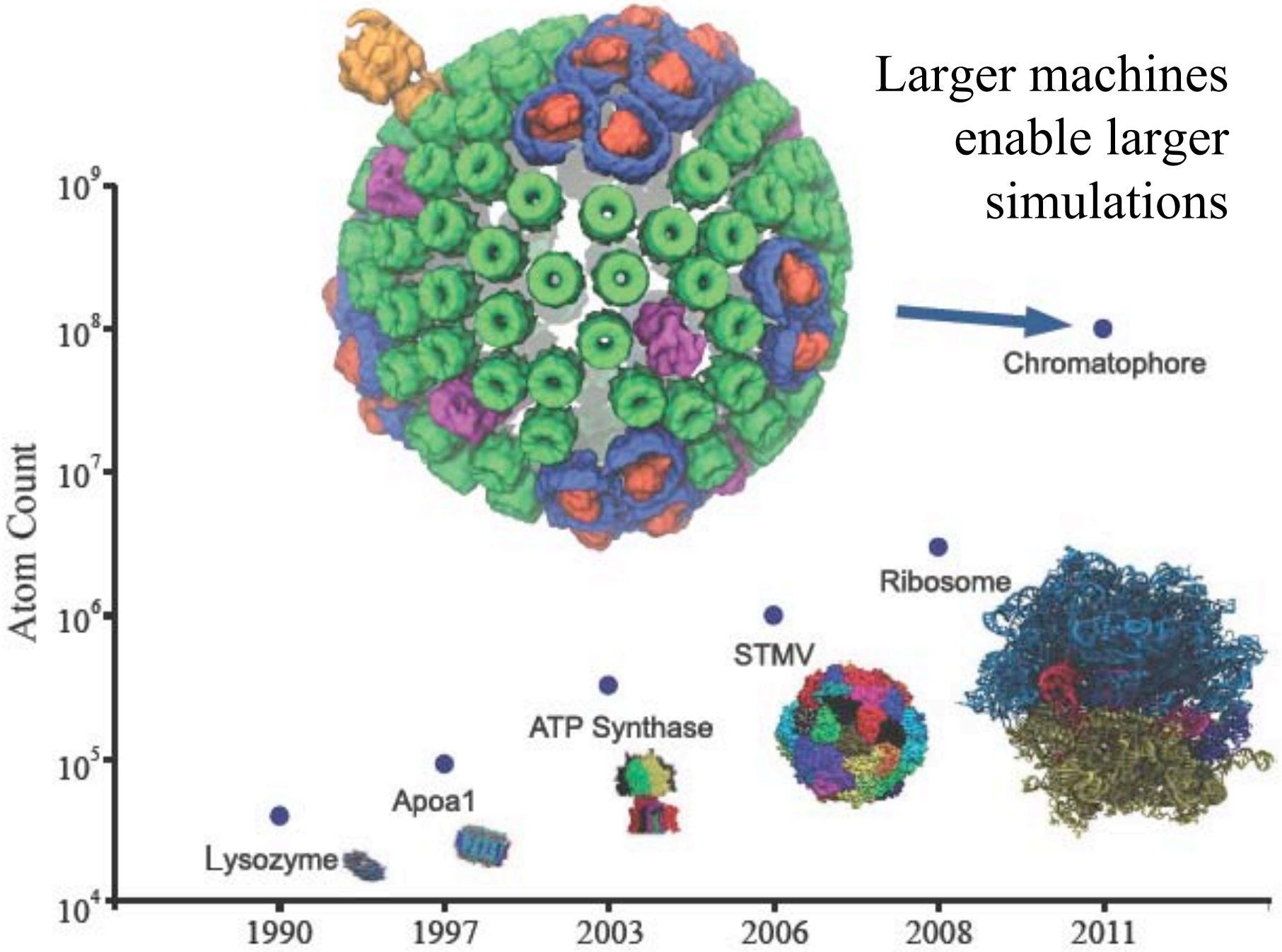
"The first all-atom structure of an **HIV virus capsid** in its tubular form, courtesy Klaus Schulten, University of Illinois at Urbana-Champaign Theoretical and Computational Biophysics Group/Beckman Institute; Angela Gronenborn and Peijun Zhang, University of Pittsburgh School of Medicine Center for HIV Protein Interactions/Department of Structural Biology."



# NAMD Petascale Preparations (2010)

# NAMD Petascale Preparations (2012)


# Larger machines enable larger simulations



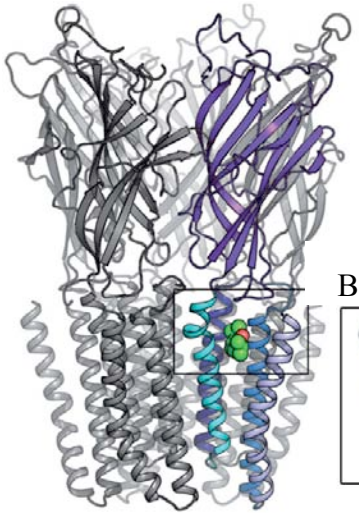
# NAMD impact is broad and deep

- Comprehensive, industrial-quality software
  - Integrated with VMD for simulation setup and analysis
  - Portable extensibility through Tcl scripts (also used in VMD)
  - Consistent user experience from laptop to supercomputer
- Large user base – 51,000 users
  - 9,100 (18%) are NIH-funded; many in other countries
  - 14,100 have downloaded more than one version
- Leading-edge simulations
  - "most-used software" on NICS Cray XT5 (largest NSF machine)
  - "by far the most used MD package" at TACC (2nd and 3rd largest)
  - NCSA Blue Waters early science projects and acceptance test
  - Argonne Blue Gene/Q early science project

# Outside researchers choose NAMD and succeed

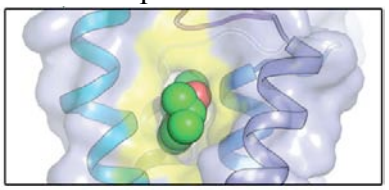
**2100 external citations since 2007**

Corringer, et al., *Nature*, 2011

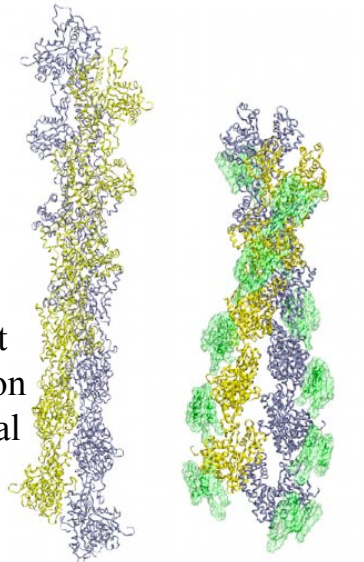


180K-atom 30 ns study of anesthetic binding to bacterial ligand-gated ion channel provided "complementary interpretations…that could not have been deduced from the static structure alone."

Bound Propofol Anesthetic



Voth, et al., *PNAS*, 2010



Bare actin Cofilactin

500K-atom 500 ns investigation of effect of actin depolymerization factor/cofilin on mechanical properties and conformational dynamics of actin filament.

## Recent NAMD Simulations in *Nature*

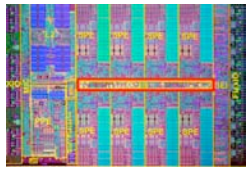
- **M. Koeksal, et al.,** *Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis*. (2011)
- **C.-C. Su, et al.**, *Crystal structure of the CusBA heavy-metal efflux complex of Escherichia coli.* (2011)
- **D. Slade, et al.**, *The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase*. (2011)
- **F. Rose, et al.,** *Mechanism of copper(II)-induced misfolding of Parkinson's disease protein*. (2011)
- **L. G. Cuello, et al.**, *Structural basis for the coupling between activation and inactivation gates in K(+) channels*. (2010)
- **S. Dang, et al.,**, *Structure of a fucose transporter in an outward-open conformation.* (2010)
- **F. Long, et al.**, *Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport*. (2010)
- **R. H. P. Law, et al.**, *The structural basis for membrane binding and pore formation by lymphocyte perforin.* (2010)
- **P. Dalhaimer and T. D. Pollard**, *Molecular Dynamics Simulations of Arp2/3 Complex Activation.* (2010)
- **J. A. Tainer, et al.**, *Recognition of the Ring-Opened State of Proliferating Cell Nuclear Antigen by Replication Factor C Promotes Eukaryotic Clamp-Loading*. (2010)
- **D. Krepkiy, et al.,**, *Structure and hydration of membranes embedded with voltage-sensing domains*. (2009)
- **N. Yeung, et al.,**, *Rational design of a structural and functional nitric oxide reductase*. (2009)
- **Z. Xia, et al.**, *Recognition Mechanism of siRNA by Viral p19 Suppressor of RNA Silencing: A Molecular Dynamics Study*. (2009)

# Our Goal: Practical Acceleration

- Broadly applicable to scientific computing
  - Programmable by domain scientists
  - Scalable from small to large machines
- Broadly available to researchers
  - Price driven by commodity market
  - Low burden on system administration
- Sustainable performance advantage
  - Performance driven by Moore's law
  - Stable market and supply chain

# Acceleration Options for NAMD

- Outlook in 2005-2006:
  - FPGA reconfigurable computing (with NCSA)
    - Difficult to program, slow floating point, expensive
  - Cell processor (NCSA hardware)
    - Relatively easy to program, expensive
  - ClearSpeed (direct contact with company)
    - Limited memory and memory bandwidth, expensive
  - MDGRAPE
    - Inflexible and expensive
  - Graphics processor (GPU)
    - Program must be expressed as graphics operations

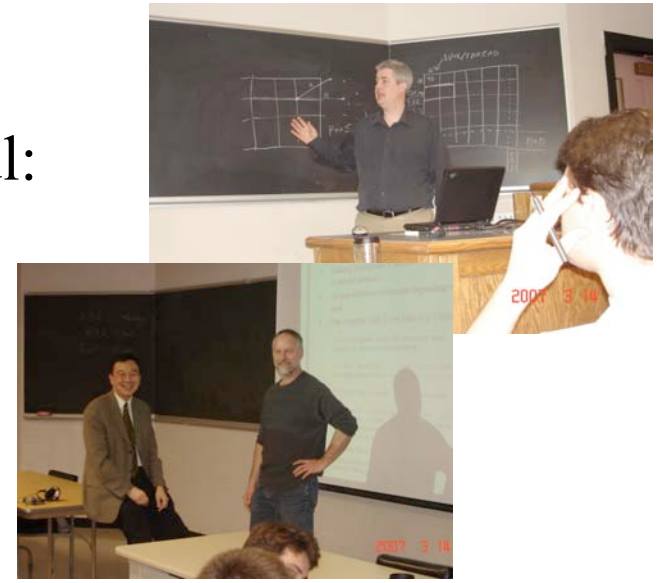


# CUDA: Practical Performance

*November 2006: NVIDIA announces CUDA for G80 GPU.*

- CUDA makes GPU acceleration usable:
  - Developed and supported by NVIDIA.
  - No masquerading as graphics rendering.
  - New shared memory and synchronization.
  - No OpenGL or display device hassles.
  - Multiple processes per card (or vice versa).
- Resource and collaborators make it useful:
  - Experience from VMD development
  - David Kirk (Chief Scientist, NVIDIA)
  - Wen-mei Hwu (ECE Professor, UIUC)


Fun to program (and drive)



Stone *et al.*, *J. Comp. Chem.* **28**:2618-2640, 2007.

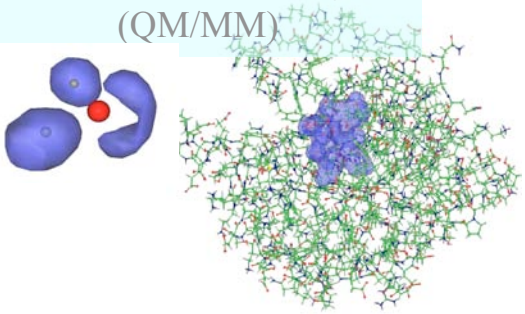
# Parallel Programming Lab
## University of Illinois at Urbana-Champaign

Siebel Center for Computer Science

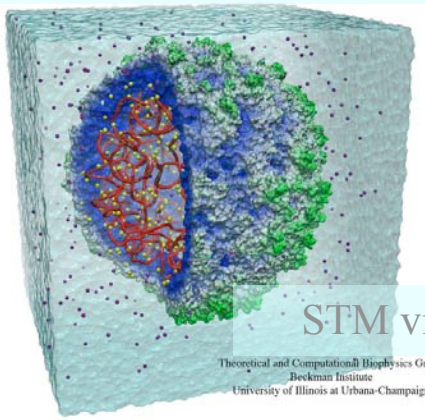
http://charm.cs.illinois.edu/

# Develop abstractions in context of full-scale applications

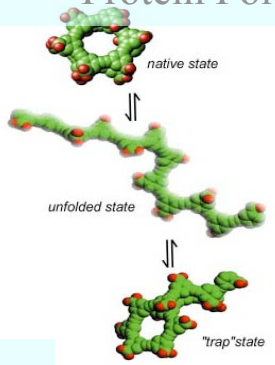
Quantum Chemistry (QM/MM)



NAMD: Molecular Dynamics

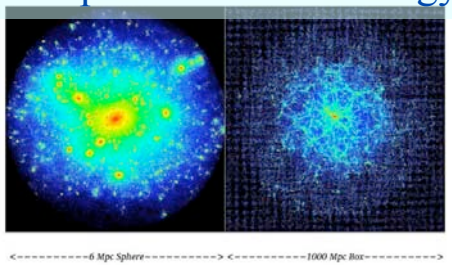


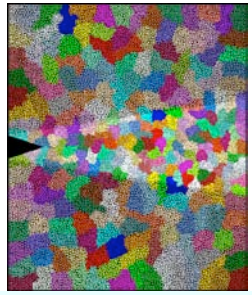

Protein Folding



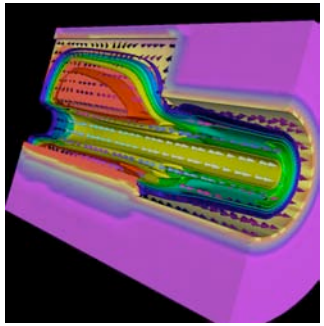

STM virus simulation

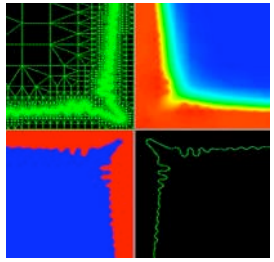
Computational Cosmology



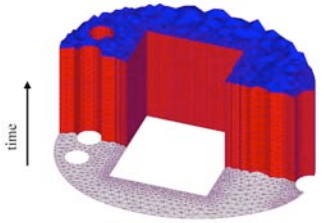



Parallel Objects,
Adaptive Runtime System
Libraries and Tools





Crack Propagation




Rocket Simulation

Dendritic Growth

Space-time meshes

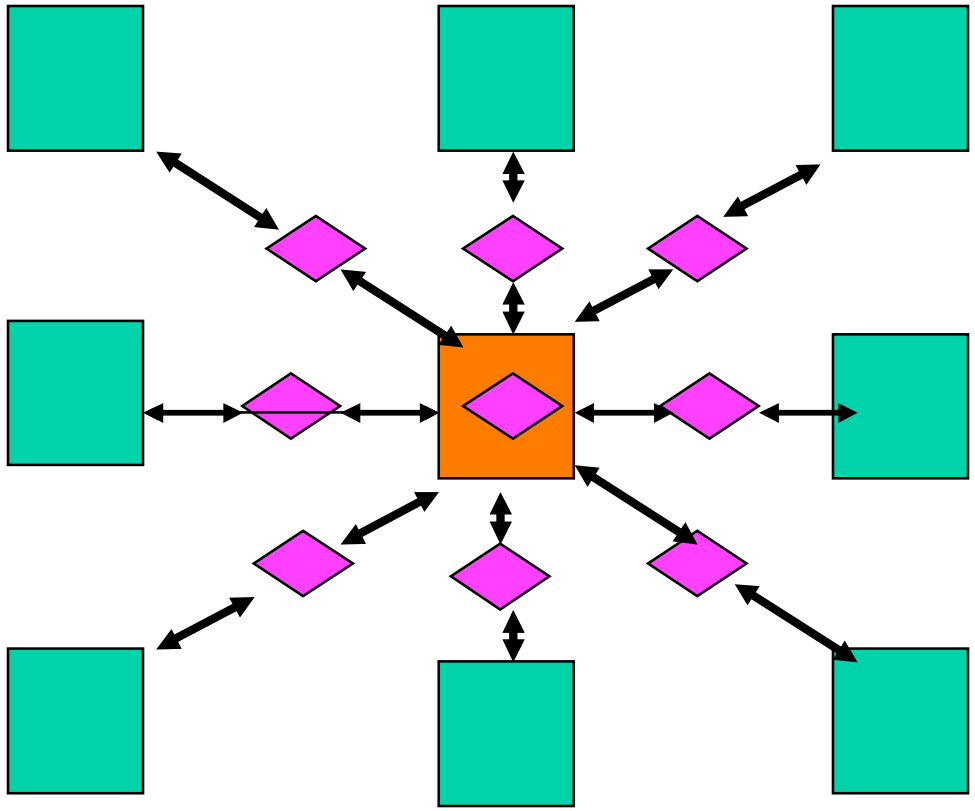
The enabling CS technology of parallel objects and intelligent Runtime systems has led to several collaborative applications in CSE

# 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.

# NAMD Hybrid Decomposition

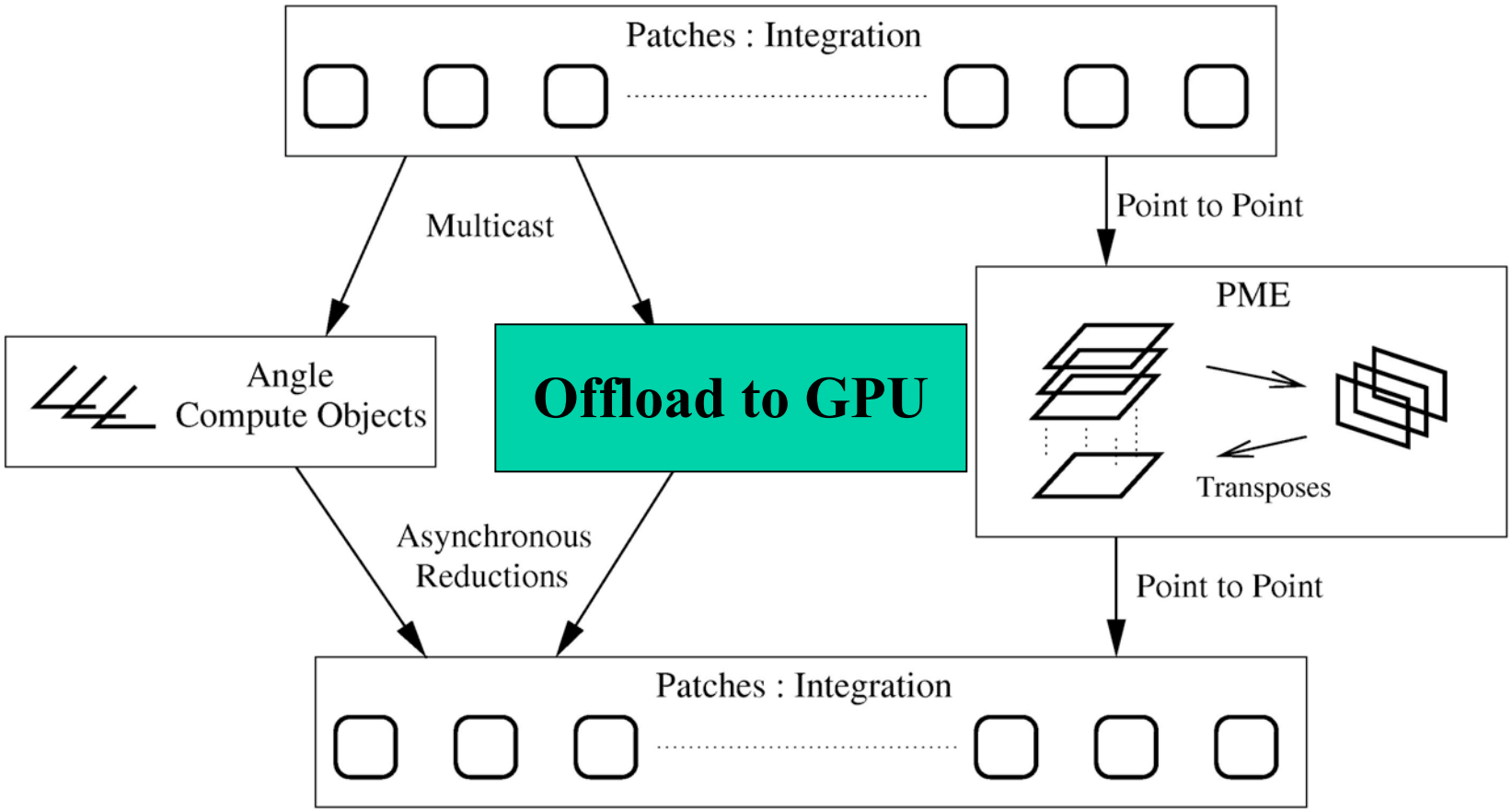
Kale *et al., J. Comp. Phys.* **151**:283-312, 1999.



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

# NAMD Overlapping Execution

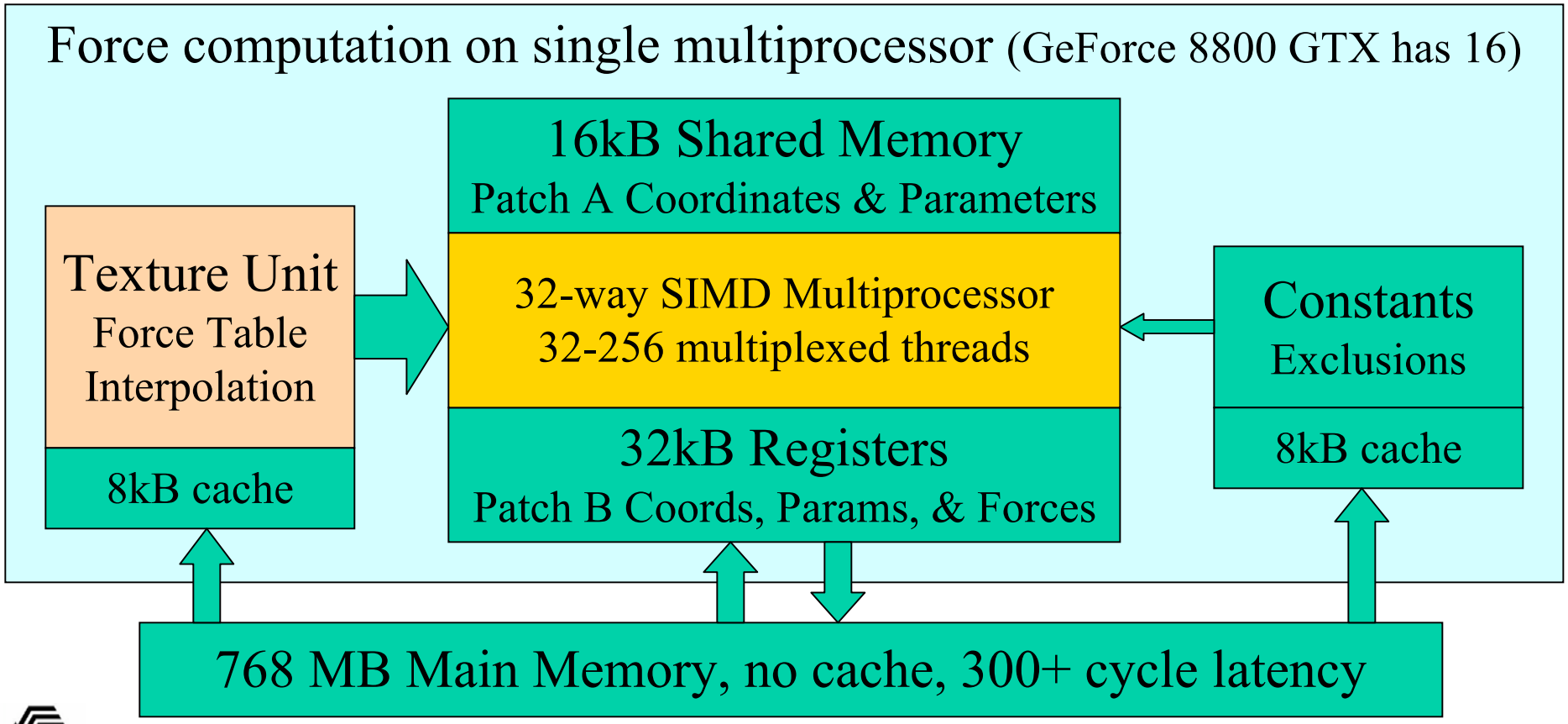
Phillips *et al., SC2002*.




Objects are assigned to processors and queued as data arrives.

# Nonbonded Forces on CUDA GPU

- Start with most expensive calculation: direct nonbonded interactions.
- Decompose work into pairs of patches, identical to NAMD structure.
- GPU hardware assigns patch-pairs to multiprocessors dynamically.



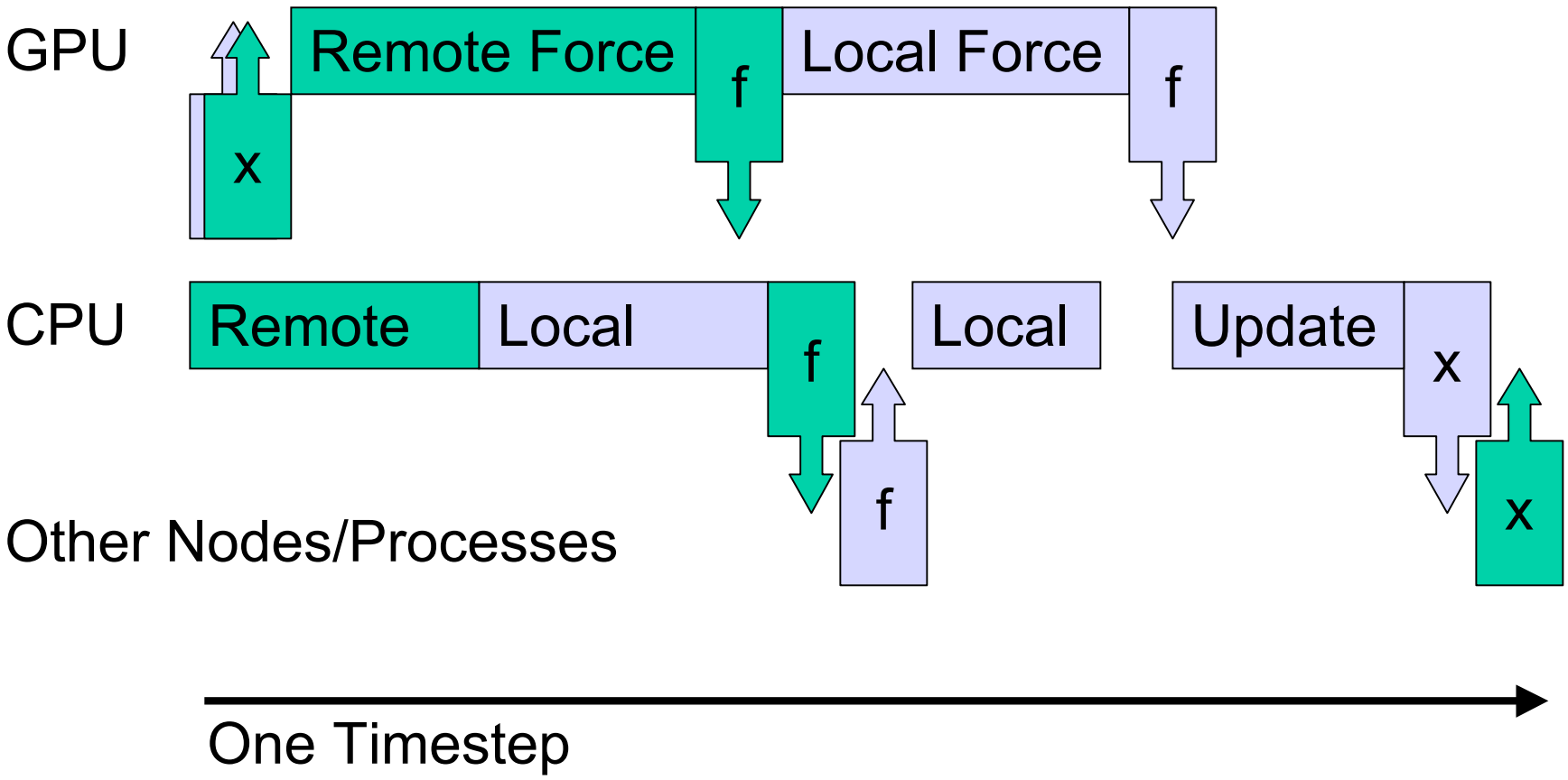

National Center for Research Resources

Stone *et al.*, *J. Comp. Chem.* **28**:2618-2640, 2007.

# Nonbonded Forces CUDA Code

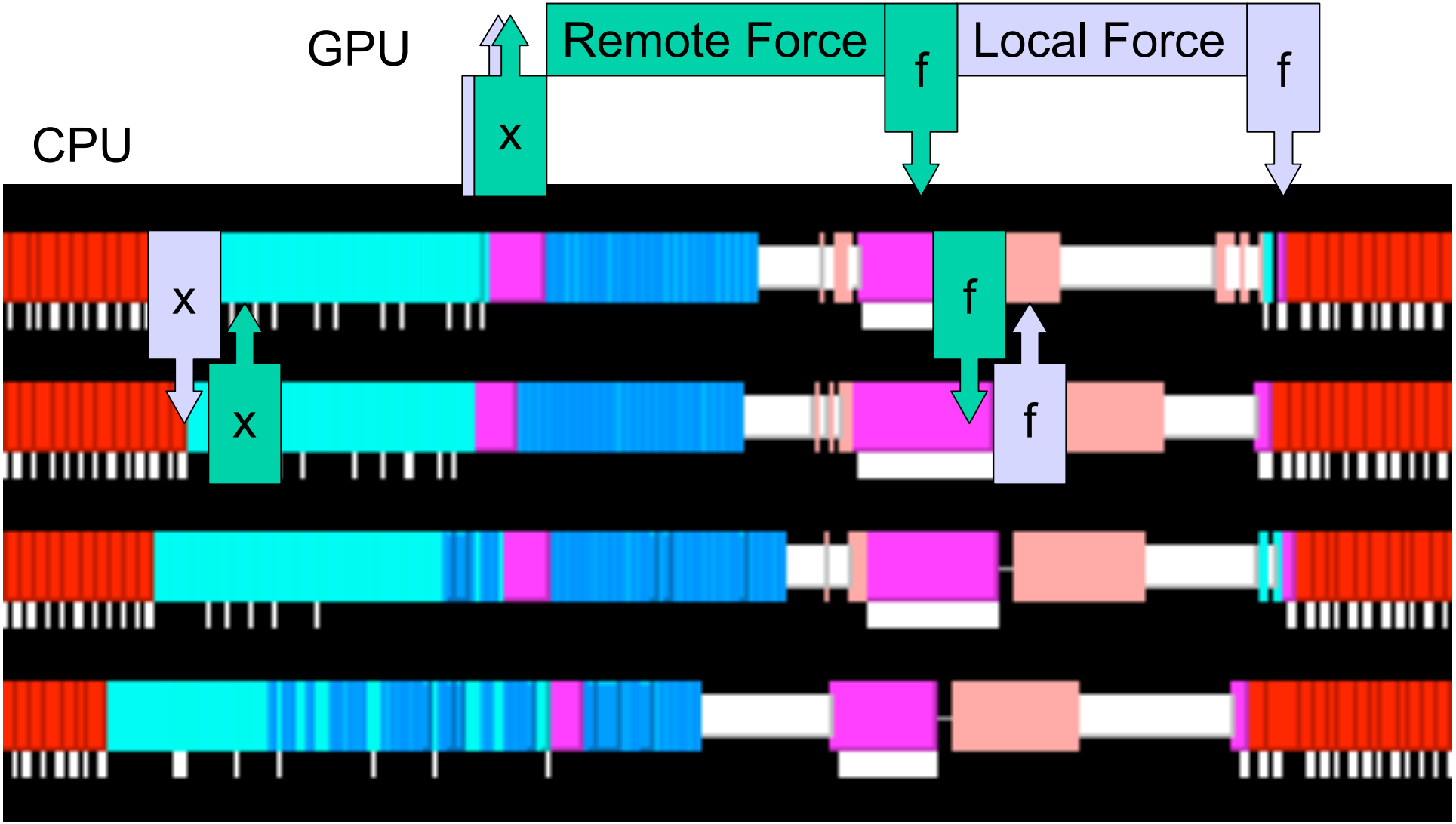
```
texture<float4> force_table;
__constant__ unsigned int exclusions[];
__shared__ atom jatom[];
atom iatom;      // per-thread atom, stored in registers
float4 iforce;   // per-thread force, stored in registers
for ( int j = 0; j < jatom_count; ++j ) {
  float dx = jatom[j].x - iatom.x;   float dy = jatom[j].y - iatom.y;  float dz = jatom[j].z - iatom.z;
  float r2 = dx*dx + dy*dy + dz*dz;
  if ( r2 < cutoff2 ) {
    float4 ft = texfetch(force_table, 1.f/sqrt(r2));                          // Force Interpolation
    bool excluded = false;
    int indexdiff = iatom.index - jatom[j].index;                             // Exclusions
    if ( abs(indexdiff) <= (int) jatom[j].excl_maxdiff ) {
      indexdiff += jatom[j].excl_index;
      excluded = ((exclusions[indexdiff>>5] & (1<<(indexdiff&31))) != 0);
    }
    float f = iatom.half_sigma + jatom[j].half_sigma;  // sigma                 // Parameters
    f *= f*f;  // sigma^3
    f *= f;  // sigma^6
    f *= ( f * ft.x + ft.y );  // sigma^12 * fi.x - sigma^6 * fi.y
    f *= iatom.sqrt_epsilon * jatom[j].sqrt_epsilon;
    float qq = iatom.charge * jatom[j].charge;
    if ( excluded ) { f = qq * ft.w; }  // PME correction
    else { f += qq * ft.z; }  // Coulomb
    iforce.x += dx * f;   iforce.y += dy * f;    iforce.z += dz * f;          // Accumulation
    iforce.w += 1.f;  // interaction count or energy
  }
}
```

# Overlapping GPU and CPU with Communication



# Actual Timelines from NAMD

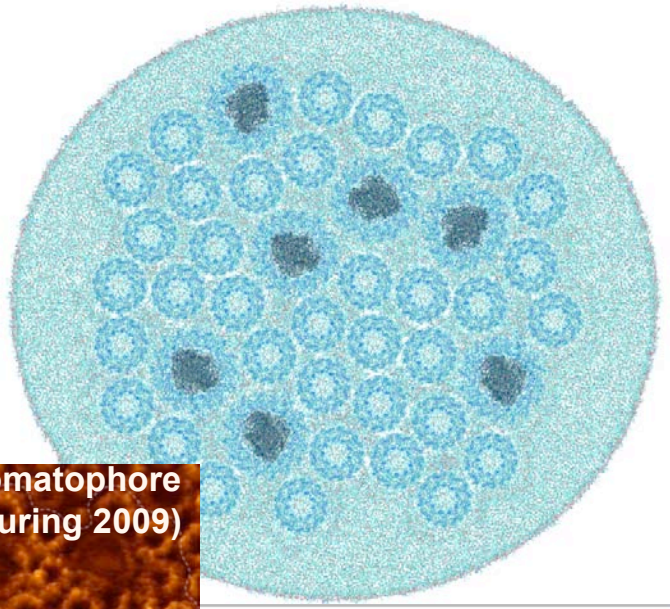
Generated using Charm++ tool “Projections” http://charm.cs.uiuc.edu/

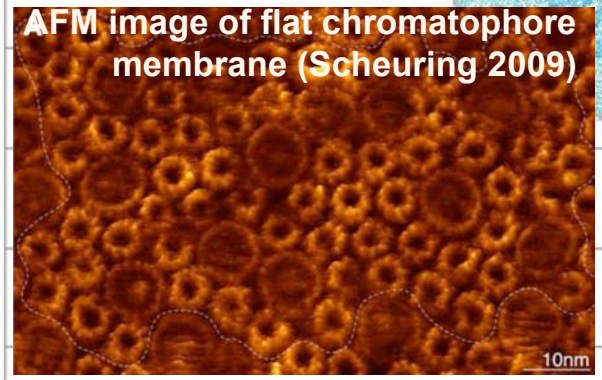


# Tsubame (Tokyo) Application of GPU Accelerated NAMD (fall 2011)

*20 million atom proteins + membrane*

8400 cores





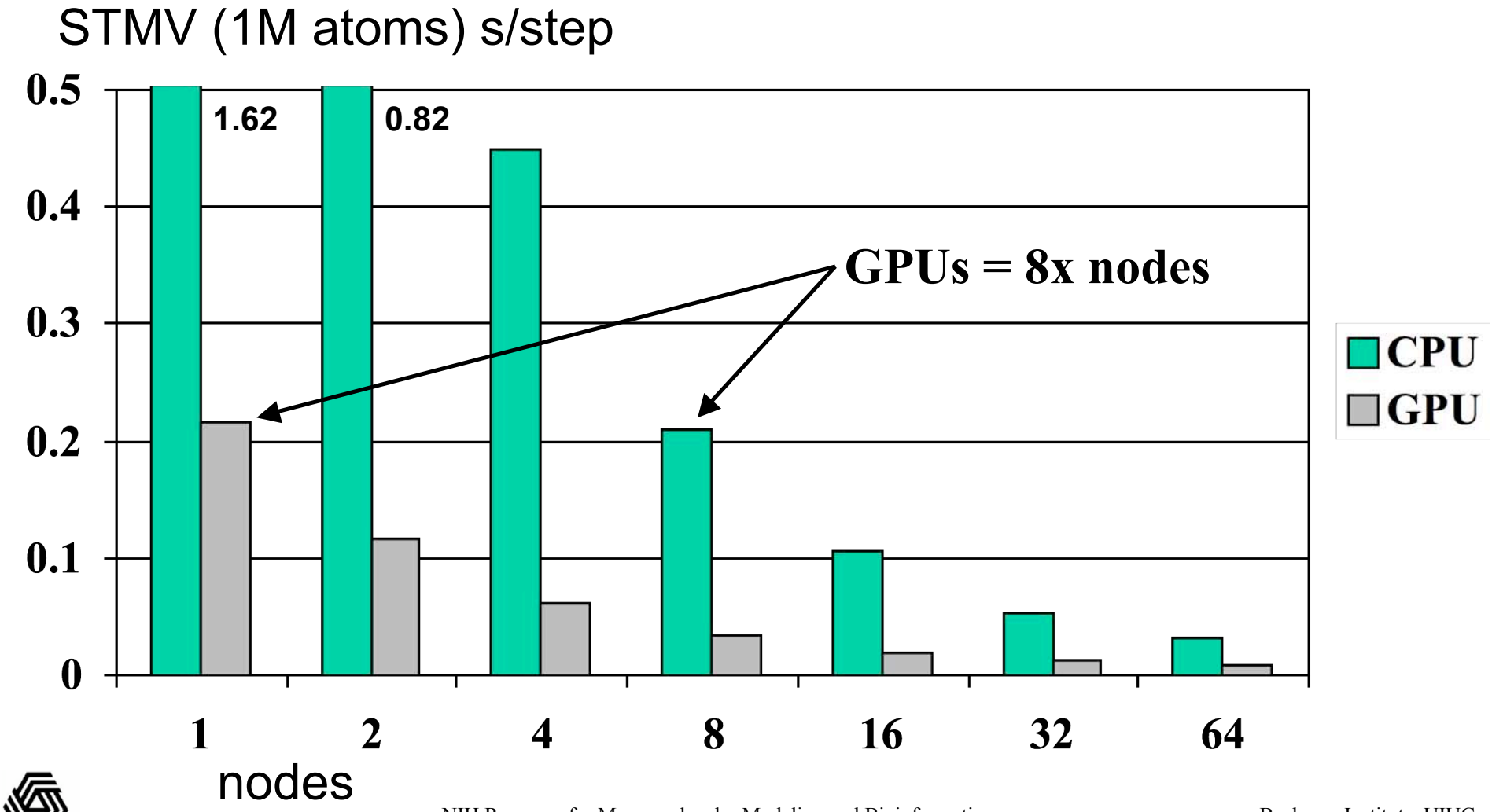




| | 4nodes | 8nodes | 32nodes | 64nodes | 128nodes | 256nodes | 512nodes | 700nodes |
|---|---|---|---|---|---|---|---|---|
| CPU 12 cores | 0.02 | 0.04 | 0.14 | 0.23 | 0.41 | 0.92 | 1.69 | 2.23 |
| CPU 12 cores + 1 GPU | N/A | N/A | 0.42 | 0.80 | 1.50 | 2.93 | 5.01 | 5.98 |
| CPU 12 cores + 2 GPUs | N/A | N/A | 0.62 | 1.17 | 2.00 | 3.68 | 6.18 | 7.38 |
| CPU 12 cores + 3 GPUs | 0.10 | 0.19 | 0.65 | 1.25 | 2.11 | 3.93 | 6.42 | 8.00 |

# NAMD 2.9 on Keeneland ID

(12 Intel cores and 3 Telsa M2070 GPUs per node)



# Trends Affecting Performance

- GPU performance increasing
  - Performance limit will be code on CPU
  - Most highly tuned CPU code moved to GPU
  - Remaining CPU code is also less efficient
  - Therefore CPU must run serial code well
- CPU serial performance static
- CPU core counts increasing

# Suggested Strategy

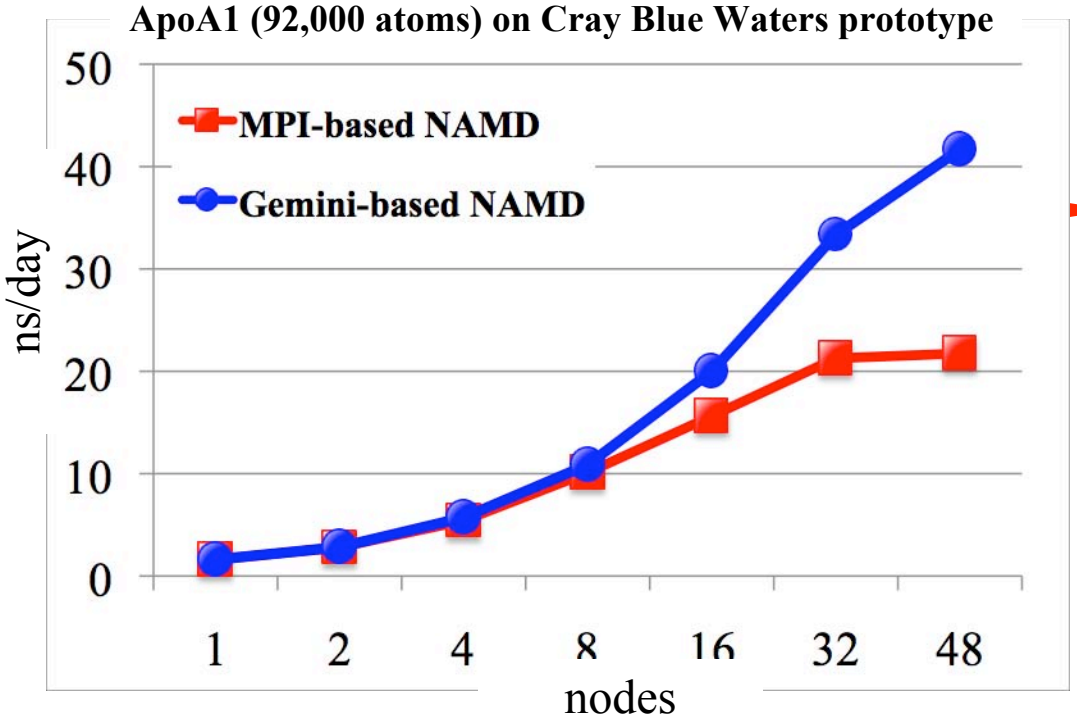
- Focus on CPU-side code
  - Port to GPU or optimize/paralellize on CPU
  - Stream results off GPU to increase overlap
  - Use CPUs with best single-thread performance
- Focus on communication
  - Reduce communication overhead on CPU
  - Deal with multithreaded MPI issues
  - General parallel scalablity improvements

# Streaming GPU Results to CPU

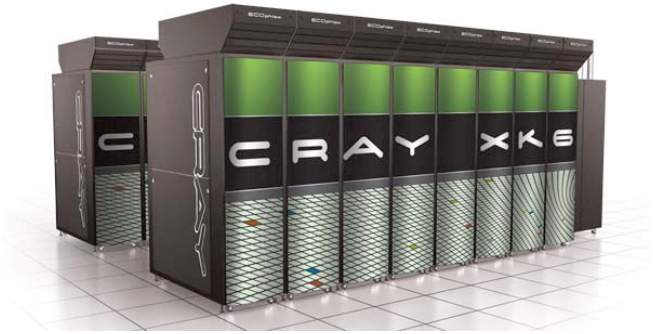
- Allows incremental results from a single grid to be processed on CPU before grid finishes on GPU
- GPU side:
  - Write results to host-mapped memory
  - __threadfence_system() and __syncthreads()
  - Atomic increment for next output queue location
  - Write result index to output queue
- CPU side:
  - Poll end of output queue (int array) in host memory

# Cray Gemini Optimization

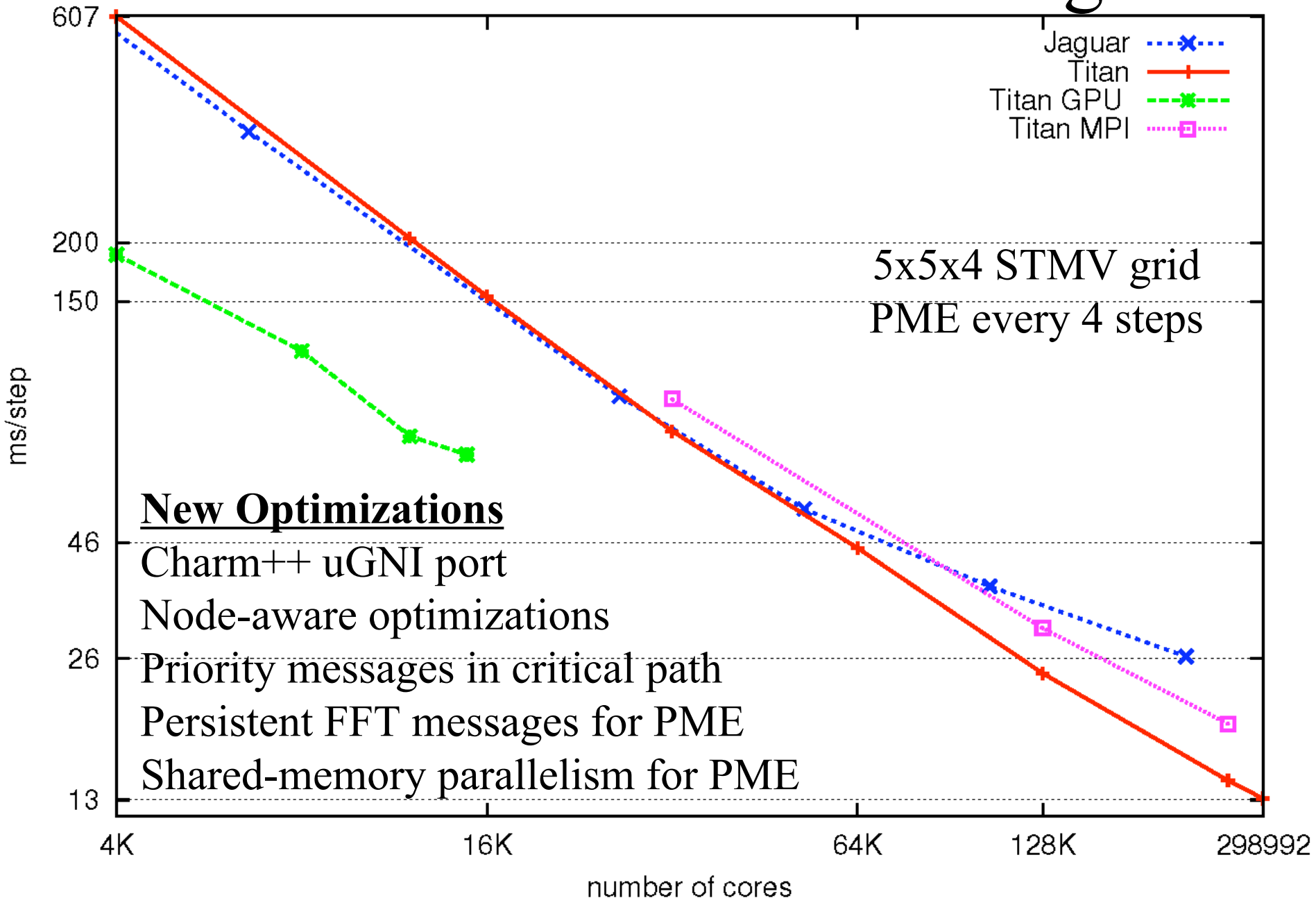
- The new Cray machine has a better network (called Gemini)
- MPI-based NAMD scaled poorly
- BTRC implemented direct port of Charm++ to Cray
  - *uGNI* is the lowest level interface for the Cray Gemini network
  - Removes MPI from NAMD call stack



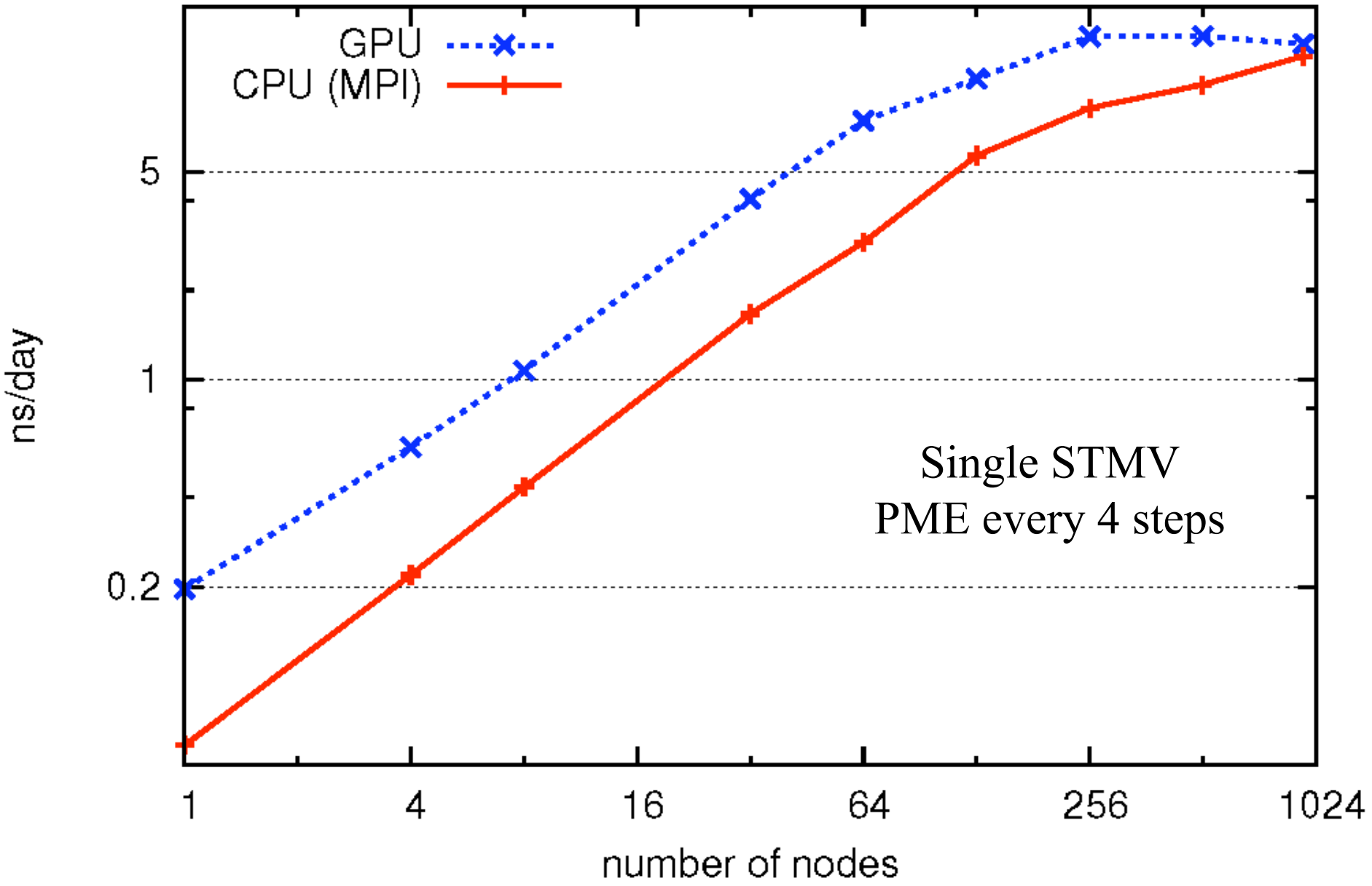

Gemini provides at least 2x increase in usable nodes for strong scaling



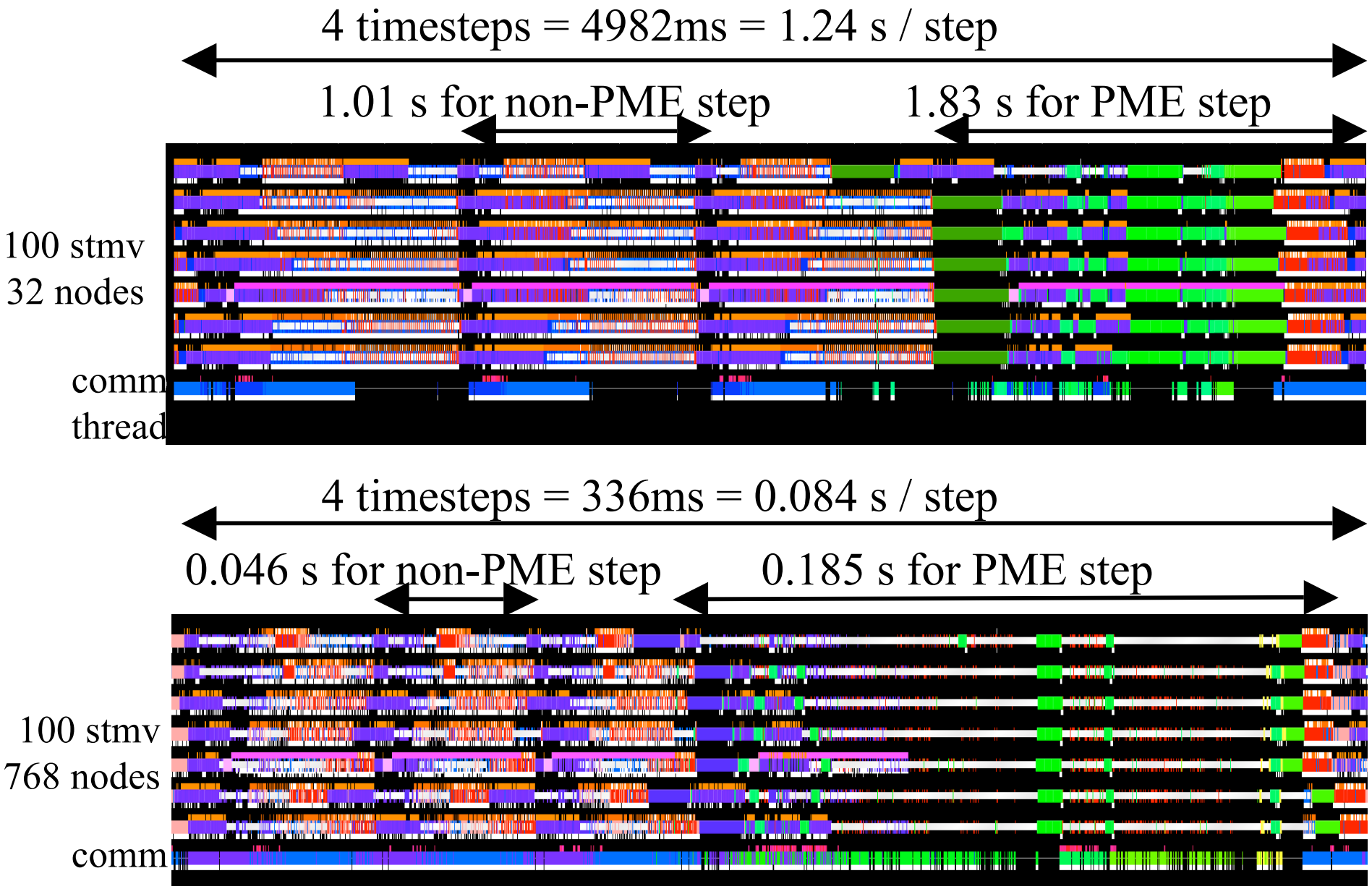
# 100M Atoms on Titan vs Jaguar



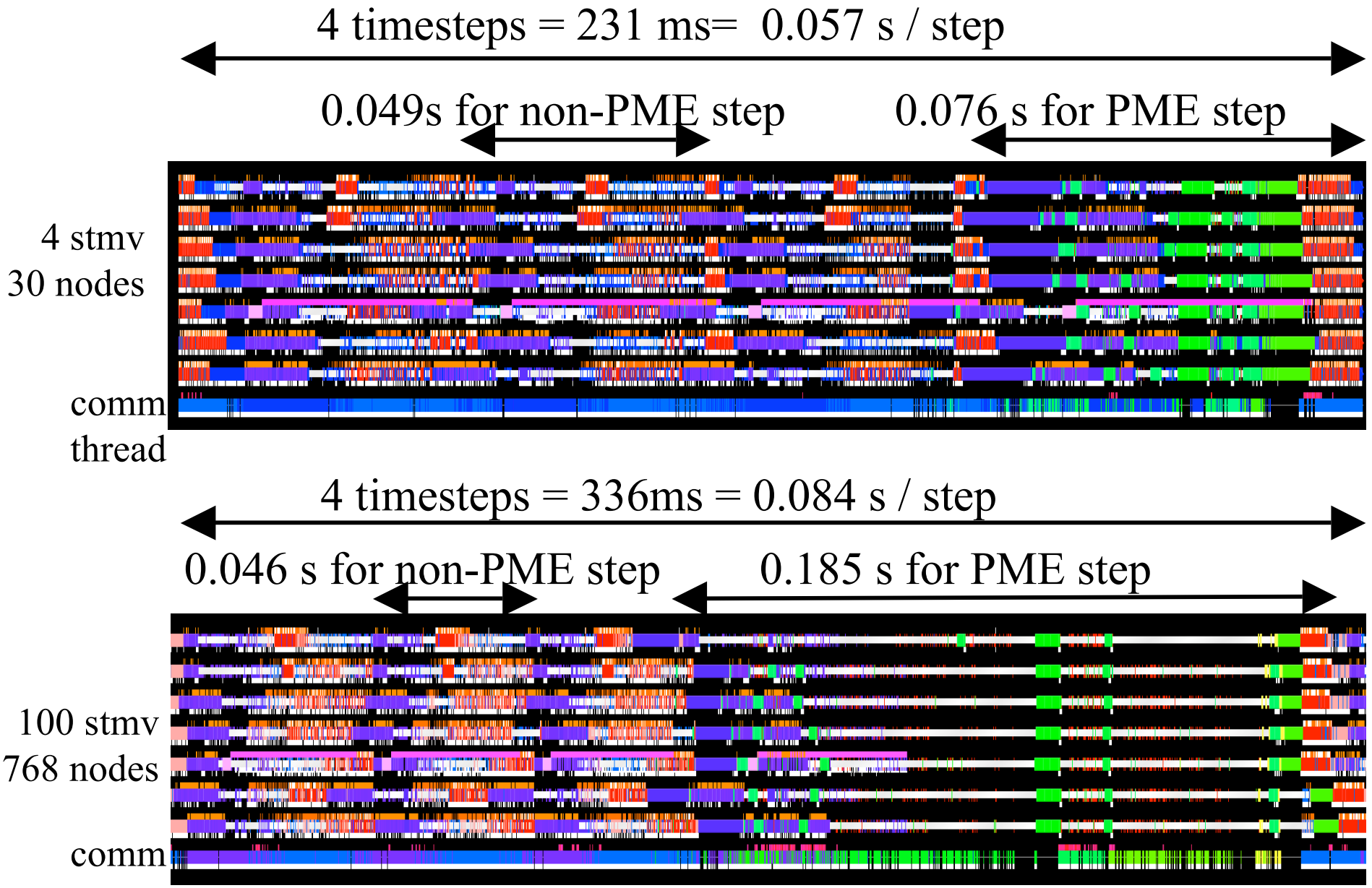
# 1M Atom Virus on TitanDev GPU



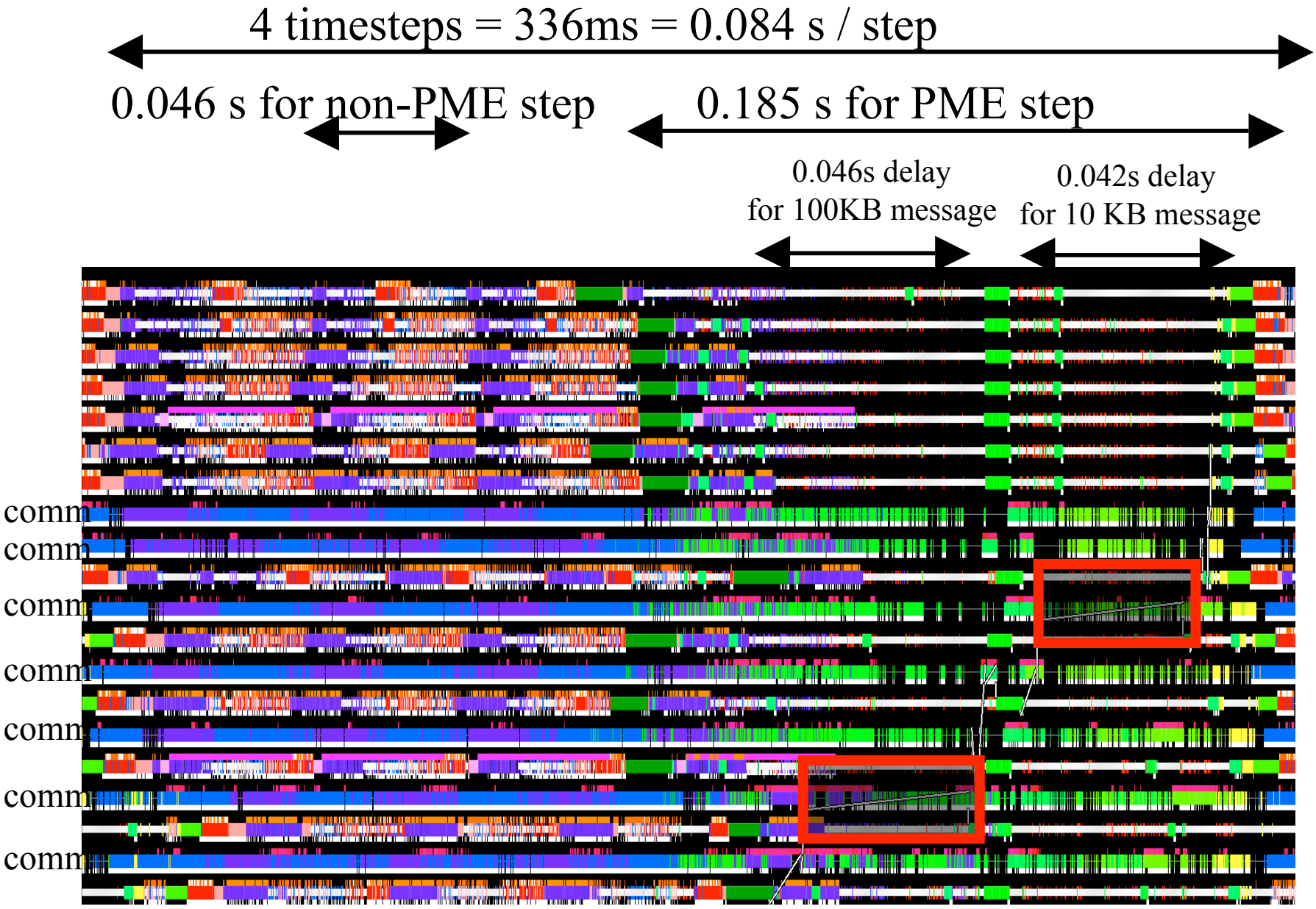
# TitanDev Strong Scaling



# TitanDev Weak Scaling



# PME delays – tracing data needed for one ungrid calculation



# Strategy to improve scalability

- Fix issues with communication
  - 23x16x2 topology limits bisection bandwidth
- Coarsen PME grid
  - Reduces communication
  - Increases short-range work for GPU
- Push PME work to the GPU
  - Charge gridding overlaps coordinate receive
- Start GPU work sooner
  - Currently waiting for all coordinate receives
  - Use streams to launch work as data arrives

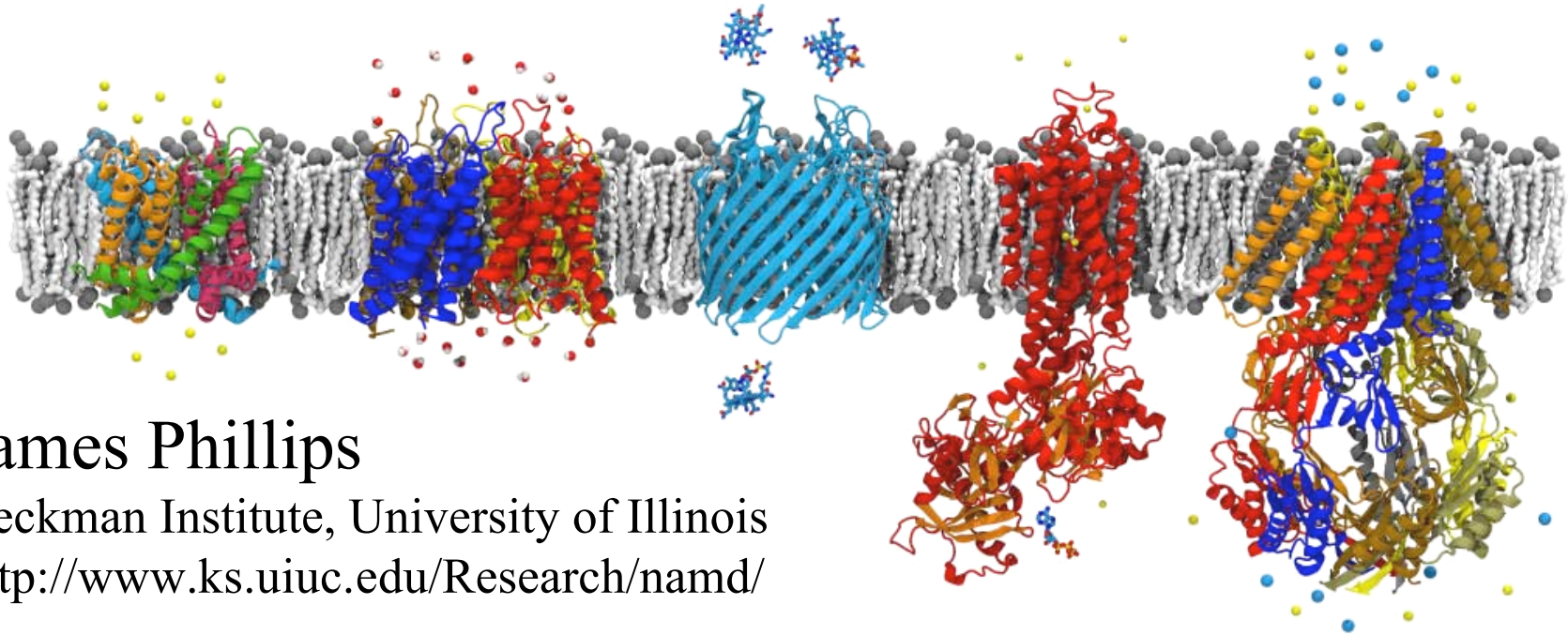
# NAMD 2.9 Scalable Replica Exchange

- Easier to use *and* more efficient:
  - Eliminates complex, machine-specific launch scripts
  - Scalable pair-wise communication between replicas
  - Fast communication via high-speed network
- Basis for many enhanced sampling methods:
  - Parallel tempering (temperature exchange)
  - Umbrella sampling for free-energy calculations

  } Released in NAMD 2.9

  - Hamiltonian exchange (alchemical or conformational)
  - Finite Temperature String method
  - Nudged elastic band
- Great power *and* flexibility:
  - **Enables petascale simulations of modestly sized systems**
  - Leverages features of Collective Variables module
  - Tcl scripts can be highly customized and extended

Thanks to: NIH, NSF, DOE,
NVIDIA (**Sarah Tariq**, Sky Wu,
Justin Luitjens, Nikolai Sakharnykh),
Cray (Sarah Anderson, Ryan Olson),
PPL (Eric Bohm, Yanhua Sun, Gengbin Zheng)
and 17 years of NAMD and Charm++
developers and users.



James Phillips
Beckman Institute, University of Illinois
http://www.ks.uiuc.edu/Research/namd/