In-Situ Visualization and Analysis of Petascale Molecular Dynamics Simulations with VMD

John Stone

Theoretical and Computational Biophysics Group

Beckman Institute for Advanced Science and Technology

University of Illinois at Urbana-Champaign

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

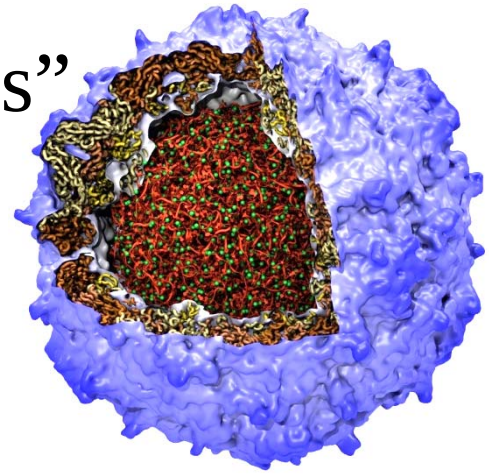
Accelerated HPC Symposium,

San Jose Convention Center,

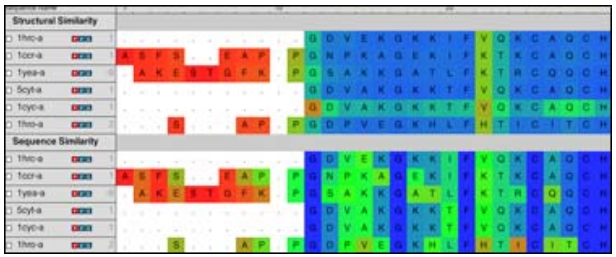
San Jose, CA, May 17, 2012

VMD – “Visual Molecular Dynamics”

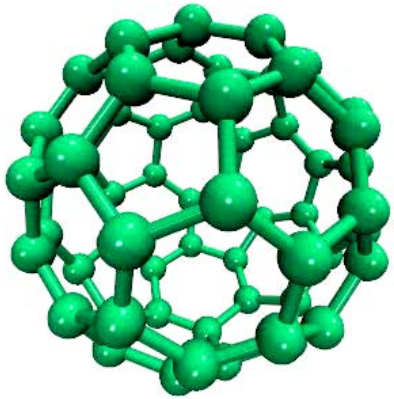
- Visualization and analysis of:
  - molecular dynamics simulations
  - quantum chemistry calculations
  - particle systems and whole cells
  - sequence data
- User extensible w/ scripting and plugins
- http://www.ks.uiuc.edu/Research/vmd/



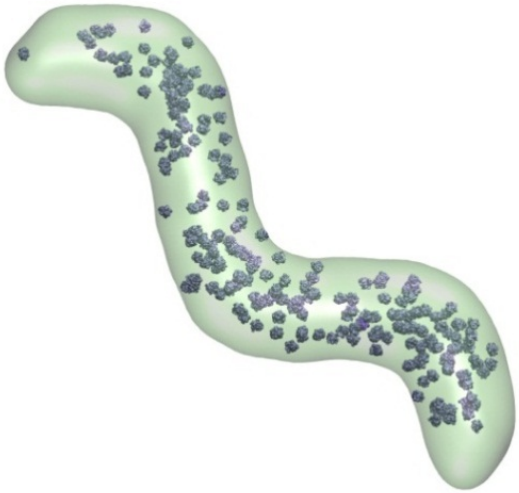
Poliovirus



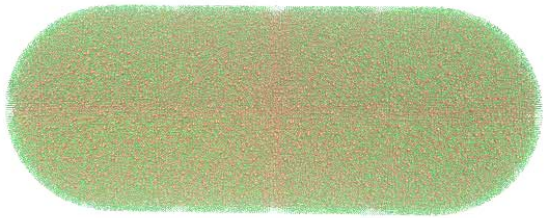
Ribosome Sequences



Electrons in
Vibrating Buckyball



Cellular Tomography,
Cryo-electron Microscopy



Whole Cell Simulations

Molecular Visualization and Analysis Challenges for Petascale Simulations

- Very large structures (10M to over 100M atoms)
  - 12-bytes per atom per trajectory frame
  - 100M atom trajectory frame: 1200MB!
- Long-timescale simulations produce huge trajectories
  - MD integration timesteps are on the femtosecond timescale (10^{-15} sec) but many important biological processes occur on microsecond to millisecond timescales
  - Even storing trajectory frames infrequently, resulting trajectories frequently contain millions of frames
- Petabytes of data to analyze, far too large to move
- Viz and analysis must be done primarily on the supercomputer where the data already resides

Approaches for Visualization and Analysis of Petascale Molecular Simulations with VMD

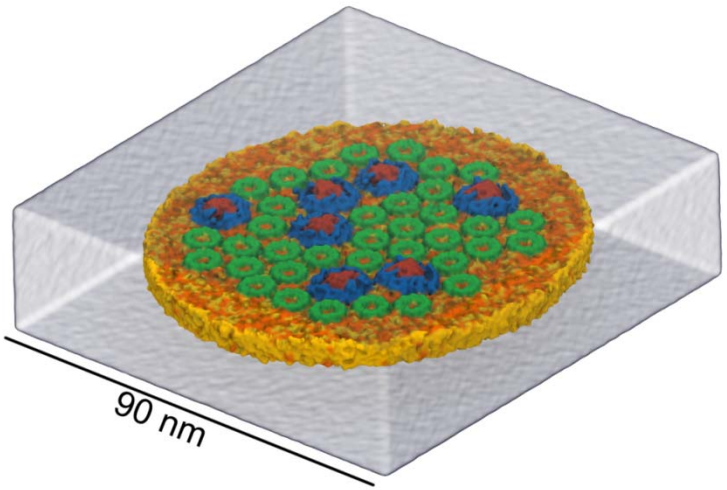
- Abandon conventional approaches, e.g. bulk download of trajectory data to remote viz/analysis machines
  - In-place processing of trajectories on the machine running the simulations
  - Use remote visualization techniques: Split-mode VMD with remote front-end instance, and back-end viz/analysis engine running in parallel on supercomputer
- Large-scale parallel analysis and visualization via distributed memory MPI version of VMD
- Exploit GPUs and other accelerators to increase per-node analytical capabilities, e.g. NCSA Blue Waters Cray XK6
- In-situ on-the-fly viz/analysis and event detection through direct communication with running MD simulation

Improved Support for Large Datasets in VMD

- New structure building tools, file formats, and data structures enable VMD to operate efficiently up to 150M atoms
  - Up to 30% more memory efficient
  - Analysis routines optimized for large structures, up to 20x faster for calculations on 100M atom complexes where molecular structure traversal can represent a significant amount of runtime
  - New and revised graphical representations support smooth trajectory animation for multi-million atom complexes; VMD remains interactive even when displaying surface reps for 20M atom membrane patch
- Uses multi-core CPUs and GPUs for the most demanding computations

20M atoms:

membrane patch and solvent



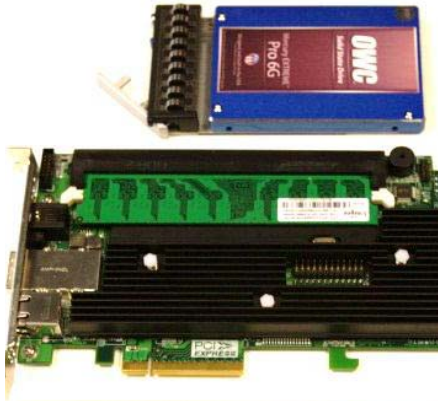
New Interactive Display & Analysis of Terabytes of Data:

Out-of-Core Trajectory I/O w/ Solid State Disks

450MB/sec

to 4GB/sec

A DVD movie per second!

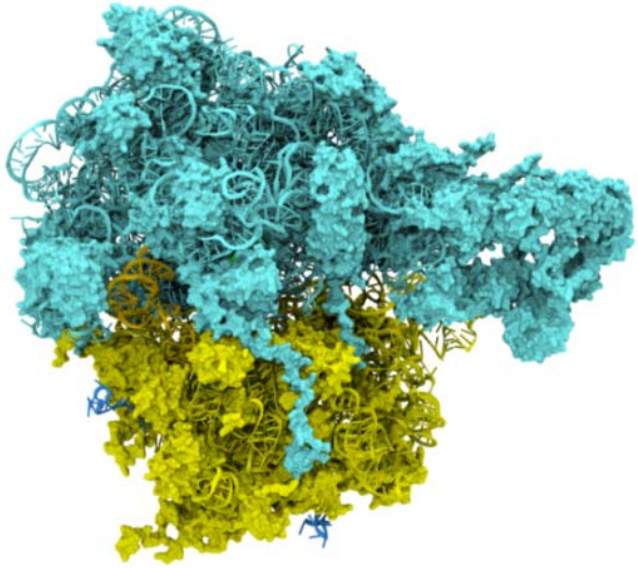


Commodity SSD, SSD RAID

- Timesteps loaded on-the-fly (out-of-core)
  - Eliminates memory capacity limitations, even for multi-terabyte trajectory files
  - High performance achieved by new trajectory file formats, optimized data structures, and efficient I/O
- Analyze long trajectories significantly faster
- New SSD Trajectory File Format 2x Faster vs. Existing Formats

Immersive out-of-core visualization of large-size and long-timescale molecular dynamics trajectories. J. Stone, K. Vandivort, and K. Schulten. *Lecture Notes in Computer Science*, 6939:1-12, 2011.

VMD Out-of-Core Trajectory I/O Performance: SSD-Optimized Trajectory Format, 8-SSD RAID

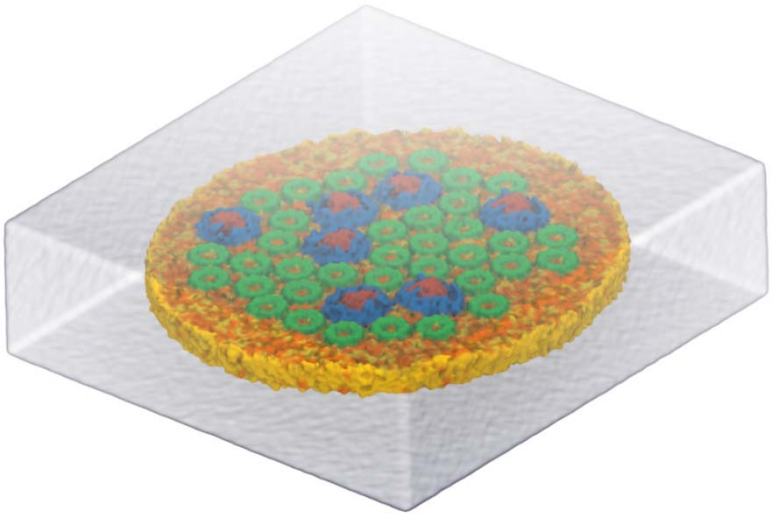


Ribosome w/ solvent

3M atoms

3 frames/sec w/ HD

60 frames/sec w/ SSDs



Membrane patch w/ solvent

20M atoms

0.4 frames/sec w/ HD

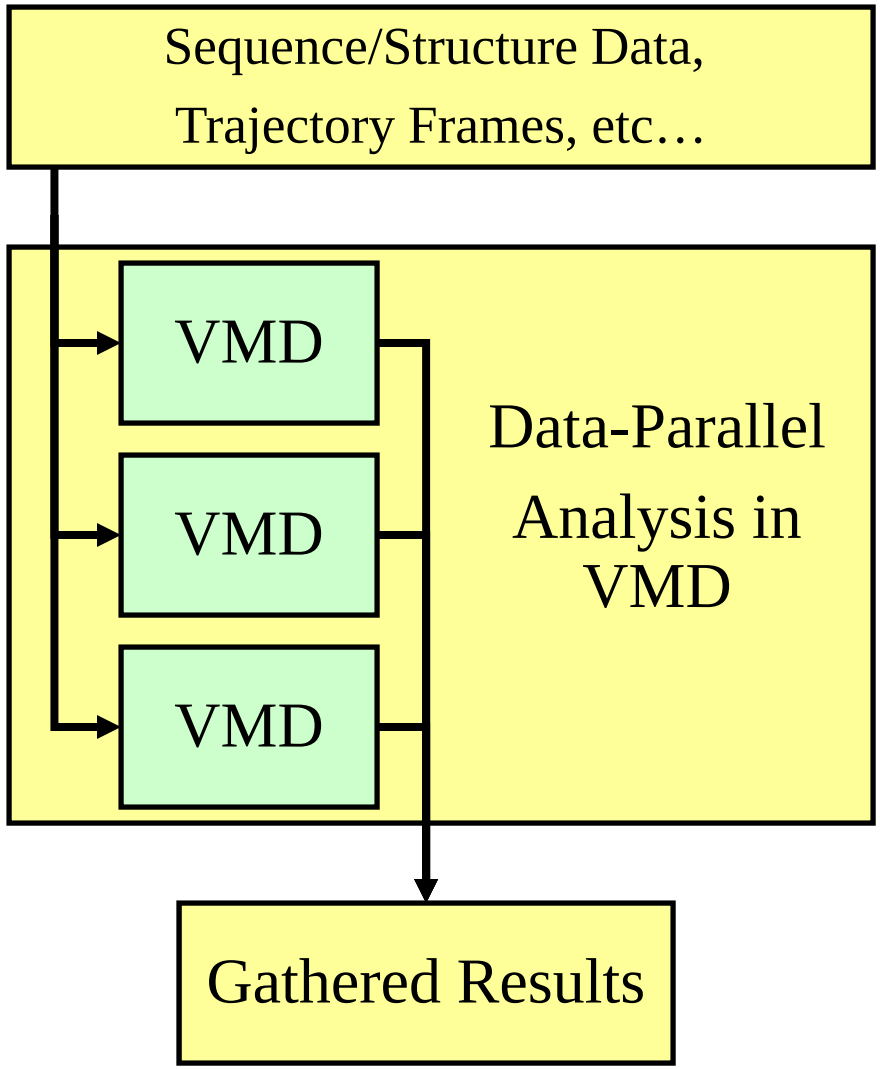
8 frames/sec w/ SSDs

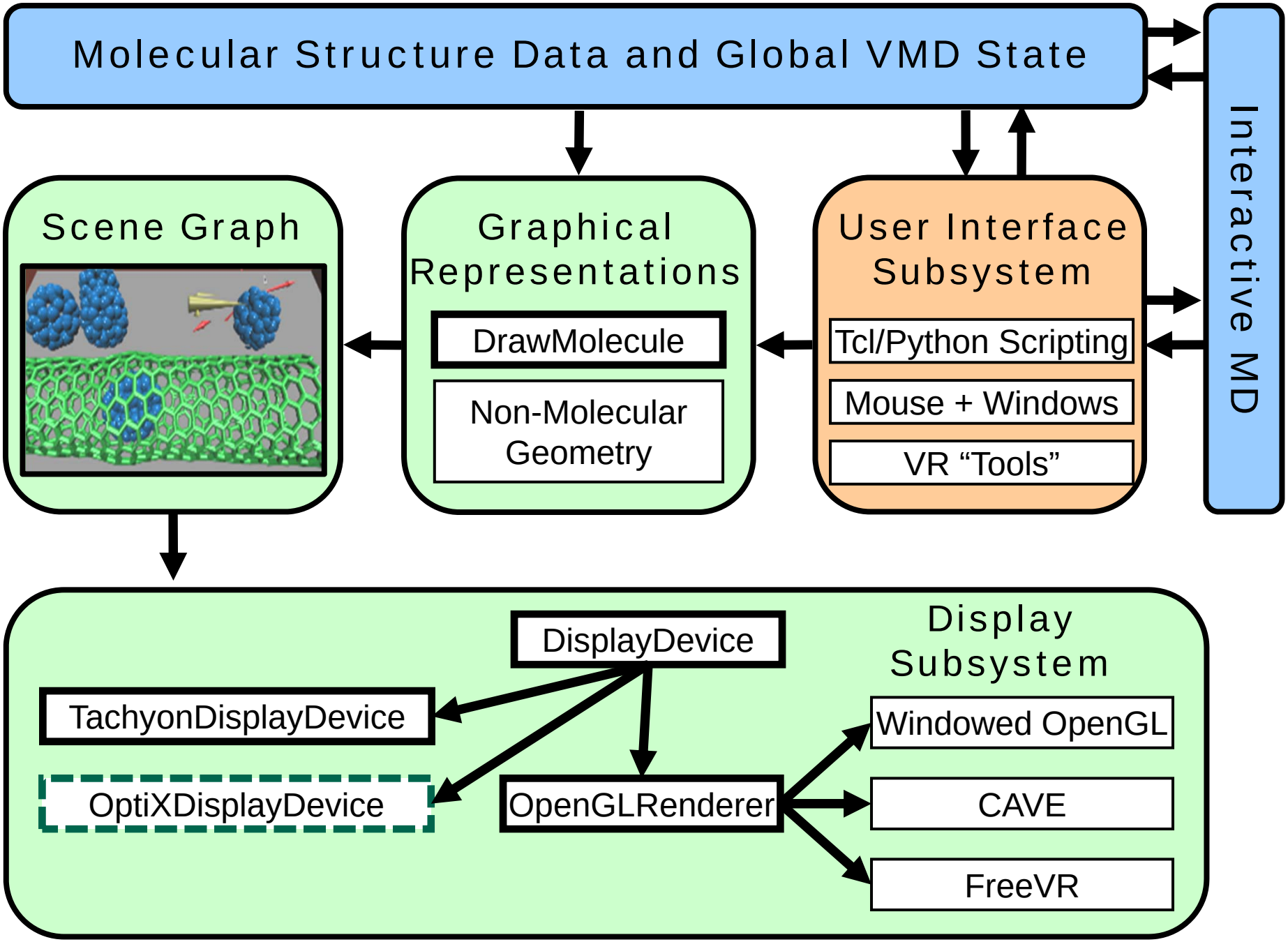
New SSD Trajectory File Format 2x Faster vs. Existing Formats

VMD I/O rate ~2.1 GB/sec w/ 8 SSDs

Parallel VMD Analysis w/ MPI

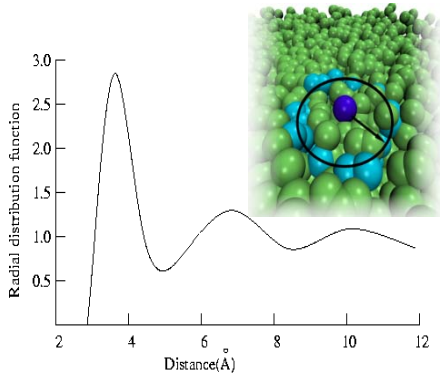
- Analyze trajectory frames, structures, or sequences in parallel supercomputers:
  - Parallelize user-written analysis scripts with minimum difficulty
  - Parallel analysis of independent trajectory frames
  - Parallel structural analysis using custom parallel reductions
  - Parallel rendering, movie making
- Dynamic load balancing:
  - Recently tested with up to 15,360 CPU cores
- **Supports GPU-accelerated clusters and supercomputers**

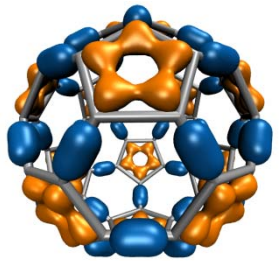



Molecular Structure Data and Global VMD State
Interactive MD
Scene Graph
Graphical Representations
DrawMolecule
Non-Molecular Geometry
User Interface Subsystem
Tcl/Python Scripting
Mouse + Windows
VR “Tools”
DisplayDevice
Display Subsystem
TachyonDisplayDevice
OptiXDisplayDevice
OpenGLRenderer
Windowed OpenGL
CAVE
FreeVR

GPU Accelerated Trajectory Analysis and Visualization in VMD

GPU-Accelerated Feature	Speedup vs. single CPU core
Molecular orbital display	120x
Radial distribution function	92x
Electrostatic field calculation	44x
Molecular surface display	40x
Ion placement	26x
MDFF density map synthesis	26x
Implicit ligand sampling	25x
Root mean squared fluctuation	25x
Radius of gyration	21x
Close contact determination	20x
Dipole moment calculation	15x





NCSA Blue Waters Early Science System
Cray XK6 nodes w/ NVIDIA Tesla X2090

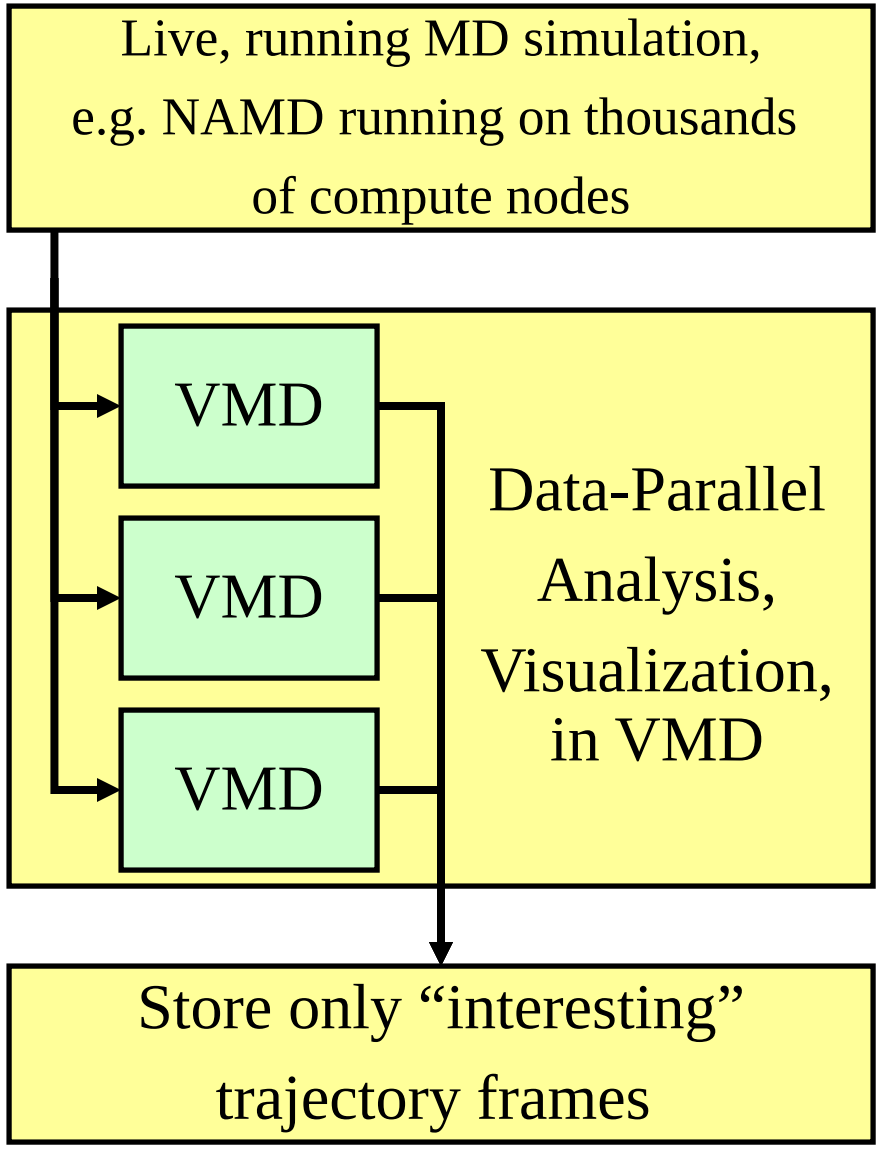
Time-Averaged Electrostatics Analysis on NCSA Blue Waters Early Science System

NCSA Blue Waters Node Type	**Seconds per trajectory frame for one compute node**
Cray XE6 Compute Node: 32 CPU cores (2xAMD 6200 CPUs)	9.33
Cray XK6 GPU-accelerated Compute Node: 16 CPU cores + **NVIDIA X2090 (Fermi) GPU**	2.25
Speedup for GPU XK6 nodes vs. CPU XE6 nodes	**GPU nodes are 4.15x faster overall**

Preliminary performance for VMD time-averaged electrostatics w/ Multilevel Summation Method on the NCSA Blue Waters Early Science System

In-Situ Visualization and Analysis with VMD

- Early prototype and testing phase
- VMD supports live socket connection to running MD code
- Custom user-written analysis scripts are triggered by callbacks as incoming frames arrive
- Separate threads handle async. network I/O between MD code and “master” VMD instance, MPI broadcast or decomposition among peer VMD nodes
- Perform real-time analysis processing of incoming frames to find rare events
  - Store the most “interesting” timesteps
  - Build summary analyses useful for accelerating interactive “Timeline” displays, and subsequent detailed batch mode analysis runs



Acknowledgements

- Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign
- NCSA Blue Waters Team
- NCSA Innovative Systems Lab
- NVIDIA CUDA Center of Excellence, University of Illinois at Urbana-Champaign
- The CUDA team at NVIDIA
- NIH support: P41-RR005969

GPU Computing Publications

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

- **Fast Visualization of Gaussian Density Surfaces for Molecular Dynamics and Particle System Trajectories.** M. Krone, J. Stone, T. Ertl, and K. Schulten. *In proceedings EuroVis 2012*, 2012. (In-press)
- **Immersive Out-of-Core Visualization of Large-Size and Long-Timescale Molecular Dynamics Trajectories.** J. Stone, K. Vandivort, and K. Schulten. G. Bebis et al. (Eds.): *7th International Symposium on Visual Computing (ISVC 2011)*, LNCS 6939, pp. 1-12, 2011.
- **Fast Analysis of Molecular Dynamics Trajectories with Graphics Processing Units – Radial Distribution Functions.** B. Levine, J. Stone, and A. Kohlmeyer. *J. Comp. Physics*, 230(9):3556-3569, 2011.

GPU Computing Publications
http://www.ks.uiuc.edu/Research/gpu/

- **Quantifying the Impact of GPUs on Performance and Energy Efficiency in HPC Clusters.** J. Enos, C. Steffen, J. Fullop, M. Showerman, G. Shi, K. Esler, V. Kindratenko, J. Stone, J Phillips. *International Conference on Green Computing*, pp. 317-324, 2010.
- **GPU-accelerated molecular modeling coming of age. J. Stone, D. Hardy, I. Ufimtsev, K. Schulten.** *J. Molecular Graphics and Modeling*, 29:116-125, 2010.
- **OpenCL: A Parallel Programming Standard for Heterogeneous Computing. J. Stone, D. Gohara, G. Shi.** *Computing in Science and Engineering*, 12(3):66-73, 2010.
- **An Asymmetric Distributed Shared Memory Model for Heterogeneous Computing Systems**. I. Gelado, J. Stone, J. Cabezas, S. Patel, N. Navarro, W. Hwu. *ASPLOS '10: Proceedings of the 15th International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 347-358, 2010.

GPU Computing Publications
http://www.ks.uiuc.edu/Research/gpu/

- **GPU Clusters for High Performance Computing**. V. Kindratenko, J. Enos, G. Shi, M. Showerman, G. Arnold, J. Stone, J. Phillips, W. Hwu. *Workshop on Parallel Programming on Accelerator Clusters (PPAC)*, In Proceedings IEEE Cluster 2009, pp. 1-8, Aug. 2009.
- **Long time-scale simulations of in vivo diffusion using GPU hardware**. E. Roberts, J. Stone, L. Sepulveda, W. Hwu, Z. Luthey-Schulten. In *IPDPS'09: Proceedings of the 2009 IEEE International Symposium on Parallel & Distributed Computing*, pp. 1-8, 2009.
- **High Performance Computation and Interactive Display of Molecular Orbitals on GPUs and Multi-core CPUs**. J. Stone, J. Saam, D. Hardy, K. Vandivort, W. Hwu, K. Schulten, *2nd Workshop on General-Purpose Computation on Graphics Pricessing Units (GPGPU-2), ACM International Conference Proceeding Series*, volume 383, pp. 9-18, 2009.
- **Probing Biomolecular Machines with Graphics Processors**. J. Phillips, J. Stone. *Communications of the ACM*, 52(10):34-41, 2009.
- **Multilevel summation of electrostatic potentials using graphics processing units**. D. Hardy, J. Stone, K. Schulten. *J. Parallel Computing*, 35:164-177, 2009.

GPU Computing Publications
http://www.ks.uiuc.edu/Research/gpu/

- **Adapting a message-driven parallel application to GPU-accelerated clusters**. J. Phillips, J. Stone, K. Schulten. *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, IEEE Press, 2008.
- **GPU acceleration of cutoff pair potentials for molecular modeling applications**. C. Rodrigues, D. Hardy, J. Stone, K. Schulten, and W. Hwu. *Proceedings of the 2008 Conference On Computing Frontiers*, pp. 273-282, 2008.
- **GPU computing**. J. Owens, M. Houston, D. Luebke, S. Green, J. Stone, J. Phillips. *Proceedings of the IEEE*, 96:879-899, 2008.
- **Accelerating molecular modeling applications with graphics processors**. J. Stone, J. Phillips, P. Freddolino, D. Hardy, L. Trabuco, K. Schulten. *J. Comp. Chem.*, 28:2618-2640, 2007.
- **Continuous fluorescence microphotolysis and correlation spectroscopy**. A. Arkhipov, J. Hüve, M. Kahms, R. Peters, K. Schulten. *Biophysical Journal*, 93:4006-4017, 2007.