Petascale Molecular Ray Tracing: Accelerating VMD/Tachyon with OptiX

John E. Stone

Theoretical and Computational Biophysics Group

Beckman Institute for Advanced Science and Technology

University of Illinois at Urbana-Champaign

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

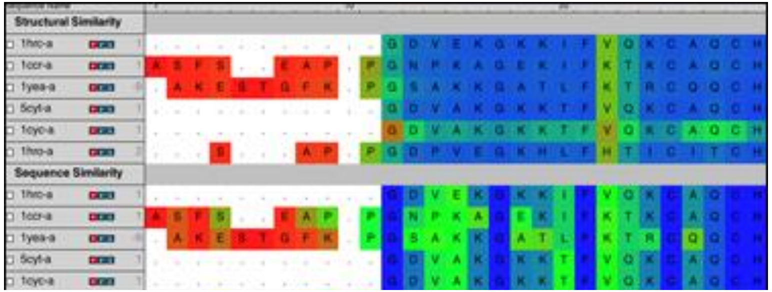
S4400, GPU Technology Conference

10:00-10:25, Room LL21C, San Jose Convention Center,

San Jose, CA, Thursday March 27, 2014

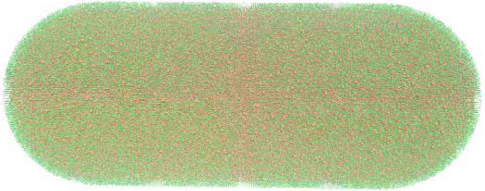
VMD – “Visual Molecular Dynamics”

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

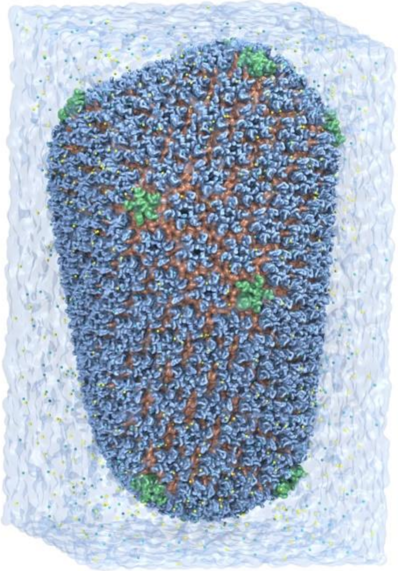


```
Structural Similarity
1hrc-a                 G D V E K G K K I F V Q K C A Q C H
1ccr-a  A S F S     E A P   P G N P K A G E K I F K T K C A Q C H
1yea-a    A K E S T G F K   P G S A K K G A T L F K T R C Q Q C H
5cyt-a                 G D V A K G K K T F V Q K C A Q C H
1cyc-a                 G D V A K G K K T F V Q K C A Q C H
1hro-a        S       A P   P G D P V E G K H L F H T I C I T C H
Sequence Similarity
1hrc-a                 G D V E K G K K I F V Q K C A Q C H
1ccr-a  A S F S     E A P   P G N P K A G E K I F K T K C A Q C H
1yea-a    A K E S T G F K   P G S A K K G A T L F K T R C Q Q C H
5cyt-a                 G D V A K G K K T F V Q K C A Q C H
1cyc-a                 G D V A K G K K T F V Q K C A Q C H
```

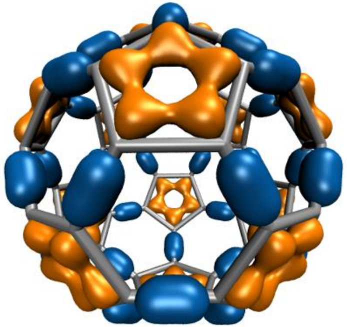
Sequence Data



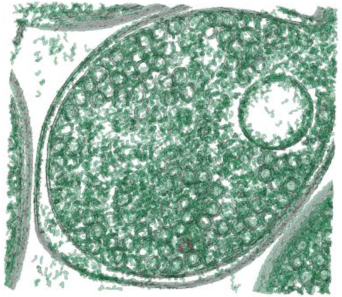
Whole Cell Simulation



MD Simulations



Quantum Chemistry

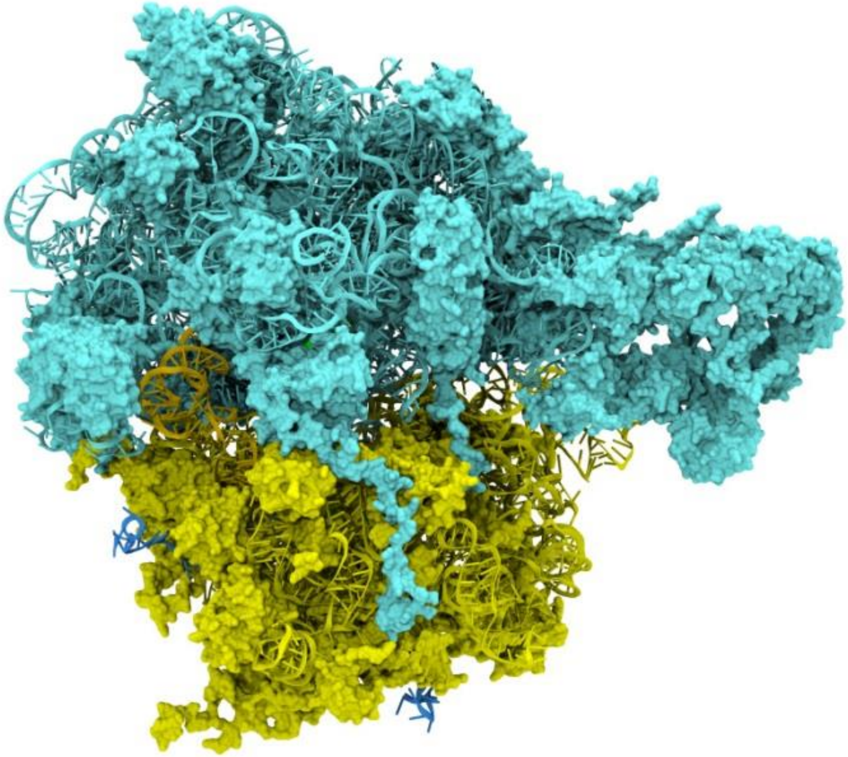


CryoEM, Cellular Tomography

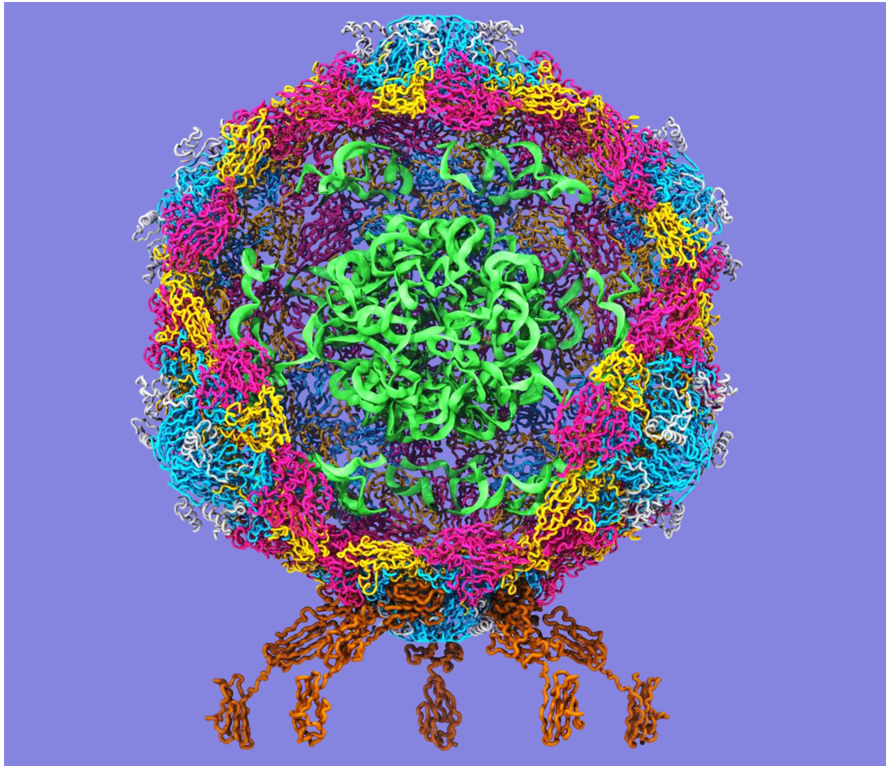
Goal: A Computational Microscope

Study the molecular machines in living cells

Ribosome: target for antibiotics



Poliovirus

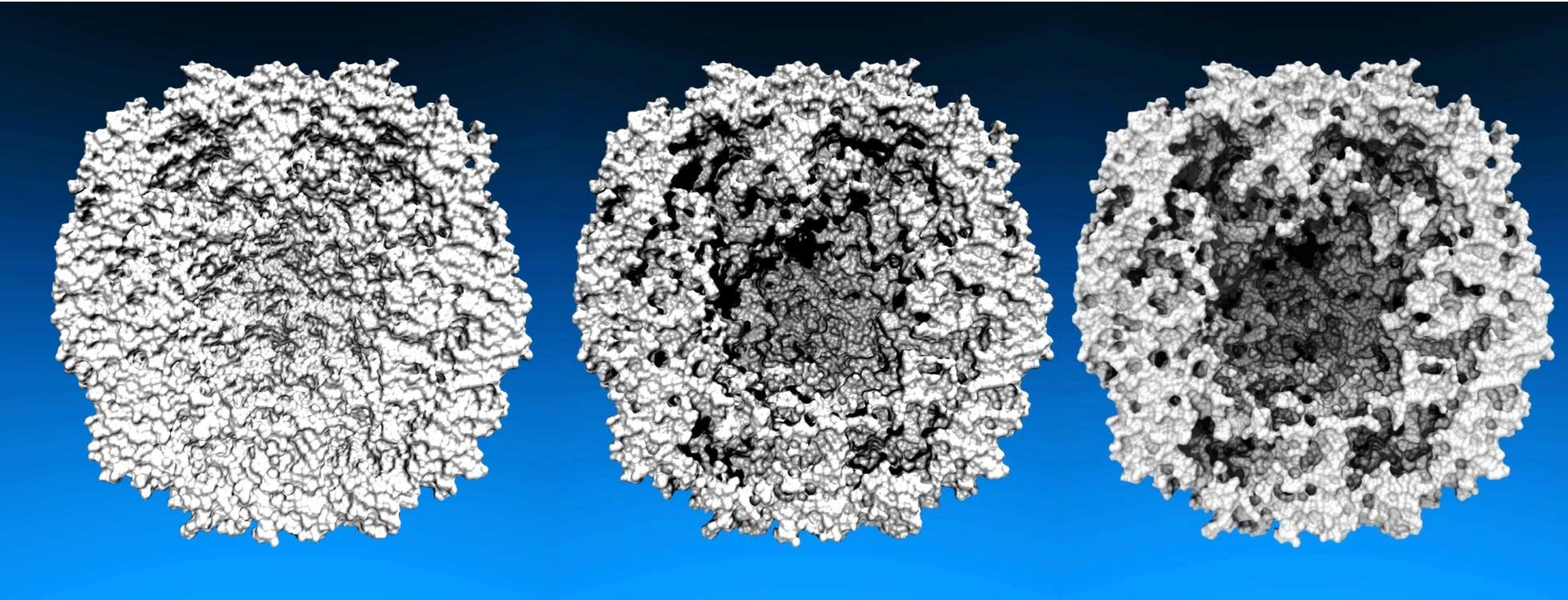


Lighting Comparison

Two lights, no shadows

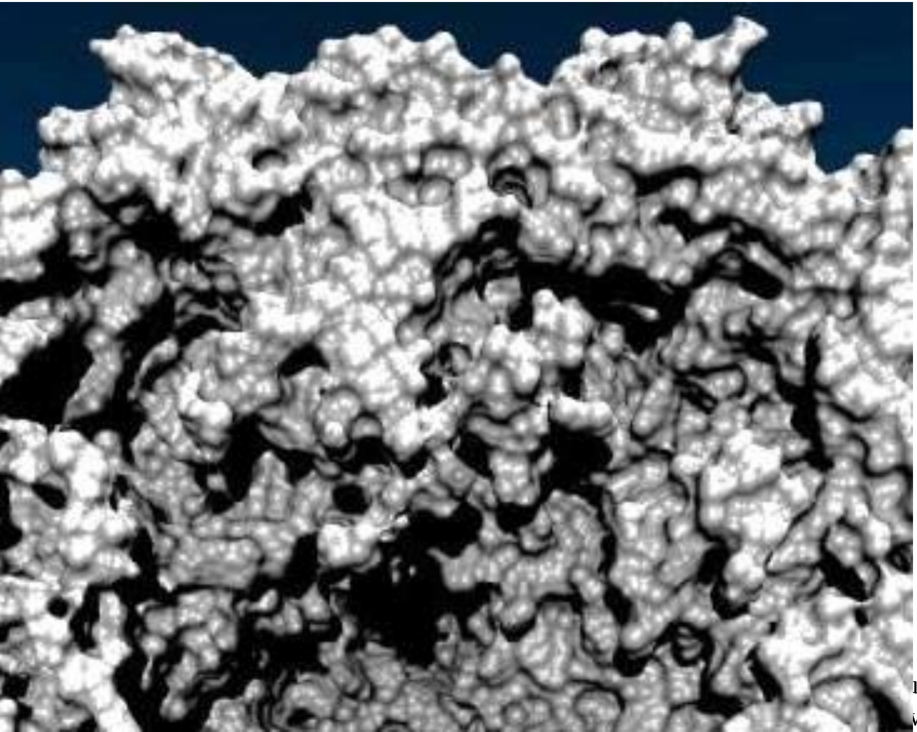
Two lights, hard shadows, 1 shadow ray per light

Ambient occlusion + two lights, 144 AO rays/hit

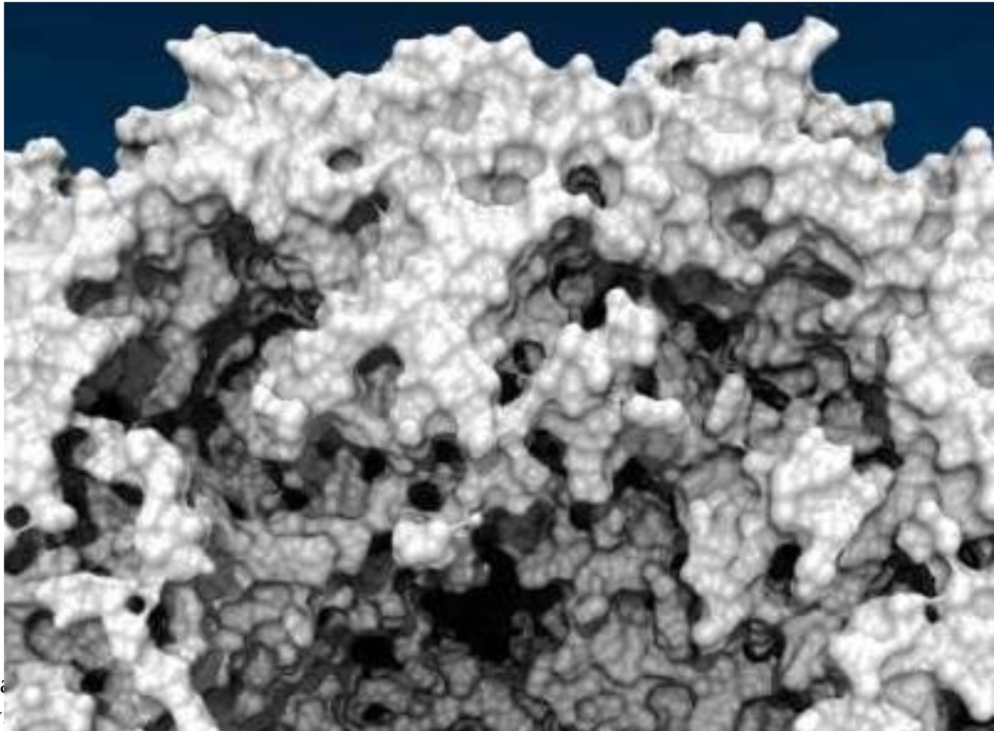


“My Lights are Always in the Wrong Place…”

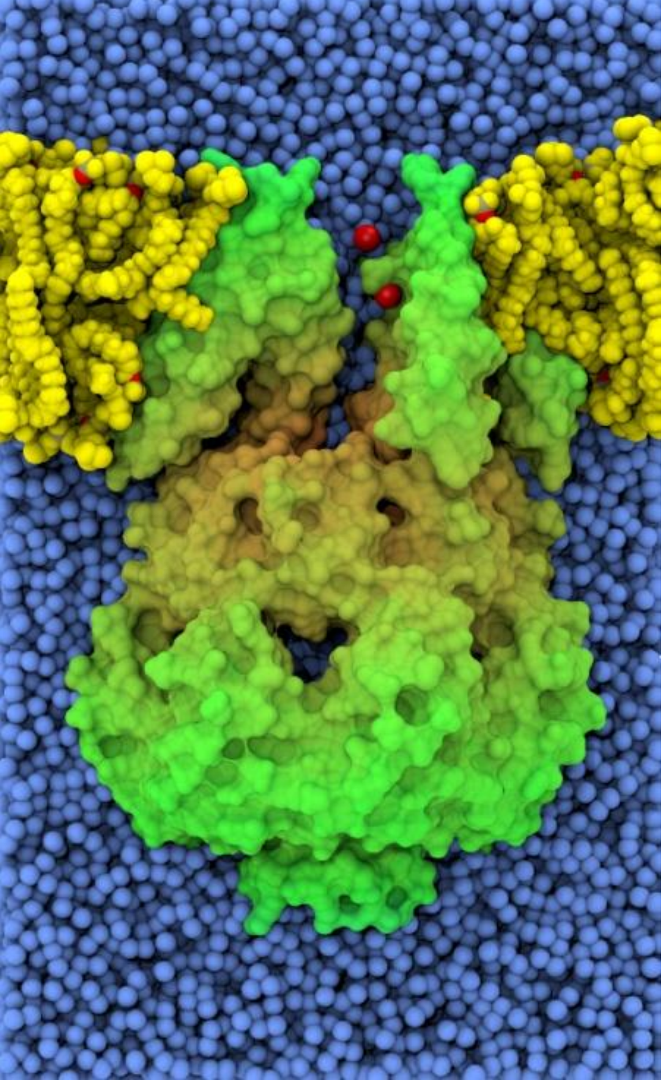
**Two lights,
harsh shadows,
1 shadow ray per light per hit**

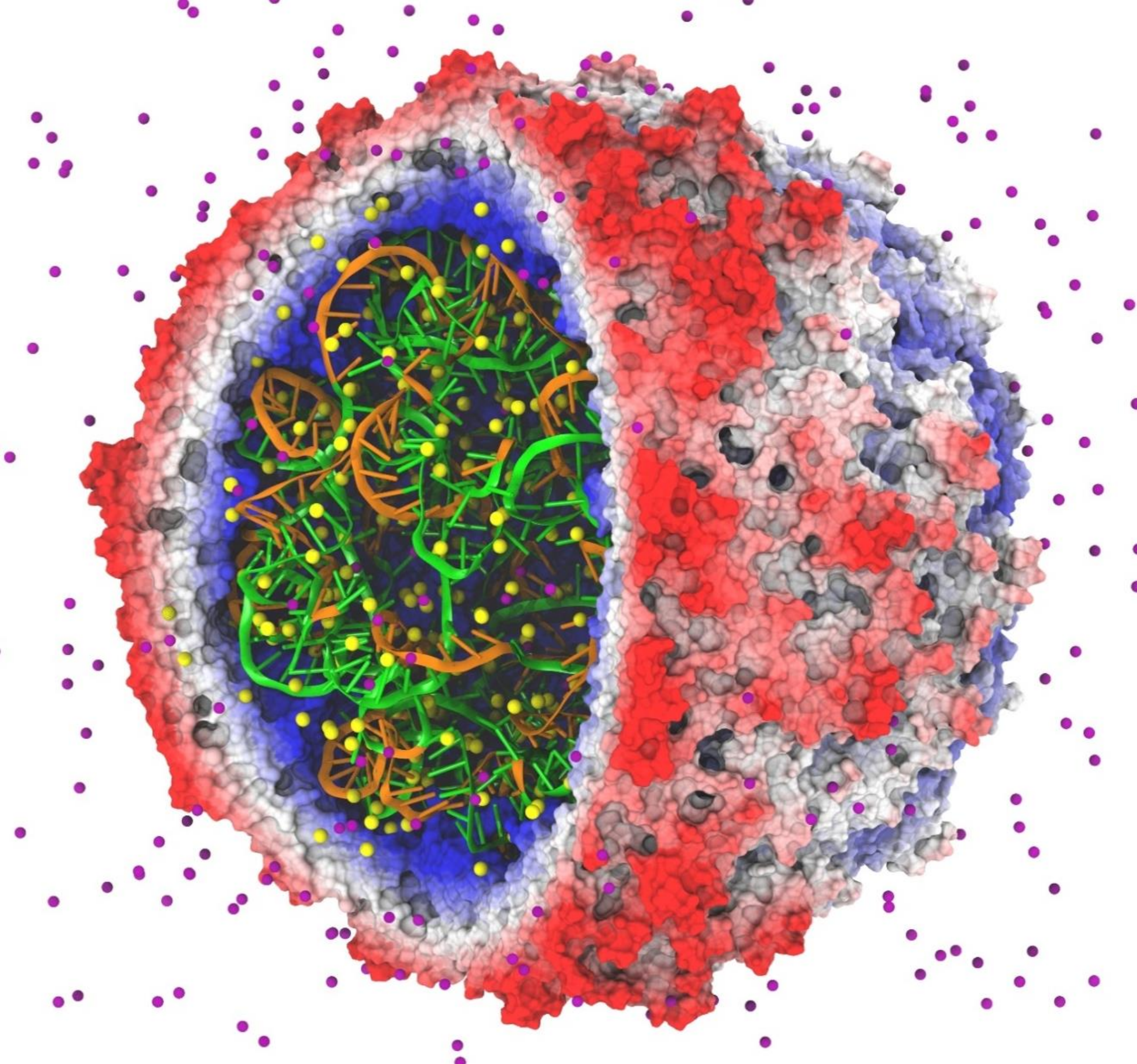


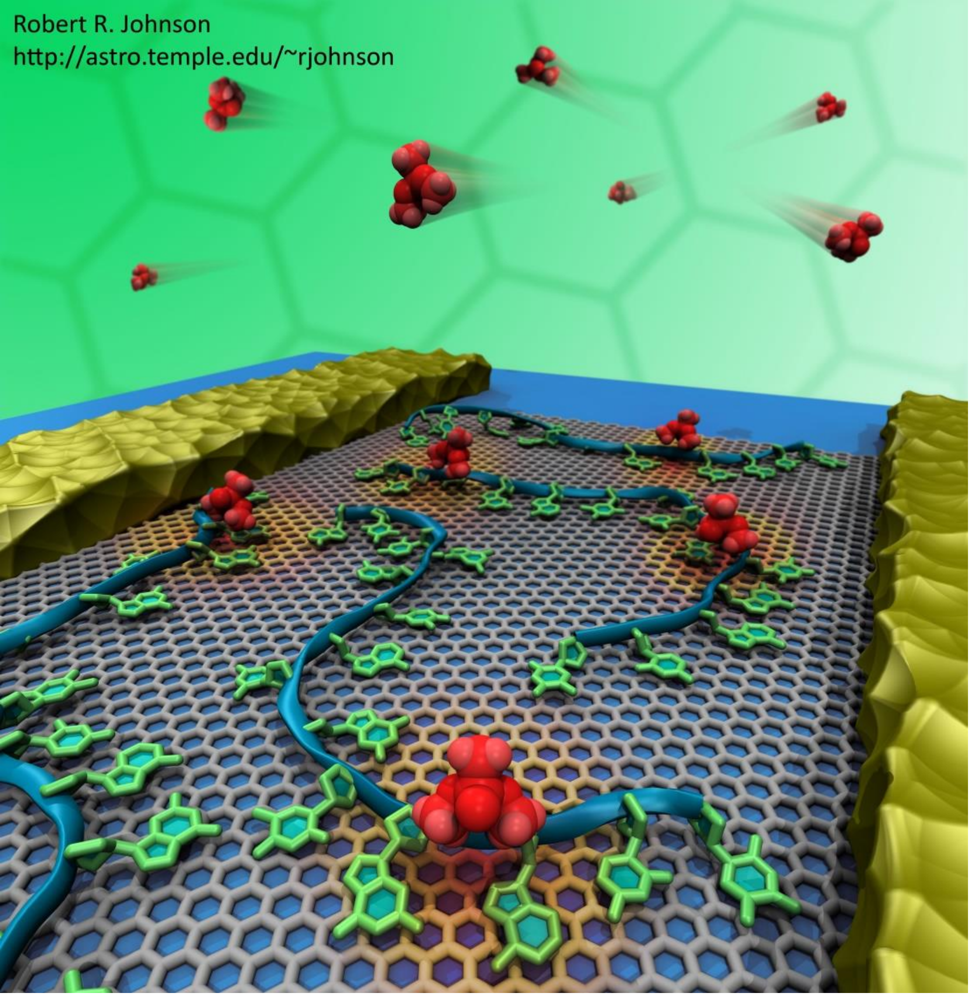
**Ambient occlusion (~80%)
+ two lights (~20%),
144 AO rays/hit**

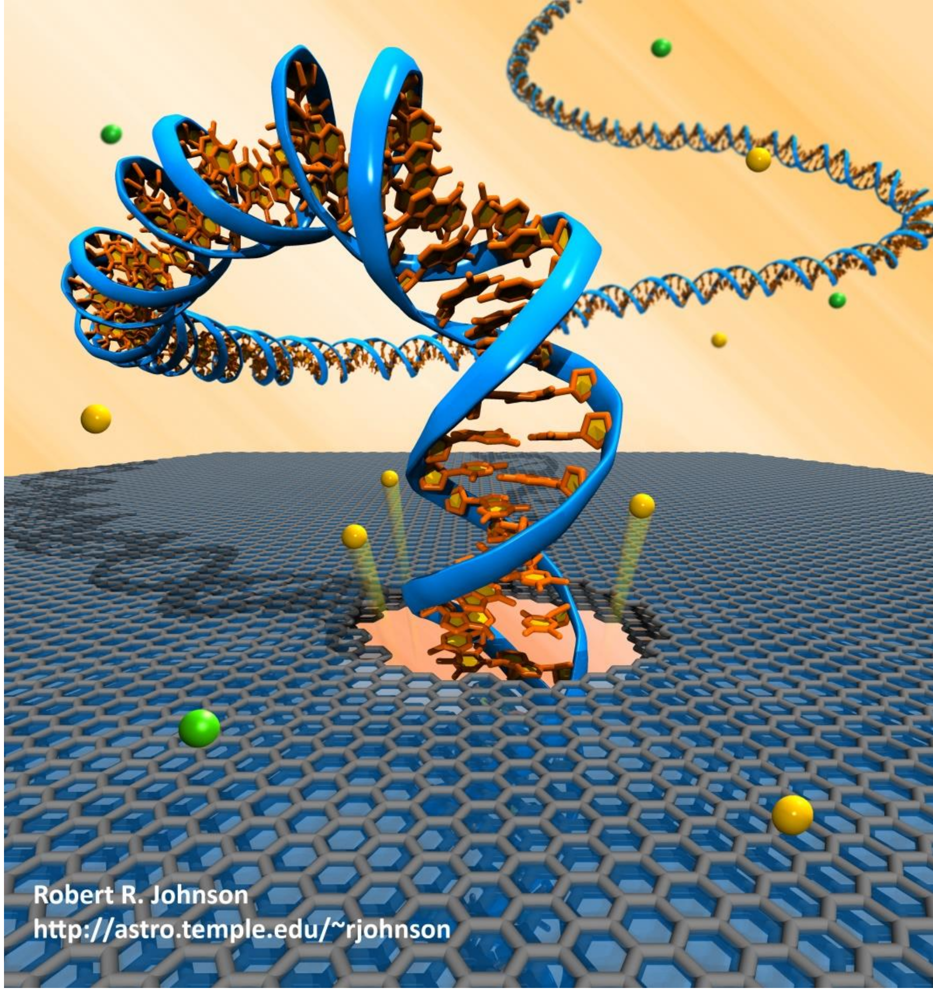


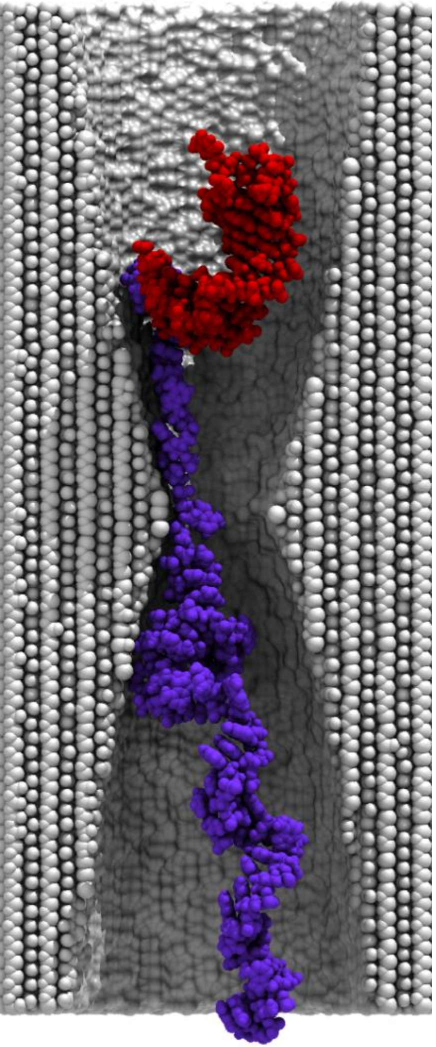
ila
w.

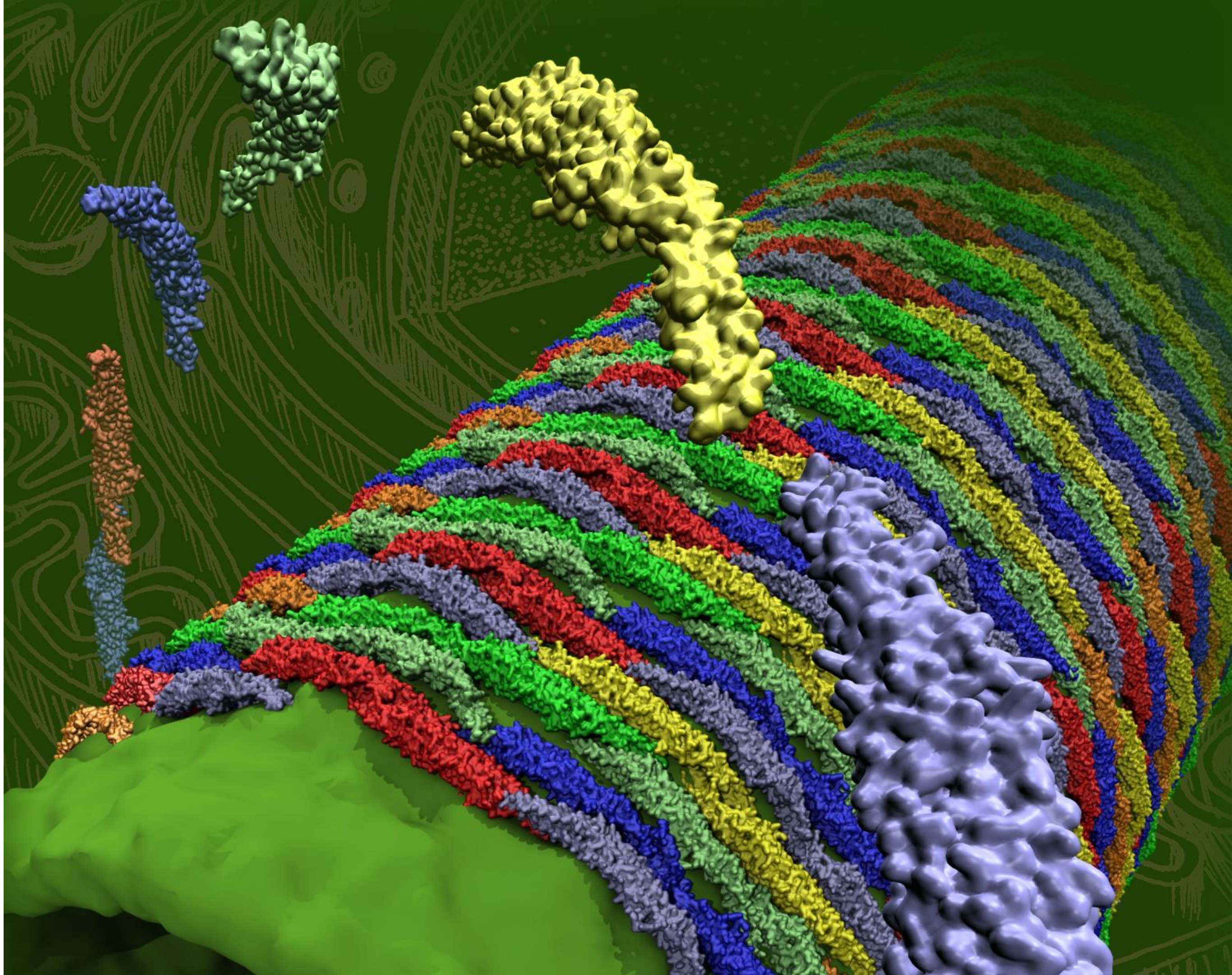


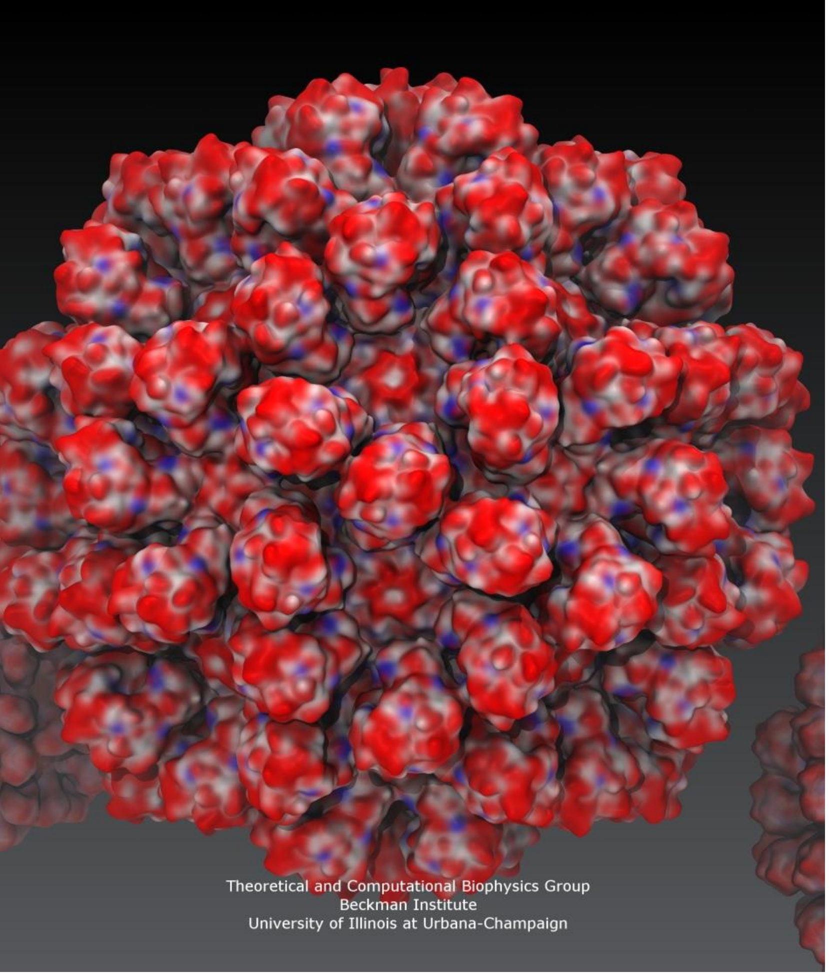



Robert R. Johnson
http://astro.temple.edu/~rjohnson

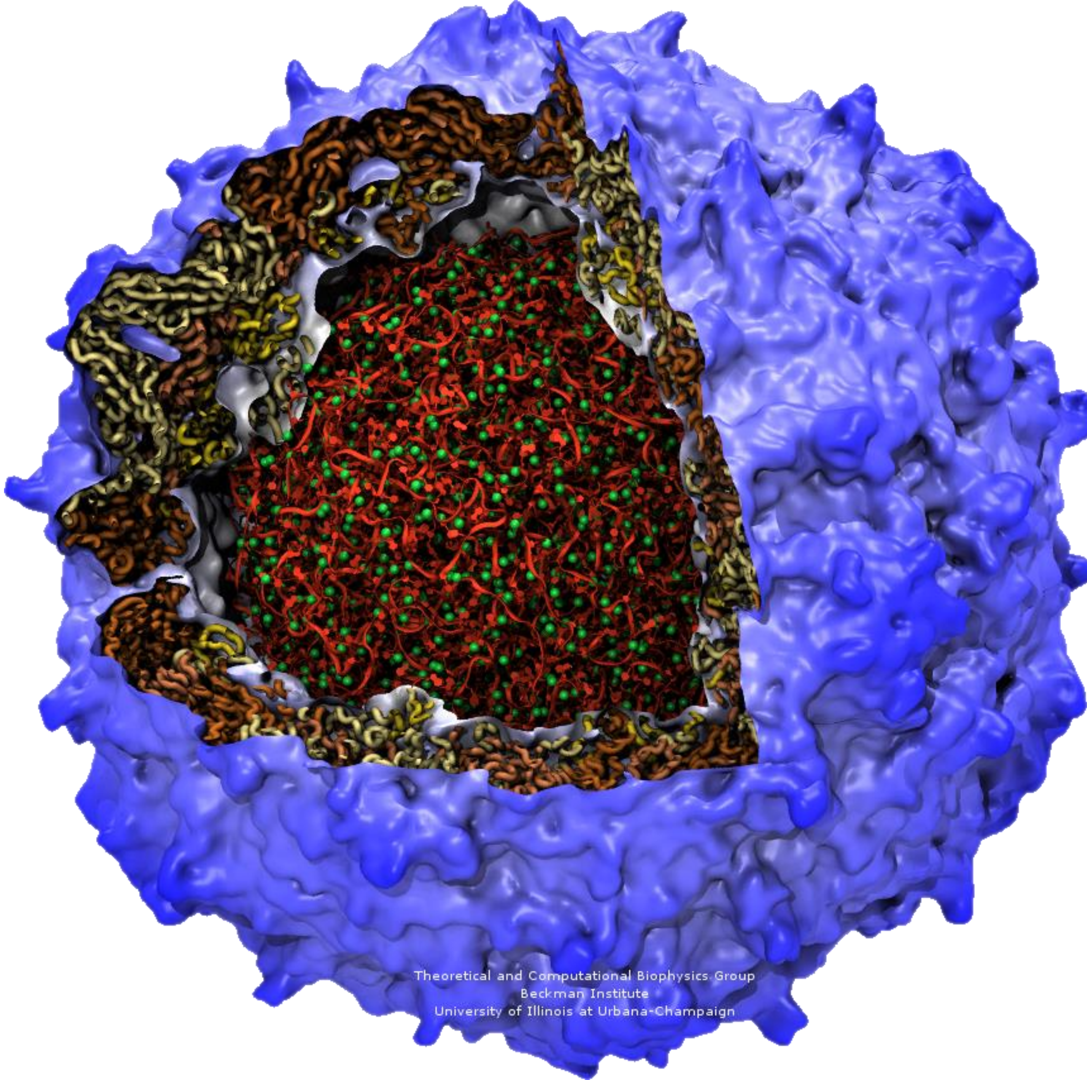

Robert R. Johnson
http://astro.temple.edu/~rjohnson





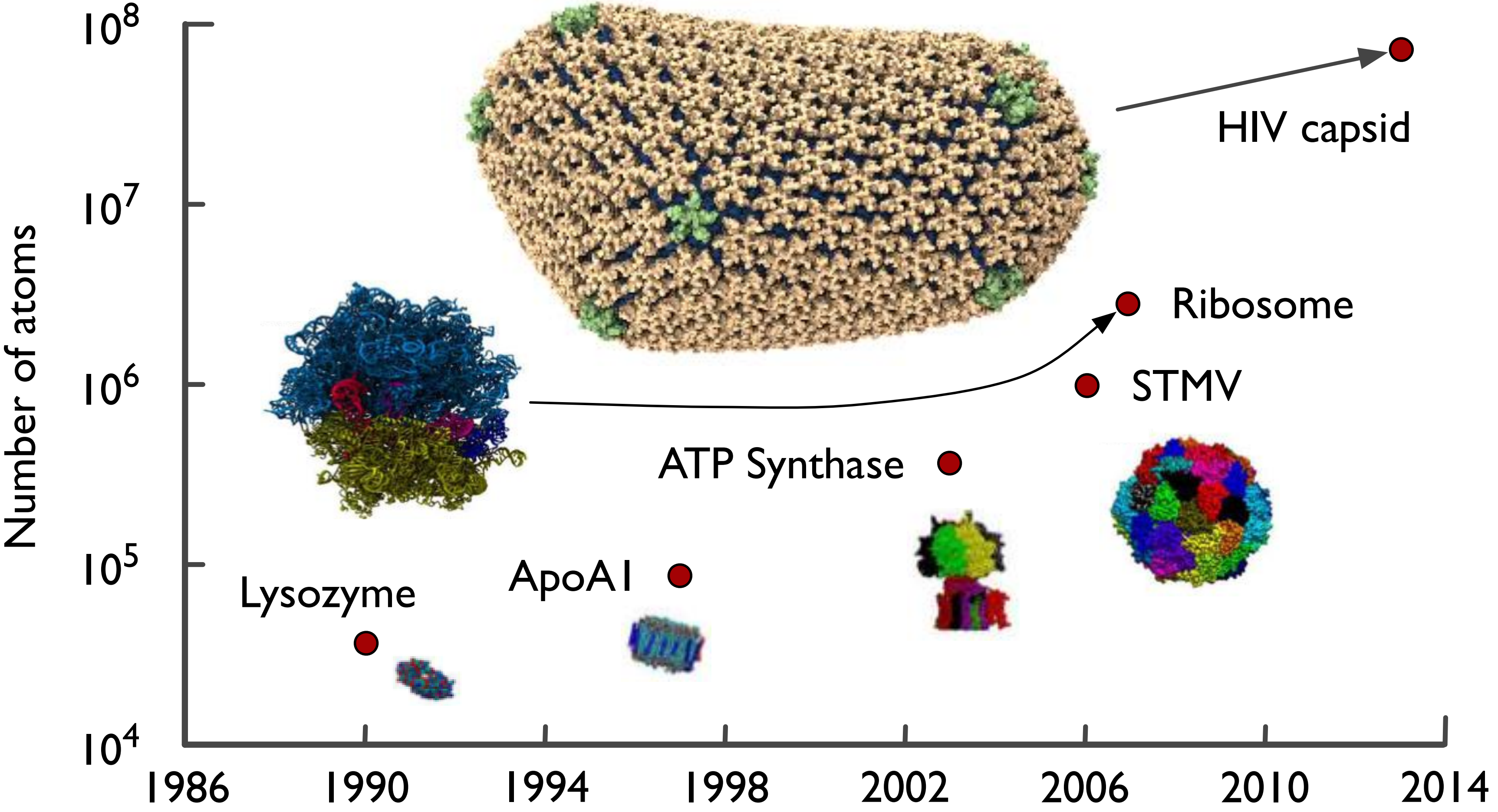


Theoretical and Computational Biophysics Group
Beckman Institute
University of Illinois at Urbana-Champaign



Theoretical and Computational Biophysics Group
Beckman Institute
University of Illinois at Urbana-Champaign

Computational Biology's Insatiable Demand for Processing Power



Visualization Goals, Challenges

- Increased GPU acceleration for visualization of **petascale molecular dynamics trajectories**
- **Overcome GPU memory capacity limits**, enable high quality visualization of >100M atom systems
- Use GPU to accelerate not only interactive-rate visualizations, but also photorealistic ray tracing with **artifact-free ambient occlusion lighting**, etc.
- Maintain **ease-of-use**, intimate link to VMD analytical features, atom selection language, etc.

Why Built-In VMD Ray Tracing Engines?

- **No disk I/O** or communication to outboard renderers
- **Eliminate unnecessary data replication and host-GPU memory transfers**
- Directly operate on VMD internal molecular scene, **quantized/compressed data formats**
- Implement all **curved surface primitives**, volume rendering, texturing, shading features required by VMD
- **Same scripting, analysis, atom selection**, and rendering features are available on all platforms, **graceful CPU fallback**

VMD GPU-Accelerated Ray Tracing Engine:
“TachyonL-OptiX”

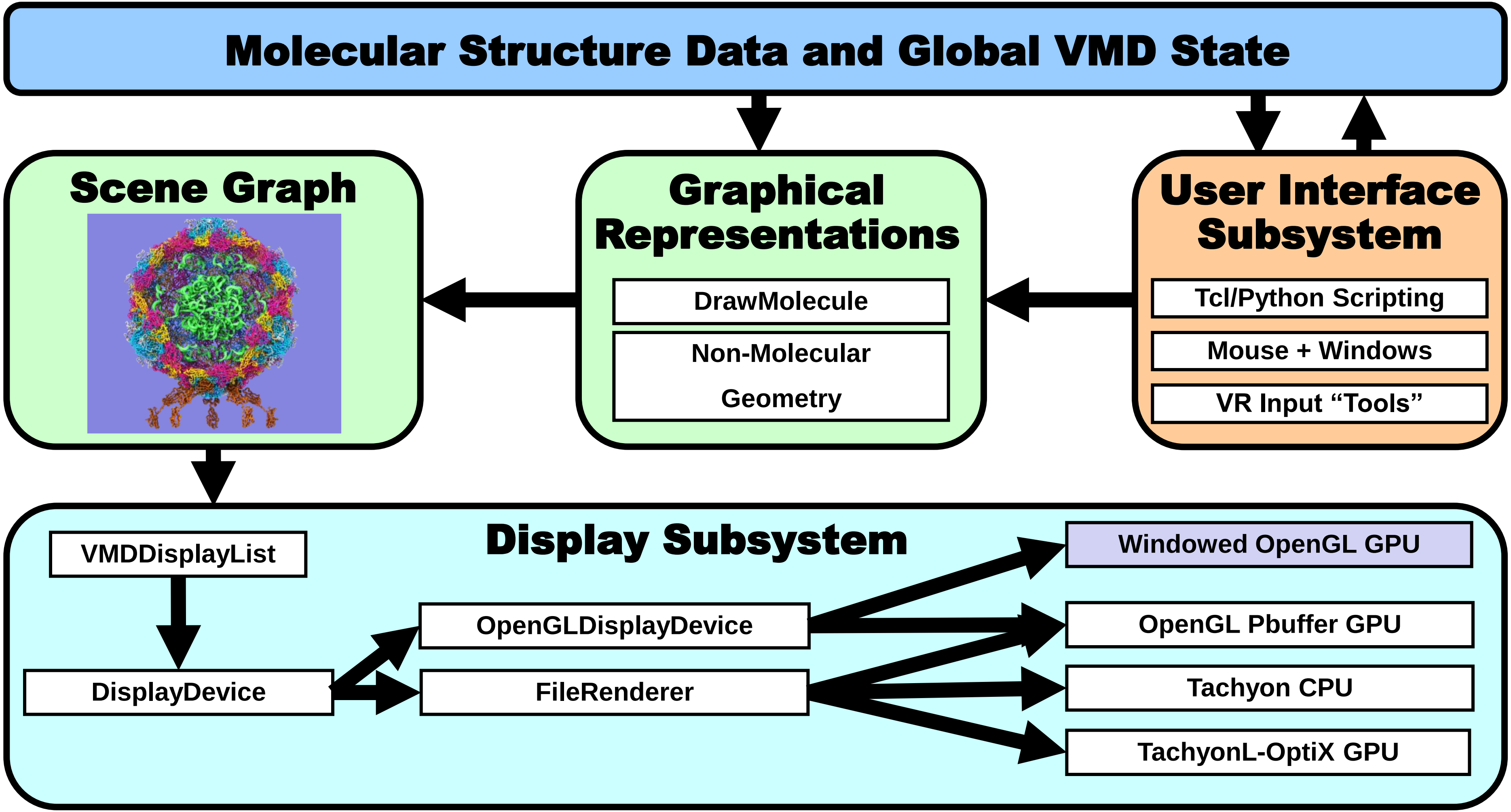
- Complementary to VMD OpenGL GLSL renderer that uses fast, interactivity-oriented rendering techniques
- Key ray tracing benefits: ambient occlusion lighting, shadows, high quality transparent surfaces, …
  - Subset of Tachyon parallel ray tracing engine in VMD
  - GPU acceleration w/ CUDA+OptiX ameliorates long rendering times associated with advanced lighting and shading algorithms
    - **Ambient occlusion generates large shadow test workload**
    - **Transparent surfaces and transmission rays can increase secondary ray counts by another order of magnitude**
  - Adaptation of Tachyon to the GPU required careful avoidance of GPU branch divergence, use of GPU memory layouts, etc.

VMD w/ OpenGL GLSL vs. GPU Ray Tracing

- GPU Ray Tracing:
  - Entire scene resident in GPU on-board memory for speed
  - RT performance is **heavily dependent on BVH** acceleration, particularly for scenes with large secondary ray workloads – shadow rays, ambient occlusion shadow feelers, transmission rays
  - RT **BVH structure regenerated / updated each trajectory timestep**, for some petascale visualizations BVH gen. can take up to ~25 sec!
- OpenGL GLSL:
  - No significant per-frame preprocessing required
  - Minimal persistent GPU memory footprint
  - Implements point sprites, ray cast spheres, pixel-rate lighting, …

TachyonL-Optix GPU Ray Tracing w/ OptiX+CUDA

- OptiX/CUDA kernels can only run for about 2 seconds uninterrupted
- GPU RT therefore cannot go wild with uninterrupted recursion, internal looping within shading code, or **GPU timeout will occur and kernel will be terminated** by OS/driver
- Complex ray tracing algorithms broken out into **multi-pass algorithms**:
  - Many GPU kernel launches (up to hundreds in some cases)
  - Intermediate rendering state written to GPU memory at end of each pass
  - Intermediate rendering state is reloaded at the start of the next pass
  - **Examples: state of multiple random number generators, color accumulation buffers, are stored and reloaded in our current implementation**



Molecular Structure Data and Global VMD State

Scene Graph

Graphical Representations

DrawMolecule

Non-Molecular Geometry

User Interface Subsystem

Tcl/Python Scripting

Mouse + Windows

VR Input “Tools”

Display Subsystem

VMDDisplayList

DisplayDevice

OpenGLDisplayDevice

FileRenderer

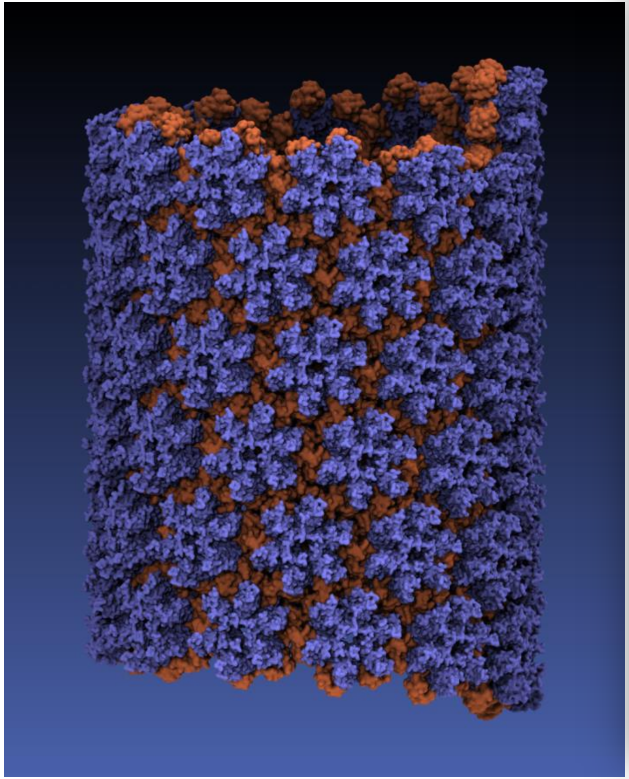
Windowed OpenGL GPU

OpenGL Pbuffer GPU

Tachyon CPU

TachyonL-OptiX GPU

VMD “QuickSurf” Representation



nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

THE HIV-1 CAPSID

Atomic structure of the AIDS pathogen’s protein coat

PAGE 643

COSMOLOGY

THE FIRST LIGHT

In pursuit of the most distant galaxies

PAGE 554

CITATION

CROSSING THE BORDERS

International collaborations make the most impact

PAGE 557

ANTICANCER DRUGS

A SITTING TARGET

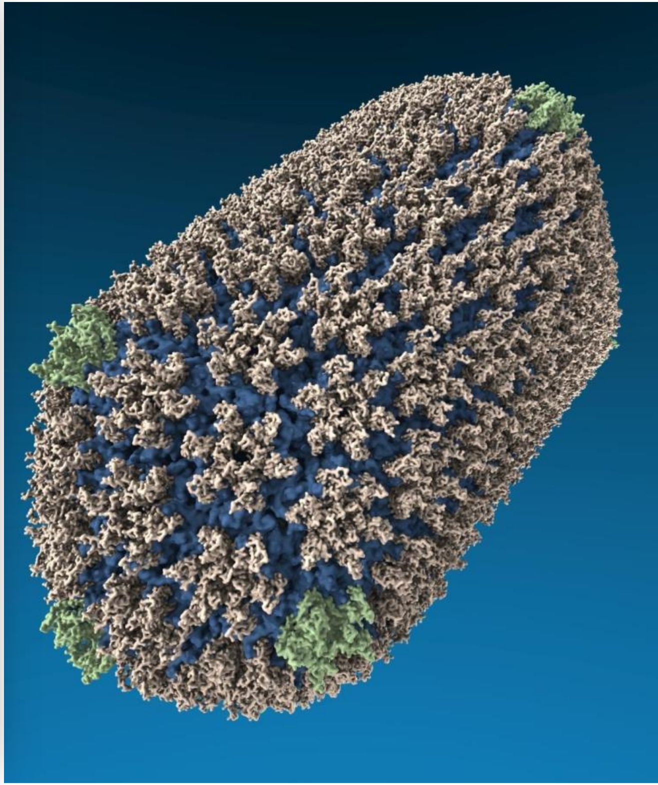
An indirect hit on ‘undruggable’ KRAS protein

PAGES 577 & 638

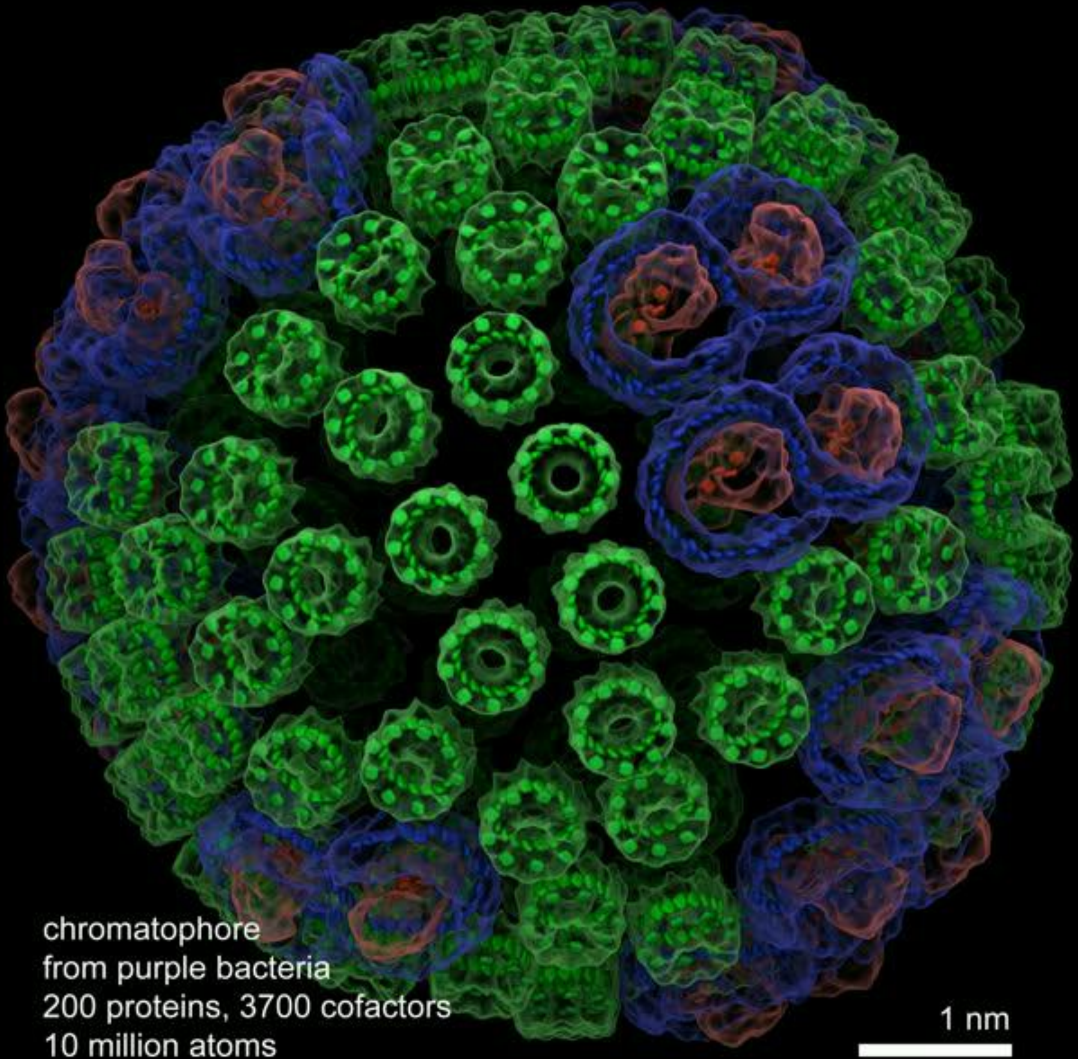
NATURE.COM/NATURE

30 May 2013

$10.00US $12.99CAN



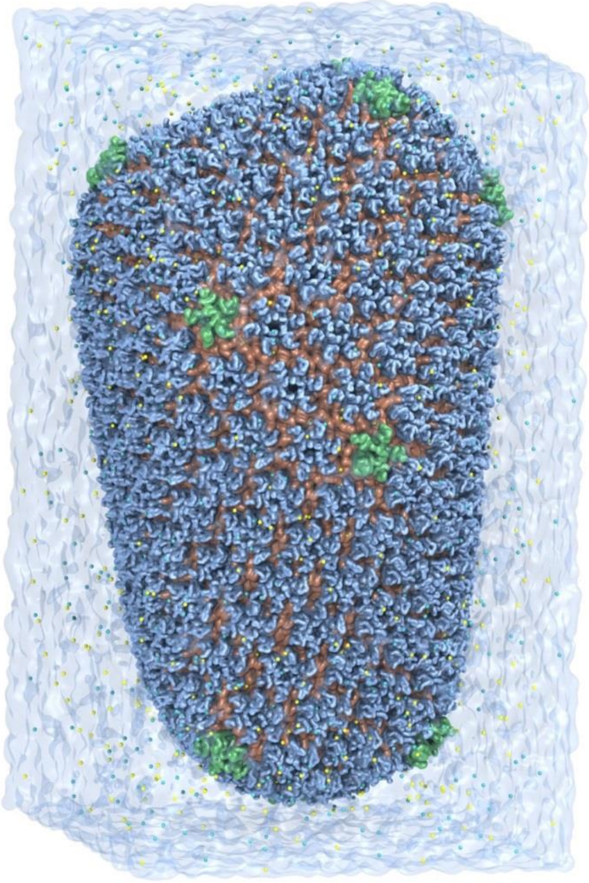
All-atom HIV capsid simulations w/ up to 64M atoms



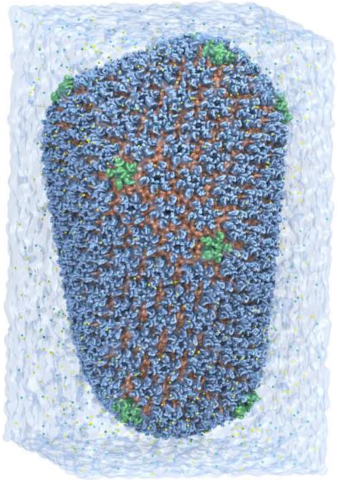
chromatophore
from purple bacteria
200 proteins, 3700 cofactors
10 million atoms

GPU Ray Tracing of HIV-1 on Blue Waters

- 64M atom simulation, 1079 movie frames
- **Ambient occlusion lightin**g, shadows, transparency, antialiasing, depth cueing, **144 rays/pixel minimum**
- GPU memory capacity hurdles:
  - Surface calc. and ray tracing each use **over 75% of K20X 6GB on-board GPU memory** even with quantized/compressed colors, surface normals, …
  - Evict non-RT GPU data to host prior to ray tracing
  - Eviction was **still required** on a test machine with a **12GB Quadro K6000** GPU – the multi-pass surface algorithm grows the per-pass chunk size to reduce the number of passes

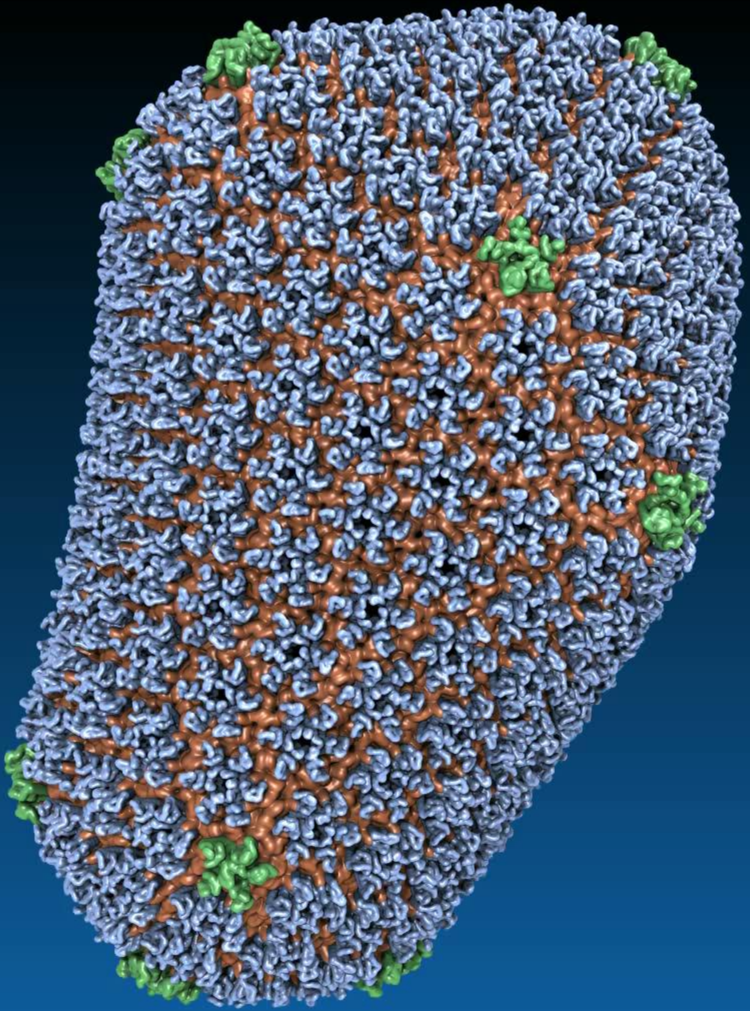


HIV-1 Parallel HD Movie Rendering on Blue Waters Cray XE6/XK7



New "TachyonL-OptiX" on XK7 vs. Tachyon on XE6:
K20X GPUs yield **up to eight times** geom+ray tracing speedup

Node Type and Count	Script Load Time	State Load Time	Geometry + Ray Tracing	Total Time
256 XE6 CPUs	7 s	160 s	**1,374 s**	**1,541 s**
512 XE6 CPUs	13 s	211 s	808 s	1,032 s
64 XK7 Tesla K20X GPUs	2 s	38 s	655 s	695 s
128 XK7 Tesla K20X GPUs	4 s	74 s	331 s	410 s
256 XK7 Tesla K20X GPUs	**7 s**	**110 s**	**171 s**	**288 s**



Future Work

- Improve multi-pass ray casting implementation
- Improve GPU BVH regen speed for time-varying geometry, MD trajectories
- Performance improvements for ambient occlusion sampling strategy
- Interactive RT in VMD
- Continue tuning of GPU-specific RT intersection routines, memory layout
- Add GPU-accelerated movie encoder back-end

Acknowledgements

- Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign
- NCSA Blue Waters Team
- NVIDIA CUDA Center of Excellence, University of Illinois at Urbana-Champaign
- NVIDIA OptiX team – especially James Bigler
- NVIDIA CUDA team
- Funding:
  - NSF OCI 07-25070
  - NSF PRAC “The Computational Microscope”
  - NIH support: 9P41GM104601, 5R01GM098243-02

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

GPU Computing Publications

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

- **GPU-Accelerated Molecular Visualization on Petascale Supercomputing Platforms.** J. Stone, K. L. Vandivort, and K. Schulten. UltraVis'13: Proceedings of the 8th International Workshop on Ultrascale Visualization, pp. 6:1-6:8, 2013.
- **Early Experiences Scaling VMD Molecular Visualization and Analysis Jobs on Blue Waters.** J. E. Stone, B. Isralewitz, and K. Schulten. In proceedings, Extreme Scaling Workshop, 2013.
- **Lattice Microbes: High-performance stochastic simulation method for the reaction-diffusion master equation.**
  E. Roberts, J. E. Stone, and Z. Luthey-Schulten.
  J. Computational Chemistry 34 (3), 245-255, 2013**.**
- **Fast Visualization of Gaussian Density Surfaces for Molecular Dynamics and Particle System Trajectories.** M. Krone, J. E. Stone, T. Ertl, and K. Schulten. *EuroVis Short Papers*, pp. 67-71, 2012.
- **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.

GPU Computing Publications

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

- **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.
- **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.