Visualization of Petascale Molecular Dynamics Simulations

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/

Imaging at Illinois:
Computational Imaging and Visualization
Beckman Institute, University of Illinois, June 1, 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/

Electrons in Vibrating Buckyball

Cellular Tomography, Cryo-electron Microscopy

Whole Cell Simulations

Polio virus

Ribosome Sequences
Goal: A Computational Microscope

- Study the molecular machines in living cells

Ribosome: synthesizes proteins from genetic information, target for antibiotics

Silicon nanopore: bionanodevice for sequencing DNA efficiently
Meeting the Diverse Needs of the Molecular Modeling Community

• Over 212,000 registered users
  – 18% (39,000) are NIH-funded
  – Over 49,000 have downloaded multiple VMD releases

• Over 6,600 citations

• User community runs VMD on:
  – MacOS X, Unix, Windows operating systems
  – Laptops, desktop workstations
  – Clusters, supercomputers

• VMD user support and service efforts:
  – 20,000 emails, 2007-2011
  – Develop and maintain VMD tutorials and topical mini-tutorials; 11 in total
  – Periodic user surveys
VMD Interoperability –
Linked to Today’s Key Research Areas

- Unique in its interoperability with a broad range of modeling tools: AMBER, CHARMM, CPMD, DL_POLY, GAMESS, GROMACS, HOOMD, LAMMPS, NAMD, and many more …

- Supports key data types, file formats, and databases, e.g. electron microscopy, quantum chemistry, MD trajectories, sequence alignments, super resolution light microscopy

- Incorporates tools for simulation preparation, visualization, and analysis
Support for Diverse Display Hardware:
Stereoscopic Displays, 6-DOF Input
Support for Diverse Display Hardware: Stereoscopic Projection for Presentations to Large Groups
Support for Diverse Display Hardware:
CAVE, 3-D Workbench, Tiled Projector Arrays
Immersive Visualization in VMD: CAVE, 6-DOF Input w/ Stereo Display
Challenges for Immersive Visualization of Dynamics of Large Structures

• Graphical representations re-generated for each animated simulation trajectory frame:
  – Dependent on user-defined atom selections

• Although visualizations often focus on interesting regions of substructure, fast display updates require rapid traversal of molecular data structures

• Optimized atom selection traversal:
  – Increased performance of per-frame updates by ~10x for 116M atom BAR case with 200,000 selected atoms

• New GLSL point sprite sphere shader:
  – Reduce host-GPU bandwidth for displayed geometry
  – Over 20x faster than old GLSL spheres drawn using display lists — drawing time is now inconsequential

• Optimized all graphical representation generation routines for large atom counts, sparse selections

116M atom BAR domain test case: 200,000 selected atoms, stereo trajectory animation 70 FPS, static scene in stereo 116 FPS
VMD “QuickSurf” Representation

- Large biomolecular complexes are difficult to interpret with atomic detail graphical representations
- Even secondary structure representations become cluttered
- Surface representations are easier to use when greater abstraction is desired, but are computationally costly
- Existing surface display methods incapable of animating dynamics of large structures
VMD “QuickSurf” Representation

- Displays continuum of structural detail:
  - All-atom models
  - Coarse-grained models
  - Cellular scale models
  - Multi-scale models: All-atom + CG, Brownian + Whole Cell
  - Smoothly variable between full detail, and reduced resolution representations of very large complexes

Fast Visualization of Gaussian Density Surfaces for Molecular Dynamics and Particle System Trajectories.

VMD “QuickSurf” Representation

- Uses multi-core CPUs and GPU acceleration to enable **smooth real-time animation** of MD trajectories
- Linear-time algorithm, scales to millions of particles, as limited by memory capacity
QuickSurf Representation of Lattice Cell Models

Continuous particle based model – often 70 to 300 million particles

Discretized lattice models derived from continuous model shown in a surface representation
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
Timeline Plugin: Analyze MD Trajectories for Events

VMD Timeline plugin: live 2D plot linked to 3D structure
- Single picture shows changing properties across entire structure+trajectory
- Explore time vs. per-selection attribute, linked to molecular structure
- Many analysis methods available; user-extendable

Recent progress:
- Faster analysis with new VMD SSD trajectory formats, GPU acceleration
- Per-secondary-structure native contact and density correlation graphing
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

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
Molecular Visualization and Analysis Challenges for Petascale Simulations

• Very large structures (10M to over 100M atoms)
  – 12-bytes per atom per trajectory frame
  – One 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

• Terabytes to petabytes of data, 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
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
# GPU Accelerated Trajectory Analysis and Visualization in VMD

<table>
<thead>
<tr>
<th>GPU-Accelerated Feature</th>
<th>Speedup vs. single CPU core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular orbital display</td>
<td>120x</td>
</tr>
<tr>
<td>Radial distribution function</td>
<td>92x</td>
</tr>
<tr>
<td>Electrostatic field calculation</td>
<td>44x</td>
</tr>
<tr>
<td>Molecular surface display</td>
<td>40x</td>
</tr>
<tr>
<td>Ion placement</td>
<td>26x</td>
</tr>
<tr>
<td>MDFF density map synthesis</td>
<td>26x</td>
</tr>
<tr>
<td>Implicit ligand sampling</td>
<td>25x</td>
</tr>
<tr>
<td>Root mean squared fluctuation</td>
<td>25x</td>
</tr>
<tr>
<td>Radius of gyration</td>
<td>21x</td>
</tr>
<tr>
<td>Close contact determination</td>
<td>20x</td>
</tr>
<tr>
<td>Dipole moment calculation</td>
<td>15x</td>
</tr>
</tbody>
</table>
Time-Averaged Electrostatics Analysis on Energy-Efficient GPU Cluster

- **1.5 hour** job (CPUs) reduced to **3 min** (CPUs+GPU)
- Electrostatics of thousands of trajectory frames averaged
- Per-node power consumption on NCSA “AC” GPU cluster:
  - CPUs-only: 299 watts
  - CPUs+GPUs: 742 watts
- GPU Speedup: **25.5x**
- Power efficiency gain: **10.5x**

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

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

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