Interactive Supercomputing for State-of-the-art Biomolecular Simulation

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

Interactivity in Supercomputing BoF,
12:15pm-1:15pm, Rooms 503-504, Colorado Convention Center
Denver, CO, Tuesday Nov 14th, 2017
VMD – “Visual Molecular Dynamics”

- Visualization and analysis of:
  - Molecular dynamics simulations
  - Lattice cell simulations
  - Quantum chemistry calculations
  - Sequence information

- User extensible scripting and plugins
- http://www.ks.uiuc.edu/Research/vmd/
Goal: A Computational Microscope

Study the molecular machines in living cells

Ribosome: target for antibiotics

Poliovirus
How Does Interactivity in HPC Benefit Scientific Productivity

- Until 2008, MD simulation performance was the major limiting factor on biomolecular modeling productivity, maybe 95% factor.
- In 2017: Batch MD sim code performance is likely only a 50% factor on overall scientific productivity.
- An extremely simplified outline:
  - Months or years of science team effort go into preparing state-of-the-art biomolecular simulations w/ experimental structures before MD sims begin.
  - During the initial and early phases of a simulation campaign, hundreds of modeling problems must be identified and addressed.
  - Interactivity is a fundamental requirement during these phases.
  - After sim campaign has produced required outputs, extensive analytical work begins, and science teams then make key decisions about outcomes, potentially conducting further sim studies, developing manuscripts with detailed analyses, figures, and movies of dynamics of systems under study.
Proto-Cell Modeling Project

- Initial model of minimal cell membrane and proteins
- Candidate for our ORNL Summit early science project
- ~1 billion atoms presently with explicit water, ions, etc.
- Main structure file used for modeling and simulation tasks is currently ~60GB
- Initial simulation preparation requires many repeated build / viz / measure / simulate cycles…
- MD trajectory output : ~12GB/frame
- Need to run all tools required for science campaign at remote supercomputer sites:
  - Optimize work-efficiency of science team
  - Not just NAMD and VMD, but also other tools
- Old approach of repeatedly transferring files back and forth between UIUC lab and remote supercomputers is untenable at this scale!!!
What Do We Need?

• **Science campaign “Bullpen” interactive compute/viz**
• Interactive resources supporting the **human-intensive** parts of science team activities
• **Specific examples:**
  – **Interactive remote visualization** supporting use of arbitrary science tools, including those that are **not traditional HPC tools** (e.g. CryoEM and X-Ray crystallography pkgs)
  – **Special batch queues, or queue policies** that support the initial and early phase of simulation bring-up activities, so that **meso-scale interactive calculations** can be performed “at a mouse click” to provide rapid-turnaround of results in interactive modeling sessions to correct flaws in molecular models, simulation parameters, etc.
  – A few (10s or less) HPC systems supporting **very large memory capacities** and dense **GPU counts**, to run the **interactive supercomputing job “front end”**
VMD Petascale Visualization and Analysis

- Analyze/visualize large trajectories too large to transfer off-site:
  - User-defined parallel analysis operations, data types
  - Parallel rendering, movie making

- Supports GPU-accelerated Cray XK7 nodes for both visualization and analysis:
  - GPU accelerated trajectory analysis w/ CUDA
  - OpenGL and GPU ray tracing for visualization and movie rendering

- Parallel I/O rates up to 275 GB/sec on 8192 Cray XE6 nodes – can read in 231 TB in 15 minutes!

Parallel VMD currently available on:

ORNL Titan, NCSA Blue Waters, Indiana Big Red II, CSCS Piz Daint, and similar systems
GPUs Can Reduce MDFF Trajectory Analysis Runtimes from Hours to Minutes

GPUs enable laptops and desktop workstations to handle tasks that would have previously required a cluster, or a very long wait...

GPU-accelerated petascale supercomputers enable analyses that were previously impractical, allowing detailed study of very large structures such as viruses

GPU-accelerated MDFF Cross Correlation Timeline
Regions with poor fit
Regions with good fit
Parallel MDFF Cross Correlation Analysis on Cray XK7

Rabbit Hemorrhagic Disease Virus (RHDV)

<table>
<thead>
<tr>
<th>Traj. frames</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure component selections</td>
<td>720</td>
</tr>
<tr>
<td>Single-node XK7 (projected)</td>
<td>336 hours (14 days)</td>
</tr>
<tr>
<td>128-node XK7</td>
<td>3.2 hours 105x speedup</td>
</tr>
<tr>
<td>2048-node XK7</td>
<td>19.5 minutes 1035x speedup</td>
</tr>
</tbody>
</table>

Calculation of 7M CCs would take **5 years** using serial CPU algorithm!

VMD Tesla P100+V100 Cross Correlation Performance
Rabbit Hemorrhagic Disease Virus: 702K atoms, 6.5Å resolution

Latest results on Pascal P100, first runs on Volta V100

<table>
<thead>
<tr>
<th>Hardware platform</th>
<th>Runtime, Speedup vs. Chimera, VMD+GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimera Xeon E5-2687W (2 socket) [1]</td>
<td>15.860s, 1x</td>
</tr>
<tr>
<td>VMD-CUDA IBM Power8 + 1x Tesla K40 [2]</td>
<td>0.488s, 32x 0.9x</td>
</tr>
<tr>
<td>VMD-CUDA Intel Xeon E5-2687W + 1x Quadro K6000 [1,2]</td>
<td>0.458s, 35x 1.0x</td>
</tr>
<tr>
<td>VMD-CUDA Intel Xeon E5-2698v3 + 1x Tesla P100</td>
<td>0.090s, 176x 5.1x</td>
</tr>
<tr>
<td>VMD-CUDA IBM Power8 “Minsky” + 1x Tesla P100</td>
<td>0.080s, 198x 5.7x</td>
</tr>
<tr>
<td><strong>VMD-CUDA Intel Xeon E5-2697Av4 + 1x Tesla V100</strong></td>
<td><strong>0.050s, 317x 9.2x</strong></td>
</tr>
</tbody>
</table>


Clustering Analysis of Molecular Dynamics Trajectories

Interactive Ray Tracing

- Remote visualization tasks on very large macromolecular complexes
- Large-scale parallel rendering: in situ or post hoc visualization tasks

GPU-Accelerated Molecular Visualization on Petascale Supercomputing Platforms.

Visualization of Energy Conversion Processes in a Light Harvesting Organelle at Atomic Detail.
M. Sener, et al. SC’14 Visualization and Data Analytics Showcase, 2014.


Atomic Detail Visualization of Photosynthetic Membranes with GPU-Accelerated Ray Tracing.


VMD/OptiX GPU Ray Tracing of all-atom Chromatophore w/ lipids.
VMD 1.9.3 supports EGL for in-situ and parallel rendering on clouds, clusters, and supercomputers

• Eliminate dependency on windowing systems
• Simplified deployment of parallel VMD builds supporting off-screen rendering
• Maintains 100% of VMD OpenGL shaders and rendering features
• Support high-quality vendor-supported commercial OpenGL implementations in HPC systems that were previously limited to Mesa
EGL Is Supported Now!

- Cloud+Workstations with most recent NVIDIA drivers
- VMD on HPC systems w/ latest Tesla P100 GPUs:
  - Cray XC50, CSCS Piz Daint, driver 375.39
  - IBM OpenPOWER, drivers 375.66 and later support both GLX and EGL
“When I was a young man, my goal was to look with mathematical and computational means at the inside of cells, one atom at a time, to decipher how living systems work. That is what I strived for and I never deflected from this goal.” – Klaus Schulten
Acknowledgements

• Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign

• Funding:
  – NIH support: P41GM104601
  – DOE INCITE, ORNL Titan: DE-AC05-00OR22725
  – NSF Blue Waters:
    NSF OCI 07-25070, PRAC “The Computational Microscope”, ACI-1238993, ACI-1440026