Harnessing GPUs to Probe Biomolecular Machines at Atomic Detail

John E. Stone
Theoretical and Computational Biophysics Group
University of Illinois at Urbana-Champaign
http://www.ks.uiuc.edu/Research/gpu/
NVIDIA GPU Technology Theater
4:30pm, Salt Palace Convention Center,
Salt Lake City, UT, Wednesday Nov 16th, 2016
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/

Structure

- X-ray, cryo-EM, cryo-ET, NMR

Parameterization

Refinement with MDFF

Analysis

Preparation with QwikMD

Cell-Scale Modeling

MD Simulation

Remote Visualization

MD/Cell Simulation

LM

NAMD

Amber

Gromacs
VMD Interoperability Serves Many Communities

- Uniquely interoperable with a broad range of tools:
  - AMBER, CHARMM, CPMD, DL_POLY, GAMESS, GROMACS, HOOMD, LAMMPS, NAMD, and many more …

- Supports key data types, file formats, and databases

- Incorporates tools for simulation preparation, visualization, and analysis
9.5 Years of GPU Computing in VMD

- Has stood the test of time
- Modeling, Visualization, Rendering, and Analysis

Blast from the past:
CUDA starting with version 0.7 !!!
Quad core Intel QX6700, three NVIDIA GeForce 8800GTX GPUs, RHEL4 Linux

Adaptation of VMD to EGL for in-situ and parallel rendering on clouds, clusters, and supercomputers

- Eliminate dependency on windowing systems
- Easy deployment of parallel VMD builds w/ off-screen rendering
- Maintains 100% of VMD OpenGL shaders and rendering features
- High-quality commercial OpenGL implementations in HPC systems
- Easier management of multi-GPU nodes and NUMA affinity issues

Poliovirus

High Performance Molecular Visualization: In-Situ and Parallel Rendering with EGL
OpenGL: GLX vs. EGL

Diagram showing the difference between GLX and EGL in the OpenGL environment.
Molecular Structure Data and Global VMD State

Scene Graph

Display Subsystem

VMDDisplayList

DisplayDevice

OpenGLRenderer

Graphical Representations

DrawMolecule

Non-Molecular Geometry

User Interface Subsystem

Tcl/Python Scripting

Mouse + Windows

6DoF Input “Tools”
VMD EGL rendering:

- Supports all VMD shaders and associated OpenGL features:
  - Pixel-rate lighting
  - Ray-cast spheres w/ GLSL
  - 3-D texture mapping
  - Text rendering
  - Multisample antialiasing
  - And much more...

**VMD EGL Performance on Amazon EC2 Cloud**

<table>
<thead>
<tr>
<th>MPI Ranks</th>
<th>EC2 “G2.8xlarge” GPU Instances</th>
<th>HIV-1 movie rendering time (sec), (I/O %) 3840x2160 resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>626s (10% I/O)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>347s (19% I/O)</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>221s (31% I/O)</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>141s (46% I/O)</td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>107s (64% I/O)</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>90s (76% I/O)</td>
</tr>
</tbody>
</table>

Performance at 32 nodes reaches ~48 frames per second


64M atom HIV-1 capsid simulation rendered via EGL
VMD 1.9.3+OptiX 4.0

- Interactive RT on laptops, desktops, and cloud
- Large-scale parallel rendering: in situ or post hoc visualization tasks
- Remote RT on NVIDIA VCA clusters
- Stereoscopic panoramic and full-dome projections
- Omnidirectional VR for YouTube, VR HMDs


VMD/OptiX GPU Ray Tracing of all-atom Chromatophore w/ lipids.
Interactive RT of All-Atom Minimal Cell Envelope

- 200 nm spherical envelope
- Membrane with ~50% occupancy by proteins (2000x Aquaporin channels)
- 42M atoms in membrane
- Interactive RT w/ 2 dir. lights and AO on Kepler GeForce Titan X @ ~12 FPS
- Complete model with correct proteins, solvent, etc, will contain billions of atoms
Proto-Cell Rendered with VMD+OptiX

• 113M particles
• 1,397 copies of 14 different membrane proteins
• Preparing for simulations on pre-exascale computers
Stereoscopic Panorama Ray Tracing w/ OptiX

- Render 360° images and movies for VR headsets such as Oculus, Vive, GearVR, Google Cardboard, and YouTube VR
- Ray trace panoramic stereo spheremaps or cubemaps for very high-frame-rate display via OpenGL texturing onto simple geometry
- Stereo requires spherical camera projections poorly suited to rasterization
- Benefits from OptiX multi-GPU rendering and load balancing, remote visualization
Immersive Molecular Visualization with Omnidirectional Stereoscopic Ray Tracing and Remote Rendering.
Molecular Dynamics Flexible Fitting (MDFF)

X-ray crystallography → MDFF → Electron microscopy

APS at Argonne

Evaluating Quality-of-Fit for Structures Solved by Hybrid Fitting Methods

Compute Pearson correlation to evaluate quality-of-fit between a reference cryo-EM density map and a simulated density map produced from an all-atom structure.
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
Padding optimizes global memory performance, guaranteeing coalesced global memory accesses.

Small 8x8x2 CUDA thread blocks afford large per-thread register count, shared memory.

Each thread computes 4 z-axis density map lattice points and associated CC partial sums.

3-D density map decomposes into 3-D grid of 8x8x8 tiles containing CC partial sums and local CC values.

Grid of thread blocks:

- 0,0
- 0,1
- ...
- 1,0
- 1,1
- ...
- ...

Fusion of density and CC calculations into a single CUDA kernel!!!

Spatial CC map and overall CC value computed in a single pass.

Threads producing results that are used

Inactive threads, region of discarded output

Single-Pass MDFF GPU Cross-Correlation
VMD Tesla P100 Cross Correlation Performance
Rabbit Hemorrhagic Disease Virus: 702K atoms, 6.5Å resolution

P100 Die-Stacked Mem Accelerates Bandwidth Intensive Calculation

<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-CPU IBM Power8 (2 socket) [2]</td>
<td>1.334s, 12x</td>
</tr>
<tr>
<td>VMD-CPU Intel Xeon E5-2660v3 (2 socket) [2]</td>
<td>0.905s, 17x</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><strong>VMD-CUDA Intel Xeon E5-2698v3 + 1x Tesla P100</strong></td>
<td><strong>0.090s, 176x 5.1x</strong></td>
</tr>
<tr>
<td><strong>VMD-CUDA IBM Power8 “Minsky” + 1x Tesla P100</strong></td>
<td><strong>0.080s, 198x 5.7x</strong></td>
</tr>
</tbody>
</table>

## VMD Tesla P100 Performance for C_{60} Molecular Orbitals, 516x519x507 grid

<table>
<thead>
<tr>
<th>Hardware platform</th>
<th>Runtime,</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM Power8 (2 socket) (ORNL ‘crest’) [1]</td>
<td>8.03s,</td>
<td>0.4x</td>
</tr>
<tr>
<td>Intel Xeon E5-2660v3 (2 socket) [1]</td>
<td>7.14s,</td>
<td>0.5x</td>
</tr>
<tr>
<td>IBM Power8 (ORNL ‘crest’) + 1x Tesla K40 [1]</td>
<td>3.49s,</td>
<td>1.0x</td>
</tr>
<tr>
<td>Intel Xeon E5-2698v3 + 1x Tesla P100</td>
<td>1.35s,</td>
<td>2.5x</td>
</tr>
<tr>
<td><strong>IBM Power8 “Minsky” + 1x Tesla P100</strong></td>
<td><strong>1.09s</strong>,</td>
<td><strong>3.3x</strong></td>
</tr>
<tr>
<td>IBM Power8 (ORNL ‘crest’) + 4x Tesla K40 [1]</td>
<td>0.91s,</td>
<td>3.8x</td>
</tr>
<tr>
<td>Intel Xeon E5-2698v3 + 4x Tesla P100</td>
<td>0.37s,</td>
<td>9.4x</td>
</tr>
<tr>
<td><strong>IBM Power8 “Minsky” + 4x Tesla P100</strong></td>
<td><strong>0.30s</strong>,</td>
<td><strong>11.6x</strong></td>
</tr>
</tbody>
</table>

Come See the VMD+Unreal Chromatophore VR Demo in the NVIDIA VR Room!
Acknowledgements

• Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign
• CUDA Center of Excellence, University of Illinois at Urbana-Champaign
• NVIDIA CUDA, OptiX, GL/EGL teams, and PSG cluster admins
• IBM and NVIDIA for access to “Minsky” Power8+P100 hardware
• NCSA Blue Waters Team, ORNL Titan Team, ORNL CAAR
• Funding:
  – DOE INCITE, ORNL Titan: DE-AC05-00OR22725
  – NIH support: 9P41GM104601, 5R01GM098243-02
NIH BTRC for Macromolecular Modeling and Bioinformatics
1990-2017

Beckman Institute
University of Illinois at Urbana-Champaign
Related Publications
http://www.ks.uiuc.edu/Research/gpu/


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


- **Winner of the SC’14 Visualization and Data Analytics Showcase**


- **Unlocking the Full Potential of the Cray XK7 Accelerator.** M. D. Klein and J. E. Stone. Cray Users Group, Lugano Switzerland, May 2014.


Related Publications

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


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


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


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


