GPU-Accelerated Analysis and Visualization of Large Structures Solved by Molecular Dynamics Flexible Fitting

John E. Stone, Ryan McGreevy, Barry Isralevitz, and Klaus Schulten
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/
http://www.ks.uiuc.edu/Research/gpu/
Faraday Discussion 169: Molecular Simulations and Visualization
Nottingham, UK, May 8, 2014
Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics.
Molecular Dynamics Flexible Fitting - Theory

Two terms are added to the MD potential

$$U_{total} = U_{MD} + U_{EM} + U_{SS}$$

An external potential derived from the EM map is defined on a grid as

$$U_{EM}(R) = \sum_j w_j V_{EM}(r_j)$$

$$V_{EM}(r) = \begin{cases} 
\xi \left(1 - \frac{\Phi(r) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}\right) & \text{if } \Phi(r) \geq \Phi_{thr}, \\
\xi & \text{if } \Phi(r) < \Phi_{thr}.
\end{cases}$$

A mass-weighted force is then applied to each atom

$$f_{i}^{EM} = -\nabla U_{EM}(R) = -w_i \frac{\partial V_{EM}(r_i)}{\partial r_i}$$
Structural Route to the all-atom HIV-1 Capsid

1st TEM (1999) 1st tomography (2003) Crystal structures of separated hexamer and pentamer

Briggs et al. *EMBO J*, 2003
Briggs et al. *Structure*, 2006

cryo-ET (2006)


High res. EM of hexameric tubule, tomography of capsid, all-atom model of capsid by MDFF w/ NAMD & VMD, NSF/NCSA Blue Waters computer at Illinois

hexameric tubule

Li et al., *Nature*, 2000
Byeon et al., *Cell* 2009

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

Compute Pearson correlation to evaluate the fit of a reference cryo-EM density map with a simulated density map produced from an all-atom structure.
GPUs Can Reduce 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 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
## CUDA GPU-Accelerated Trajectory Analysis and Visualization in VMD

<table>
<thead>
<tr>
<th>VMD GPU-Accelerated Feature or Kernel</th>
<th>Typical speedup vs. multi-core CPU (e.g. 4-core CPU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular orbital display</td>
<td>30x</td>
</tr>
<tr>
<td>Radial distribution function</td>
<td>23x</td>
</tr>
<tr>
<td>Molecular surface display</td>
<td>15x</td>
</tr>
<tr>
<td>Electrostatic field calculation</td>
<td>11x</td>
</tr>
<tr>
<td>Ray tracing w/ shadows, AO lighting</td>
<td>7x</td>
</tr>
<tr>
<td>Ion placement</td>
<td>6x</td>
</tr>
<tr>
<td>MDFF density map synthesis</td>
<td>6x</td>
</tr>
<tr>
<td>Implicit ligand sampling</td>
<td>6x</td>
</tr>
<tr>
<td>Root mean squared fluctuation</td>
<td>6x</td>
</tr>
<tr>
<td>Radius of gyration</td>
<td>5x</td>
</tr>
<tr>
<td>Close contact determination</td>
<td>5x</td>
</tr>
<tr>
<td>Dipole moment calculation</td>
<td>4x</td>
</tr>
</tbody>
</table>
Peak Arithmetic Performance Trend
Simulation of reaction diffusion processes over biologically relevant size and time scales using multi-GPU workstations
Michael J. Hallock, John E. Stone, Elijah Roberts, Corey Fry, and Zaida Luthey-Schulten.
http://dx.doi.org/10.1016/j.parco.2014.03.009

NIH BTRC for Macromolecular Modeling and Bioinformatics
http://www.ks.uiuc.edu/ Beckman Institute, U. Illinois at Urbana-Champaign
Padding optimizes global memory performance, guaranteeing coalesced global memory accesses.

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

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.

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

Threads producing results that are used.

Inactive threads, region of discarded output.

Grid of thread blocks.
# VMD GPU Cross Correlation Performance

<table>
<thead>
<tr>
<th></th>
<th>RHDV</th>
<th>Mm-cpn open</th>
<th>GroEL</th>
<th>Aquaporin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>6.5</td>
<td>8</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td><strong>Atoms</strong></td>
<td>702K</td>
<td>61K</td>
<td>54K</td>
<td>1.6K</td>
</tr>
<tr>
<td><strong>VMD-CUDA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quadro K6000</td>
<td>0.458s</td>
<td>0.06s</td>
<td>0.034s</td>
<td>0.007s</td>
</tr>
<tr>
<td></td>
<td>34.6x</td>
<td>25.7x</td>
<td>36.8x</td>
<td>55.7x</td>
</tr>
<tr>
<td><strong>VMD-CPU-SSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32-threads, 2x Xeon E5-2687W</td>
<td>0.779s</td>
<td>0.085s</td>
<td>0.159s</td>
<td>0.033s</td>
</tr>
<tr>
<td></td>
<td>20.3x</td>
<td>18.1x</td>
<td>7.9x</td>
<td>11.8x</td>
</tr>
<tr>
<td><strong>Chimera</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-thread Xeon E5-2687W</td>
<td>15.86s</td>
<td>1.54s</td>
<td>1.25s</td>
<td>0.39s</td>
</tr>
<tr>
<td></td>
<td>1.0x</td>
<td>1.0x</td>
<td>1.0x</td>
<td>1.0x</td>
</tr>
<tr>
<td><strong>VMD CPU-SEQ (plugin)</strong></td>
<td>62.89s</td>
<td>2.9s</td>
<td>1.57s</td>
<td>0.04s</td>
</tr>
<tr>
<td>1-thread Xeon E5-2687W</td>
<td>0.25x</td>
<td>0.53x</td>
<td>0.79x</td>
<td>9.7x</td>
</tr>
</tbody>
</table>
## VMD RHDV Cross Correlation Timeline on Cray XK7

<table>
<thead>
<tr>
<th></th>
<th>RHDV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atoms</td>
<td>702K</td>
</tr>
<tr>
<td>Traj. Frames</td>
<td>10,000</td>
</tr>
<tr>
<td>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</td>
</tr>
<tr>
<td></td>
<td>105x speedup</td>
</tr>
<tr>
<td>2048-node XK7</td>
<td>19.5 minutes</td>
</tr>
<tr>
<td></td>
<td>1035x speedup</td>
</tr>
</tbody>
</table>

Calculation would take **5 years** using original serial VMD CC plugin on a workstation!
Future Work, Invitation for Comments

• Interactive MDFF with real-time CC and spatial CC coloring
• Incorporate fast CC computation into MDFF simulations to optimize the efforts of the MDFF simulation protocol
• Programming approaches to simplify future single-pass kernels that combine complex compute/reduction ops
• **2x to 8x** performance increases possible:
  – Faster GPU **sort, scan** (parallel prefix sum)
  – Kepler GPU **shuffle** instructions for parallel reductions
  – **Multi-GPU** implementation (4x should be straightforward)
  – Intel/AMD **AVX2** 8-way vector instructions
  – Precompute **summed area tables (SAT)** for cheaper calculation of reference mean, count of non-excluded voxels, etc:
    CPU Yes?, GPU Maybe?
Acknowledgements

- Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign
- NVIDIA CUDA Center of Excellence, University of Illinois at Urbana-Champaign
- NVIDIA CUDA team
- NVIDIA OptiX team
- NCSA Blue Waters Team
- Funding:
  - NSF OCI 07-25070
  - NSF PRAC “The Computational Microscope”
  - NIH support: 9P41GM104601, 5R01GM098243-02
GPU Computing Publications
http://www.ks.uiuc.edu/Research/gpu/


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


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


NIH BTRC for Macromolecular Modeling and Bioinformatics
http://www.ks.uiuc.edu/ Beckman Institute,
U. Illinois at Urbana-Champaign
GPU Computing Publications
http://www.ks.uiuc.edu/Research/gpu/

- **Adapting a message-driven parallel application to GPU-accelerated clusters.**  

- **GPU acceleration of cutoff pair potentials for molecular modeling applications.**  

- **GPU computing.**  

- **Accelerating molecular modeling applications with graphics processors.**  

- **Continuous fluorescence microphotolysis and correlation spectroscopy.**  