ATPESC Track 4: Scalable Molecular Visualization and Analysis Tools in VMD

Mariano Spivak PhD
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/
Argonne Training Program on Exascale Computing (ATPESC)
9:15am-10:00am,
Monday August 8th, 2022
VMD – “Visual Molecular Dynamics”

- 125,000 active users worldwide
- Visualization and analysis of:
  - Molecular dynamics simulations
  - Lattice cell simulations
  - Quantum chemistry calculations
  - Cryo-EM densities, volumetric data
- User extensible scripting and plugins
- http://www.ks.uiuc.edu/Research/vmd/

Cell-Scale Modeling

MD Simulation

**Structure**

Parameterization

Refinement with **MDFF**

Preparation with **QwikMD**

Analysis

Parallel Analysis

Remote Visualization

MD/Cell Simulation

LM

NAMD

Amber

Gromacs
VMD Hands-On Tutorials

- [http://www.ks.uiuc.edu/Training/Tutorials/#vmd](http://www.ks.uiuc.edu/Training/Tutorials/#vmd)
  - Main VMD tutorial
  - QwikMD simulation preparation and analysis plugin
  - VMD images and movies tutorial
  - Structure check
  - VMD quantum chemistry visualization tutorial
  - Visualization and analysis of CPMD data with VMD
  - Parameterizing small molecules using ffTK
Goal: A Computational Microscope
Study the molecular machines in living cells

Ribosome: target for antibiotics
Poliovirus
VMD: Building A Next Generation Modeling Platform

- Provide tools for simulation preparation, visualization, and analysis
  - Reach cell-scale modeling w/ all-atom MD, coarse grained, Lattice Microbes
  - Improved performance, visual fidelity, exploit advanced technologies (GPUs, VR HMDs)
- Enable hybrid modeling and computational electron microscopy
  - Load, filter, process, interpret, visualize multi-modal structural information
- Connect key software tools to enable state-of-the-art simulations
  - Support new data types, file formats, software interfaces
- Openness, extensibility, and interoperability are VMD hallmarks
  - Reusable algorithms made available in NAMD, for other tools
QwikMD: Guided MD Simulation and Training

Smooths initial learning curve (non-expert users)

Speed up tedious simulation preparation tasks (expert users)

Reproducibility: detailed log of all steps

Interactive preparation, simulation, and analysis

## Selected VMD Plugins: Center Developed, and User Developed

### Analysis
- APBSRun
- CatDCD
- Contact Map
- GofRGUI
- HeatMapper
- ILSTools
- IRSpecGUI
- MultiSeq
- NAMD Energy
- NAMD Plot
- NetworkView
- NMWiz
- ParseFEP
- PBCTools
- PMEpot
- PropKa GUI
- RamaPlot
- RMSD Tool
- RMSD Trajectory Tool
- RMSD Visualizer Tool
- Salt Bridges
- Sequence Viewer
- Symmetry Tool
- Timeline
- TorsionPlot
- VolMap

### Modeling
- AutoIonize
- AutoPSF
- Chirality
- Cionize
- Cispeptide
- CGTools
- Dowser
- ftTK
- Inorganic Builder
- MDFG
- Membrane
- Merge Structs
- Molefacture
- Mutator
- Nanotube
- Psfgen
- RESPTool
- RNAView
- Solvate
- SSRestraints
- Topotools

### Visualization
- Clipping Plane Tool
- Clone Rep
- DemoMaster
- Dipole Watcher
- Intersurf
- Navigate
- NavFly
- MultiMolAnim
- Color Scale Bar
- Remote
- Palette Tool
- ViewChangeRender
- ViewMaster
- Virtual DNA Viewer
- VMD Movie Maker

### Simulation
- AlaScan
- AutoIMD
- IMDMenu
- NAMD GUI
- NAMD Server
- QMTool

### Collaboration
- Remote Control

### Data Import and Plotting
- Data Import
- Multiplot
- PDBTool
- MultiText

### Externally Hosted Plugins and Extensions
- Check sidechains
- MultiMSMS
- Interactive Essential Dynamics
- Mead Ionize
- Clustering Tool
- iTrajComp
- Swap RMSD
- Intervor
- SurfVol
- vmdICE

75 MolFile I/O Plugins: structure, trajectory, sequence, and density map

http://www.ks.uiuc.edu/Research/vmd/plugins/
VMD Visualization Concepts
VMD Approach to Visualization

- Molecular scene is composed of “graphical representations”
- Each representation encapsulates a group of selected atoms, a drawing style, coloring style, and other parameters
- Representations are independent of each other, can be toggled on/off easily, allowing molecular scenes to be built-up incrementally
- VMD atom selection language is shared with its analytical and scripting interfaces
Selection, Filtering

- Most viz tools allow interactive visual picking, menu-driven selections of structure components to display or operate on.
- VMD also extensively uses a text-based selection language (think google):
  - water within 10 of protein and z > 0
  - nucleic or protein or ions
  - segname BR
  - name “C.*”
  - Allows selection on user-defined data fields
  - *Promotes synergy between interactive and scripting interfaces, visualization and quantitative analysis tasks*
  - Works well with huge time-varying structures
Structure Visualization

Molecular representations provide different levels of abstraction, atomic detail vs. higher level organizational information

- Atoms, VdW spheres, bonds, ball-stick, …
- Coarse-grained “beads”
- Ribbons, secondary structure, “cartoon” reps, RNA/DNA
- Molecular surfaces
- Molecular orbitals (quantum chemistry)
Computed Properties

- Smoothing of thermal noise
- Secondary structure
- Hydrogen bonds, salt bridges
- Forces, energies, stress, strain
- Time averaging of electrostatic fields, occupancy maps
- Quality-of-fit cross correlation with cryo-EM density maps
- Normal modes, principal component analysis, essential dynamics
- Cluster simulation trajectory timesteps by structural similarity

Chemoreceptor trimer-of-dimers analysis with Bendix plugin in VMD
Display of Computed Properties on Structures

Per-residue solvent-accessible surface area of Ubiquitin

PME electrostatic potential contour for a helicase on a volumetric slice plane
CheA kinase PCA: first principal component porcupine plot
VMD Shading Comparison: EF-Tu

Outline Shader

“Goodsell” Shader

Glossy Shader

Ray Tracing: Ambient Occlusion, Shadowing
Trajectory Analysis and Visualization
Visualization of MD Trajectories

- Allow researchers study trajectories by analyzing force profiles, energies, structural changes, etc.
- **Visualization selections, graphics, structure properties, colors can all be recomputed for each trajectory timestep!**
Time-Averaged Volumetric Properties

- Compute density, distance, occupancy, potential maps for a frame or averaged over a trajectory
- Example: display binding sites for diffusively bound ions as probability density isosurfaces

trNA magnesium ion occupancy: VMD volmap plugin
Large System Analysis and Visualization
The diagram illustrates the number of atoms over time for various biological structures. The x-axis represents the year, ranging from 1990 to 2020. The y-axis shows the number of atoms, ranging from $10^4$ to $10^9$. Key structures include:

- **HIV capsid**: The HIV capsid, a viral capsid, is depicted and labeled with $(100\text{nm})^3$.
- **Protocell**: The protocell, another biological structure, is represented with $(200\text{nm})^3$.
- **ATP synthase**: This is a cellular enzyme involved in energy conversion.
- **Aquaporin**: An aquaporin, which facilitates water transport, is shown.
- **Lysozyme**: Lysozyme, an enzyme that breaks down bacterial cell walls, is also included.
- **STMV**: STMV is a plant virus with $(2\text{nm})^3$.

These structures are visualized with a logarithmic scale, highlighting the exponential growth of complexity over time in biological systems.
VMD Petascale Visualization and Analysis

- Combination of growing system sizes and timescales of simulation trajectories poses a major data size challenge for molecular visualization and analysis
- Parallel I/O rates up to **275 GB/sec** on 8192 Cray XE6 nodes – can read in **231 TB in 15 minutes**!
- Analyze/visualize large trajectories too large to transfer off-site:
  - User-defined parallel analysis operations, data types
  - Parallel rendering, movie making
- Supports GPU-accelerated compute nodes for both visualization and analysis tasks:
  - GPU accelerated trajectory analysis w/ CUDA
  - OpenGL and GPU ray tracing for visualization and movie rendering

NCSA Blue Waters Hybrid Cray XE6 / XK7
22,640 XE6 dual-Opteron CPU nodes
4,224 XK7 nodes w/ Telsa K20X GPUs

Parallel VMD currently available on: ORNL Summit and Titan, NCSA Blue Waters, IU Big Red II, CSCS Piz Daint, many similar systems
Petascale Molecular Dynamics I/O and Storage Challenges

- NAMD simulations can produce up to 10TB/day @ 1024 nodes (~20%) of ORNL Summit, more as optimizations raise NAMD performance further
- Petascale science campaigns require months of simulation runs
- Long-term storage of large-fractional petabytes impractical
- Historical “download output files for analysis and visualization” approach is a non-starter at this scale
- Demands visualization and analysis operate on the data in-place on the HPC system, whether post-hoc, in-transit, or in-situ
- Analyses must identify salient features of structure, dynamics, cull data that don’t contribute to biomolecular processes of interest
NAMD on Summit, May 2018

NAMD simulations can generate up to 10TB of output per day on 20% of Summit.
Proto-Cell Data Challenges

• 1B-atom proto-cell requires nodes with more than TB RAM to build complete model…
• 1B-atom proto-cell binary structure file: 63GB
• Trajectory frame atomic coordinates: 12GB, 1.2TB/ns of simulation (1 frame per 10ps)
• Routine modeling and visualization tasks are a big challenge at this scale
  – Models contain thousands of atomic-detail components that must work together in harmony
  – Exploit persistent memory technologies to enable “instant on” operation on massive cell-scale models – eliminate several minutes of startup during analysis/visualization of known structure
  – Sparse output of results at multiple timescales will help ameliorate visualization and analysis I/O
  – Data quantization, compression, APIs like ZFP
Running VMD on Distributed Memory Clouds, Clusters, and Supercomputers with MPI
Using VMD MPI Builds

- See “Running VMD on Supercomputers” section in the VMD tutorial

- Run one MPI rank (VMD process) per compute node
  - Each rank uses all CPU cores and all GPU accelerators they find
  - GPUs are shared between graphics/visualization and computing tasks

- Parallel VMD Scripting APIs:
  - “parallel” script subcommands, e.g., “parallel for ...”:
    - “nodename”, “noderank”, “nodecount”, “barrier”, “allgather”, “allreduce”, “for”
  - Available even in non-MPI builds, for single-node script compatibility

- Work scheduling:
  - Intra-node work scheduled via multithreading, CUDA, etc.
  - Inter-node distributed memory work scheduling of user analysis/viz scripts is provided by built-in load balancing implementation for VMD
Parallel MDFF Cross Correlation Analysis on Cray XK7

Rabbit Hemorrhagic Disease Virus (RHDV)

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traj. frames</td>
<td>10,000</td>
</tr>
<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 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>1</td>
<td>221s (31% I/O)</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>141s (46% I/O)</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>107s (64% I/O)</td>
</tr>
<tr>
<td>32</td>
<td>8</td>
<td>90s (76% I/O)</td>
</tr>
</tbody>
</table>

Performance at 32 nodes reaches ~48 FPS


64M atom HIV-1 capsid simulation rendered via EGL
High Fidelity Ray Tracing for Interactive and Cinematic Visualization
High Fidelity Ray Tracing

- Advanced rendering techniques save scientists time, produce images that are easier to interpret
- Ambient Occlusion, Depth of Field, high quality transparency, instancing, ....
- **Interactive RT** on laptops, desk, cloud, and **remote supercomputers**
- **Large-scale MPI parallel rendering:** in situ or post hoc visualization tasks
- Stereoscopic panorama and full-dome projections
- **Omnidirectional VR:** YouTube, HMDs
- **Built-in ray tracing engines:**
  - **Tachyon:** cross-platform RT
  - **NVIDIA OptiX:** GPU RTX-accelerated
  - **Intel OSPRay:** CPU x86-optimized

VMD/OptiX all-atom Chromatophore
Lighting Comparison, STMV Capsid

Two lights, no shadows

Ambient occlusion + two lights, 144 AO rays/hit
Geometrically Complex Scenes

Ray tracing techniques well matched to molecular viz. needs:

- Curved geometry, e.g. spheres, cylinders, toroidal patches, easily supported
- Greatly reduced memory footprint vs. polygonalization
- Runtime scales only moderately with increasing geometric complexity
- Occlusion culling is “free”, RT acceleration algorithms do this and much more
Goal: Intuitive interactive viz. in crowded molecular complexes

Results from 64 M atom, 1 μs sim!

Close-up view of chloride ions permeating through HIV-1 capsid hexameric centers
Satellite Tobacco Mosaic Virus

Biomedical Technology Research Center for Macromolecular Modeling and Bioinformatics
Beckman Institute, University of Illinois at Urbana-Champaign - www.ks.uiuc.edu
VMD w/ OptiX RTX Ray Tracing

- Interactive RT on laptops, desktops, and cloud
- Large-scale parallel rendering: in situ or post hoc visualization
- Remote ray tracing with NvPipe video streaming
- Stereoscopic panoramic and full-dome projections
- Omnidirectional VR for YouTube, VR HMDs
- VMD+OptiX NGC container: https://ngc.nvidia.com/registry/
- GPU memory sharing via NVLink
- In-progress:
  Denoising: faster turnaround w/ AO, DoF, etc

VMD/OptiX RTX Acceleration
VMD OptiX RT performance on Quadro RTX 6000

<table>
<thead>
<tr>
<th></th>
<th>Chromatophore @ 4Kx4K</th>
<th>Chrom Cell, 512x DoF @ 1080p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadro GV100</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2x Quadro GV100</td>
<td>1.97</td>
<td>1.95</td>
</tr>
<tr>
<td>Quadro RTX 6000</td>
<td>8.02</td>
<td>8.18</td>
</tr>
</tbody>
</table>
VMD Application Examples and Implementation Details
AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics.
https://dx.doi.org/10.1177/10943420211006452
Visualizing High-Dimensional Data: t-SNE plot of SARS-CoV2 Simulation Campaign

https://dx.doi.org/10.1177/10943420211006452
The Coronavirus Unveiled

Latching on to an ACE2 receptor, in yellow, allows the coronavirus to enter human cells. 
Cori Graf, Hamed Khatay, UC San Diego.
VMD Cinematic Molecular Visualization and Rendering: “Birth of Planet Earth” Fulldome Show
https://www.youtube.com/watch?v=NTgAok6n7I4

VMD RTX Performance Gains for “BoPE” Content:
- BoPE production used Quadro M6000 (Maxwell) GPUs
- BoPE w/ OptiX 6.5 on Quadro RTX 6000 up to 15x faster!

Multiscale modeling and cinematic visualization of photosynthetic energy conversion processes from electronic to cell scales.
Omnidirectional Stereoscopic Ray Tracing

- Ray trace 360° images and movies for Desk and VR HMDs: Oculus, Vive, Cardboard
- Stereo spheremaps or cubemaps allow very high-frame-rate interactive OpenGL display
- AO lighting, depth of field, shadows, transparency, curved geometry, …
- Summit 6x Tesla V100 GPU nodes:
  - Render many omni-stereo viewpoints, no acceleration structure rebuilds, tens of frames/sec per-node!
  - OptiX multi-GPU rendering, NVLink compositing and data distribution, etc…
  - Future: AI for warping between views

VMD Examples from In-Progress ANARI Renderers

https://www.khronos.org/anari

Tachyon Ray Tracer

OSPRay Path Tracer

OptiX Path Tracer
VMD+Folding@Home w/ NVIDIA Omniverse+Maya

Density Map Segmentation

VMD GPU-accelerated density map segmentation of GroEL

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 from an all-atom structure.

MDFF Cross Correlation Timeline

Regions with poor fit

Regions with good fit
VMD Tesla V100 Cross Correlation Performance

Rabbit Hemorrhagic Disease Virus: 702K atoms, 6.5Å resolution
Volta GPU architecture almost 2x faster than previous gen Pascal:

<table>
<thead>
<tr>
<th>Application and 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>
<tr>
<td><strong>VMD-CUDA IBM Power9 “Newell” + 1x Tesla V100</strong></td>
<td><strong>0.049s, 323x 9.3x</strong></td>
</tr>
</tbody>
</table>


Clustering Analysis of Molecular Dynamics Trajectories: Requires I/O+Memory for All-Pairs of Trajectory Frames

Acknowledgements

• Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign
• CUDA Center of Excellence, University of Illinois at Urbana-Champaign
• NVIDIA CUDA and OptiX teams
• Intel OSPRay team
• Funding:
  – NIH support: P41-GM104601
  – DOE INCITE, ORNL Summit
  – NSF Blue Waters:
    NSF OCI 07-25070, PRAC “The Computational Microscope”, ACI-1238993, ACI-1440026
“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