End-to-End Study of Parallel Volume Rendering on the IBM Blue Gene/P

Tom Peterka\textsuperscript{1}, Hongfeng Yu\textsuperscript{2}, Robert Ross\textsuperscript{1}, Kwan-Liu Ma\textsuperscript{2}, Rob Latham\textsuperscript{1}

Volume rendering of x-velocity in time-step 1530 of a hydrodynamics simulation of a core-collapse supernova.

\textsuperscript{1} Argonne National Laboratory
\textsuperscript{2} University of California at Davis

Tom Peterka
tpeterka@mcs.anl.gov
Mathematics and Computer Science Division
A Growing Rift

We are computing more data, faster than we can manage.

Total data of selected 2008 INCITE awards as of June 2008

<table>
<thead>
<tr>
<th>Domain</th>
<th>Data size (TB)</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astrophysics</td>
<td>375</td>
<td>Lamb</td>
</tr>
<tr>
<td>Climate</td>
<td>355</td>
<td>Washington</td>
</tr>
<tr>
<td>Materials</td>
<td>105</td>
<td>Wolverton</td>
</tr>
<tr>
<td>Fusion</td>
<td>54</td>
<td>Klasky</td>
</tr>
</tbody>
</table>

Normalized Storage / Compute Metrics

<table>
<thead>
<tr>
<th>Machine</th>
<th>Storage B/W (GB/s)</th>
<th>Storage Size (PB)</th>
<th>FLOPS (Pflop/s)</th>
<th>Norm. Storage B/W Byte/s/flop/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLNL BG/L</td>
<td>43</td>
<td>2</td>
<td>0.6</td>
<td>$O(10^{-4})$</td>
</tr>
<tr>
<td>Jaguar XT4</td>
<td>42</td>
<td>0.6</td>
<td>0.3</td>
<td>$O(10^{-4})$</td>
</tr>
<tr>
<td>Intrepid BG/P</td>
<td>50</td>
<td>5</td>
<td>0.6</td>
<td>$O(10^{-4})$</td>
</tr>
<tr>
<td>Roadrunner</td>
<td>50</td>
<td>5</td>
<td>1.0</td>
<td>$O(10^{-5})$</td>
</tr>
<tr>
<td>Jaguar XT5</td>
<td>42</td>
<td>5</td>
<td>1.4</td>
<td>$O(10^{-5})$</td>
</tr>
</tbody>
</table>

DOE science applications generate approximately .03 bytes perflop. Ref. Murphy et al. ICS’05

Percent Saved of Computed Data

<table>
<thead>
<tr>
<th>Code</th>
<th>Domain</th>
<th>% Saved</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLASH</td>
<td>Astrophysics</td>
<td>10</td>
<td>Ricker</td>
</tr>
<tr>
<td>Nek5000</td>
<td>CFD</td>
<td>1</td>
<td>Fischer</td>
</tr>
<tr>
<td>CCSM</td>
<td>Climate</td>
<td>1</td>
<td>Jacob</td>
</tr>
<tr>
<td>GCRM</td>
<td>Climate</td>
<td>10</td>
<td>Cram</td>
</tr>
<tr>
<td>S3D</td>
<td>Combustion</td>
<td>1-5</td>
<td>Bennett</td>
</tr>
</tbody>
</table>

Ref: CScADS Scientific Data Analysis & Visualization Workshop '09
The current workflow will not scale indefinitely. "Models that can currently be run on typical supercomputing platforms produce data in amounts that make storage expensive, movement cumbersome, visualization difficult, and detailed analysis impossible. The result is a significantly reduced scientific return from the nation's largest computational efforts." -Mark Rast, Laboratory for Atmospheric and Space Physics, University of Colorado

One solution: Large scale parallel visualization on HPC machines

The increasing demands for analysis and visualization can be met by performing more analysis and visualization tasks directly on supercomputers traditionally reserved for simulation.

Potential benefits: Increased performance, reduced cost, tighter integration of analysis and visualization in computational science.
Applications
The science behind the computer science

Volume rendering of shock wave formation in core-collapse supernova dataset, courtesy of John Blondin, NCSU. Structured grid of $1120^3$ data elements, 5 variables per cell.

Pressure at time-step 1530

Angular momentum at time-step 1492

Angular momentum at time-step 1403

Entropy over 100 time-steps

Entropy at time-step 1518
Other Optimizations

Our related work

Parallel structure for volume rendering algorithm consists of 3 stages performed in parallel

Parallel pipelining and I/O subsetting

Assessing Improvements to the Parallel Volume Rendering Pipeline at Large Scale. SC08 Ultrascale Visualization Workshop.

Sort-last parallel rendering requires compositing resulting images into one final image.

A Configurable Algorithm for Parallel Image-Compositing Applications. SC09.

Stereo parallel volume rendering

Display of Large-Scale Scientific Visualization. SPIE’09
Hybrid of Systems and Visualization Research

Sometimes we are systems people,

3D torus interconnect offers high bandwidth and low latency.

PVFS-2 parallel file system provides 50 GB/s peak aggregate b/w and 5 PB total capacity.

The Blue Gene/P features a highly scalable compute architecture composed of 160,000 compute cores. Peak performance is 557 TF.

And other times we are applications people.

Domain Decomposition: Grid topology, decomp. strategy, neighbor cells, load balance, static / dynamic distr.

Scalability: Strong, weak scaling, max. effective number of processes, efficiency, isoefficiency

Performance: Overall time to completion, component time, time distribution

Data Movement: Nature of algorithm, communication signature, storage patterns

ICPP 2009 Conference September 25, 2009 Tom Peterka tpeterka@mcs.anl.gov
Performance
Total and component time

Total frame time and individual component times. Raw data format, 1120^3, image size 1600^2, original and improved image compositing.

The relative percentage of time in the stages of volume rendering as a function of system size. Large visualization is primarily dominated by I/O and secondarily by communication.
Performance

Large-scale results

Volume Rendering End-to-End Performance

<table>
<thead>
<tr>
<th>Grid Size</th>
<th>Time-step size (GB)</th>
<th>Image size (px)</th>
<th># Procs</th>
<th>Total time (s)</th>
<th>% I/O</th>
<th>Read B/W (GB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2240^3</td>
<td>42</td>
<td>2048^3</td>
<td>8K</td>
<td>51</td>
<td>96</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16K</td>
<td>43</td>
<td>97</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32K</td>
<td>35</td>
<td>96</td>
<td>1.3</td>
</tr>
<tr>
<td>4480^3</td>
<td>335</td>
<td>4096^3</td>
<td>8K</td>
<td>316</td>
<td>96</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16K</td>
<td>272</td>
<td>97</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32K</td>
<td>220</td>
<td>96</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Volume rendering performance at large size is dominated by I/O. While overall performance is scalable, I/O bandwidth is far below peak.

Scalability over a variety of data, image, and system sizes. A number of performance points exist for each data size.
Parallel I/O, system- and application-level optimizations can produce drastic speedups.

The organization of variables within a netCDF file. Record variables are stored in interleaved 2D format.

Changing data file layout can drastically improve I/O performance. Top, different layouts produce improved file access patterns. Bottom, benchmarks confirm improved performance.
Recap

Lessons learned and the road ahead

Successes

- Demonstrated scaling on large data and images
- Improved compositing
- Improved and benchmarked I/O

Ongoing

- Other algorithms and grid topologies
- In situ
- Adoption into tools and libraries

Take-away

- HPC has appropriate resources for visualization: massive parallelism, storage, and interconnect capability.

- Visualization algorithms can be developed that scale with the machine and problem size.
Further Reading

References


Peterka, T., Ross, R., Yu, H., Ma, K.-L.: Assessing Improvements to the Parallel Volume Rendering Pipeline at Large Scale. SC08 Ultrascale Visualization Workshop, Austin TX, November 2008.


End-to-End Study of Parallel Volume Rendering on the IBM Blue Gene/P

Acknowledgments:
John Blondin, Tony Mezzacappa
Argonne and Oak Ridge Leadership
Computing Facilities
US DOE SciDAC UltraVis Institute

Tom Peterka
tpeterka@mcs.anl.gov
Mathematics and Computer Science Division