Argonne NATIONAL LABORATORY

... for a brighter future

UltraVis
SciDAC Institute for Ultrascale Visualization
www.ultravis.org

U.S. Department of Energy

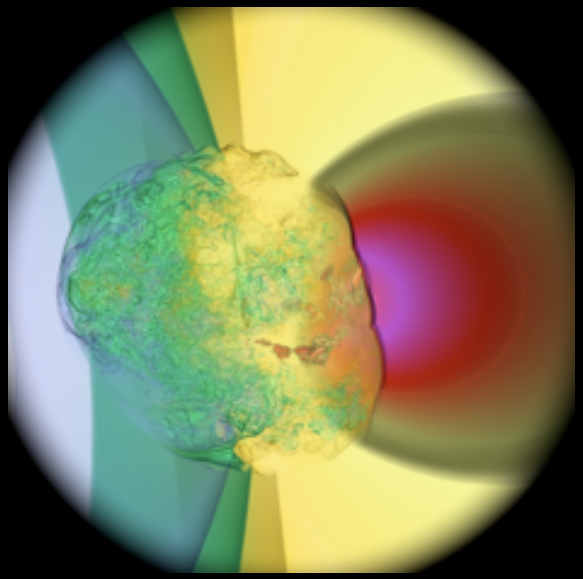
UChicago ▶ Argonne LLC

Office of Science
U.S. DEPARTMENT OF ENERGY

A U.S. Department of Energy laboratory managed by UChicago Argonne, LLC

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

Tom Peterka[1], Hongfeng Yu[2], Robert Ross[1], Kwan-Liu Ma[2], Rob Latham[1]



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

[1] Argonne National Laboratory
[2] University of California at Davis

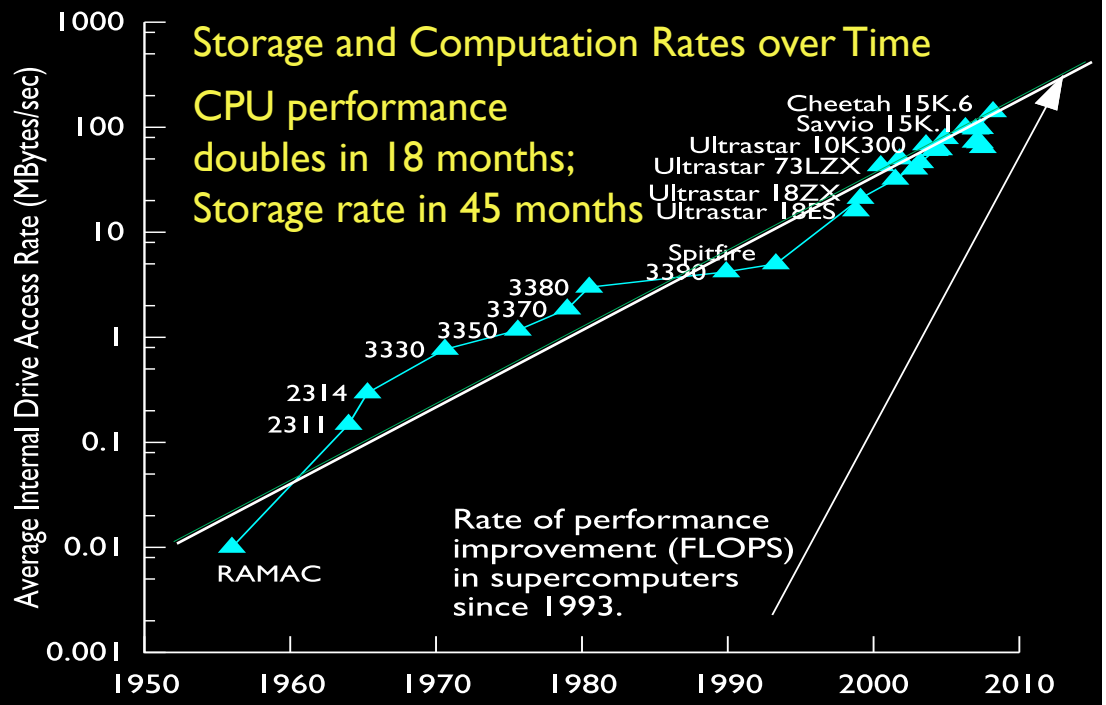
Tom Peterka
tpeterka@mcs.anl.gov
Mathematics and Computer Science Division

A Growing Rift

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

Total data of selected 2008 INCITE awards as of June 2008

Domain	Data size (TB)	PI
Astrophysics	375	Lamb
Climate	355	Washington
Materials	105	Wolverton
Fusion	54	Klasky



Normalized Storage / Compute Metrics

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

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

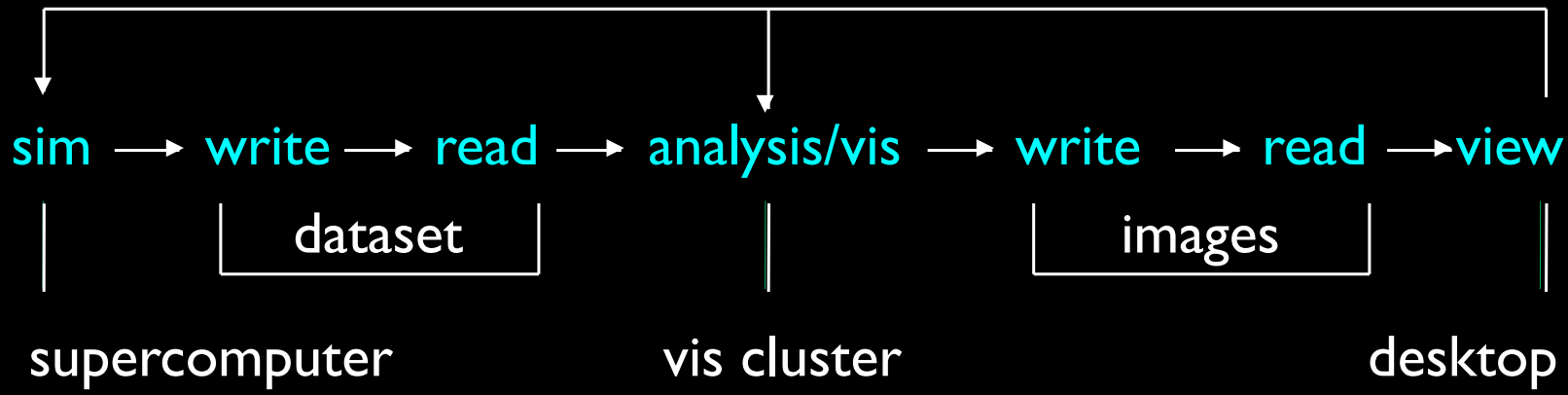
Percent Saved of Computed Data

Code	Domain	% Saved	PI
FLASH	Astrophysics	10	Ricker
Nek5000	CFD	1	Fischer
CCSM	Climate	1	Jacob
GCRM	Climate	10	Cram
S3D	Combustion	1-5	Bennett

Ref: CScADS Scientific Data Analysis & Visualization Workshop '09

Effect on Analysis and Visualization

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

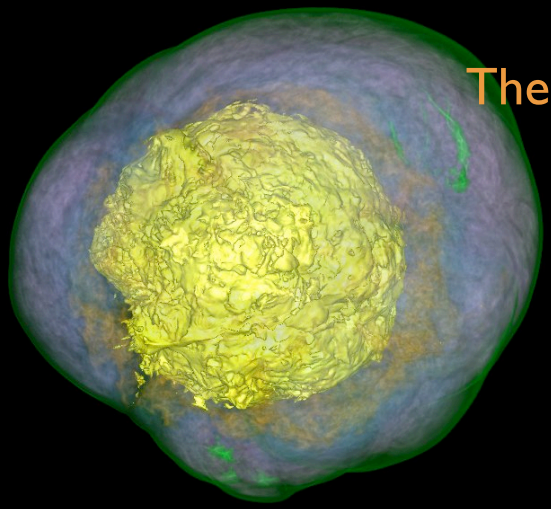
One solution: Large scale parallel visualization on HPC machines

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

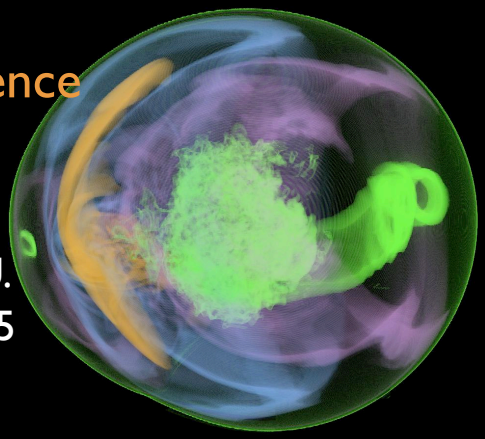
Potential benefits: Increased performance, reduced cost, tighter integration of analysis and visualization in computational science.

Applications

The science behind the computer science

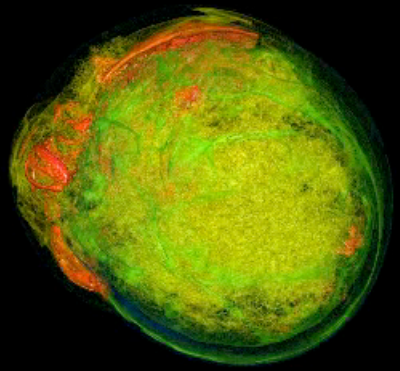


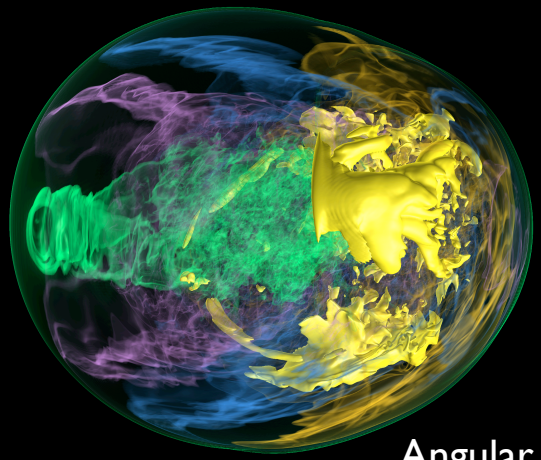
Volume rendering of shock wave formation in core-collapse supernova dataset, courtesy of John Blondin, NCSU. Structured grid of 1120^3 data elements, 5 variables per cell.



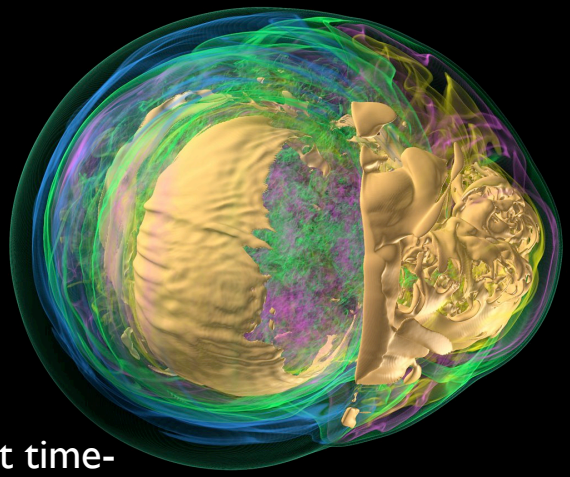
Angular momentum at time-step 1403

Pressure at time-step 1530





Entropy over 100 time-steps



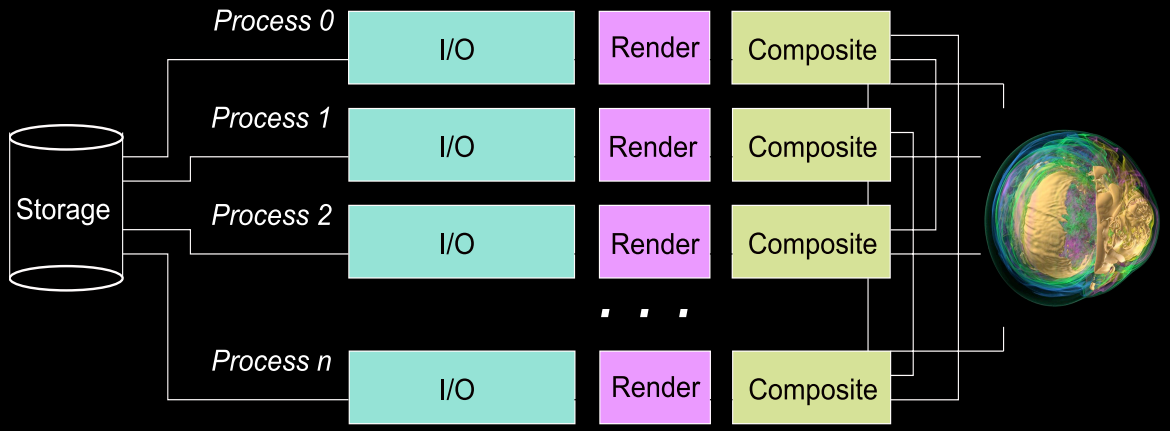
Angular momentum at time-step 1492

Entropy at time-step 1518

Other Optimizations

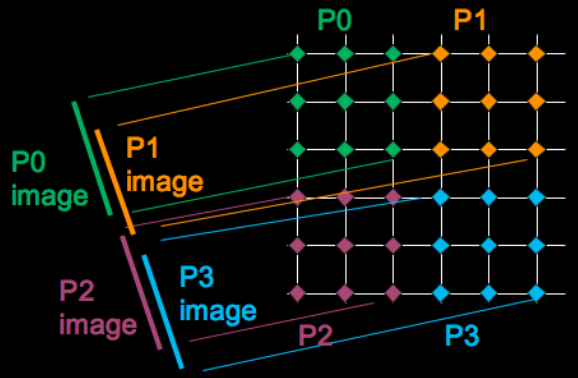
Our related work

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



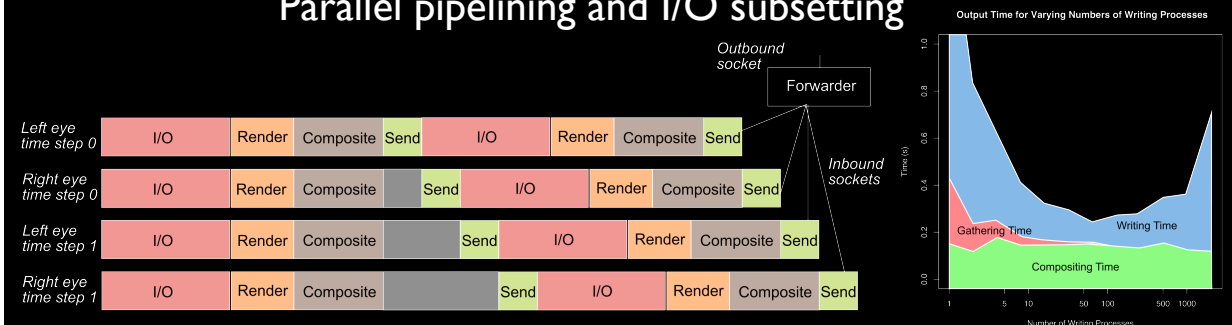
Parallel Volume Rendering on the IBM Blue Gene/P. EGPGV'08.

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



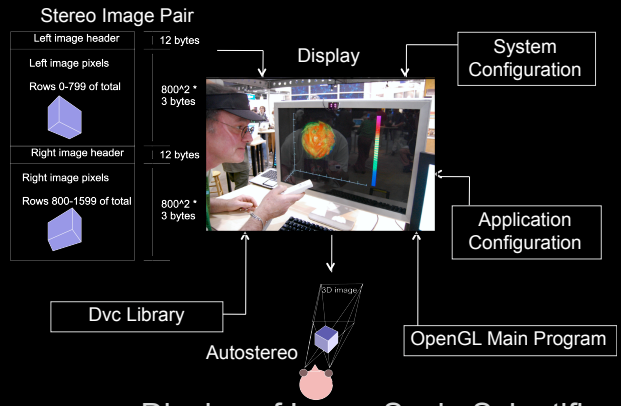
A Configurable Algorithm for Parallel Image-Compositing Applications. SC09.

Parallel pipelining and I/O subsetting



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

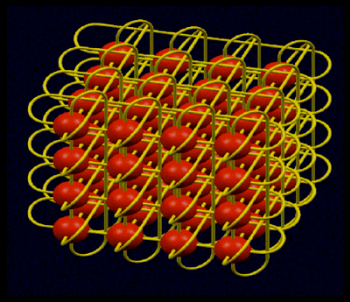
Stereo parallel volume rendering



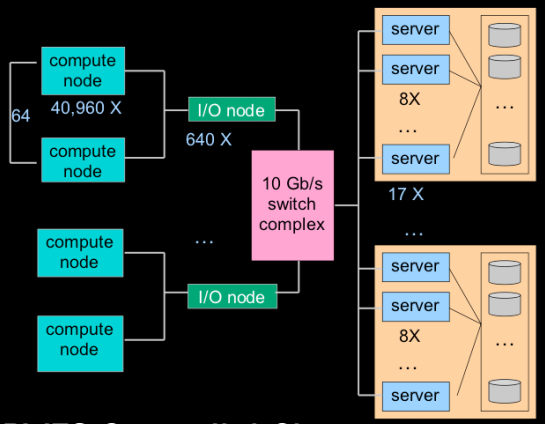
Display of Large-Scale Scientific Visualization. SPIE'09

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.

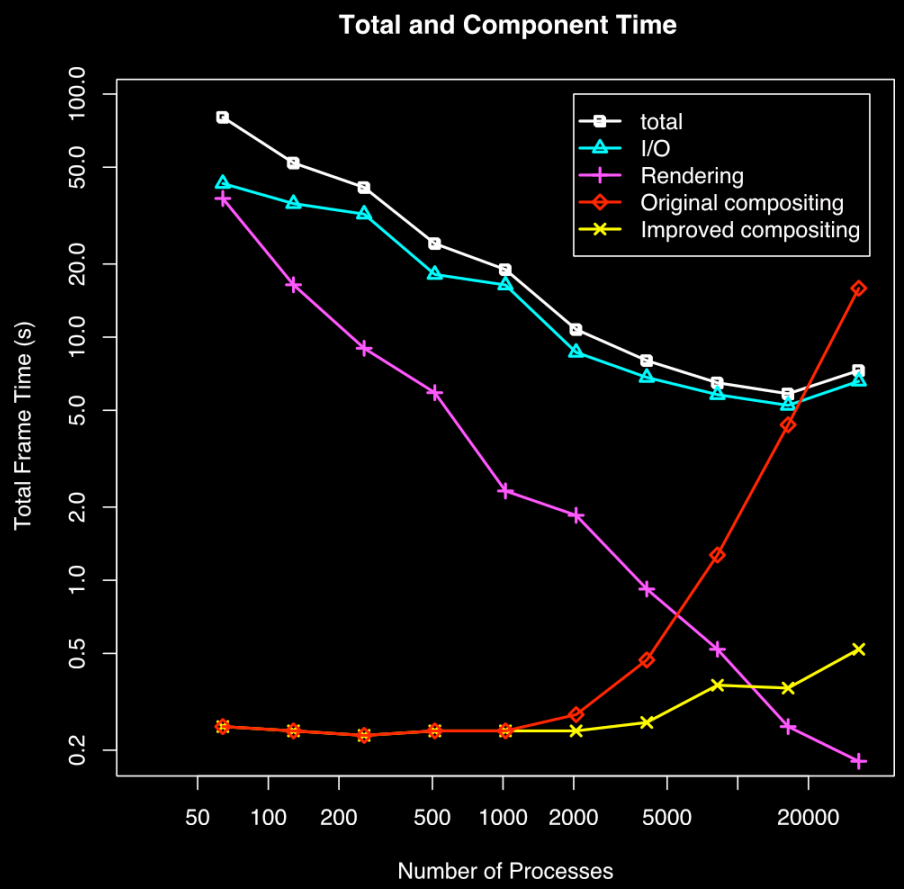
Scalability: Strong, weak scaling, max. effective number of processes, efficiency, isoefficiency

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

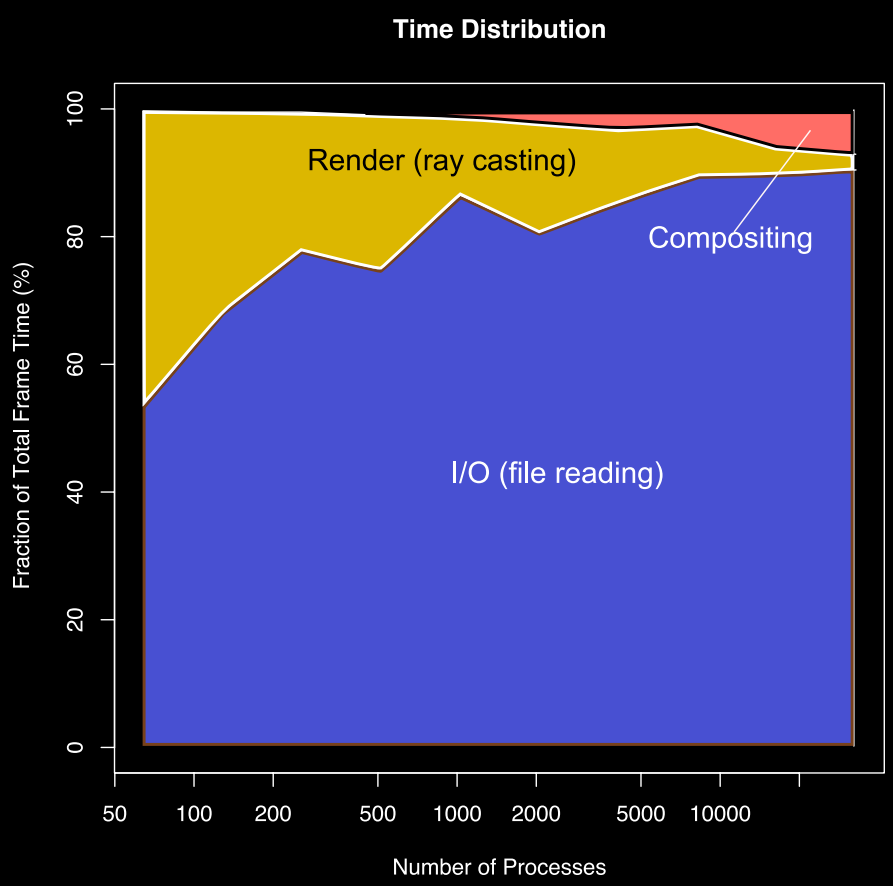
Data Movement: Nature of algorithm, communication signature, storage patterns

Performance

Total and component time



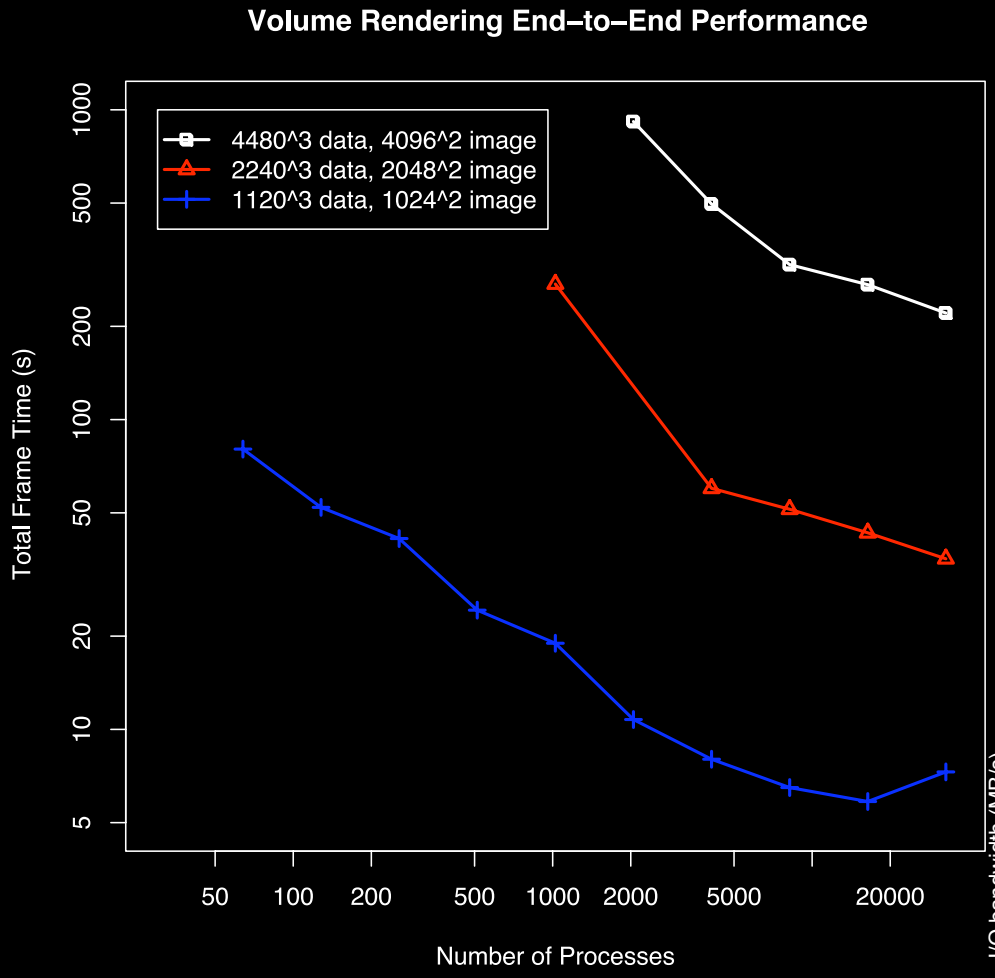
Total frame time and individual component times. Raw data format, 1120^3, image size 1600^2, original and improved image compositing.



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

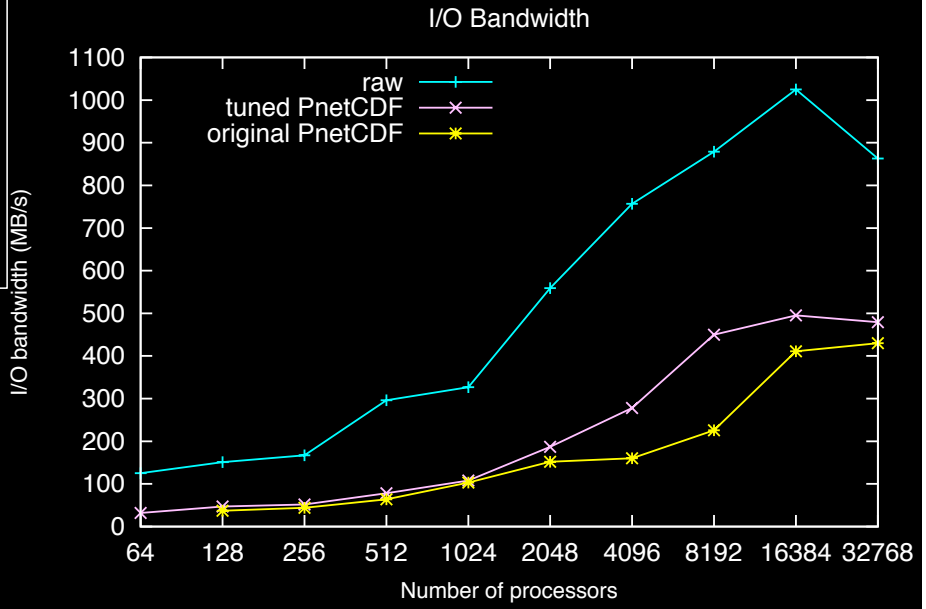
Large-scale results



Scalability over a variety of data, image, and system sizes. A number of performance points exist for each data size.

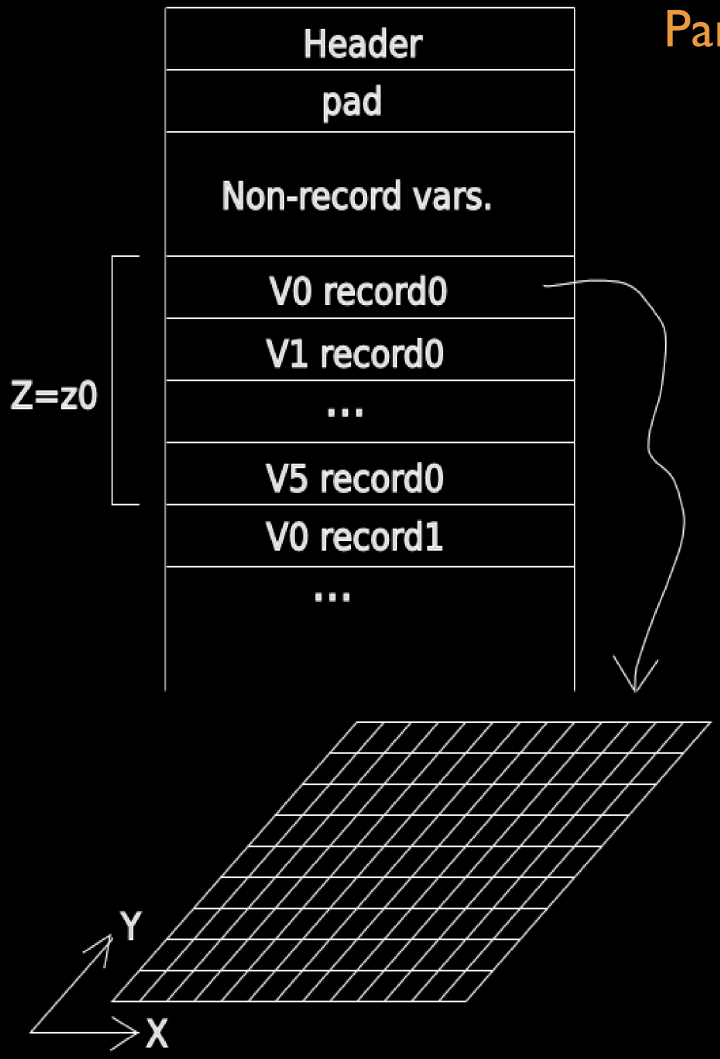
Grid Size	Time-step size (GB)	Image size (px)	# Procs	Tot. time (s)	% I/O	Read B/W (GB/s)
2240^3	42	2048^3	8K	51	96	0.9
			16K	43	97	1.0
			32K	35	96	1.3
4480^3	335	4096^3	8K	316	96	1.1
			16K	272	97	1.3
			32K	220	96	1.6

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

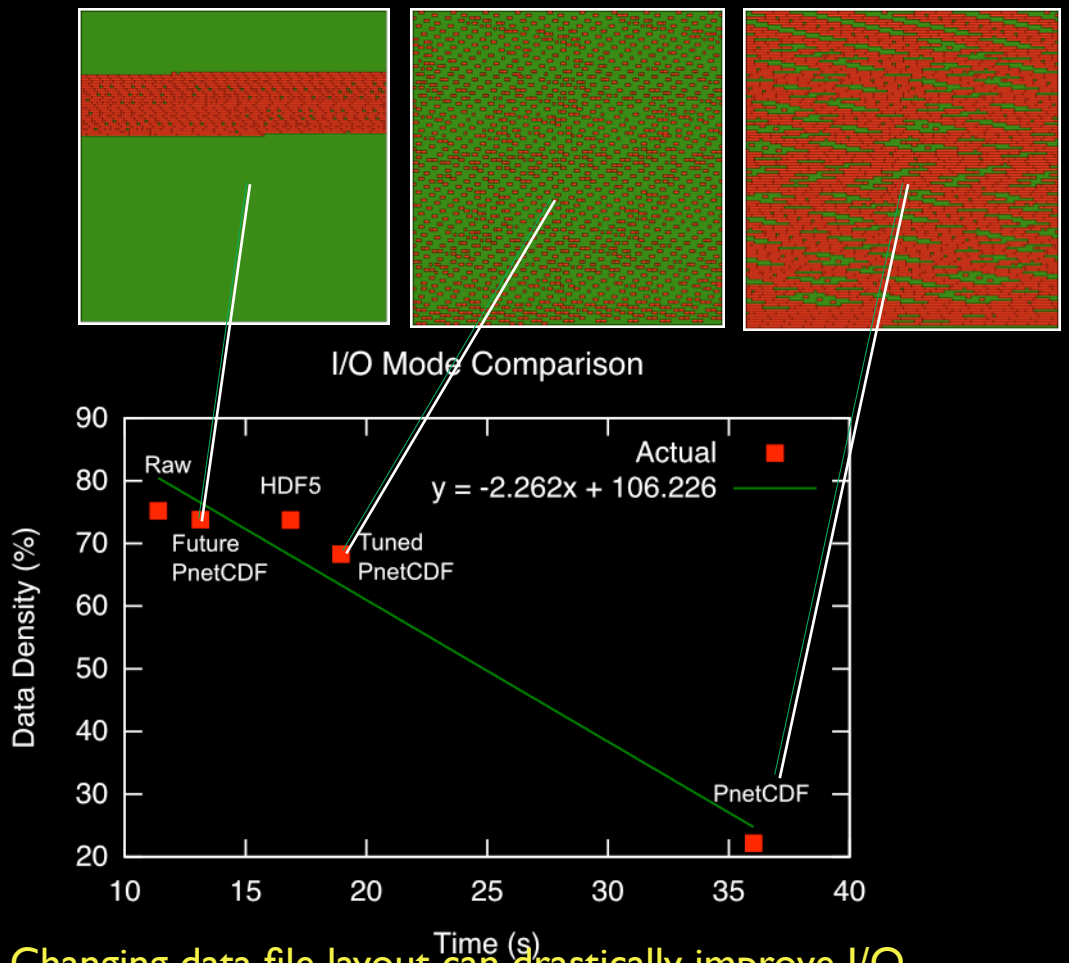


I/O Matters

Parallel I/O, system- and application-level optimizations can produce drastic speedups.



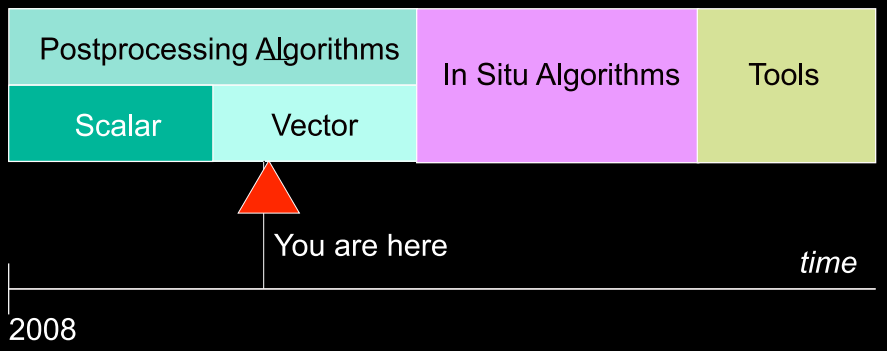
The organization of variables within a netCDF file. Record variables are stored in interleaved 2D format.



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., Goodell, D., Ross, R., Shen, H.-W., Thakur, R.: A Configurable Algorithm for Parallel Image -Compositing Applications. Proceedings of SC09, Portland OR, November 2009.

Peterka, T., Ross, R. B., Shen, H.-W., Ma, K.-L., Kendall, W., Yu, H.: Parallel Visualization on Leadership Computing Resources. Journal of Physics: Conference Series SciDAC 2009, June 2009.

Peterka, T., Ross, R., Yu, H., Ma, K.-L., and Girado, Javier: Autostereoscopic Display of Large-Scale Scientific Visualization. Proceedings of IS&T / SPIE SD&A XX Conference, San Jose CA, January 2009.

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.

Ross, R. B., Peterka, T., Shen, H.-W., Hong, Y., Ma, K.-L., Yu, H., Moreland, K.: Parallel I/O and Visualization at Extreme Scale. Journal of Physics: Conference Series SciDAC 2008, July 2008.

Peterka, T., Yu, H., Ross, R., Ma, K.-L.: Parallel Volume Rendering on the IBM Blue Gene/P. Proceedings of Eurographics Symposium on Parallel Graphics and Visualization 2008 (EGPGV'08) Crete, Greece, April 2008.

Argonne NATIONAL LABORATORY

... for a brighter future

UltraVis SciDAC Institute for Ultrascale Visualization

www.ultravis.org

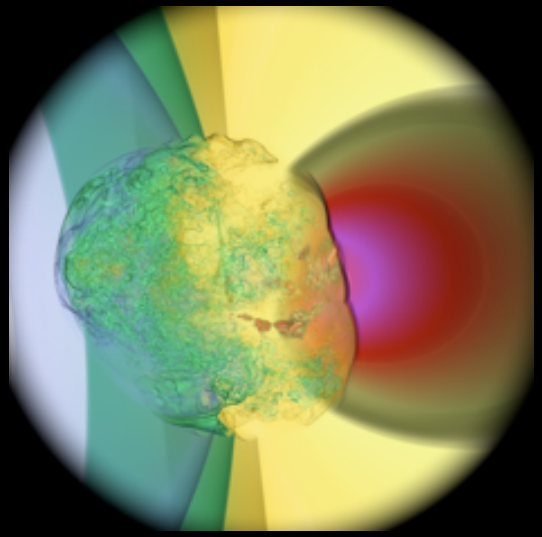
U.S. Department of Energy

UChicago ▶ Argonne LLC

Office of Science U.S. DEPARTMENT OF ENERGY

A U.S. Department of Energy laboratory managed by UChicago Argonne, LLC

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