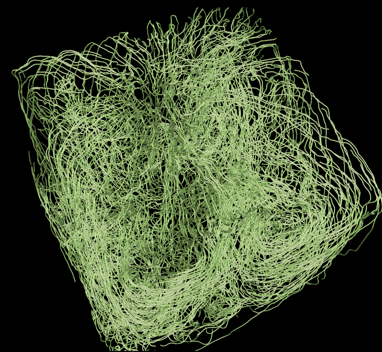
# DIY Applications

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

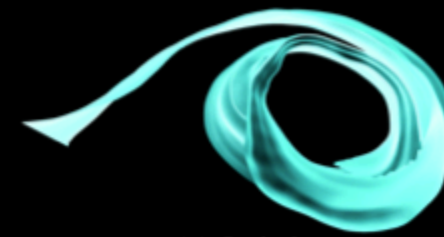
# Abstractions Matter: Blocks, not Tasks

- Block = unit of decomposition
- Block size, shape can be configured
  - From coarse to fine
  - Regular, adaptive, KD-tree
- Block placement is flexible, dynamic
  - Blocks per task
  - Tasks per block
  - Memory / storage hierarchy
- Data is first-class citizen
  - Separate operations per block
  - Thread safety

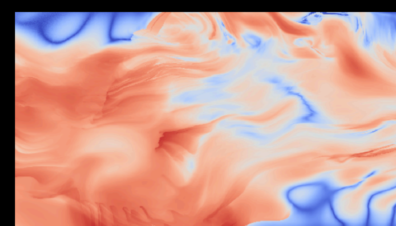
*Parallel data analysis consists of decomposing a problem into blocks, operating on them, and communicating between them.*



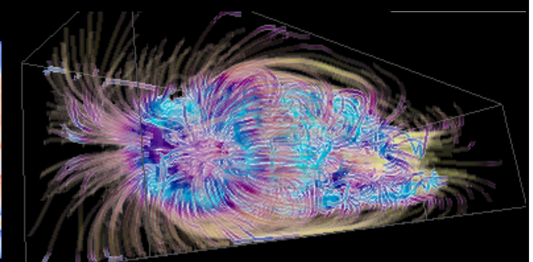
Streamlines and pathlines



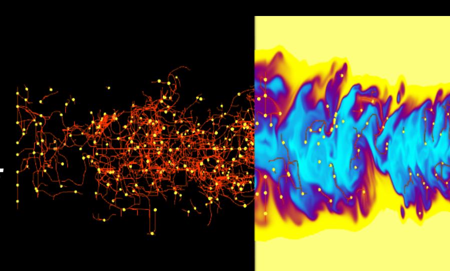
Stream surfaces



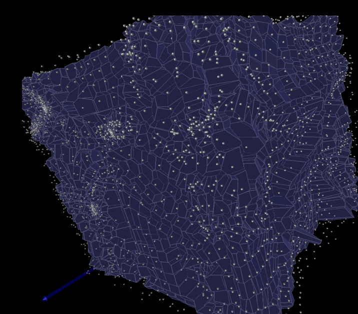
FTLE



Information entropy

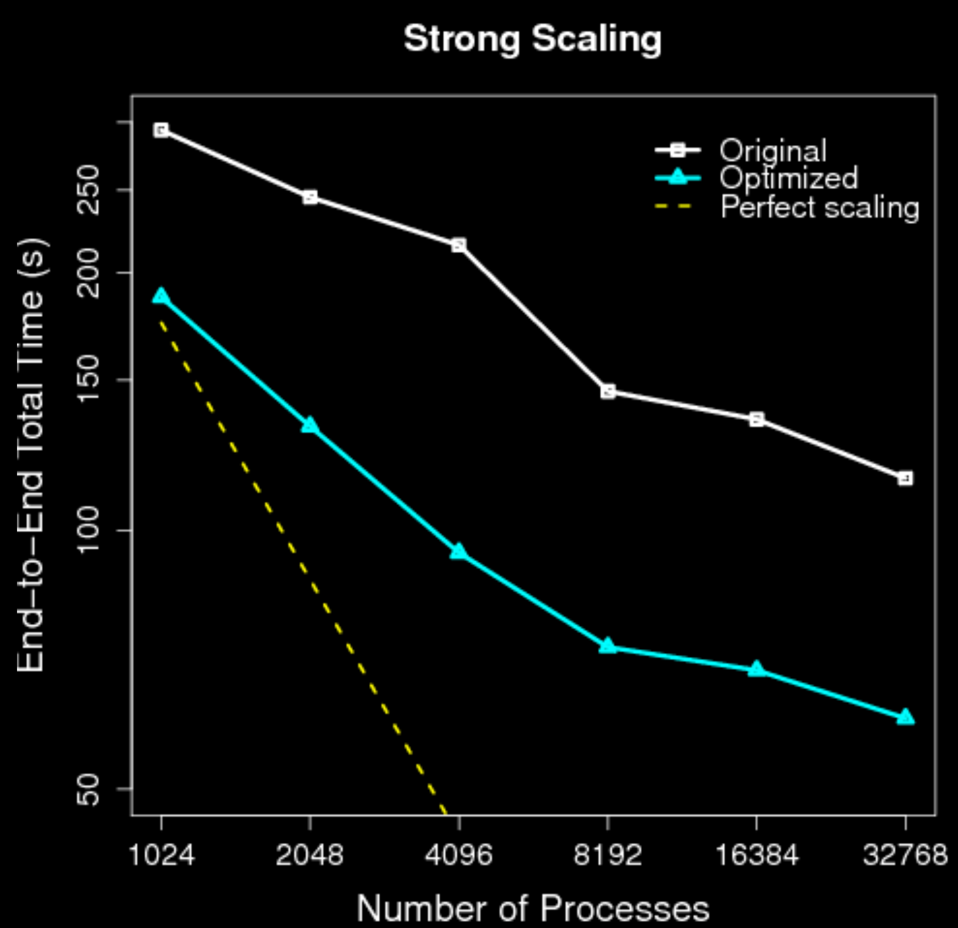


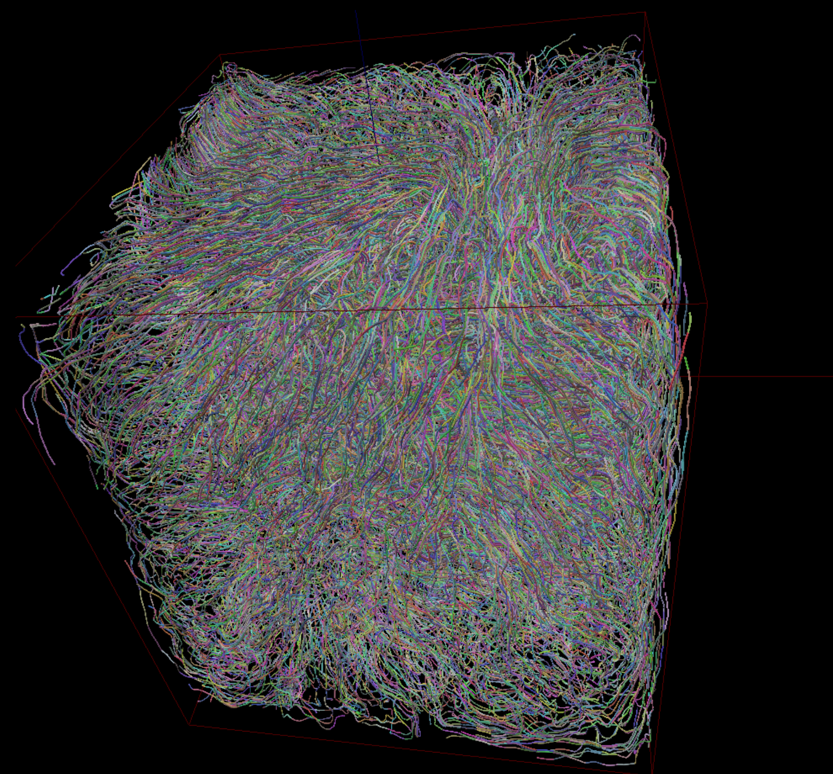
Morse-Smale complex



Voronoi Tessellation

# Particle Tracing Streamlines and Pathlines



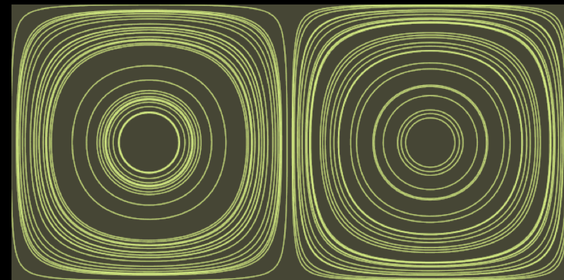



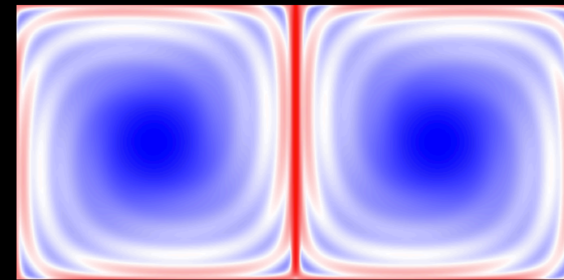
Particle tracing of ¼ million particles in a $2048^3$ thermal hydraulics dataset results in strong scaling to 32K processes and an overall improvement of 2X over earlier algorithms

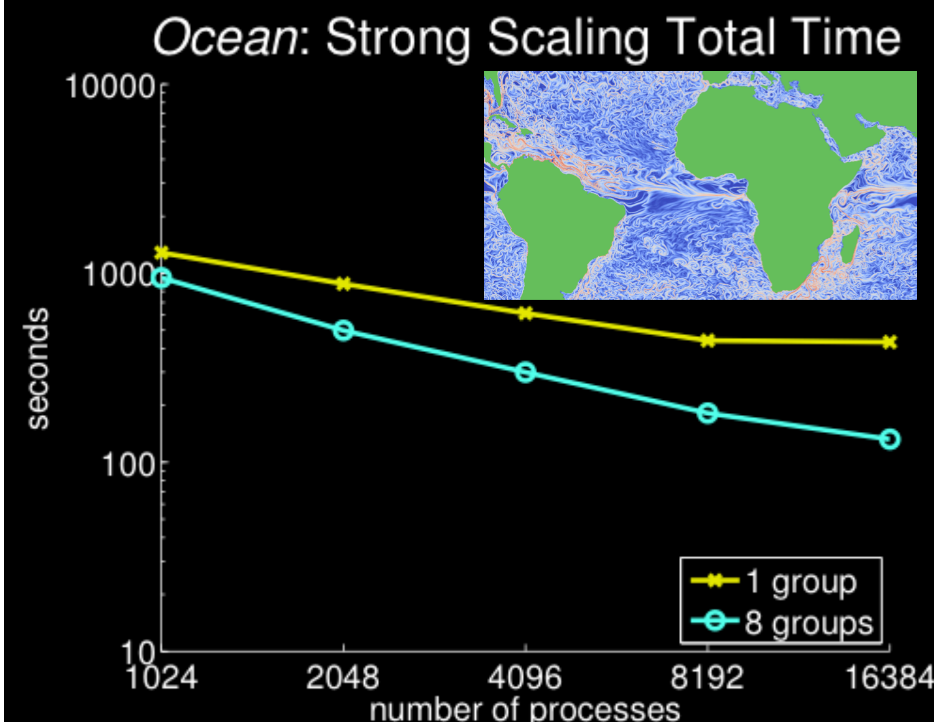
Peterka et al., A Study of Parallel Particle Tracing for Steady-State and Time-Varying Flow Fields. IPDPS '11.

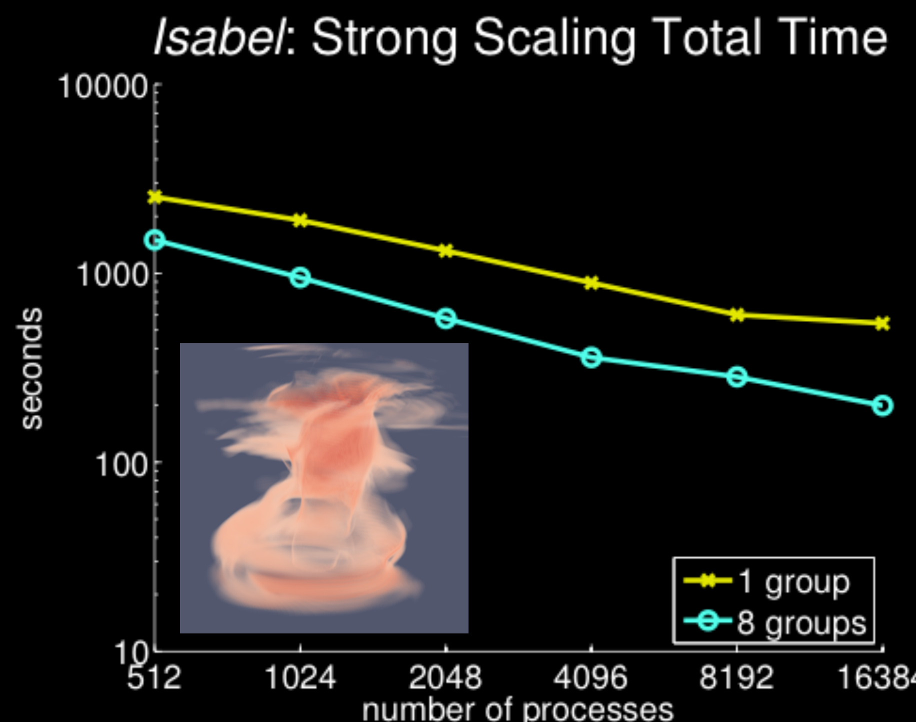
# Lagrangian Coherent Structures from FTLE

Courtesy Boonthanome Nouanesengsy







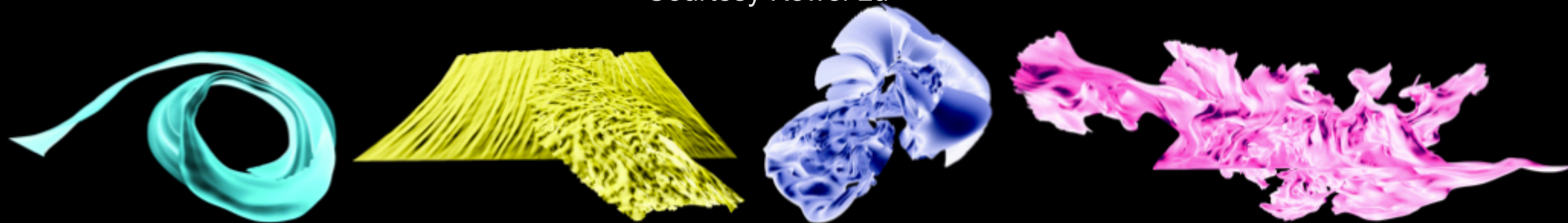




Left: Particle tracing of 288 million particles over 36 time steps in a 3600x2400x40 eddy resolving dataset. Right: 131 million particles over 48 time steps in a 500x500x100 simulation of Hurricane Isabel. Time includes I/O.

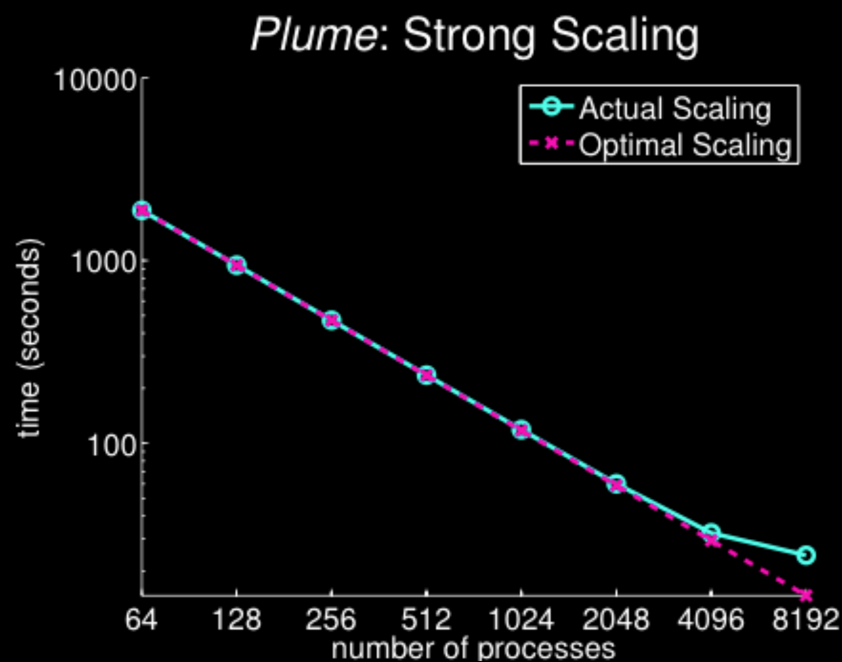
Nouanesengsy et al., Parallel Particle Advection and FTLE Computation for Time-Varying Flow Fields, SC12,

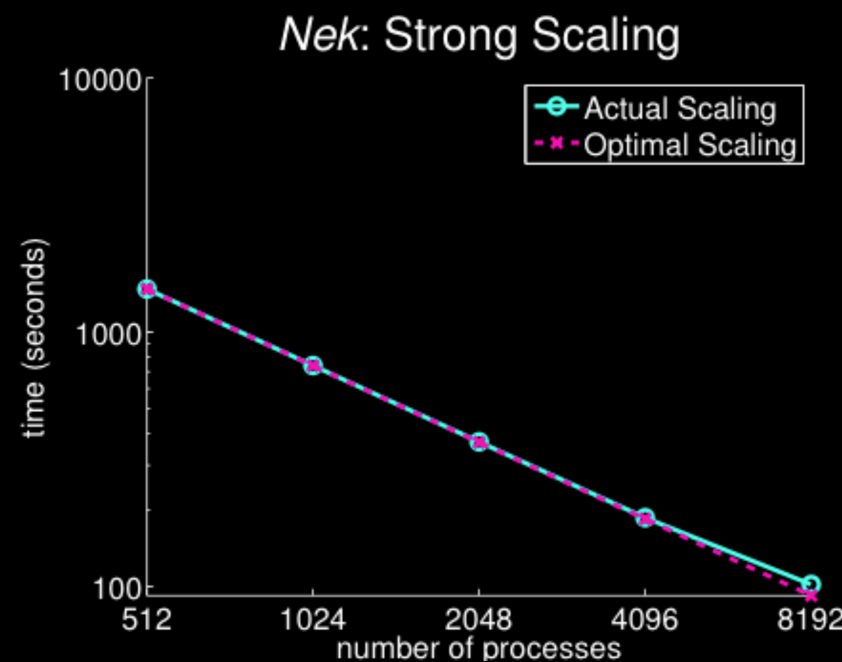
# Stream Surfaces

Courtesy Kewei Lu



(a) Isabel (b) MJO (c) Plume (d) Nek



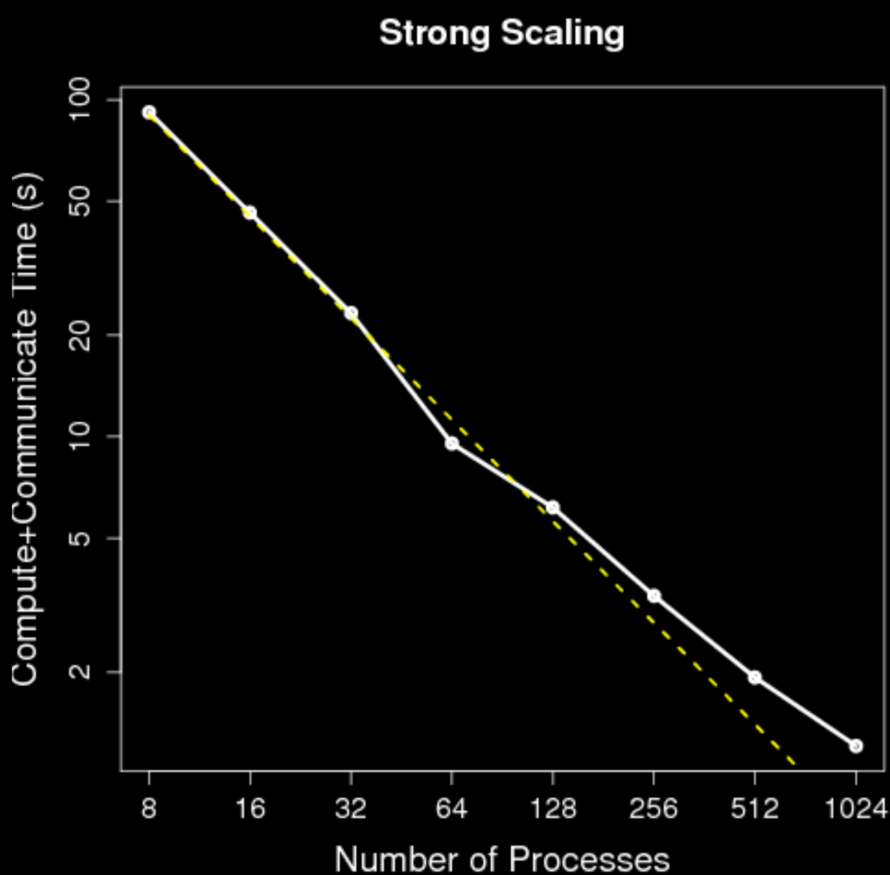



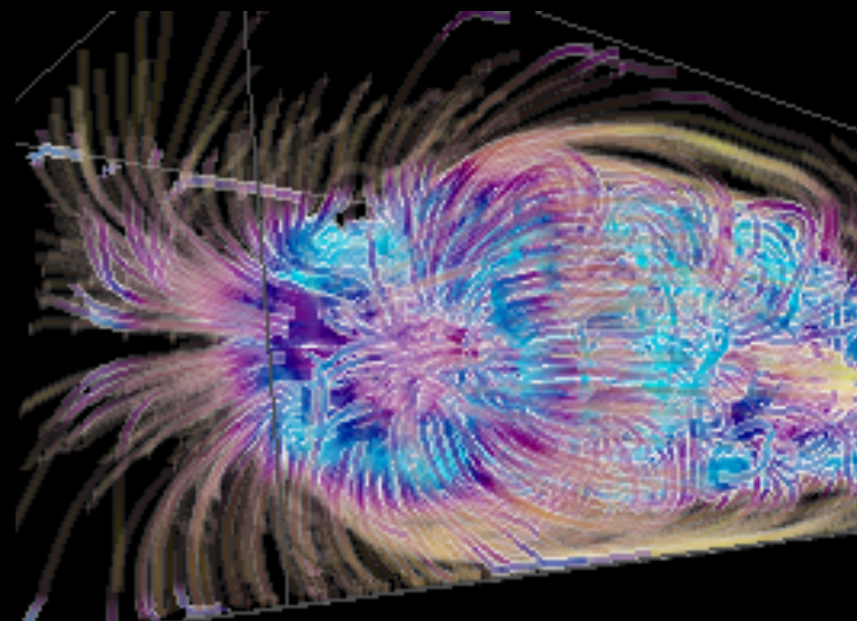

Left: 64 surfaces each seeded with 512 particles are advected in a 504x504x2048 simulation of a solar flare. Right: 64 surfaces each with 2K seeds in a 2K x 2K x 2K Nek5000 thermal hydraulics simulation. Time excludes I/O.

Lu et al., Scalable Computation of Stream Surfaces on Large Scale Vector Fields, submitted to SC14.

# Information Entropy

Courtesy Abon Chaudhuri



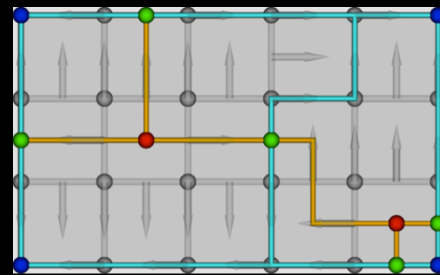



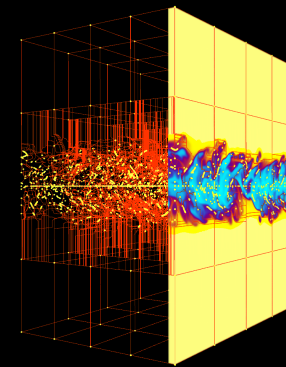
Computation of information entropy in 126x126x512 solar plume dataset shows 59% strong scaling efficiency. Time excludes I/O.

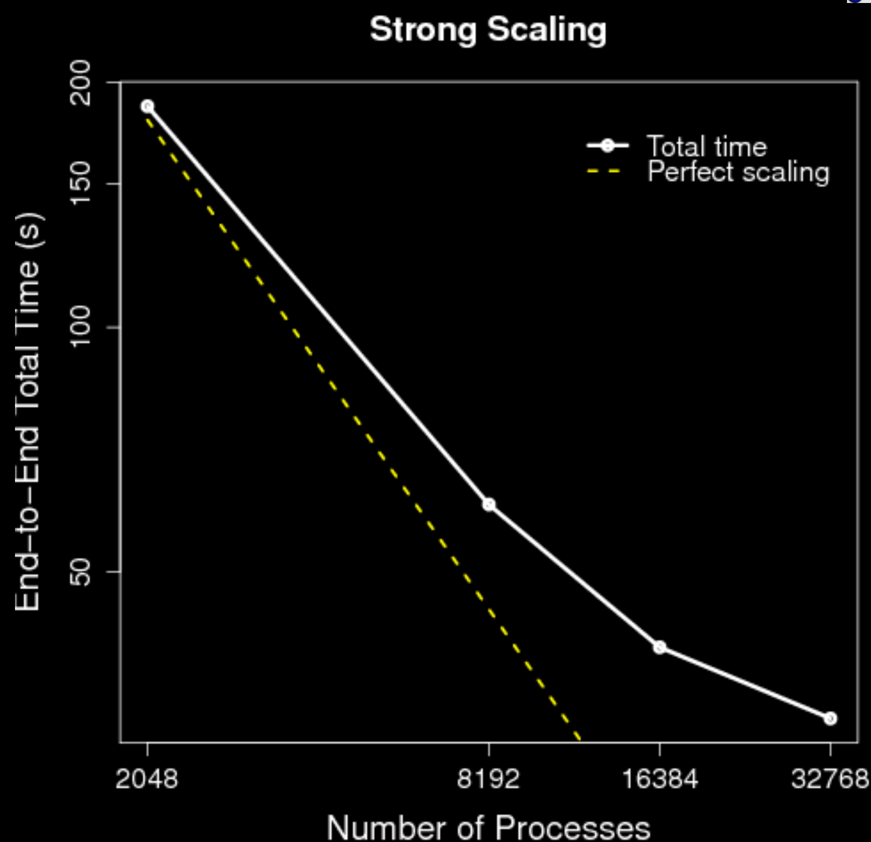
Chaudhuri et al., Scalable Computation of Distributions from Large Scale Data Sets, LDAV ’12.

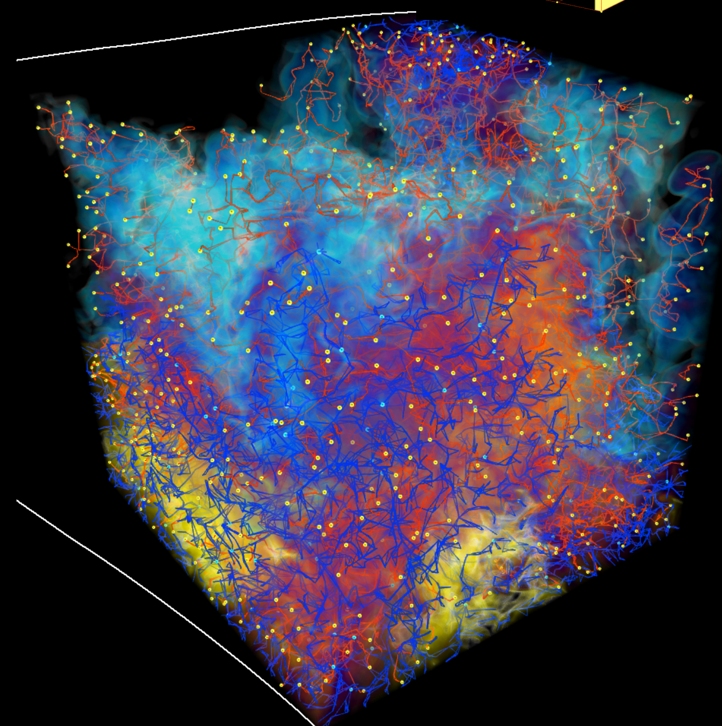
# Topological Analysis

Courtesy Attila Gyulassy







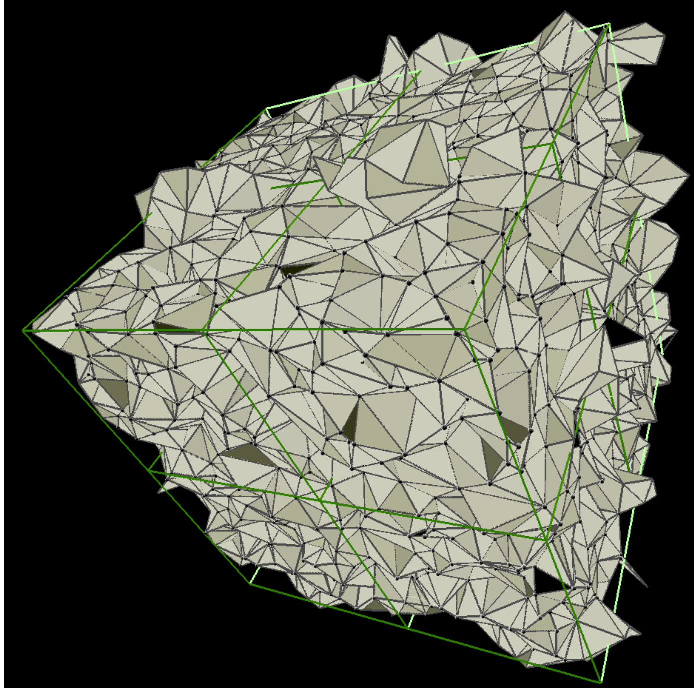



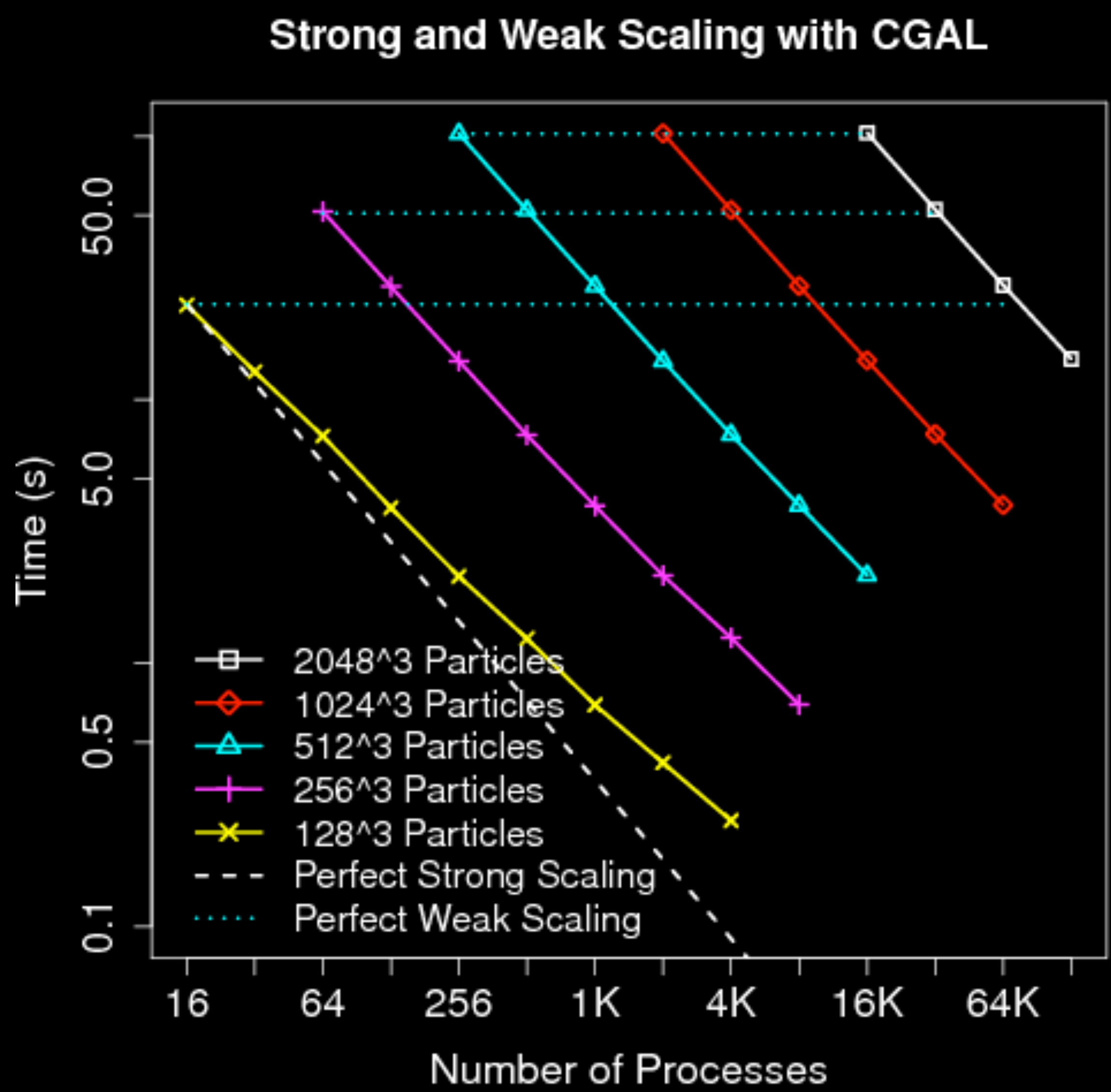
Computation of Morse-Smale complex in $1152^3$ Rayleigh-Taylor instability data set results in 35% end-to-end strong scaling efficiency, including I/O.

Gyulassy et al., The Parallel Computation of Morse-Smale Complexes, IPDPS '12.

# Computational Geometry

With Dmitriy Morozov and Carolyn Phillips

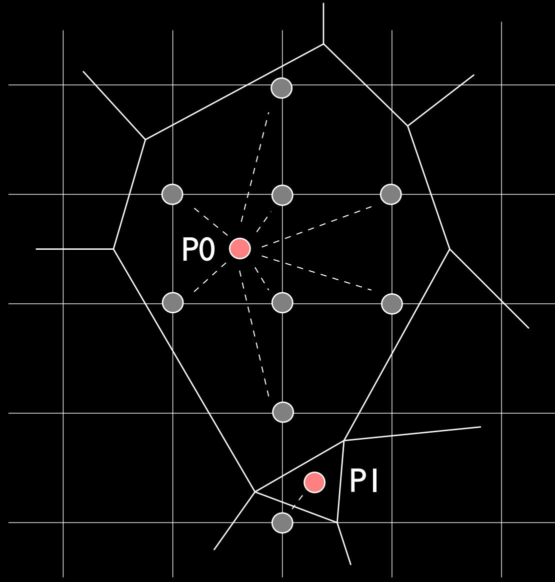


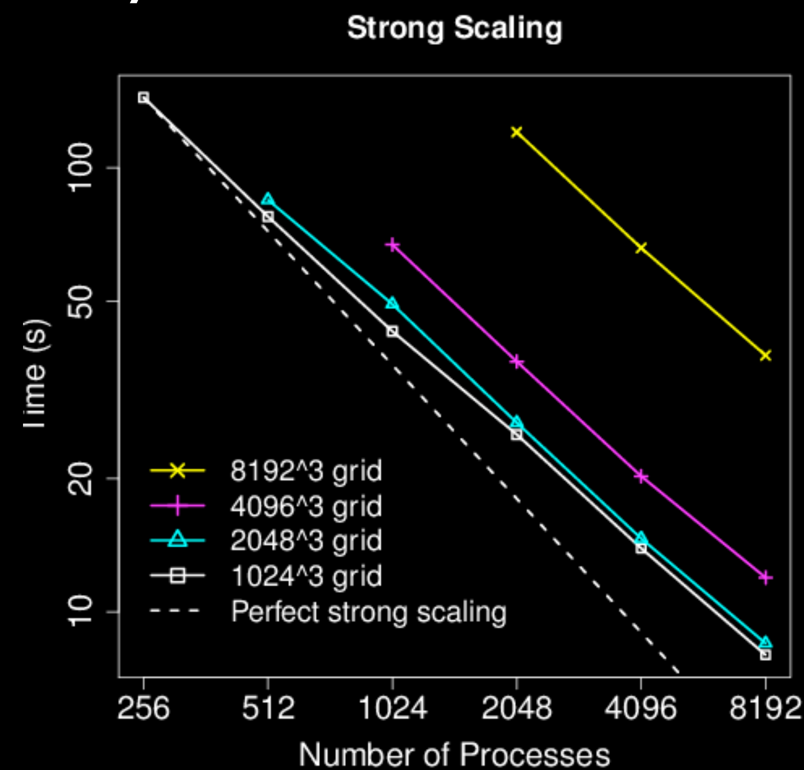


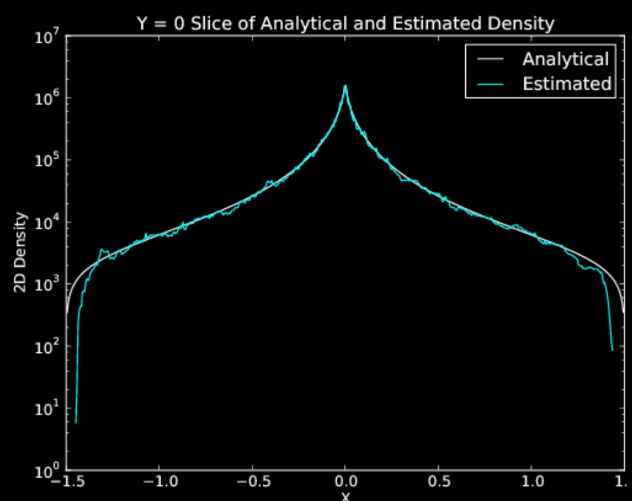

Strong and weak scaling for up to $2048^3$ synthetic particles and up to 128K processes (excluding I/O) shows up to 90% strong scaling and up to 98% weak scaling.

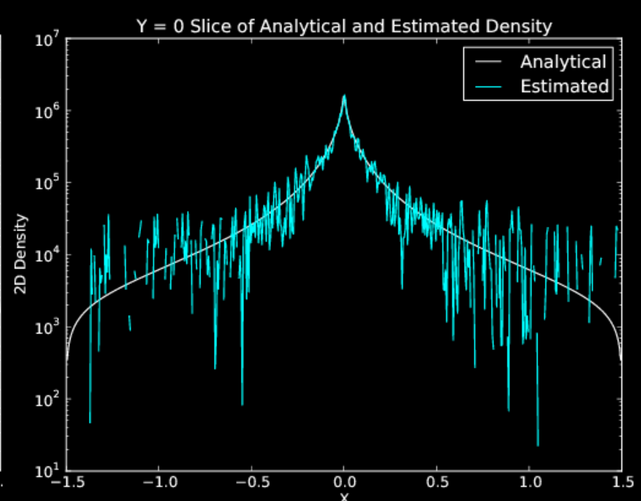
# Tessellation-Based Density Estimation

Tessellation-based density estimation is parameter free, shape free, and automatically adaptive



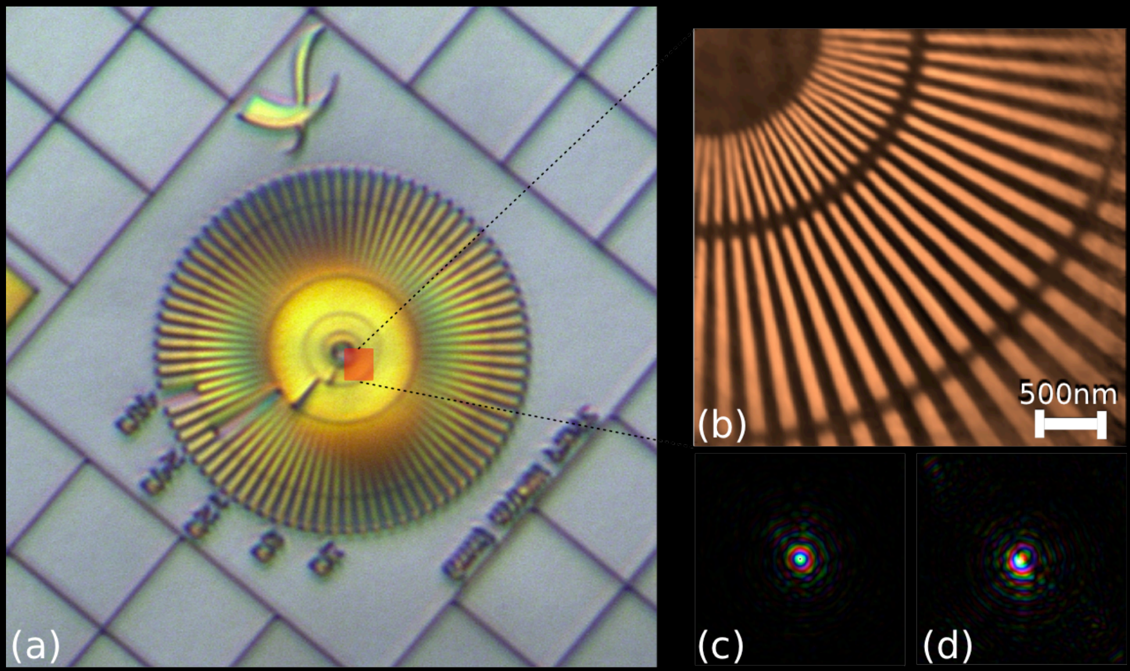

Above: Strong scaling of estimating the density of 512^3 synthetic particles onto grids of various sizes.

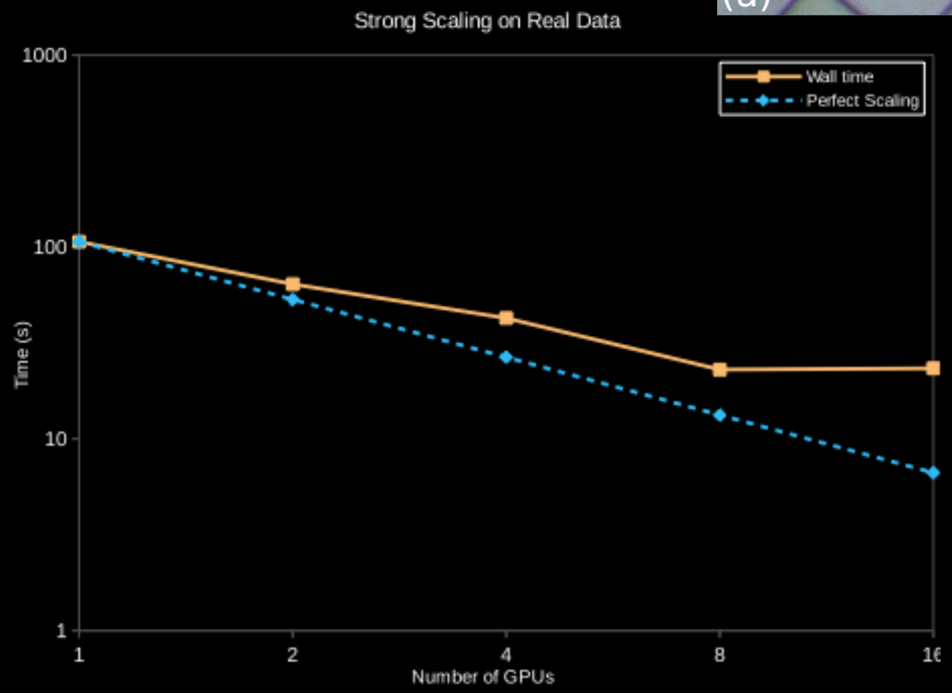
Left: comparison of tessellation-based and CIC density

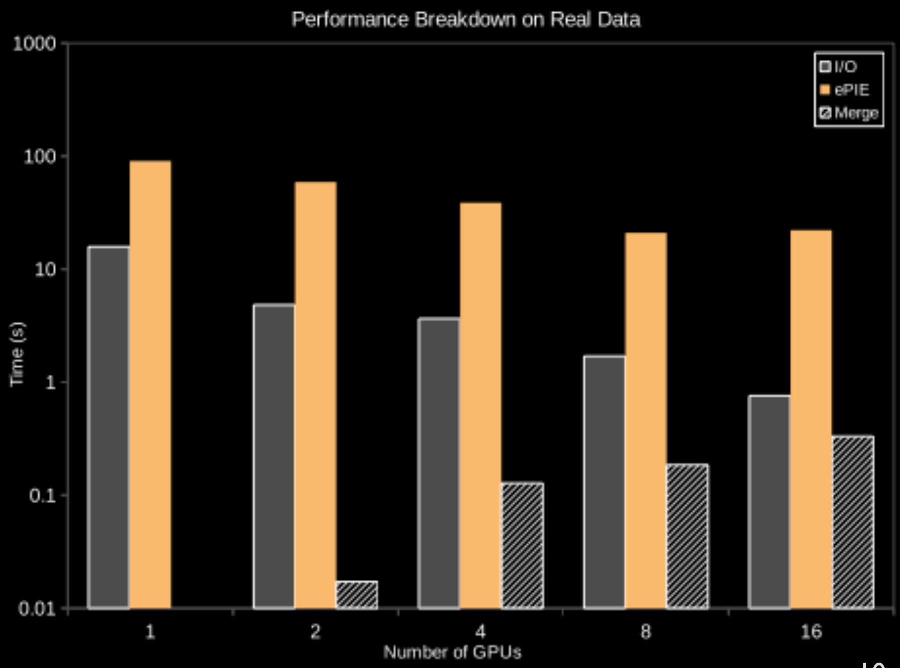
# Parallel Reconstruction Performance

*[Courtesy of Youssef Nashed, ANL]*

A gold Siemens star test pattern, with 30 nm smallest feature size, was raster scanned through a 26×26 grid using a step size of 40nm and an exposure time of 0.6s per scan point, using a 5.2 keV X-ray beam. The total scanning time was about 20 minutes.



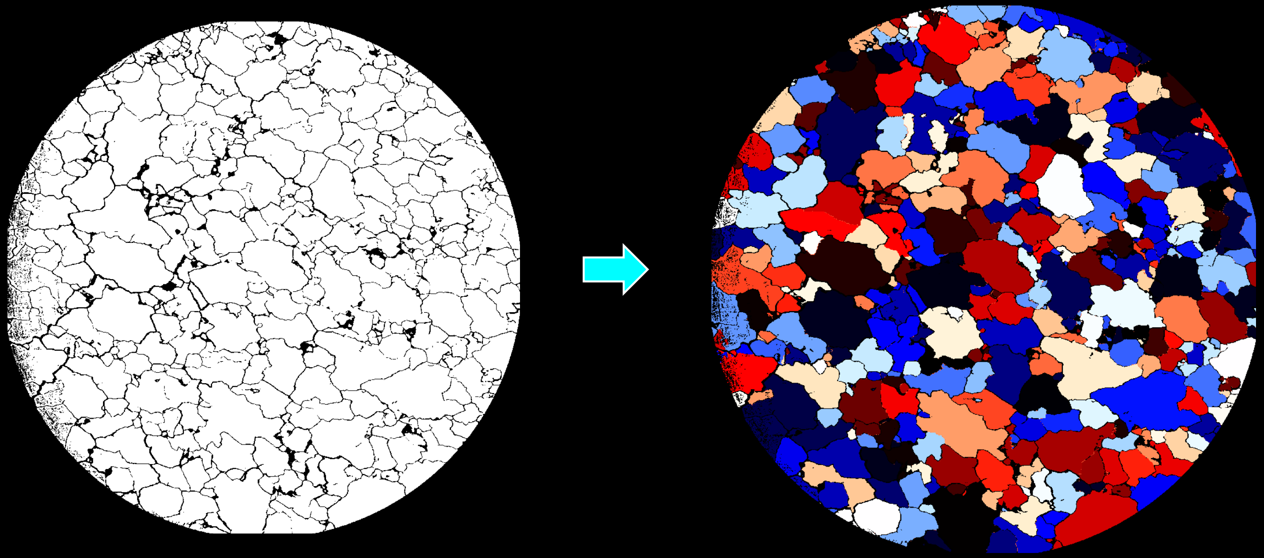



# Image Segmentation in Porous Media

Courtesy Dmitriy Morozov and Patrick O’Neil

LBL (Dmitriy Morozov and Patrick O’Neil) developed tools for segmentation and connectivity analysis of granular and porous media using diy2.



Left: 3D image of a granular material (flexible sandstone) acquired at ALS by Michael Manga and Dula Parkinson. (Data: 2560 × 2560 × 1276). Right: Watershed segmentation of the material identifies individual grains (run on Edison @ NERSC) [courtesy Morozov, O’Neil (LBL)].

# Further Reading


DIY

- Peterka, T., Ross, R., Kendall, W., Gyulassy, A., Pascucci, V., Shen, H.-W., Lee, T.-Y., Chaudhuri, A.: Scalable Parallel Building Blocks for Custom Data Analysis. Proceedings of Large Data Analysis and Visualization Symposium (LDAV'11), IEEE Visualization Conference, Providence RI, 2011.
- Peterka, T., Ross, R.: Versatile Communication Algorithms for Data Analysis. 2012 EuroMPI Special Session on Improving MPI User and Developer Interaction IMUDI'12, Vienna, AT.

DIY applications

- Peterka, T., Morozov, D., Phillips, C.: High-Performance Computation of Distributed-Memory Paral- lel 3D Voronoi and Delaunay Tessellation. Proceedings of SC14, New Orleans, LA, 2014.
- Lu, K., Shen, H.-W., Peterka, T.: Scalable Computation of Stream Surfaces on Large Scale Vector Fields. Proceedings of SC14, New Orleans, LA, 2014.
- Chaudhuri, A., Lee-T.-Y., Shen, H.-W., Peterka, T.: Efficient Indexing and Querying of Distributions for Visualizing Large-scale Data. Proceedings of IEEE PacificVis 2014, Yokohama, Japan.
- Nashed, Y., Vine, D., Peterka, T., Deng, J., Ross, R., Jacobsen, C.: Parallel Ptychographic Recon- struction. Optics Express Journal, Vol. 22, No. 26, 2014.
- Wozniak, J., Peterka, T., Armstrong, T., Dinan, J., Lusk, E., Wilde, M., Foster, I.: Dataflow Coor- dination of Data-Parallel Tasks via MPI 3.0. Proceedings of EuroMPI, Madrid, Spain, September 2013.
- Sewell, C., Meredith, J., Moreland, K., Peterka, T., DeMarle, D., Lo, Li-ta, Ahrens, J., Maynard, R., Geveci, B.: The SDAV Software Frameworks for Visualization and Analysis on Next-Generation Multi-Core and Many-Core Architectures. Proceedings of the SC12 Ultrascale Visualization Workshop, Salt Lake City, UT.
- Peterka, T., Ross, R.: Versatile Communication Algorithms for Data Analysis. 2012 EuroMPI Special Session on Improving MPI User and Developer Interaction IMUDI'12, Vienna, AT.
- Peterka, T., Ross, R., Nouanesengsey, B., Lee, T.-Y., Shen, H.-W., Kendall, W., Huang, J.: A Study of Parallel Particle Tracing for Steady-State and Time-Varying Flow Fields. Proceedings IPDPS'11, Anchorage AK, May 2011.
- Gyulassy, A., Peterka, T., Pascucci, V., Ross, R.: The Parallel Computation of Morse-Smale Complexes. Proceedings of IPDPS'12, Shanghai, China, 2012.
- Nouanesengsy, B., Lee, T.-Y., Lu, K., Shen, H.-W., Peterka, T.: Parallel Particle Advection and FTLE Computation for Time-Varying Flow Fields. Proceedings of SC12, Salt Lake, UT.
- Peterka, T., Kwan, J., Pope, A., Finkel, H., Heitmann, K., Habib, S., Wang, J., Zagaris, G.: Meshing the Universe: Integrating Analysis in Cosmological Simulations. Proceedings of the SC12 Ultrascale Visualization Workshop, Salt Lake City, UT.
- Chaudhuri, A., Lee-T.-Y., Zhou, B., Wang, C., Xu, T., Shen, H.-W., Peterka, T., Chiang, Y.-J.: Scalable Computation of Distributions from Large Scale Data Sets. Proceedings of 2012 Symposium on Large Data Analysis and Visualization, LDAV'12, Seattle, WA.
- Peterka, T., Kendall, W., Goodell, D., Nouanesengsey, B., Shen, H.-W., Huang, J., Moreland, K., Thakur, R., Ross, R.: Performance of Communication Patterns for Extreme-Scale Analysis and Visu- alization. Journal of Physics: Conference Series SciDAC 2010, 2010.

U.S. DEPARTMENT OF ENERGY

Argonne NATIONAL LABORATORY

# DIY Applications

Acknowledgments:

Facilities
Argonne Leadership Computing Facility (ALCF)
Oak Ridge National Center for Computational Sciences (NCCS)
National Energy Research Scientific Computing Center (NERSC)

Funding
DOE SDMAV Exascale Initiative
DOE Exascale Codesign Center
DOE SciDAC SDAV Institute

People
Dmitriy Morozov, Han-Wei Shen, Abon Chaudhuri, Kewei Lu, Teng-Yok Lee, Attila Gyulassy, Valerio Pascucci, Salman Habib, Boothanome Nouanesengsy, Jian Huang, Rob Ross

https://bitbucket.org/diatomic/diy

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