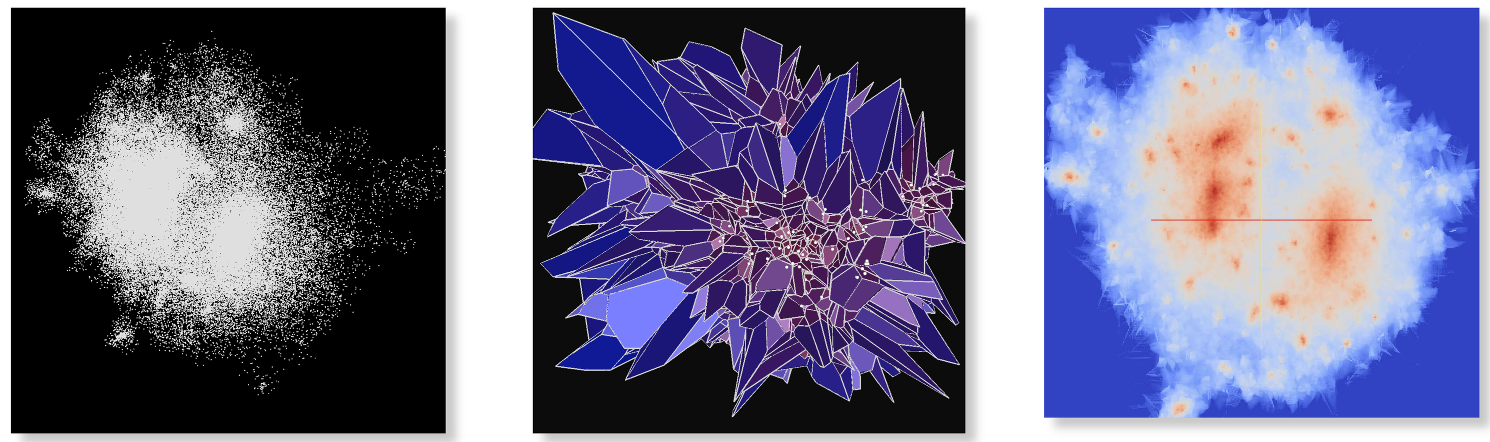
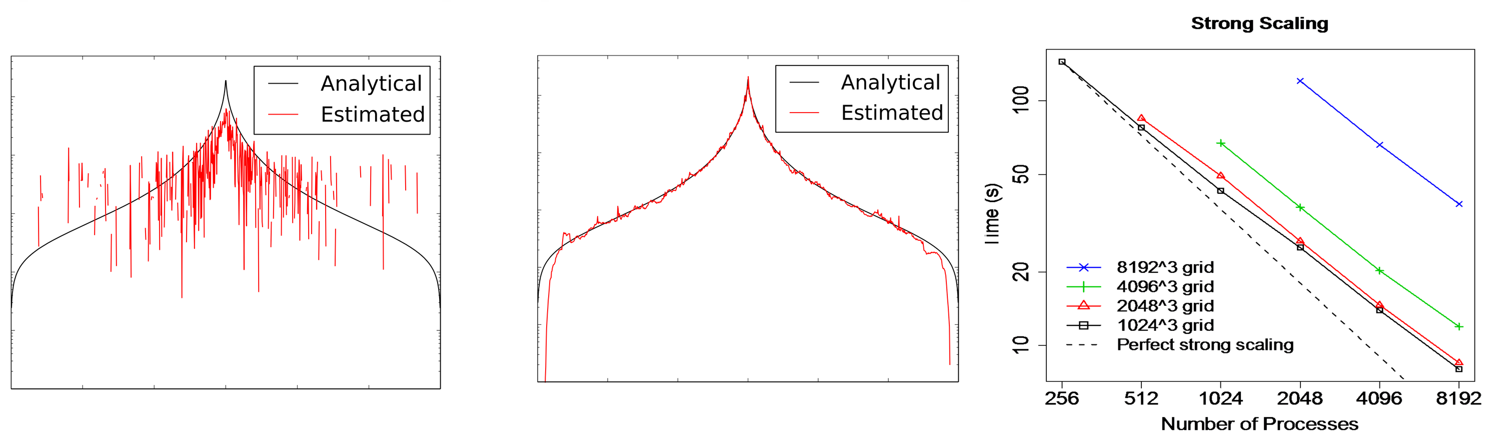
DIY Applications

SDAV technologies help scientists analyze data ranging from nanometers to gigaparsecs and femtoseconds to billions of years. DIY is a technology that enables parallelization of numerous data analysis methods at extreme scale.

Density Estimation in Cosmology and Astrophysics



Three representations of the same halo. From left to right: original raw particle data, Voronoi tessellation, and regular grid density sampling.

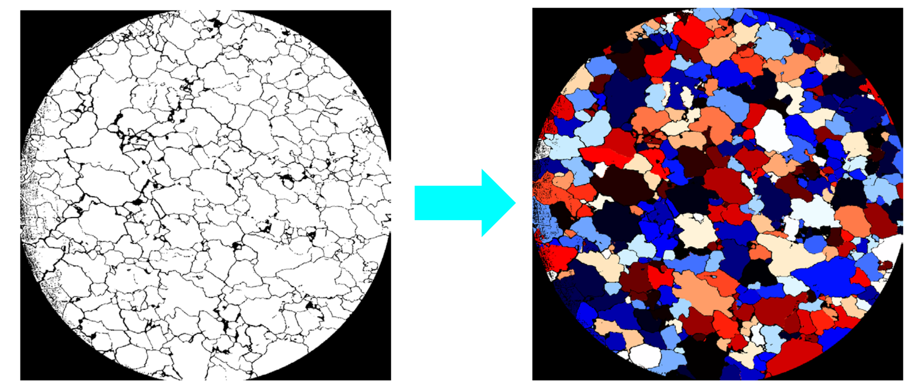


Left and center: Cross sections of CIC and TESS-based density estimators. Right: Strong scaling for grid sizes ranging from 1024^3 to 8192^3.

Image Segmentation in Porous Media

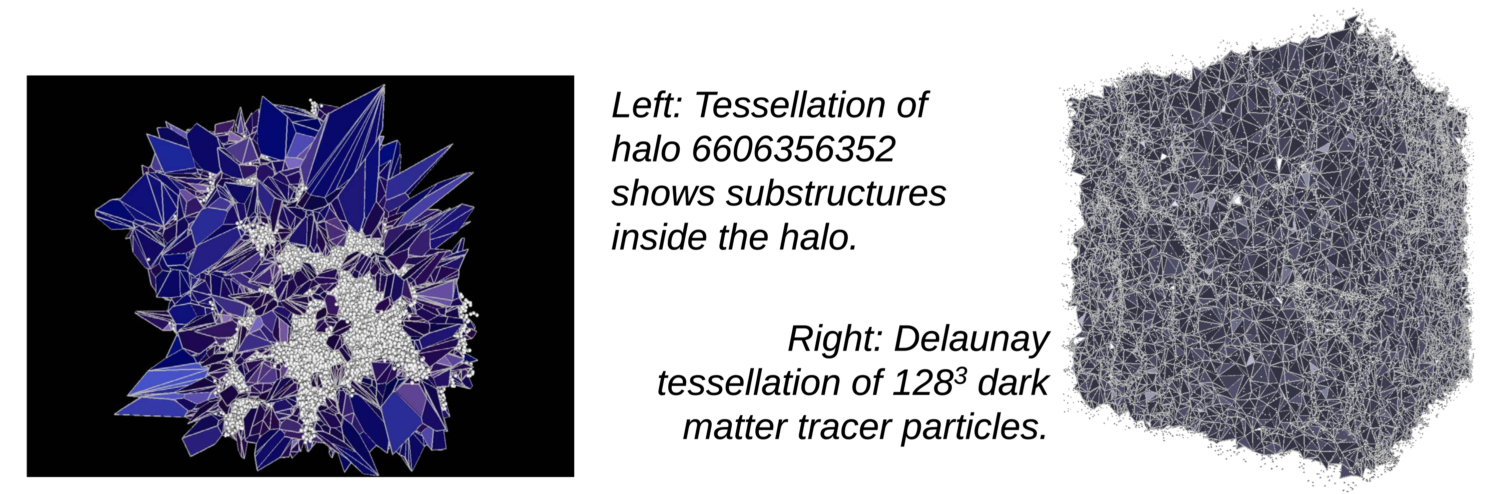
[Courtesy Dmitriy Morozov, LBL]

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



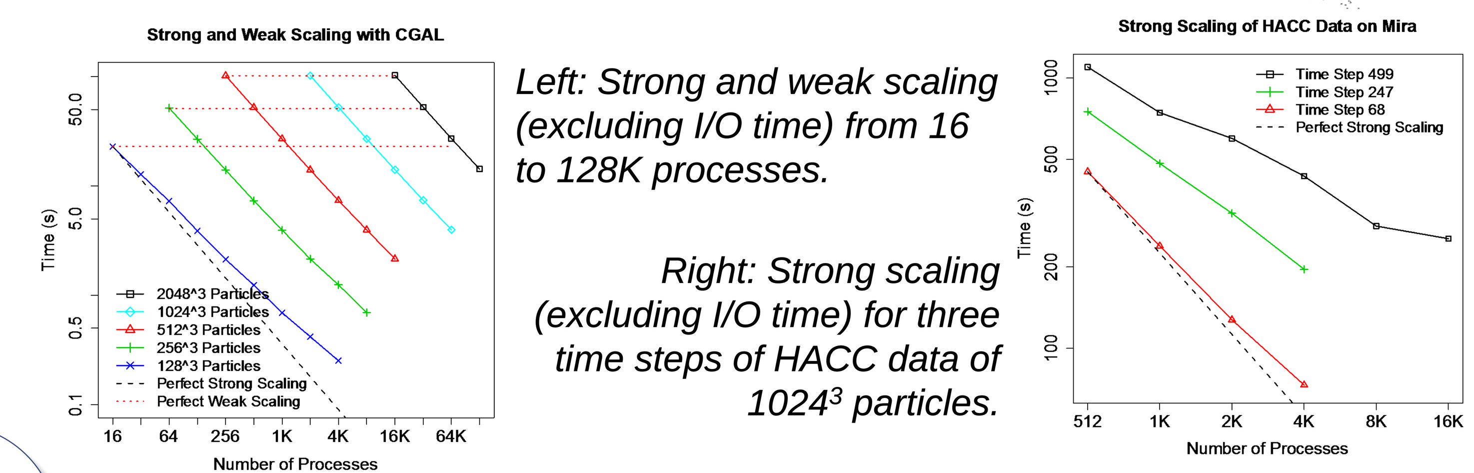
Left: 3D image of a granular material (flexible sandstone) acquired at ALS by Michael Manga and Dula Parkinson. (Data: 2560 × 2560 × 1276). Right: Watershed segmentation of the material identifies individual grains (run on Edison @ NERSC) [courtesy Morozov, O'Neil (LBL)].

Voronoi and Delaunay Tessellations in Cosmology



Left: Tessellation of halo 6606356352 shows substructures inside the halo.

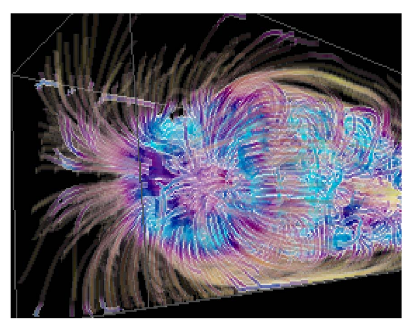
Right: Delaunay tessellation of 128^3 dark matter tracer particles.

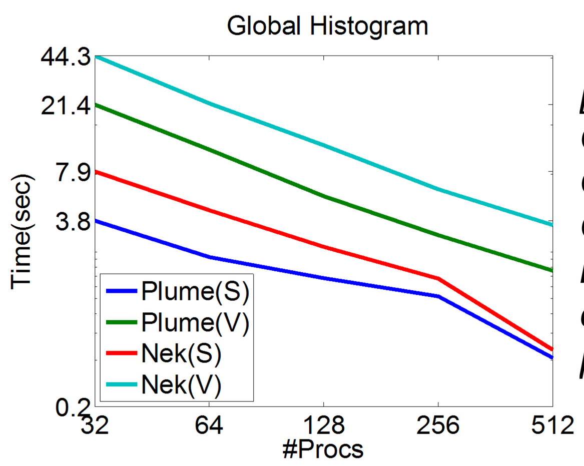


Left: Strong and weak scaling (excluding I/O time) from 16 to 128K processes.

Right: Strong scaling (excluding I/O time) for three time steps of HACC data of 1024^3 particles.

Information Entropy in Astrophysics

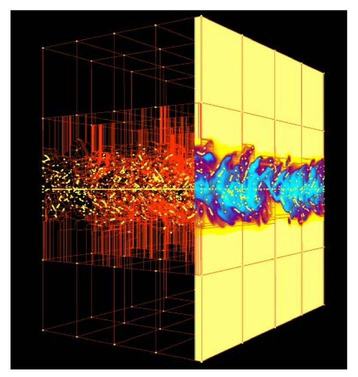


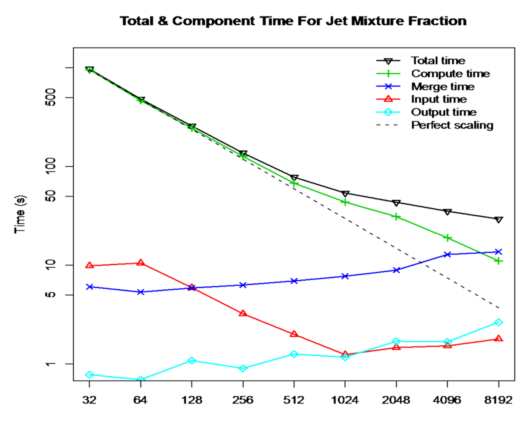


[Courtesy Abon Chaudhuri, OSU] Computation of information entropy in solar plume dataset.

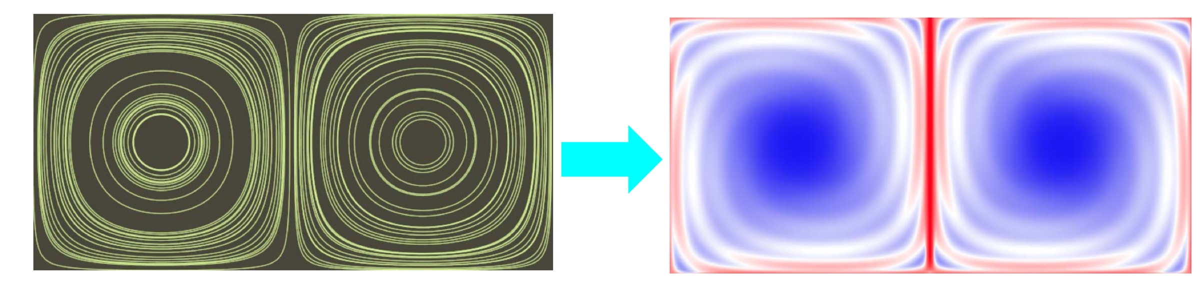
Computational Topology in Combustion

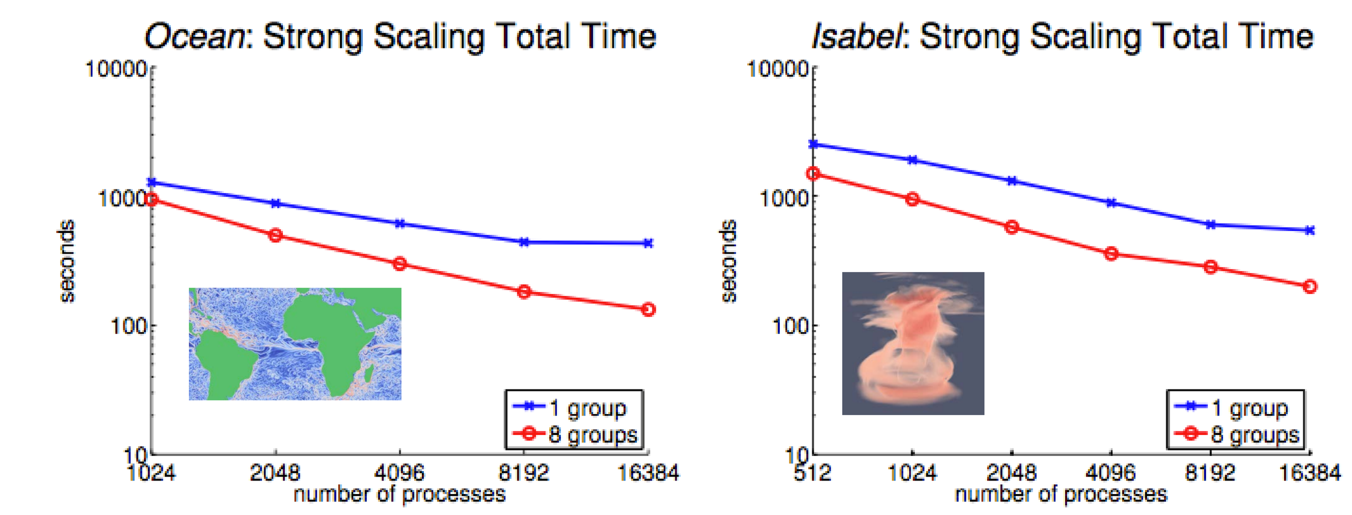
[Courtesy Attila Gyulassy, SCI] Computation of Morse-Smale complex in jet mixture fraction data set





LaGrangian Coherent Structures Based on FTLE Computation in Climate Science

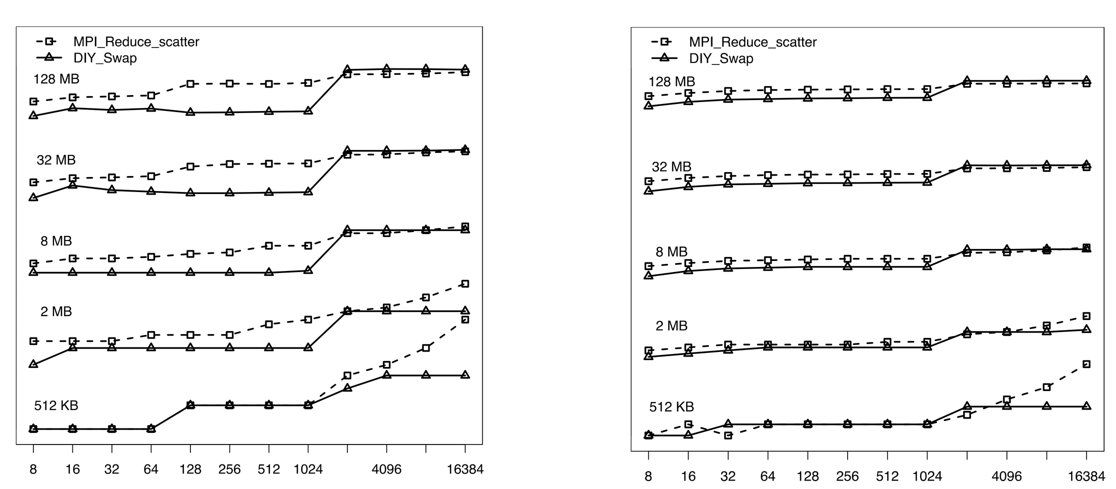




[Courtesy Boonthanome Nouanesengsy, OSU]

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.

Proxy Applications for Exascale Codesign

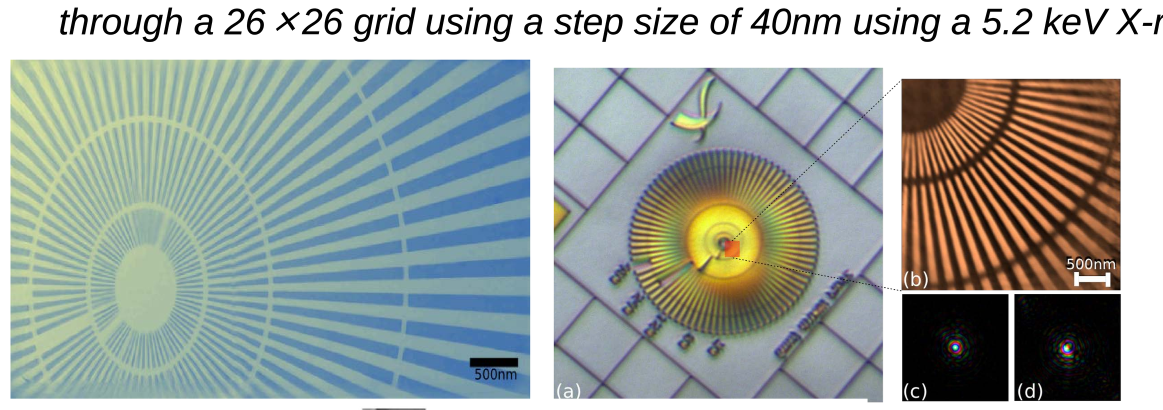


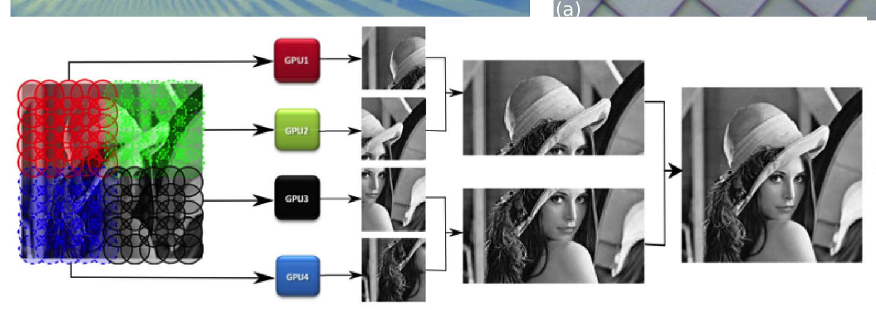
# Items	Bytes/Item	Total Bytes	# Procs	Exchange Time (s)
64	20	1 K	512	0.003
256	20	5 K	512	0.010
1 K	20	20 K	512	0.040
4K	20	80 K	512	0.148
16 K	20	320 K	512	0.627
64 K	20	1 M	512	2.516
256 K	20	5 M	512	10.144
1 M	20	20 M	512	40.629

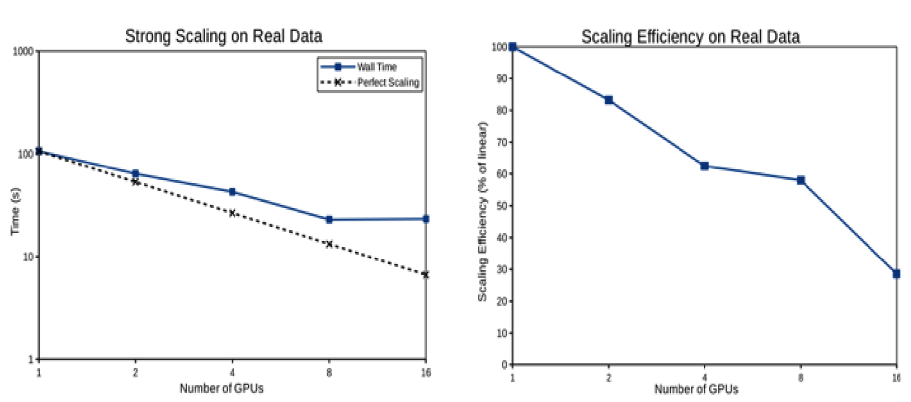
Left: Communication time only for our merge algorithm compared with MPI's reduction algorithm. Center: Our swap algorithm compared with MPI's reduce-scatter algorithm. Right: Performance of neighbor exchange algorithm at high numbers of small messages is a stress test of point to point communication. Diy2 shows linear complexity in total data size.

Parallel Ptychographic Image Reconstruction

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 using a 5.2 keV X-ray beam.

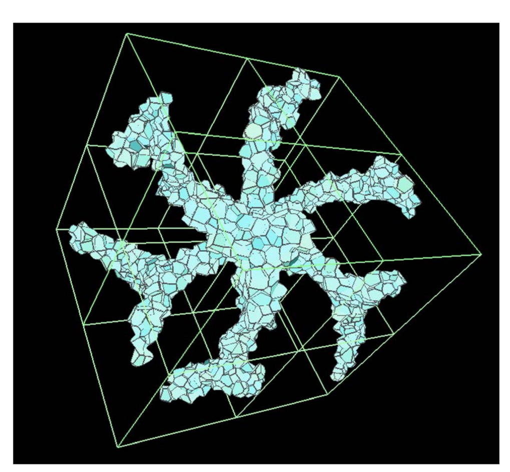






[Courtesy Youssef Nashed, ANL]. An illustration of multi-GPU phase retrieval. The diffraction patterns are subdivided and distributed among available GPUs. The initial diffraction patterns are subdivided and merged together using DIY's merge-based reduction. Strong scaling efficiency on synthetic data is 55% on 21,000 256x256 images.

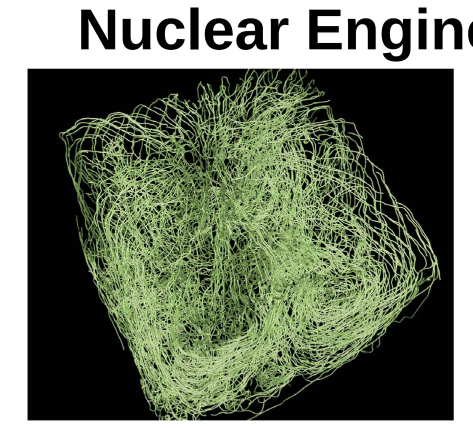
Tessellations in Molecular Dynamics

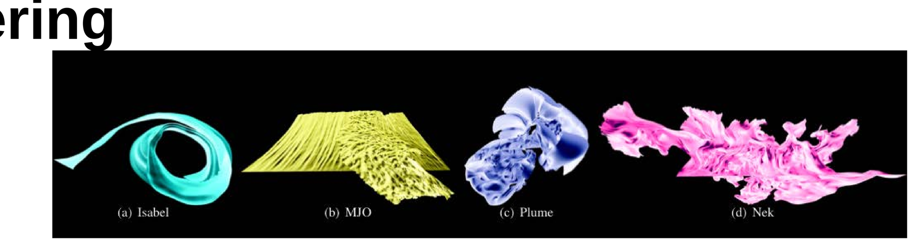


[Courtesy of Carolyn Phillips, ANL] In simulations of soft matter systems, molecules self-organize to form two or more domains. These domains can form complicated geometries such as the double gyroid. Summing the Voronoi cells of 300 time steps of a system of an A-B-A triblock copolymer suggests that the B domain dilates relative to the A domain.

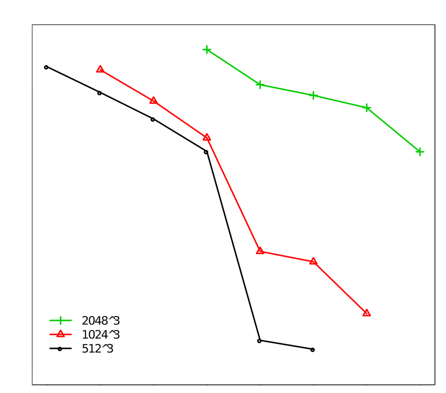
Voronoi tessellation of 1,000 A-B-C ``telechelics" composed of two nanospheres connected by polymer tether beads in a double gyroid morphology. Only the Voronoi cells associated with the A species are shown. Such surfaces are usually constructed by using isosurface methods, which require averaging over many time steps; whereas by using the tessellation, such surfaces can be constructed for every time step.

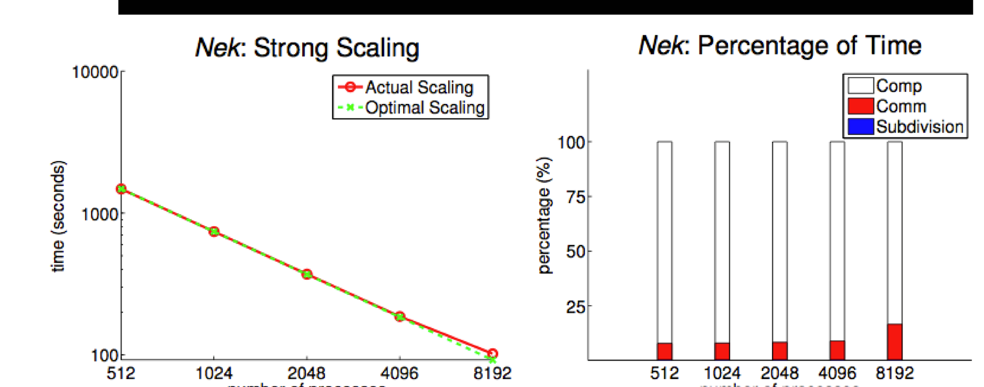
Streamlines, Pathlines, and Stream Surfaces in Nuclear Engineering





Particle tracing of thermal hydraulics data is plotted in log-log scale. The top panel shows 400 particles tracing streamlines in this flow field. In the center panel, 128 K particles are traced in three data sizes: 512^3 (134 million cells), 1024^3 (1 billion cells), and 2048^3 (8 billion cells). End-to- end results are shown, including I/O (reading the vector dataset from storage and writing the output particle traces.)





[Courtesy Kewei Lu, OSU] 64 surfaces each with 2K seeds in a 2K x 2K x 2K Nek5000 thermal hydraulics simulation. Time excludes I/O. Left: Strong scaling. Right: Percentage of time in the three stages of our algorithm.

SDAV
Scalable Data Management, Analysis, and Visualization

Tom Peterka (ANL), Dmitriy Morozov (LBNL), Carolyn Phillips (ANL), Youssef Nashed (ANL), Han-Wei Shen (OSU), Boonthanome Nouanesensy (LANL), Kewei Lu (OSU)