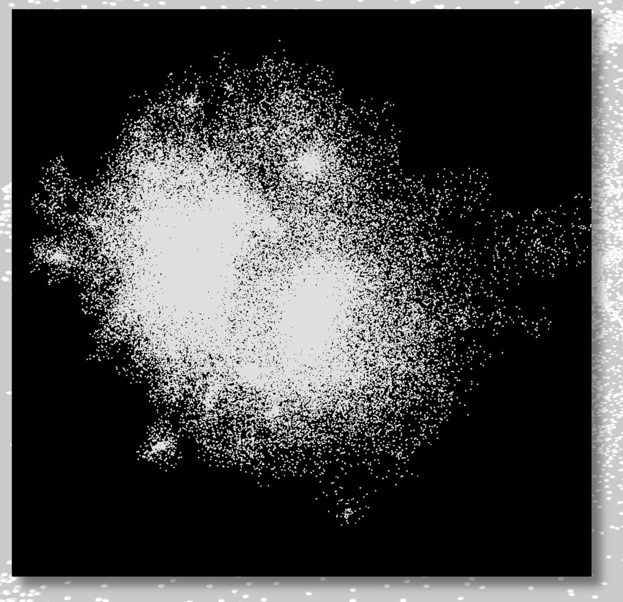
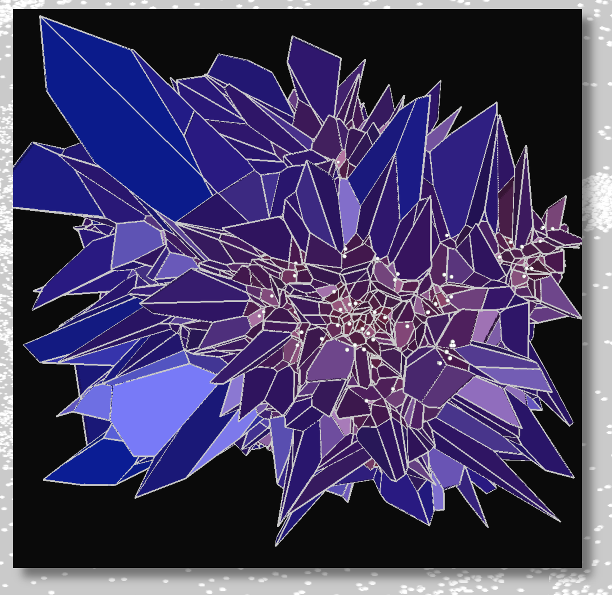
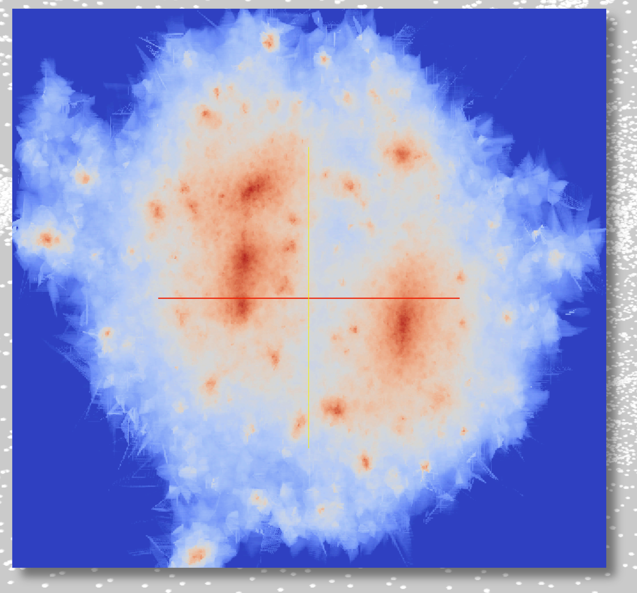
Exploring Cosmology with SDAV Technologies

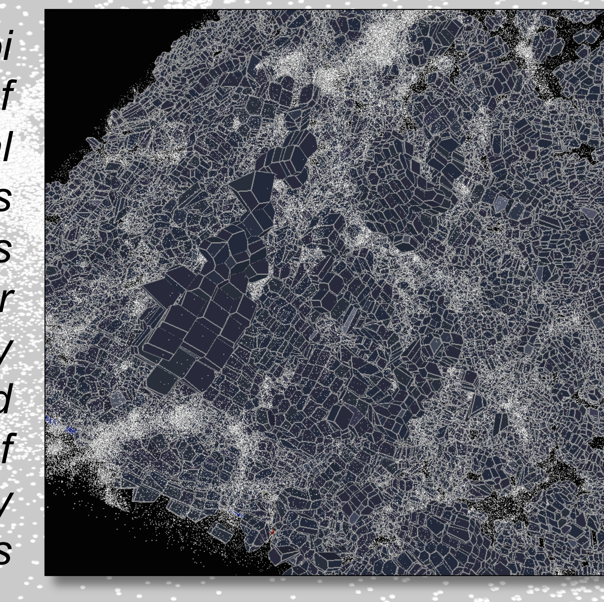




Three representations of the same halo. From left to right: original raw particle data, Voronoi tessellation, and regular grid density sampling.

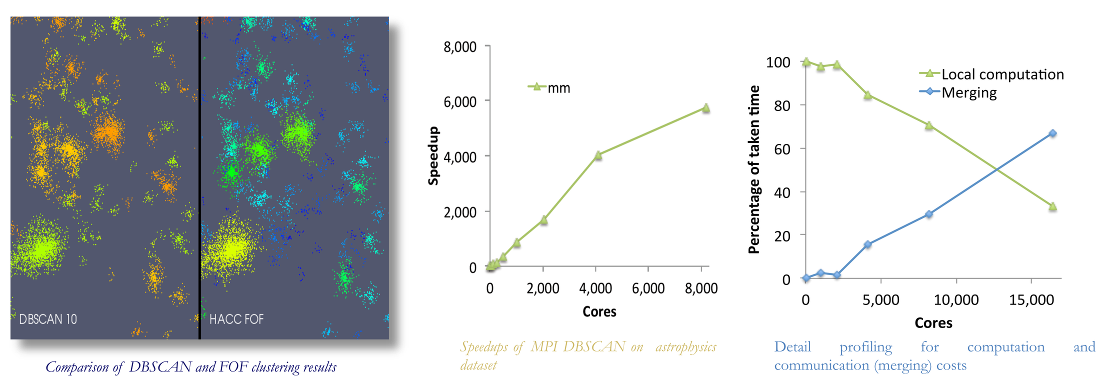
Overview

SDAV technologies aim to help cosmologists unravel the mysterious nature of dark matter and energy by transforming raw data into meaningful representations. For example, mesh tessellations help analyze point data because they transform sparse discrete samples into dense continuous functions. Similarly, large-scale structures such as halos and voids are extracted, tracked, and summarized in high-level models. The goal of SDAV's partnership with computational cosmology is to bring such methods to extreme scale.

Voronoi tessellation of cosmological simulations reveals regions of irregular low-density voids amid clusters of high-density halos

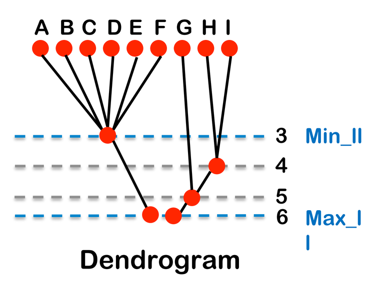


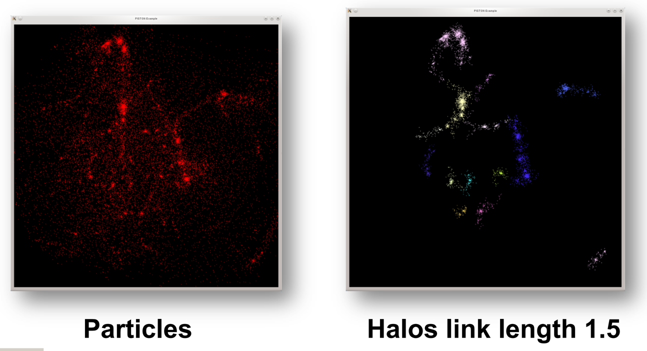
Feature Detection & Modeling

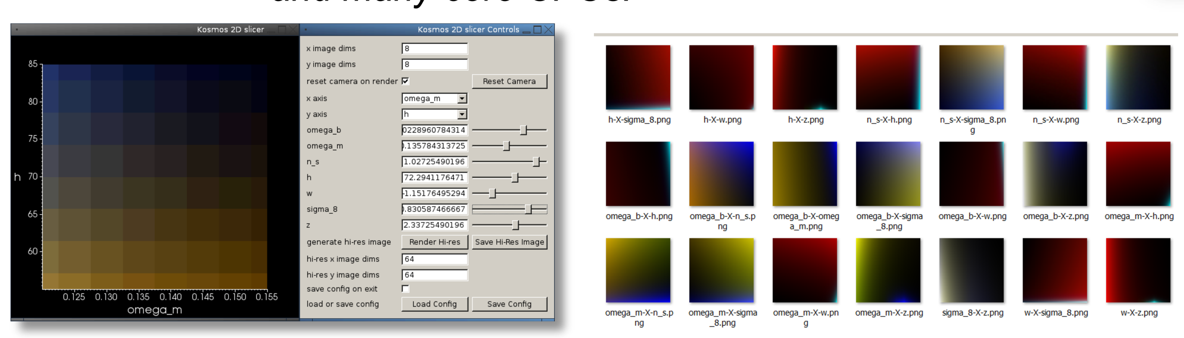


Left: Comparison of density-based clustering (DBSCAN) and HACC friends-of-friends (FOF) halo finder shows similar results but improved linking at halo edges and less false linkage. Scalability of parallel DBSCAN is shown in center and right.

Right: Data-parallel halo finder using dendrograms allows queries over a range of linking lengths. FOF halo finder implemented in PISTON leverages thread-parallelism of GPUs and many-core CPUs.



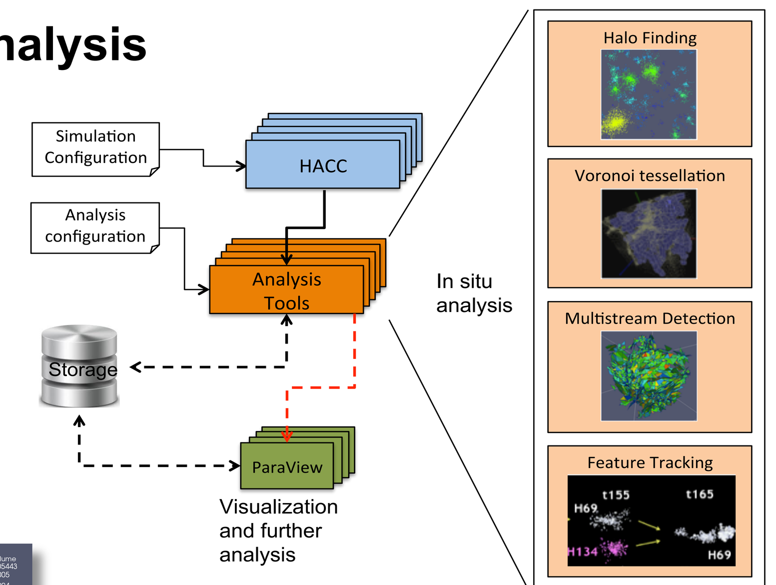


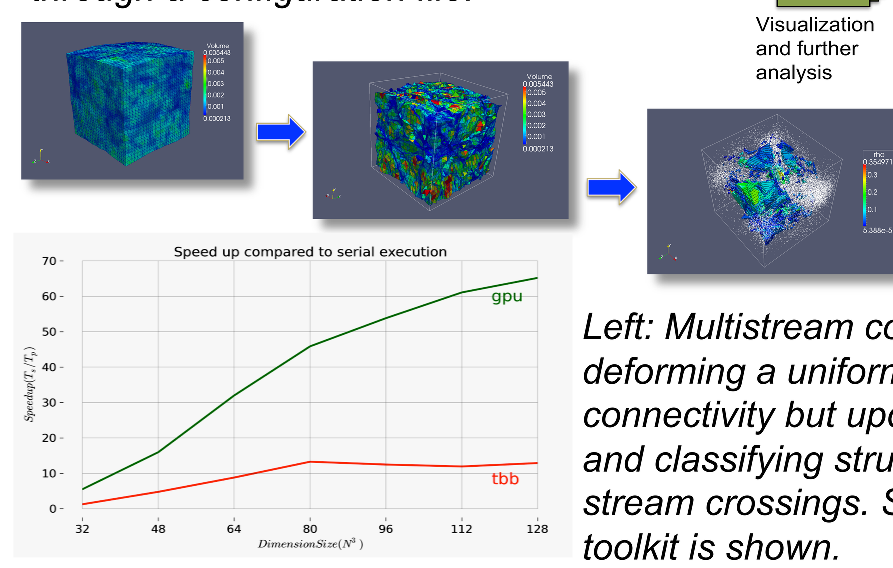


Left: Cosmic emu 2D slicer portrays small 2D multiples of a 5D space of cosmological input parameters and maps output power spectrum to color.

CosmoTools: In Situ Analysis Framework

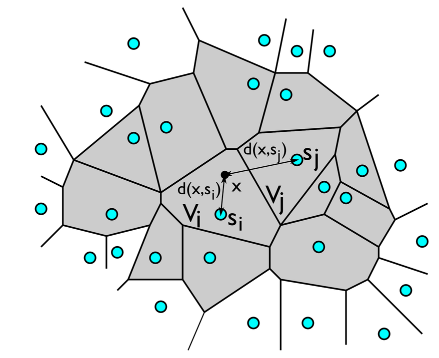
Right: A unified framework facilitates integration of new algorithms, services, and tools without modifying HACC code. A simple API, consisting of 7 main functions allows different tools to be easily controlled through a configuration file.



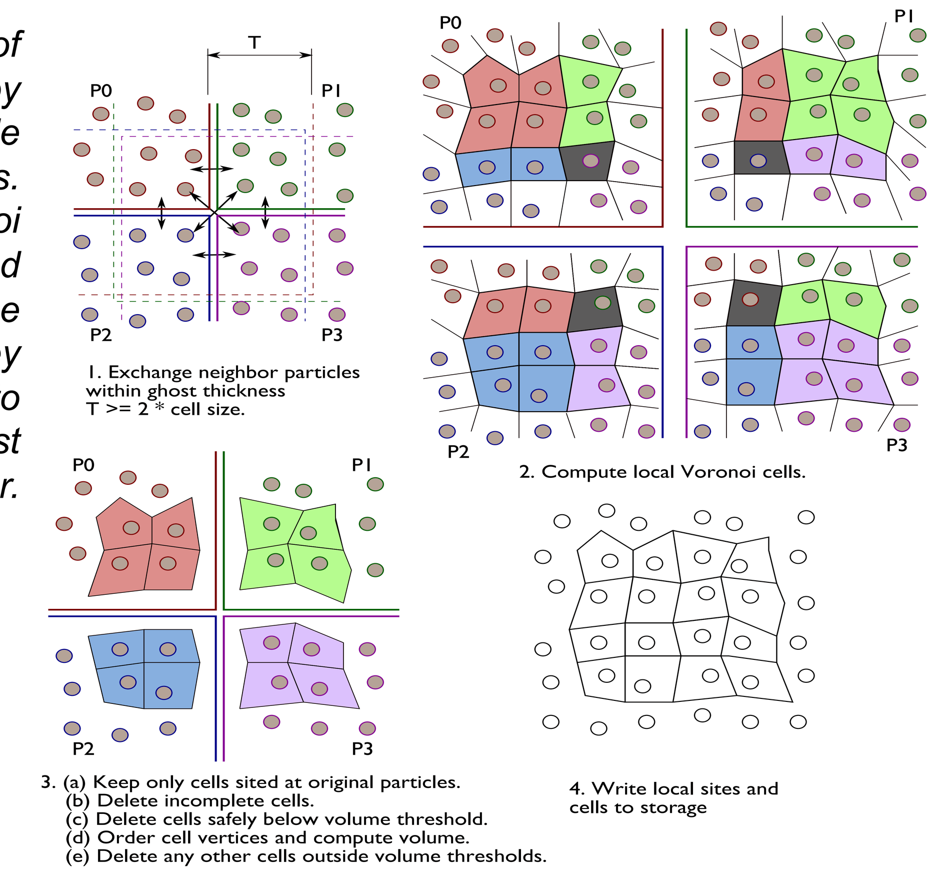


Left: Multistream counting consists of deforming a uniform mesh with fixed connectivity but updated particle positions and classifying structures by number of stream crossings. Speedup using DAX toolkit is shown.

Tess: In Situ Voronoi Tessellation Library



Right: Overview of parallel algorithm by using an example with four processes. Particles and Voronoi cells are colored according to the process where they originated prior to exchanging ghost layer.

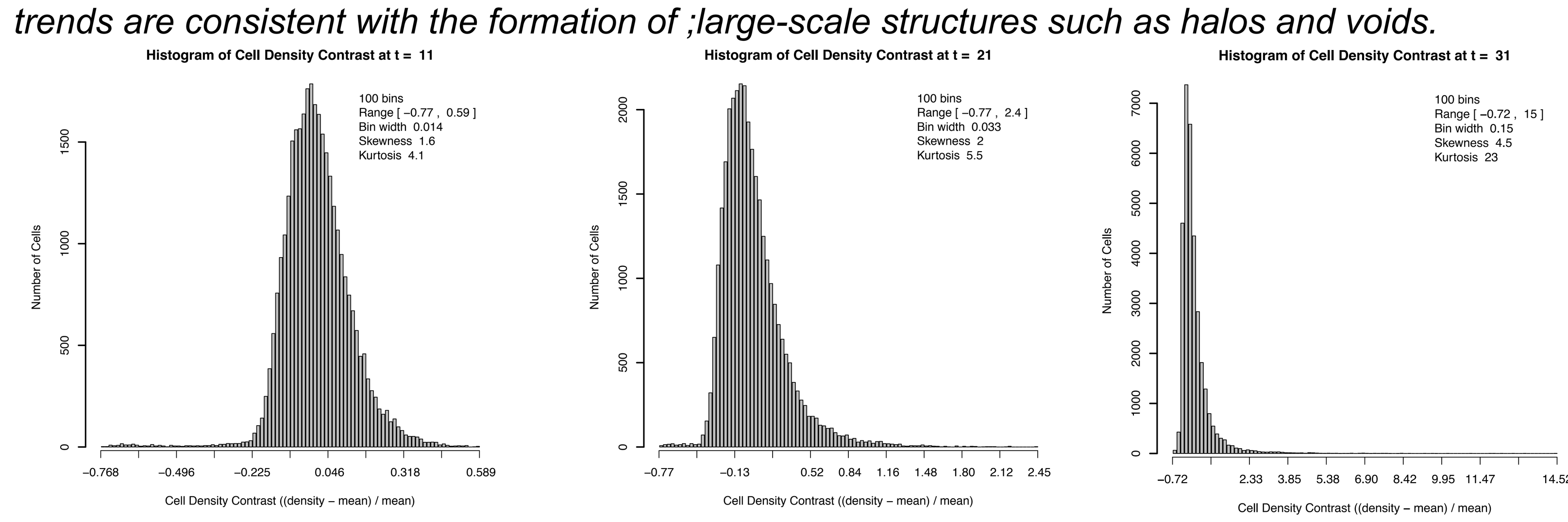


Above: Each Voronoi cell is associated with one input particle, the site of the cell. A cell consists all points closer to the site of that cell than to any other site

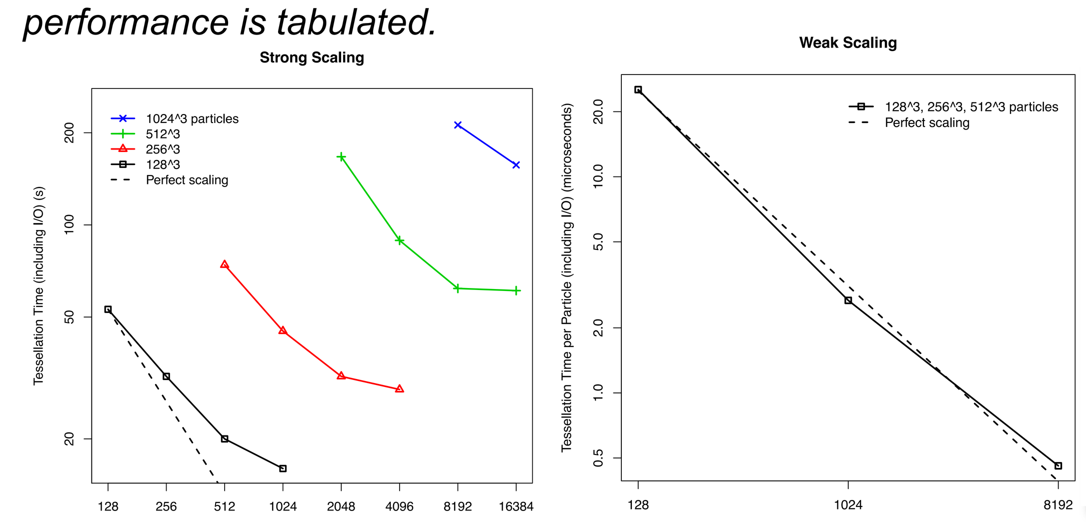
$$V_i = \{ x \mid d(x, s_i) < d(x, s_k) \} \ \forall k \neq i$$

In 3D, Voronoi cells are polyhedra; dual is Delaunay tetrahedralization.

Below: Density contrast distribution of evolving Voronoi cells at three time steps. Statistically, the trends are consistent with the formation of ;large-scale structures such as halos and voids.



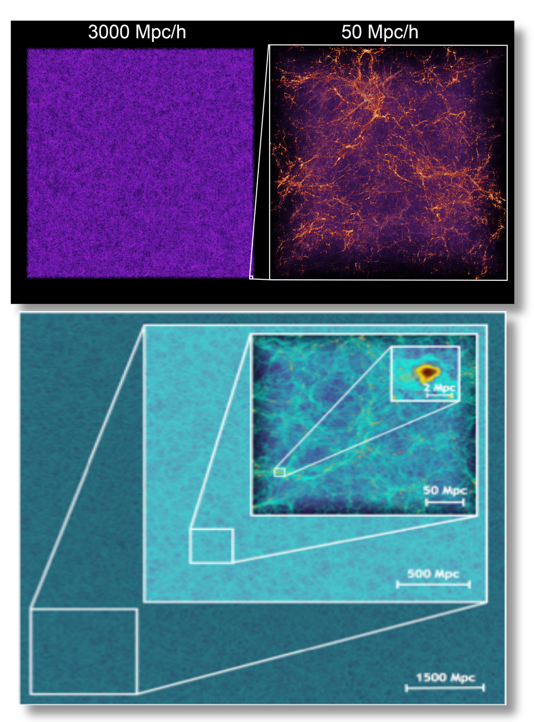
Bottom: In situ strong scaling (left) and weak scaling (center) are plotted on a log-log scale. Weak scaling time is normalized by the number of particles. Plots represent the total tessellation time, including the time to write the result to storage. Strong scaling efficiency is 41%; weak scaling efficiency is 86%. Bottom right: raw performance is tabulated.

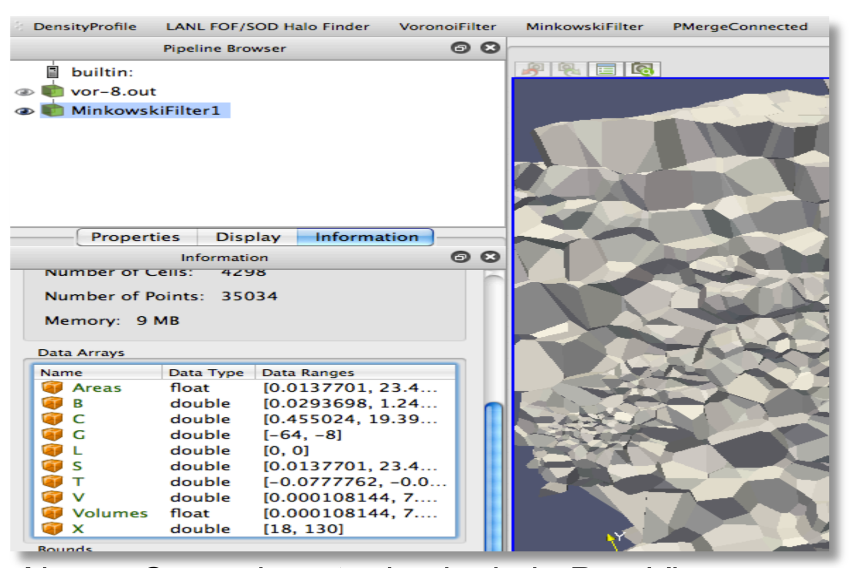


Particles	Time steps	Processes	Tot Time (s)	Sim Time (s)	Tess Tot Time (s)	Exchange Time	Tess Time (s)	Output Time (s)	Output Size(GB)
128^3	100	128	1862	1809	53	1	50	2	0.3
		256	1354	1322	32	1	29	2	
		512	1116	1096	20	1	17	2	
		1024	745	729	16	1	14	3	
256^3	100	512	3090	3016	74	2	69	3	1.7
		1024	2391	2346	45	2	39	4	
		2048	1861	1830	32	2	26	4	
		4096	1334	1305	29	2	15	12	
512^3	50	2048	3852	3684	167	4	157	6	14
		4096	2008	1918	89	3	77	9	
		8192	1784	1722	62	3	48	11	
		16384	1406	1344	61	2	32	27	
1024^3	25	8192	2331	2119	212	6	186	20	101
		16384	1446	1289	157	4	113	40	

Visualization

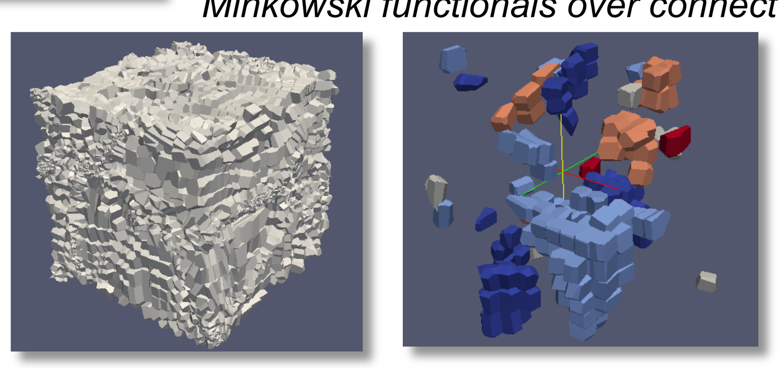
Right: VI3 is a parallel visual analysis framework for Blue Gene supercomputers, GPU-based clusters, and scientists' laptops. Its modular design scales to 16K GPU cores and supports interactive visual analysis and data exploration of 1 trillion particles.





Above: Cosmology tools plugin in ParaView promotes interactive feature exploration and includes a parallel reader for Voronoi tessellations and Minkowski functionals over connected components.

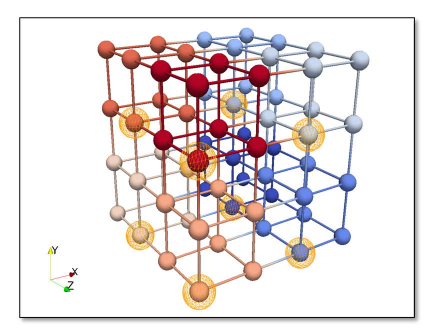
Right: Connected components of Voronoi cells that have been filtered on cell volume are further characterized.



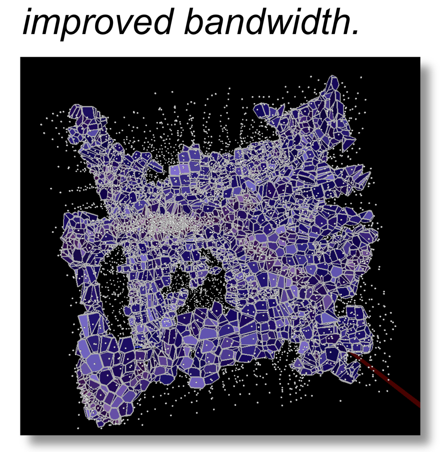
Checkpoint and Analysis I/O

Under a common generic I/O interface, we implemented efficient parallel I/O checkpoint / restart based on GLEAN directly into the HACC simulation that delivers performance at full scale, and we are working toward pnetCDF implementations of in situ analysis products.

- Scaled to the entire 768K cores of Mira BG/Q
- Enabled HACC cosmology production runs for two Gordon Bell finalist submissions
- Used in production on BG/Q (Mira) and Cray (Hopper)
- Achieved 160 GB/s for HACC I/O and up to ~10X improvement over the previous I/O mechanism on Mira
- Written and read ~10 PB of data on Mira (and counting)
- Used for all HACC inputs and outputs of production runs
- Initial lossless data compression work for I/O completed, and custom pre-conditioner development in progress.



Above: topology-aware data movement for I/O for smarter aggregation and improved bandwidth.



Above: analysis products such as Voronoi tessellations are now stored in pnetCDF.

SDAV
Scalable Data Management, Analysis, and Visualization

Tom Peterka, Venkat Vishwanath, Joe Insley, Juliana Kwan, Adrian Pope, Hal Finkel, Katrin Heitmann, Salman Habib (ANL), Jon Woodring, Chris Sewell, Jim Ahrens (LANL), George Zagaris, Robert Maynard, Berk Geveci (Kitware), Wei-keng Liao, Mostofa Potwary, Ankit Agrawal, Saba Sehrish, Alok Choudhary (NU)