The Scalable Data Management, Analysis, and Visualization Institute http://sdav-scidac.org

# Meshing the Universe: Integrating Analysis in Cosmological Simulations

## Motivation

Mesh tessellations are indispensible tools for analyzing point data because they transform sparse discrete samples into dense continuous functions. Meshing the output of petascale simulations, however, can be as data-intensive as the simulations themselves and often must be executed in parallel on the same supercomputers in order to fit in memory. To date, however, no general-purpose large-scale parallel tessellation tools exist.

## Objective

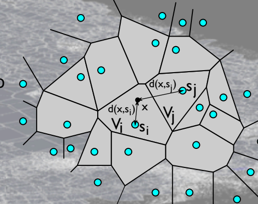
We present a method for computing a Voronoi tessellation in situ during the simulation of cosmology. In principle, similar methods can be applied to other computational geometry problems such as Delaunay tetrahedralization and convex hull in other science domains. We demonstrate the utility of our approach as part of an in situ cosmology tools framework that runs various analysis tools at selected timesteps, saves results to parallel storage, and includes visualization and further analysis in ParaView. For example, connected components of Voronoi cells are interrogated to detect and characterize cosmological voids.

## Voronoi Tessellation

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

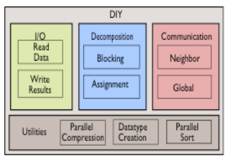
$$V_i = \{ x \mid d(x, s_i) < d(x, s_k) \} \ \forall\ k \neq i$$

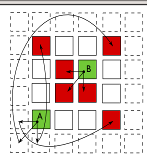
In 2D, Voronoi diagram of polygons, in 3D tessellation of polyhedra. Dual is Delaunay triangulation (2D), tetrahedralization (3D)




## Software

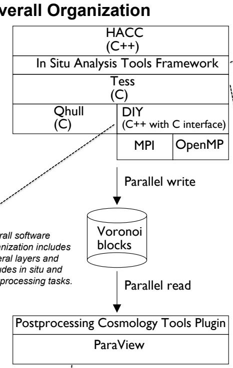
### Parallelism Infrastructure



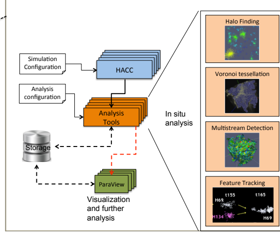

*Top: DIY provides parallel data movement functions for tess. Above: Neighborhood communication with periodic boundary conditions and neighbors that are near enough to a target point.*

### Overall Organization



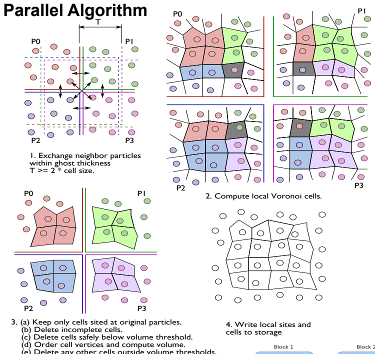

*Overall software organization includes several layers and includes in situ and postprocessing tasks.*

### In Situ Analysis Framework

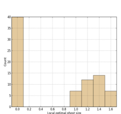


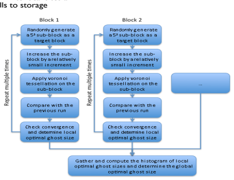

*Left: Framework enables in situ analyses at selected time steps. Analysis filters include halo finding, multistream feature classification, feature tracking, and Voronoi tessellation. The framework also connects to run-time or postprocessing visualization tools.*

### Parallel Algorithm




*Left: 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. Gray-colored cells are exchanged by multiple processes*

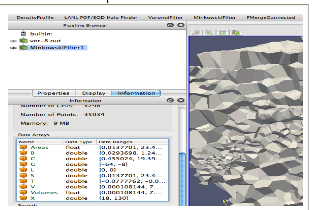



*Right: Algorithm for finding ghost size automatically. Left: Histogram of local optimal ghost sizes shows 95% percentile results in a global optimal ghost size of 1.6.*

### Postprocessing

ParaView cosmology tools plugin and toolbar includes parallel reader for Voronoi tessellations and Minkowski functionals over connected components of Voronoi cells.




*Above: Cosmology tools plugin in ParaView promotes interactive feature exploration.*

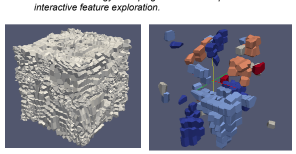
<u>Minkowski Functionals</u>

*Basic*

- Volume
- Surface area
- Extrinsic curvature
- Genus

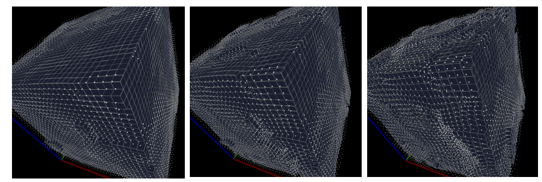
*Derived*

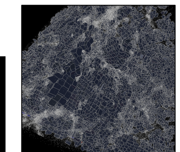
- Thickness
- Breadth
- Length



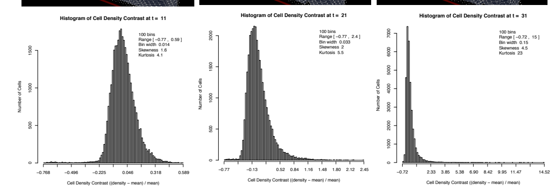
*Right: Connected components of Voronoi cells that have been filtered on cell volume are further characterized according to their Minkowski functionals.*

## Results





*Above: Voronoi tessellation of dark matter tracer particles reveals multiscale structures in the distribution of large and small cells.*

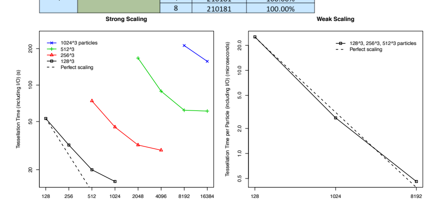



*Left: Evolving Voronoi cells at three time steps during the simulation (top) and corresponding density contrast distribution (bottom). Both visually and statistically, trends are consistent with the formation of ;large-scale structures such as halos and voids.*

| 64³ Particles | | | | |
|---|---|---|---|---|
| Ghost size | # Cells in serial version | # Blocks | # Identical Cells | Accuracy |
| 0 | 210181 | 2 | 201952 | 96.08% |
| | | 4 | 196803 | 93.64% |
| | | 8 | 192140 | 91.42% |
| 1 | | 2 | 209367 | 99.61% |
| | | 4 | 208564 | 99.23% |
| | | 8 | 207024 | 98.50% |
| 2 | | 2 | 210176 | 100.00% |
| | | 4 | 210155 | 99.99% |
| | | 8 | 210012 | 99.92% |
| 3 | | 2 | 210181 | 100.00% |
| | | 4 | 210180 | 100.00% |
| | | 8 | 210180 | 100.00% |
| 4 | | 2 | 210181 | 100.00% |
| | | 4 | 210181 | 100.00% |
| | | 8 | 210181 | 100.00% |

*Left: Accuracy of parallelism with varying number of blocks is compared to the serial version with all particles in one block. Accuracy improves with increasing ghost size until the ghost region is sufficient. We can now determine the optimal ghost size automatically.*

*Bottom: Strong scaling (left) and weak scaling (center) are plotted on a log-log scale. Weak scaling time is normalized by the number of particles, resulting in a downward path whose slope can be compared with the ideal. All graphs represent the total tessellation time, including the time to write the result to storage. The strong scaling efficiency is 41%, and the weak scaling efficiency is 86%. Bottom right: raw performance for in situ tessellation 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) |
|---|---|---|---|---|---|---|---|---|---|
| 128³ | 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 | |
| 256³ | 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 | |
| 512³ | 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 | |
| 1024³ | 25 | 8192 | 2331 | 2119 | 212 | 6 | 186 | 20 | 101 |
| | | 16384 | 1446 | 1289 | 157 | 4 | 113 | 40 | |

SDAV
Scalable Data Management, Analysis, and Visualization

Tom Peterka, Juliana Kwan, Adrian Pope, Hal Finkel, Katrin Heitmann, Salman Habib (ANL),
Jingyuan Wang (UTK), George Zagaris (Kitware)