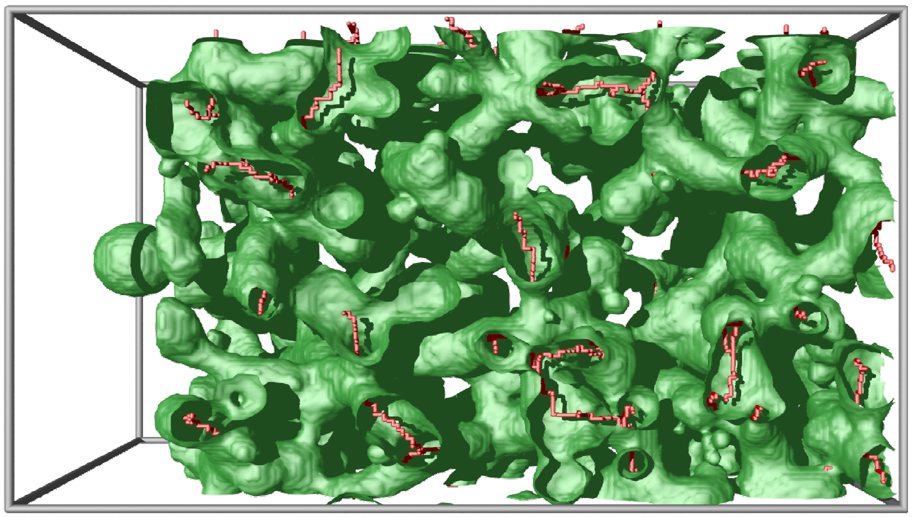
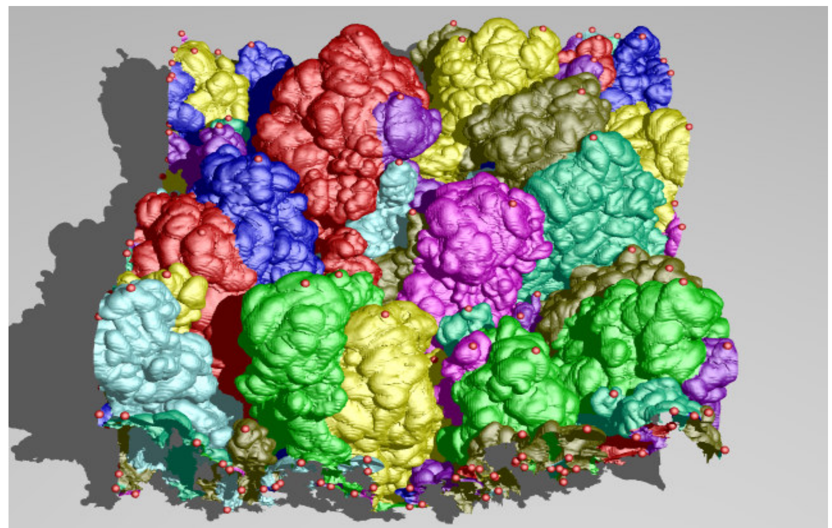
Robust extraction and tracking of topological features in scientific data: state of the art and future challenges





Valerio Pascucci

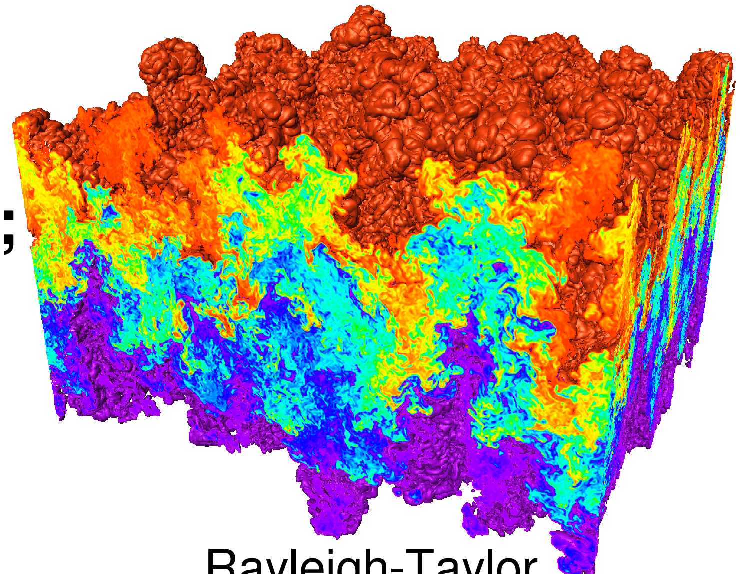
LLNL, UC Davis

CASC

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48. UCRL-PRES-200743

Traditional Data Analysis Tools Are Often Ineffective for Massive Models

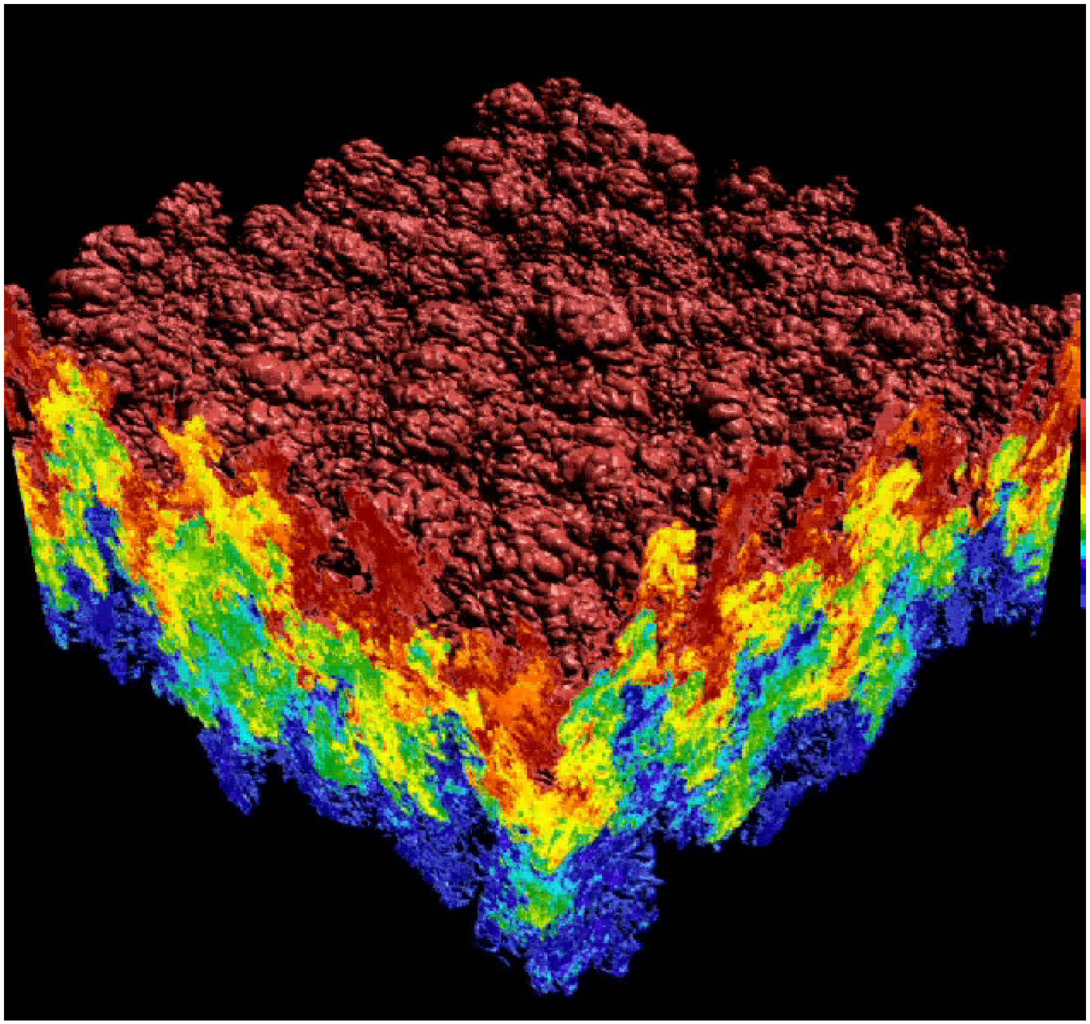
- Massive scientific models are challenging:
  - sheer volume of information;
  - complexity of the information represented.
- Tools do not scale with the data sizes.
- Difficult to capture multiple scales.



Rayleigh-Taylor instability (Miranda)

- Numerical methods unstable and sensitivity to noise.
- Difficulty in providing error bounds associated with the coarse scale analysis.
- Lack of a mathematical language makes hard to reproduce results and map them to new definitions.

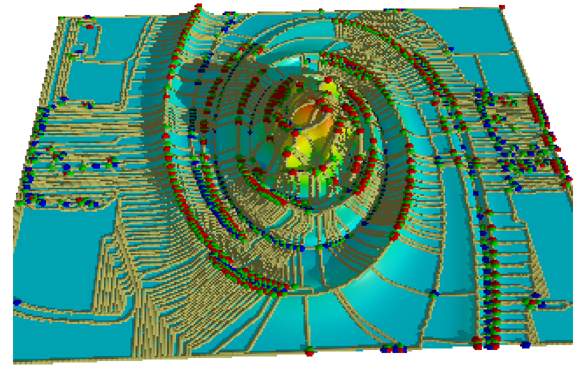
Rayleigh-Taylor Instabilities Arise in Fusion Super-novae, and Other Phenomena

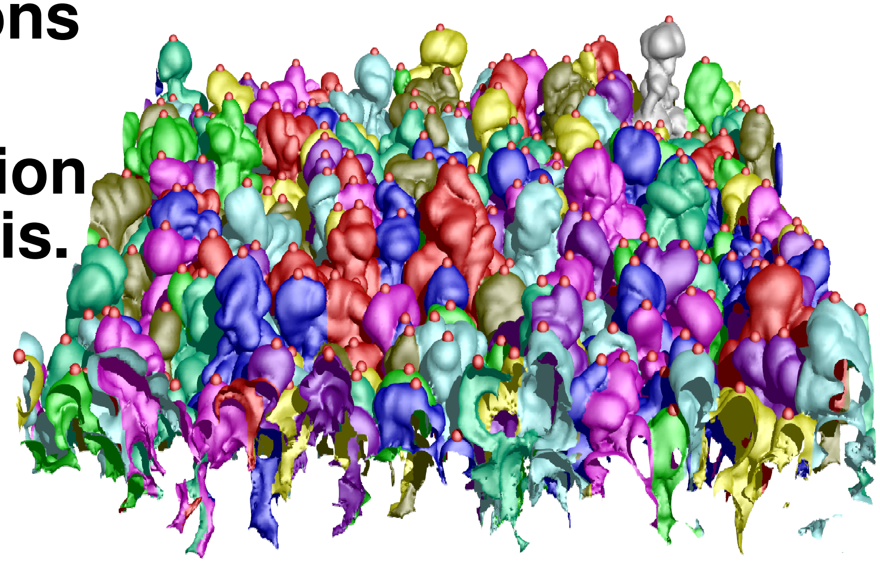


- Rayleigh-Taylor Instability:
  - heavy fluid is above mid-plane, light fluid is below;
  - gravity drives the mixing process;
  - the mixing region lies between the upper envelope surface (red) and the lower envelope surface (blue);
  - 25 to 40 TB of data from simulations.

We Introduced Robust Topological Methods for Quantitative Data Analysis

- **Provably robust computation.**
- **Provably complete feature extraction and quantification.**
- **Hierarchical topological structures used to capture multiple scales.**
- **Error bounded approximations associated with each scale.**
- **Formal mathematical definition associated with each analysis.**
- **Scalable performance in association with streaming techniques.**

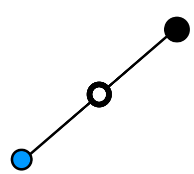




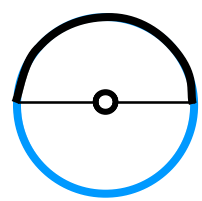
We Rewrote Morse Theory for Provably Robust and Correct Computations

	$f(x): D \rightarrow \Re$ Classical Mathematical Definitions	$F(x): S \rightarrow \Re$ Simulation of Differentiability
domain	D smooth manifold	S simplicial complex
function	f infinitely differentiable	$F(x)$ PL-extension of $f(x_i)$
critical point	$\nabla f(p) = 0$	$LowerLink(p) \neq B^{d-1}$

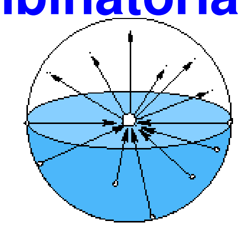
numerical **combinatorial**



1D



2D



3D

Independent local computation yield globally consistent results

We Introduced New Techniques for Critical Point Classification

type	index
Minimum	0
⋮	⋮
Saddle	d-1
Maximum	*d*

$$\nabla f(p) = 0$$

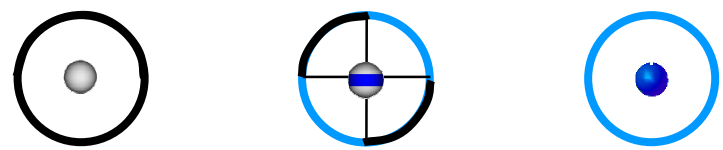
$$f(x)|_p = f(p) + \sum_{i=1}^{d-k} x_i^2 - \sum_{j=d-k+1}^{d} x_j^2$$

numerical

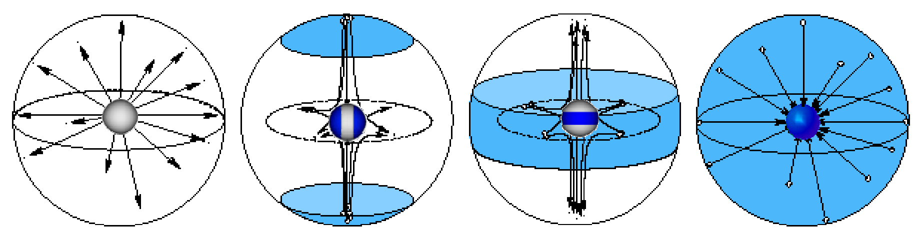
The Morse Lemma

There are *d+1* types of critical points

Maximum Minimum



Minimum Saddle Maximum

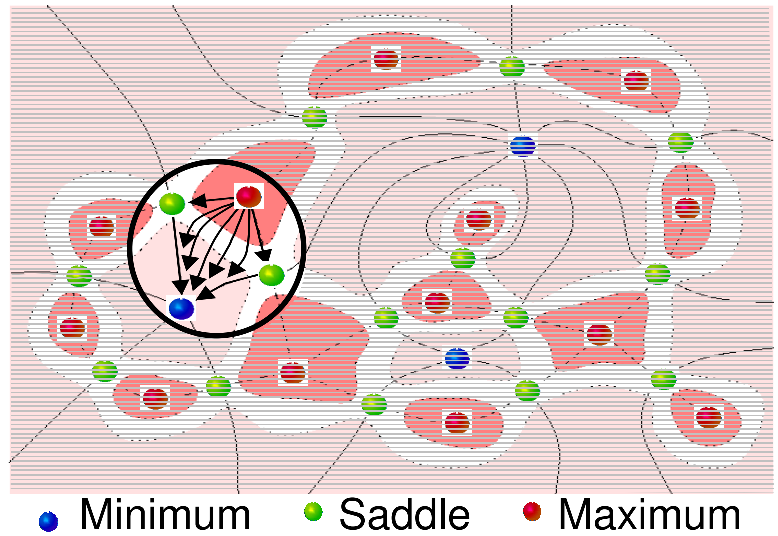


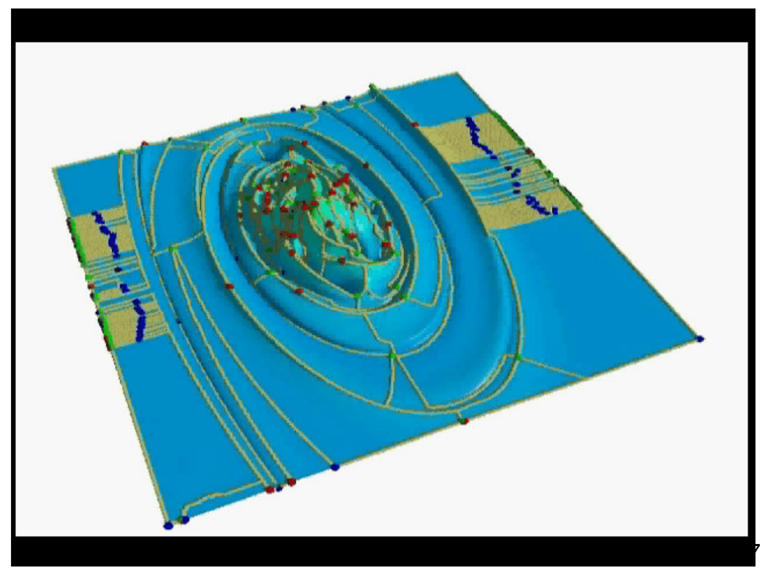
Minimum 1-saddle 2-saddle Maximum

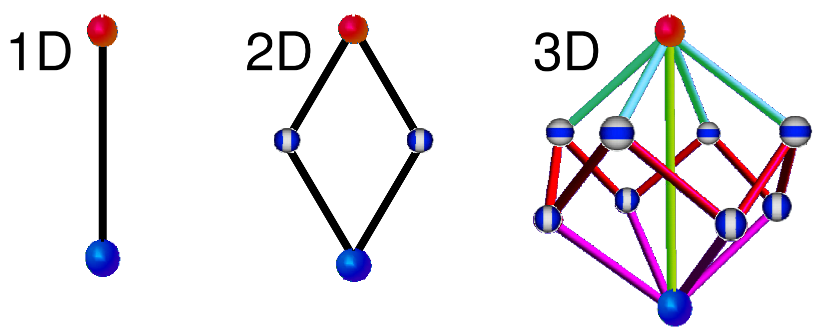
combinatorial

We Introduced the Morse-Smale Complex for Complete Data Analysis

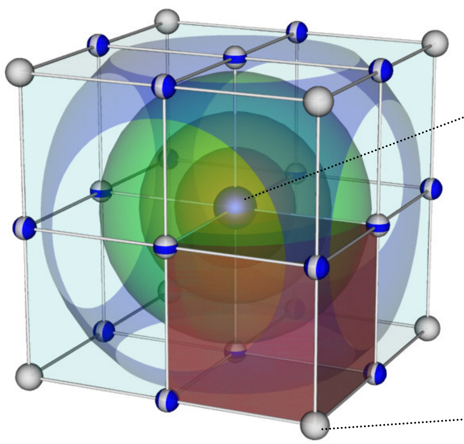
- **The Morse-Smale complex partitions the domain of *f* in regions of uniform gradient.**
- **Generalizes the notion of monotonic interval.**
- **Dimension of a region equal index difference of source and destination.**
- **Remove inconsistency of local gradient evaluations.**

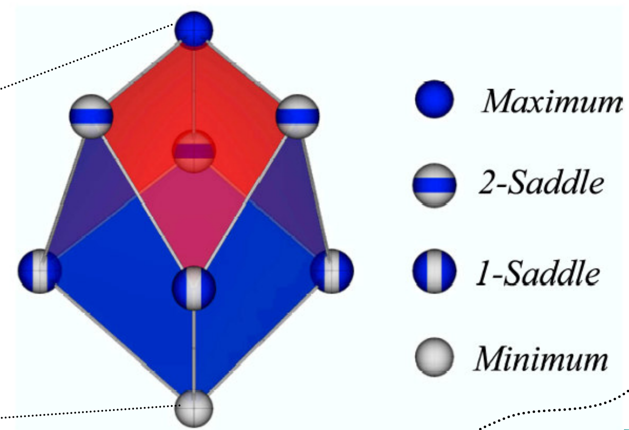


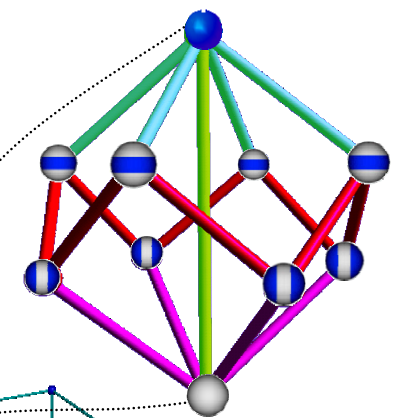




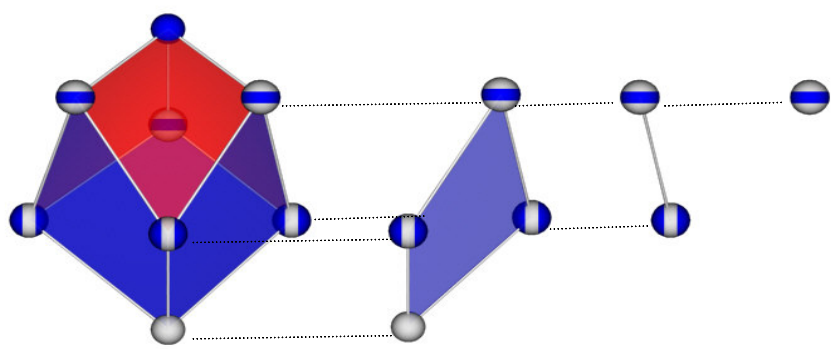
We use ascending/descending manifolds to characterize the cells of the complex.

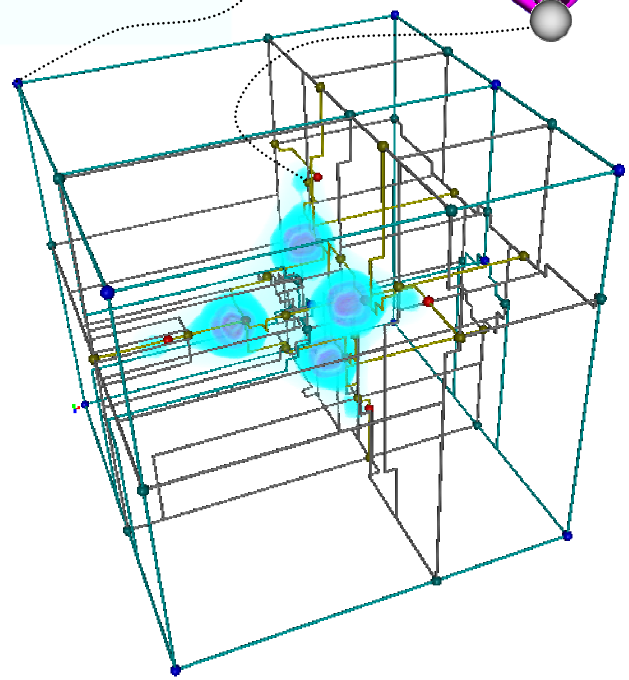






Cells of dimension i (*crystals*, *quads*, *arcs*, and *nodes*) connect critical points with index that differ by i.



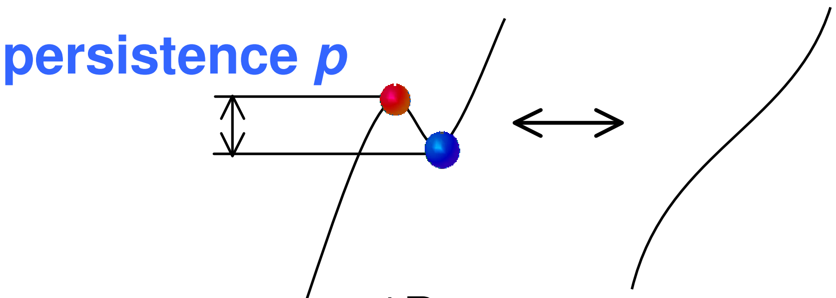


We Map the Index Lemma to a Morse-Smale Complex Simplification

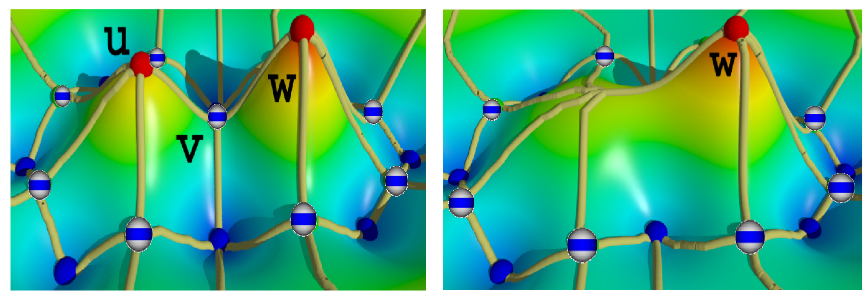
Index Lemma: critical points can be created or destroyed in pairs that differ by one in index.

Approximation: error = persistence/2 (proven lower bound).

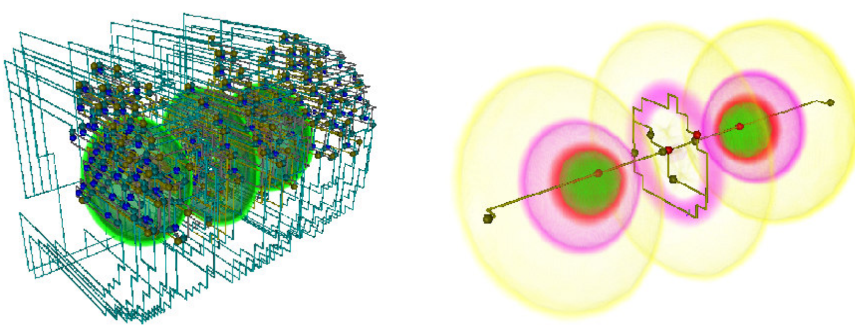
Multi-scale: consistent gradient segmentation at all scales.



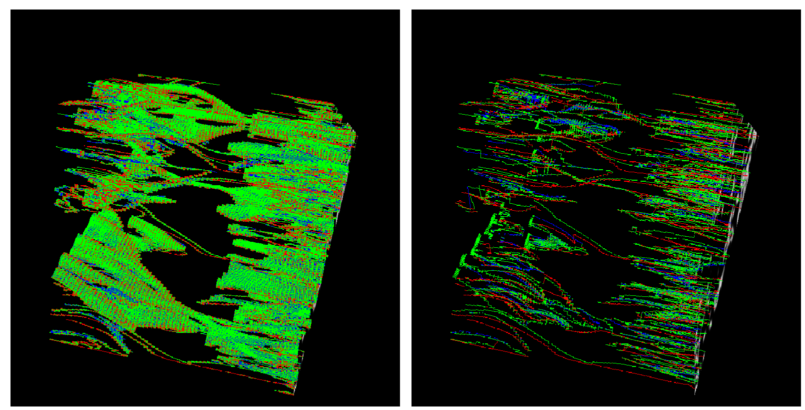
1D



2D

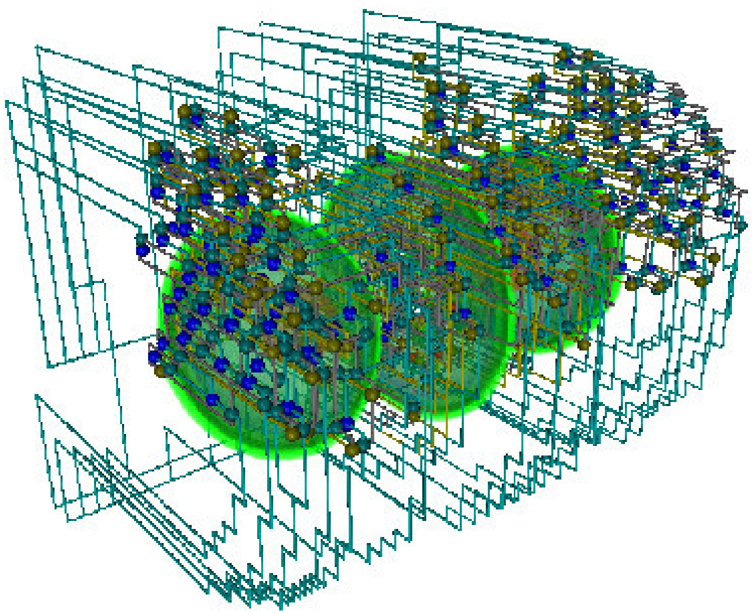


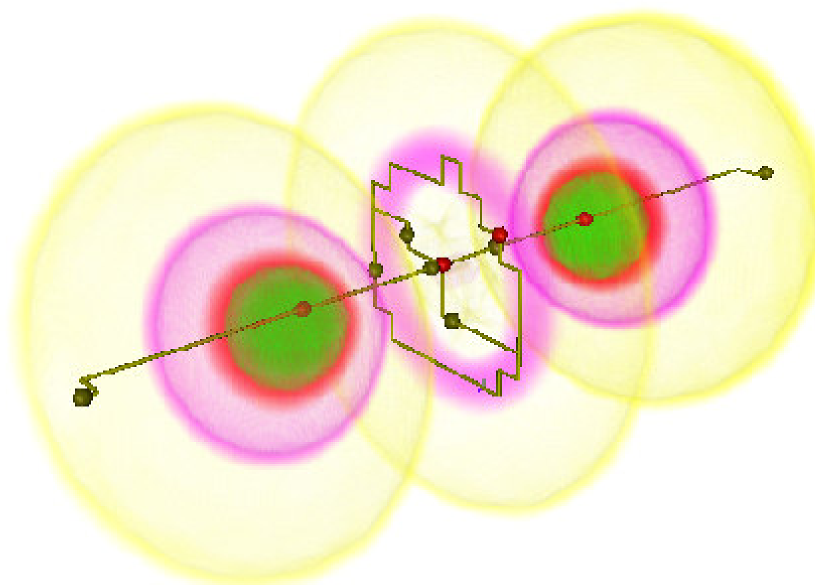
3D



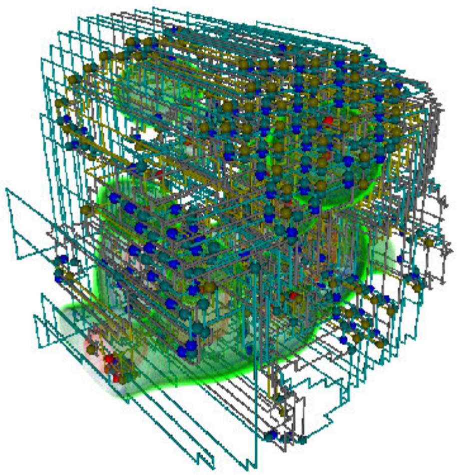
Time

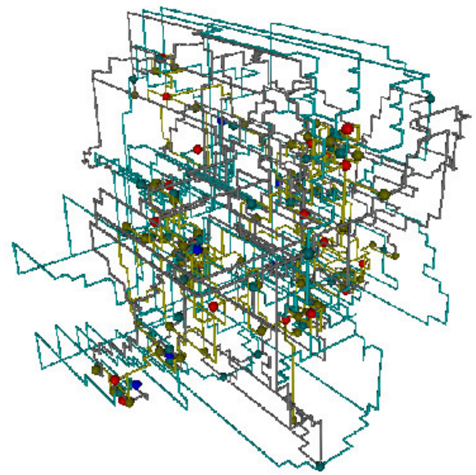
Topological Simplification of the Electron Density Distribution of an Hydrogen Atom

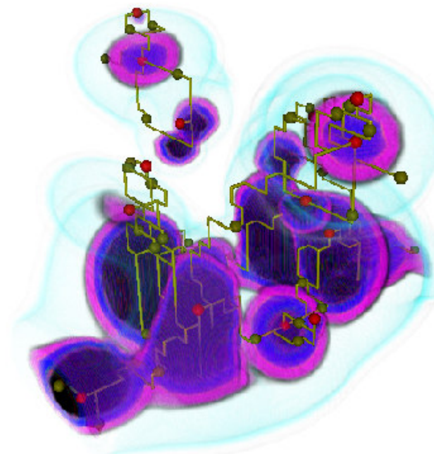




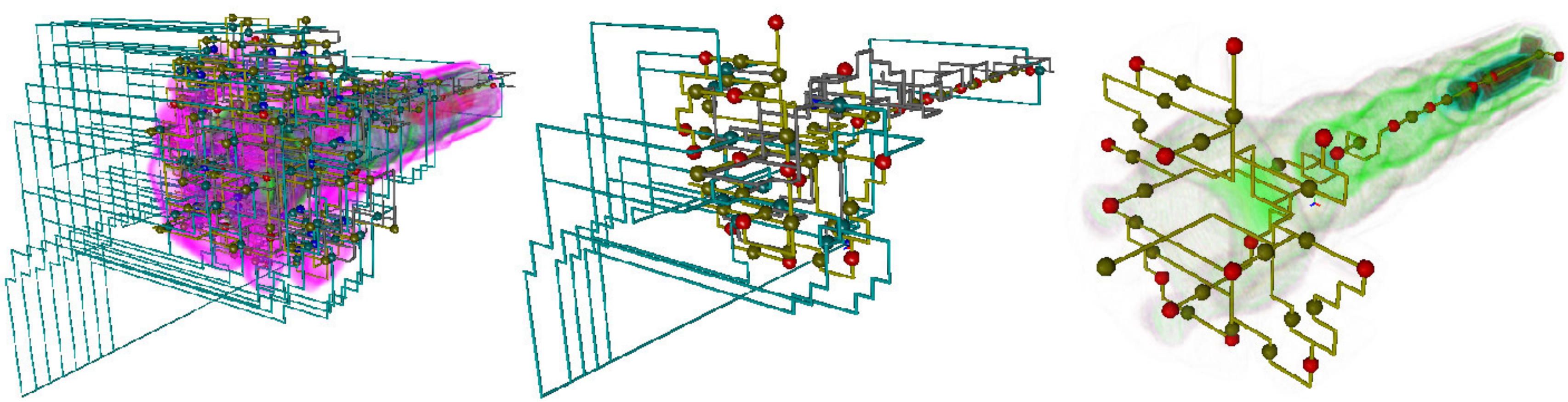
Topological Simplification for the Neghip Dataset



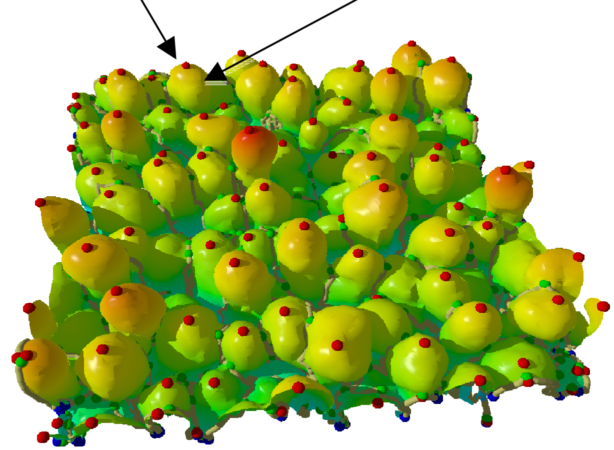




Topological Simplification for the Fuel Injection Dataset



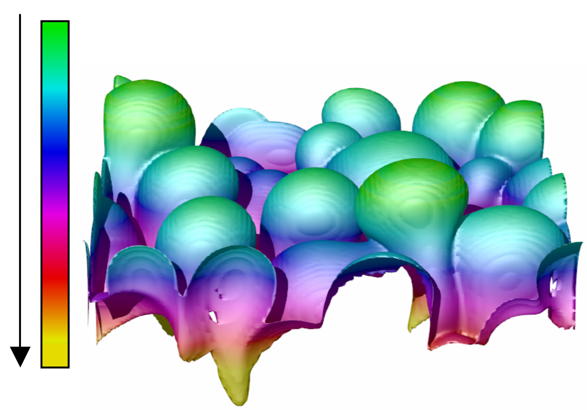
We Analyze High Resolution Rayleigh-Taylor Instability Simulations

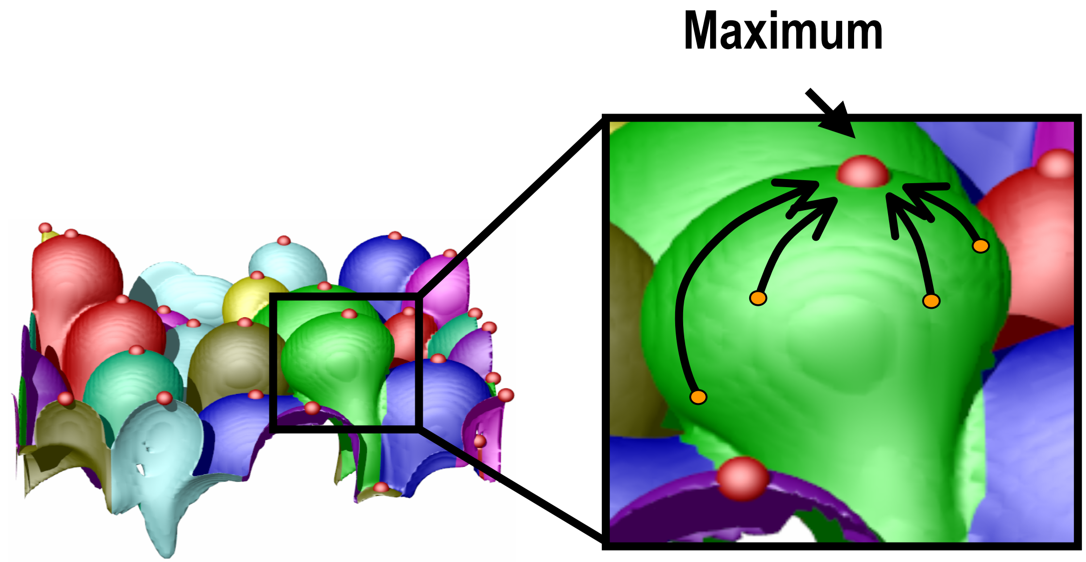


- Large Eddy Simulation run on Linux cluster: 1152 x 1152 x 1152
  - ~ 40 G / dump
  - 759 dumps, about 25 TB
- Direct Numerical Simulation run on BlueGene/L: 3072 x 3072 x Z
  - Z depends on width of mixing layer
  - Over 40 TB
- Bubble-like structures are observed in laboratory and simulations.
- Bubble dynamics are considered and important way to characterize the mixing process:
  - mixing rate $= \partial(\#bubbles)/\partial t.$
- There is no prevalent formal definition of bubbles.

We Compute the Morse-Smale Complex of the Upper Envelope Surface

F(x) = z





F(x) on the surface is aligned against the direction of gravity which drives the flow.

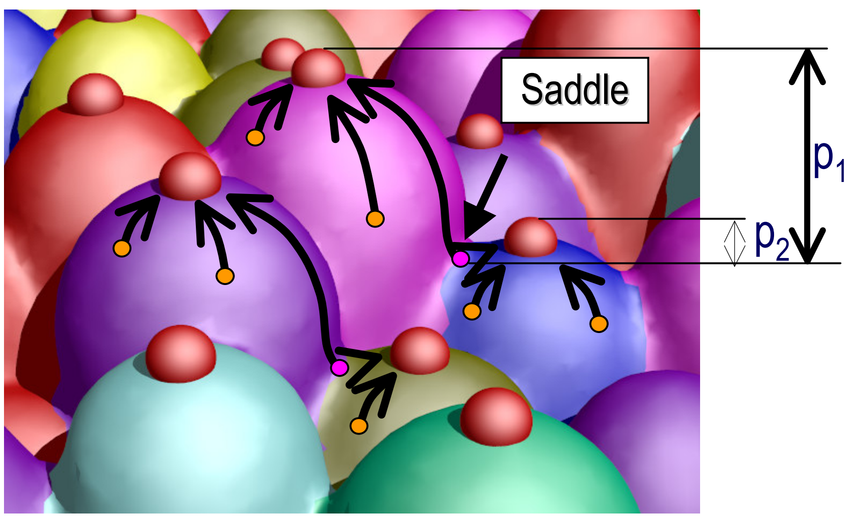
Morse complex cells drawn in distinct colors.

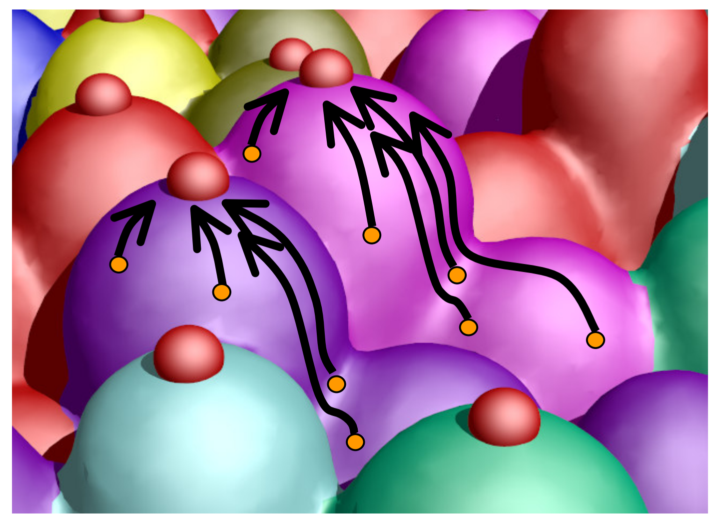
In each Morse complex cell all steepest ascending lines converge to one maximum.

A Hierarchal Model Is Generated by Simplification of Critical Points

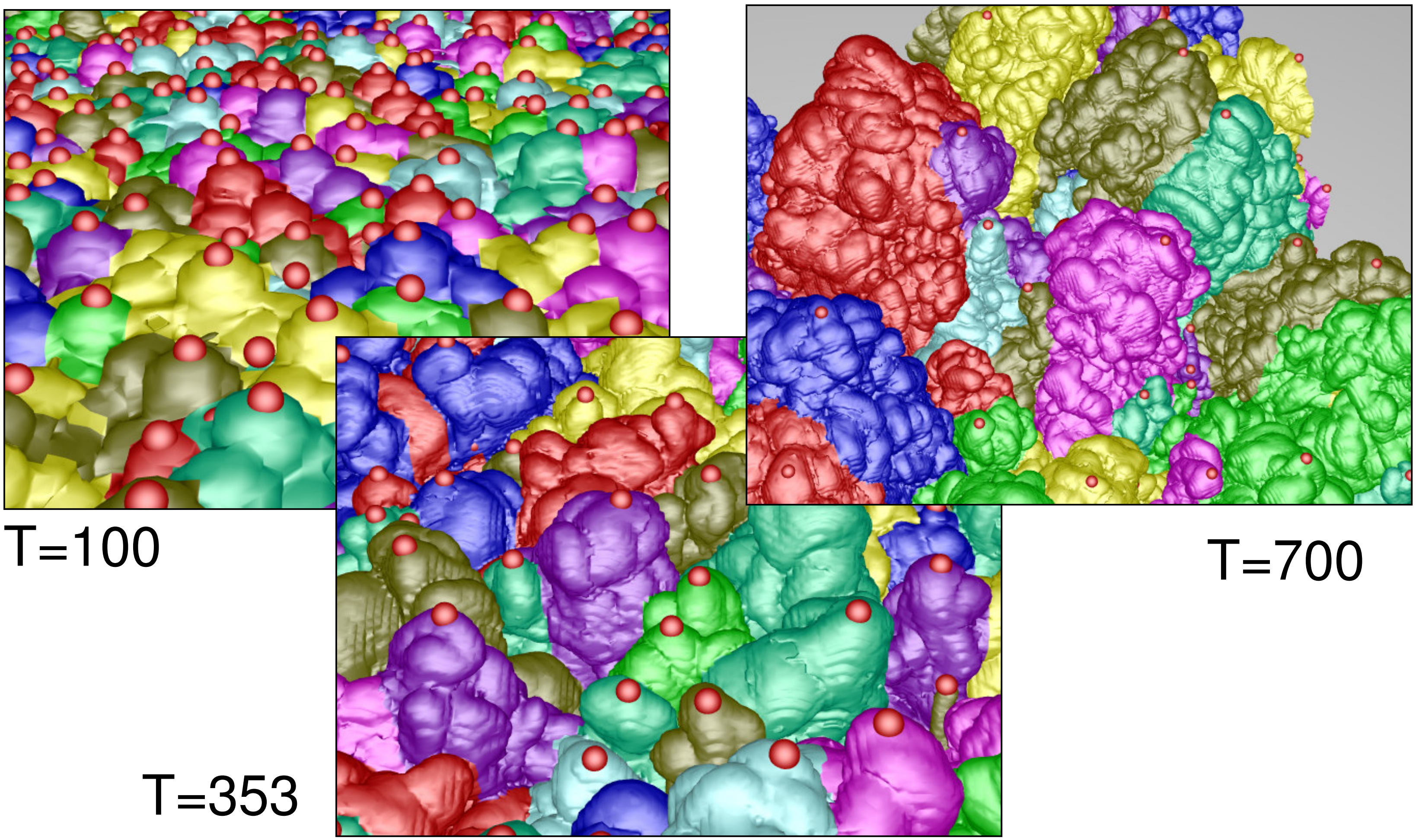
- **Persistence is varied to annihilate pairs of critical points and produce coarser segmentations.**

- **Critical points with higher persistence are preserved at the coarser scales.**





The Segmentation Method is Robust From Early Mixing to Late Turbulence



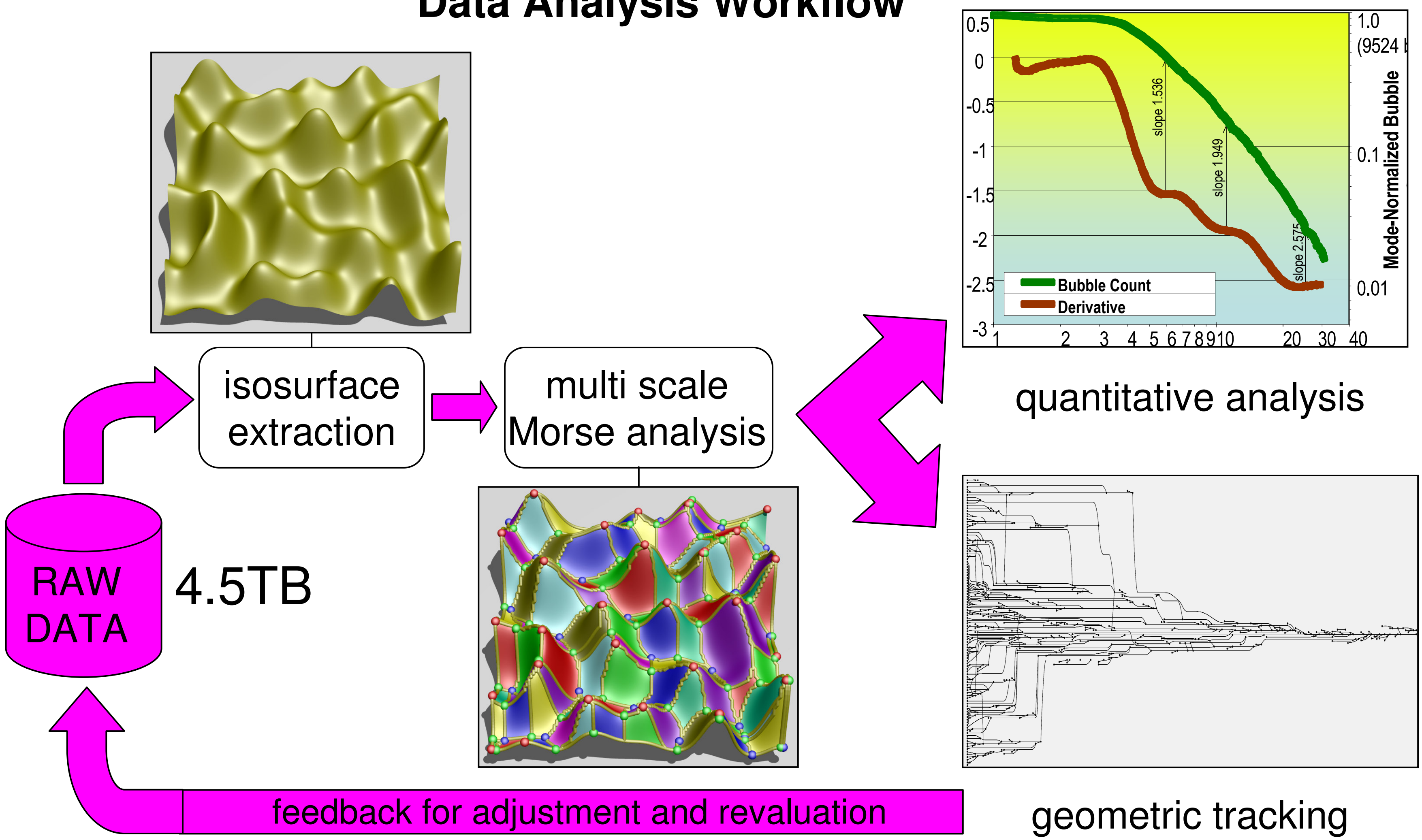
T=100

T=353

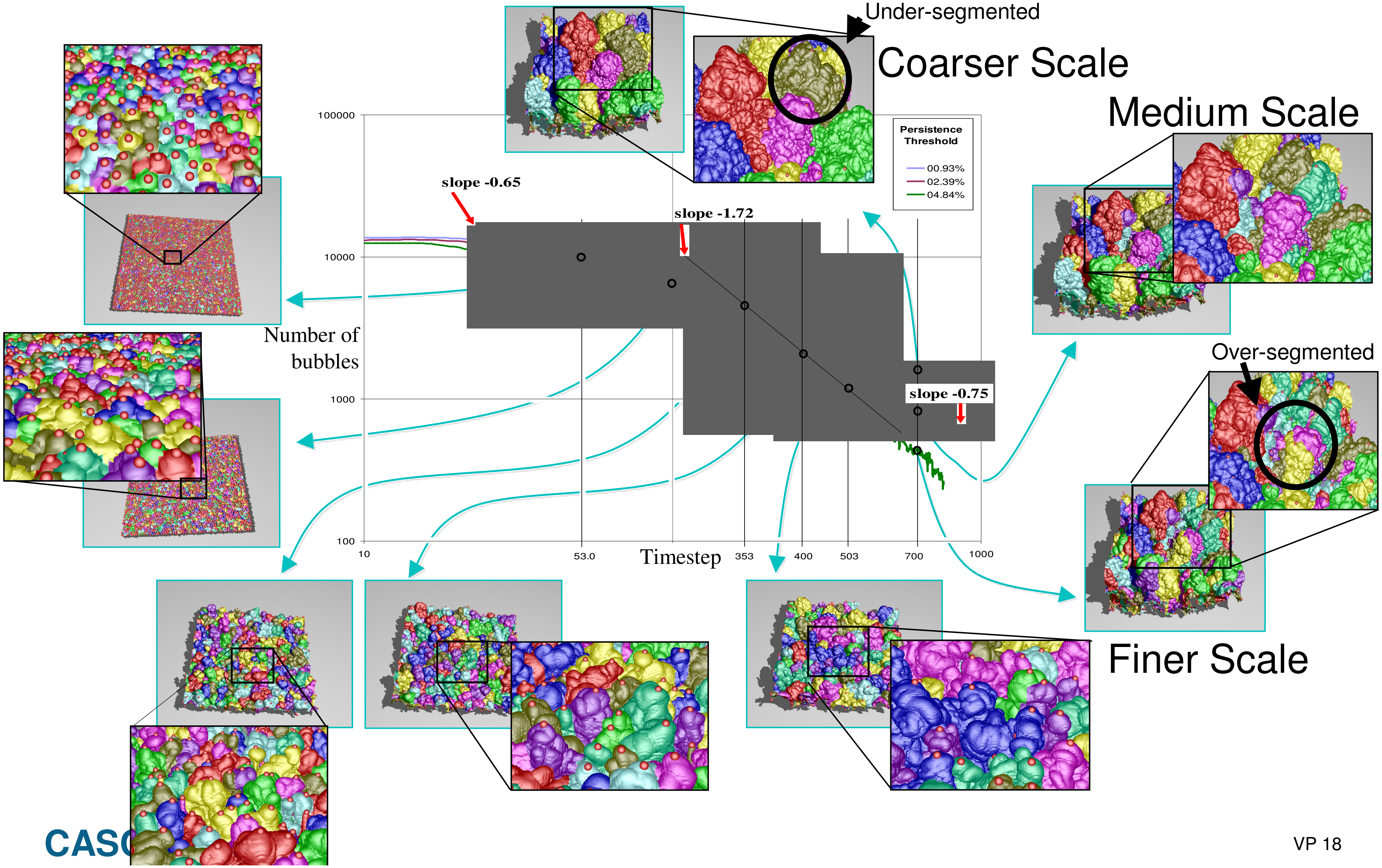
T=700

We Achieve High Performance With New Streaming Data Techniques

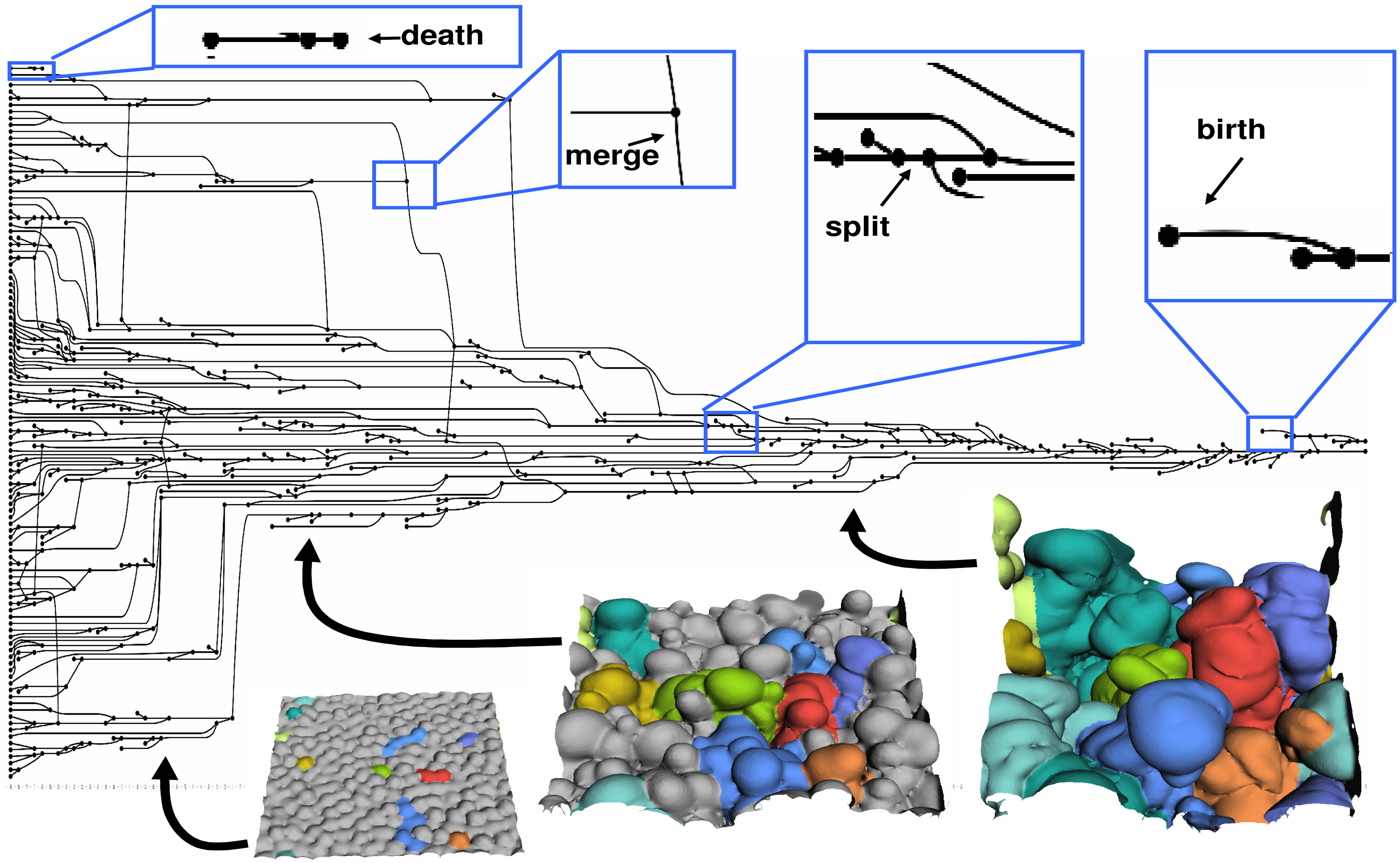
Data Analysis Workflow



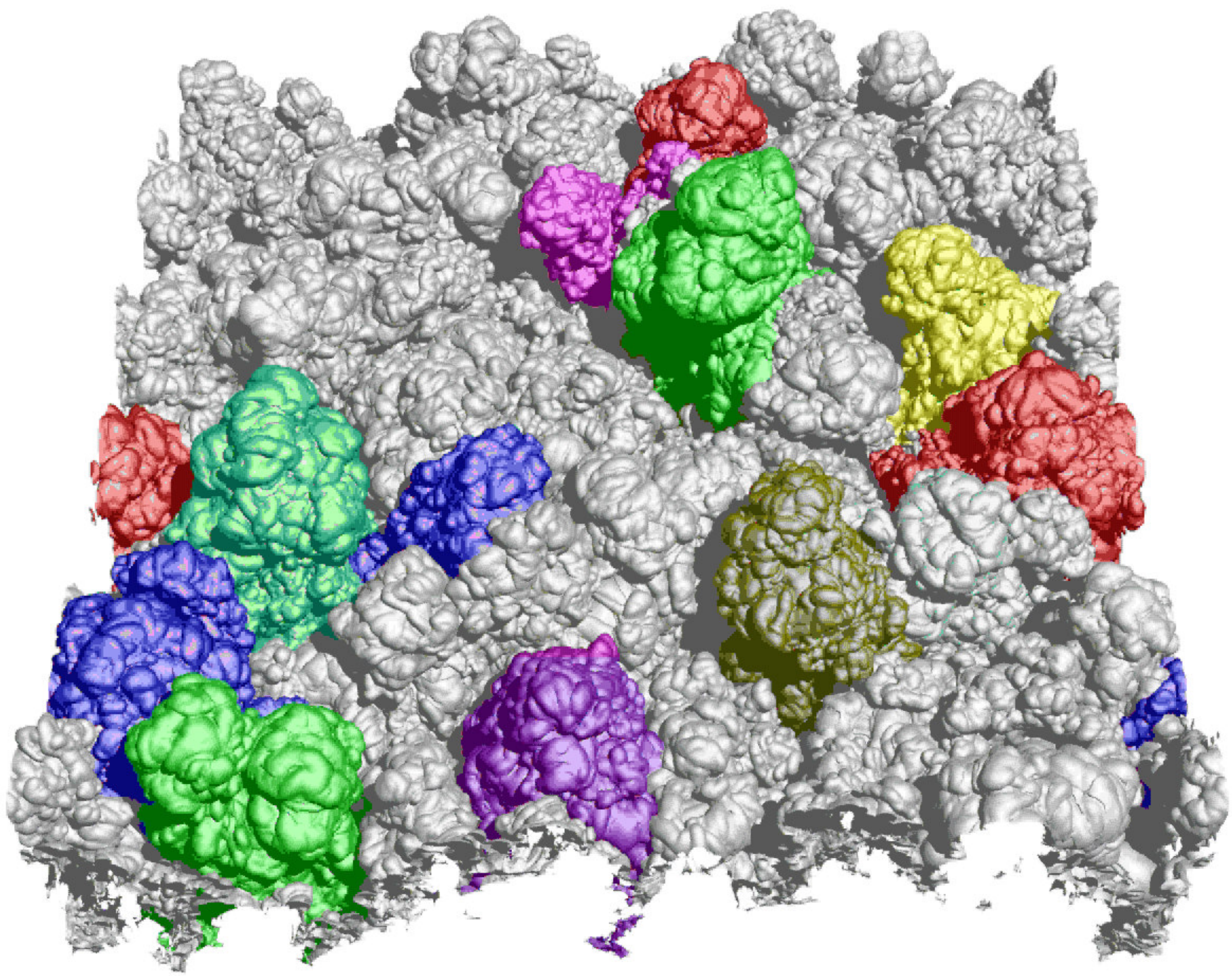
We Evaluated Our Quantitative Analysis at Multiple Scales



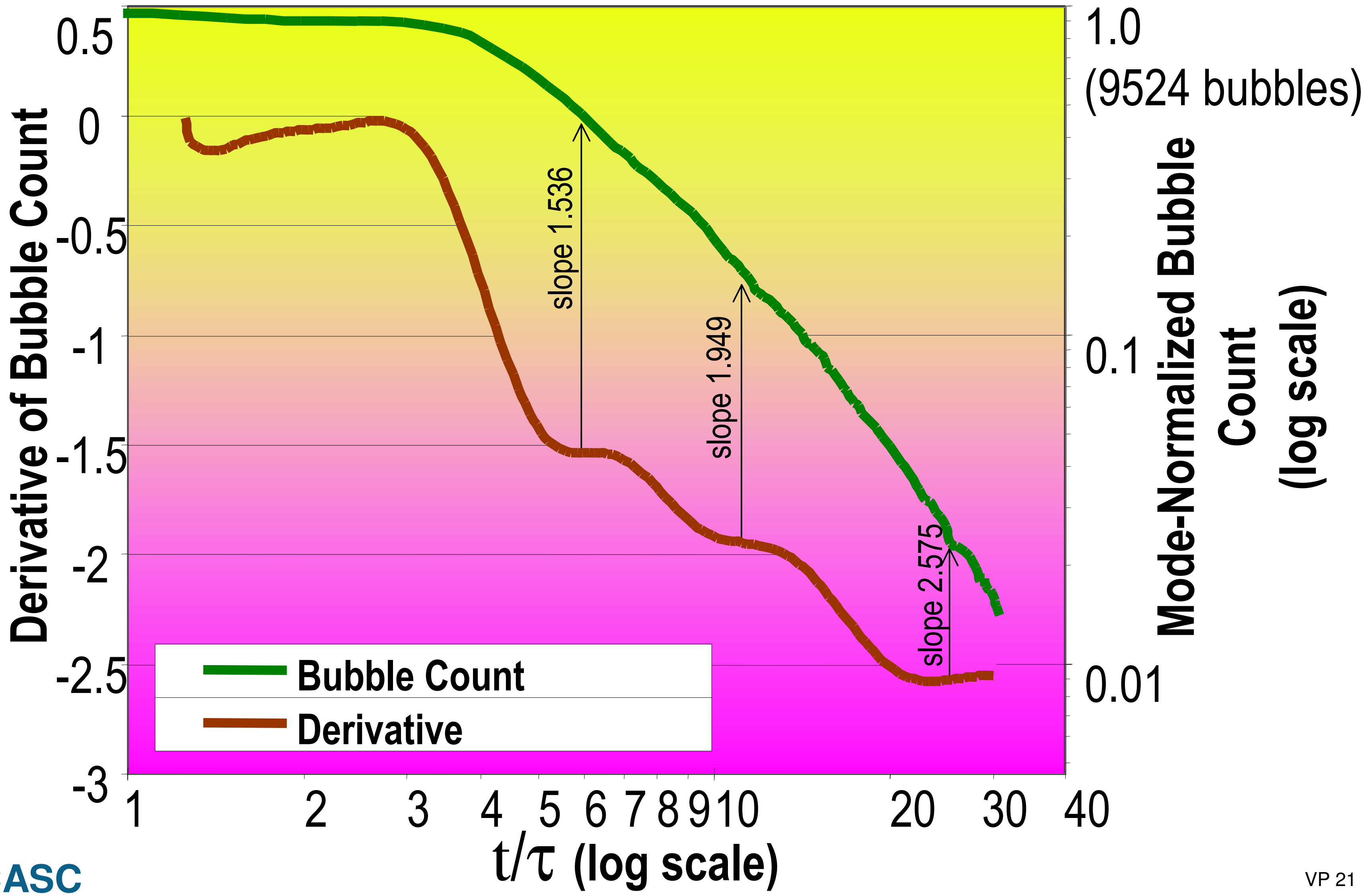
We Characterize the Events That Occur in the Mixing Process



First Robust Bubble Tracking From Beginning to Late Turbulent Stages

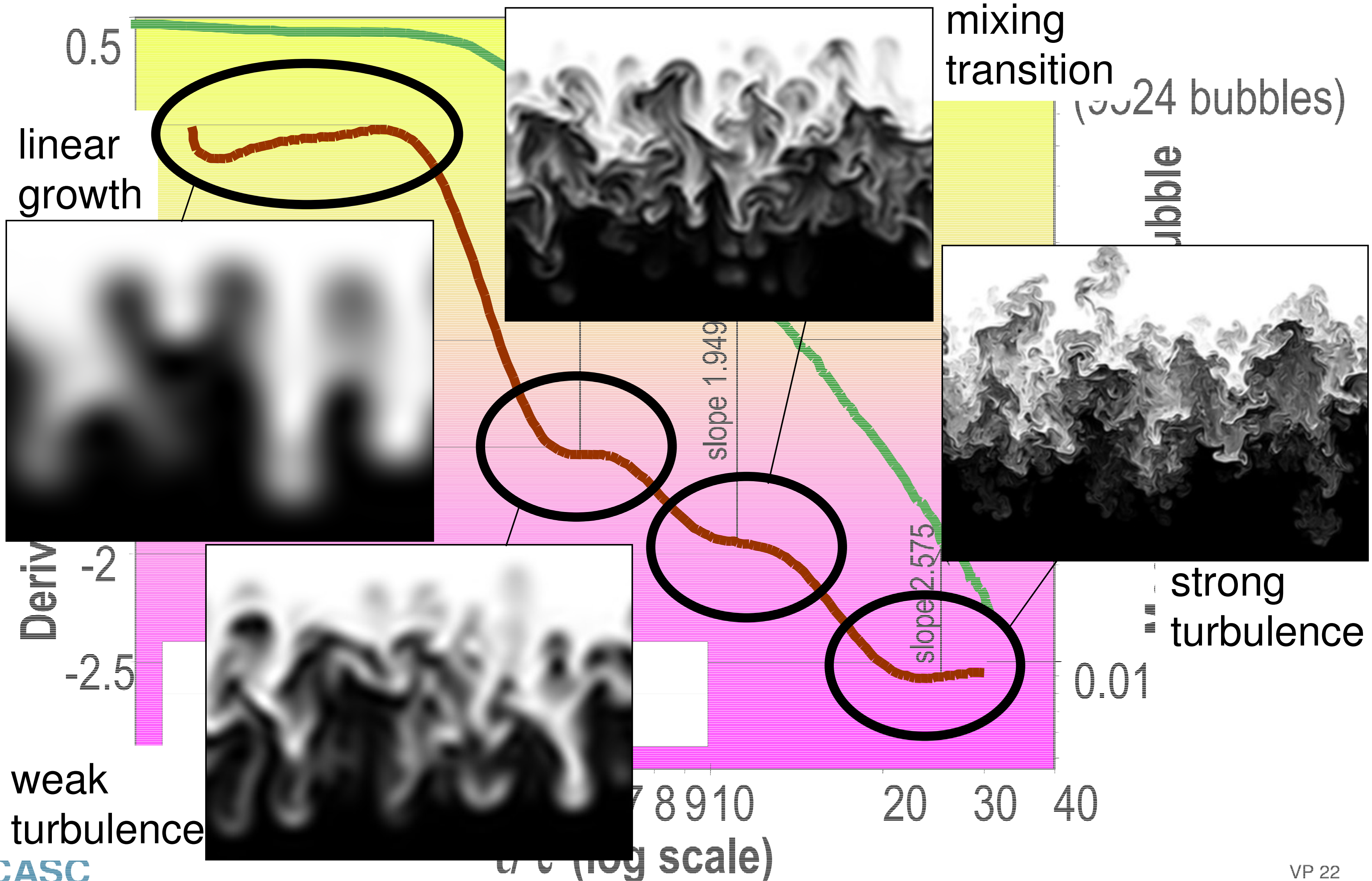


First Time That Scientists Can Quantify Robustly Mixing Rates by Bubble Count



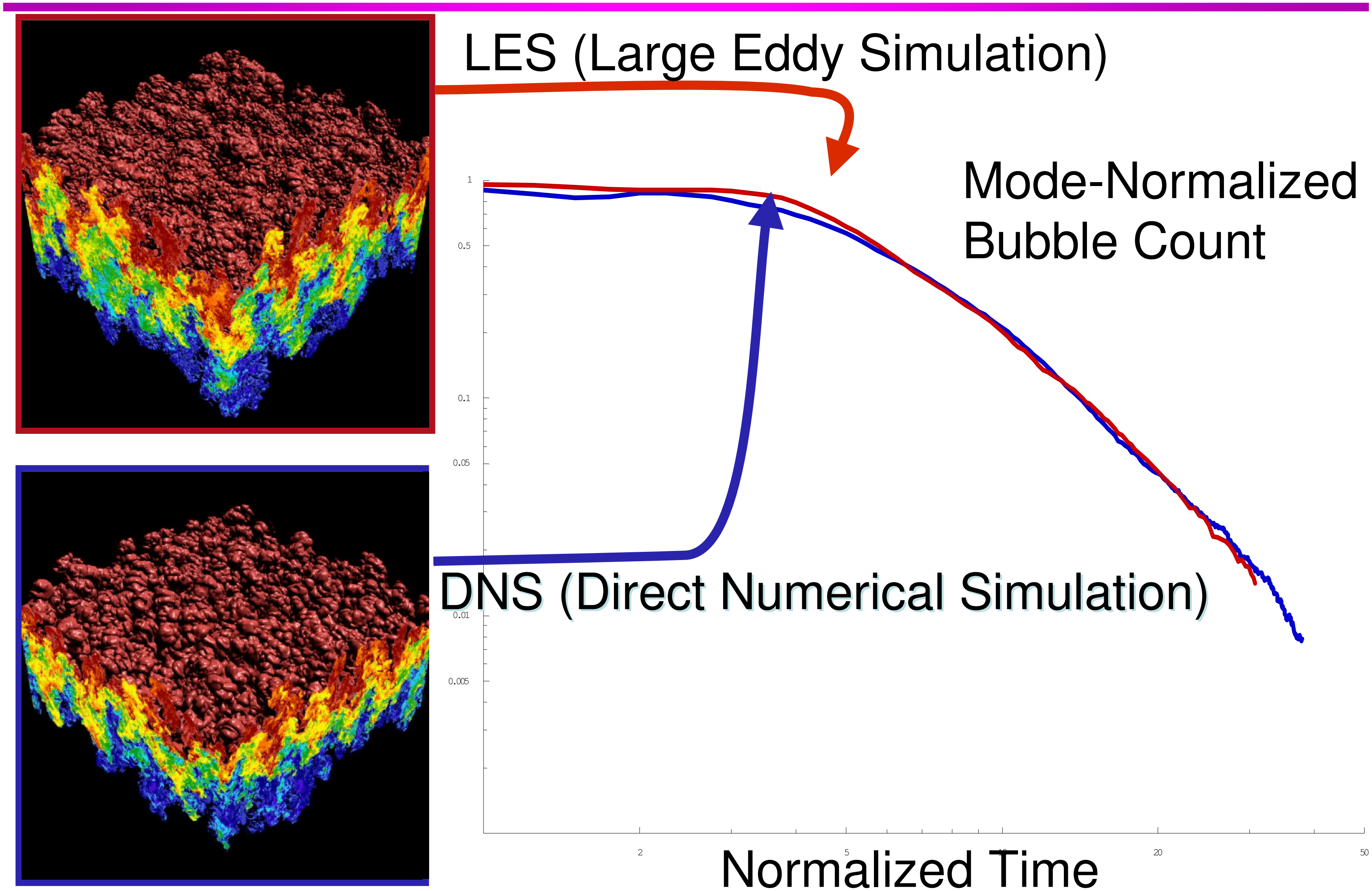
CASC

We Provide the First Quantification of Known Stages of the Mixing Process

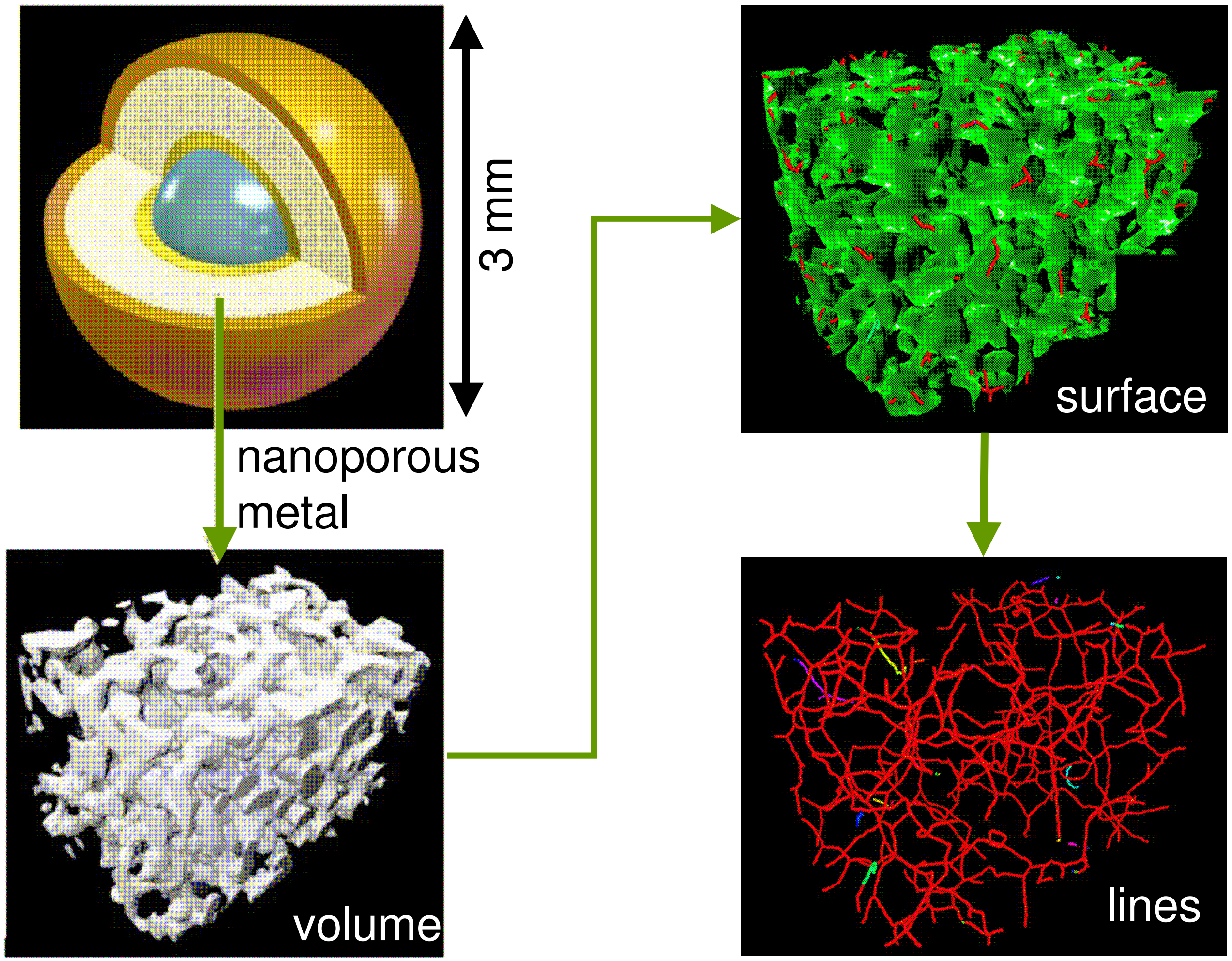


CASC

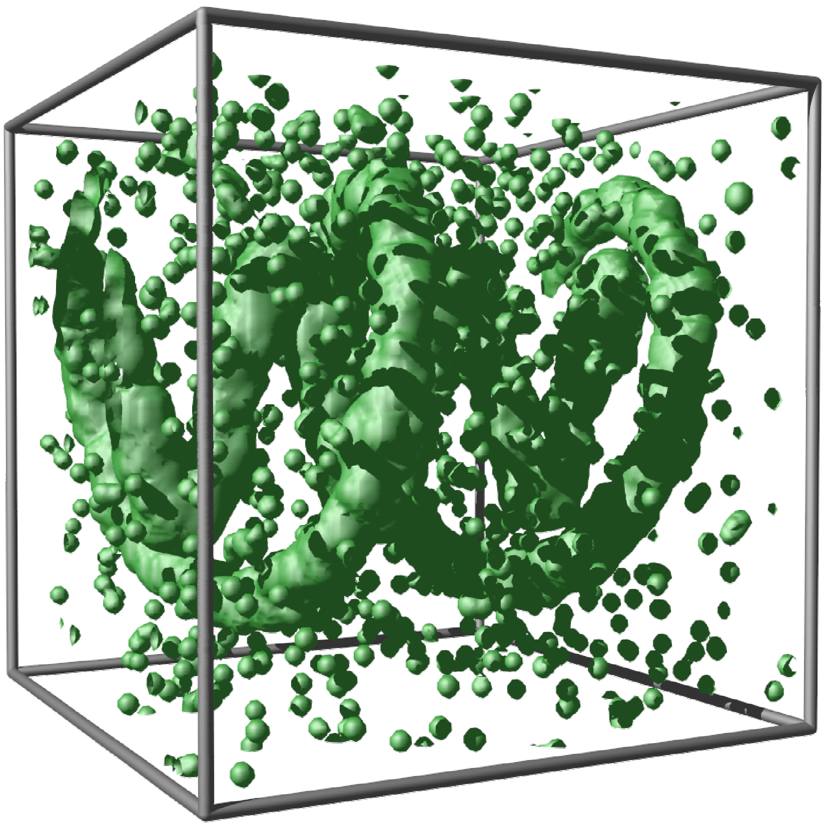
First Feature Based Comparison and Validation of a DNS with a LES

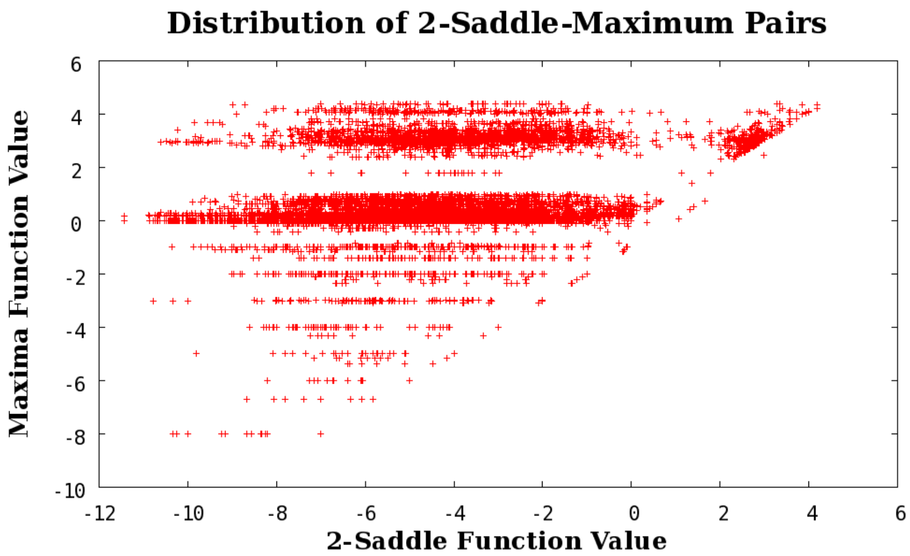


Analysis of the Structural Properties of Nonoporous Metal in the ICF Target

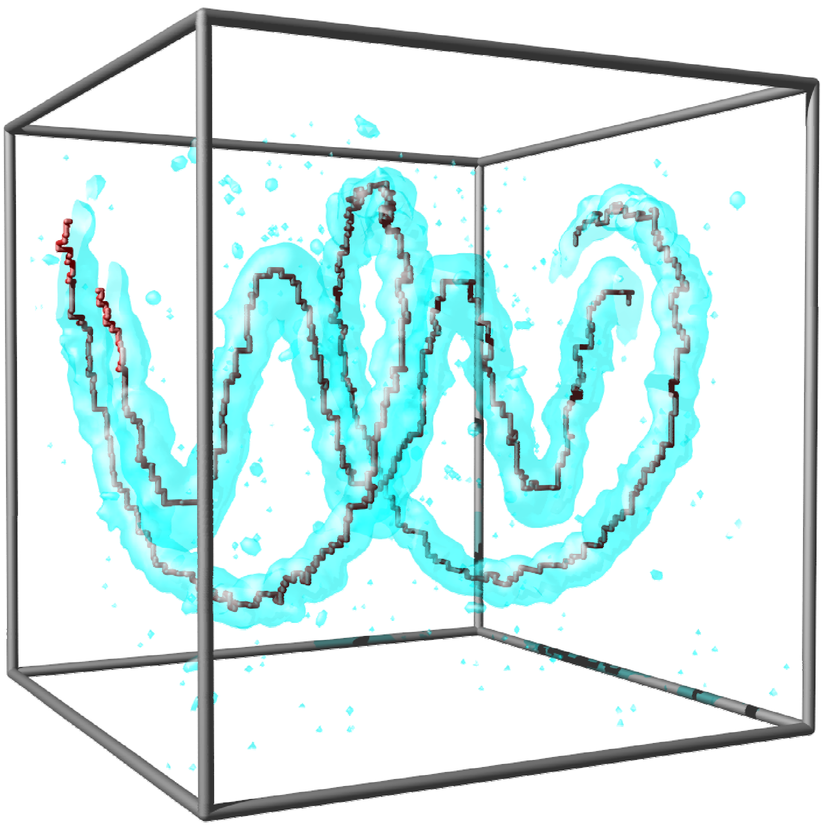


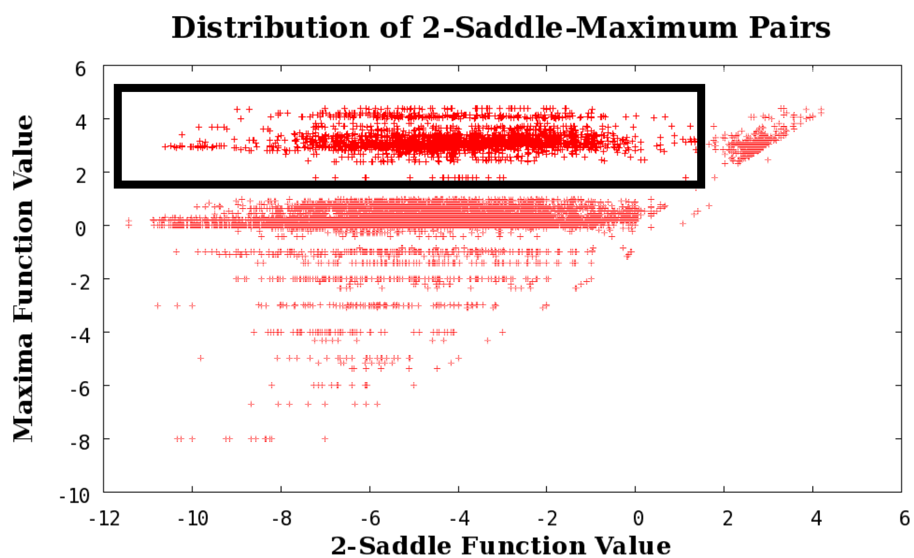
The Hierarchical Morse-Smale Complex Has Very Good Reconstruction Properties





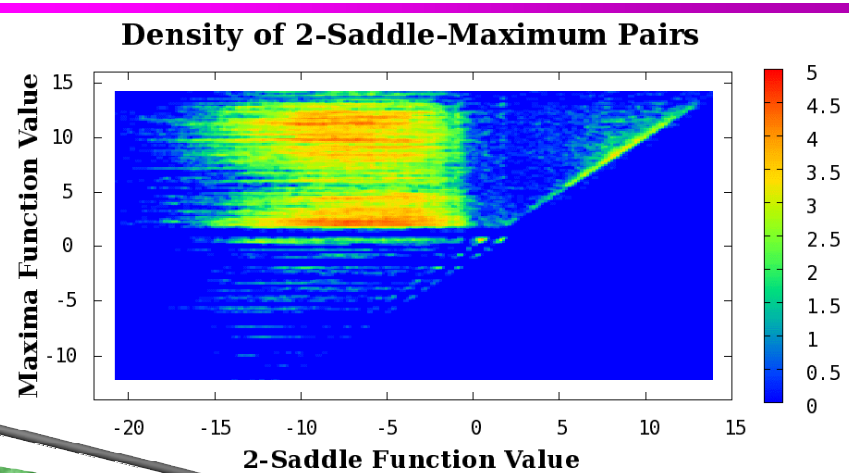
The Hierarchical Morse-Smale Complex Has Very Good Reconstruction Properties

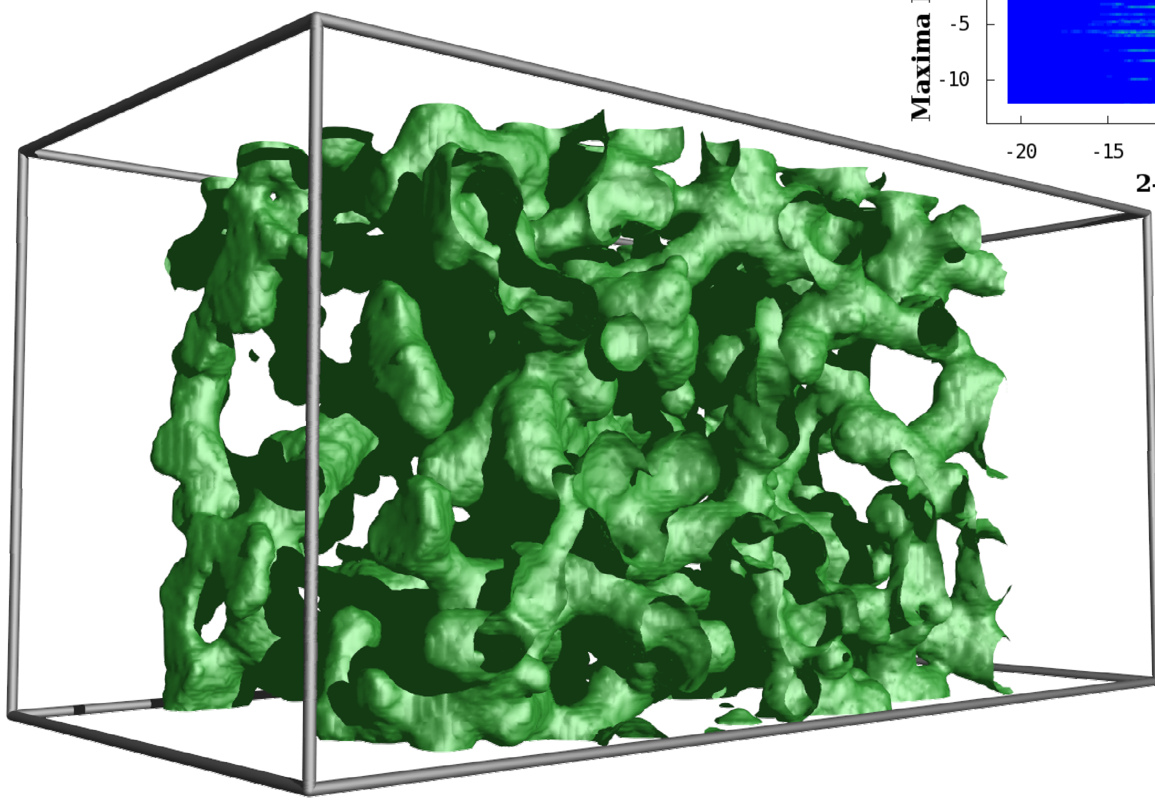




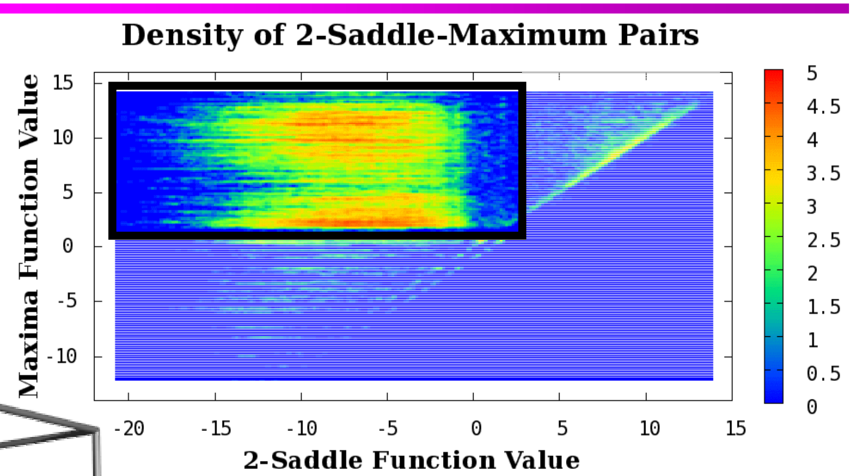
CASC

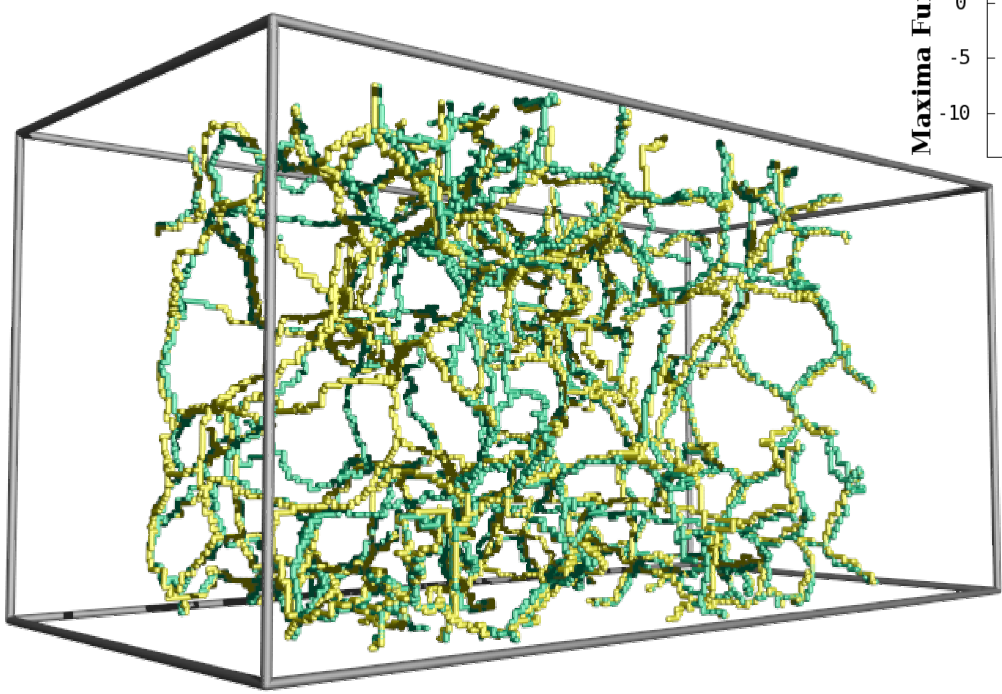
We Apply This Procedure to Nonoporous Metal From Atomistic Simulations

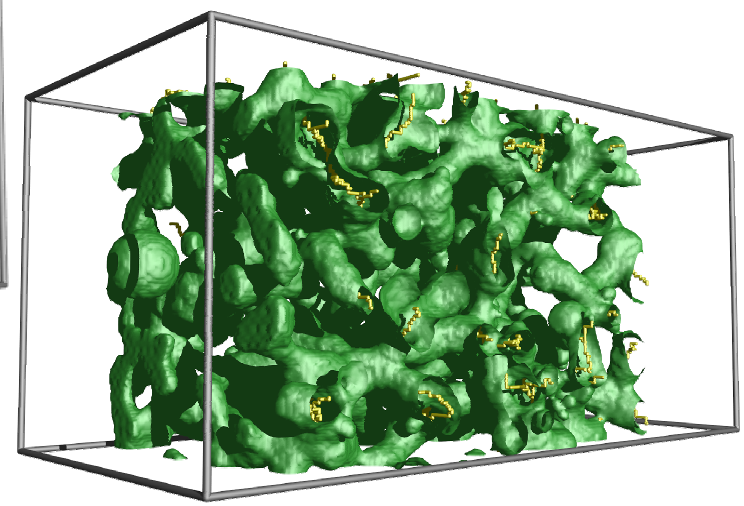




We Apply This Procedure to Nonoporous Metal From Atomistic Simulations

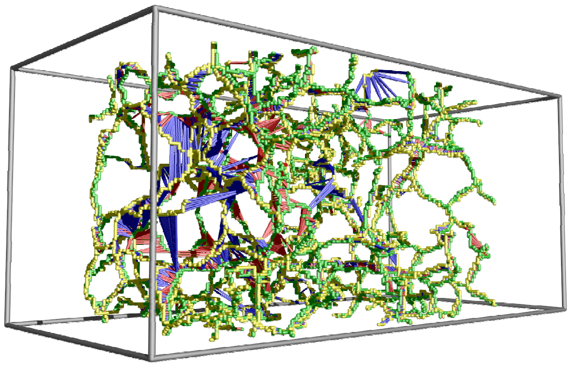


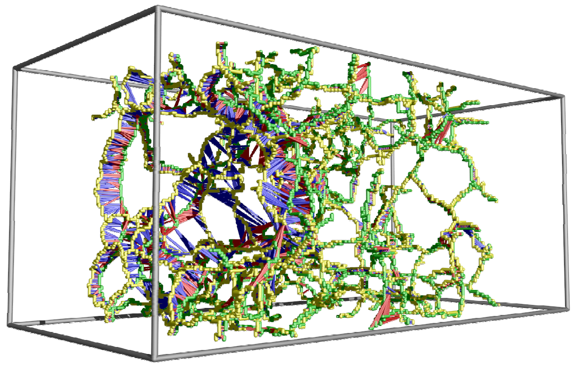


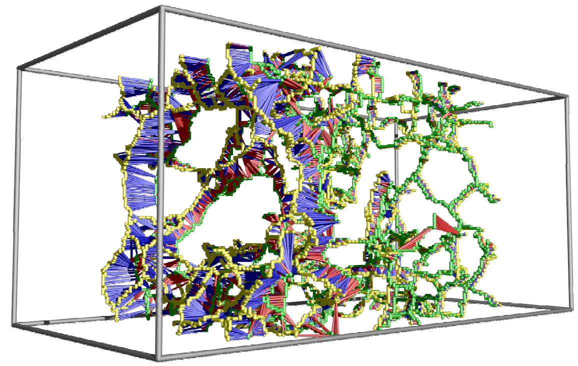


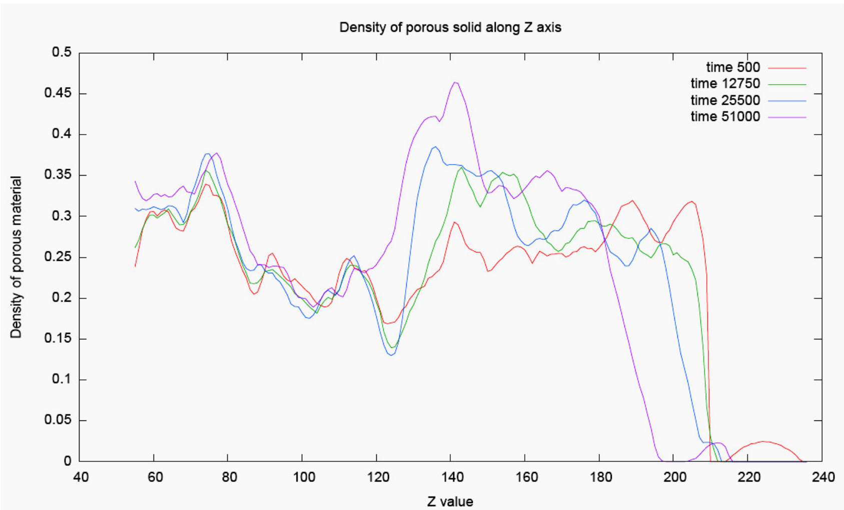
We Obtain a Consistent Reconstruction in Time With Quantitave Density Profiles

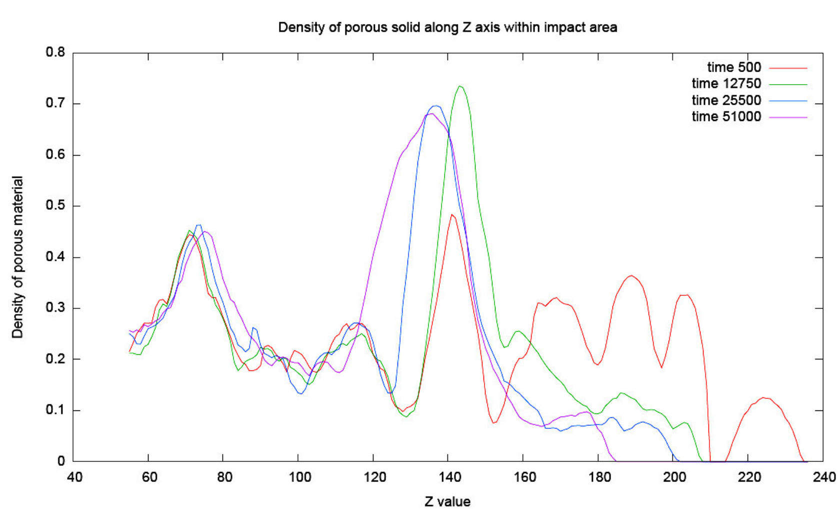
- **Simulation of structural properties during impact with a micrometeoroids.**



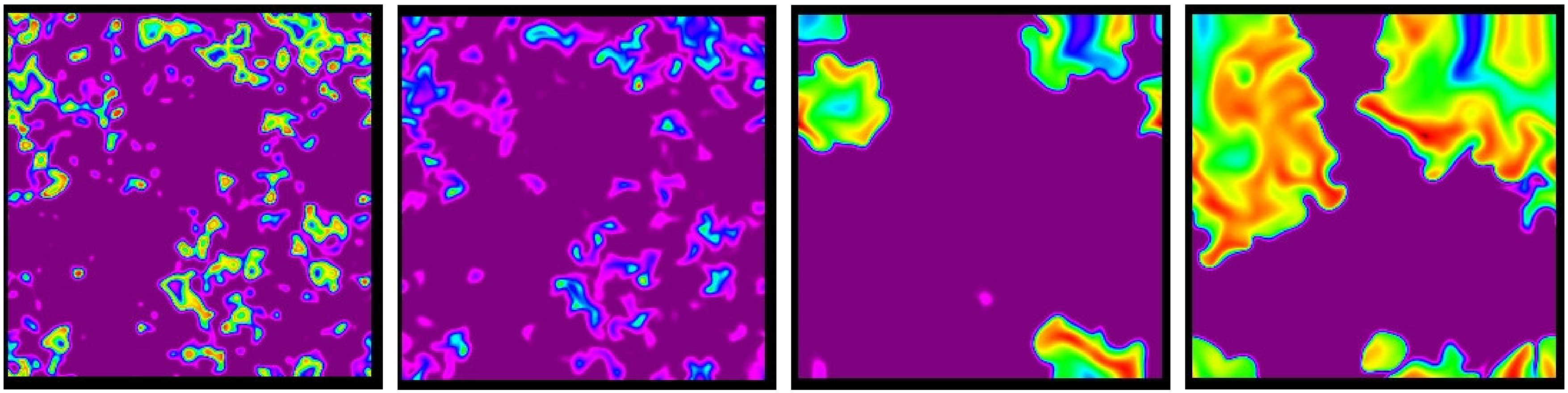


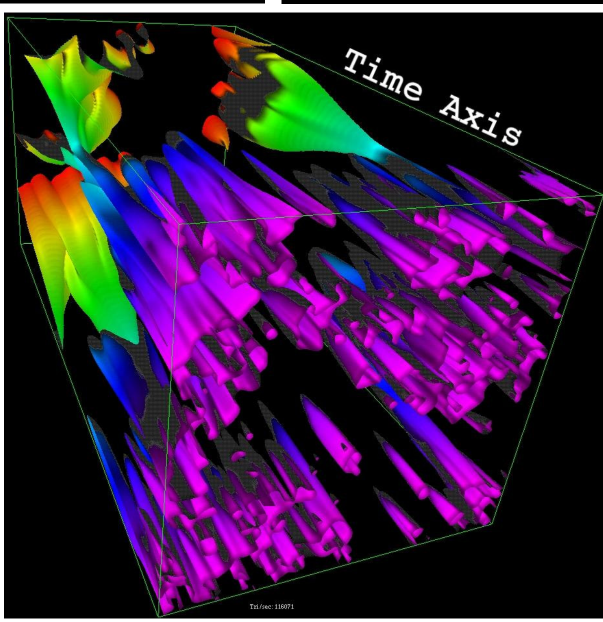


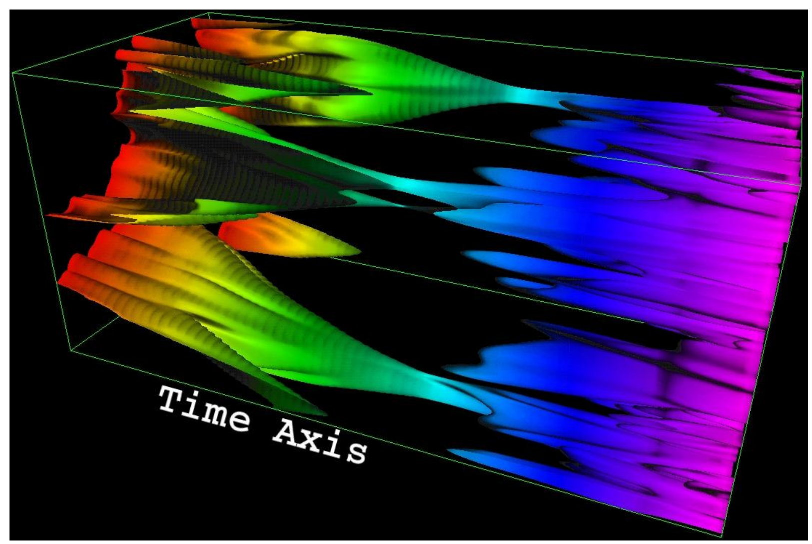




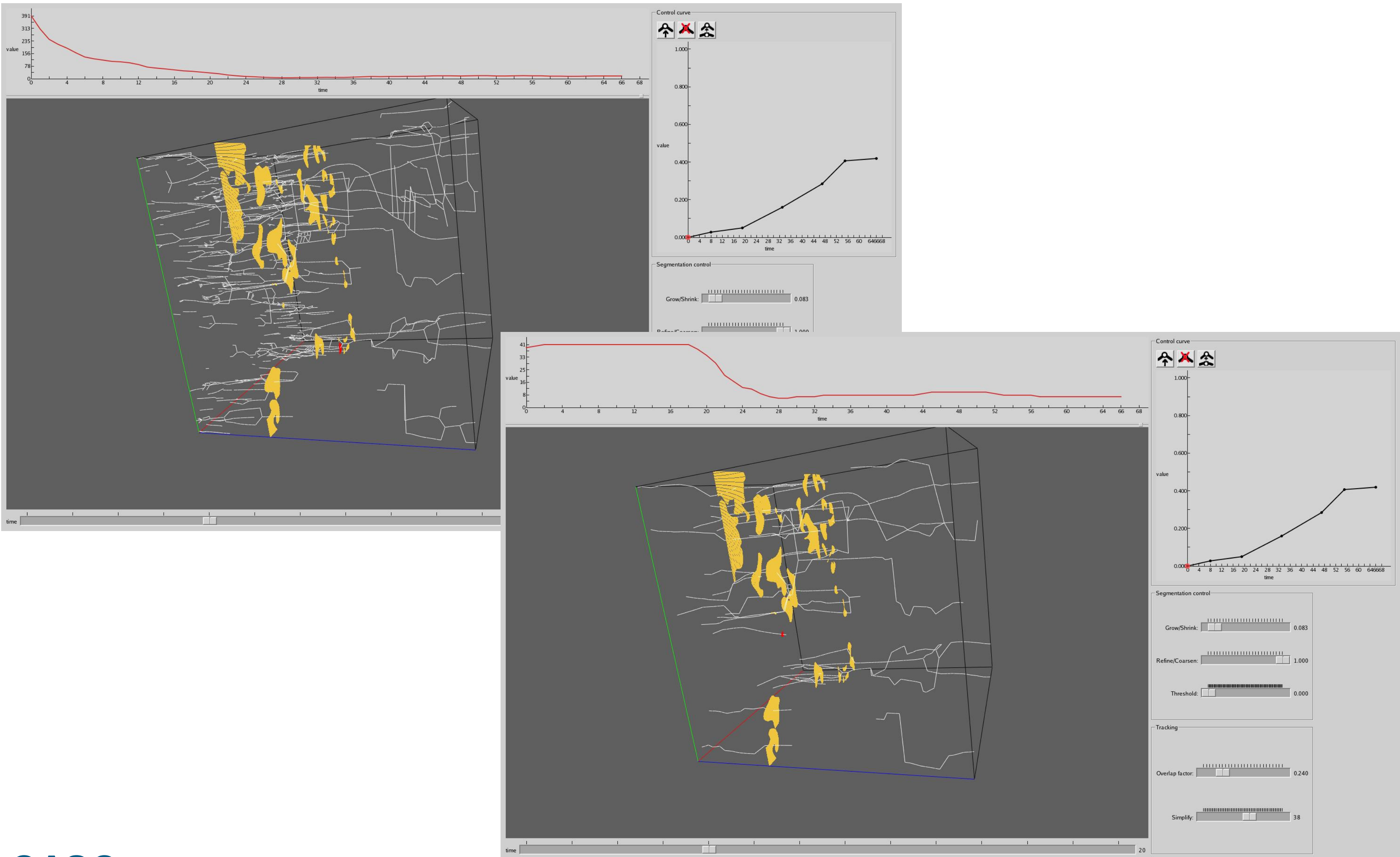
The current attempt in data analysis is focused on the problem of combustion



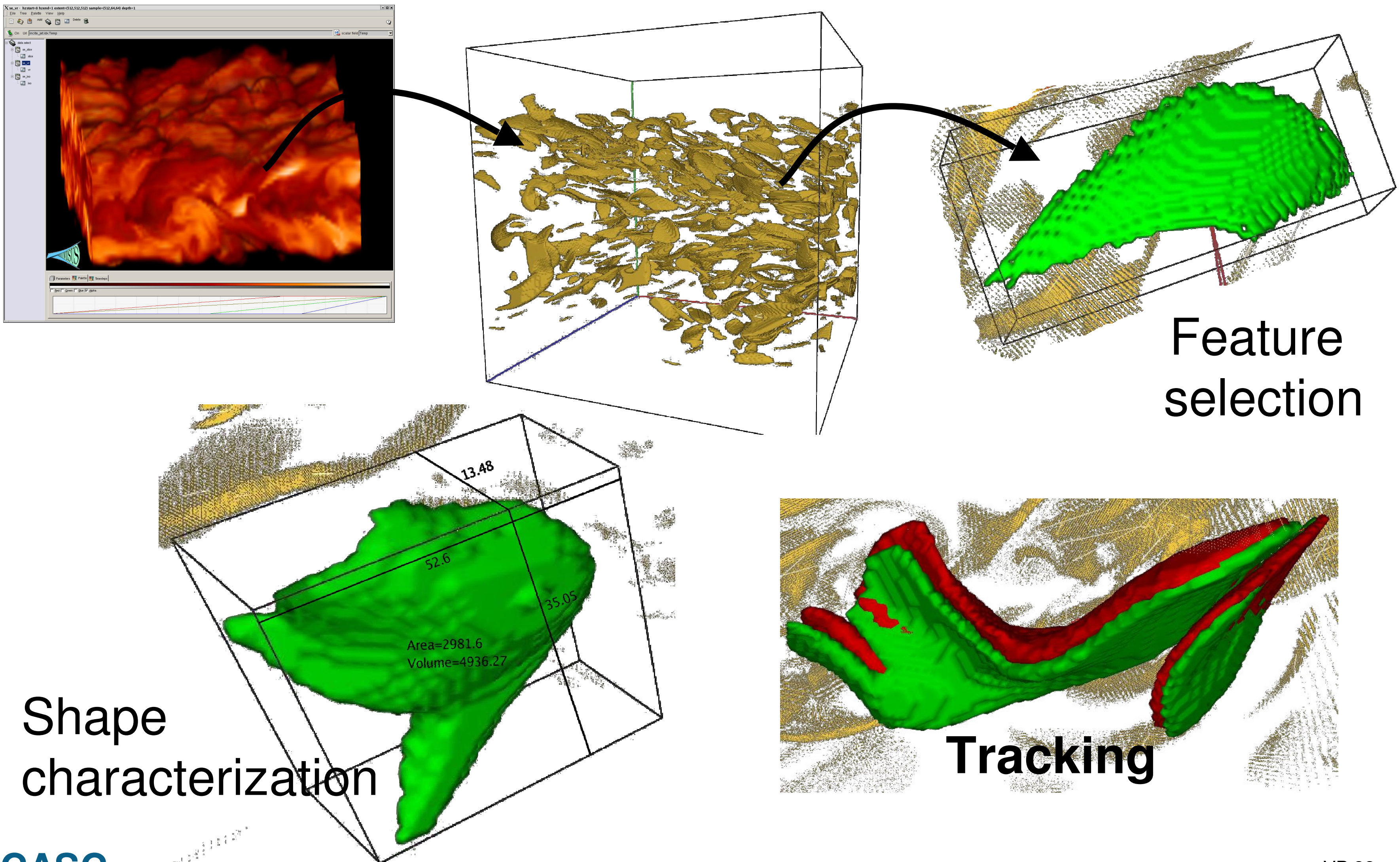




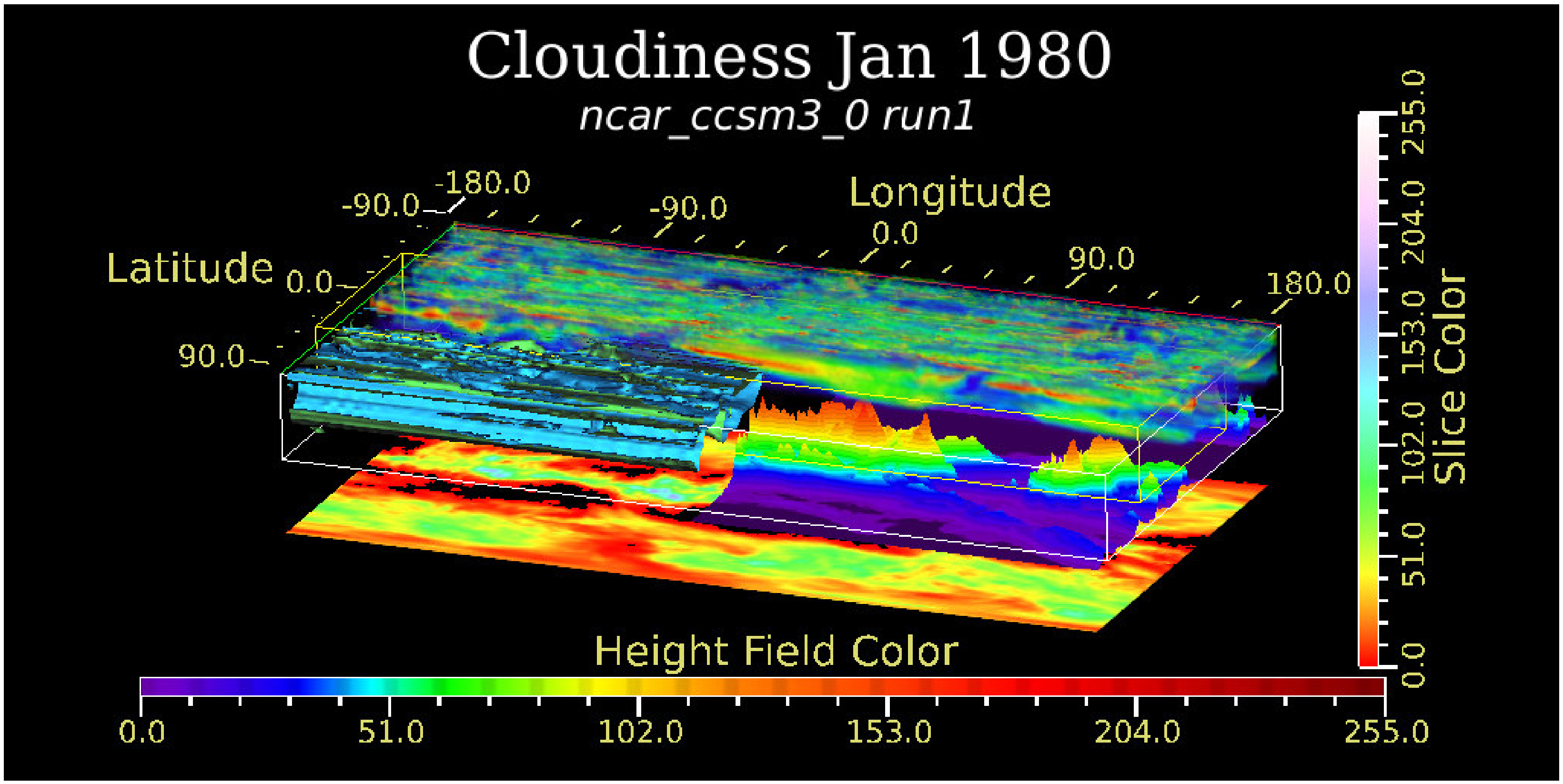
The current attempt in data analysis is focused on the problem of combustion



The current attempt in data analysis is focused on the problem of combustion



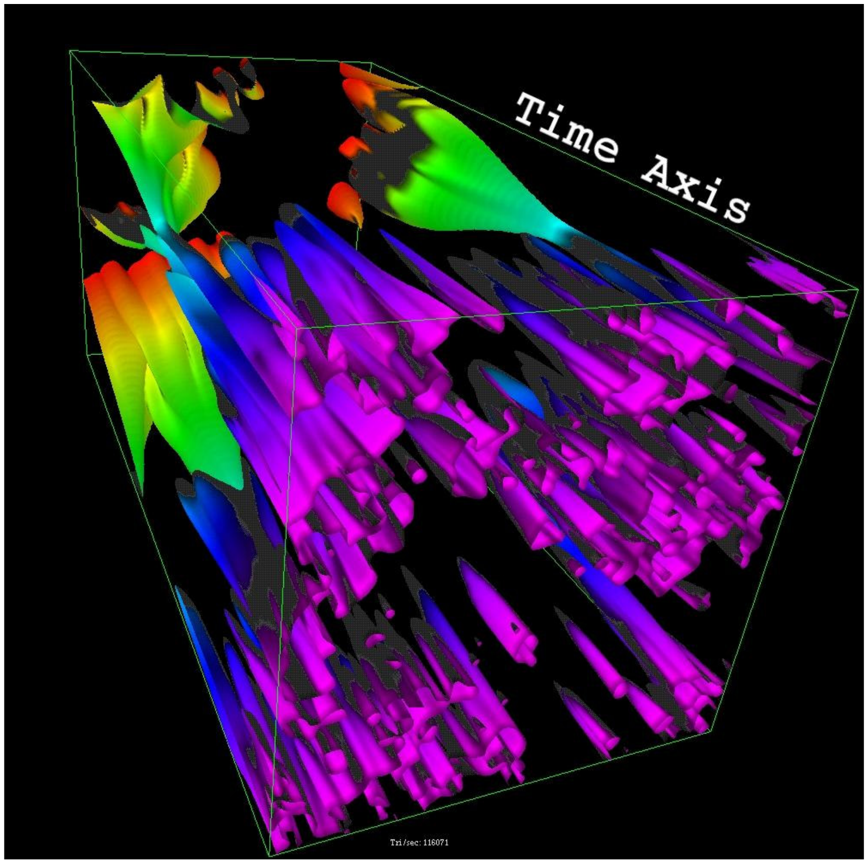
We are starting to look at climate data

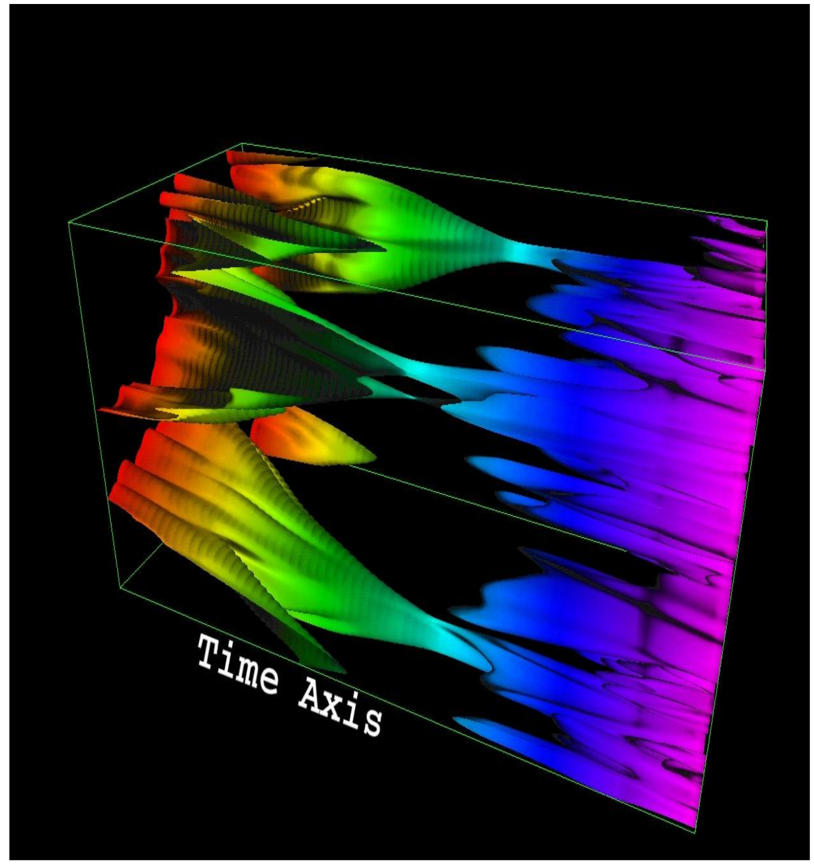


We are working on new extensions of the current formal and algorithmic foundations

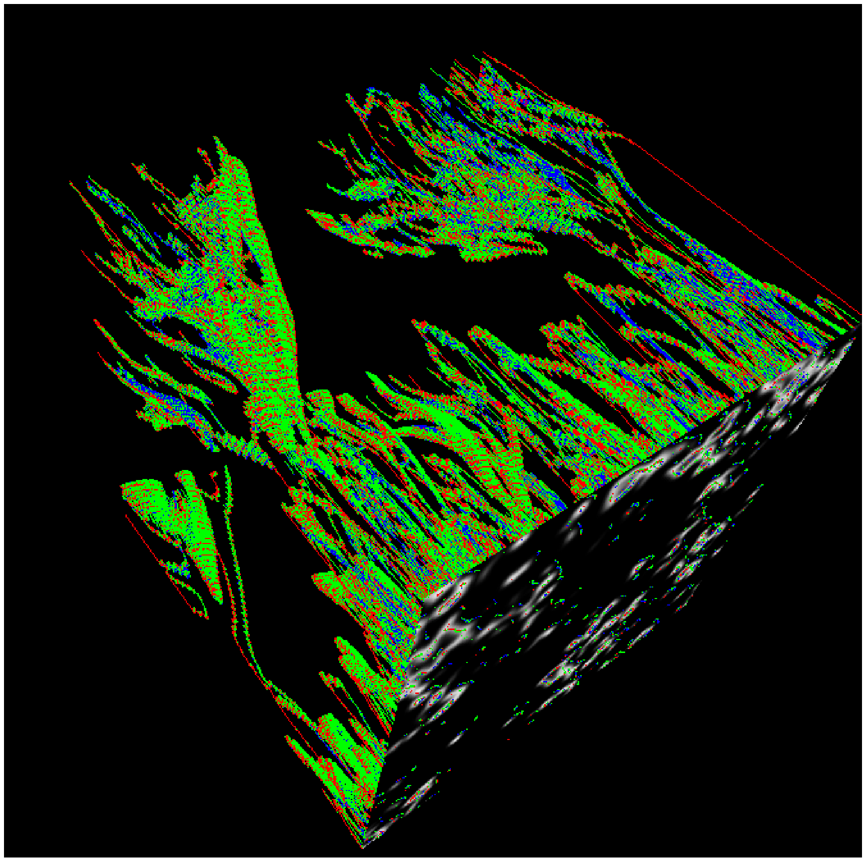
- Robust 3D multi-resolution models tightly coupled with parallel and streaming techniques.
- Tight combination of metric and topological characterization of shape.
- Combined notion of multi scale in space and time (more in general for non homogenous dimensions).
- General vector fields

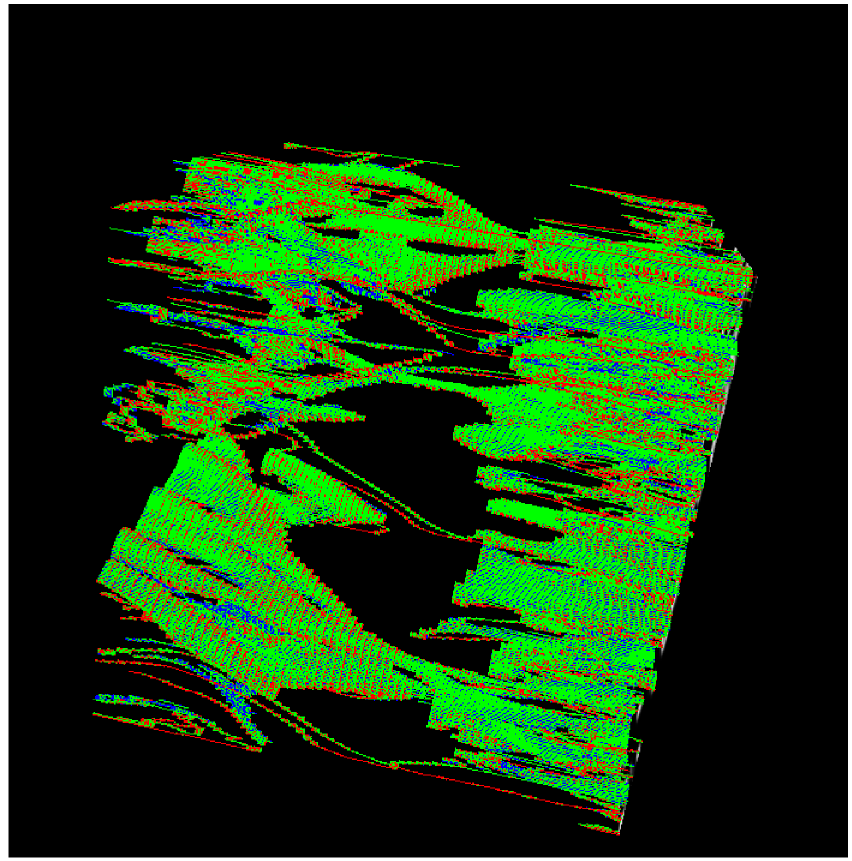
We are testing the scheme for feature tracking in combustion code



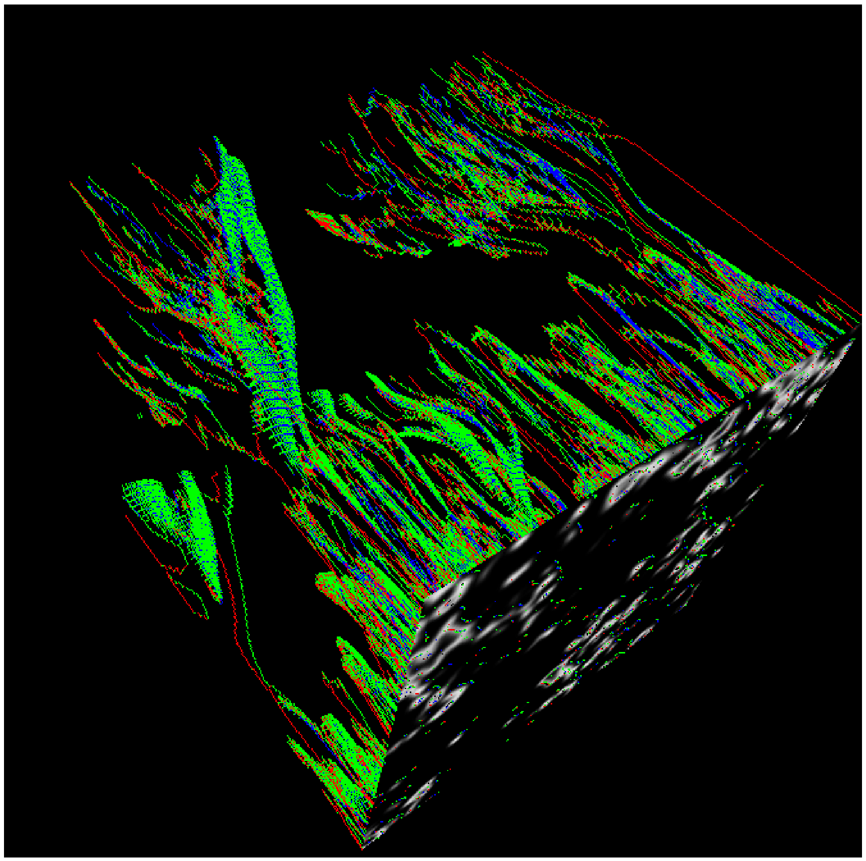


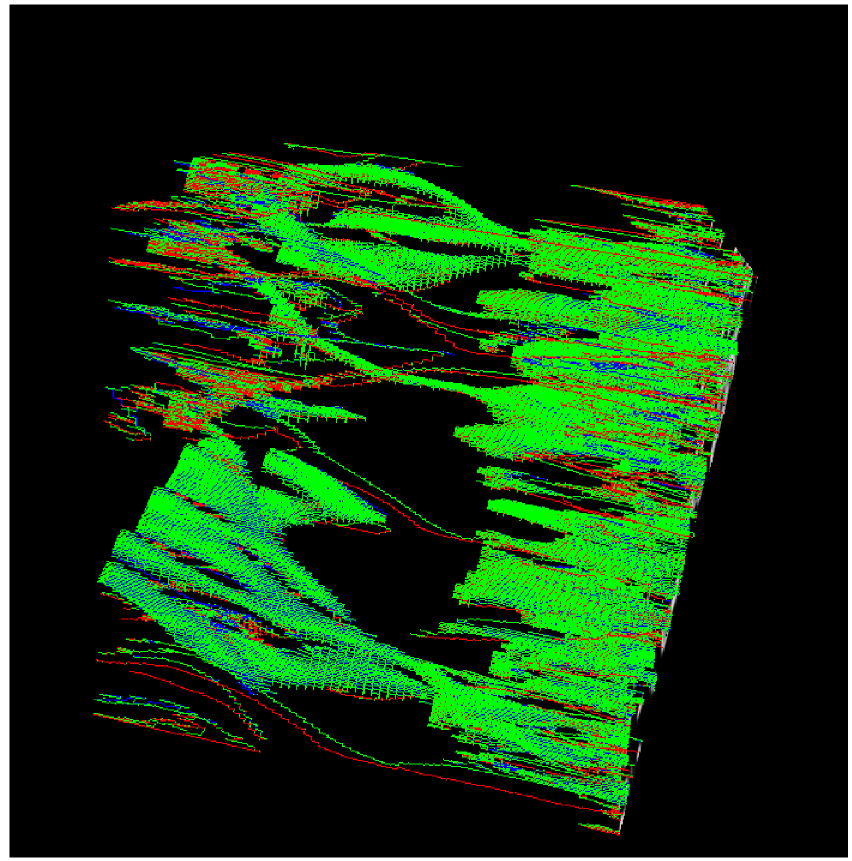
Multi-Scale Jacobi sets highlight trajectories of combustion particles



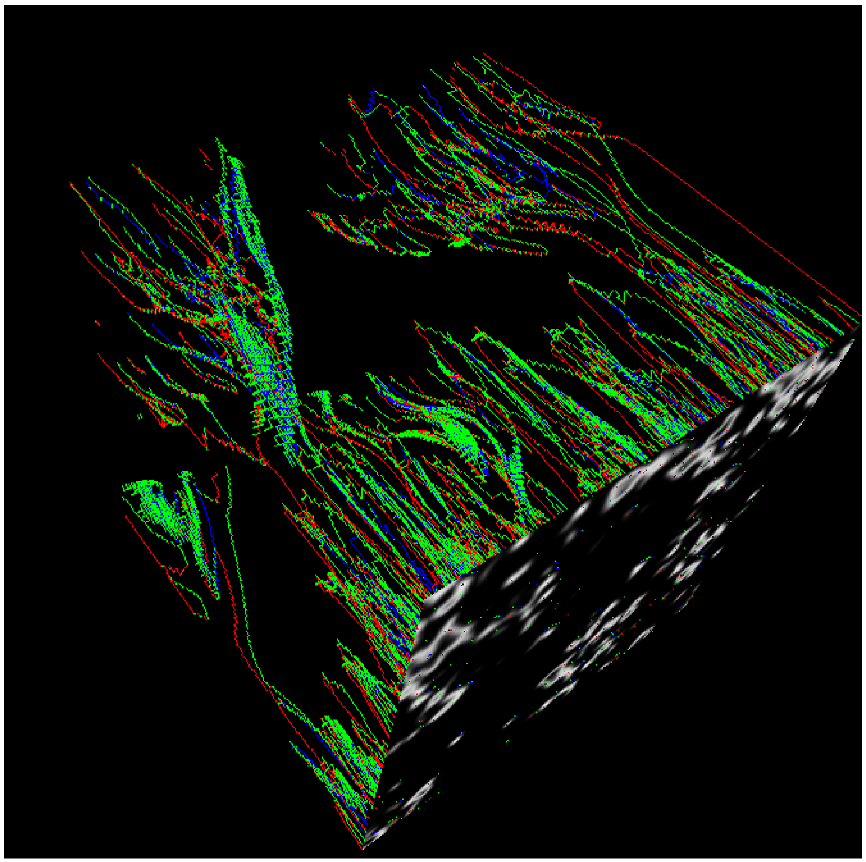


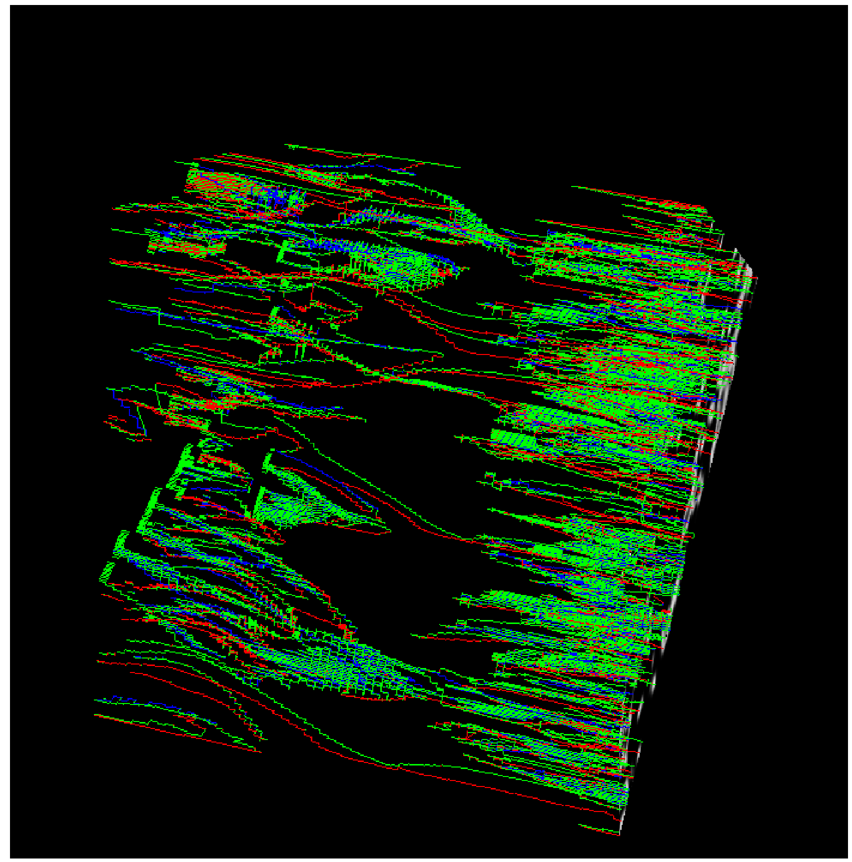
Multi-Scale Jacobi sets highlight trajectories of combustion particles



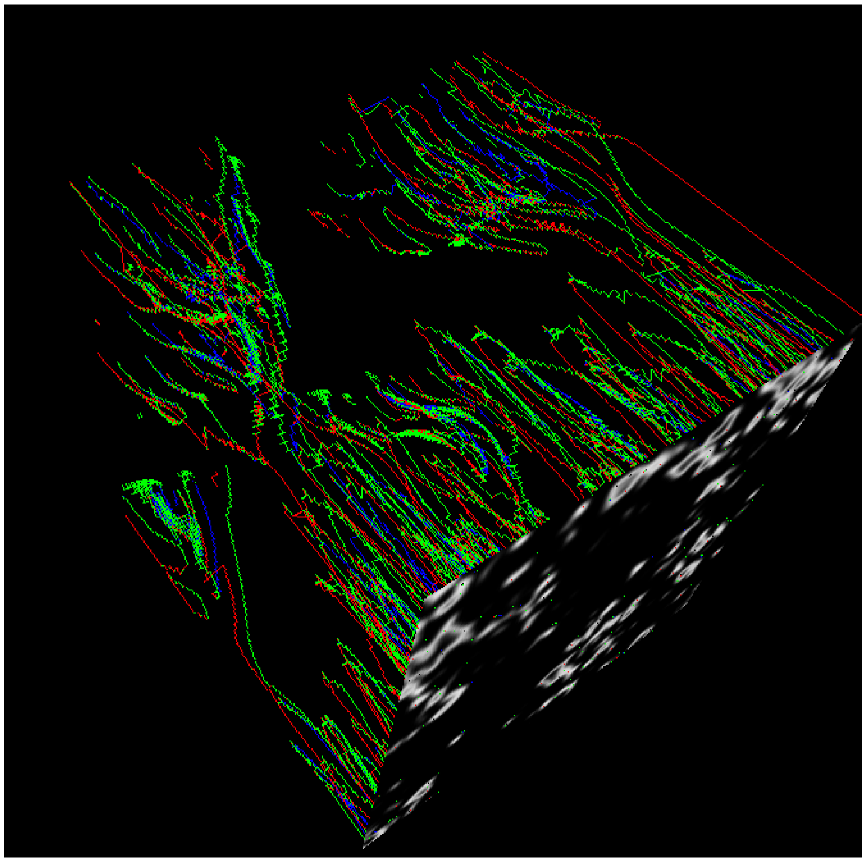


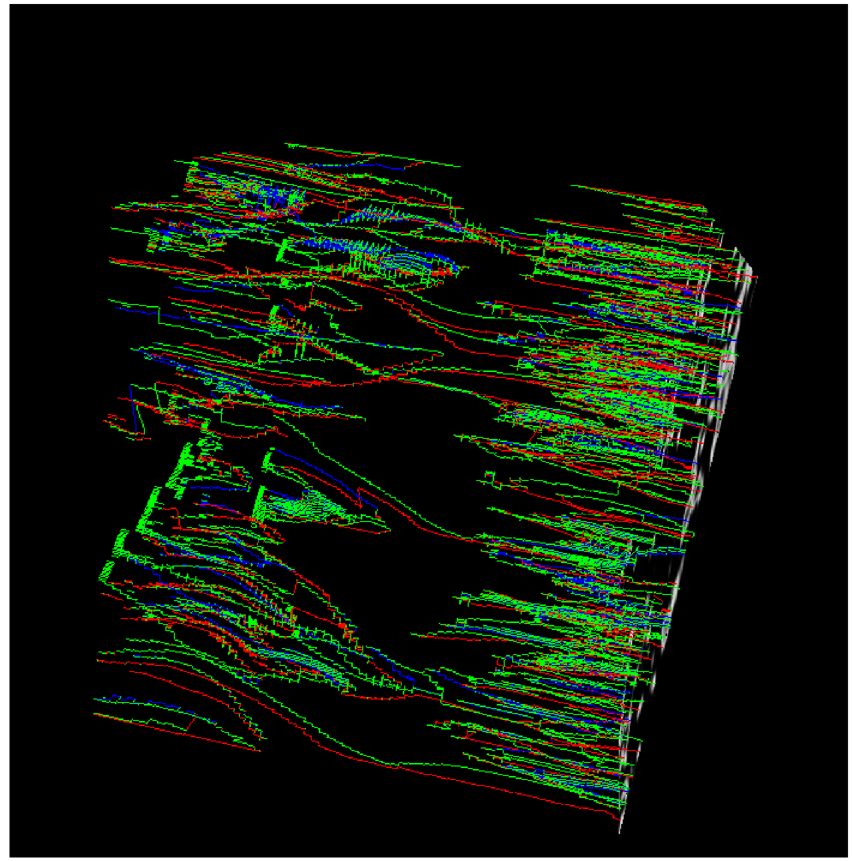
Multi-Scale Jacobi sets highlight trajectories of combustion particles





Multi-Scale Jacobi sets highlight trajectories of combustion particles





Thank You!

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory, under Contract W-7405-Eng-48.