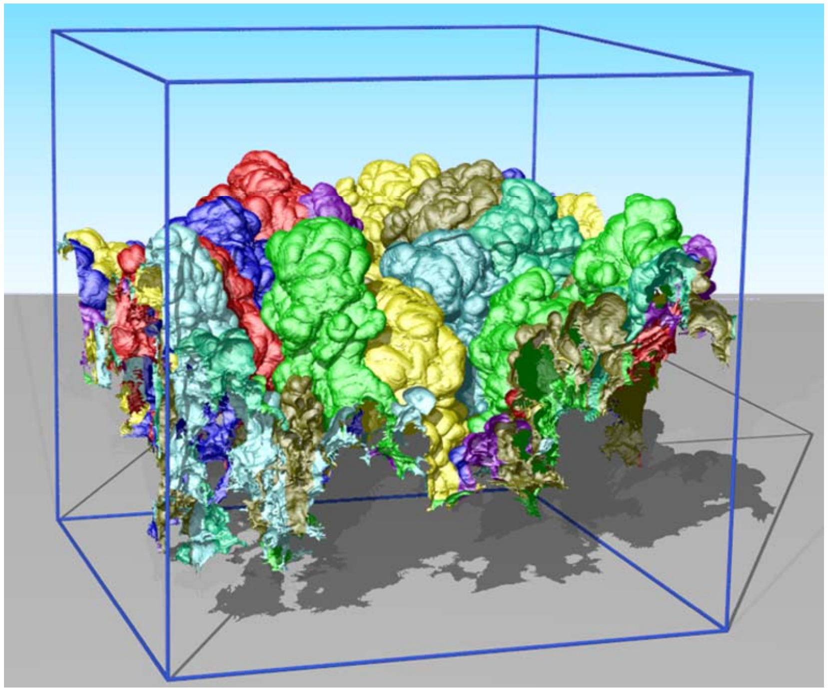
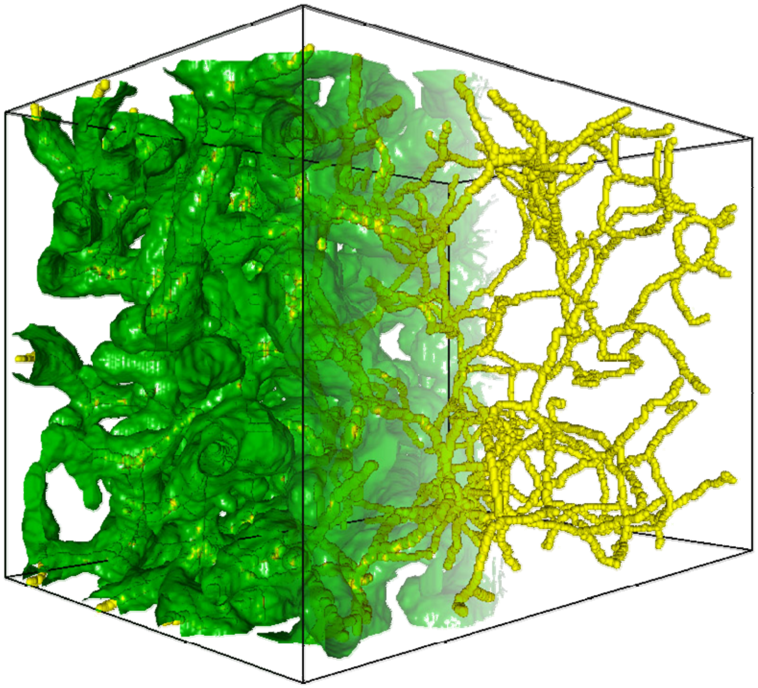
Multi-scale Morse Theory for Science Discovery



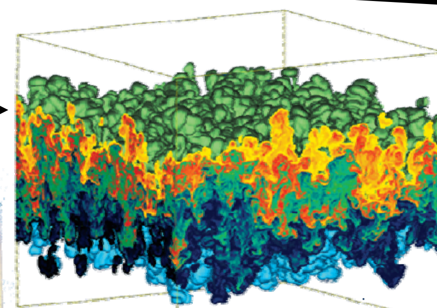


Valerio Pascucci

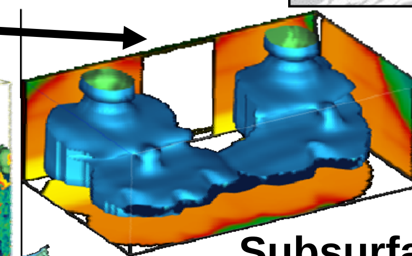
Associate Professor, Scientific Computing and Imaging Institute,
School of Computing, University of Utah
Adjunct Professor, Computer Science, UC Davis

Massive Scientific Models are Source of Great Challenges and Opportunities

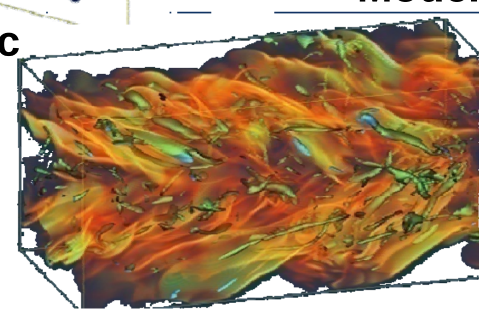
BlueGene/L



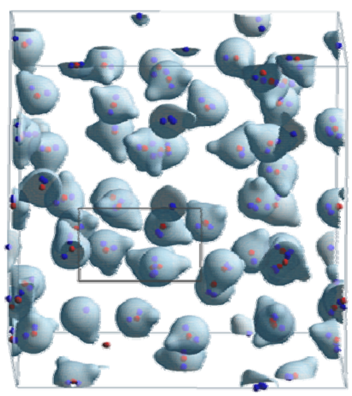
Hydrodynamic Instabilities



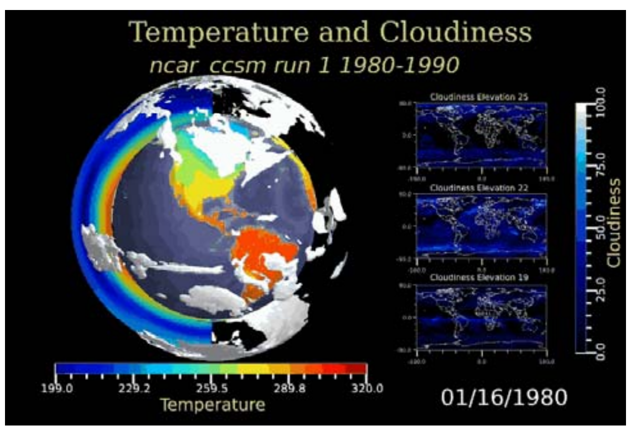
Subsurface Modeling



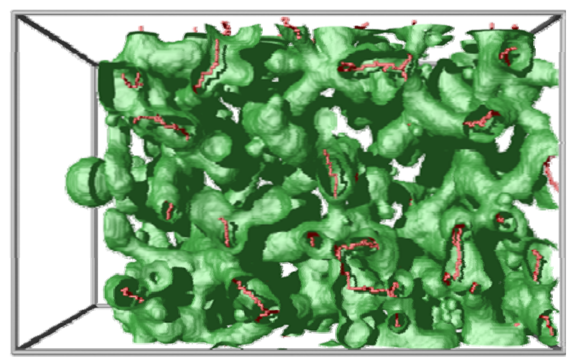
Combustion Simulations



Molecular Dynamics

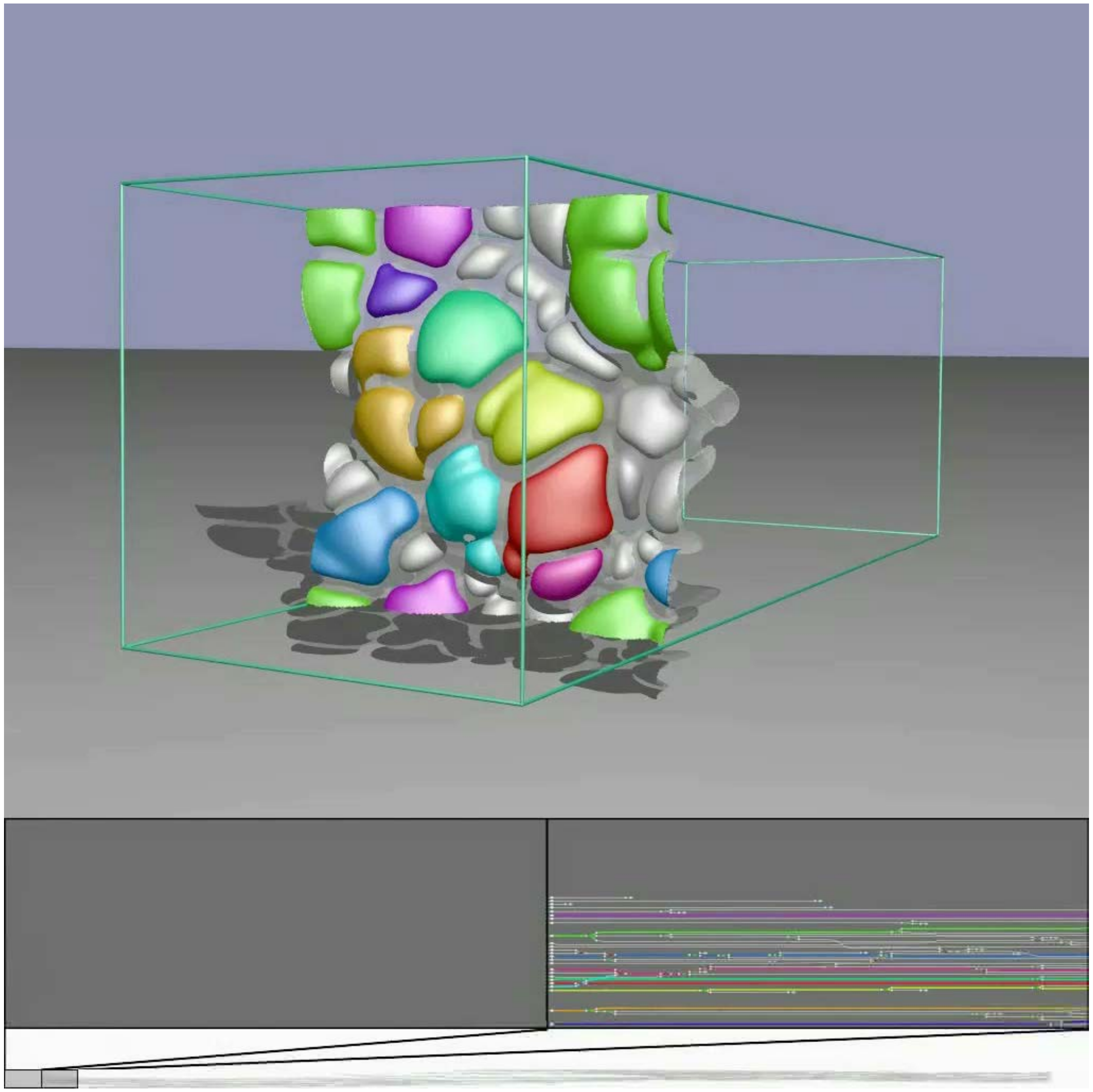


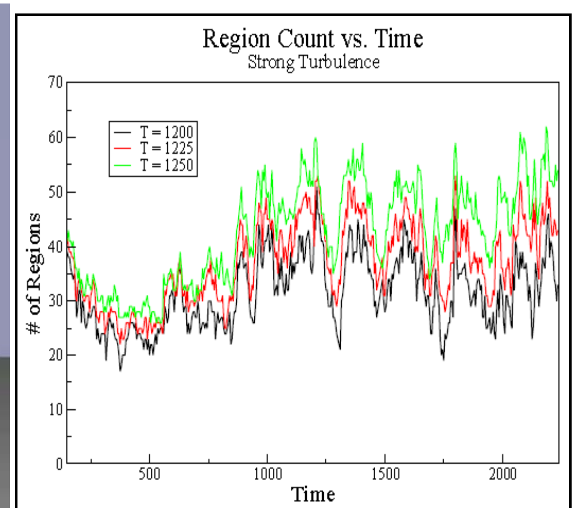
Climate Modeling

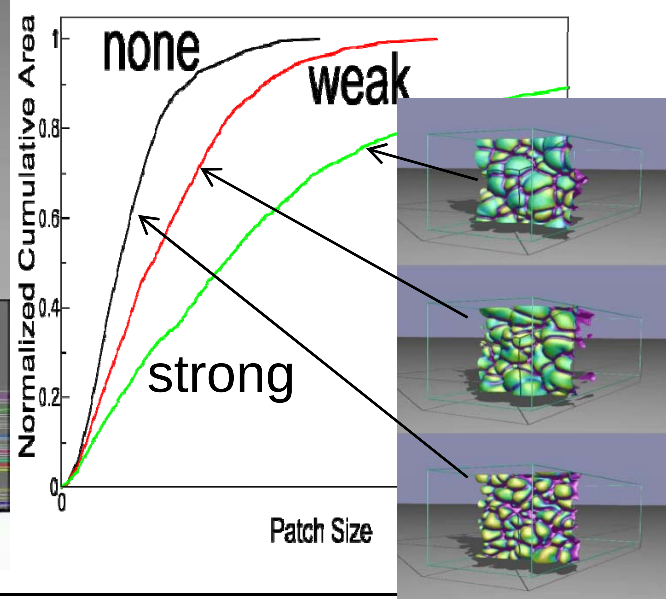


Material Sciences

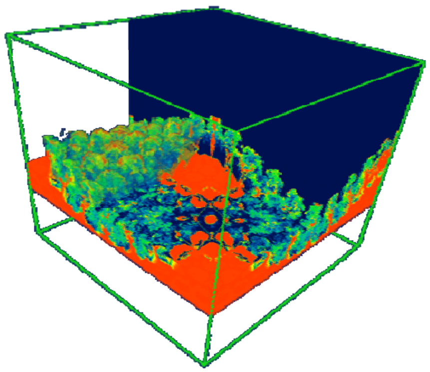
We Need to Enable Complex Data Analysis and Visualization to Aid Science Discovery

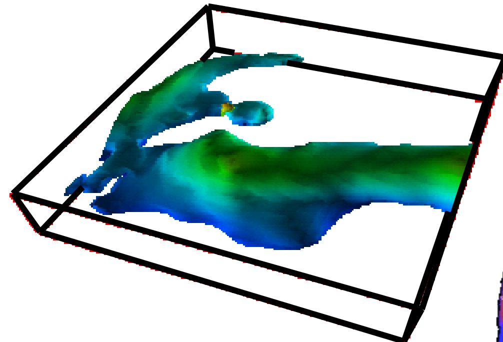


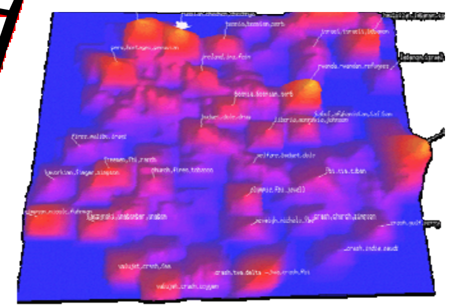


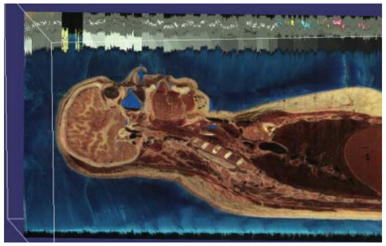


We Develop General Purpose Tools for Efficient and Reliable Data Understanding

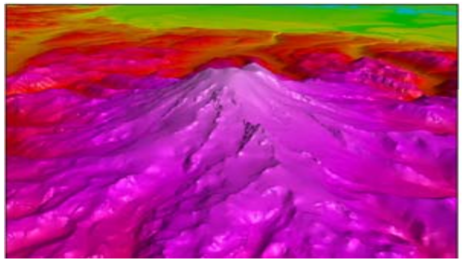


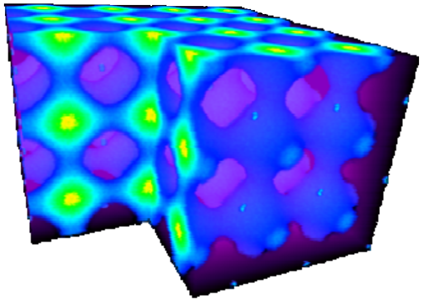


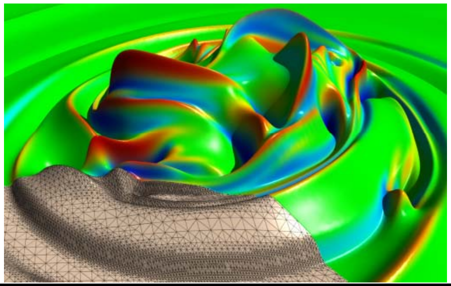


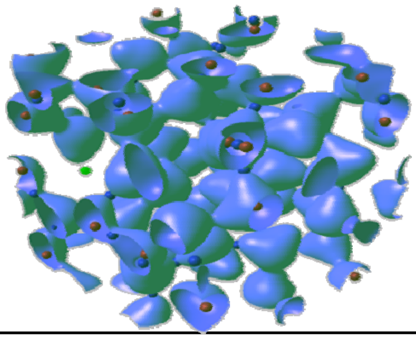


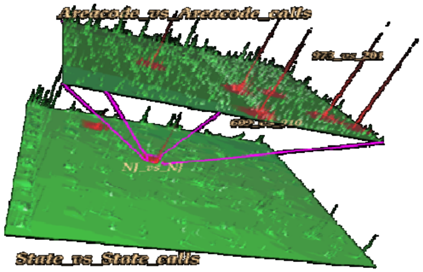
Real functions are ubiquitous in the representation of scientific information.







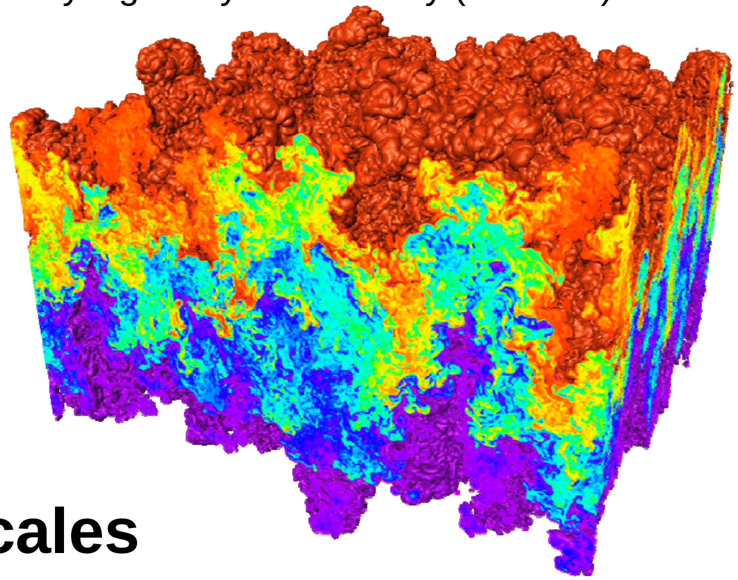




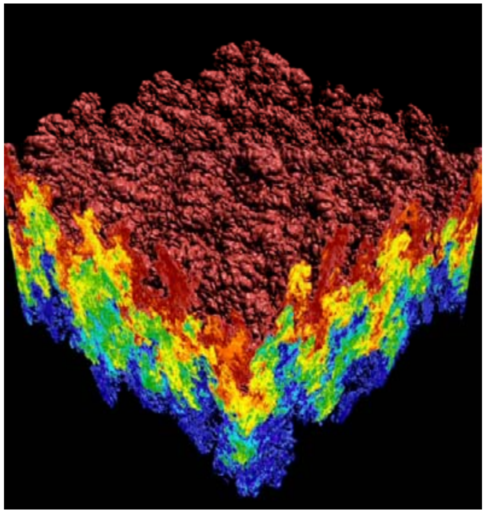
Traditional Data Analysis Tools are Often Ineffective for Massive Models

- **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**
- **Numerical methods unstable and sensitive 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 instability (Miranda)

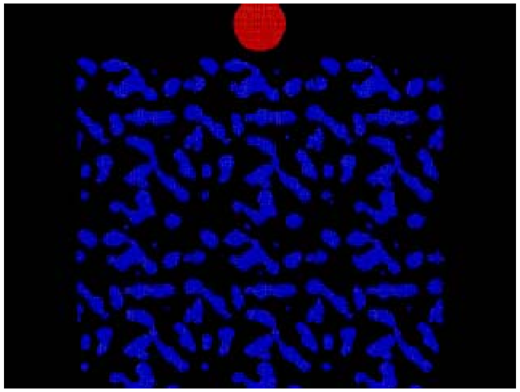


We Target Specific Data Analysis Problems in Support of Science Applications



Rayleigh-Taylor instab. arise in fusion, super-novae, and other fundamental phenomena:

- start: heavy fluid above, light fluid 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

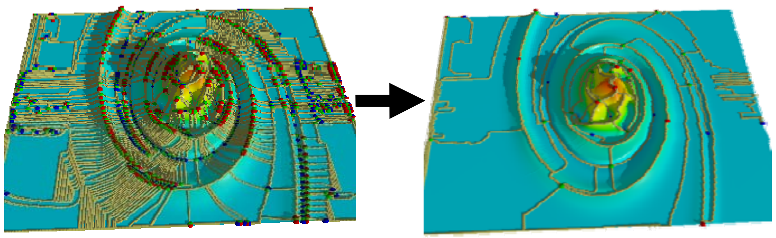


Impact of particles on porous medium arises in space missions (Stardust) collecting comet particles, impact of bullets and more:

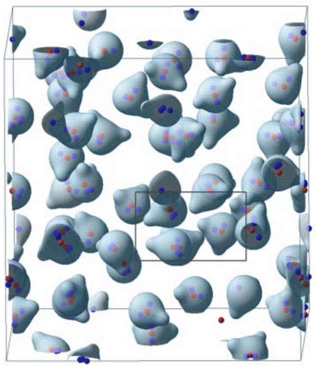
- equation of state and thermodynamic input
- atomistic simulation
- extract structure of porous medium
- quantify the deformation of the structure

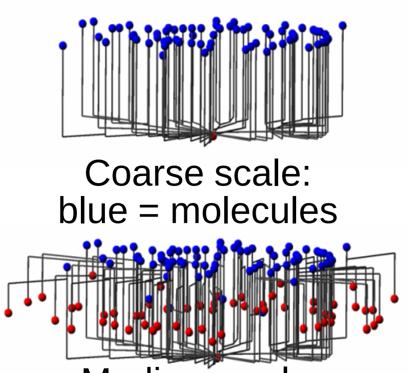
We Introduced Robust Topological Methods for Quantitative Data Analysis

- **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**



Hierarchical topology of a 2D Miranda vorticity field





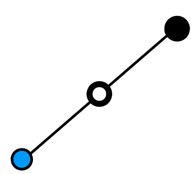
Coarse scale: blue = molecules

Medium scale: red-blue = dipoles

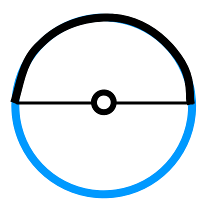
Molecular dynamics simulation (left) with abstract graph representation of its features at two scales (right)

We Rewrote Morse Theory for Provably Robust and Correct Computations

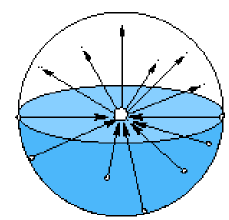
	$f(x): D \rightarrow \Re$	$F(x): S \rightarrow \Re$
	Classical mathematical definitions	Simulation of differentiability
domain	D smooth manifold	S simplicial complex
function	f infinitely differentiable	$F(x)$ PL-extension of $f(x_i)$
critical point	$\nabla f(p) = 0$ **numerical**	$LowerLink(p) \neq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

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

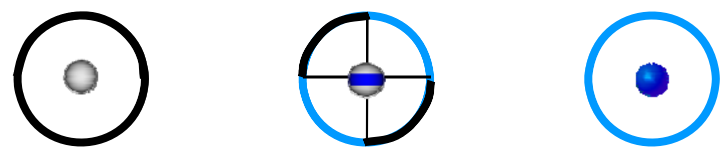
↓

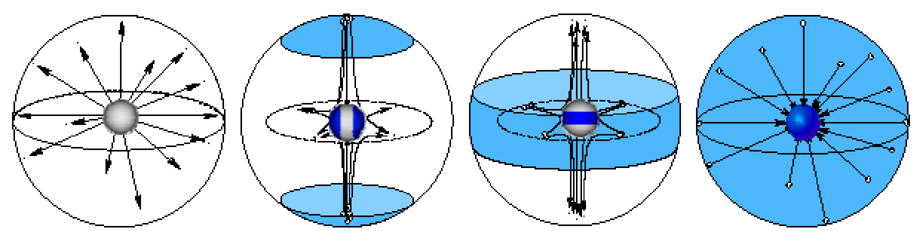
$$f(x)|_p = f(p) + \sum_{i=1}^{d-k} x_i^2 - \sum_{j=d-k+1}^{d} x_j^2$$

numerical

The Morse Lemma

There are *d+1* types
of critical points

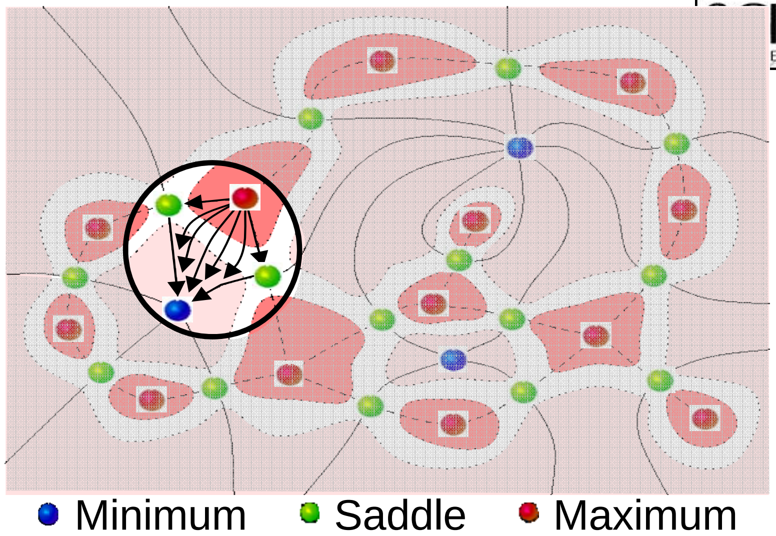


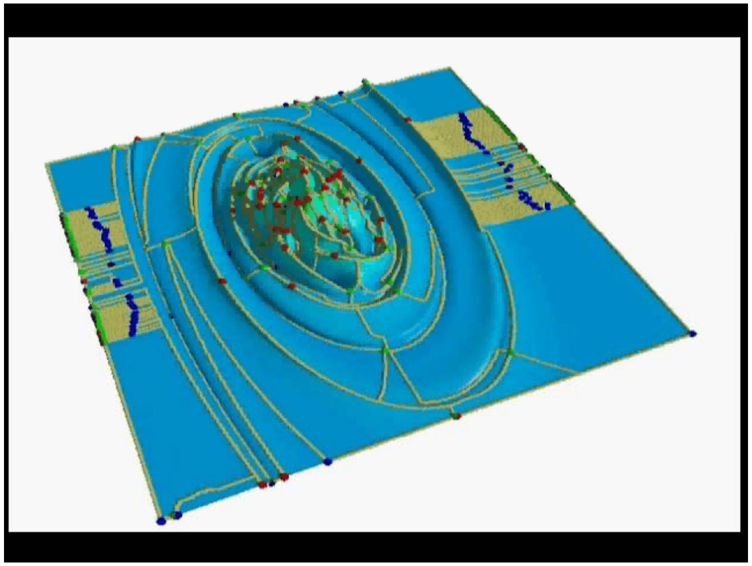


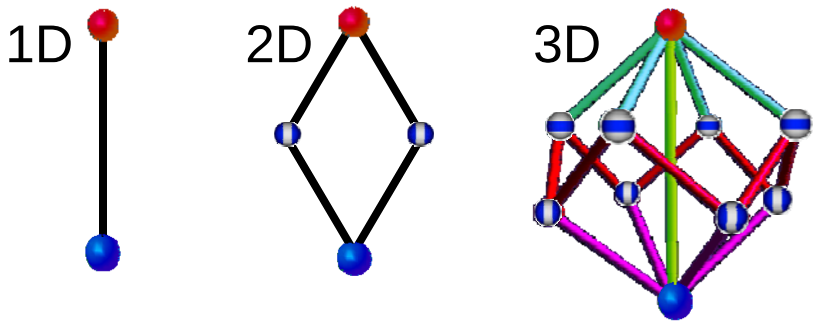
combinatorial

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**







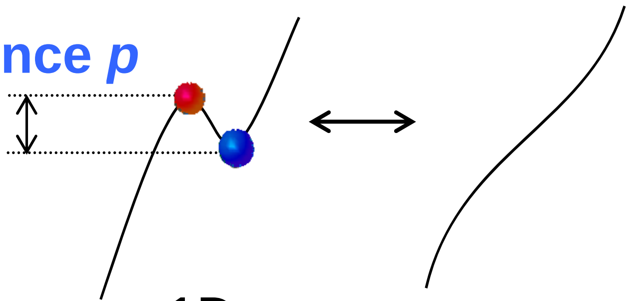
We Mapped the Index Lemma to a Morse–Smale Complex Simplification

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

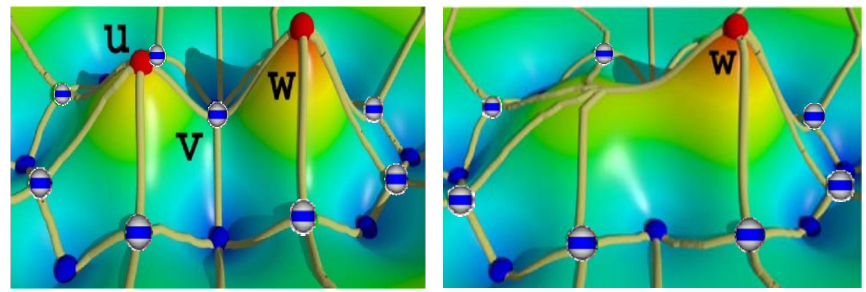
Approximation: error = persistence/2 (proven lower bound)

Multiscale: consistent gradient segmentation at all scales

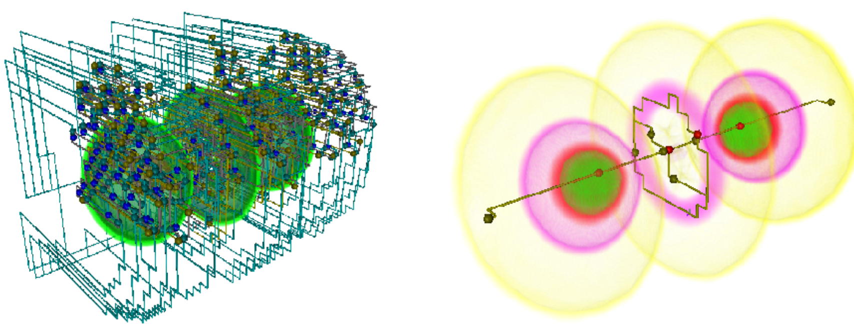
persistence *p*



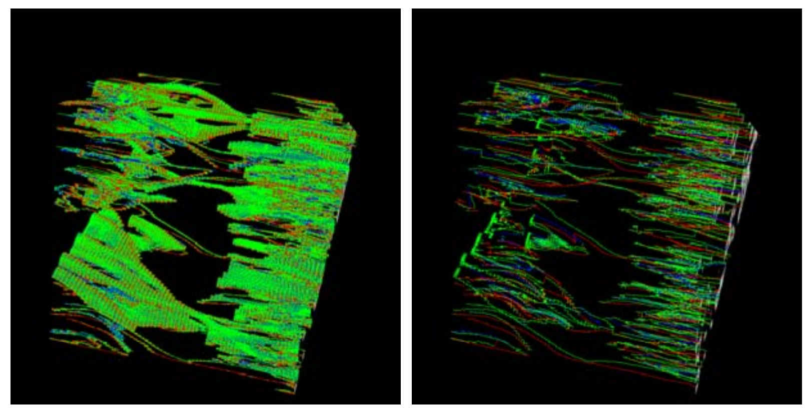
1D



2D



3D

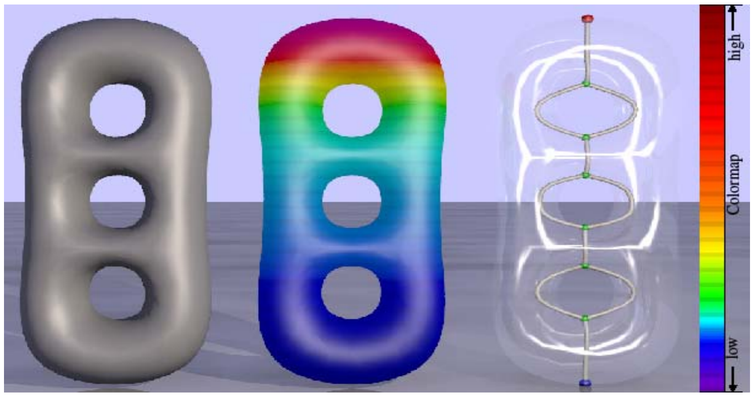


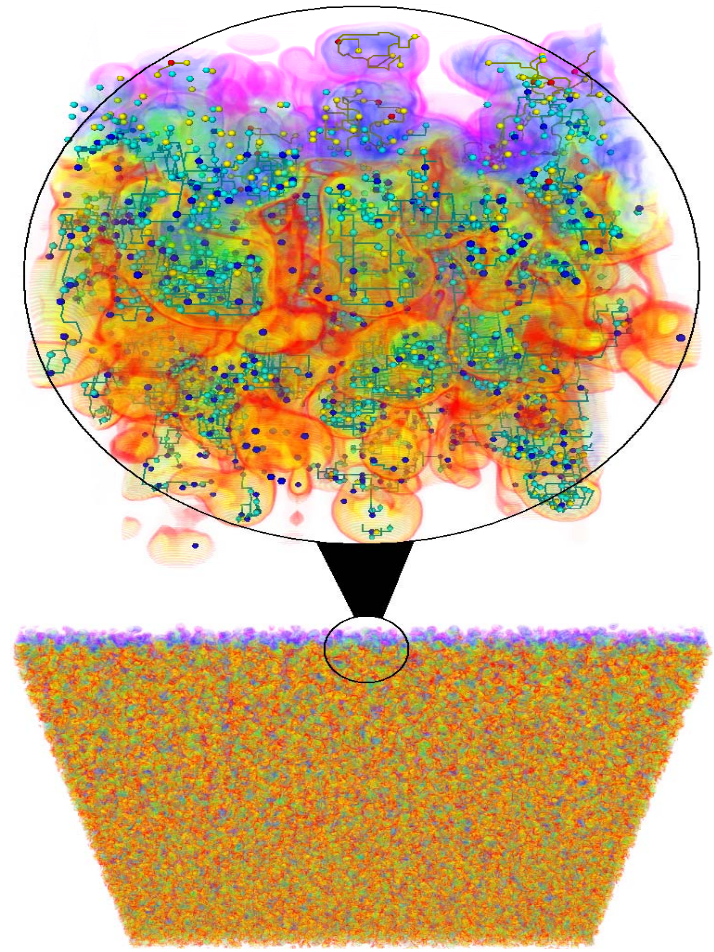
Time

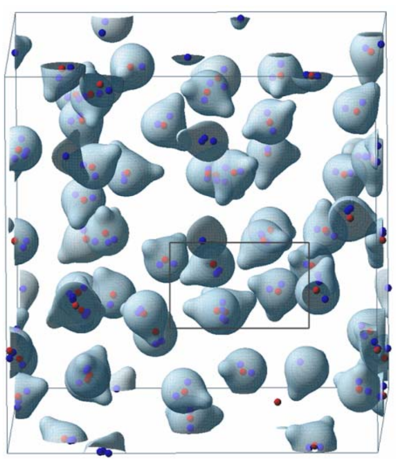
We Achieve a Theory of Provably-Correct Topological Computations

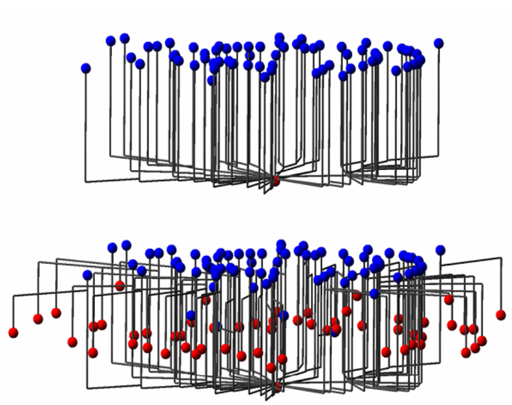
- **No approximation introduced when translating mathematical definitions to algorithms**
- **The quality of the analysis does not deteriorate when the data size increases**
- **New topological concepts that allow complete data segmentations**
- **Multi-scale representation of the input data**
- **Explicit error bounds for any approximation**

We Achieve Compact Representations with Fast and Scalable Data Processing

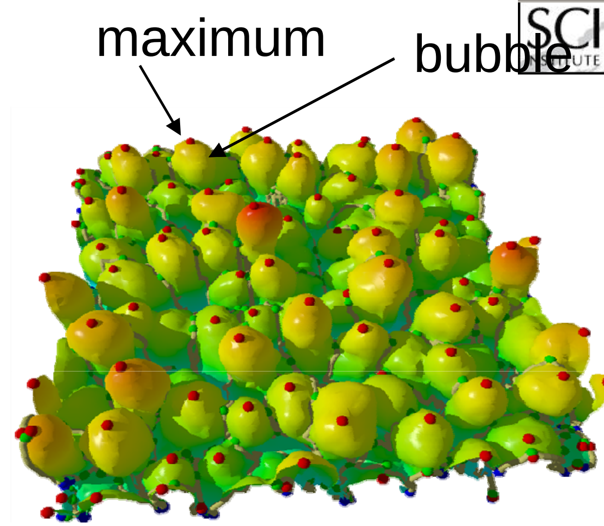








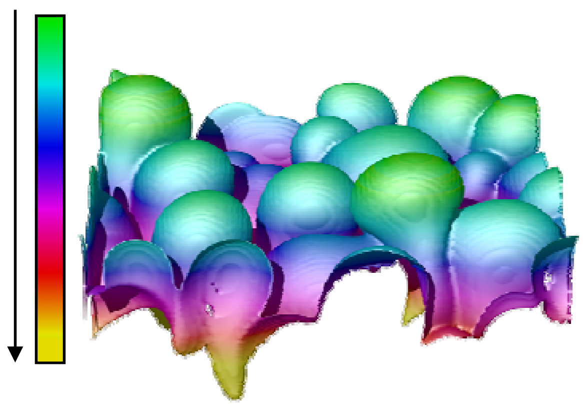
We Analyze High-Resolution Rayleigh–Taylor Instability Simulations



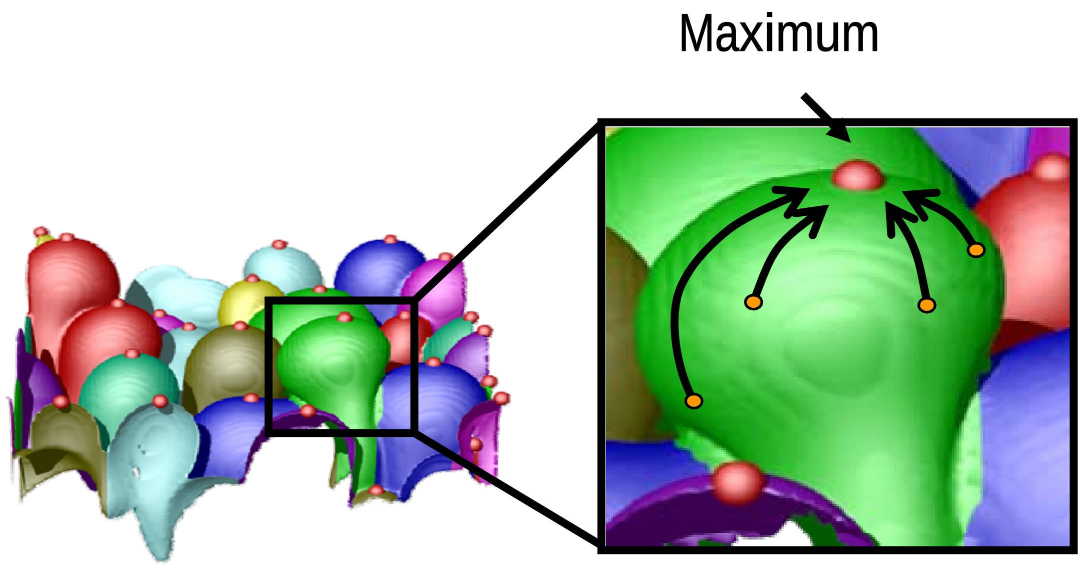
- **Large eddy simulation run on Linux cluster: 1152 x 1152 x 1152**
  - **~ 40 G / dump**
  - **759 dumps, about 25 TB**
- **Direct numerical simulation run on BlueGene/L: 3072 x 3072 x Z**
  - **Z depends on width of mixing layer**
  - **More than 40 TB**
- **Bubble-like structures are observed in laboratory and simulations**
- **Bubble dynamics are considered an 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



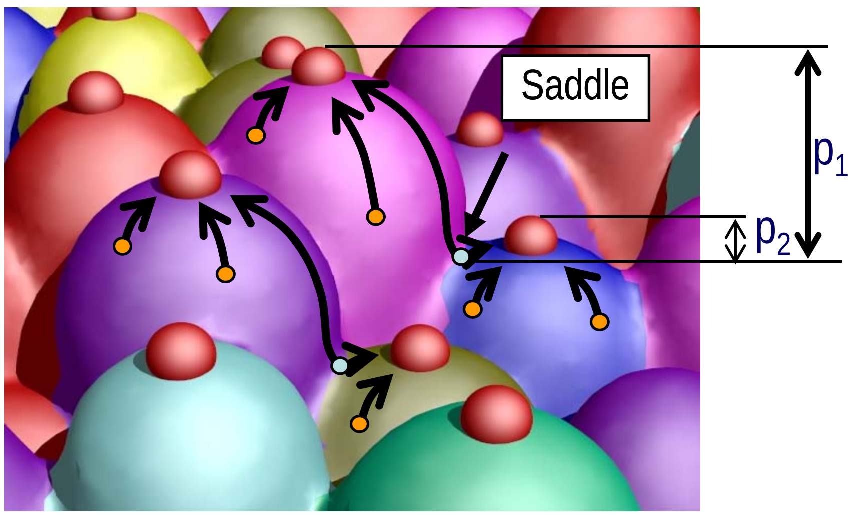
Morse complex cells drawn in distinct colors

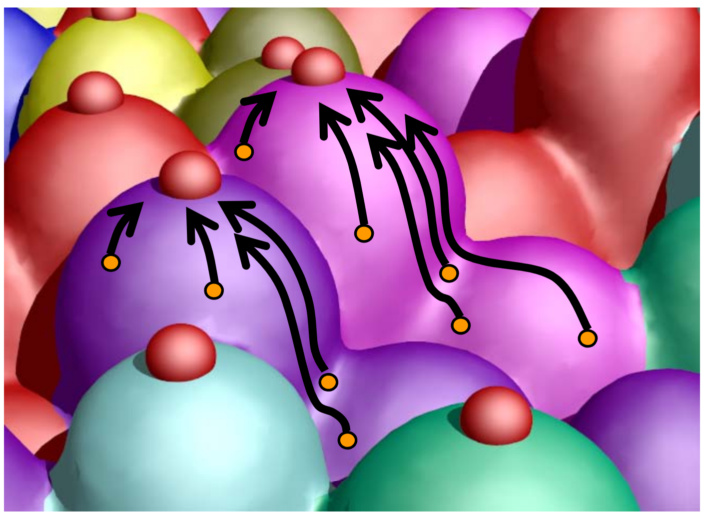
In each Morse complex cell, all steepest ascending lines converge to one maximum

A Hierarchal Model is Generated by Simplification of Critical Points

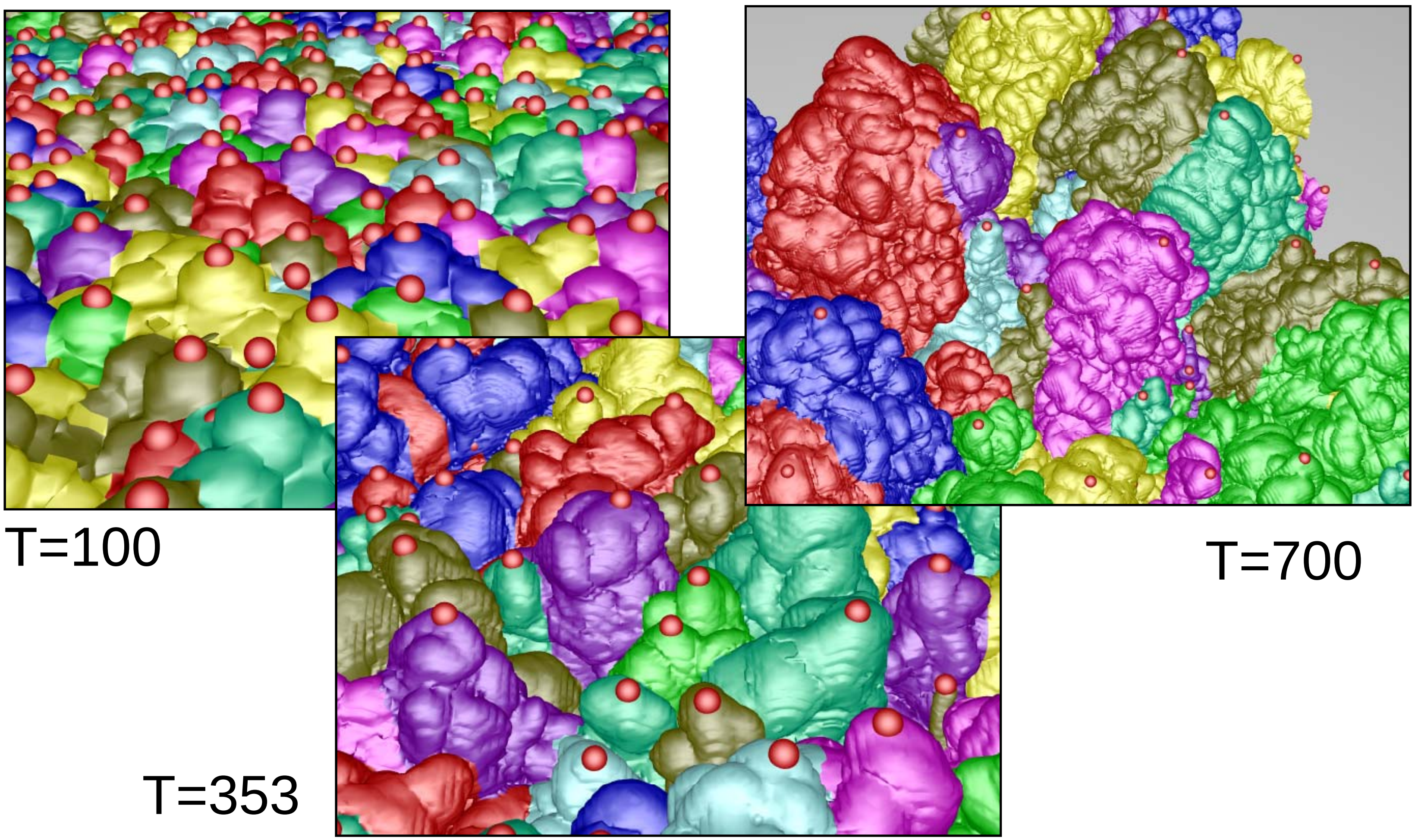
- **Persistence is varied to annihilate pairs of critical points and produce coarser segmentations**

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





The Segmentation Method is Robust From Early Mixing to Late Turbulence

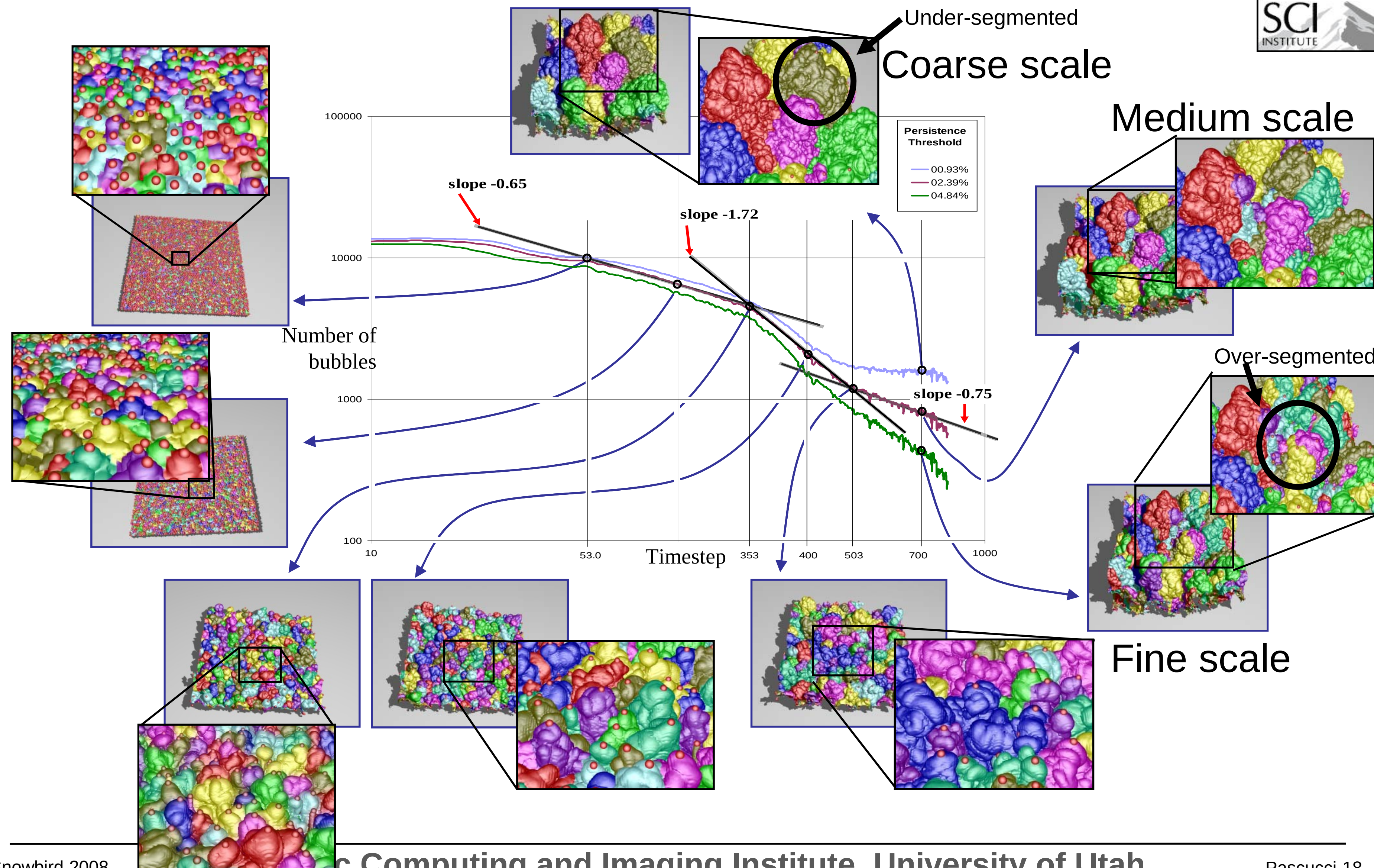


T=100

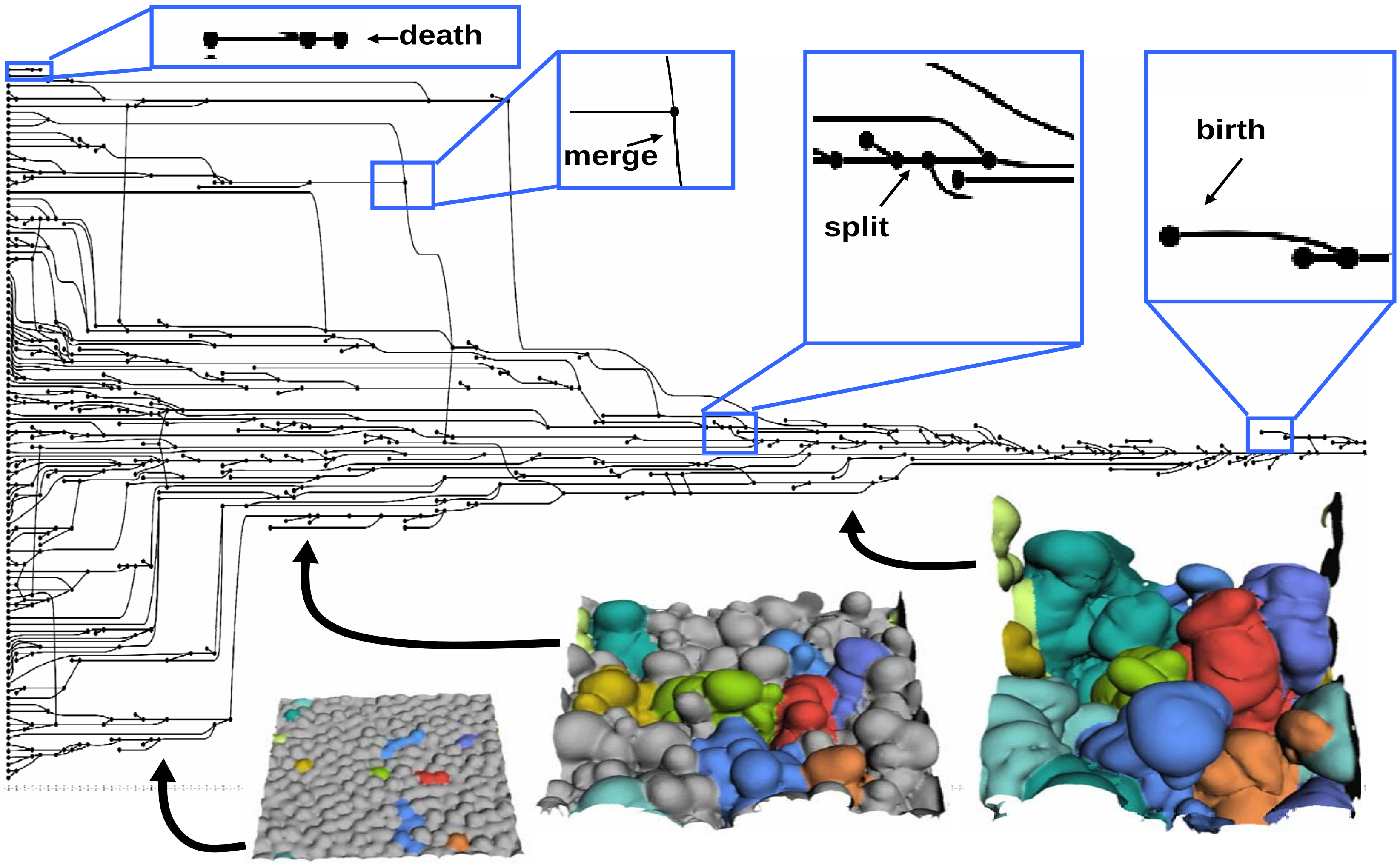
T=353

T=700

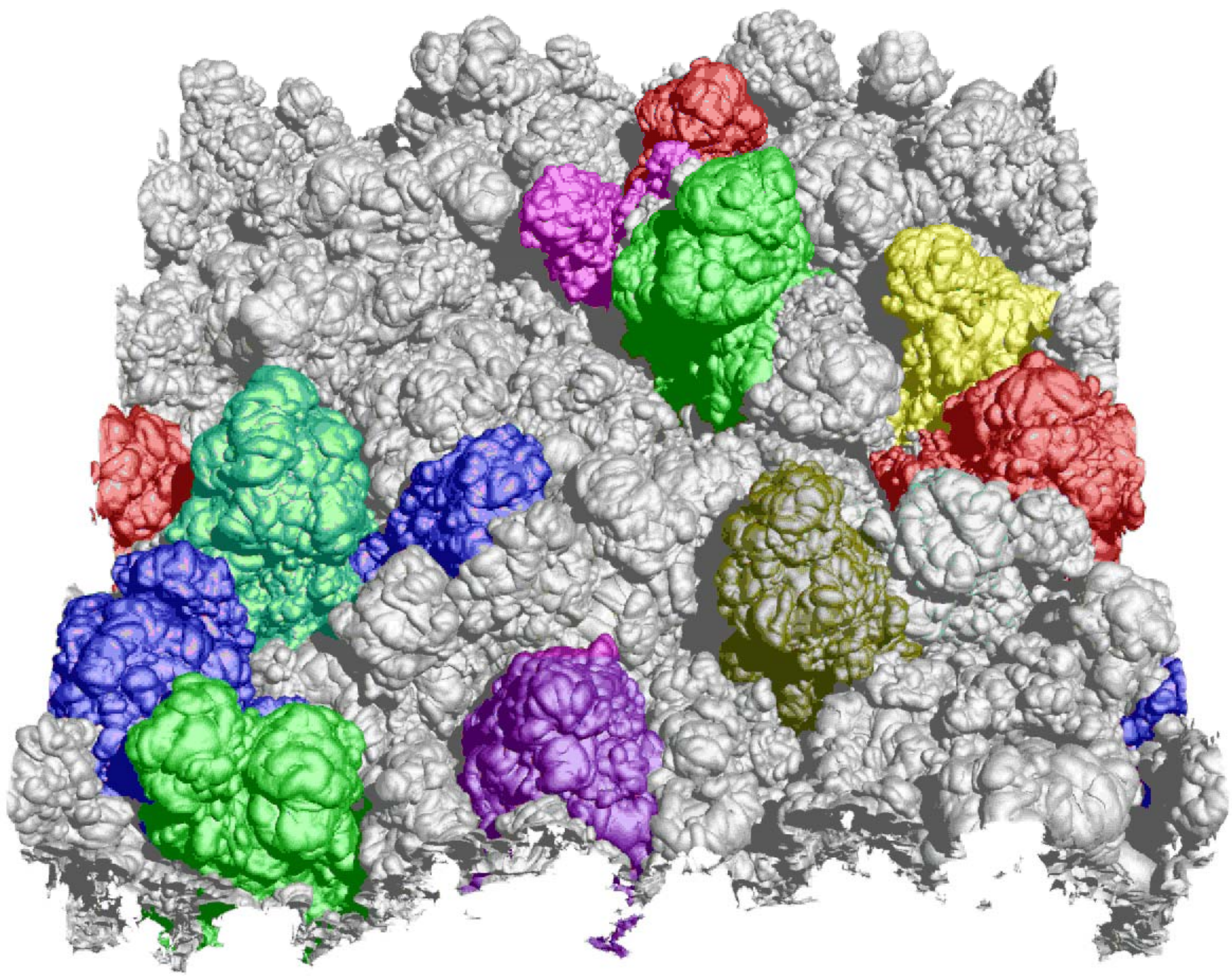
We Evaluated Our Quantitative Analysis at Multiple Scales



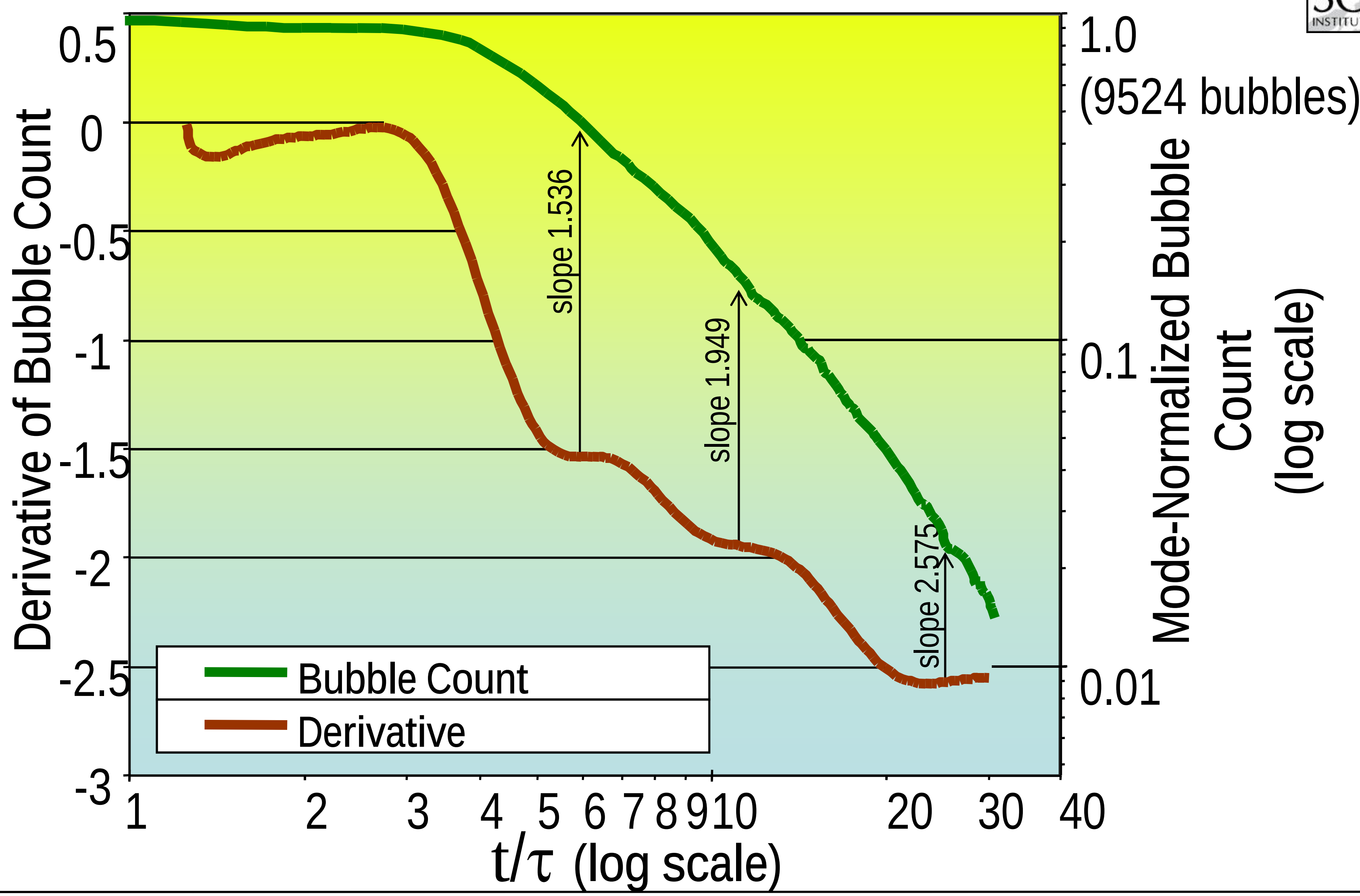
We Characterize Events that Occur in the Mixing Process



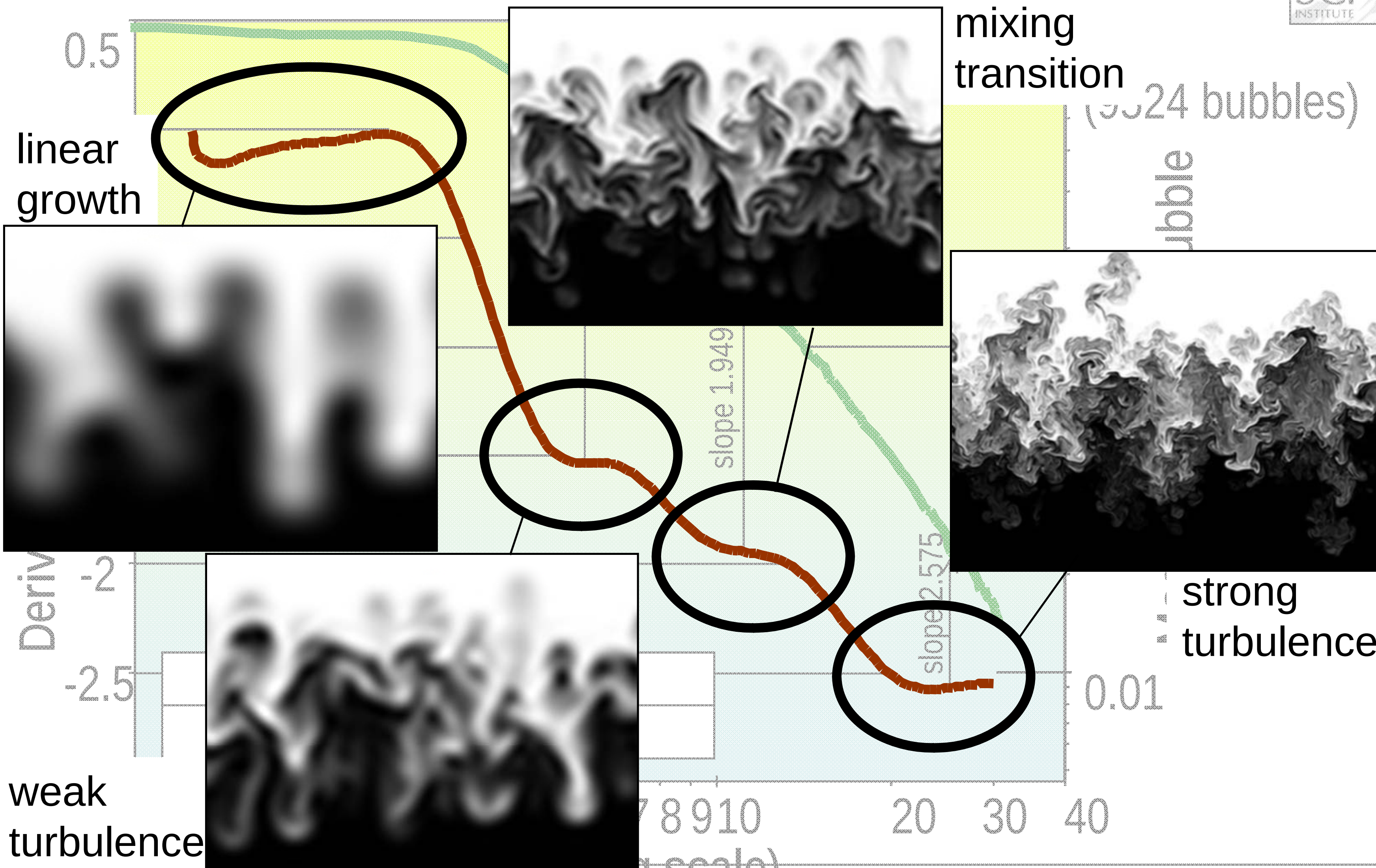
First Robust Bubble Tracking From Beginning to Late Turbulent Stages



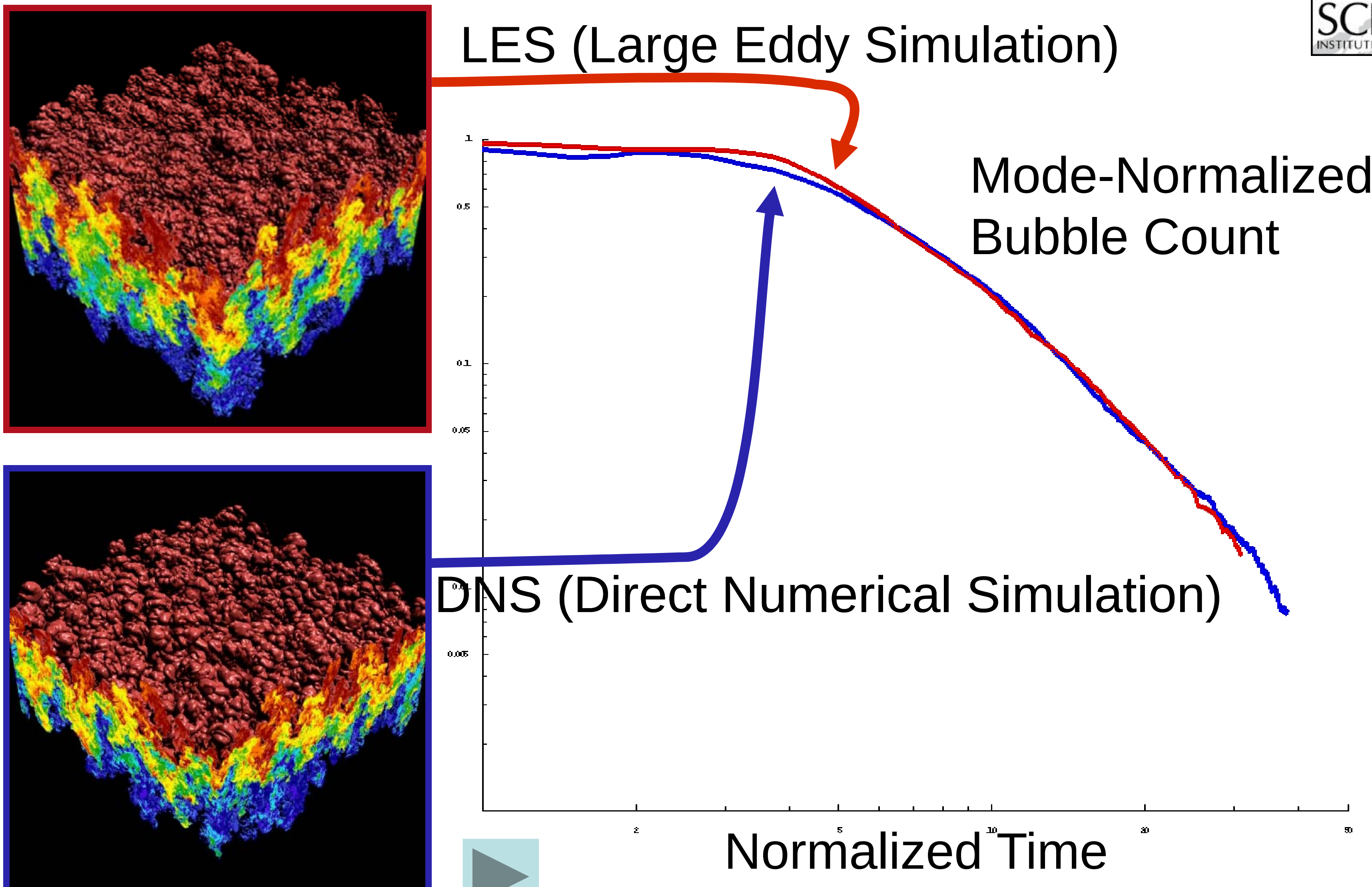
First Time Scientists Can Quantify Robustly Mixing Rates by Bubble Count



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

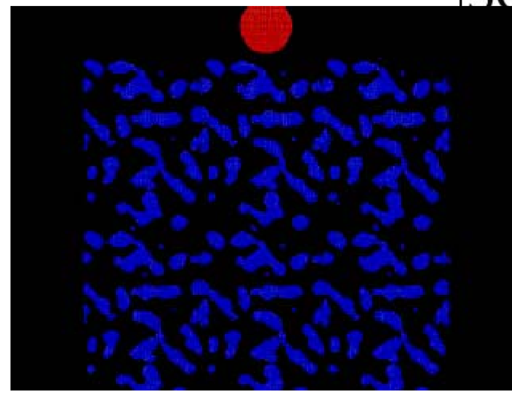


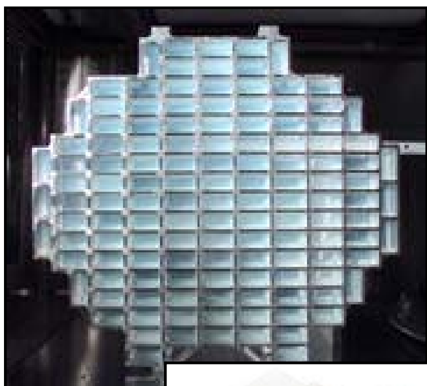
We Provided the First Feature-Based Validation of a LES with Respect to a DNS

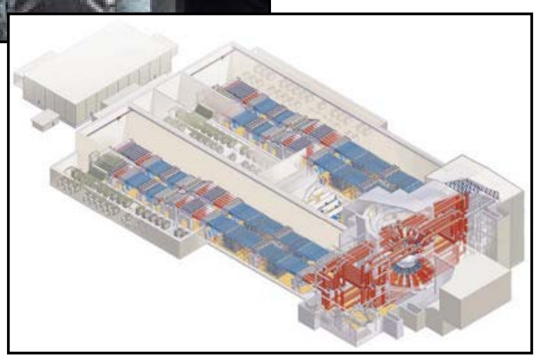


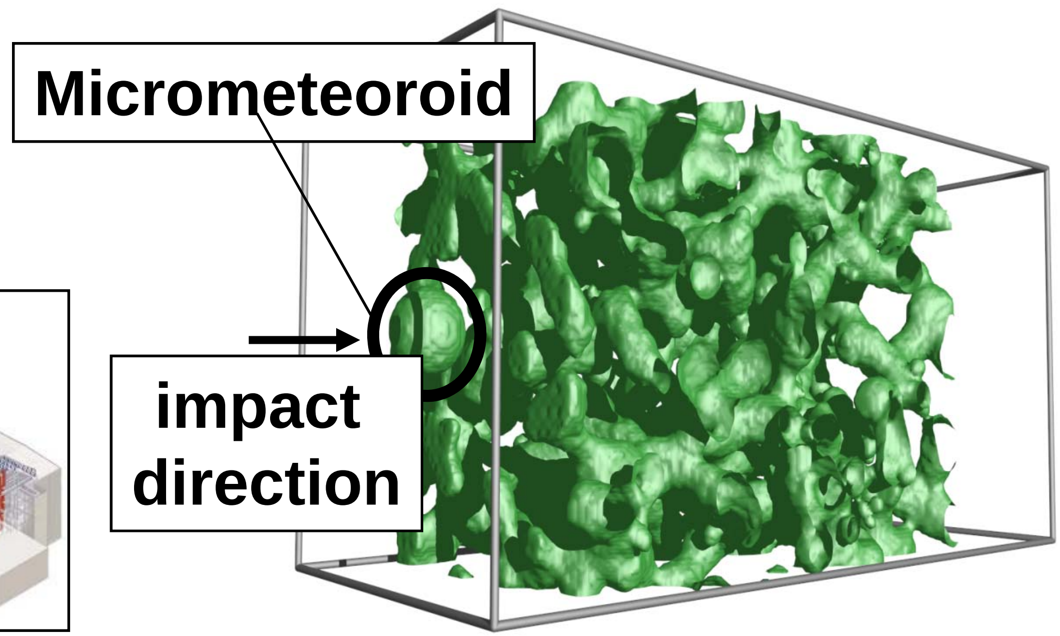
Quantitative Analysis of the Impact of a Micrometeoroid in a Porous Medium

- **Many possible applications:**
  - **NASA's Stardust Spacecraft**
  - **National Ignition Facility Targets**
  - **Light and Robust Materials**
  - **many more...**



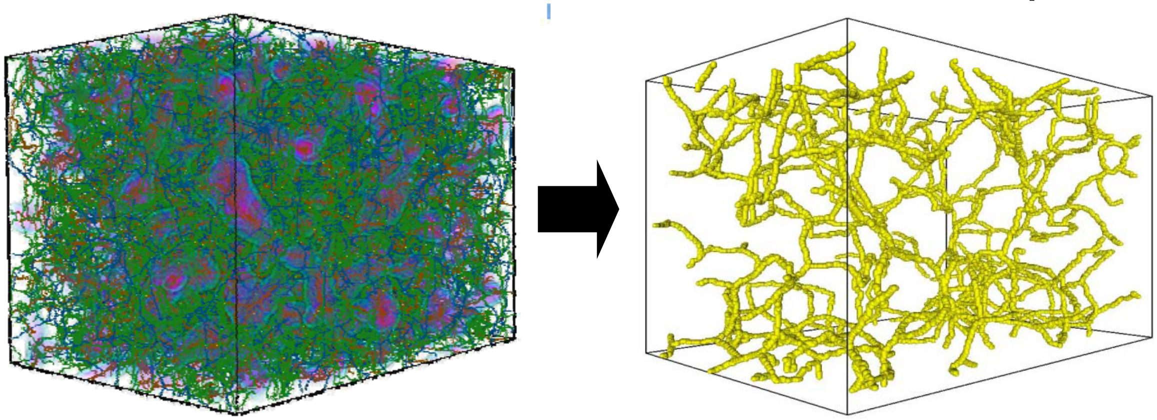




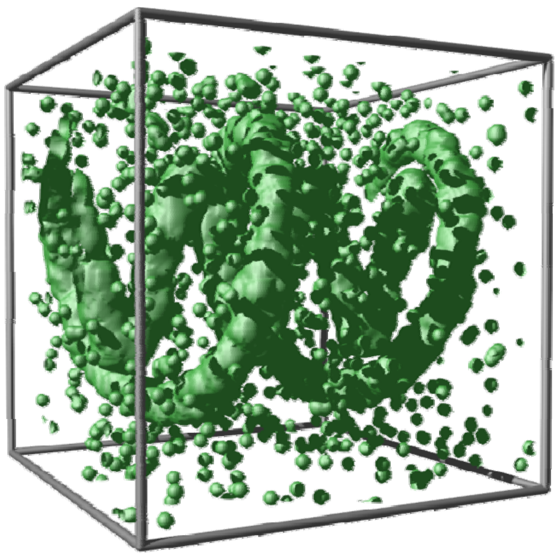


The Topological Reconstruction Method is Validated with a Controlled Test Shape

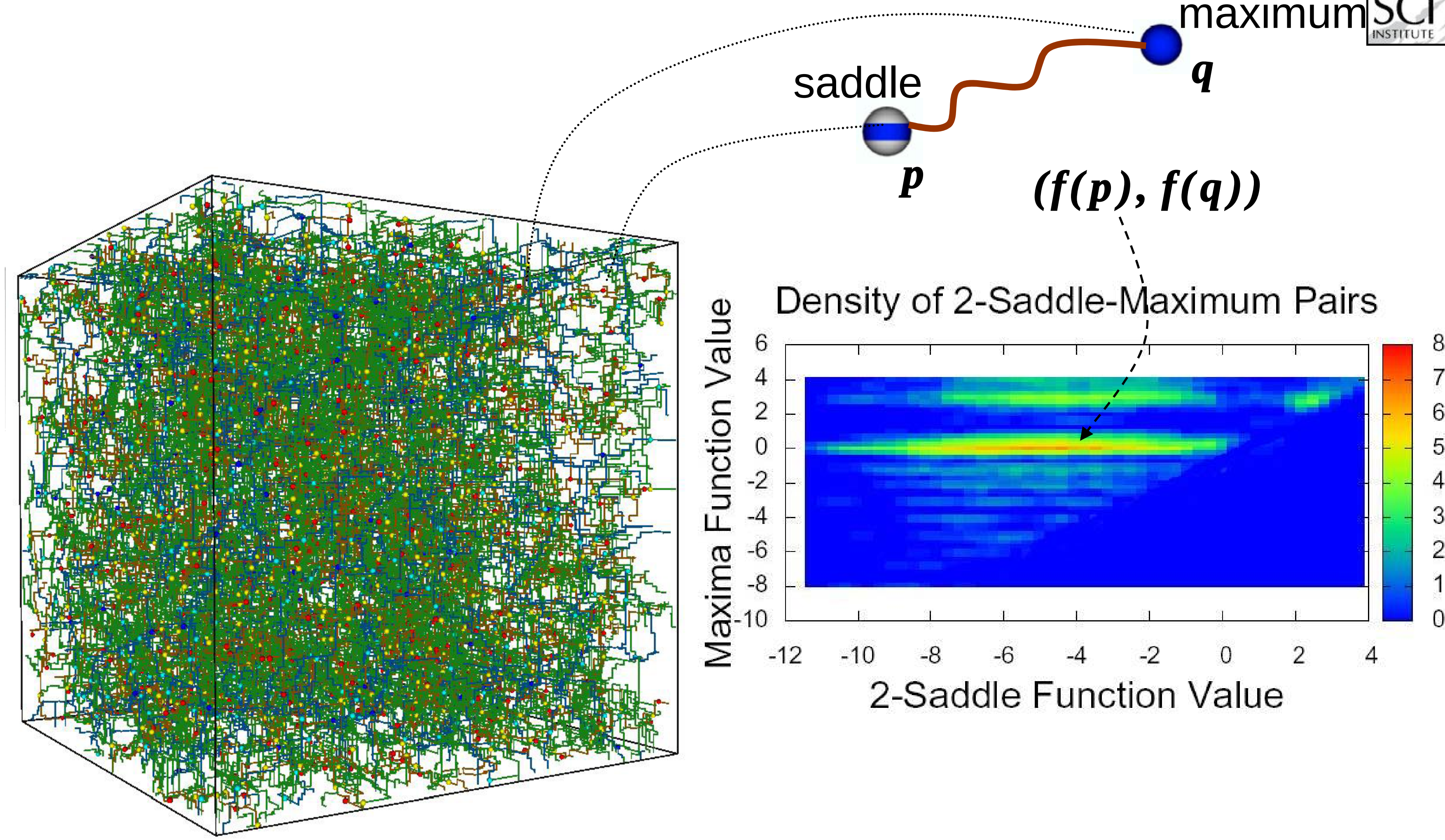
Challenge: robust reconstruction of the structure of a porous medium



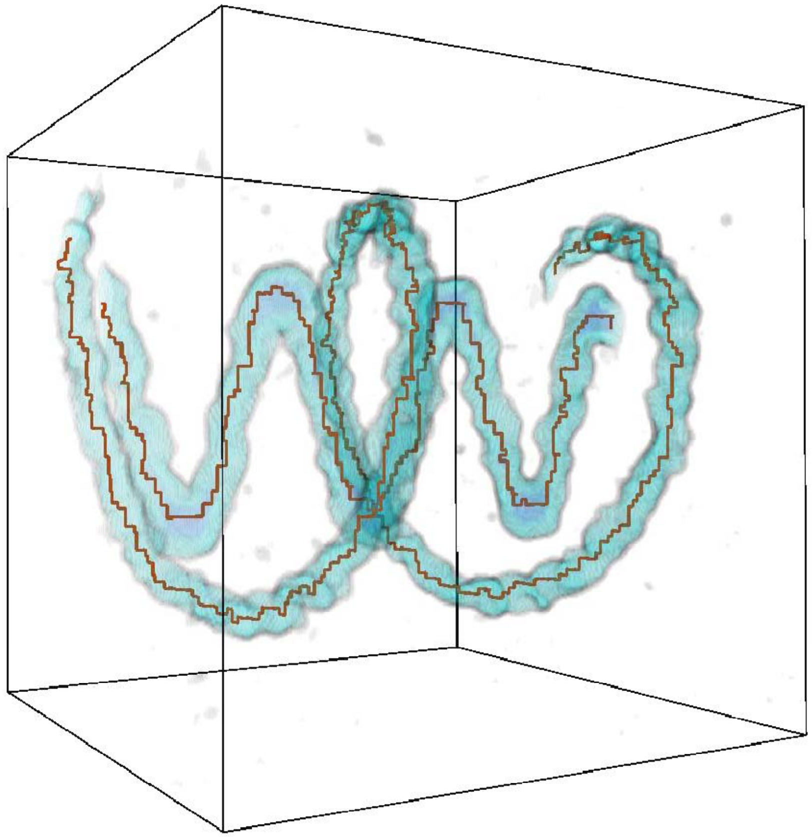
Preparation: we develop control test data to validate the approach

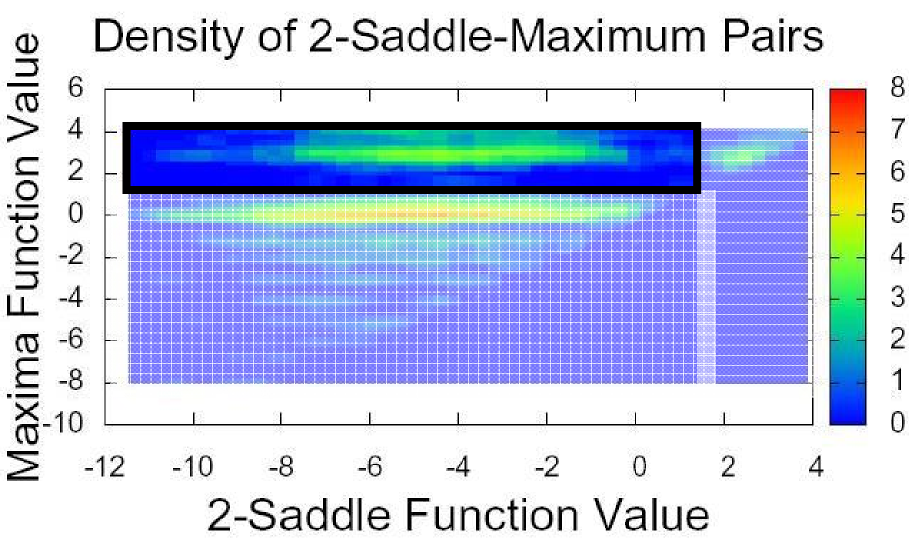


We Report the Distribution of Topological Features in the Full Resolution Data

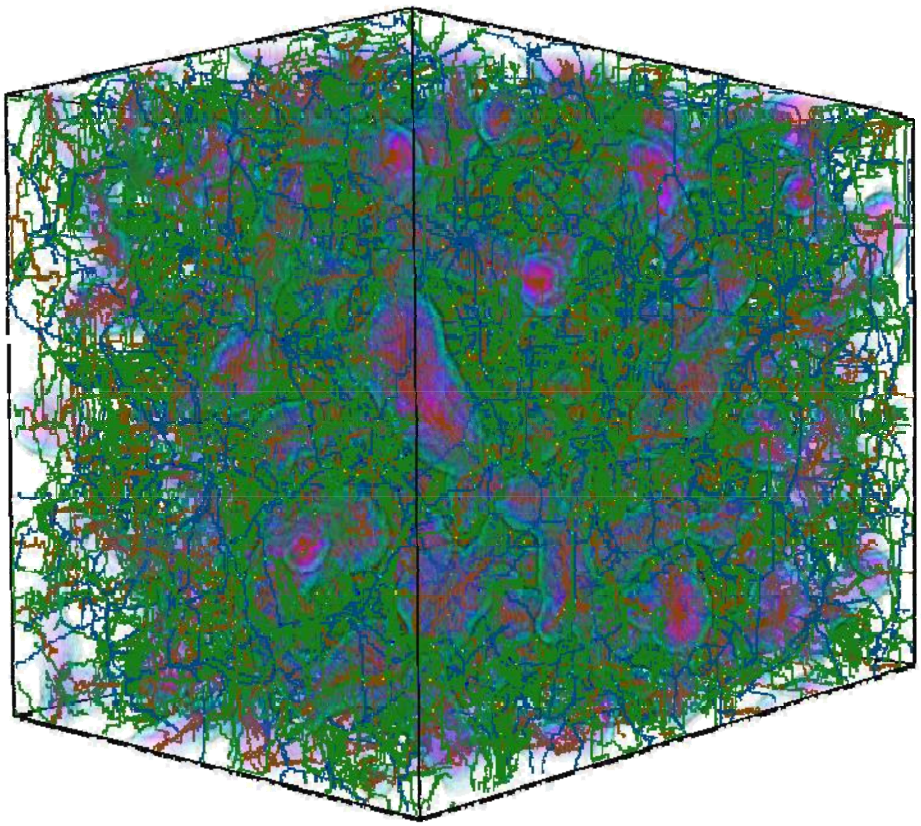


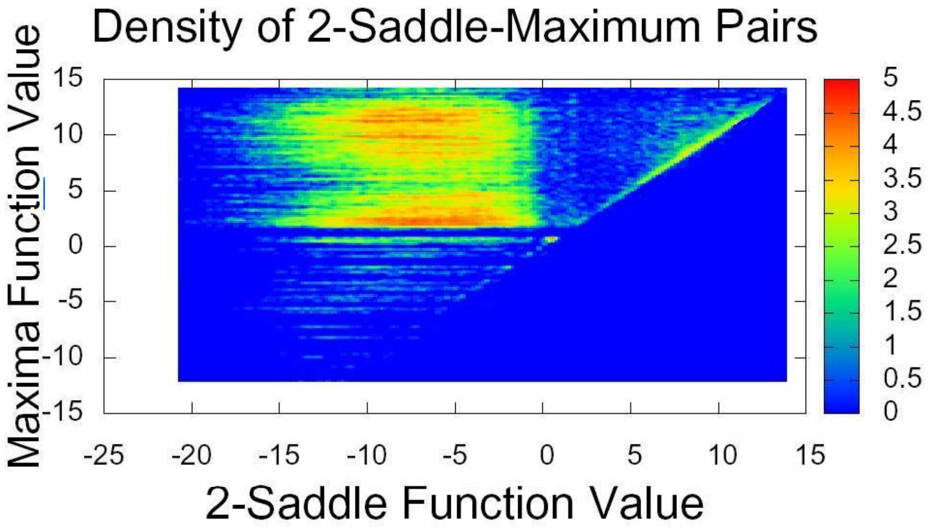
The Hierarchical Morse-Smale Complex Has Very Good Reconstruction Properties



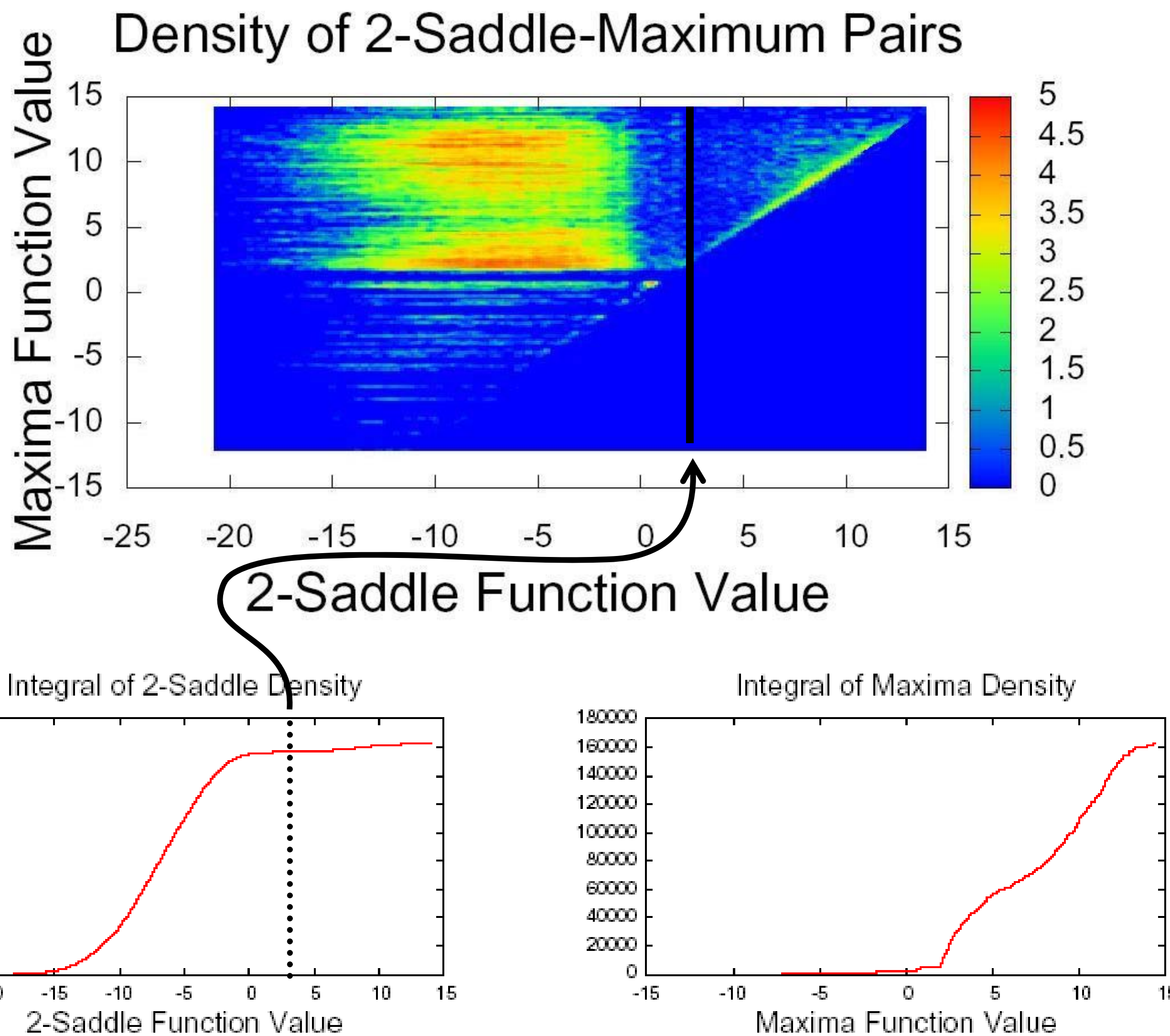


We Compute the Complete Morse-Smale Complex for the Porous Medium

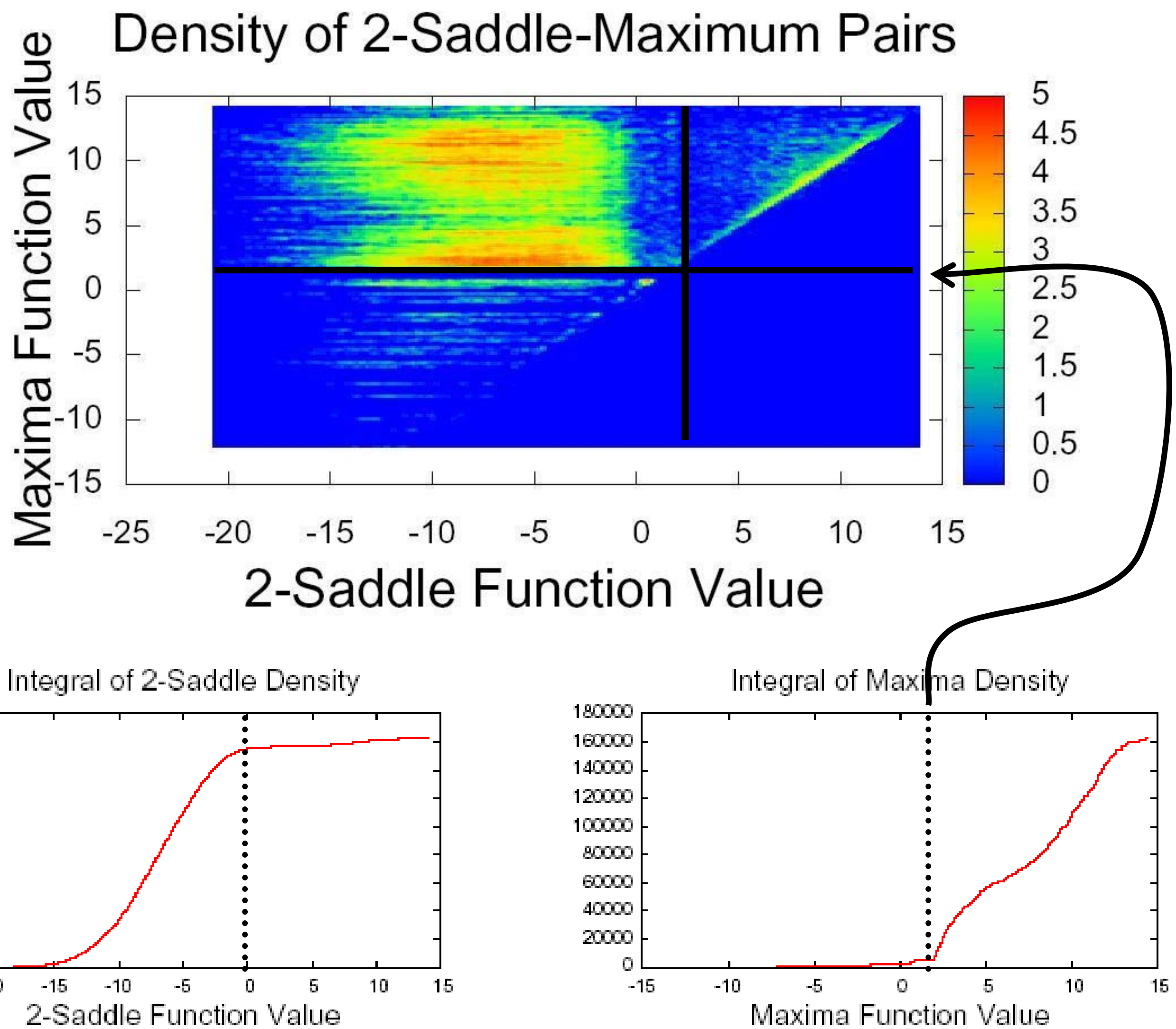




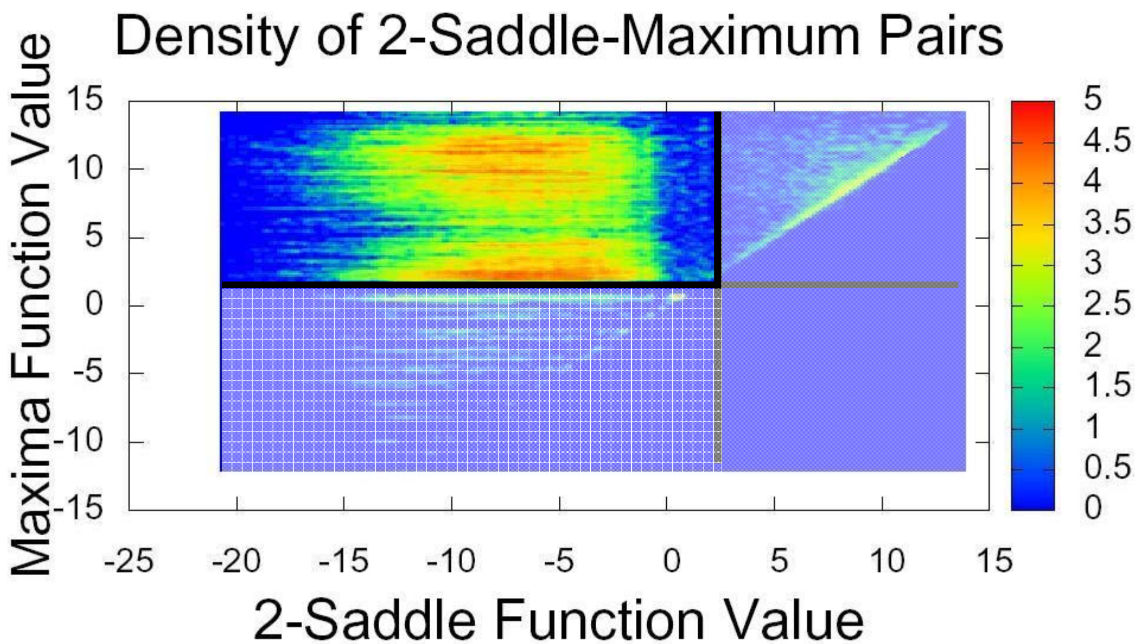
Need to Find Proper Threshold Values and Characterize the Stability of the Solution

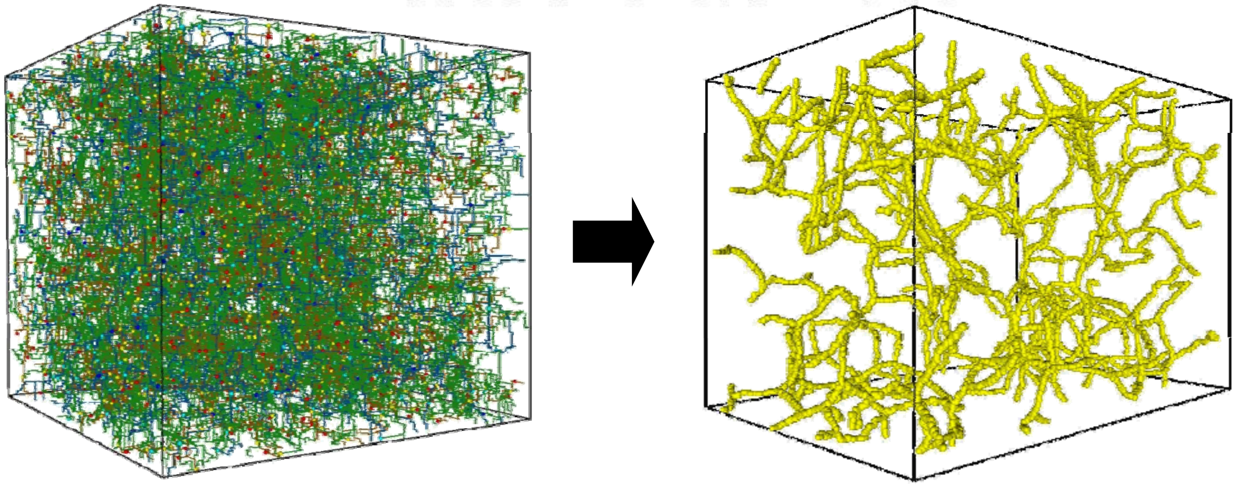


Need to Find Proper Threshold Values and Characterize the Stability of the Solution

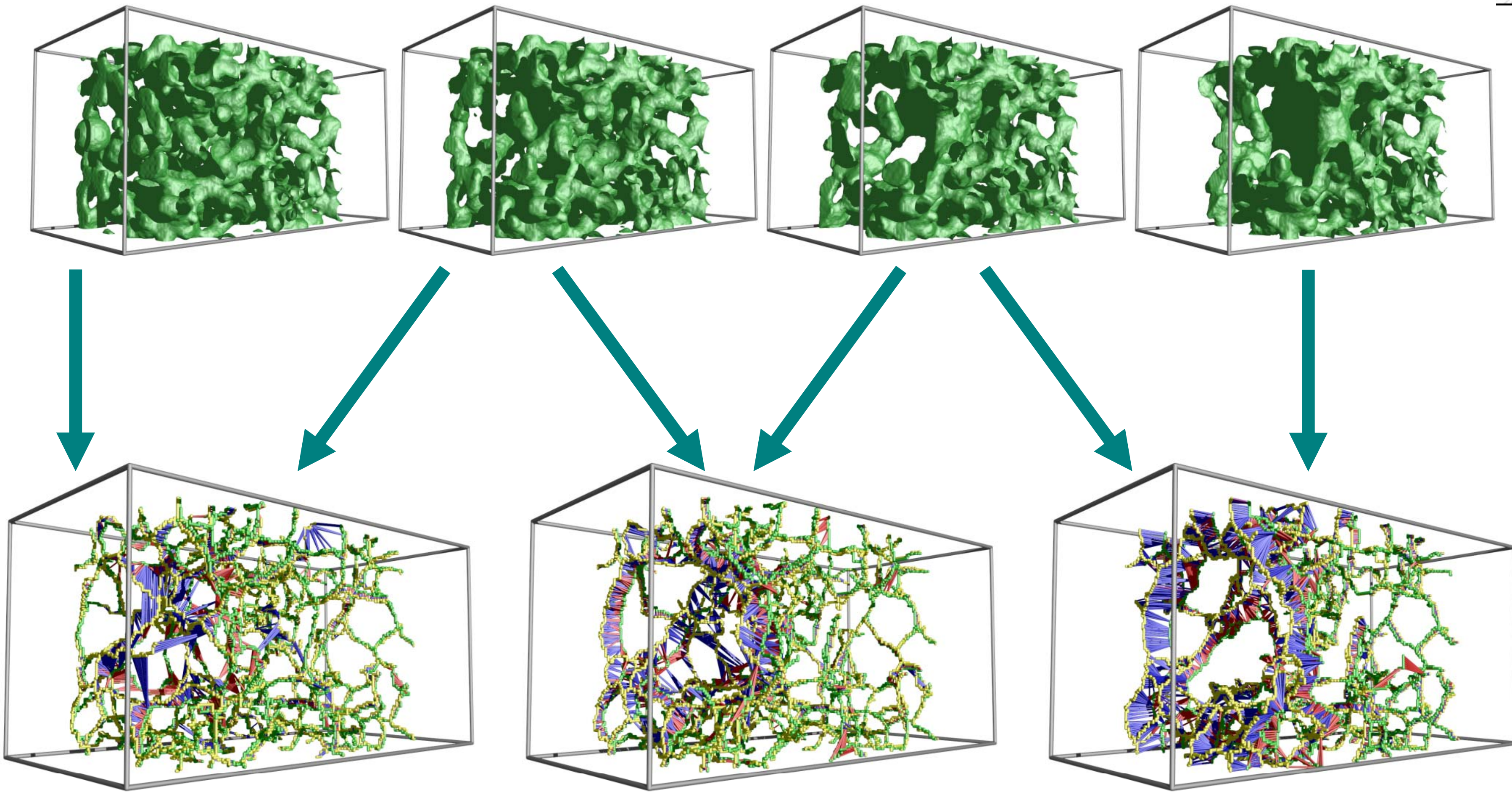


We Obtain a Robust Reconstruction of the Filament Structures in the Material





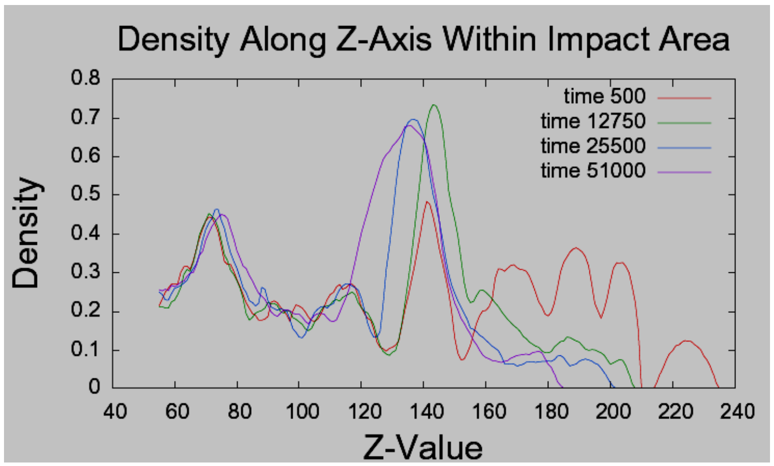
We Track the Evolution of the Filament Structure of the Material Under Impact

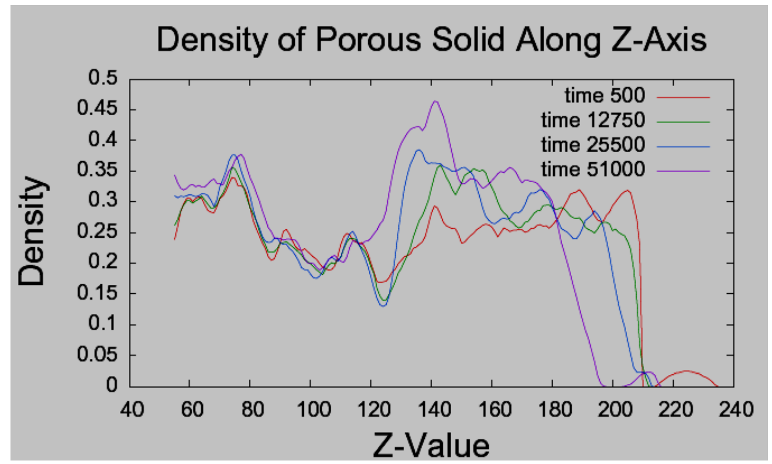


Time comparison of the reconstructions

The Extracted Structures Allow to Quantify the Change in Porosity of the Material

Density profiles





Decay in porosity of the material

Metric	t=500	t=12750	t=25500	t=51000
# Cycles	762	340	372	256
Total Length	34756	24316	23798	18912

Analysis of Massive Scientific Data Provide Great Challenges and Opportunities

- Tight cycle of :
  - basic research,
  - software deployment
  - user support
- Strong University-Lab collaboration

- Focused technical approach:
  - performance tools for fast data access
  - general purpose data exploration
  - error bounded quantitative analysis
  - feature extraction and tracking

- Interdisciplinary collaboration with domain scientists (from math to physics):
  - motivating the work
  - formal theoretical approaches
  - feedback to specific disciplines