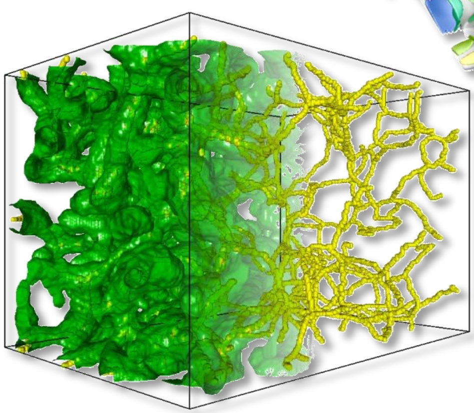
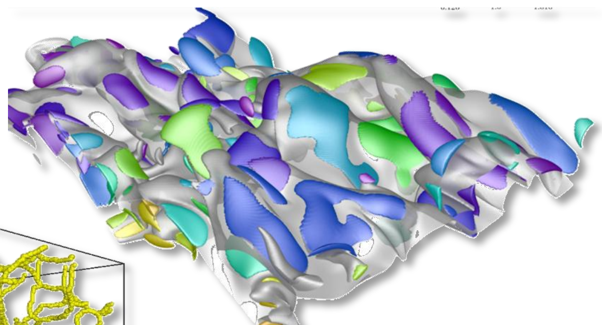
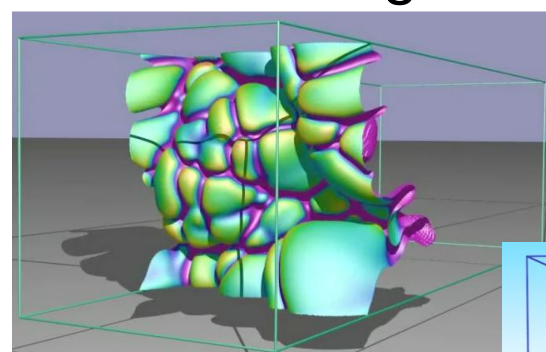
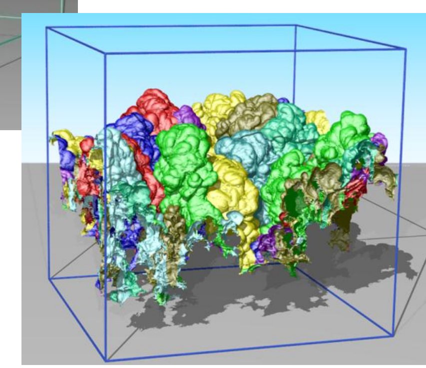
How has Computational Geometry helped Visualization and Analysis of Massive Scientific Data:
from the Design of Clean Fuels to Understanding Climate Change

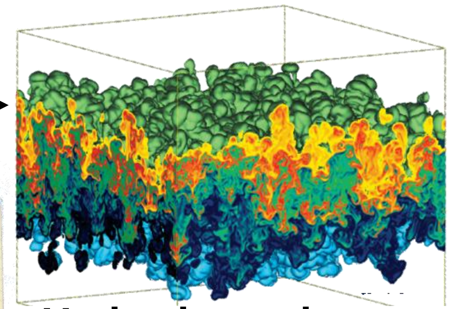





Valerio Pascucci

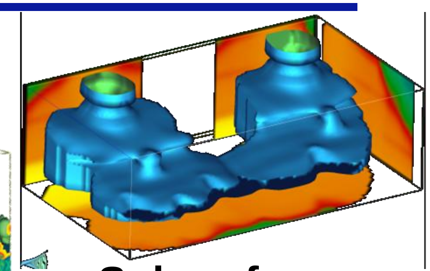
Director, Center for Extreme Data Management Analysis and Visualization
Professor, SCI institute and School of Computing, University of Utah
Laboratory Fellow, Pacific Northwest National Laboratory

Massive Scientific Models are Source of Great Challenges and Opportunities

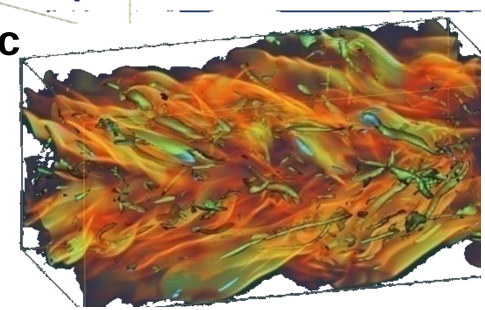
BlueGene/L



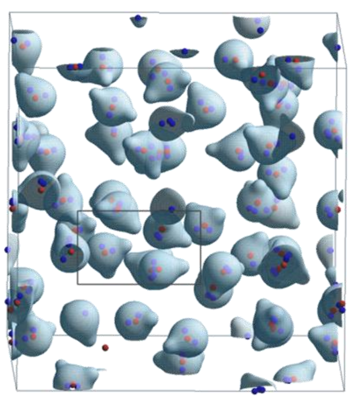
Hydrodynamic Instabilities



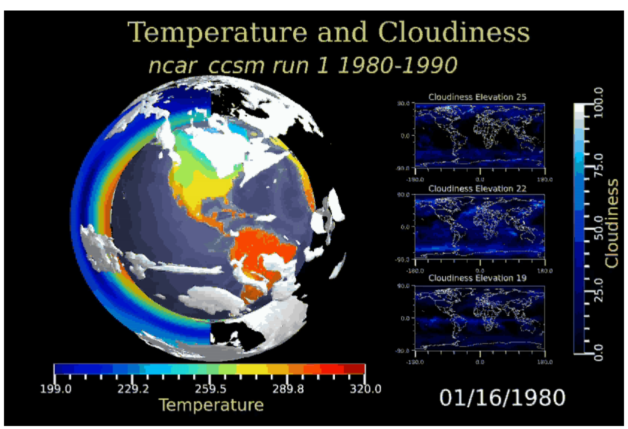
Subsurface Modeling



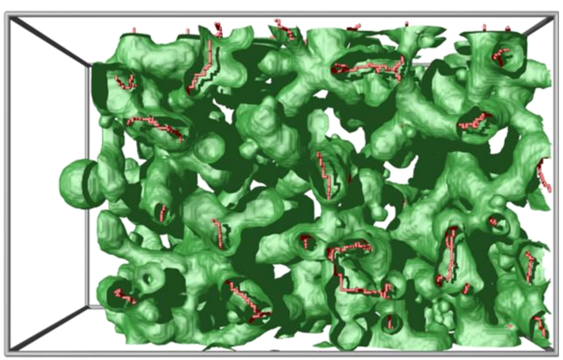
Combustion Simulations



Molecular Dynamics

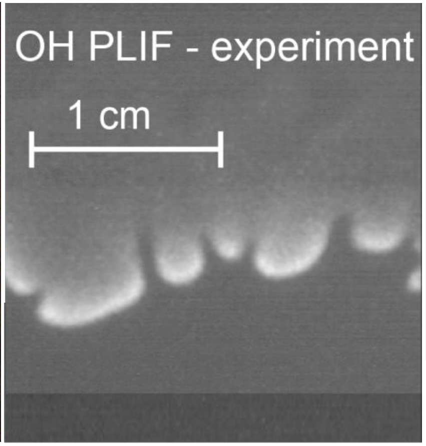


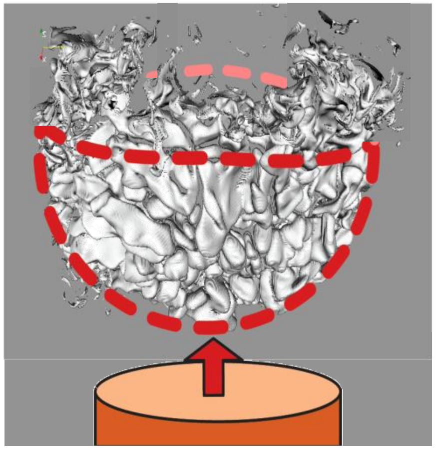
Climate Modeling

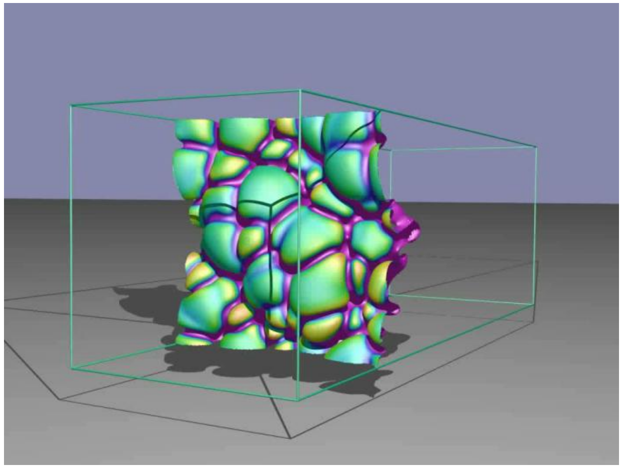


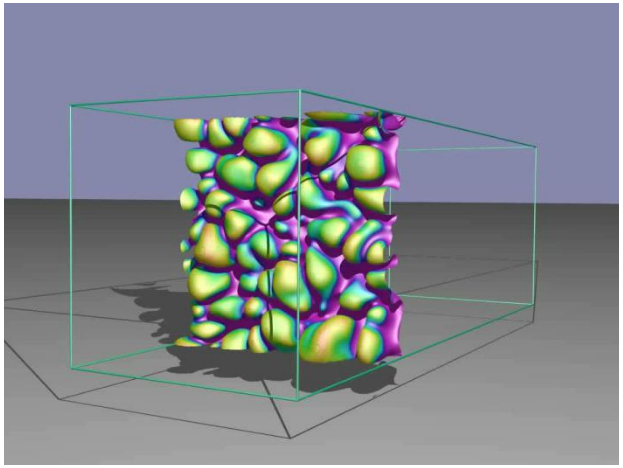
Material Sciences

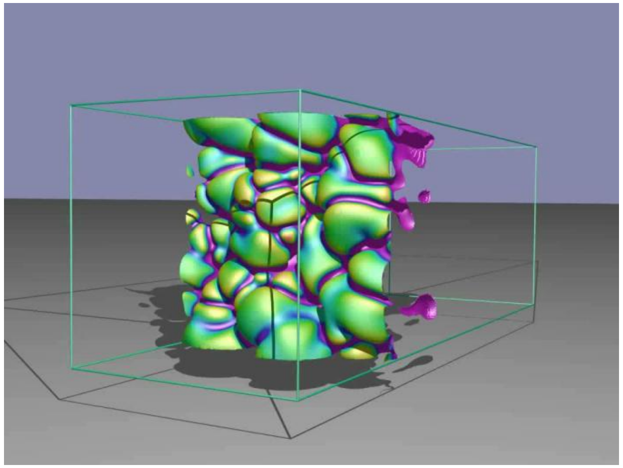
Given Three Burning Hydrogen Flames, Rank Them by Level of Turbulence



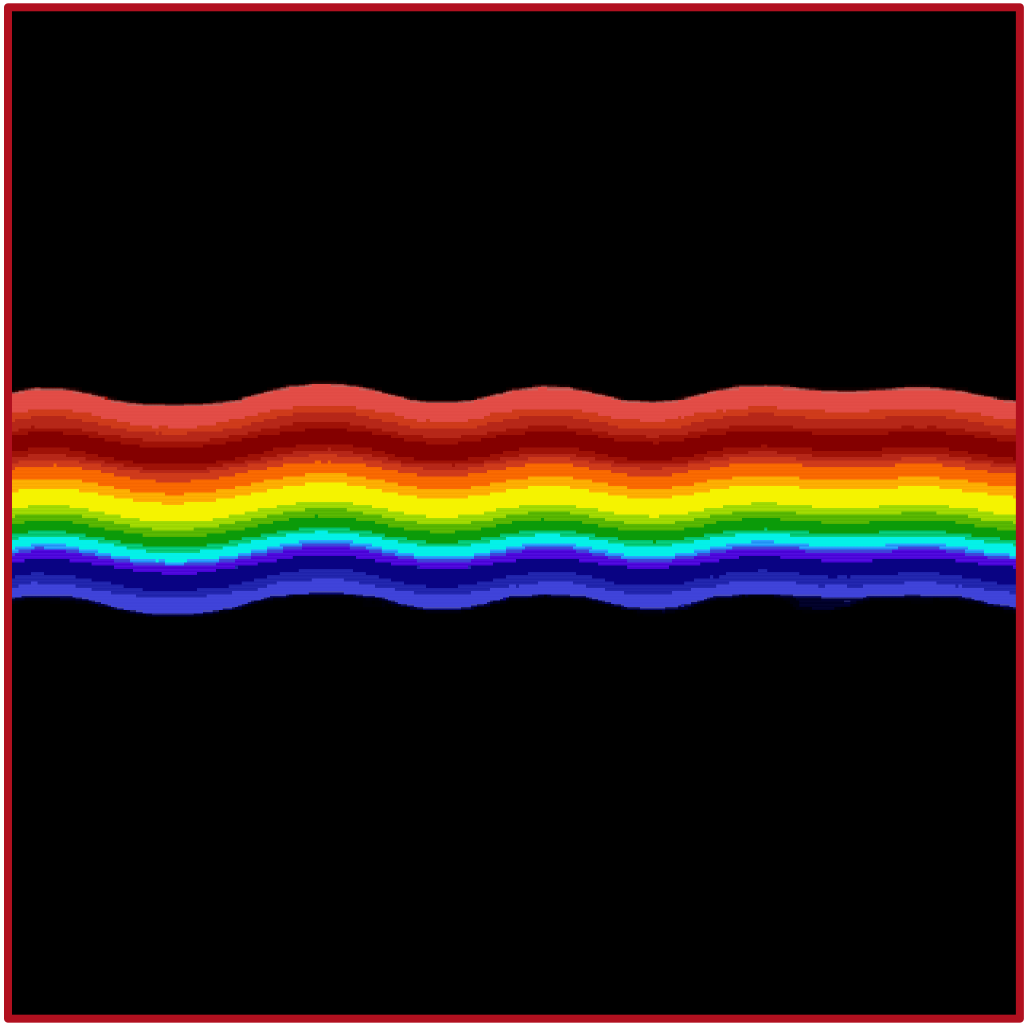








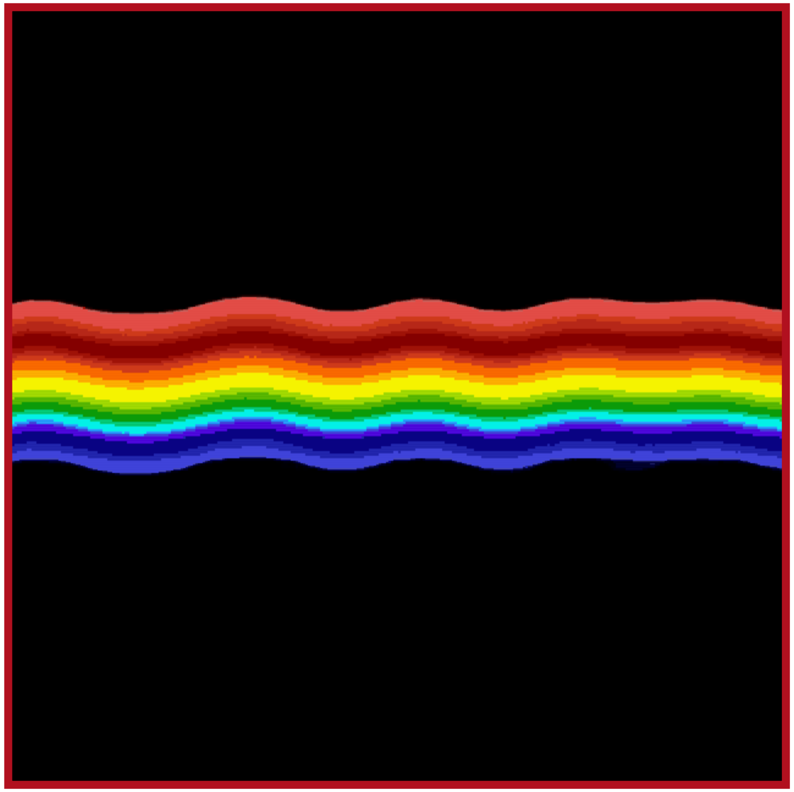
Count the Number of Bubbles in a Rayleigh–Taylor Instability



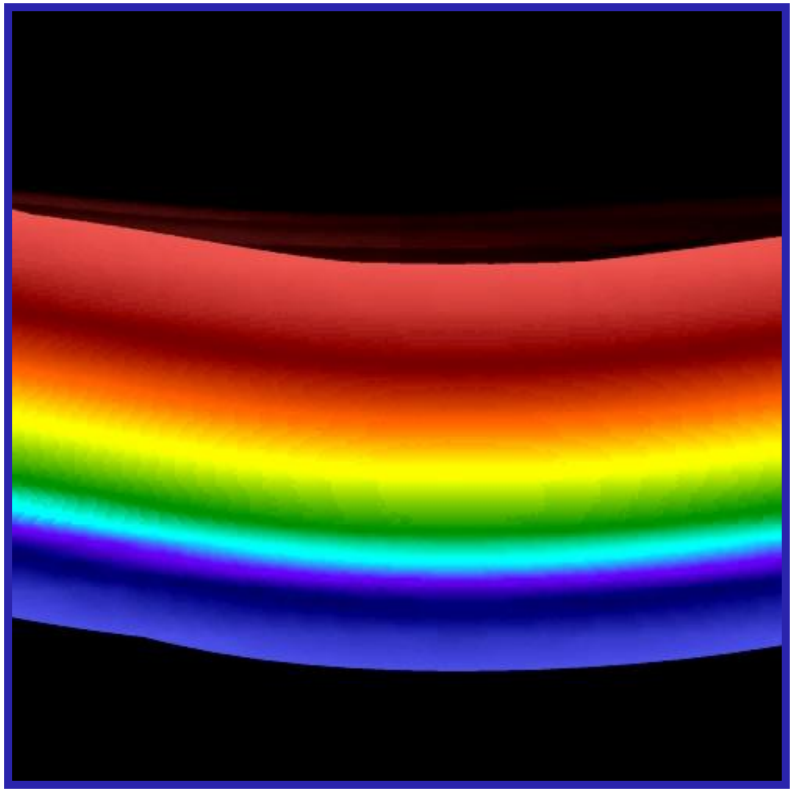
Rayleigh-Taylor instabilities arise in fusion, super-novae, and other fundamental phenomena:

- 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

Are these LES and DNS simulation codes expressing equivalent fenomena?

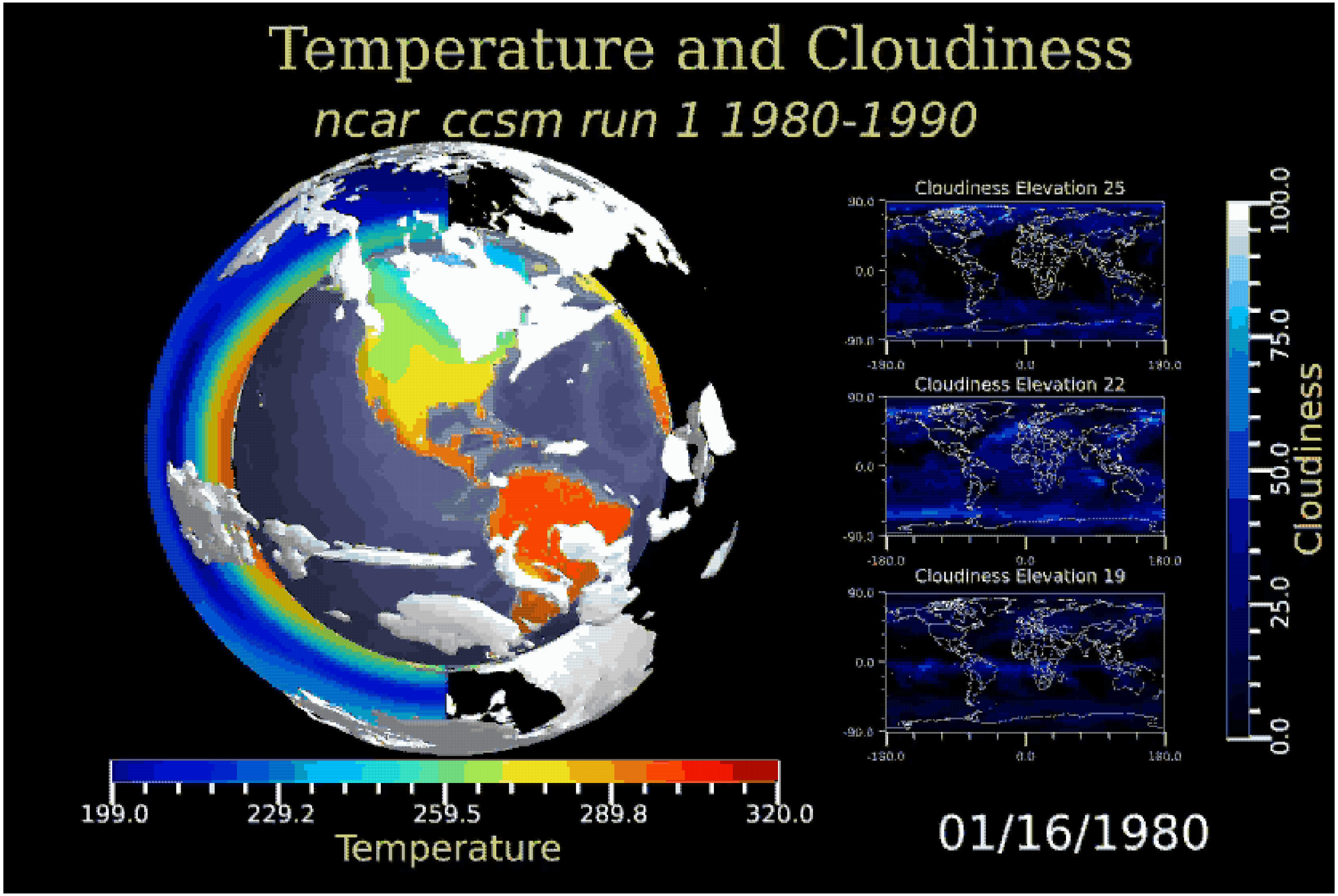


LES
Large Eddy Simulation

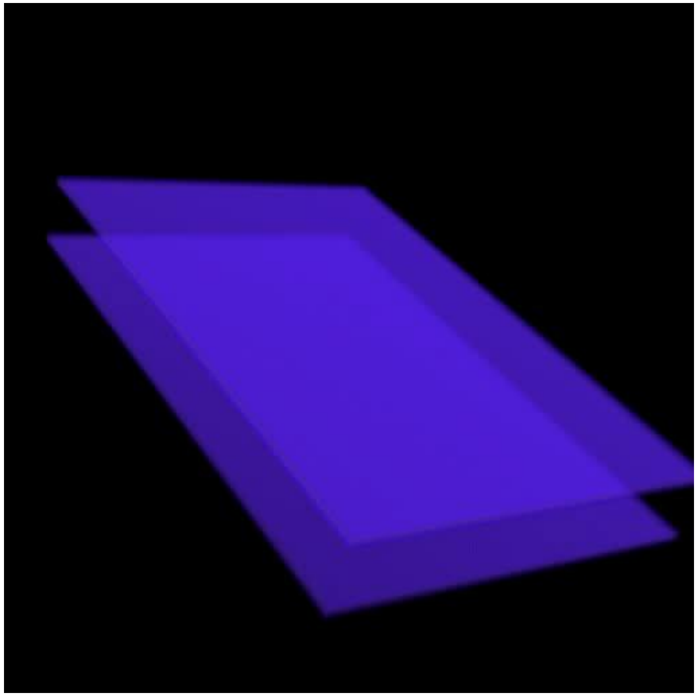


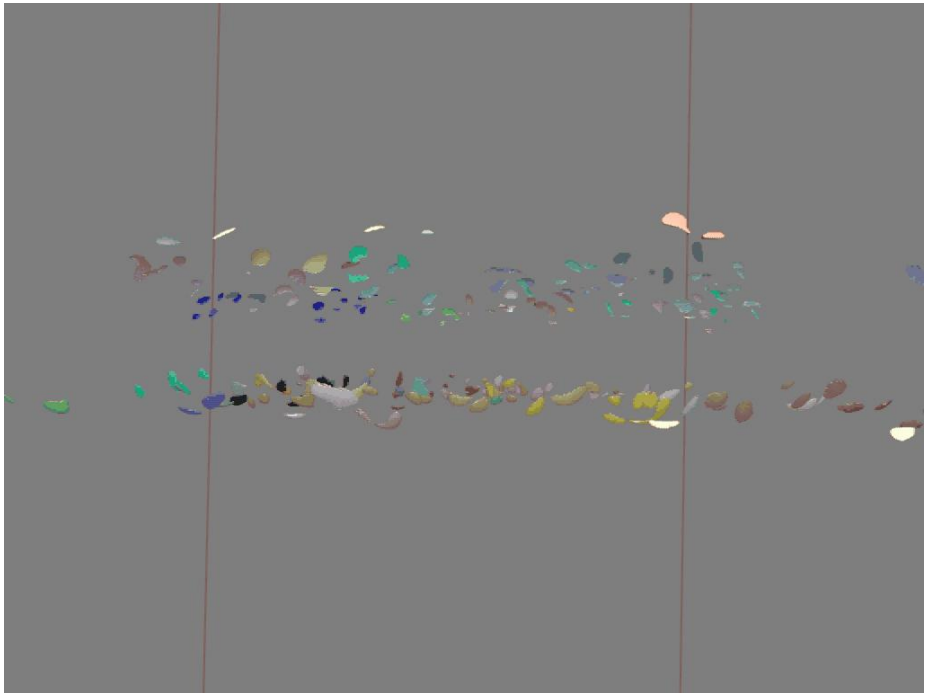
DNS
Direct Numerical Simulation

What is the Long Term Impact of Human Activities on the Global Climate?

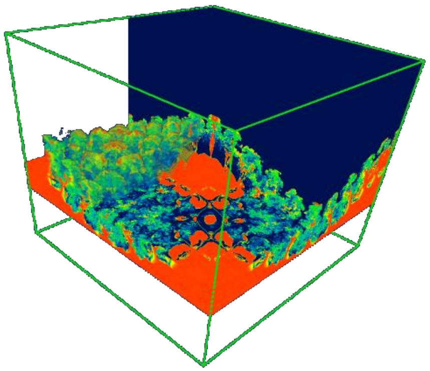
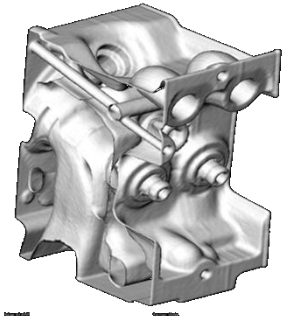
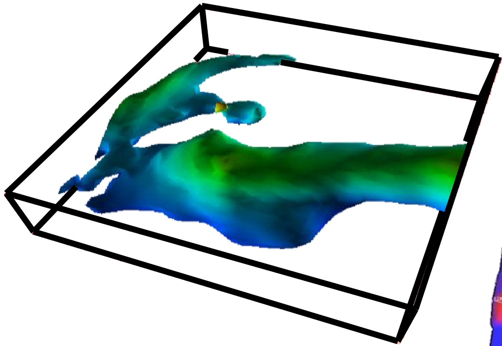
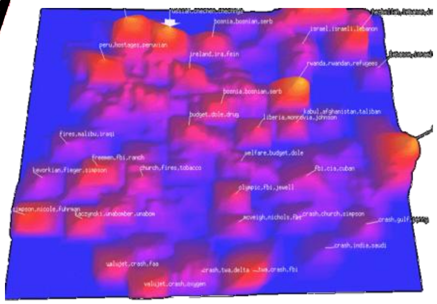
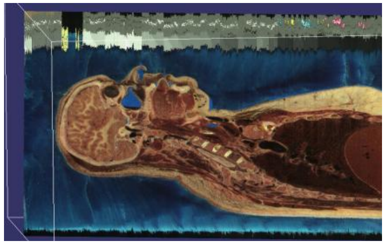


What Conditions Cause Extinction and Reignition in Premixed Hydrogen Flames

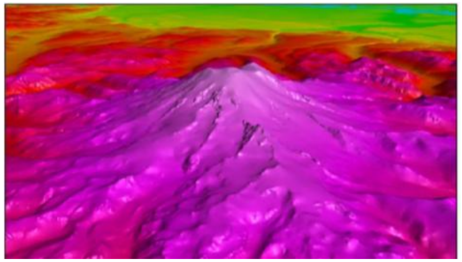
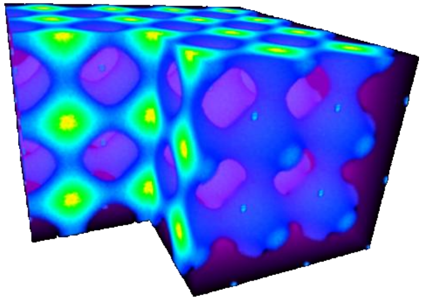
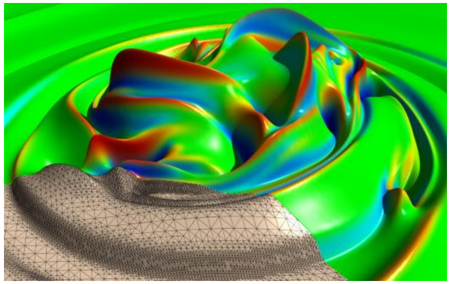
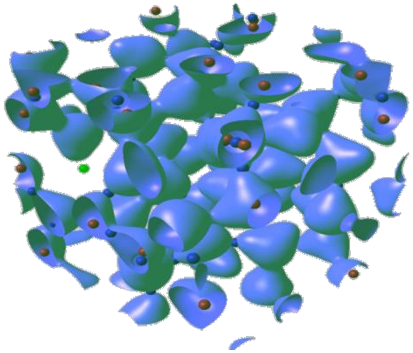
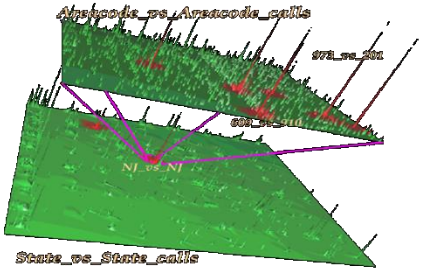




We Develop General Purpose Tools for Efficient and Reliable Data Understanding

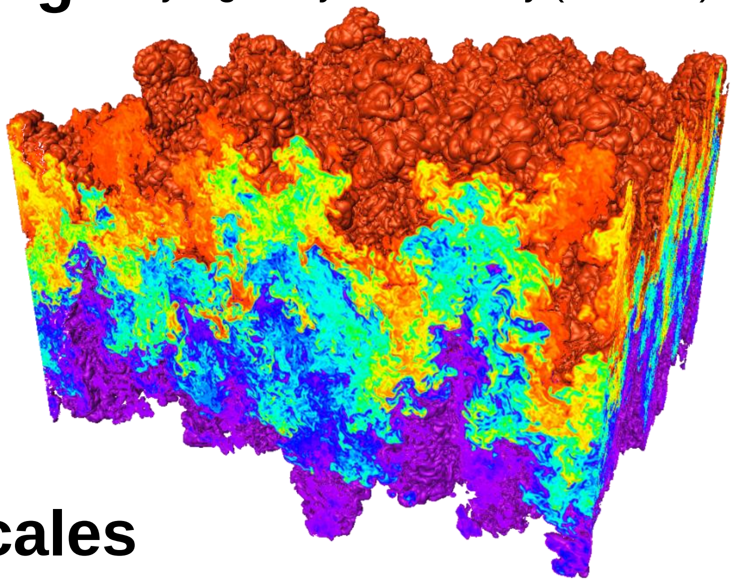
Real functions are ubiquitous in the representation of scientific information.

Traditional Data Analysis Tools are Often Ineffective for Massive Models

- **Massive models are challenging:**
  - **Sheer volume of information**
  - **Complexity of the information represented**
  - **Complexity of presentation**

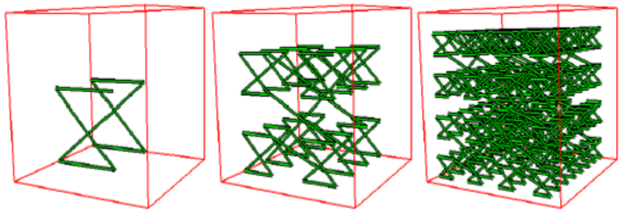
Rayleigh–Taylor instability (Miranda)



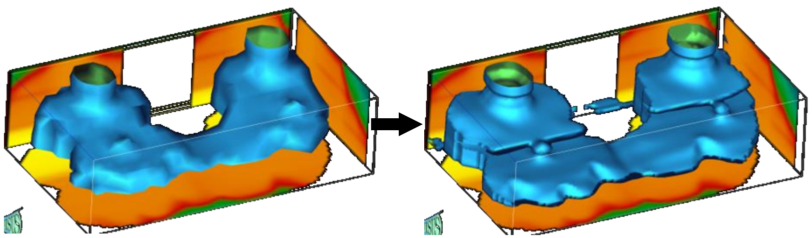
- **Tools do not scale with the data sizes**
- **Difficult to capture multiple scales**
- **Numerical methods unstable and sensitive to noise**
- **Difficulty in providing error bounds associated with the coarse scale analysis**
- **Need new abstractions and metaphors to covey information reliably and efficiently**

We Develop New Techniques for Efficient Data Management and Presentation

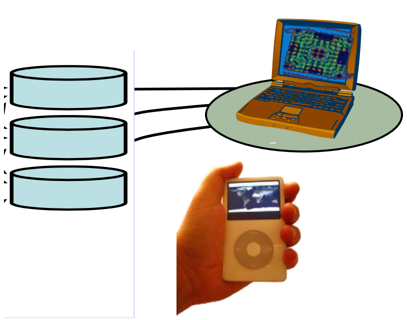
- **Advanced data storage techniques:**
  - Data re-organization.
  - Compression.



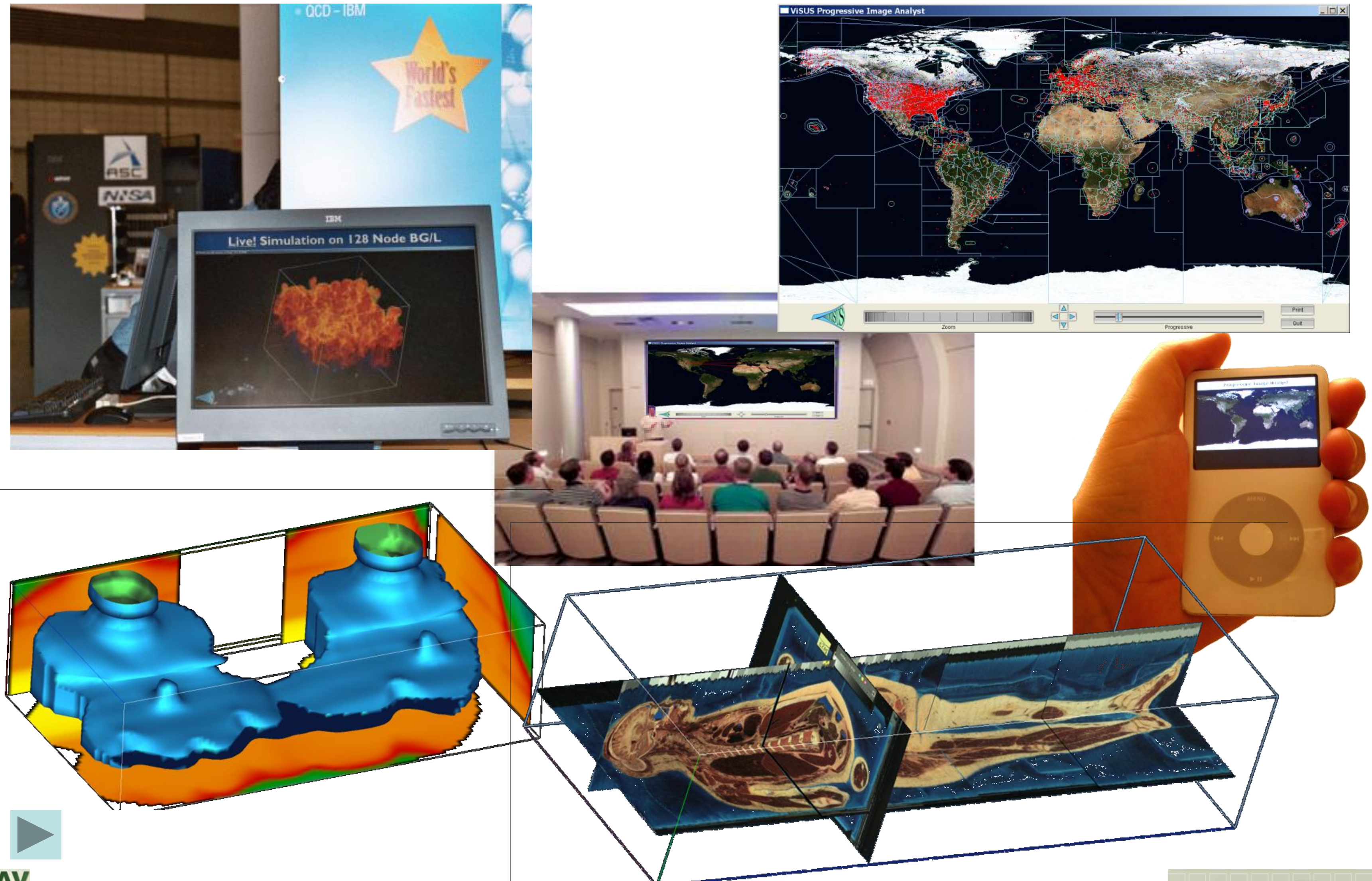
- **Advanced algorithmic techniques:**
  - Streaming.
  - Progressive multi-resolution.
  - Out of core computations.



- **Scalability across a wide range of running conditions:**
  - From laptop, to office desktop, to cluster of PC, to BG/L.
  - Memory, to disk, to remote data access.

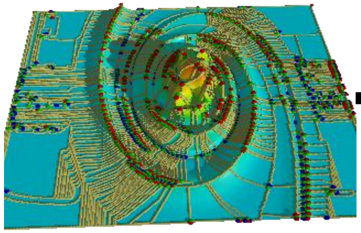


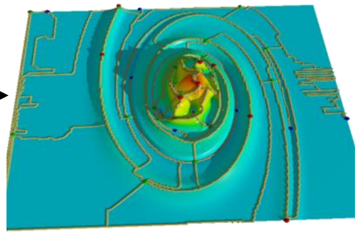
We Demonstrated Performance and Scalability in a Variety of Applications



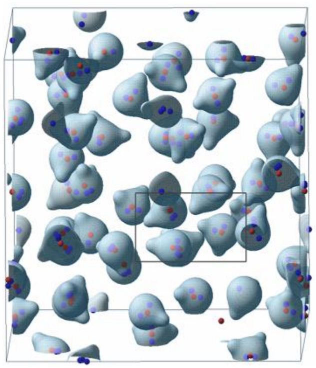
Our Framework is Based on Robust Topological Computations for Quantitative Data Analysis

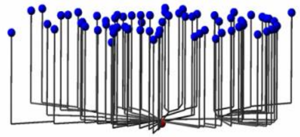
- **Provably robust computation**
- **Provably complete feature extraction and quantification**
- **Hierarchical topological structures used to capture multiple scales**
- **Error-bounded approximations associated with each scale**
- **Formal mathematical definition associated with each analysis**
-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



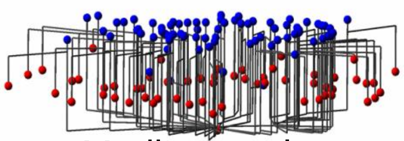


Hierarchical topology of a 2D Miranda vorticity field





Coarse scale:
blue = molecules

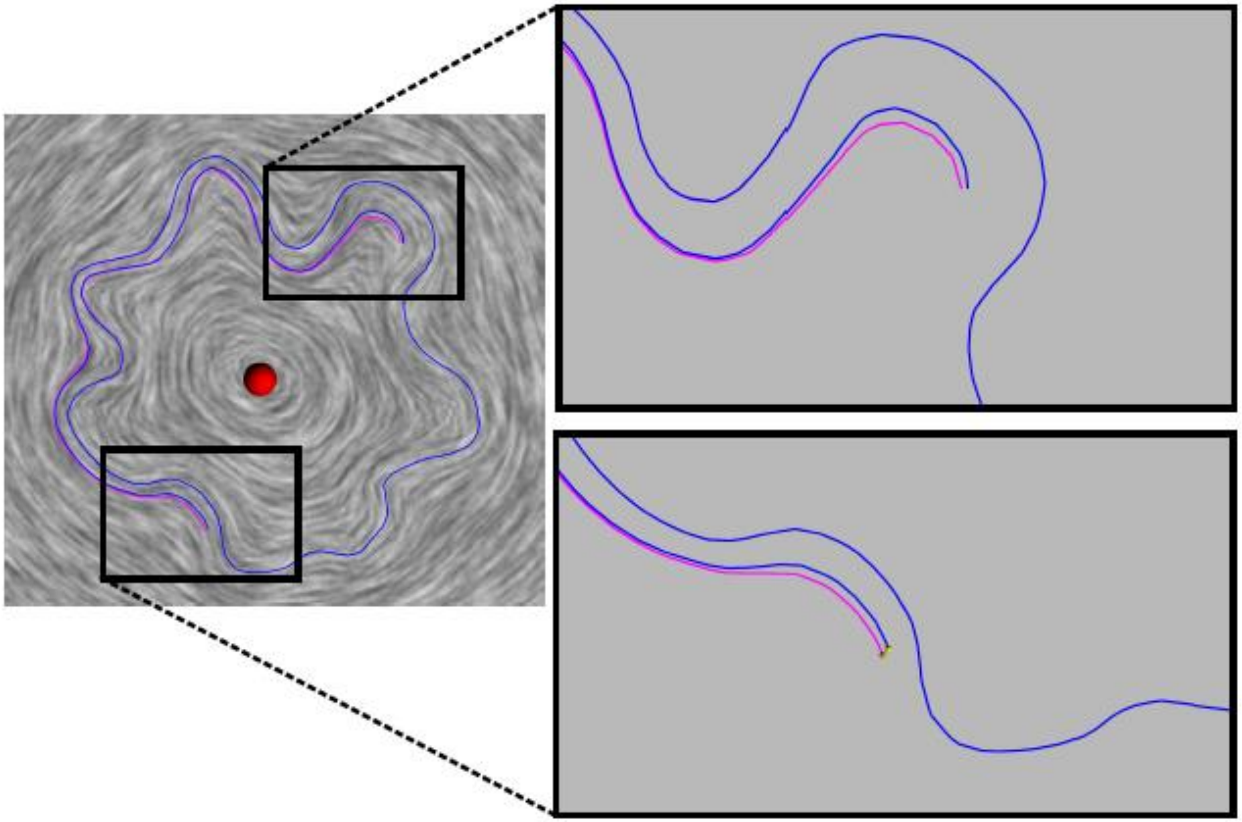


Medium scale:
red-blue = dipoles

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

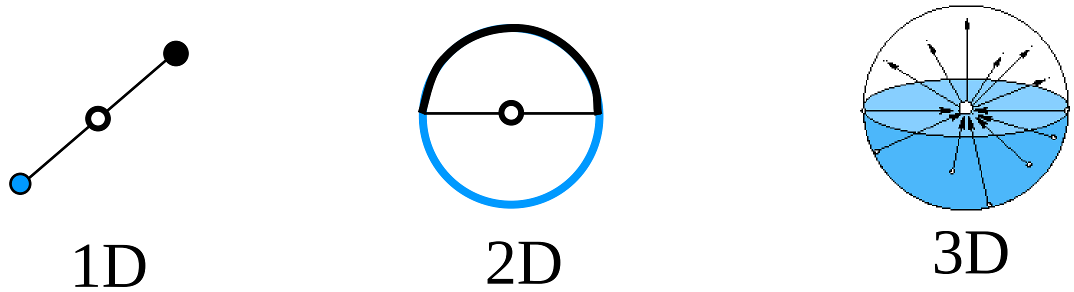
Motivation

- **The presence of computational errors in streamline computation using integration techniques can lead to inconsistent results**



We Adopted a Combinatorial Approach to Morse Theory for Provably Correct Computations

	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	**numerical**	**combinatorial**



Independent local computation yield globally consistent results

We Use Robust Techniques for Critical Point Detection and Classification

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

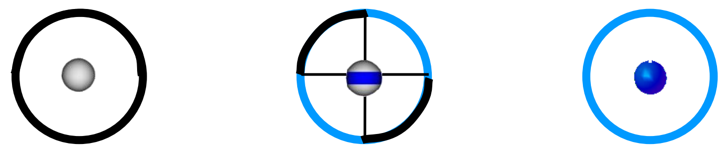
$$\nabla f(p) = 0$$

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

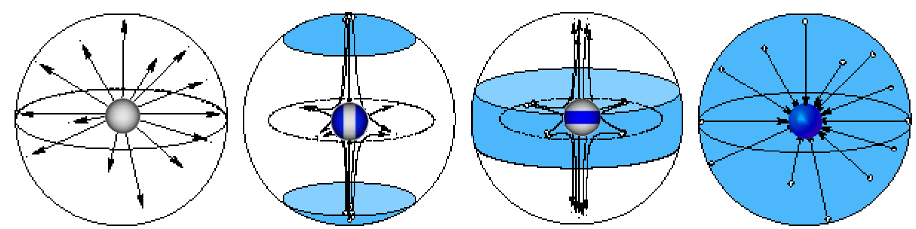
The Morse Lemma

There are *d+1* types of critical points

Minimum Maximum



Minimum Saddle Maximum

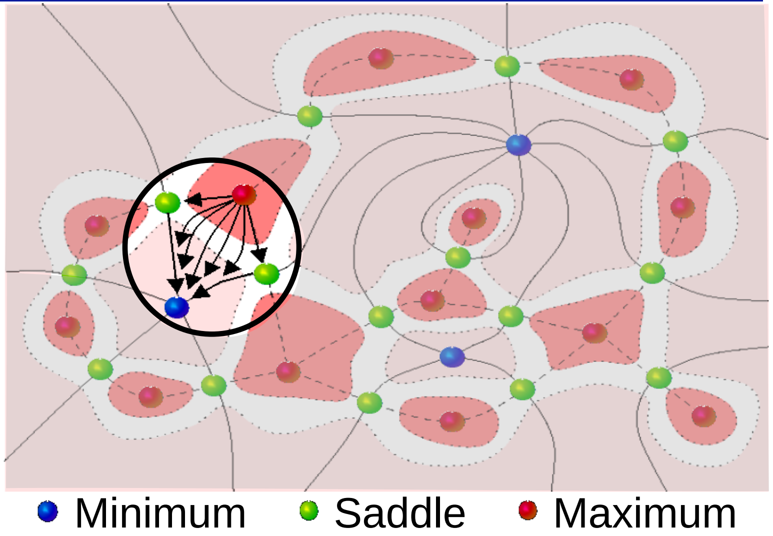


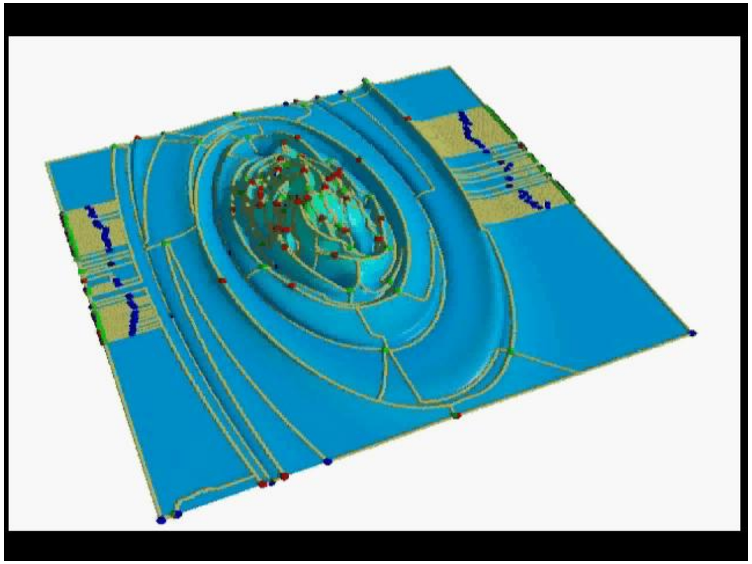
Minimum 1-saddle 2-saddle Maximum

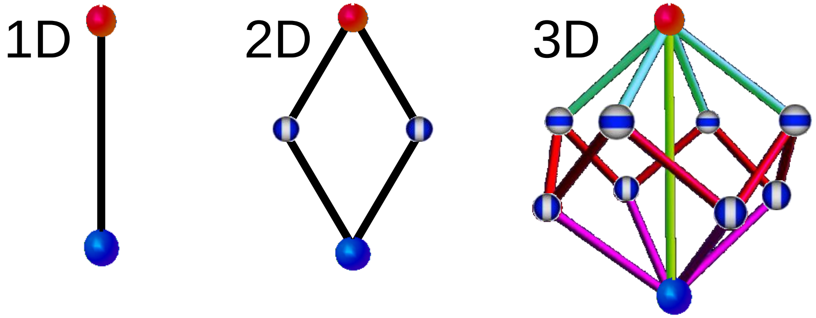
numerical **combinatorial**

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**







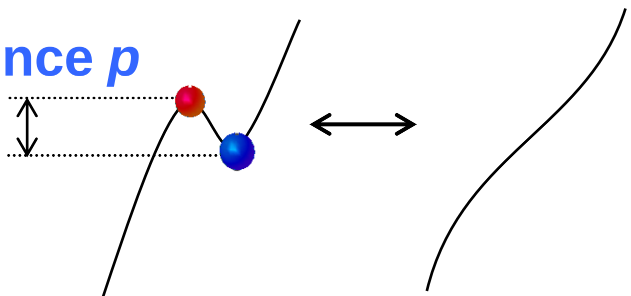
We Use Cancellations to Create a Multi-scale Representation of the Trends in the Data

Cancellations: critical points can be created or destroyed in pairs that are connected 1-manifolds

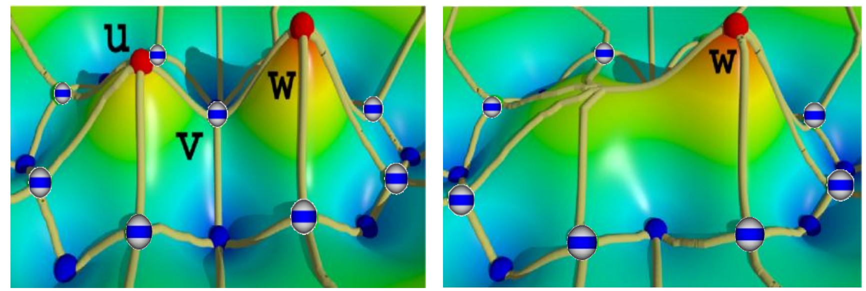
Approximation: error = persistence/2 (proven lower bound)

Multi-scale: consistent gradient segmentation at all scales

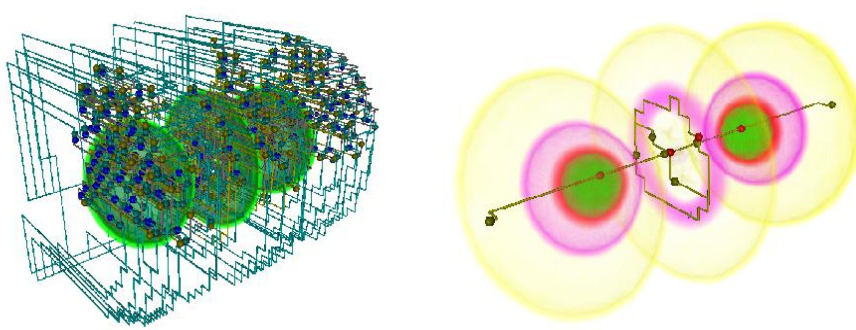
persistence *p*



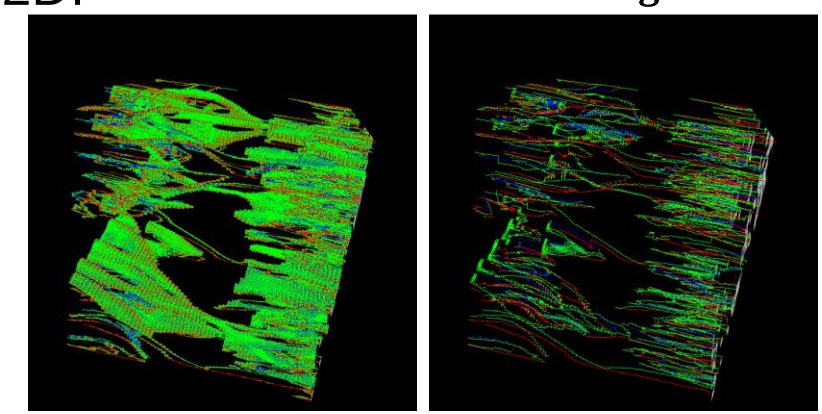
1D: **cancellation=contraction**



2D: **cancellation=contraction + edge removal**

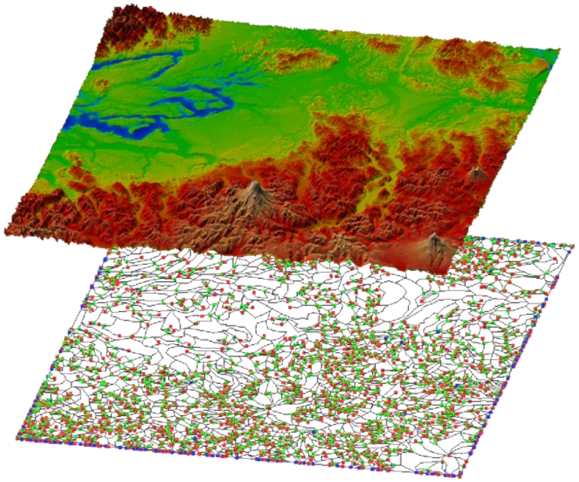


3D

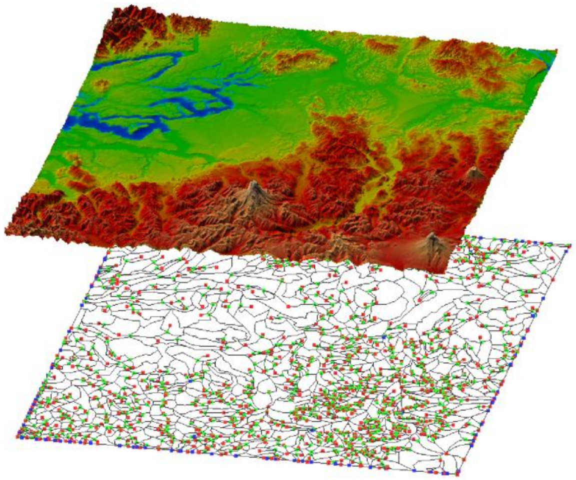


Time

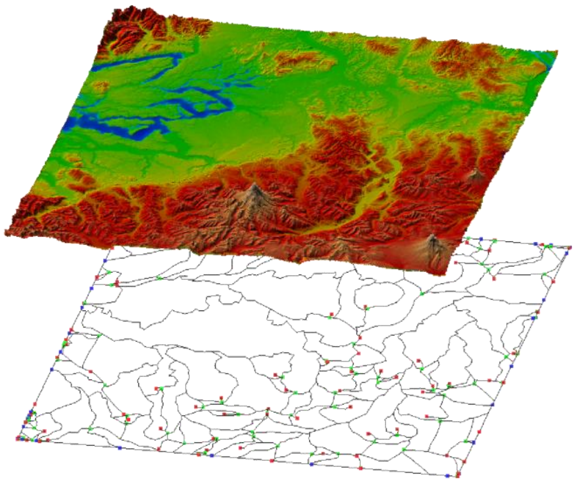
Contraction of extrema with saddles allows to simplify the MS-complex.



0.5% persistence filter

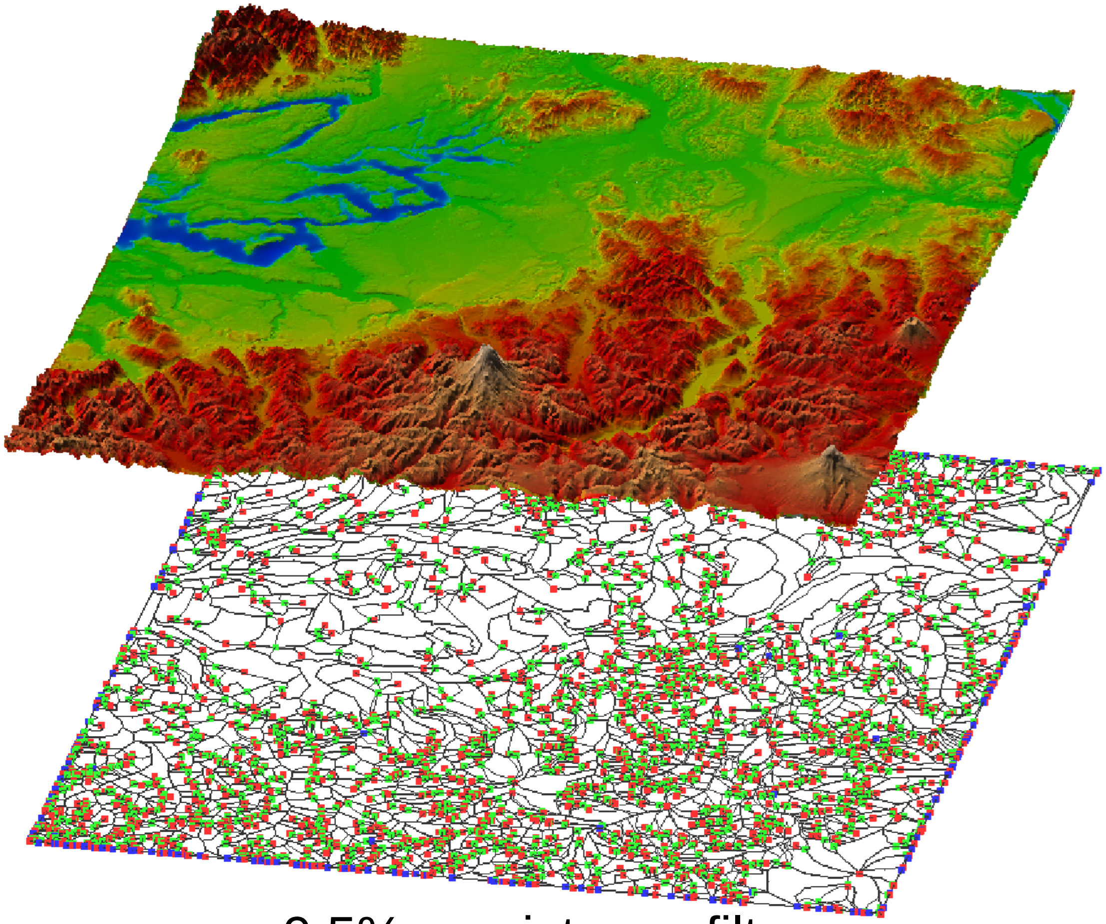


1.0% persistence filter



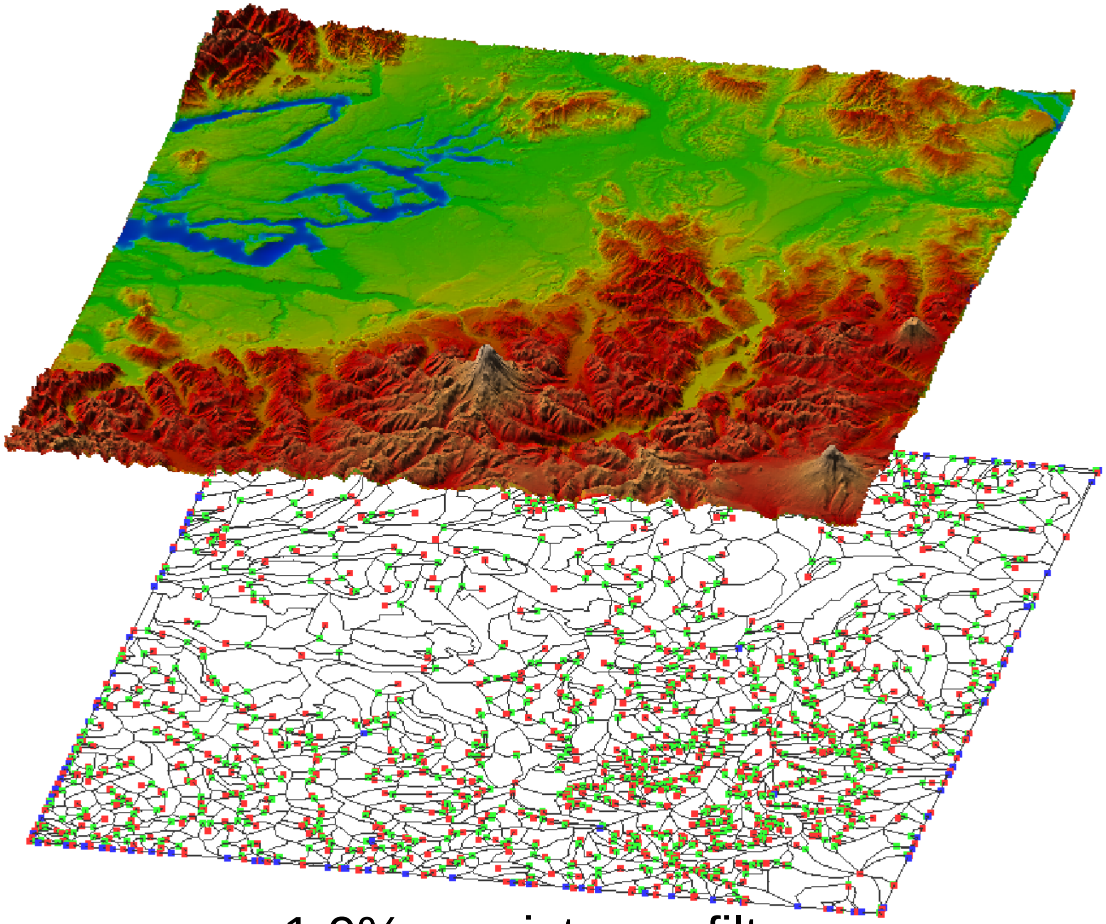
20% persistence filter

Contraction of extrema with saddles allows to simplify the MS-complex.



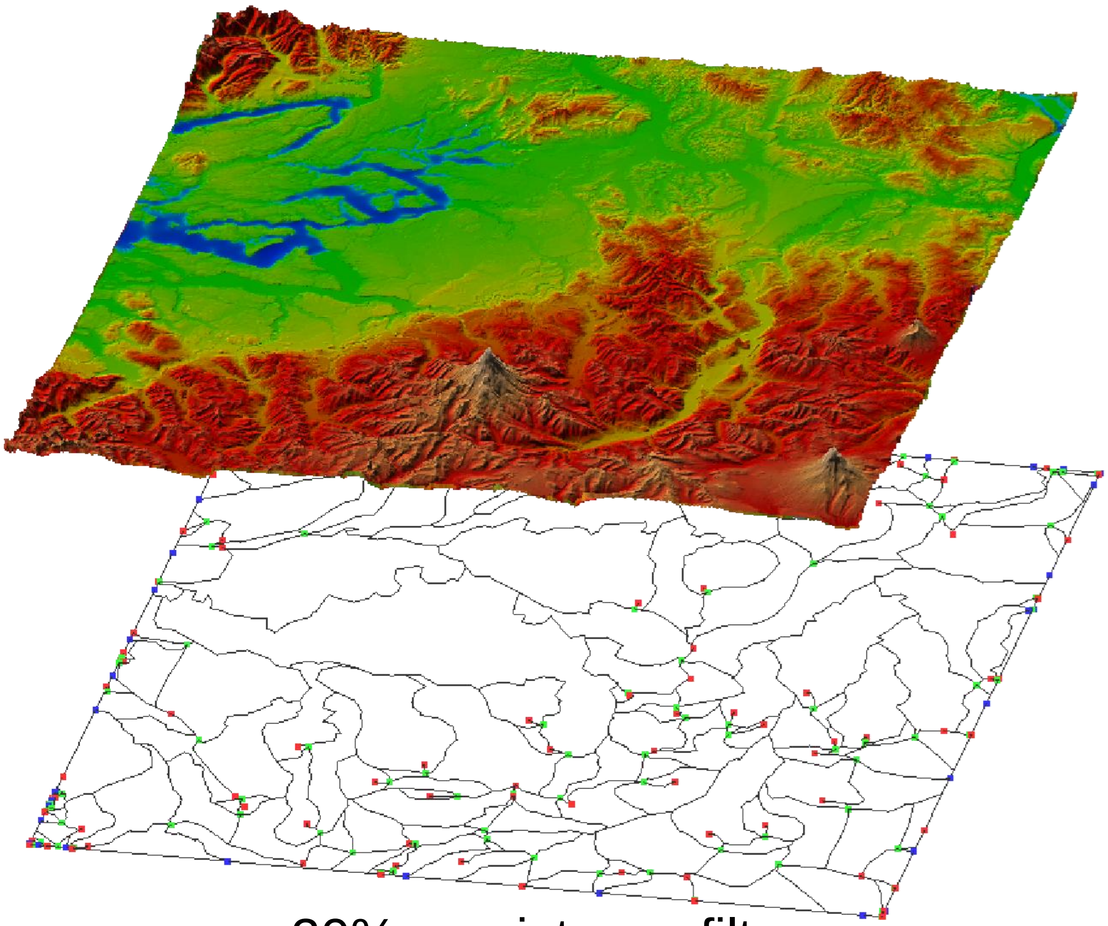
0.5% persistence filter

Contraction of extrema with saddles allows to simplify the MS-complex.



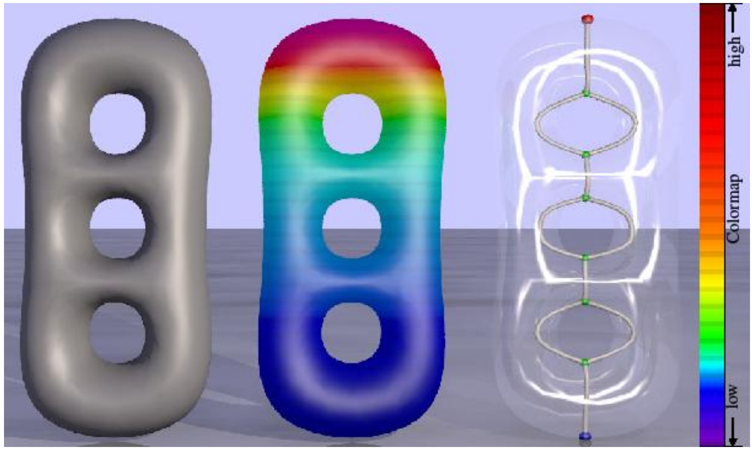
1.0% persistence filter

Contraction of extrema with saddles allows to simplify the MS-complex.

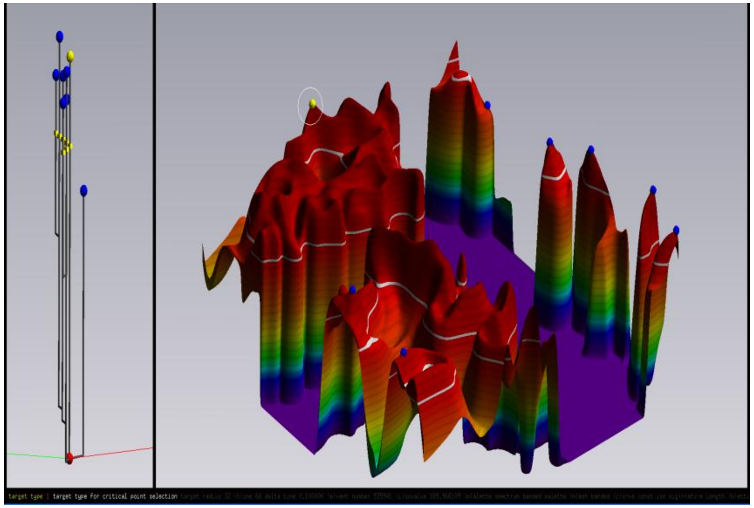


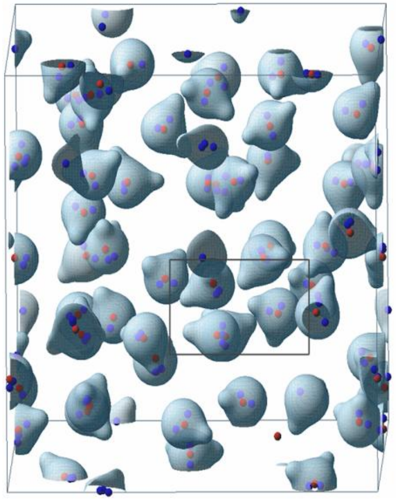
20% persistence filter

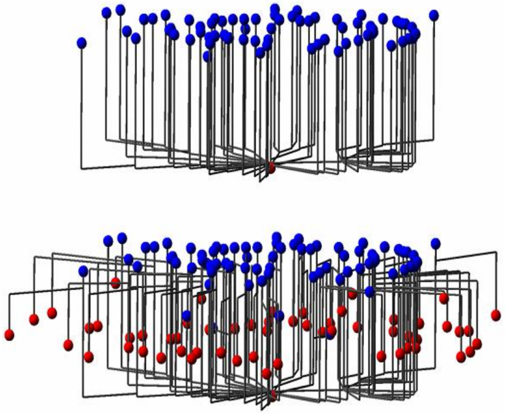
Topological Constructs Allow Building Effective and Succinct Shape Descriptors



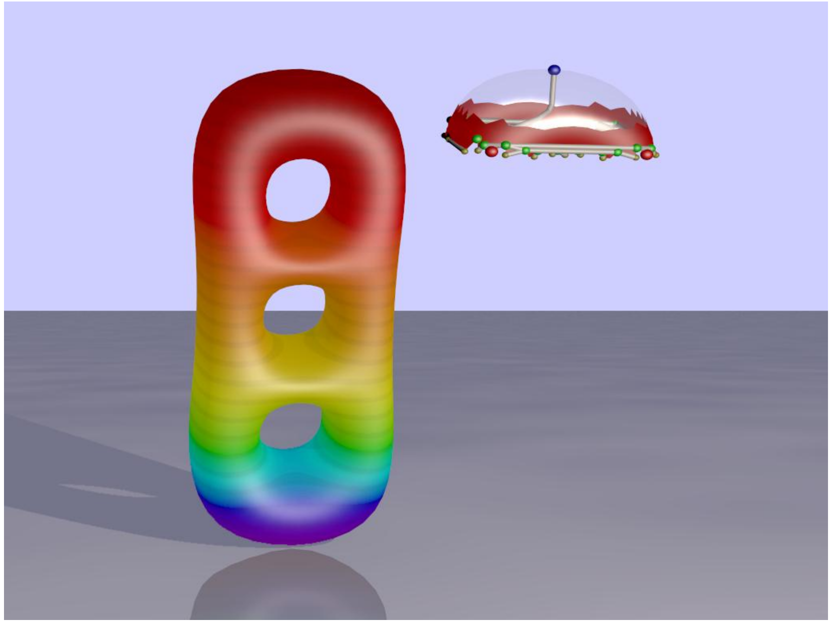
The Reeb graph is the graph obtained by continuous contraction of all the contours in a scalar field, where each contour is collapsed to a distinct point.

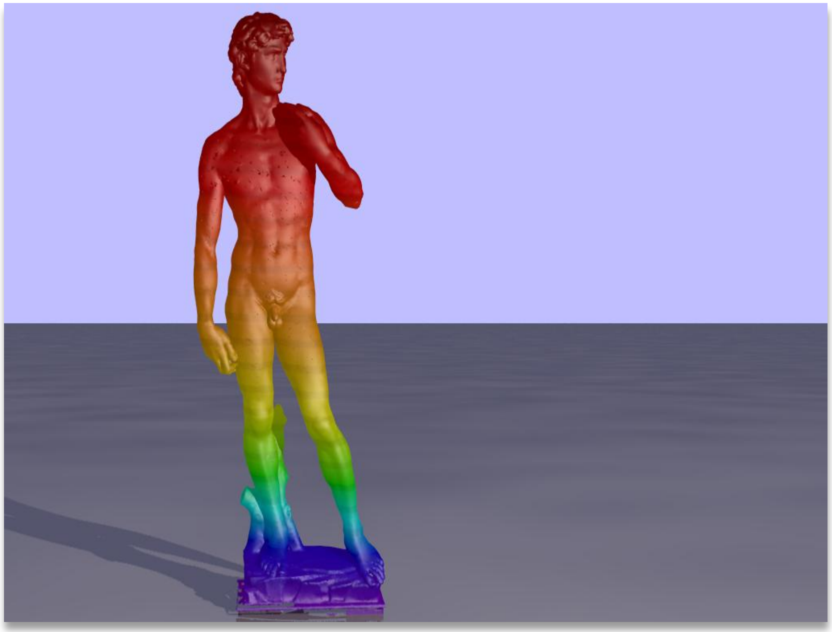






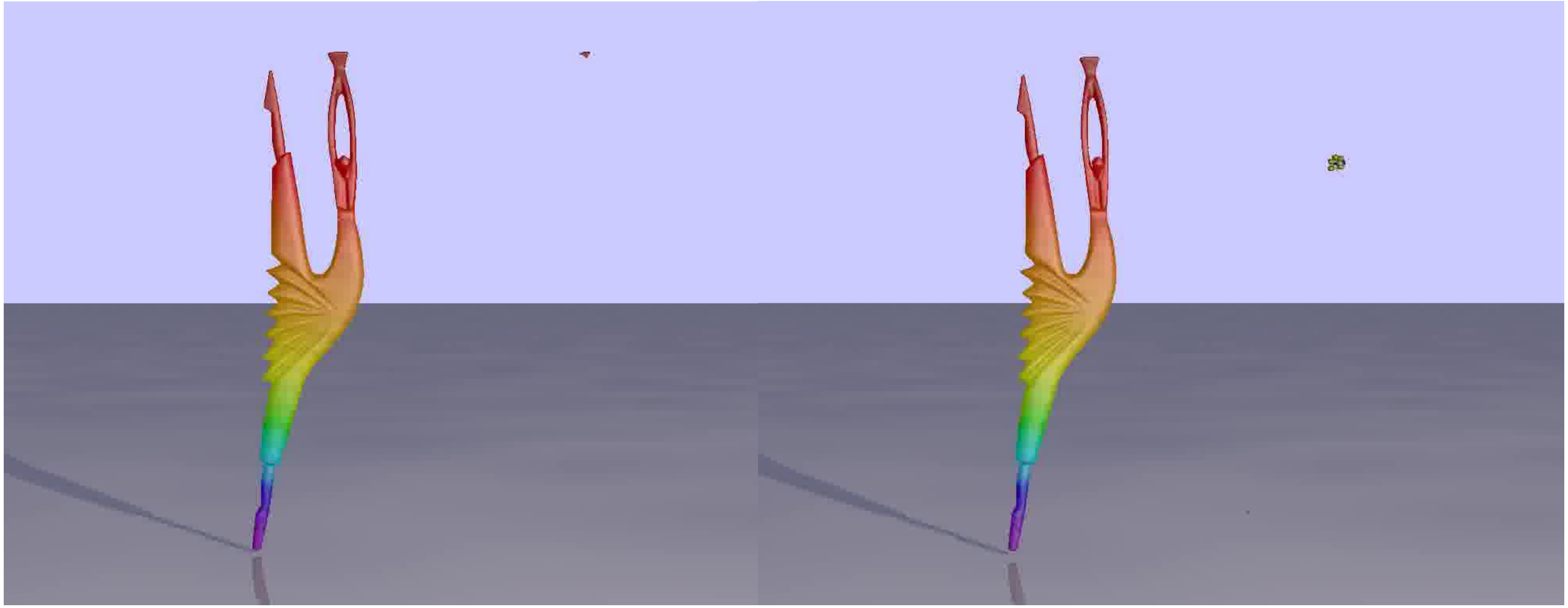
We Use Streaming Techniques to Achieve High Performance Analysis of Shapes



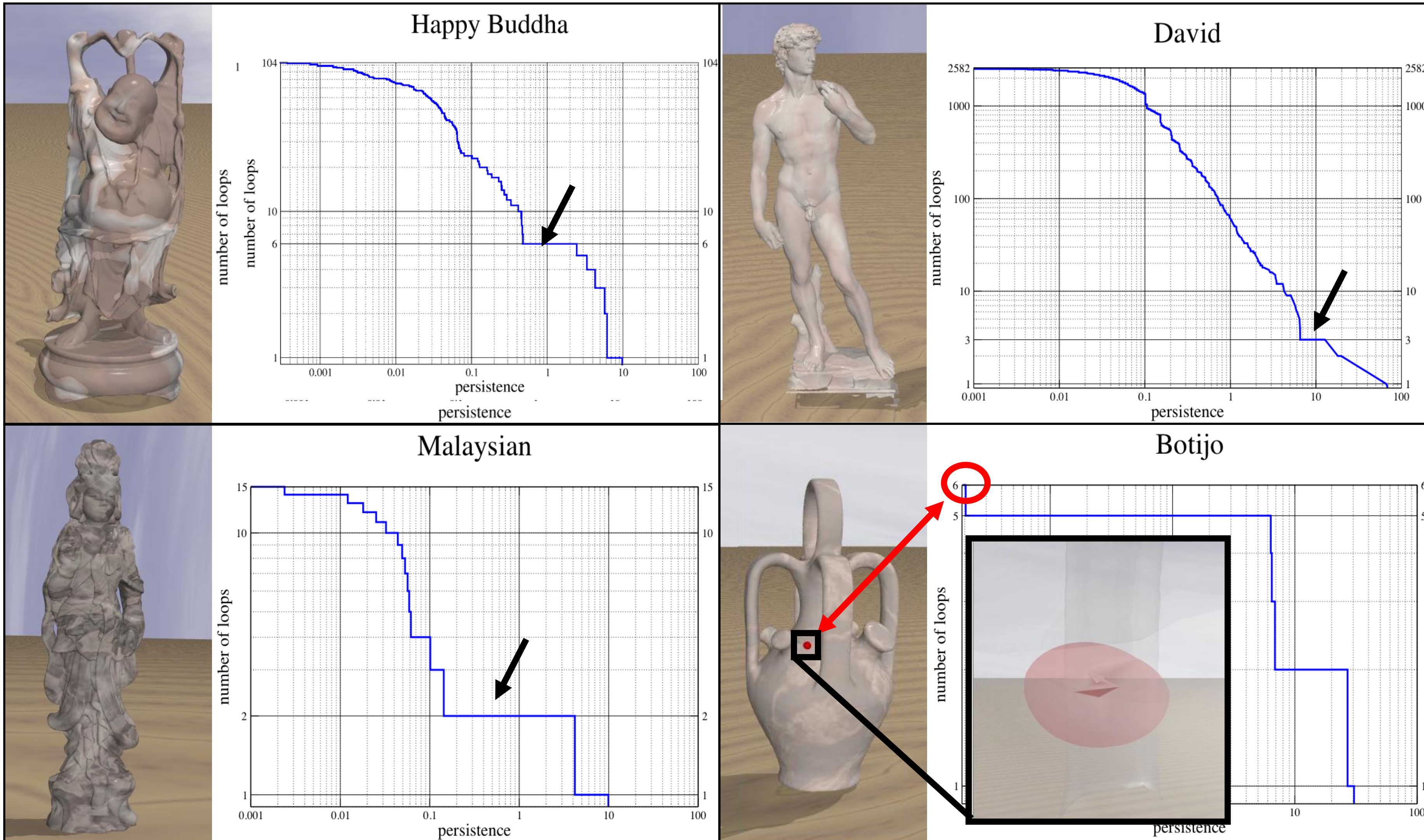


We Exploit the Existing Locality of the Input Mesh to Avoid Costly Reordering

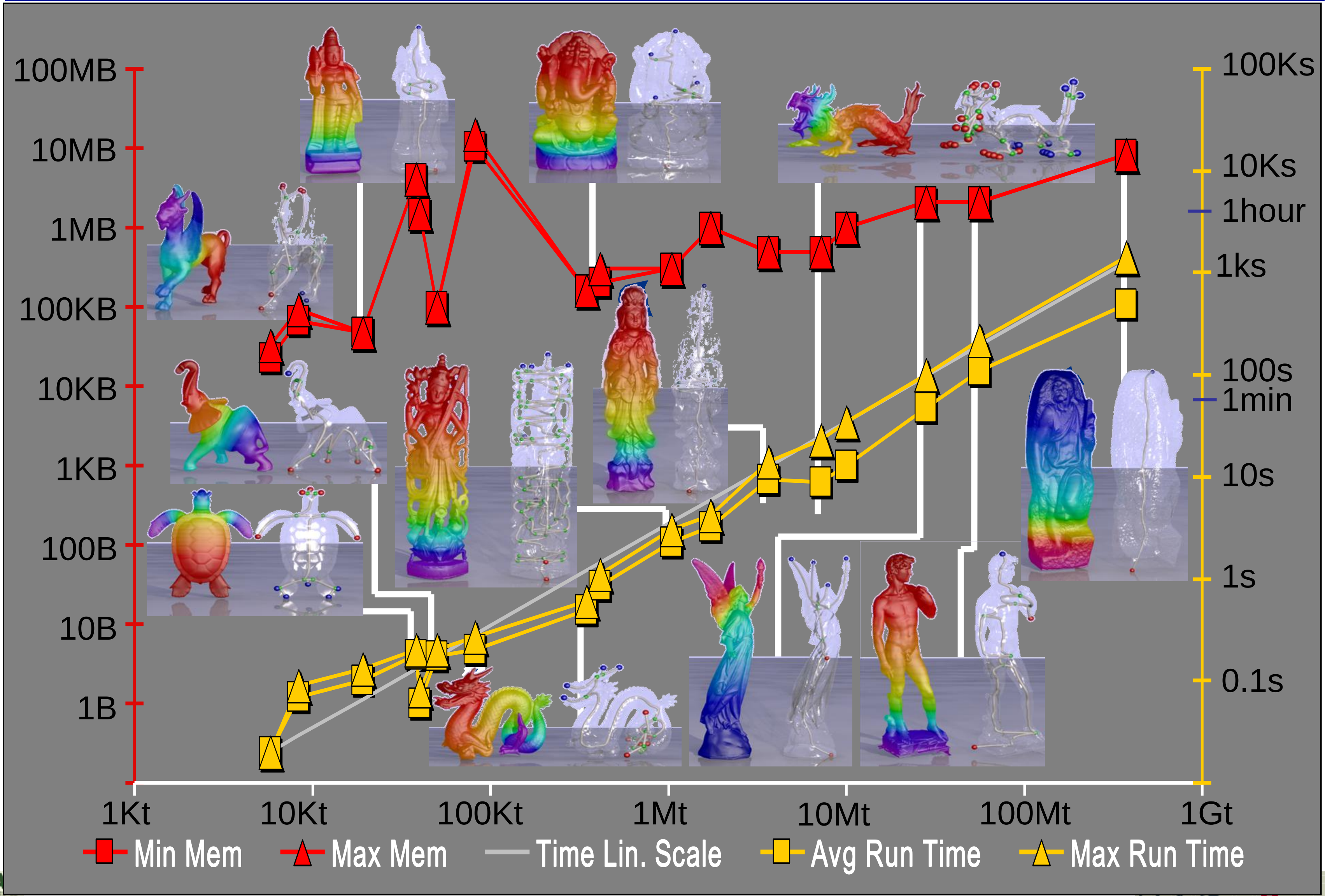
- **Computation of the Reeb graph for two layouts of the same model**



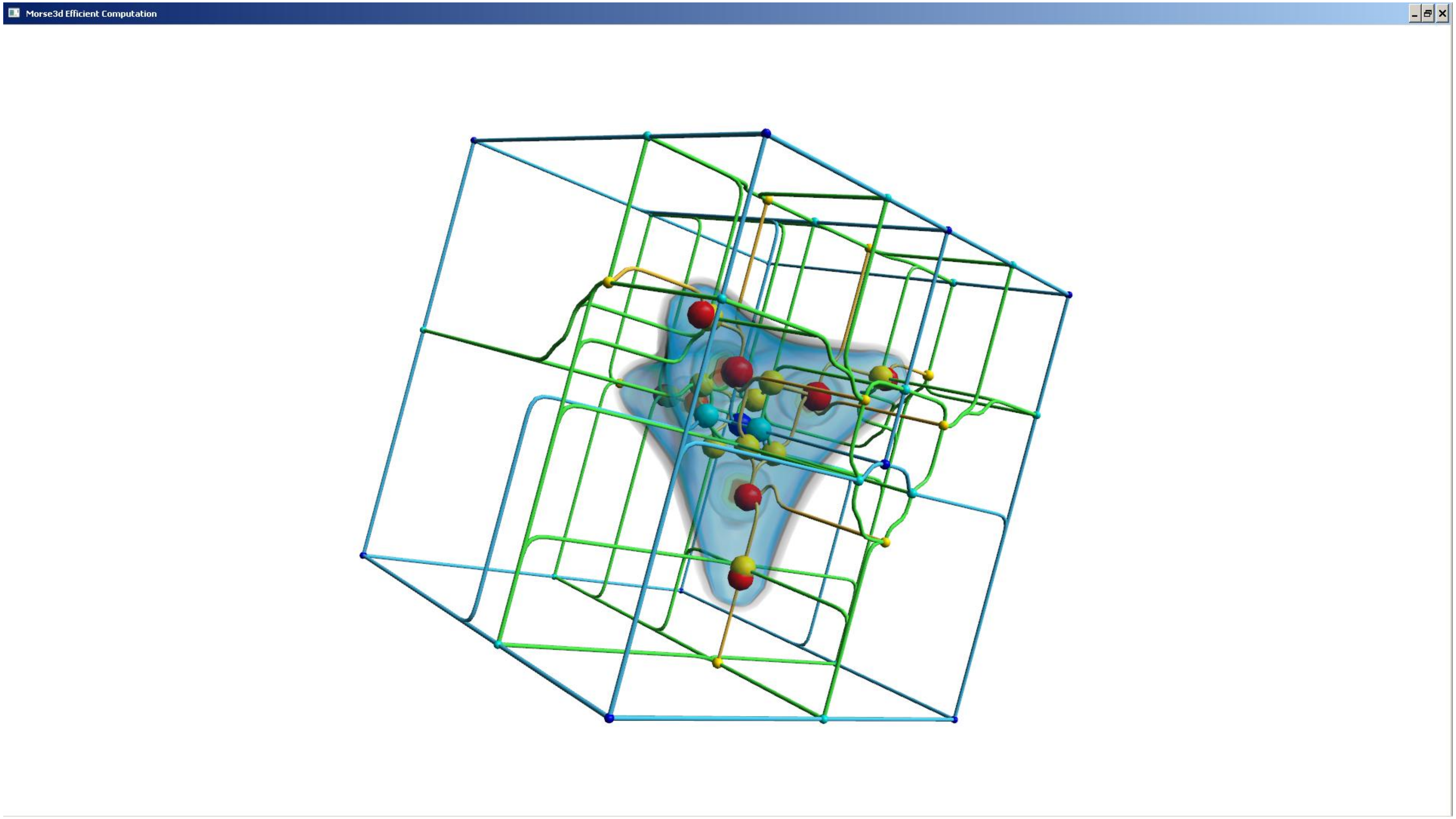
We Develop Shape Signatures to Find Defects in Large Scale Geometric Models



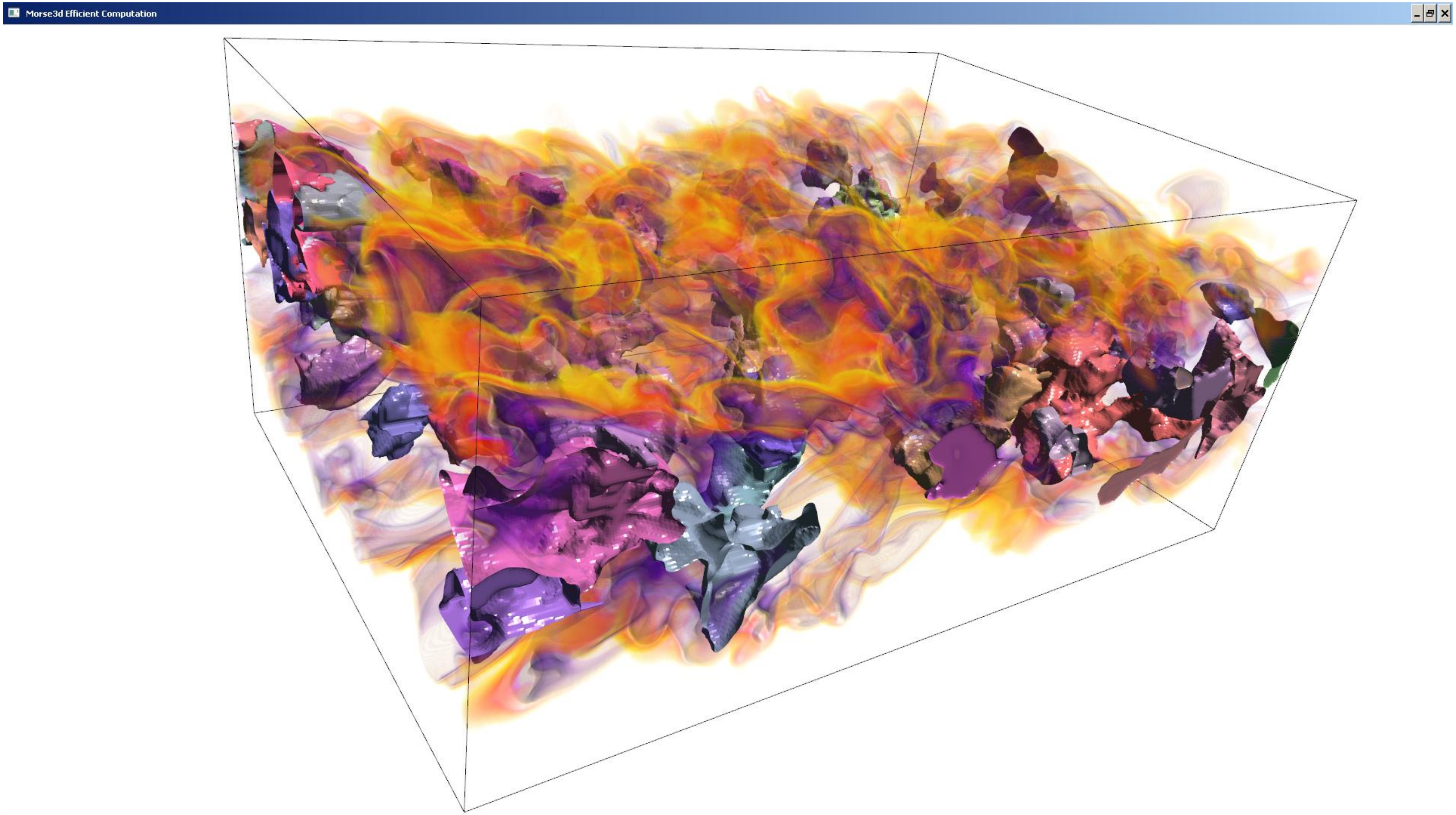
Practical Tests Confirm Robustness and Show Virtually Linear Scalability



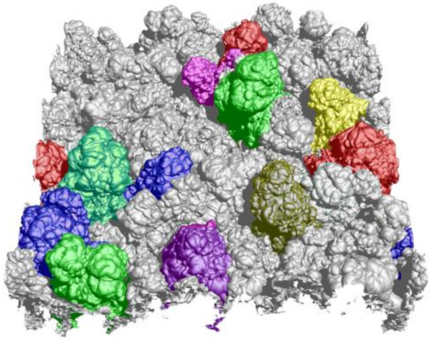
Demo C_4H_4

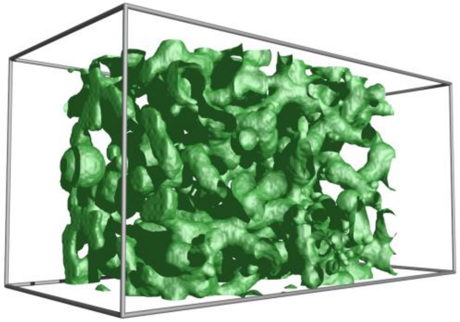


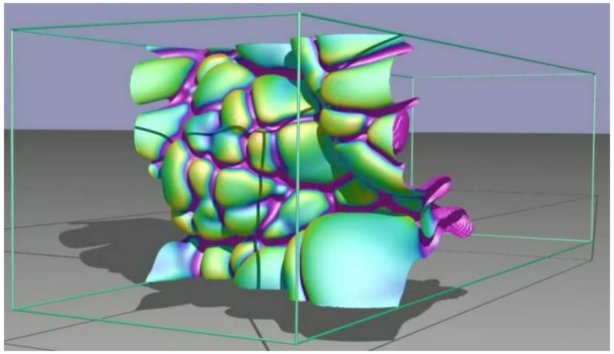
Demo S3D Combustion Simulation



Data Analysis Examples

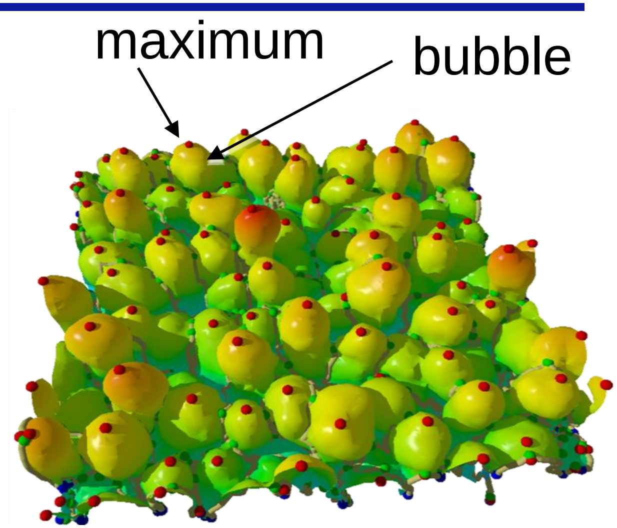






We Analyze High-Resolution Rayleigh–Taylor Instability Simulations

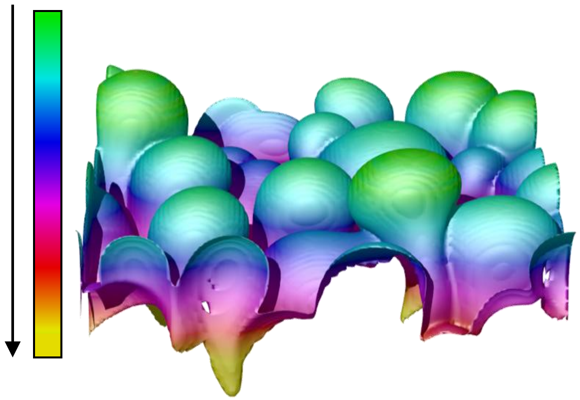
- **Large eddy simulation run on Linux cluster: 1152 x 1152 x 1152**
  - **~ 40 G / dump**
  - **759 dumps, about 25 TB**
- **Direct numerical simulation run on BlueGene/L: 3072 x 3072 x Z**
  - **Z depends on width of mixing layer**
  - **More than 40 TB**

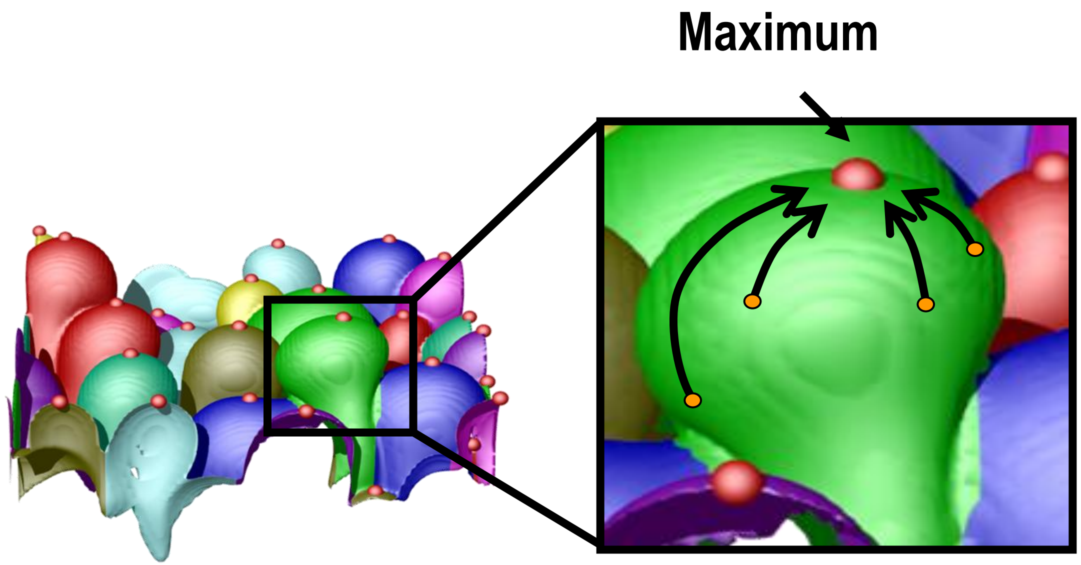


- **Bubble-like structures are observed in laboratory and simulations**
- **Bubble dynamics are considered an important way to characterize the mixing process**
  - **Mixing rate =** $\partial(\#bubbles)/\partial t.$
- **There is no prevalent formal definition of bubbles**

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

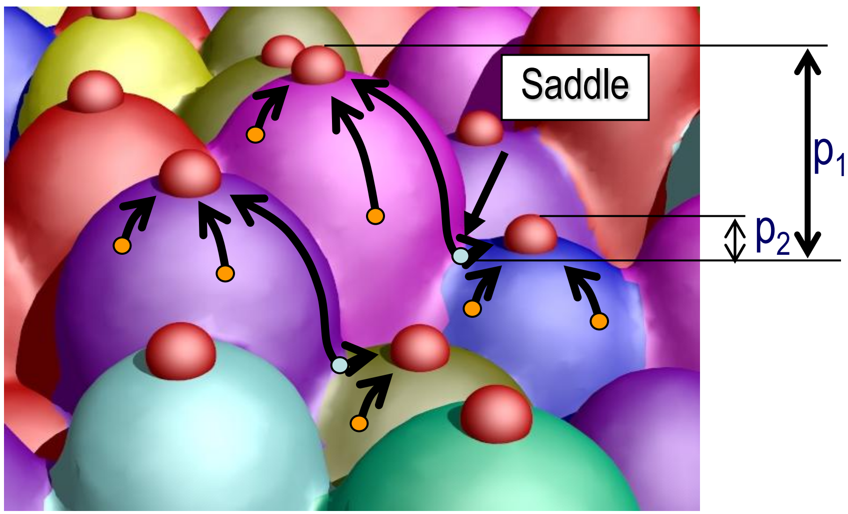
Morse complex cells drawn in distinct colors

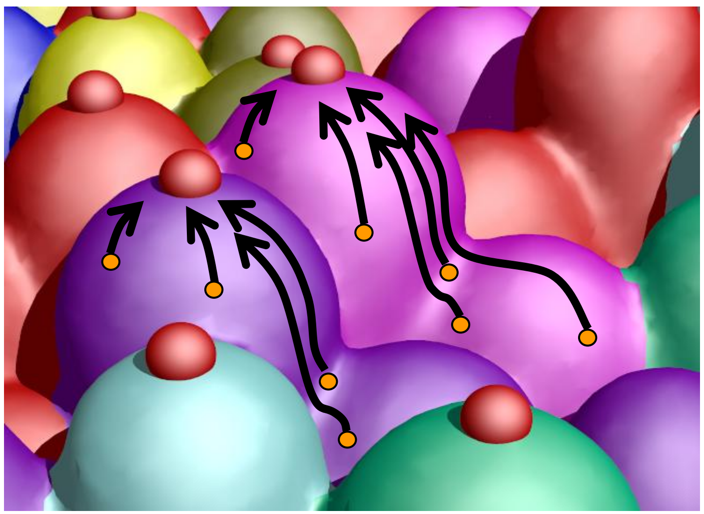
In each Morse complex cell, all steepest ascending lines converge to one maximum

A Hierarchal Model is Generated by Simplification of Critical Points

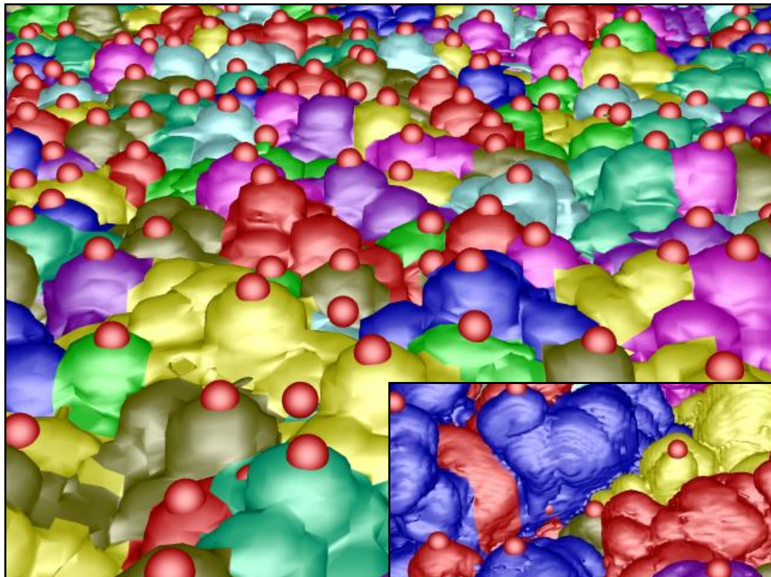
- **Persistence is varied to annihilate pairs of critical points and produce coarser segmentations**

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

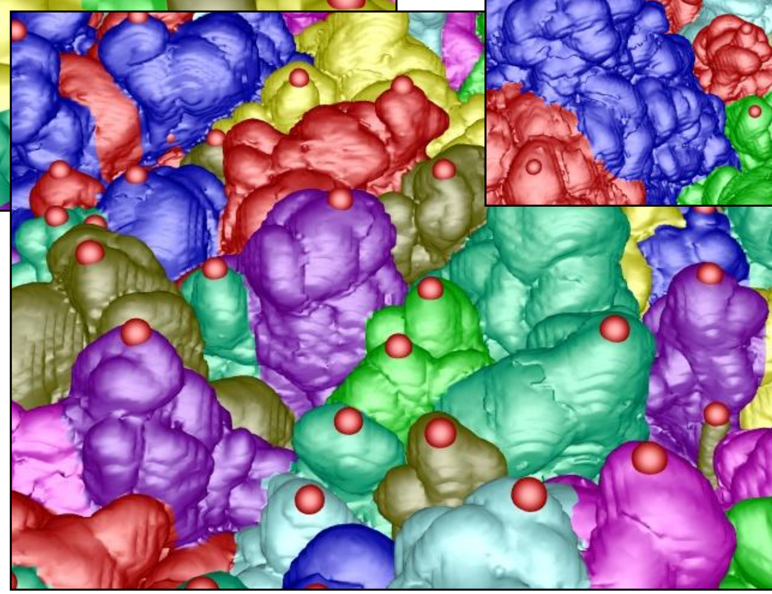




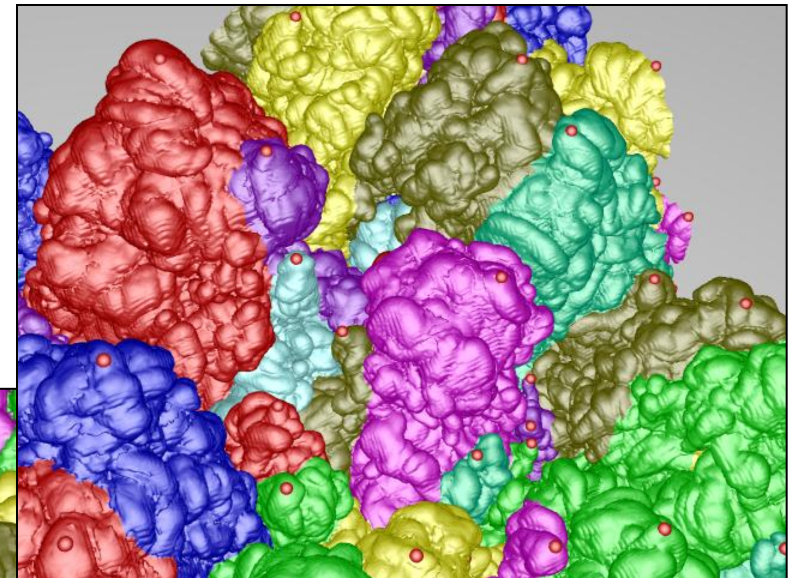
The Segmentation Method is Robust From Early Mixing to Late Turbulence



T=100

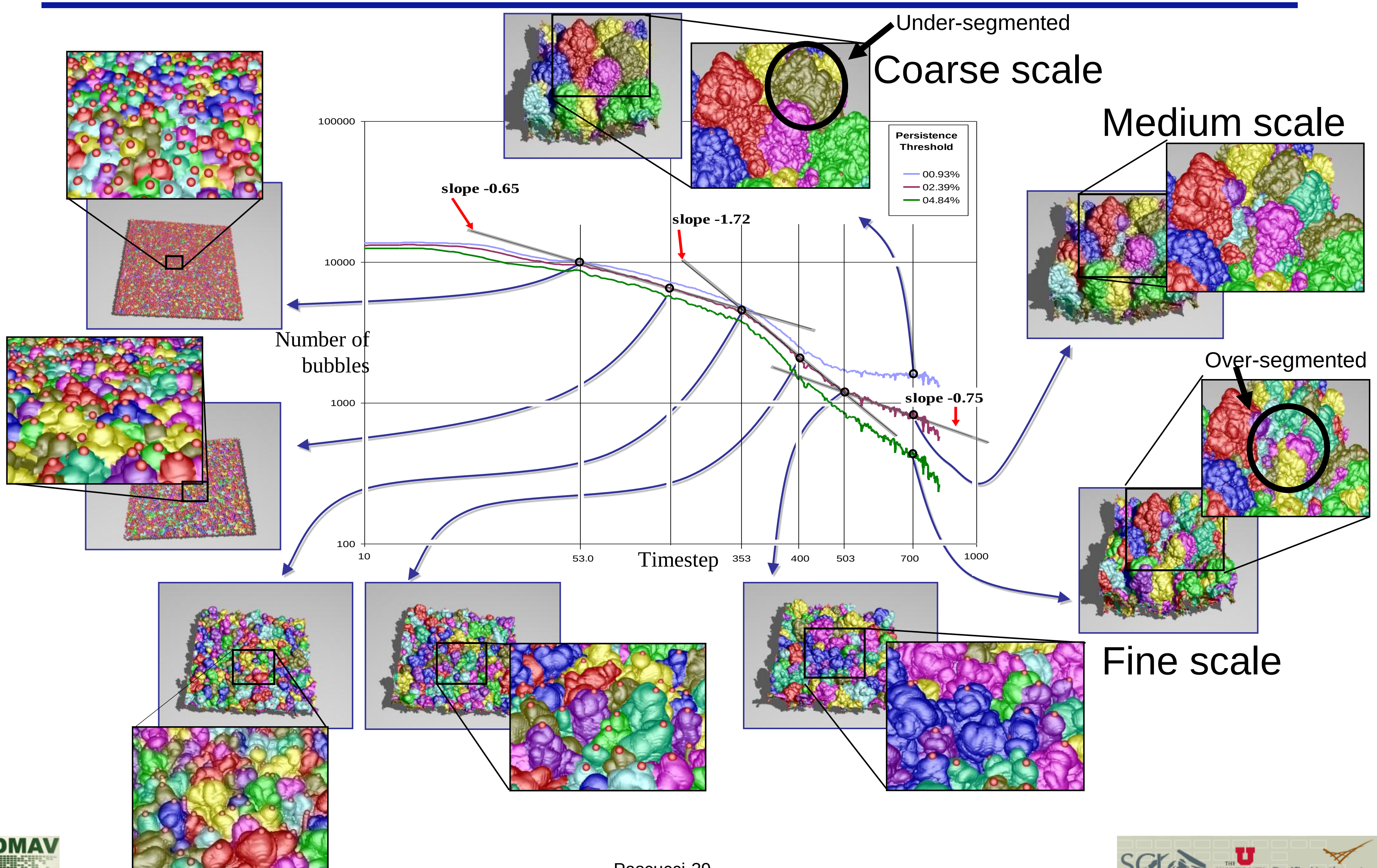


T=353

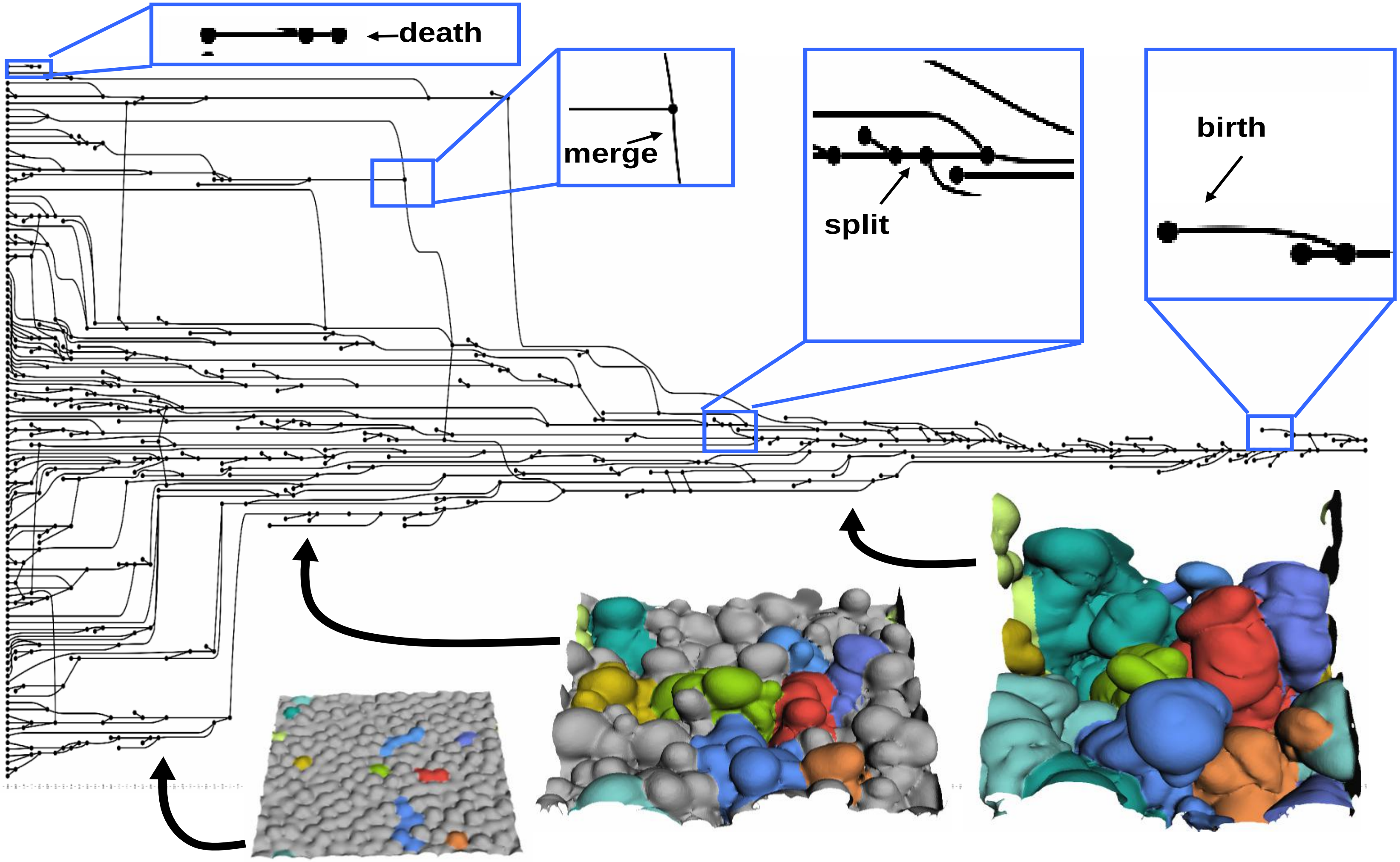


T=700

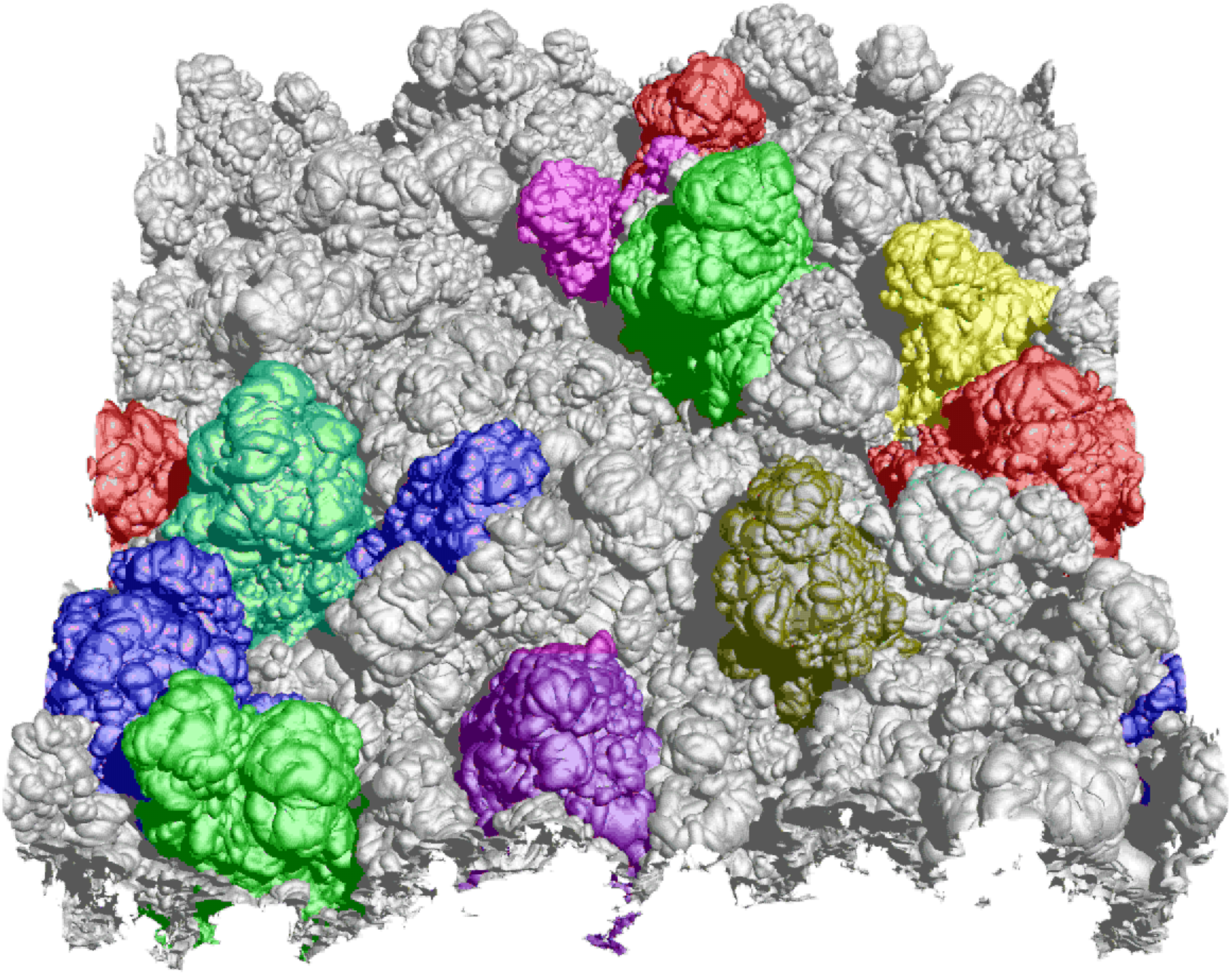
We Evaluated Our Quantitative Analysis at Multiple Scales



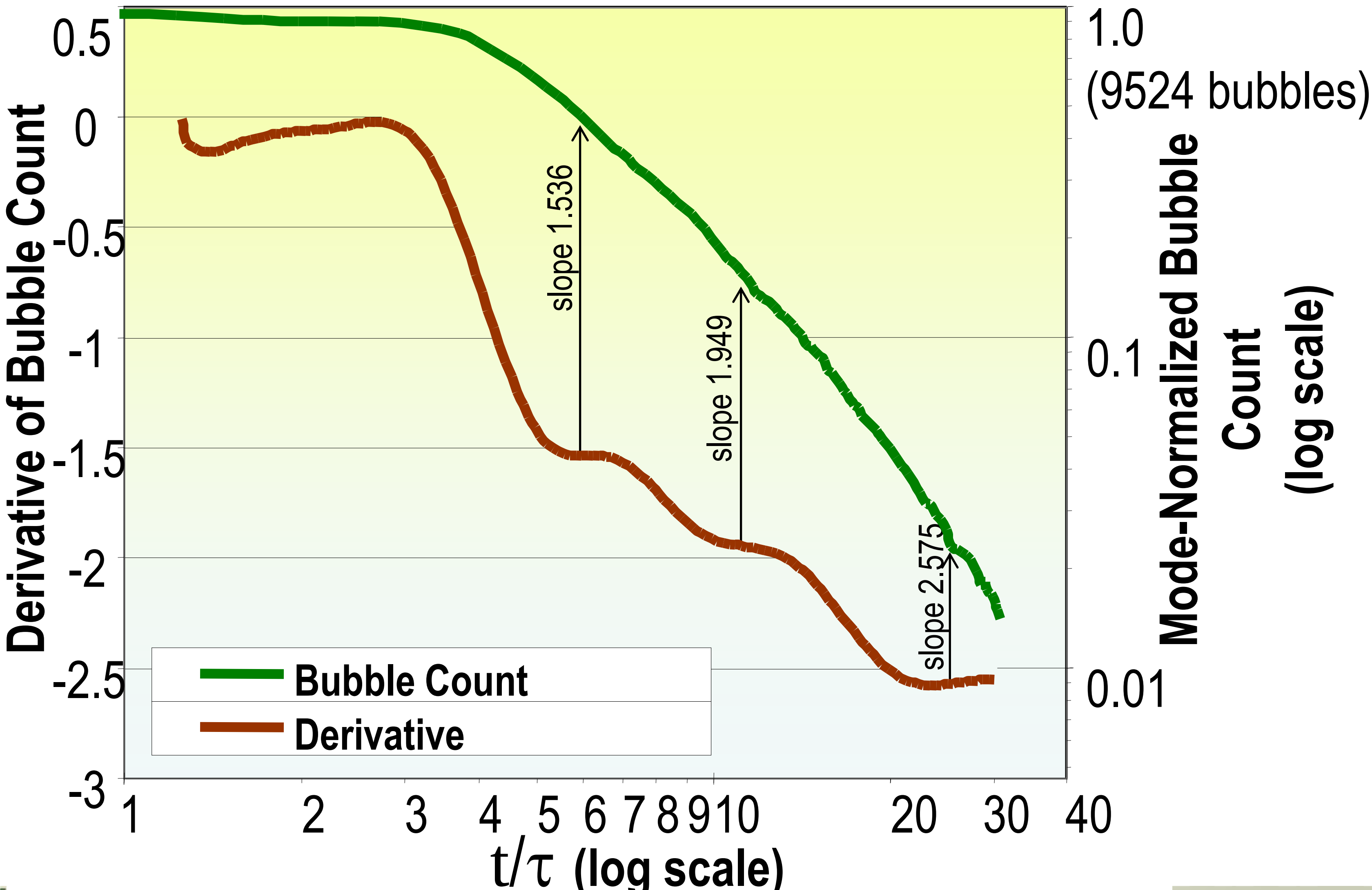
We Characterize Events that Occur in the Mixing Process



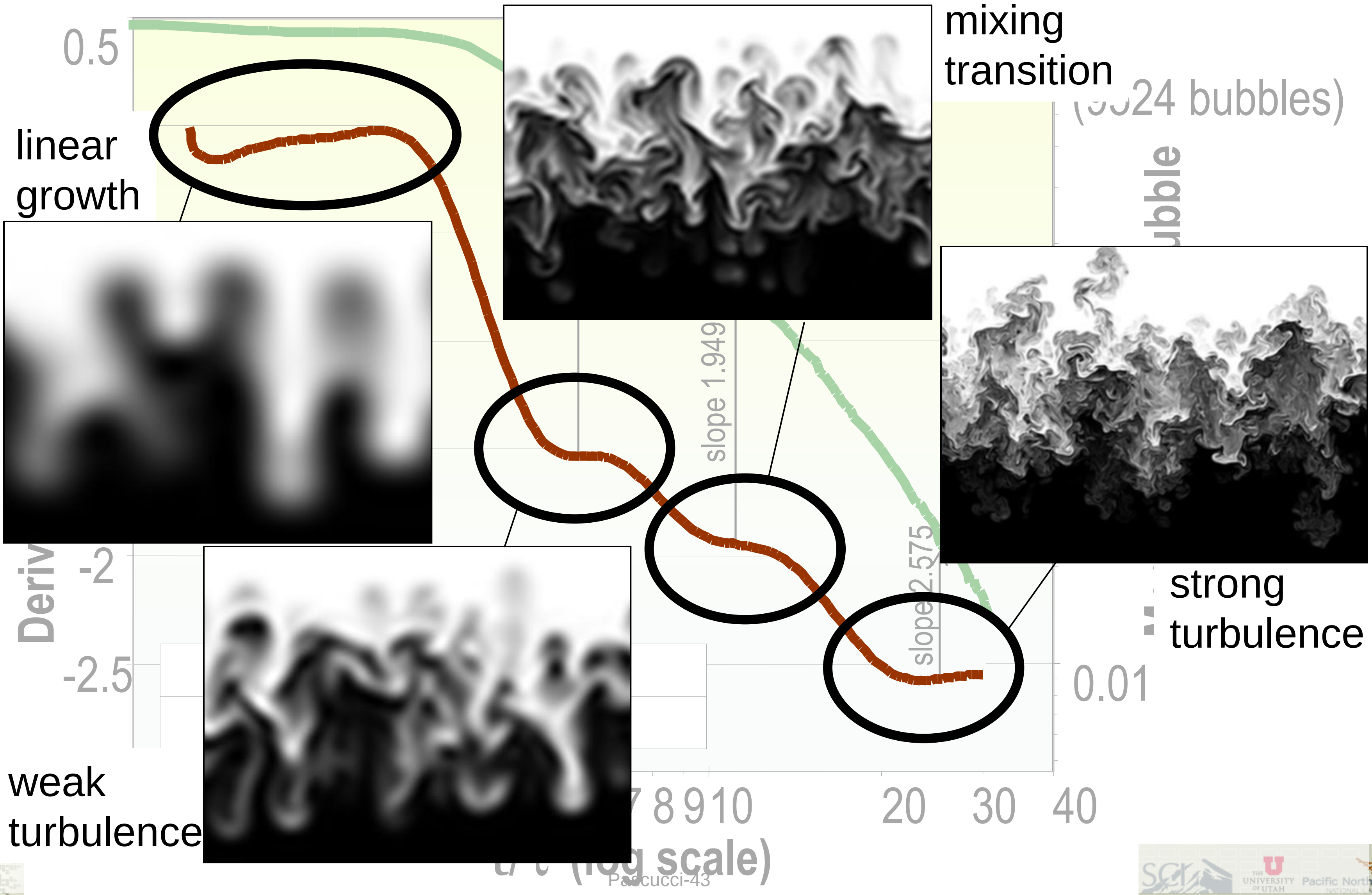
First Robust Bubble Tracking From Beginning to Late Turbulent Stages



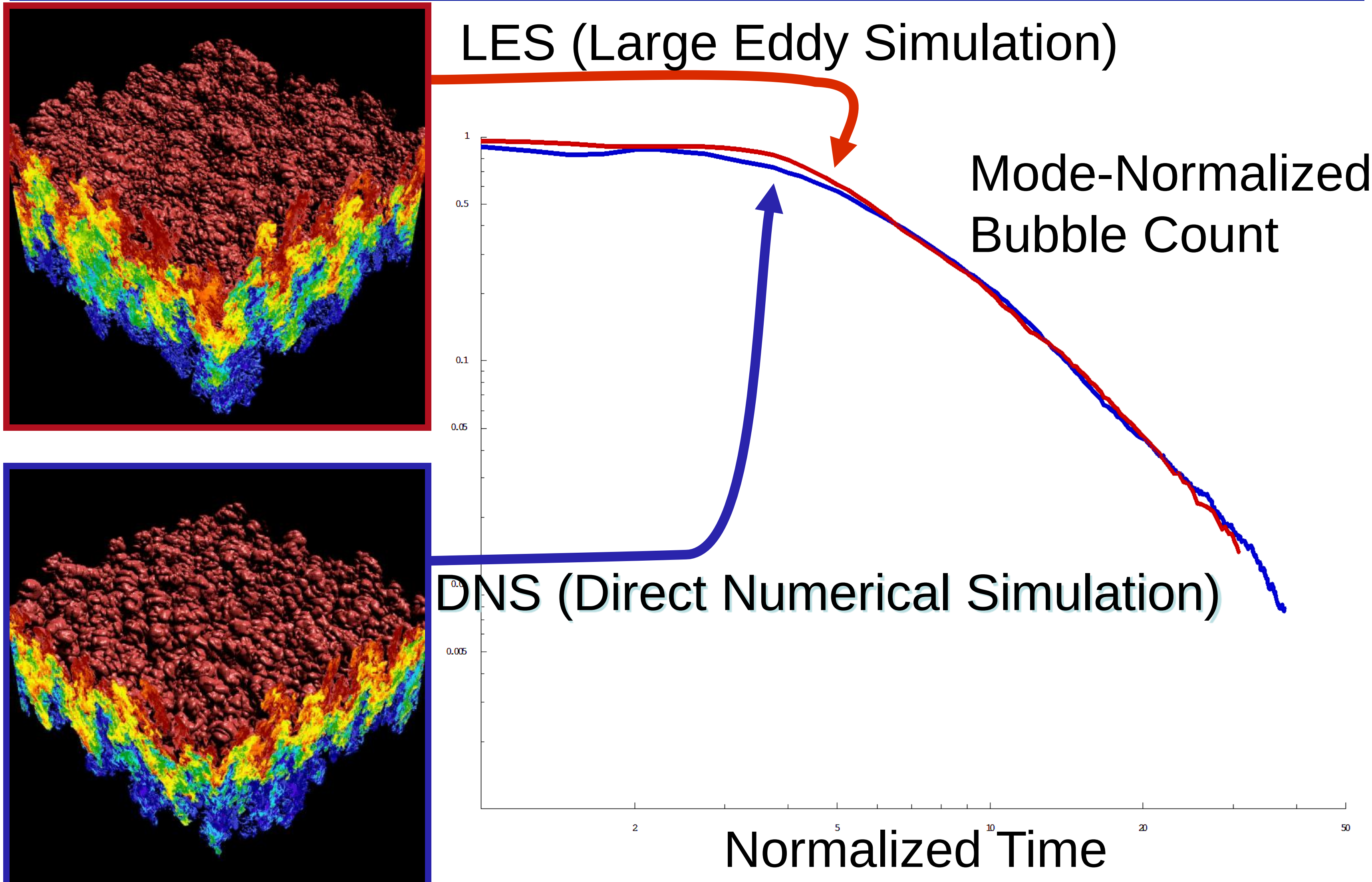
First Time Scientists Can Quantify Robustly Mixing Rates by Bubble Count



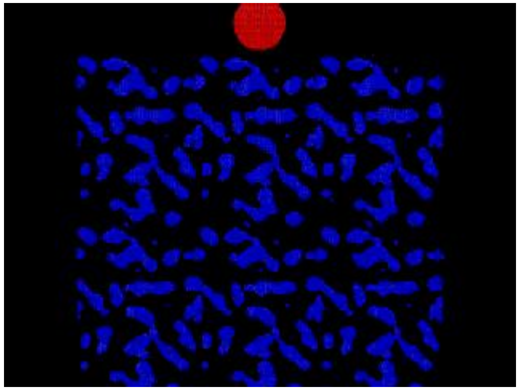
We Provide the First Quantification of Known Stages of the Mixing Process



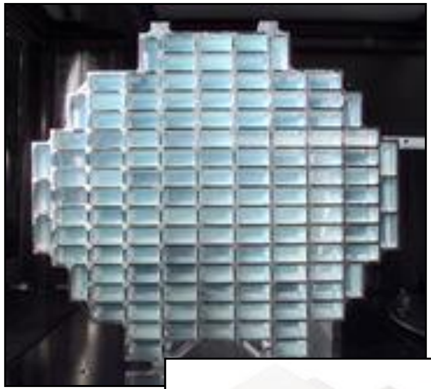
We Provided the First Feature-Based Validation of a LES with Respect to a DNS

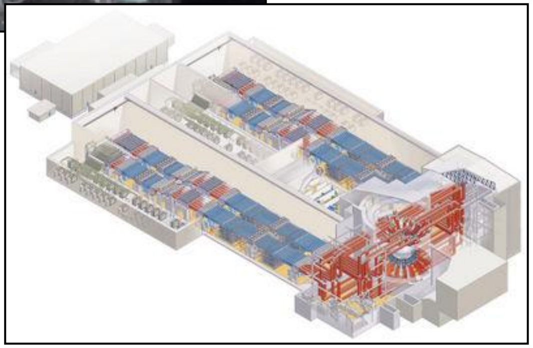


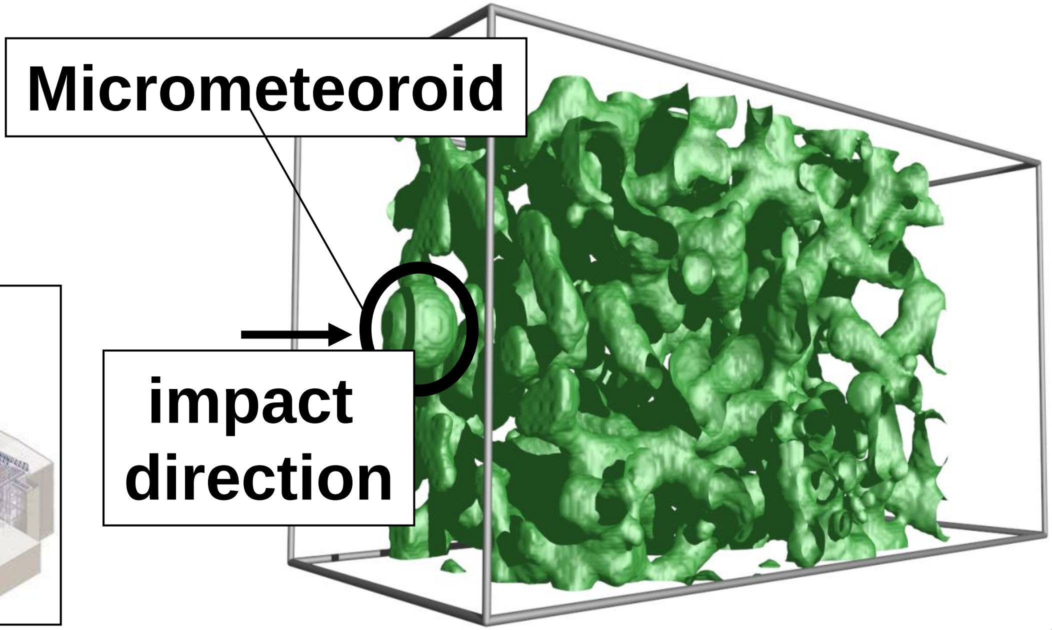
Quantitative Analysis of the Impact of a Micrometeoroid in a Porous Medium



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

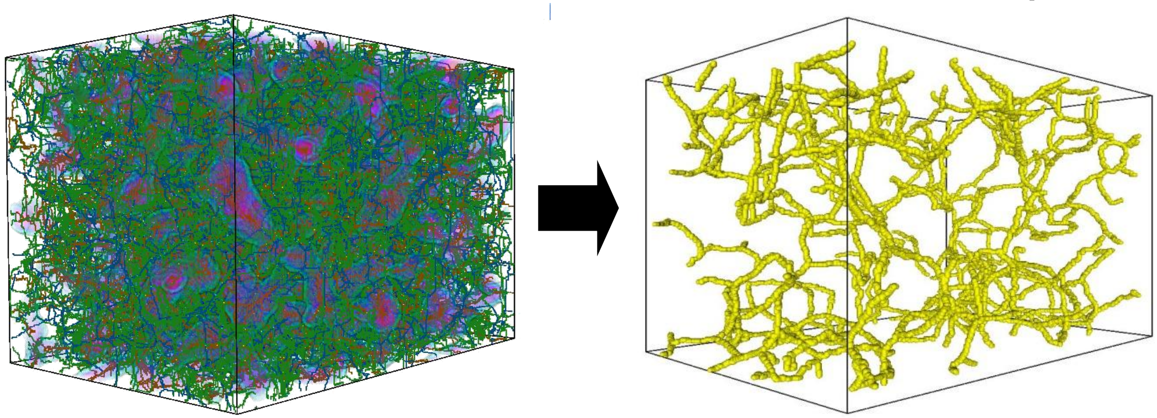




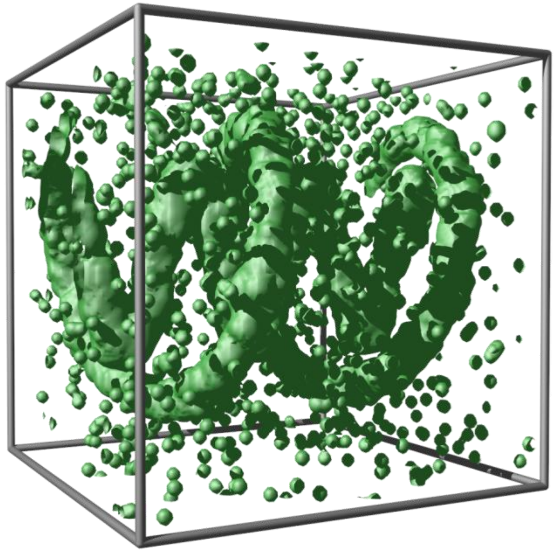


The Topological Reconstruction Method is Validated with a Controlled Test Shape

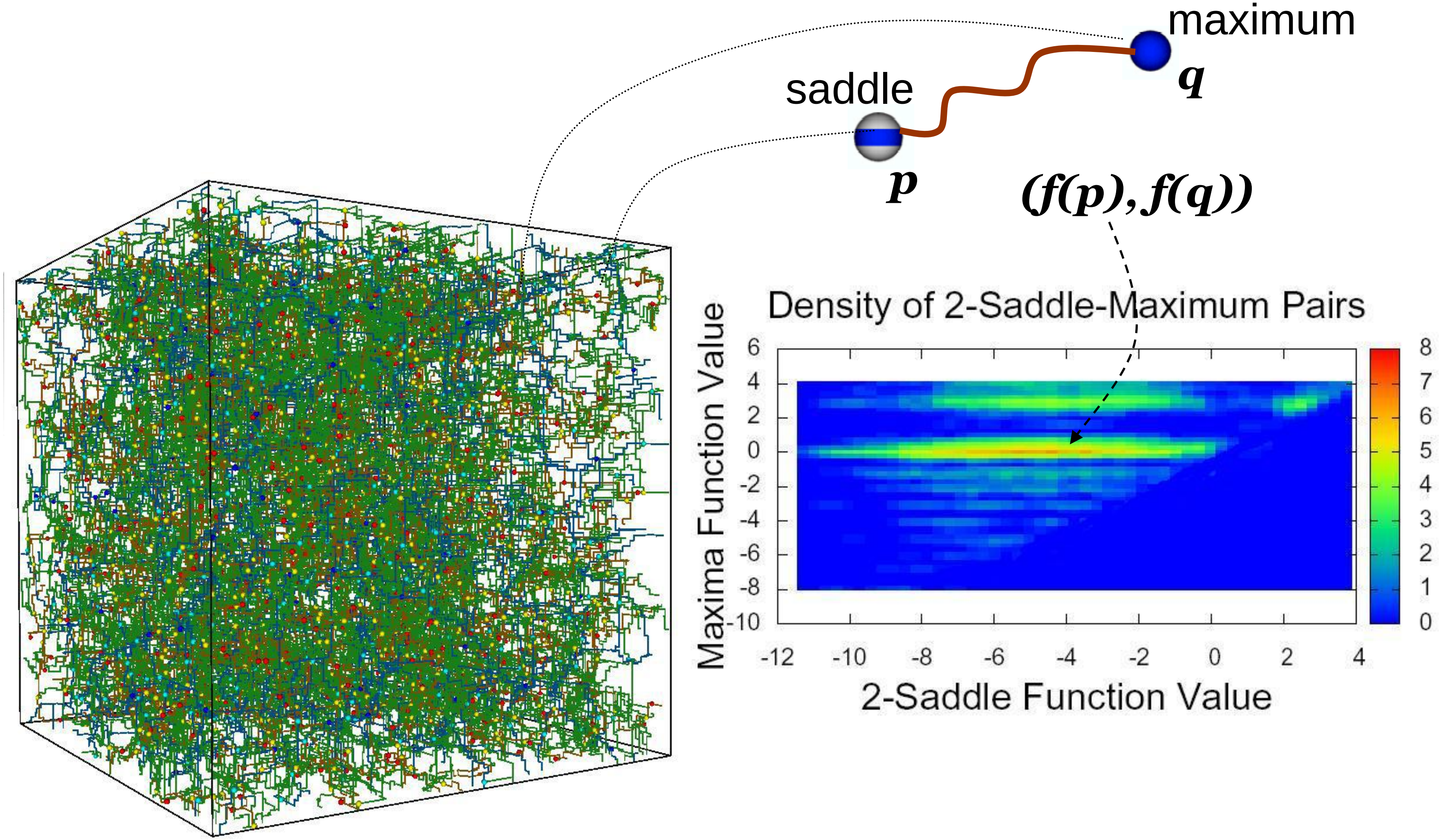
Challenge: robust reconstruction of the structure of a porous medium



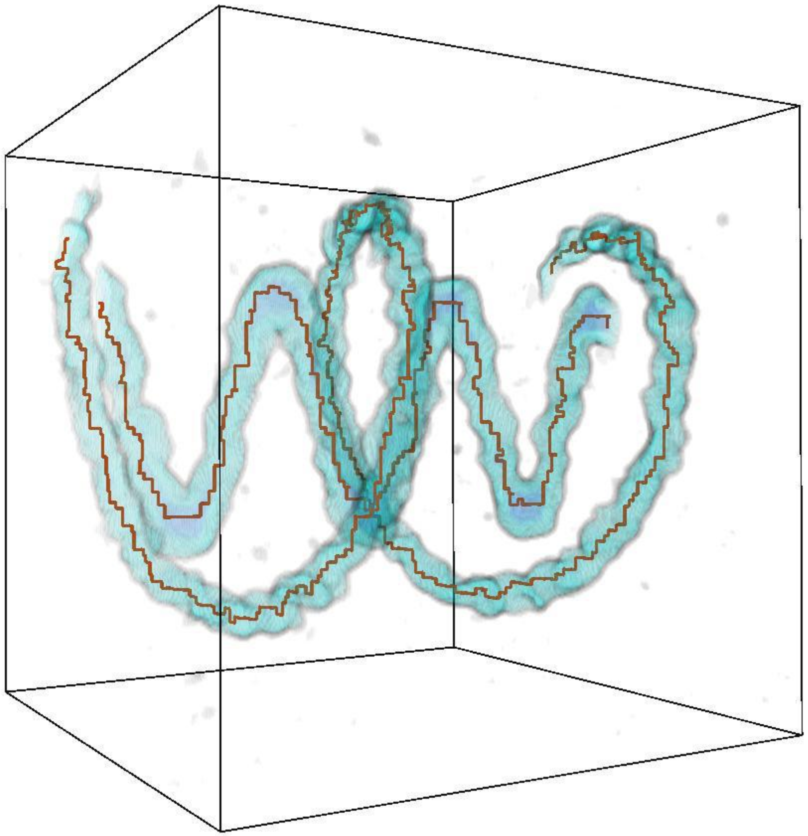
Preparation: we develop control test data to validate the approach

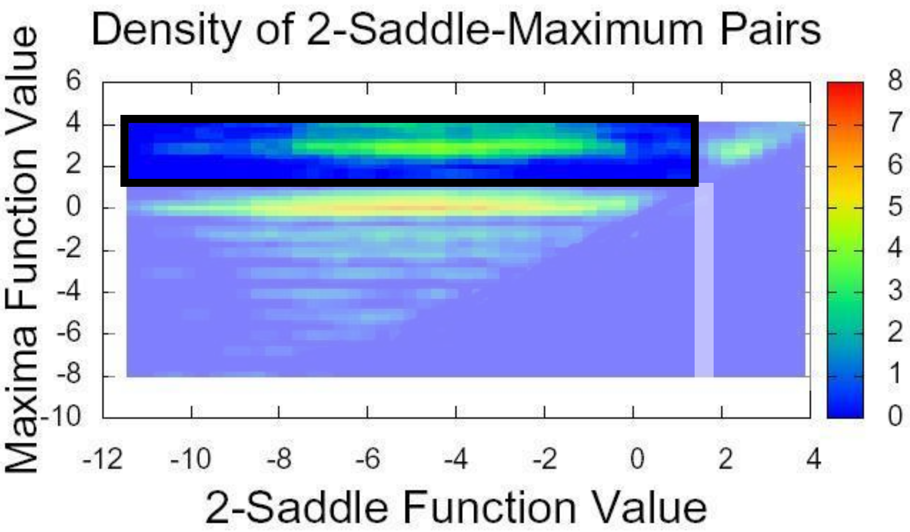


We Report the Distribution of Topological Features in the Full Resolution Data

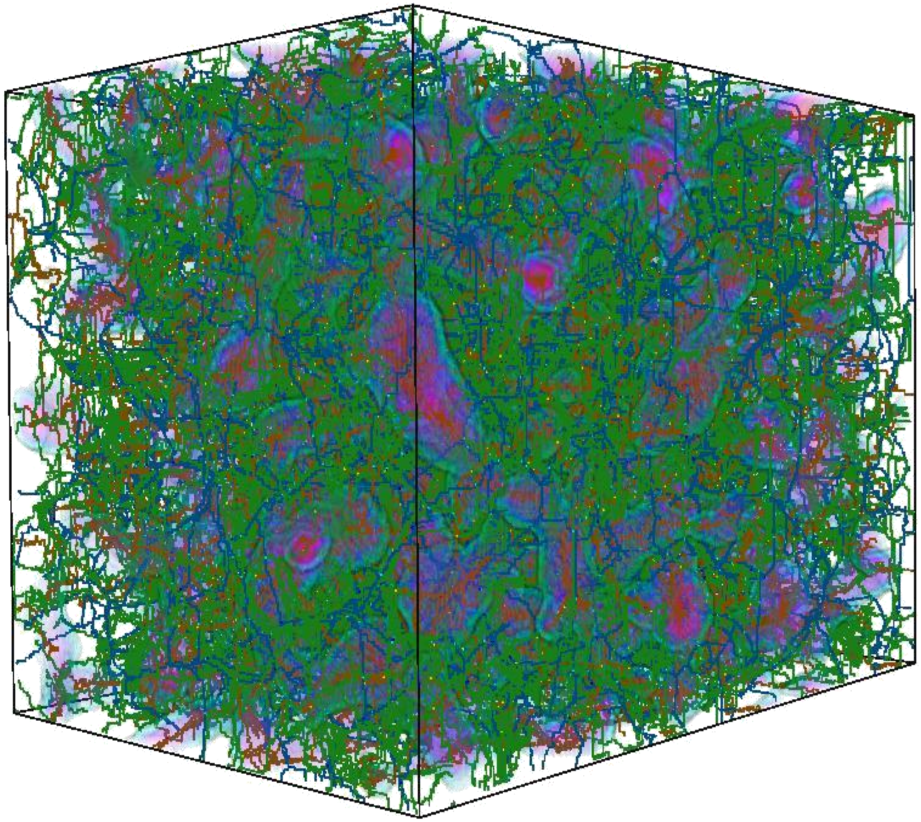


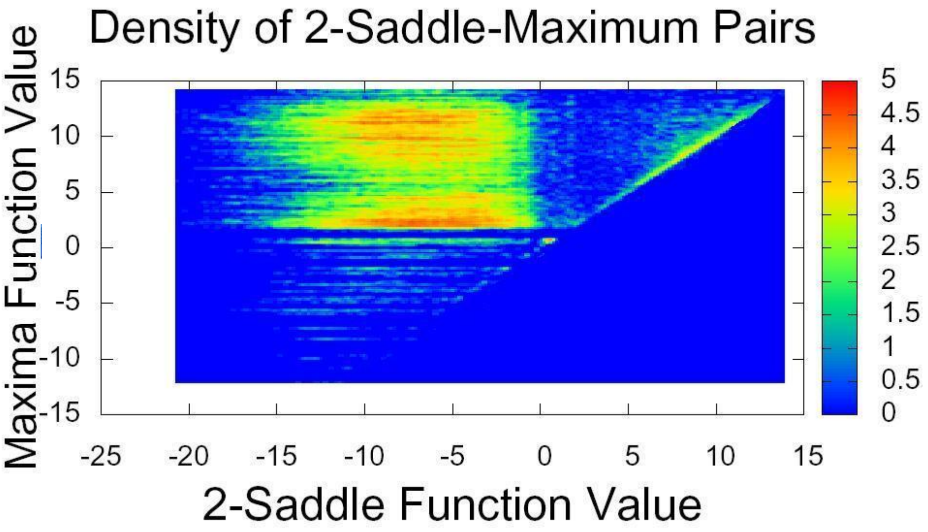
The Hierarchical Morse-Smale Complex Has Very Good Reconstruction Properties



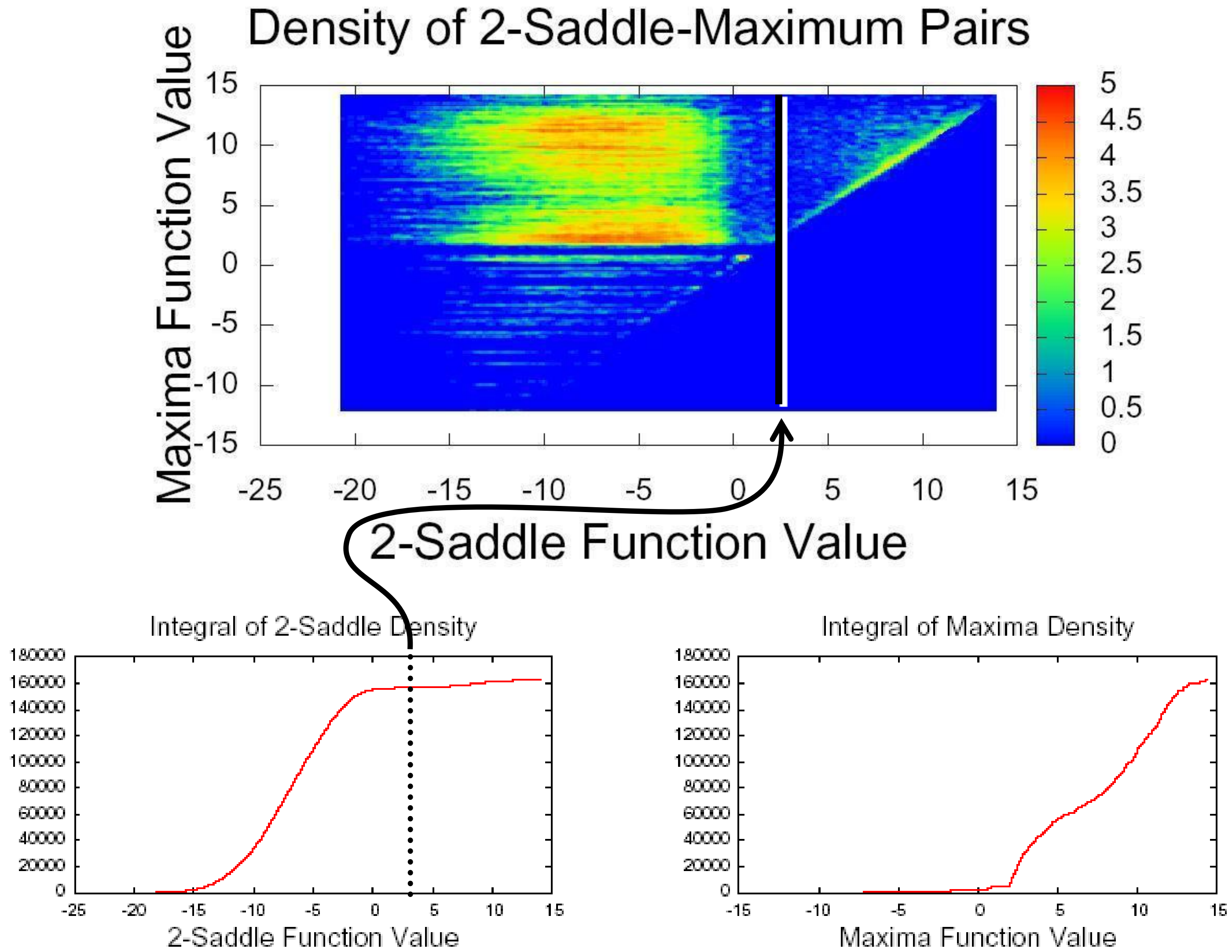


We Compute the Complete Morse-Smale Complex for the Porous Medium

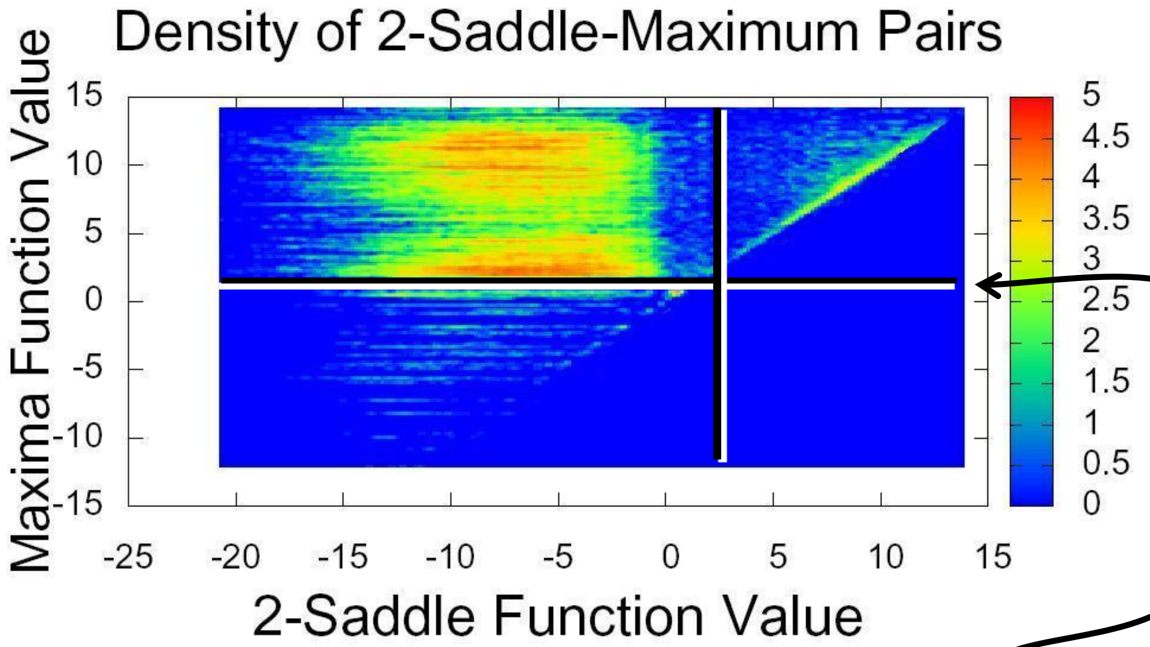


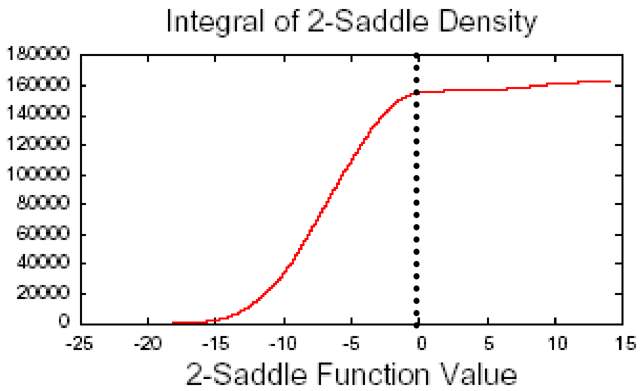


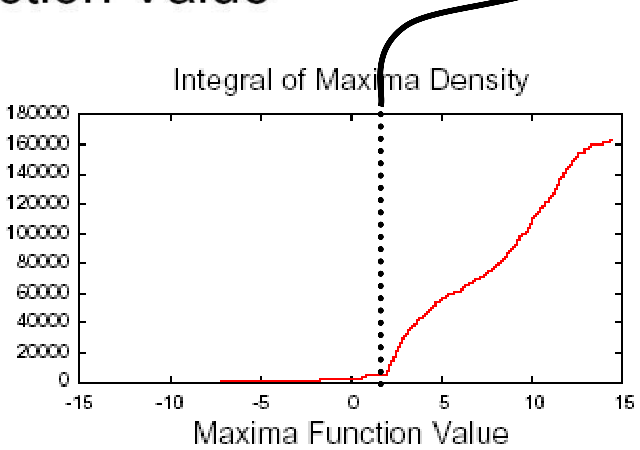
Need to Find Proper Threshold Values and Characterize the Stability of the Solution



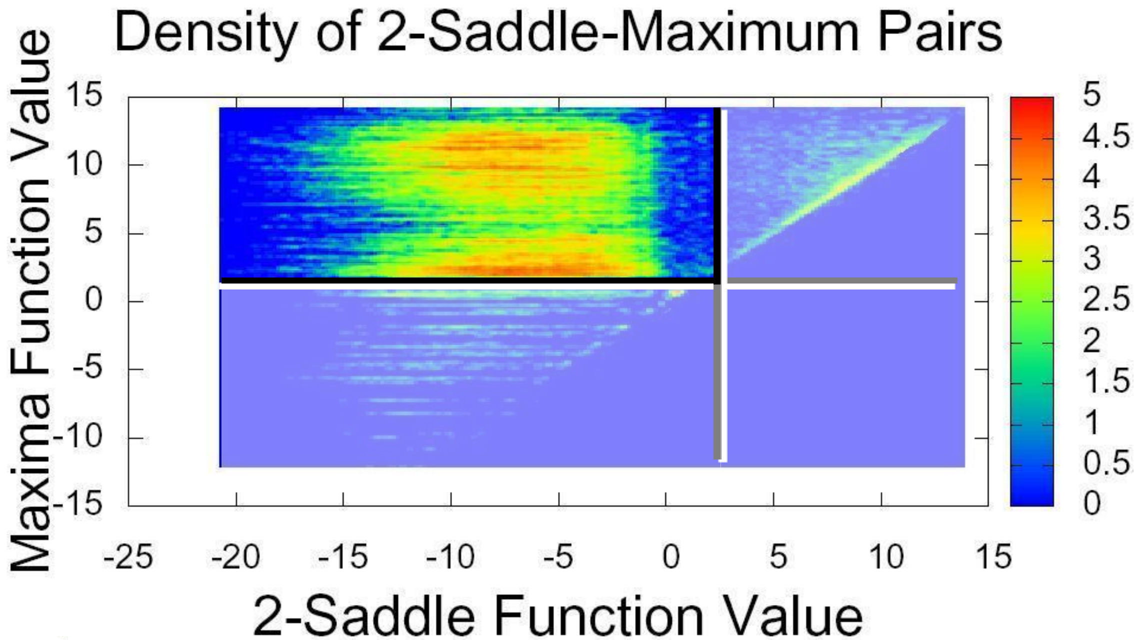
Need to Find Proper Threshold Values and Characterize the Stability of the Solution

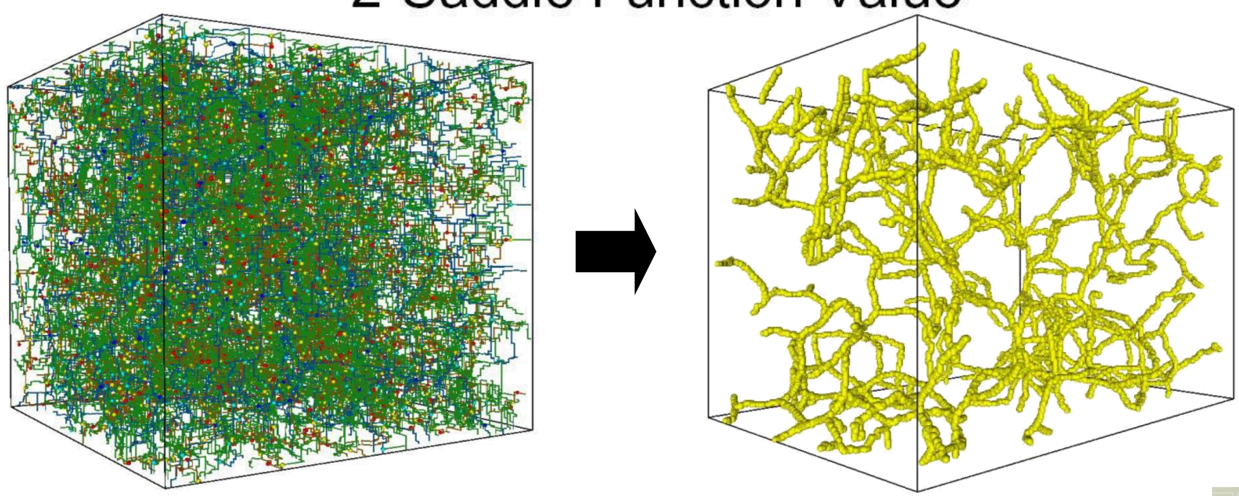




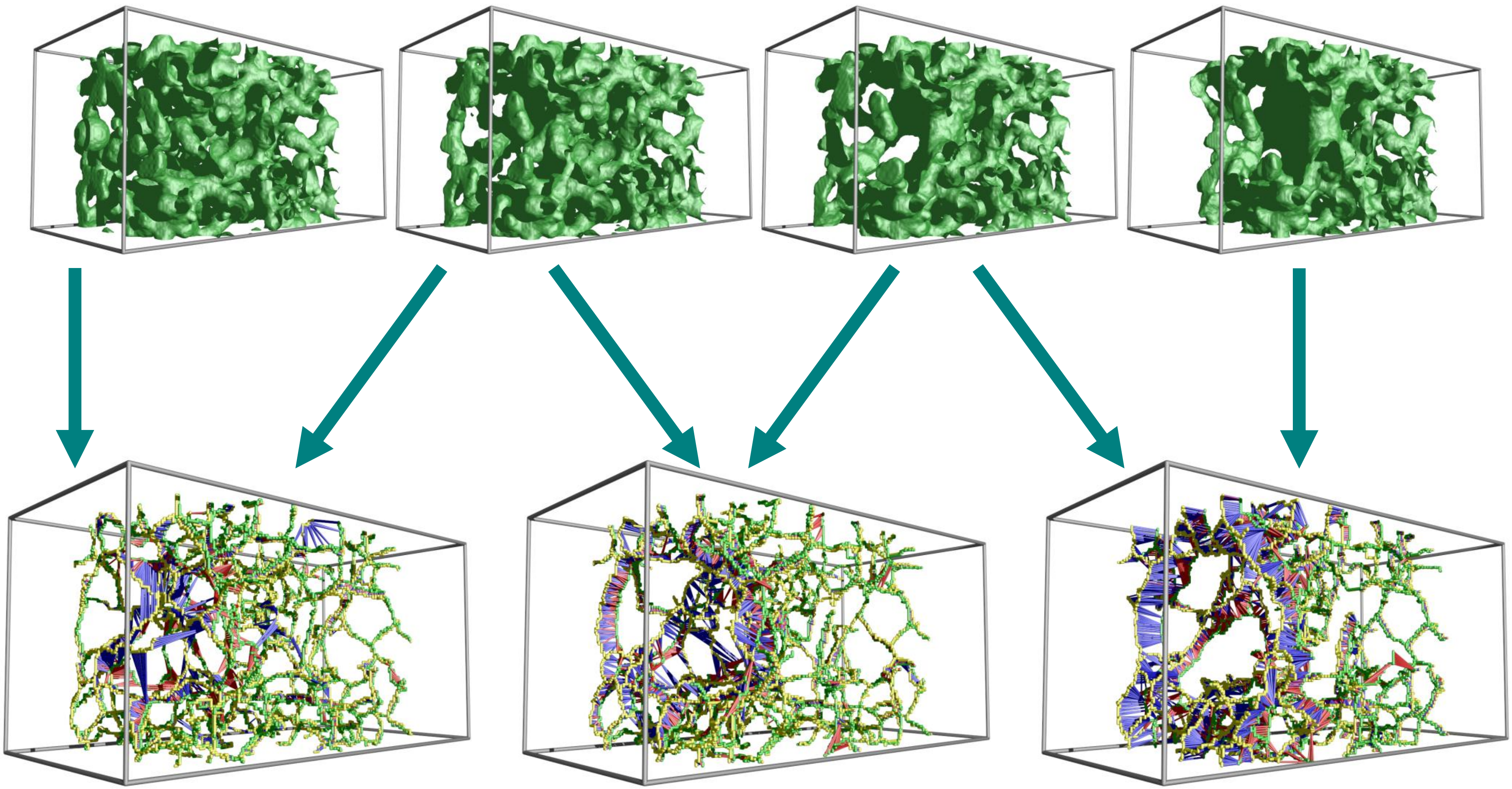


We Obtain a Robust Reconstruction of the Filament Structures in the Material



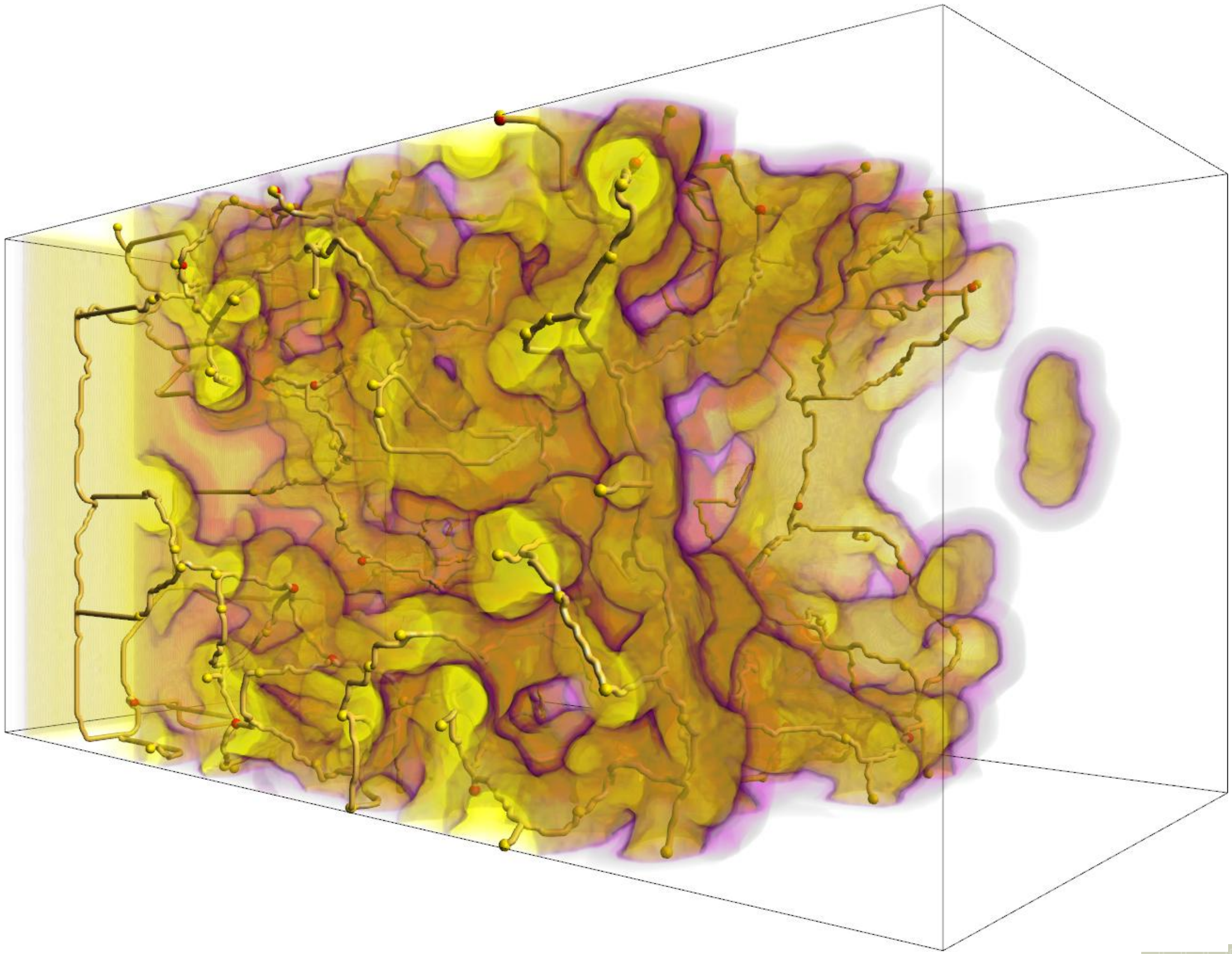


We Track the Evolution of the Filament Structure of the Material Under Impact



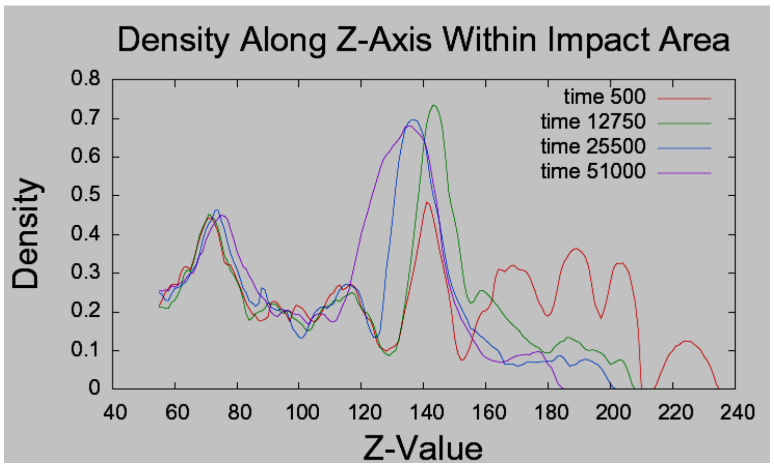
Time comparison of the reconstructions

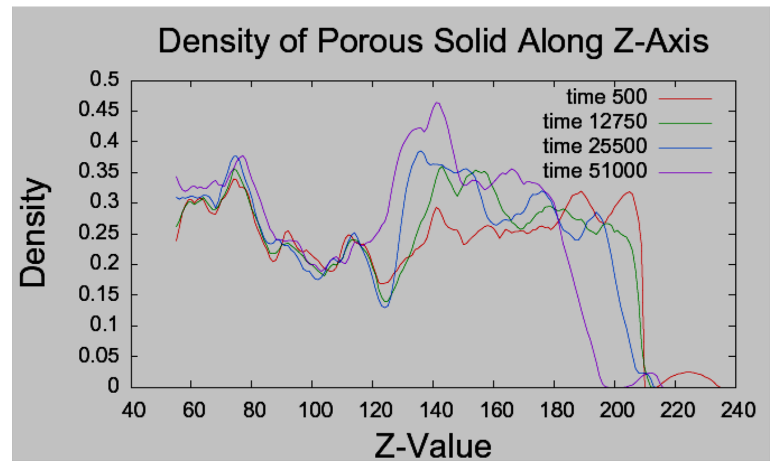
Demo Porous Medium



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

Density profiles

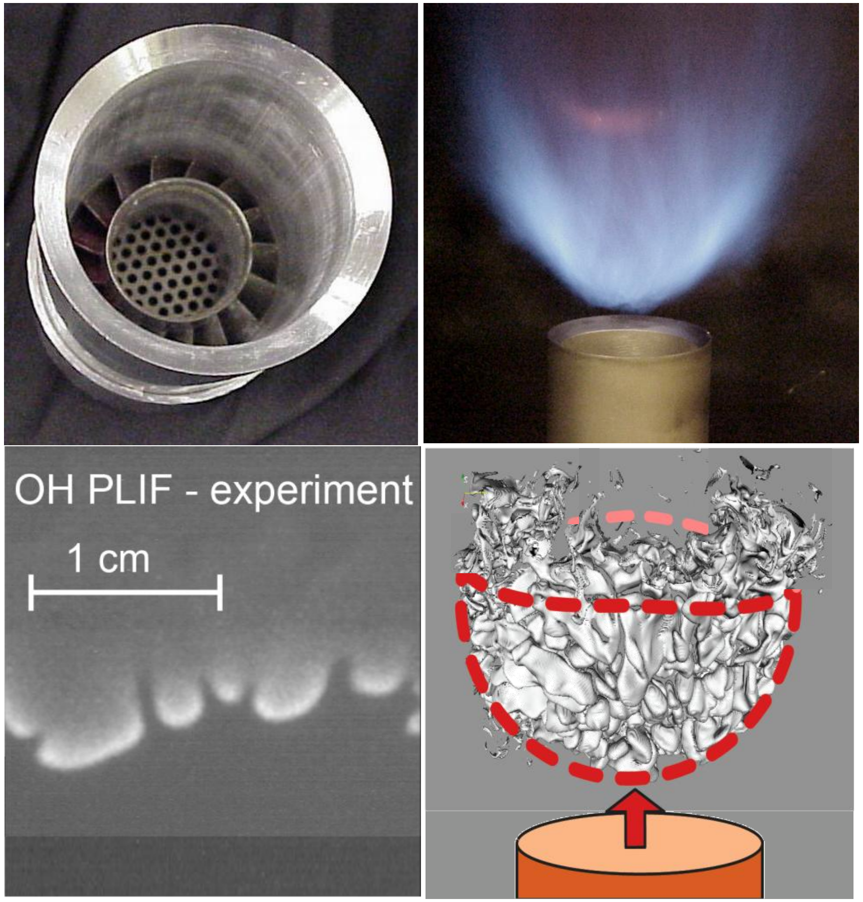




Decay in porosity of the material

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

Understanding Turbulence for Low Emission, High Efficiency Combustion



Experiment **Simulation**

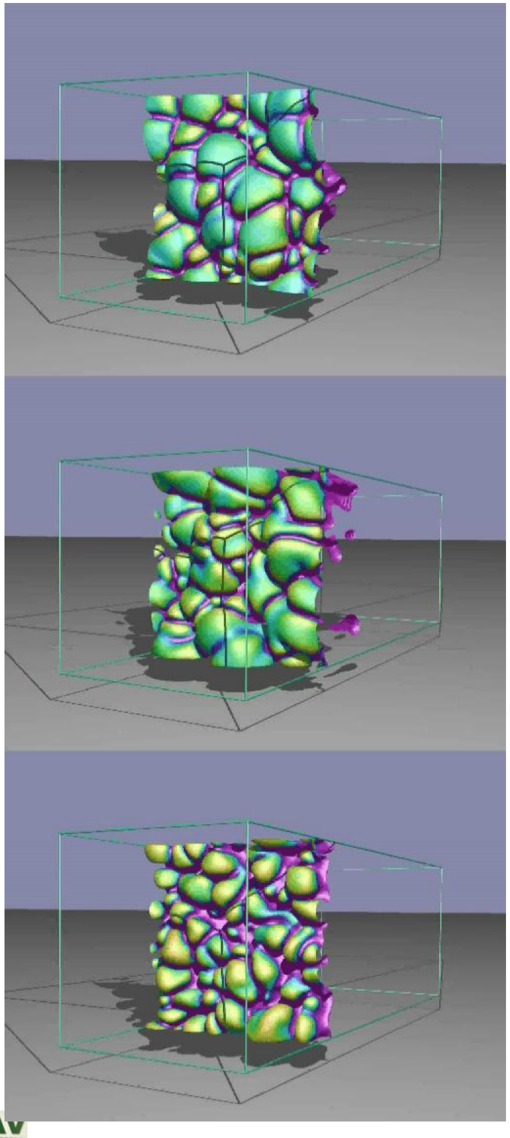
- **Lean premixed H_2 flames**
- **Low Swirl Combustion (LSC) Burners**
- **<u>Low pollution</u> in energy production**
- **<u>High Efficiency</u> in fuel consumption**
- **Scalable from residential to industrial use**
- **Each variable 3.9-4.5 TB**

1" burner (5 kW, 17 KBtu/hr) 28" burner (44 MW, 150 MBtu/hr)

CEDMAV

SCI THE UNIVERSITY OF UTAH Pacific Northwest NATIONAL LABORATORY

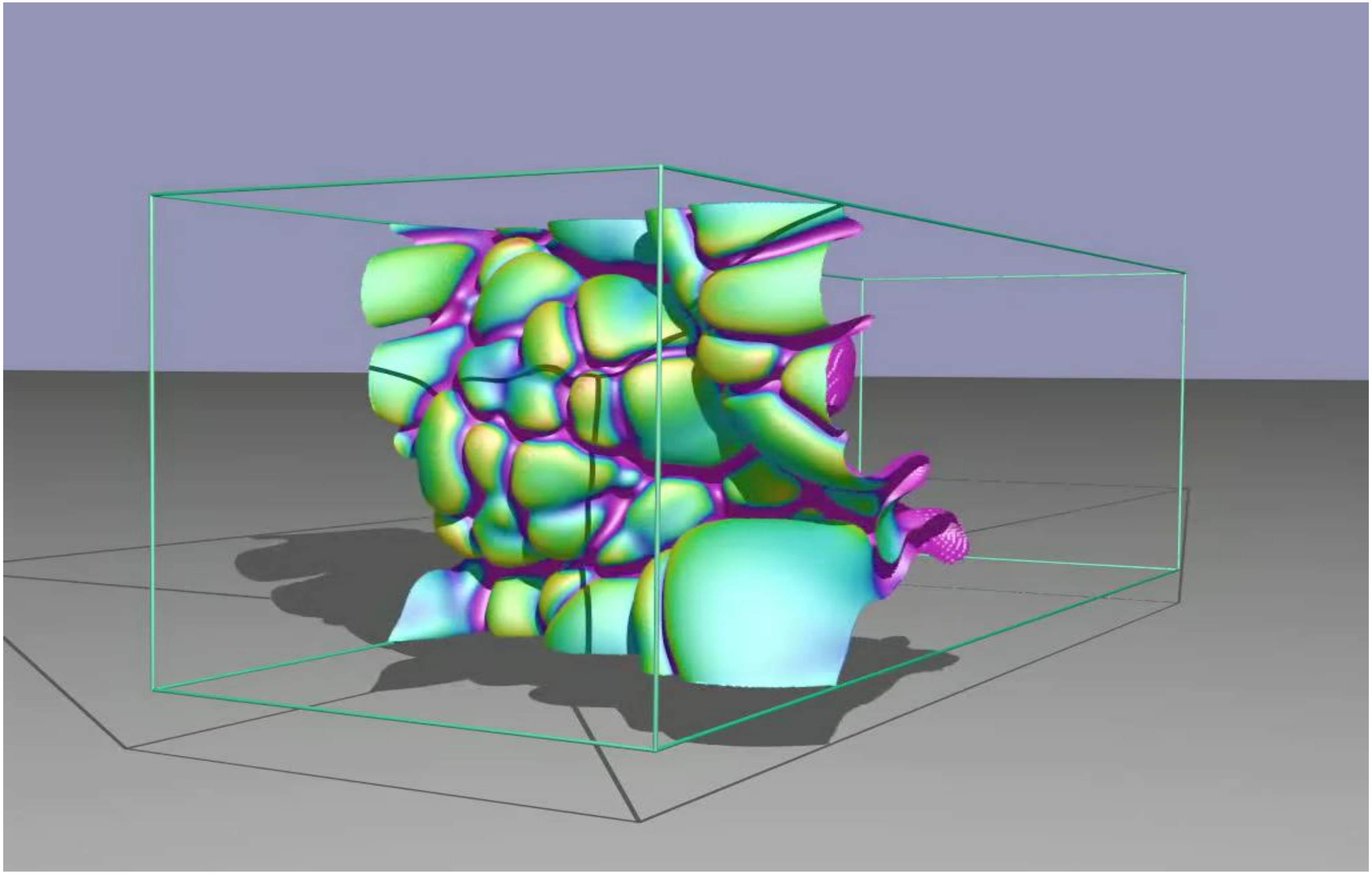
We Take on the Challenge of Developing a Quantitative Analysis Detecting Turbulence



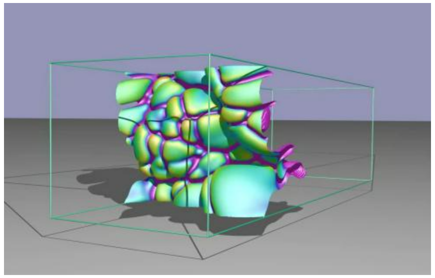
Understanding combustion processes over a broad range of burning conditions is an important problem for designing engines and power plants.

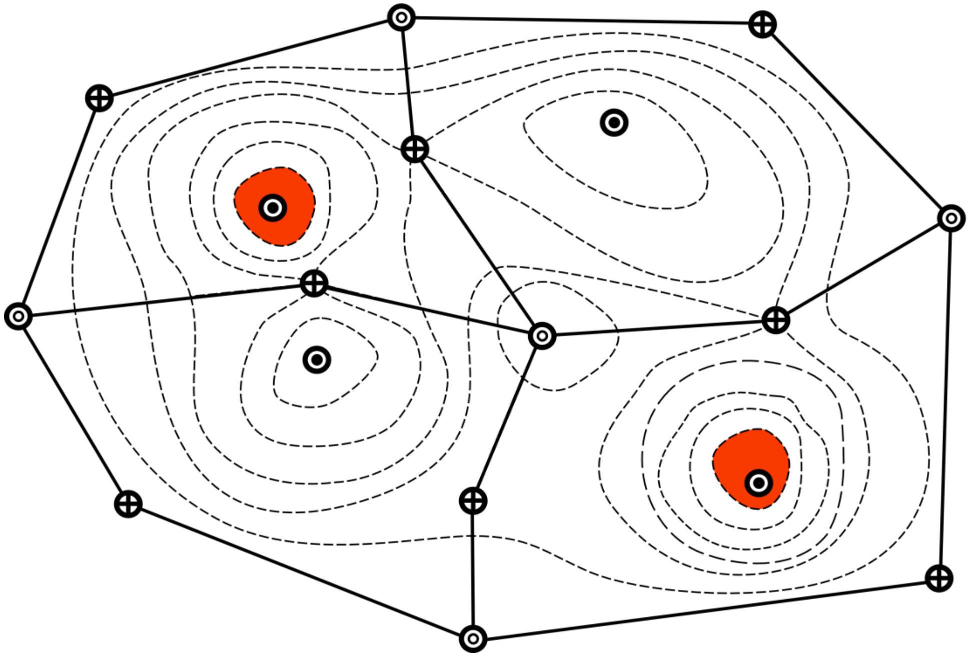
- **Simulation with AMR mesh.**
- **Simulations of lean premixed hydrogen flames with three degrees of turbulence.**
- **Can we identify precisely and track in time burning regions?**
- **Can we discriminate the degree of turbulence from a quantitative analysis?**

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces

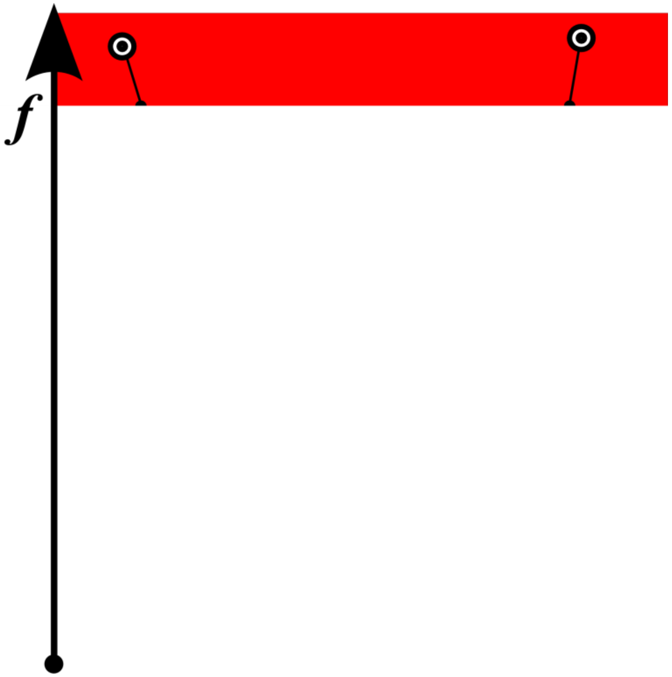


We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



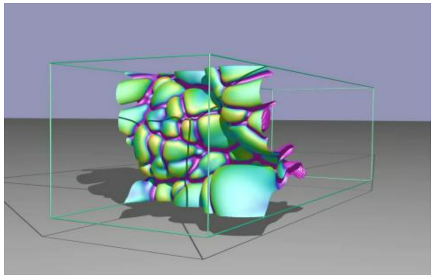


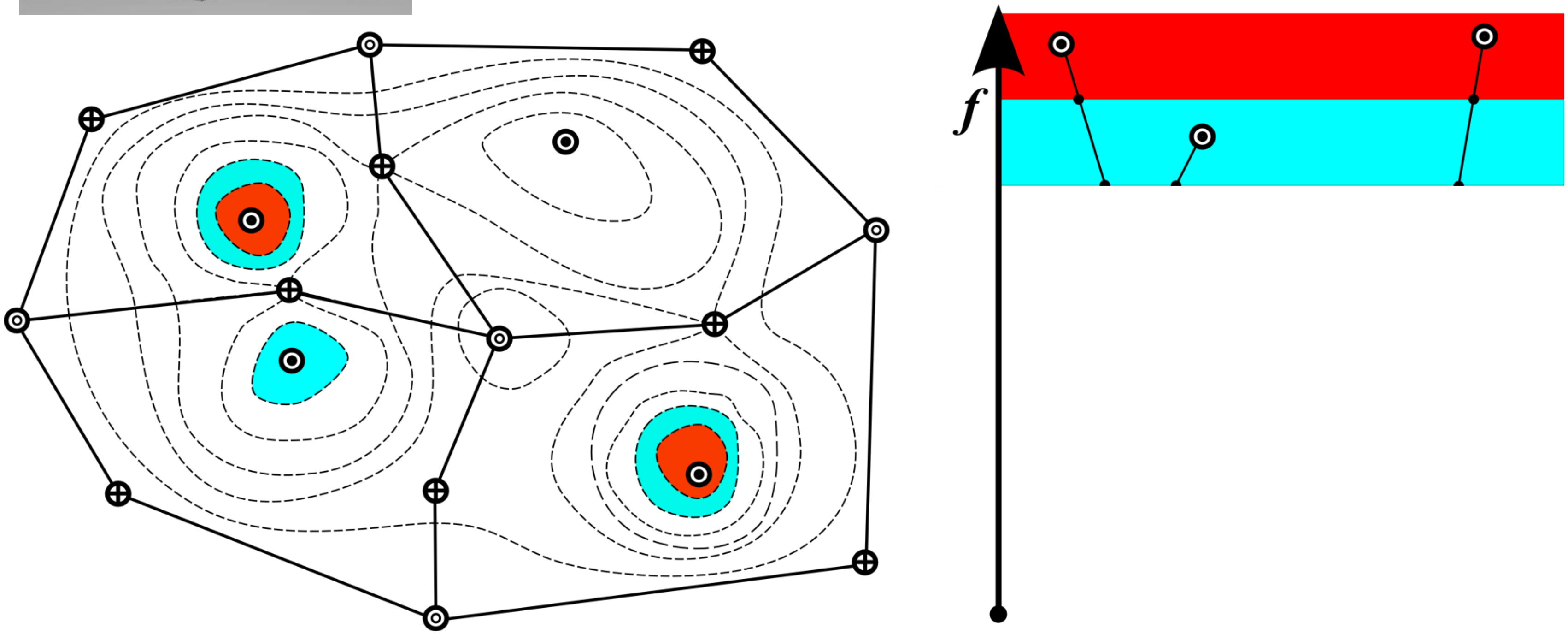
domain



hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces

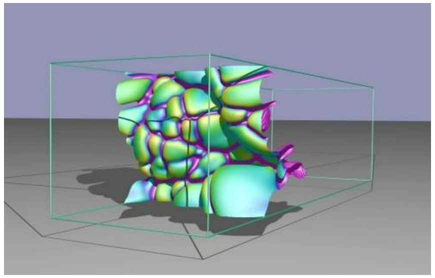


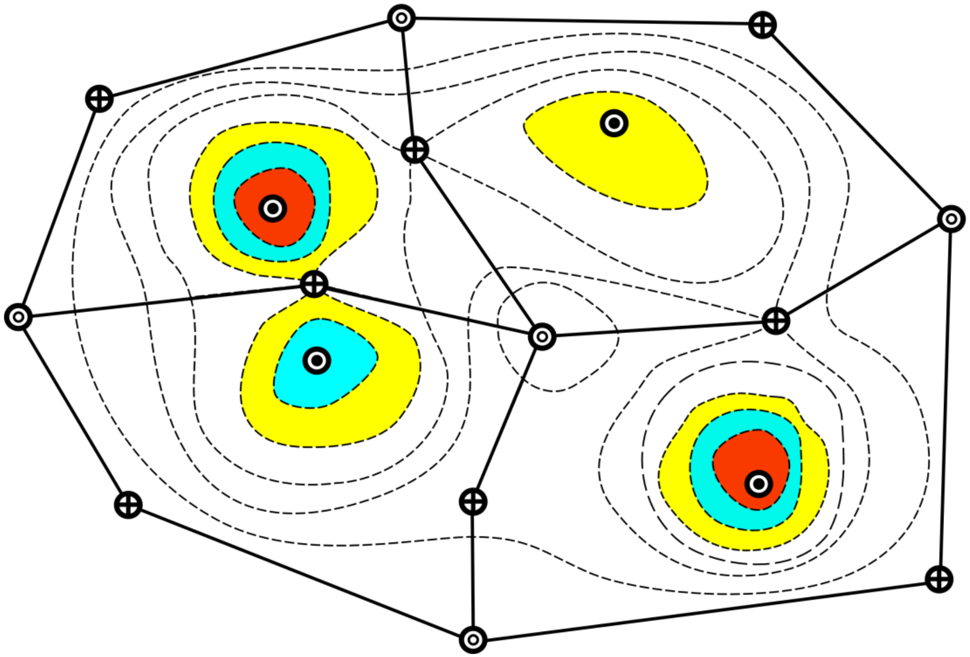


domain

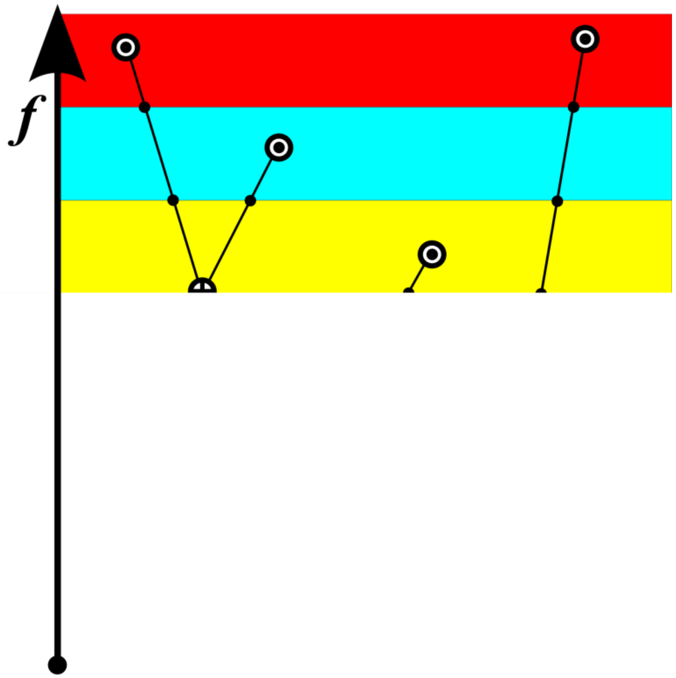
hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



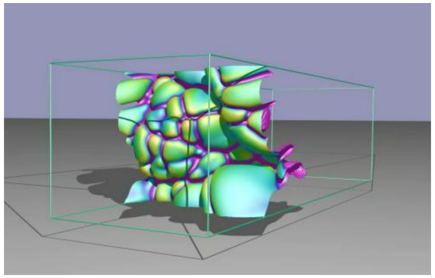


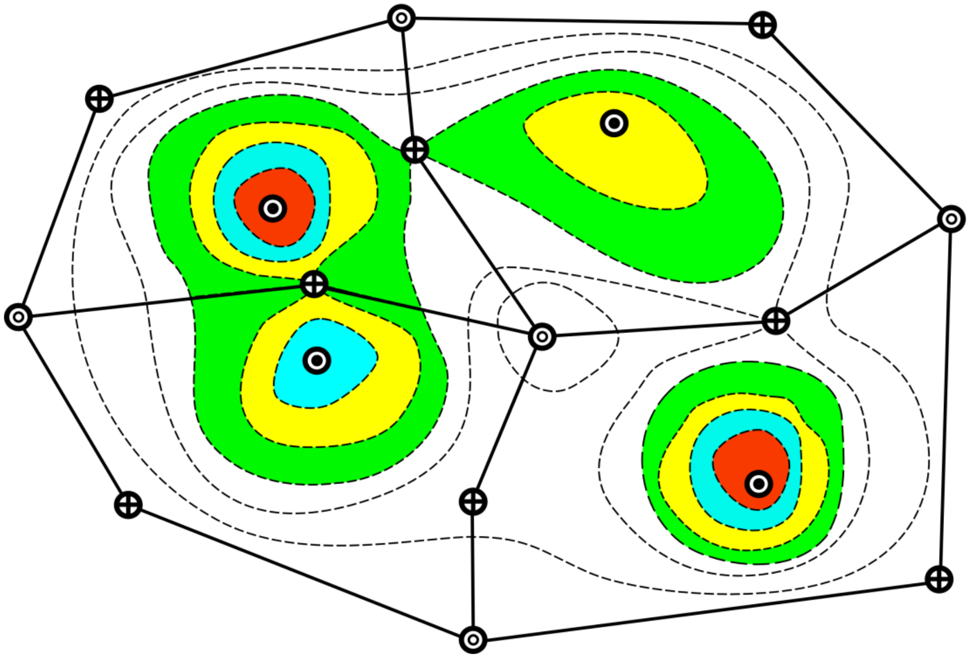
domain



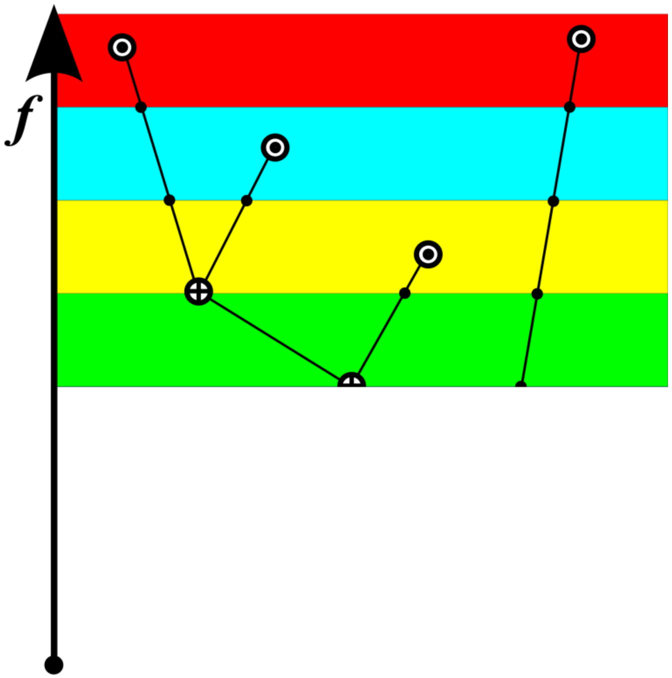
hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



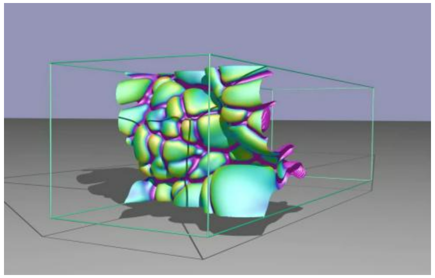


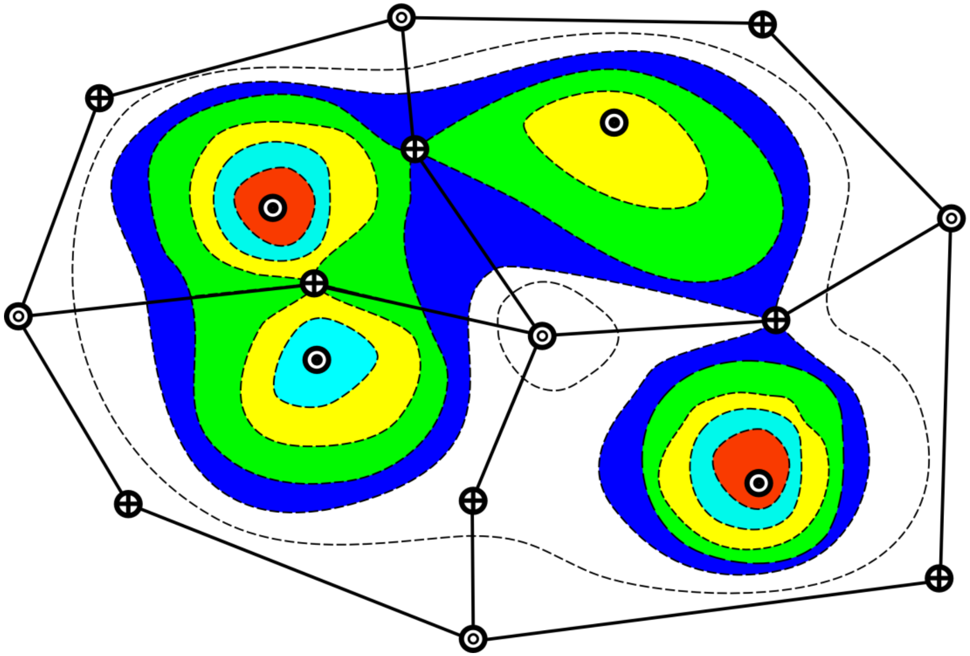
domain



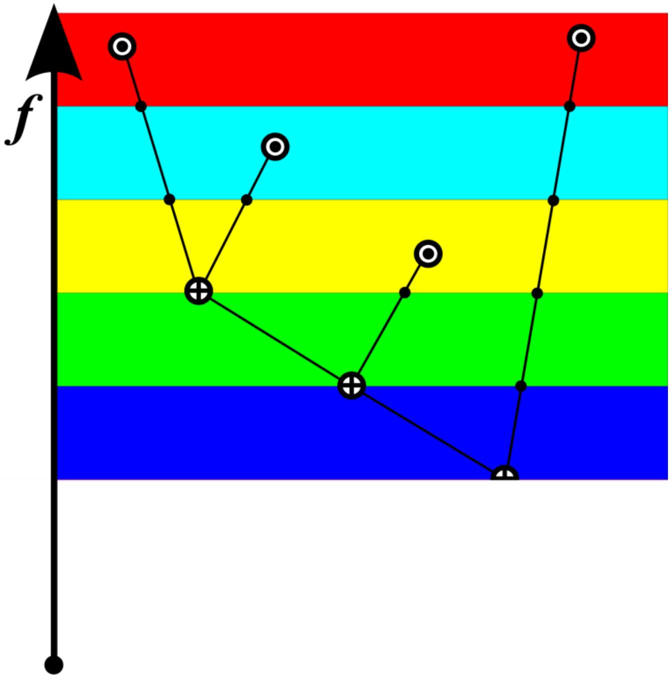
hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



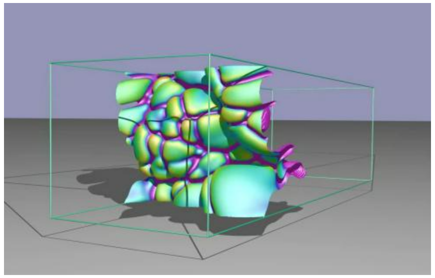


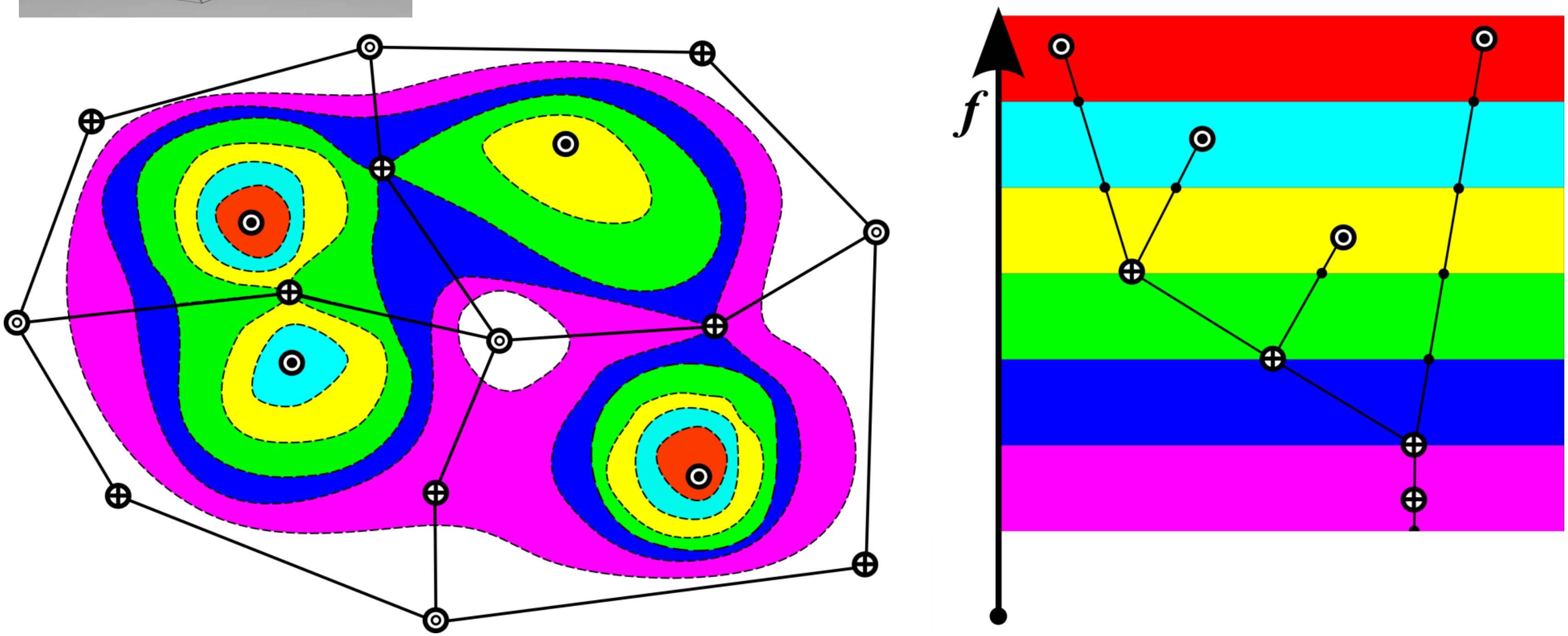
domain



hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces

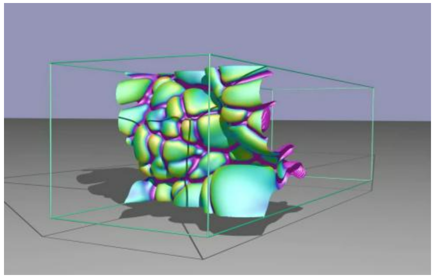


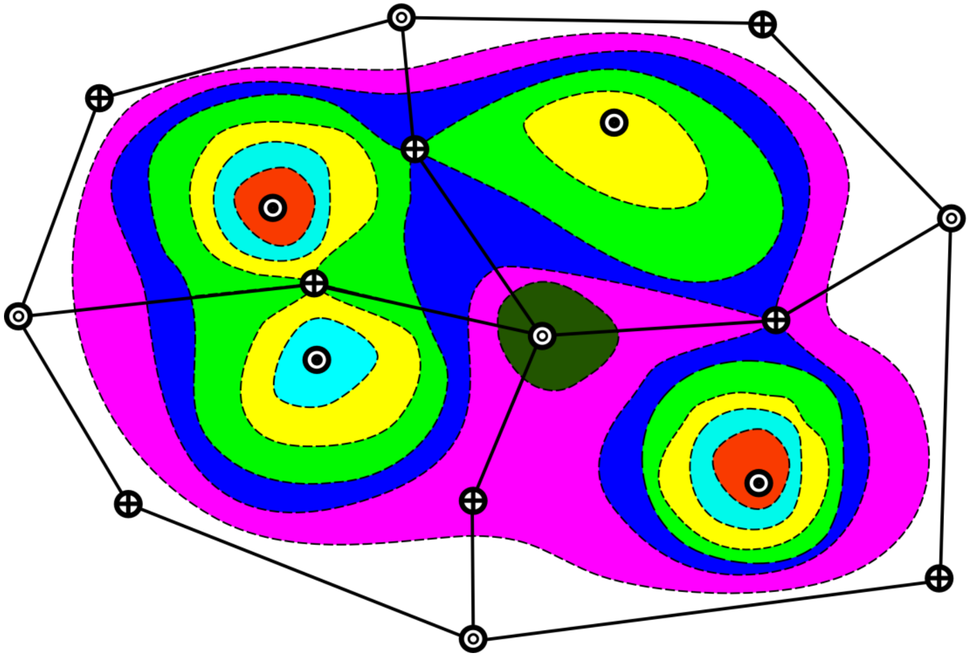


domain

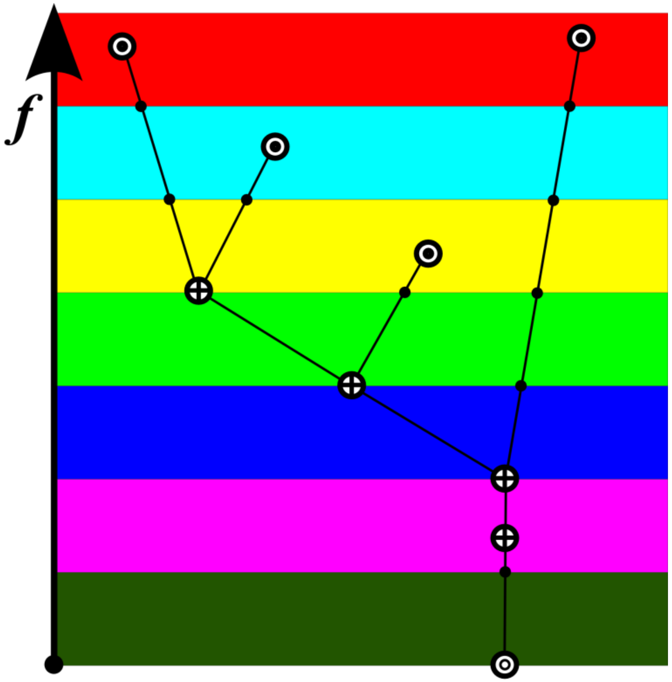
hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



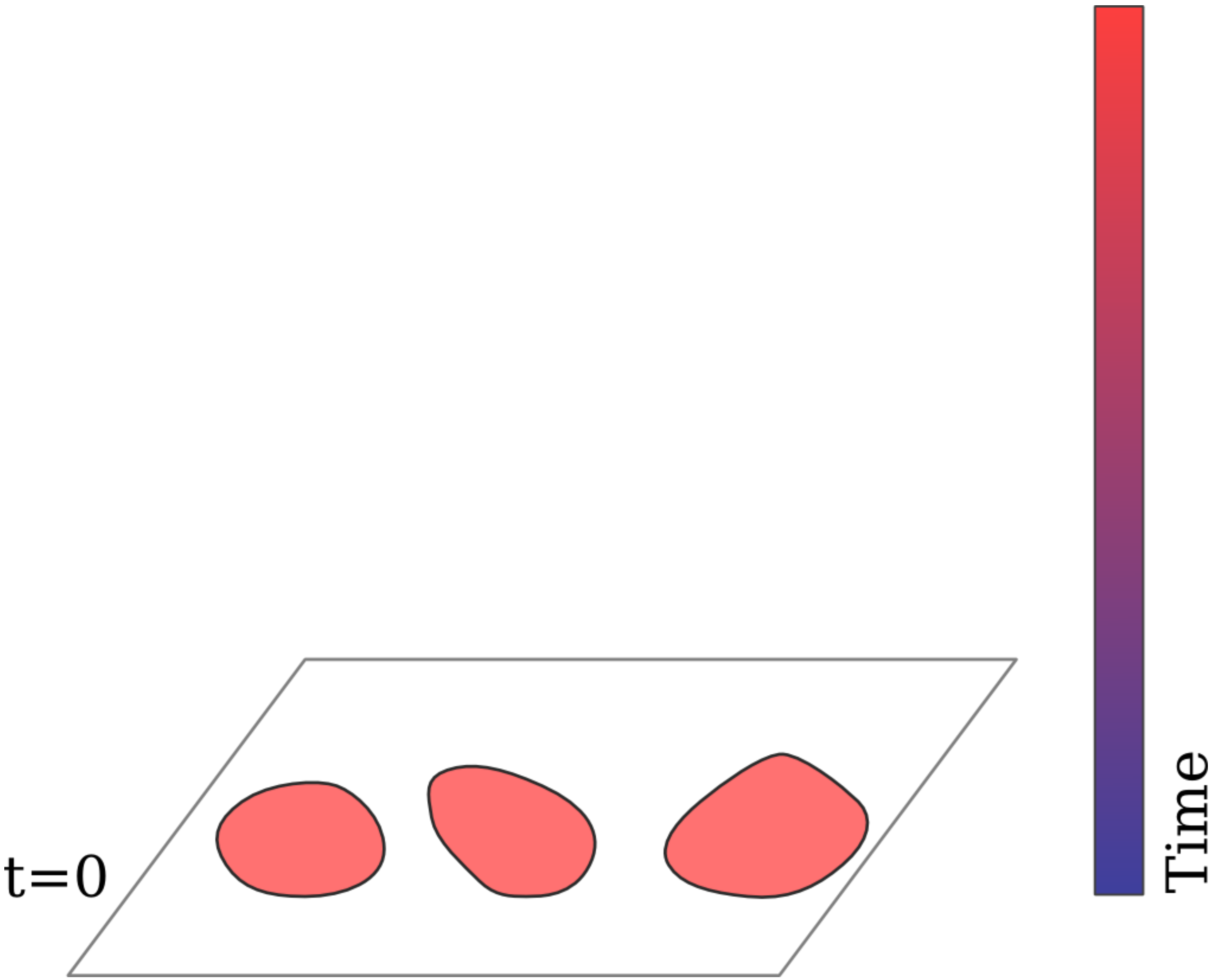


domain

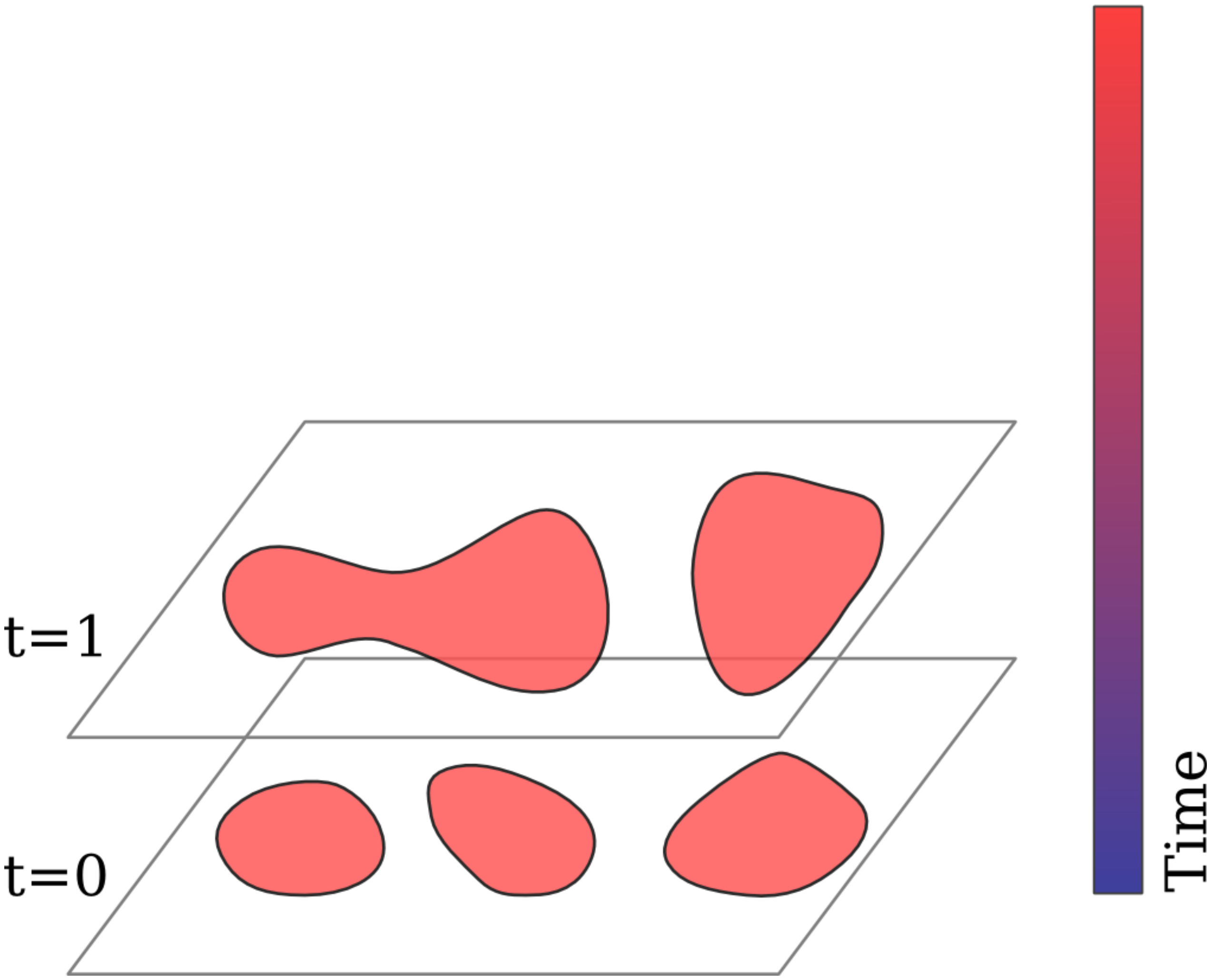


hierarchy

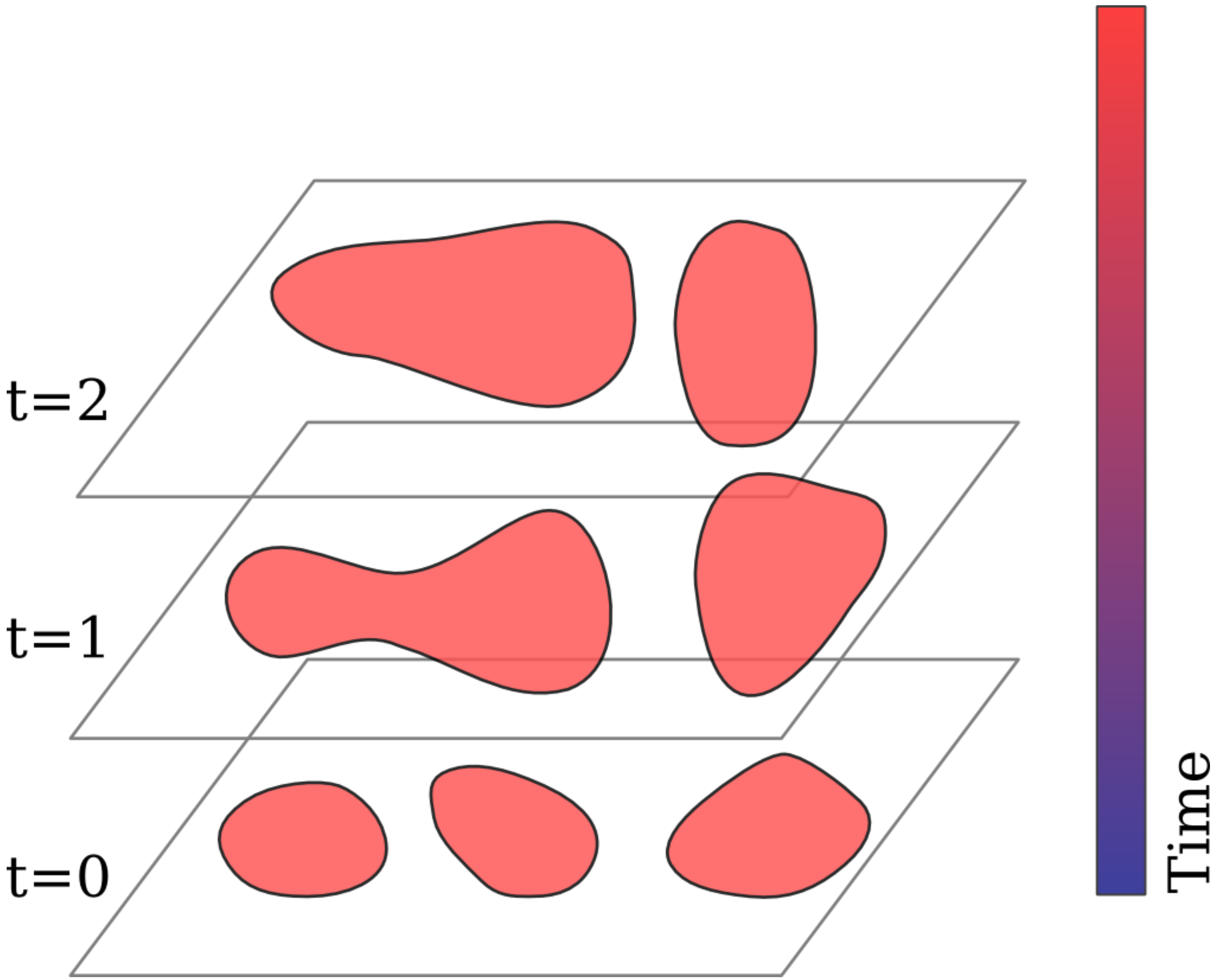
We track in time by interpolation in 4D and contraction to a Reeb Graph



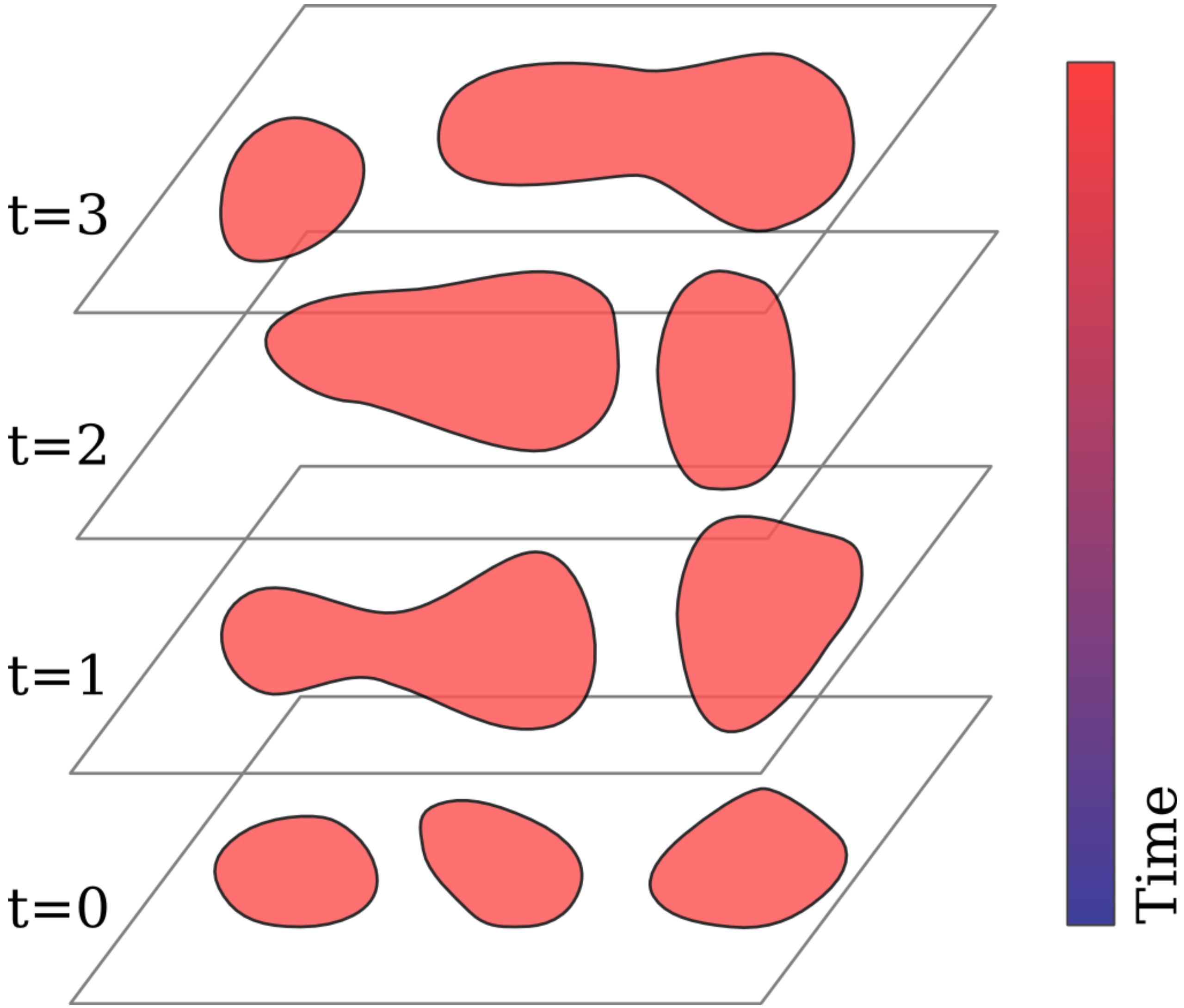
We track in time by interpolation in 4D and contraction to a Reeb Graph



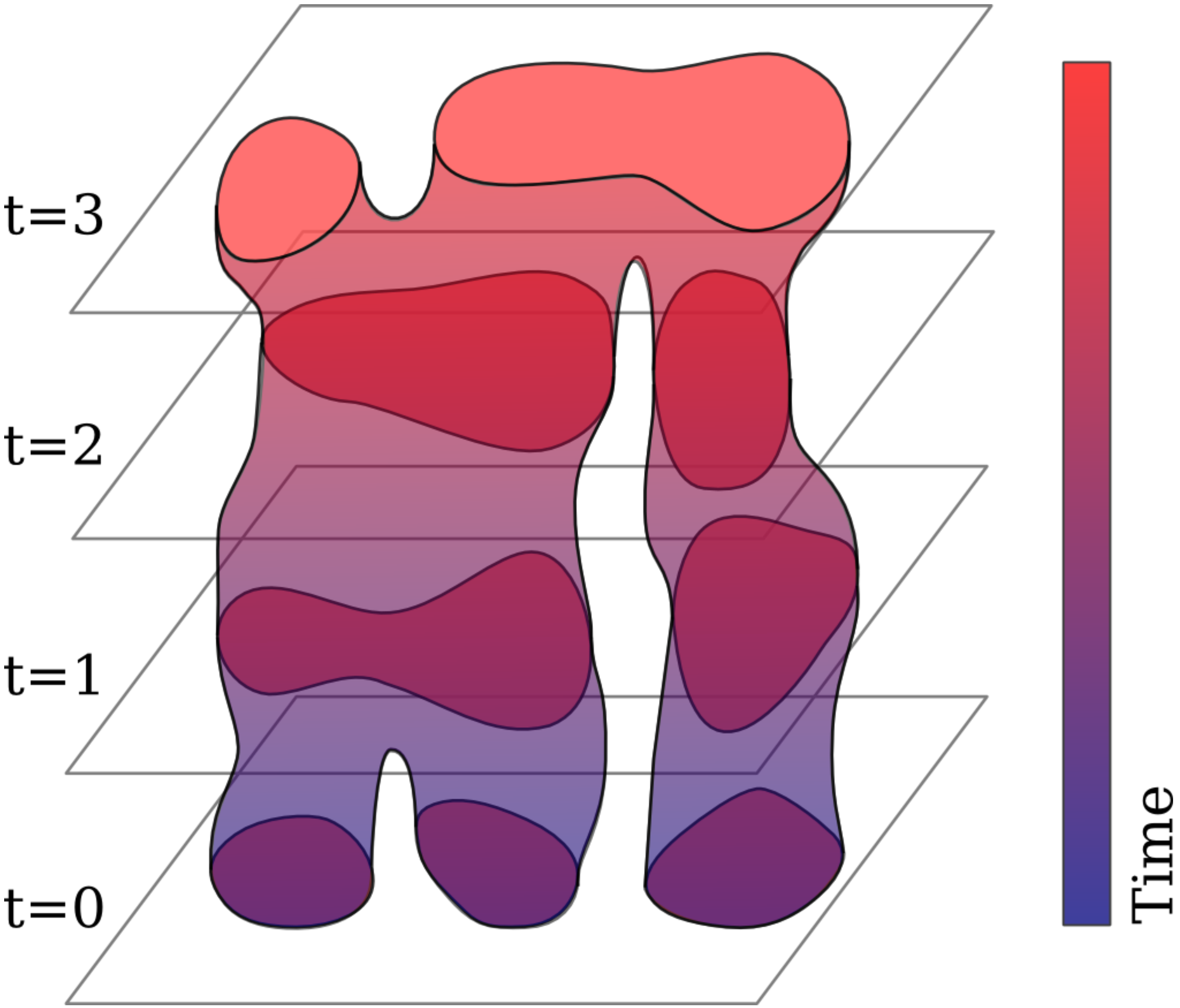
We track in time by interpolation in 4D and contraction to a Reeb Graph



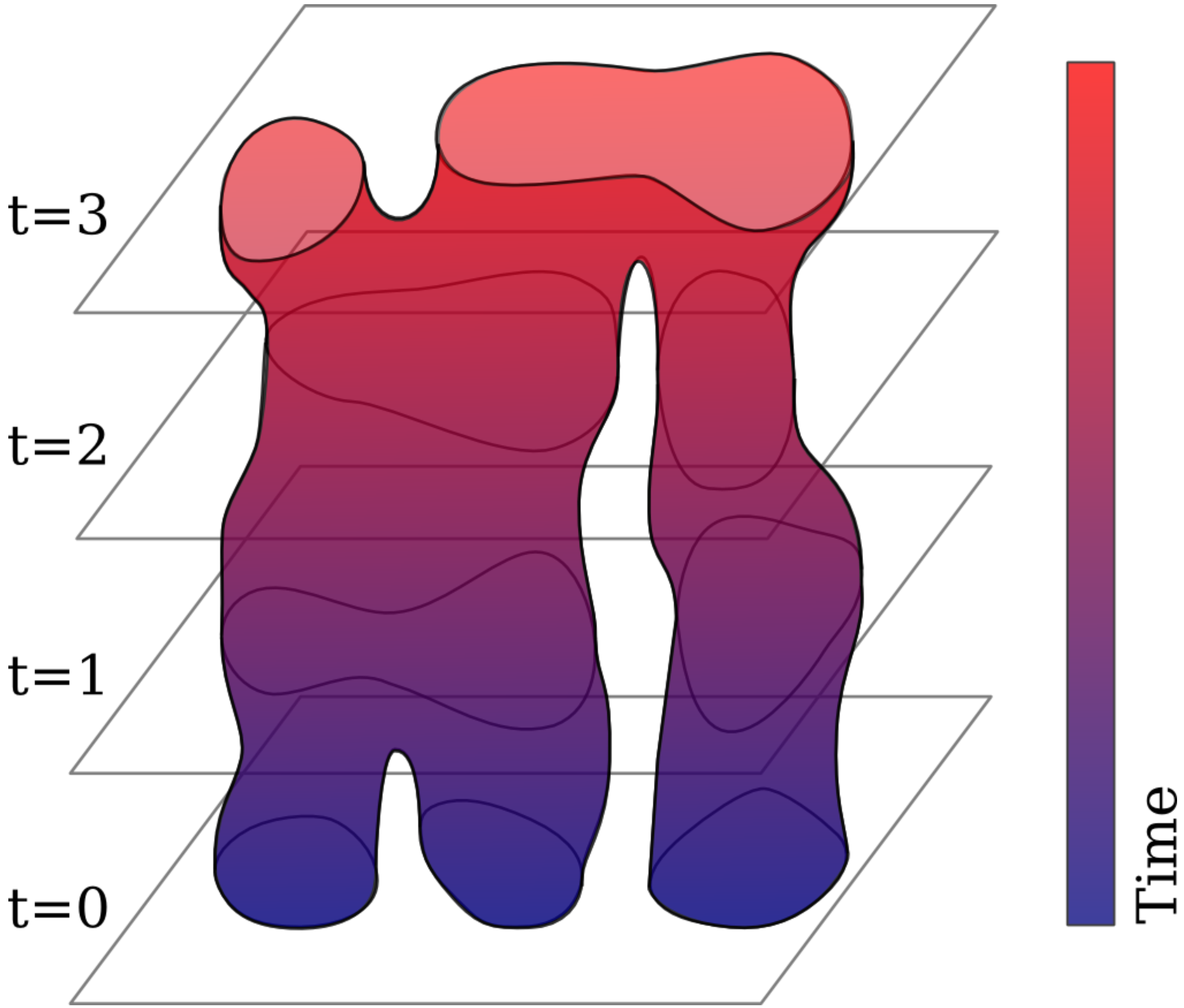
We track in time by interpolation in 4D and contraction to a Reeb Graph



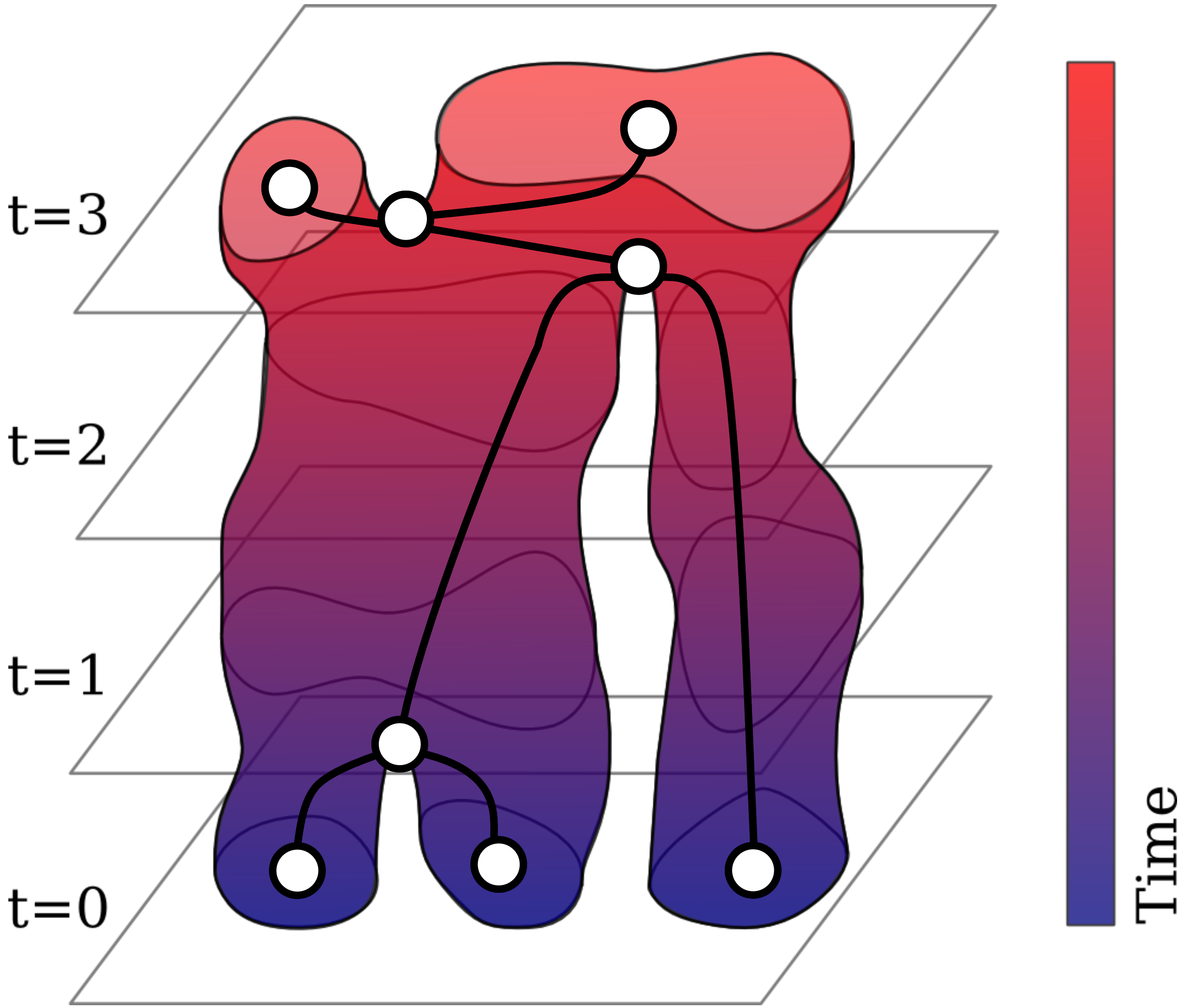
We track in time by interpolation in 4D and contraction to a Reeb Graph



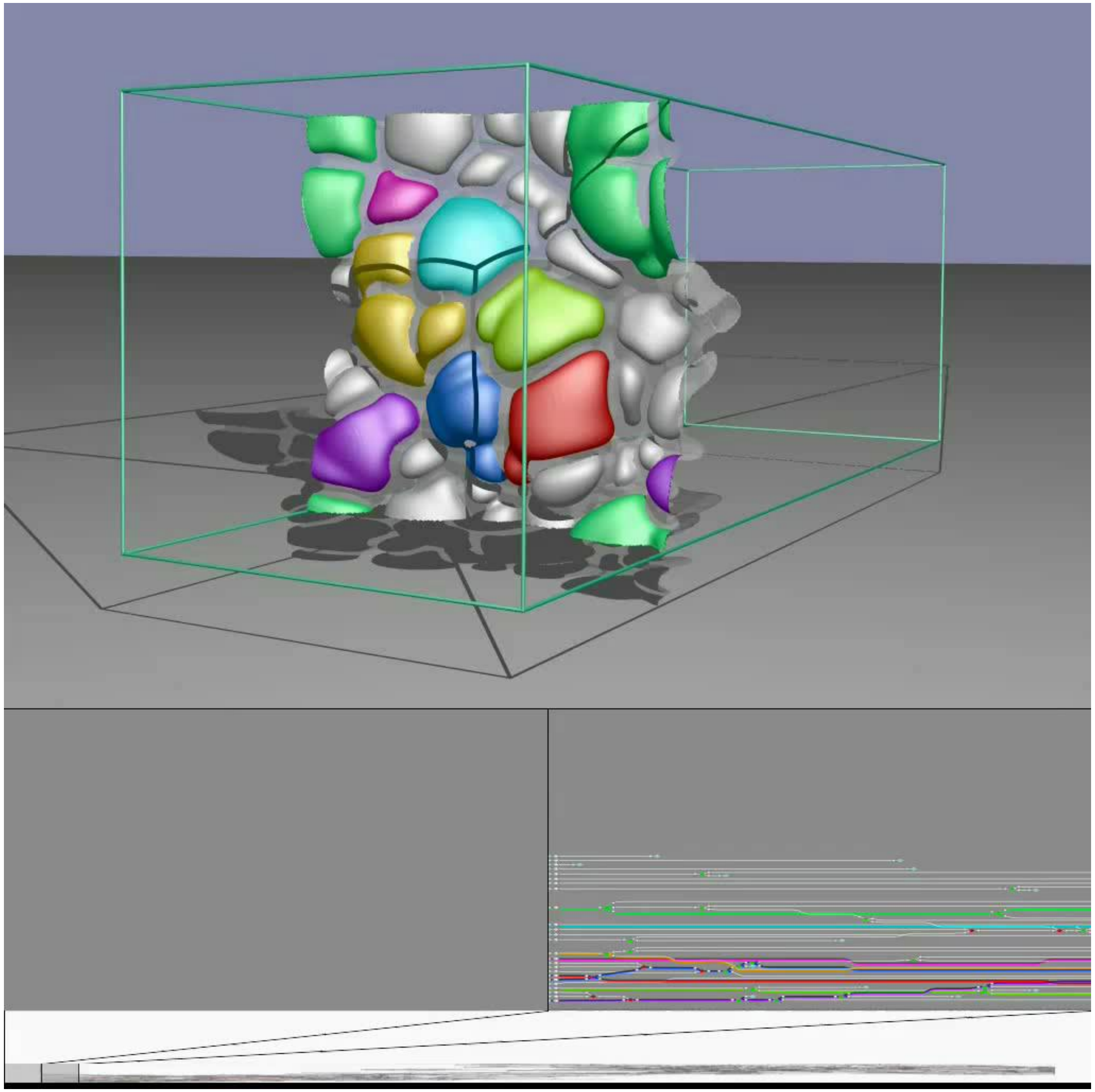
We track in time by interpolation in 4D and contraction to a Reeb Graph



We track in time by interpolation in 4D and contraction to a Reeb Graph

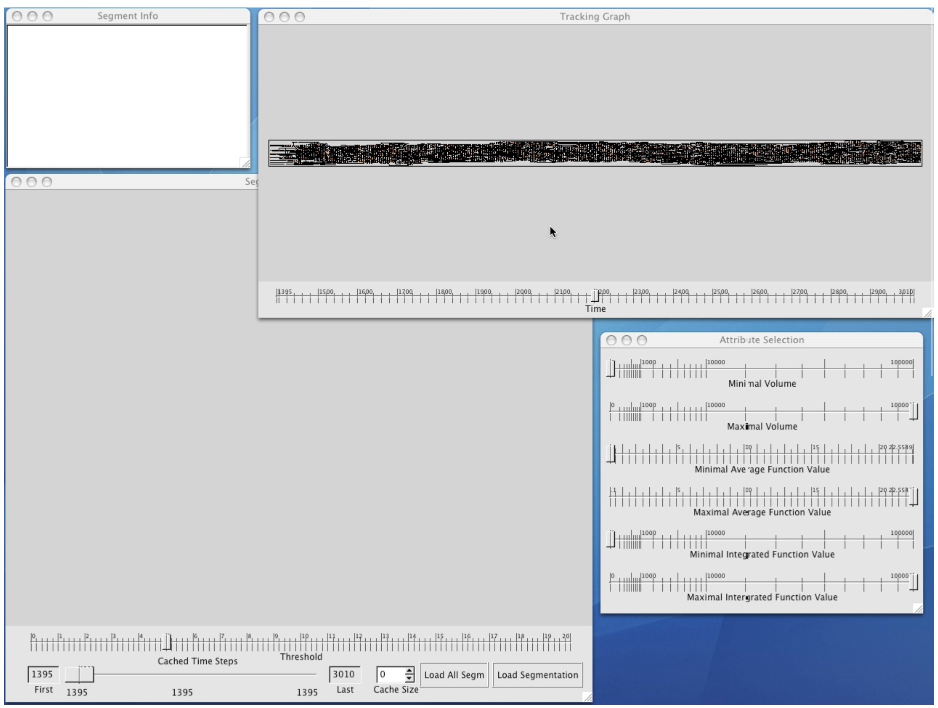


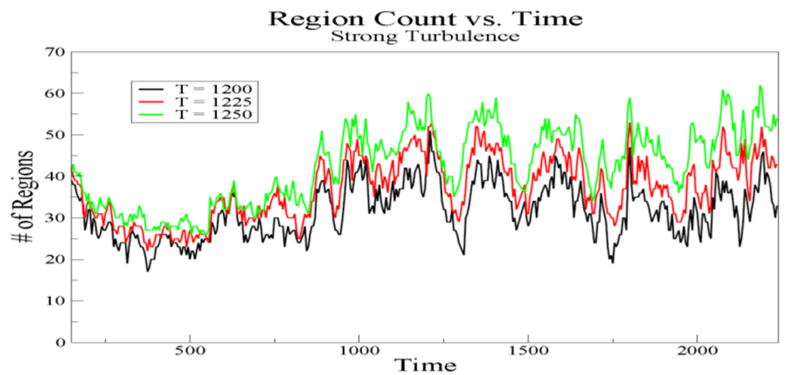
Each Set of Parameters Results in a Robust Segmentation and Tracking of Burning Cells

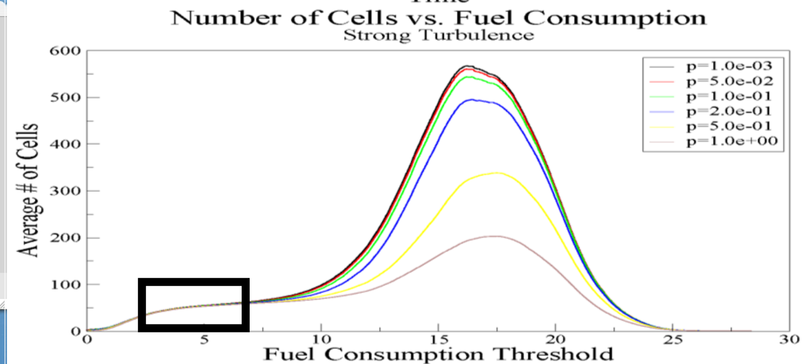


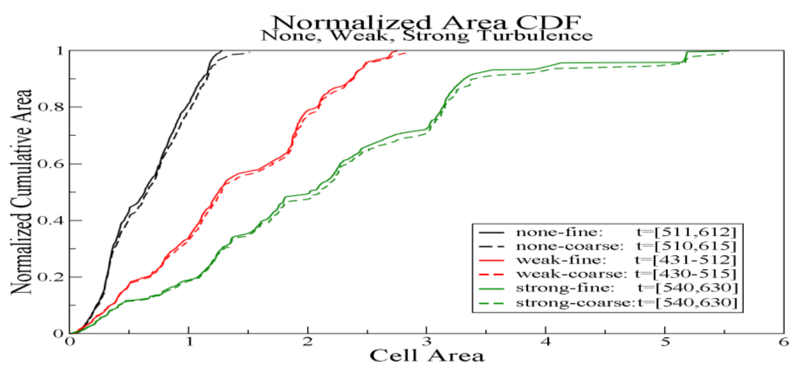
We Allow Exploration of the Full Space of Parameters Defining the Features

- **interactive exploration**
- **comparison of statistics**

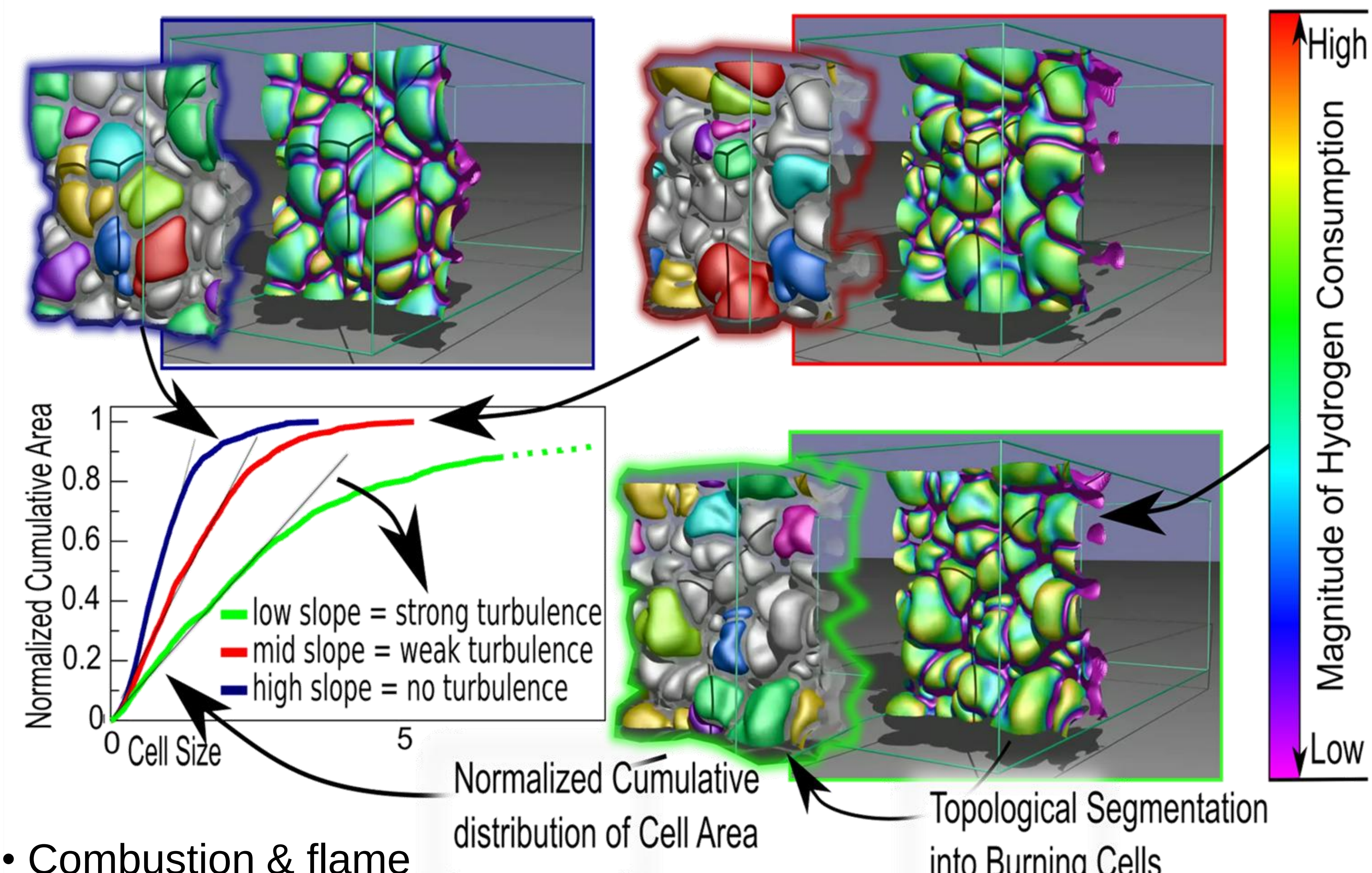








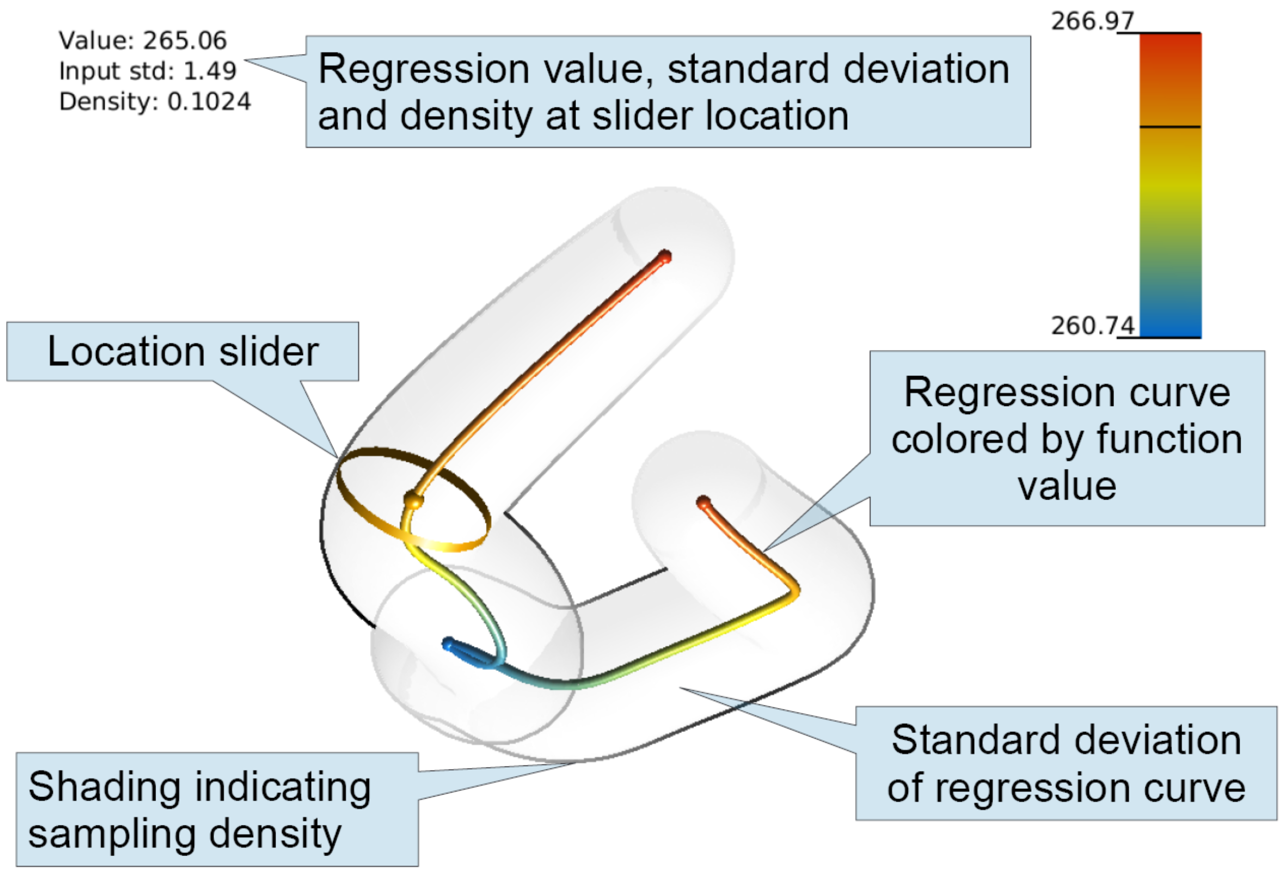
Topological Segmentation Allows to Quantify Turbulence as Slope of the Area Distributions



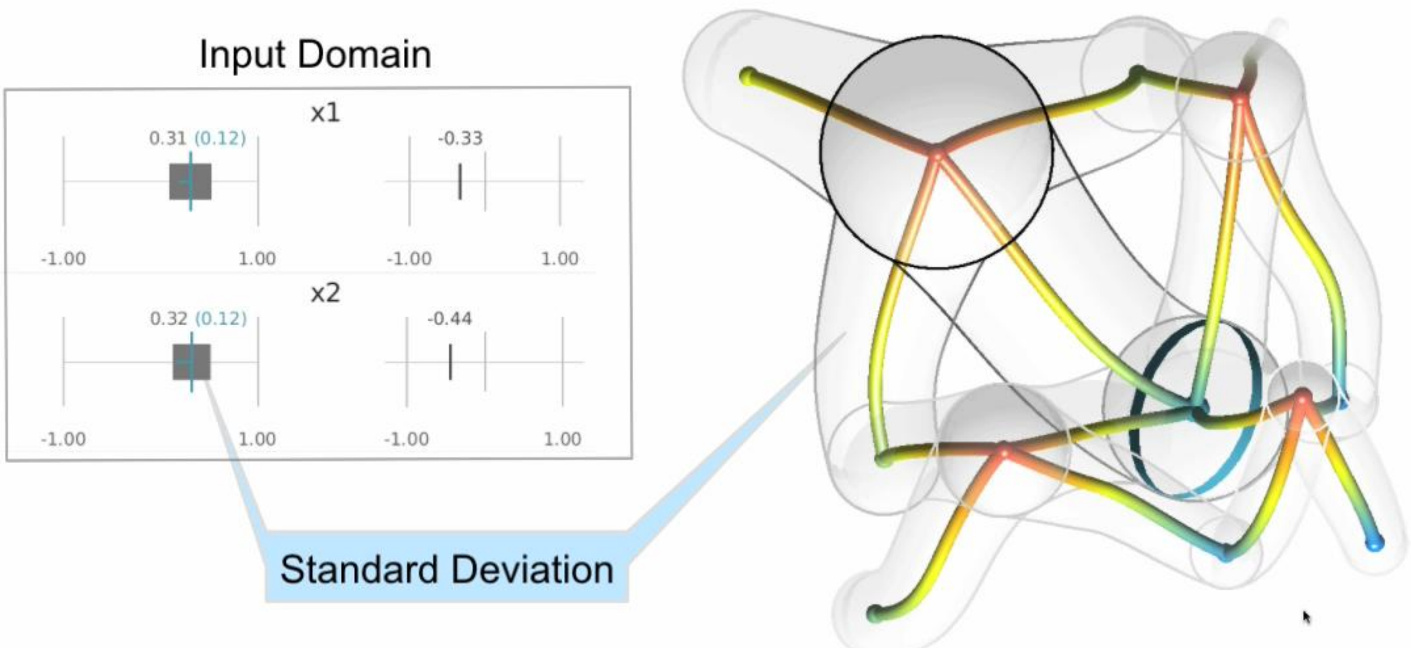
- Combustion & flame

Exploration of High Dimensional Functions for Sensitivity Analysis

Integrated presentation of statistics and topology



Exploration of High Dimensional Functions for Sensitivity Analysis



The set of regression curves provides a platform to visualize further information, such as standard deviation and sampling density. The color corresponds to the function value.

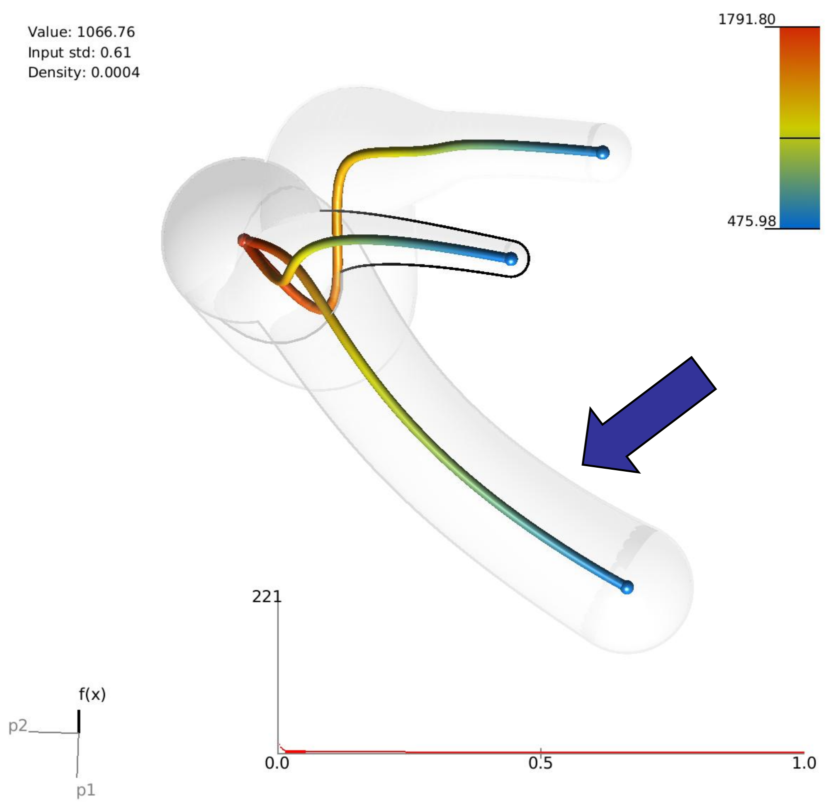
Analysis of Combustion Simulations

Combustion Simulation of Jet CO/H2-Air Flames

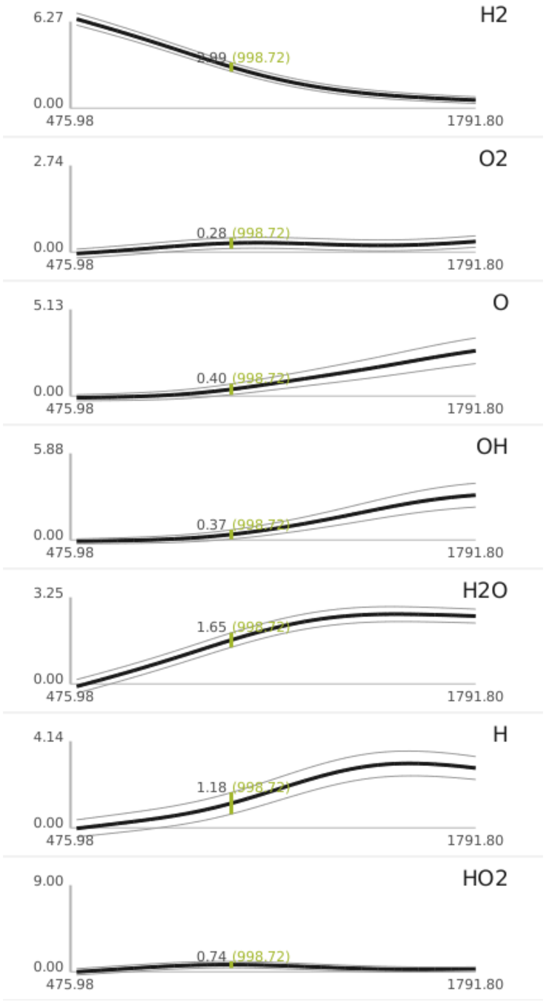
Input: Composition of 10 chemical species

Output: Temperature

The Framework Allows Detailed Visualization and Analysis of High Dimensional Functions

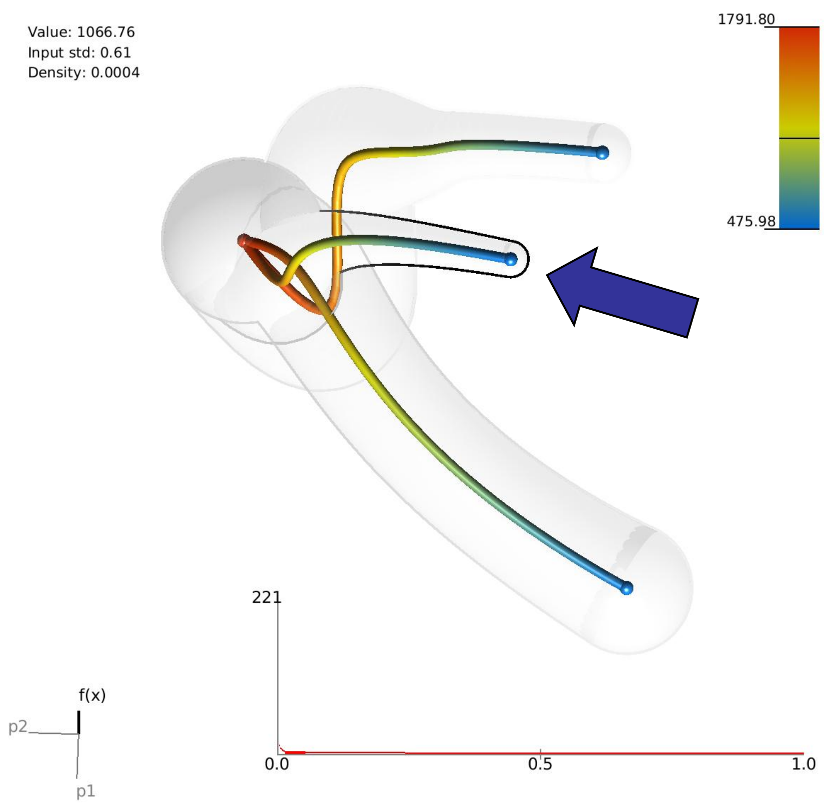


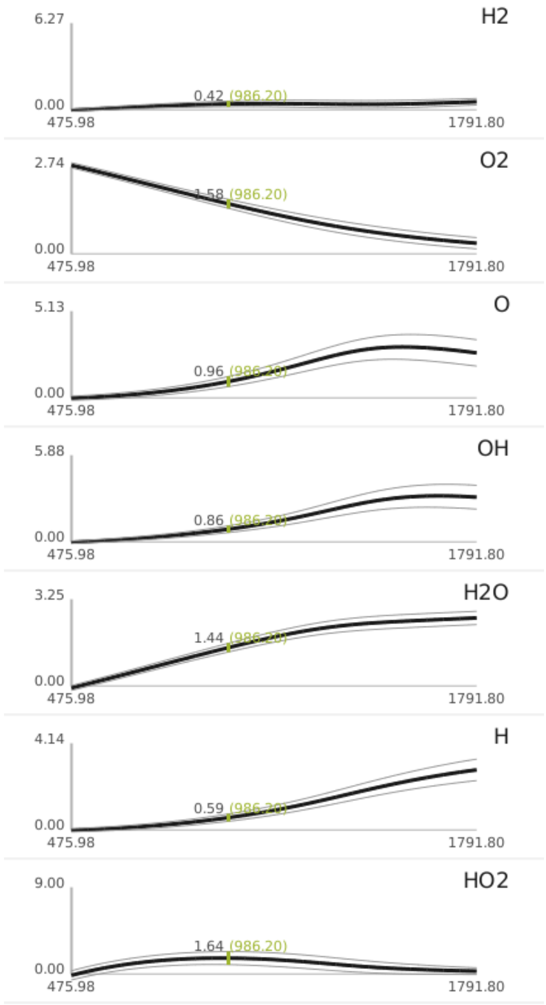
10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation



Pure fuel

The Framework Allows Detailed Visualization and Analysis of High Dimensional Functions

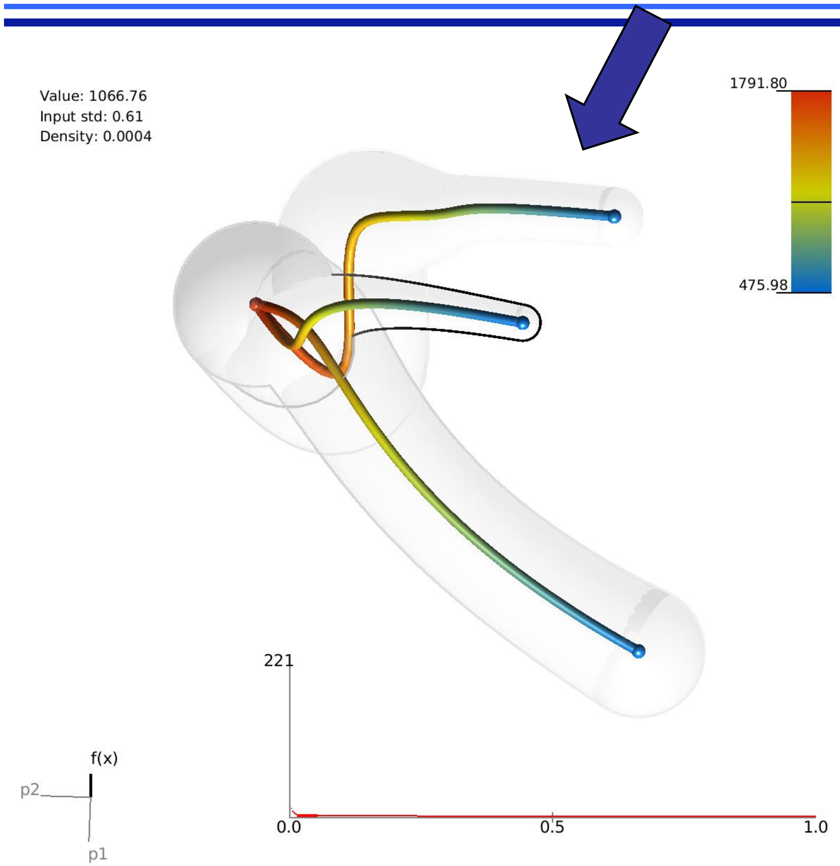


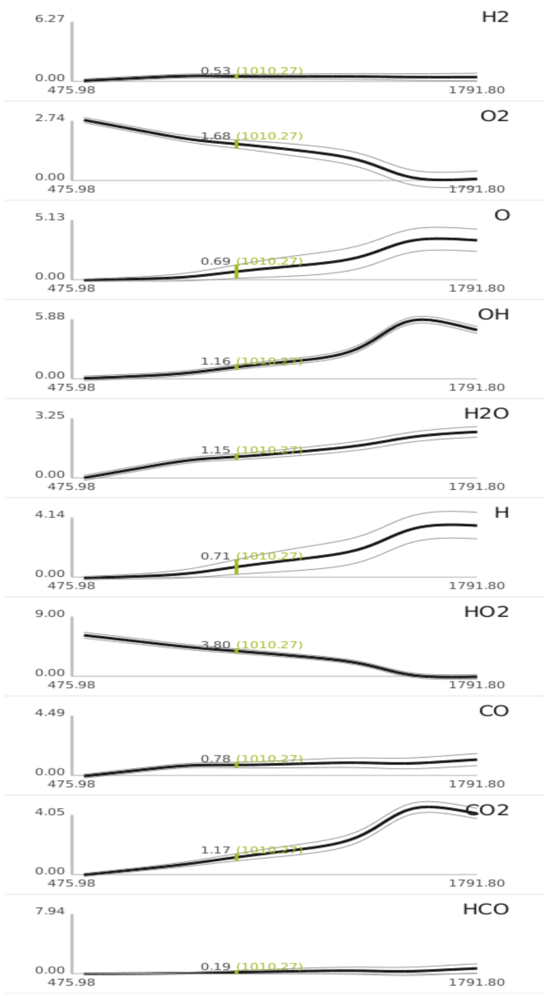


Pure oxidizer

10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation

The Framework Allows Detailed Visualization and Analysis of High Dimensional Functions





Local extinction

10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation

Combustion Simulation of Jet CO/H2-Air Flames

Input: Composition of 10 chemical species

Output: Temperature

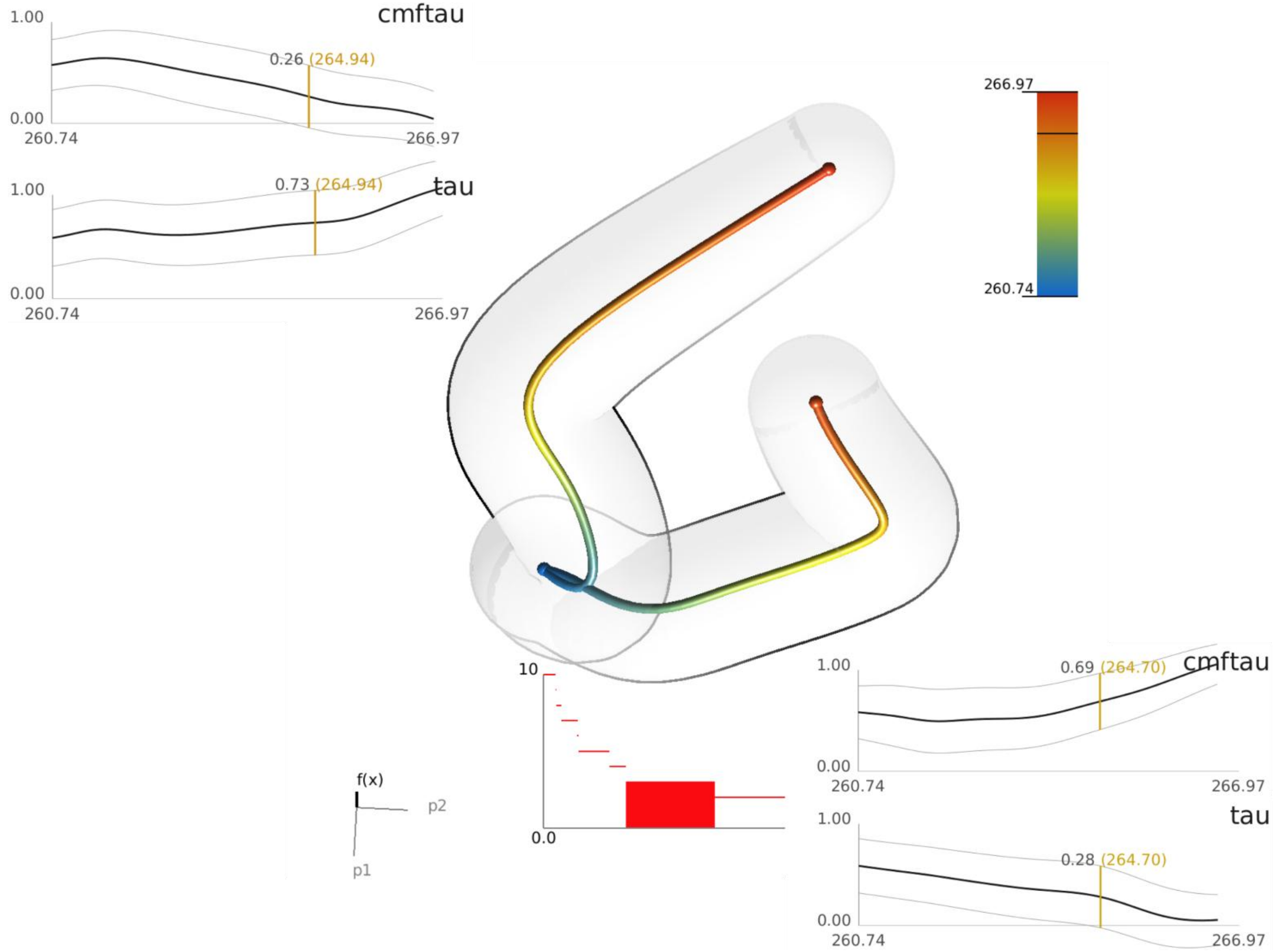
Analysis of Climate Data

Community Atmosphere Climate Model

Input: 21 parameter settings

Output: Net long wave flux (thermal radiation)

The Framework Reveals Relationship Between Convection and Global Long Wave Flux



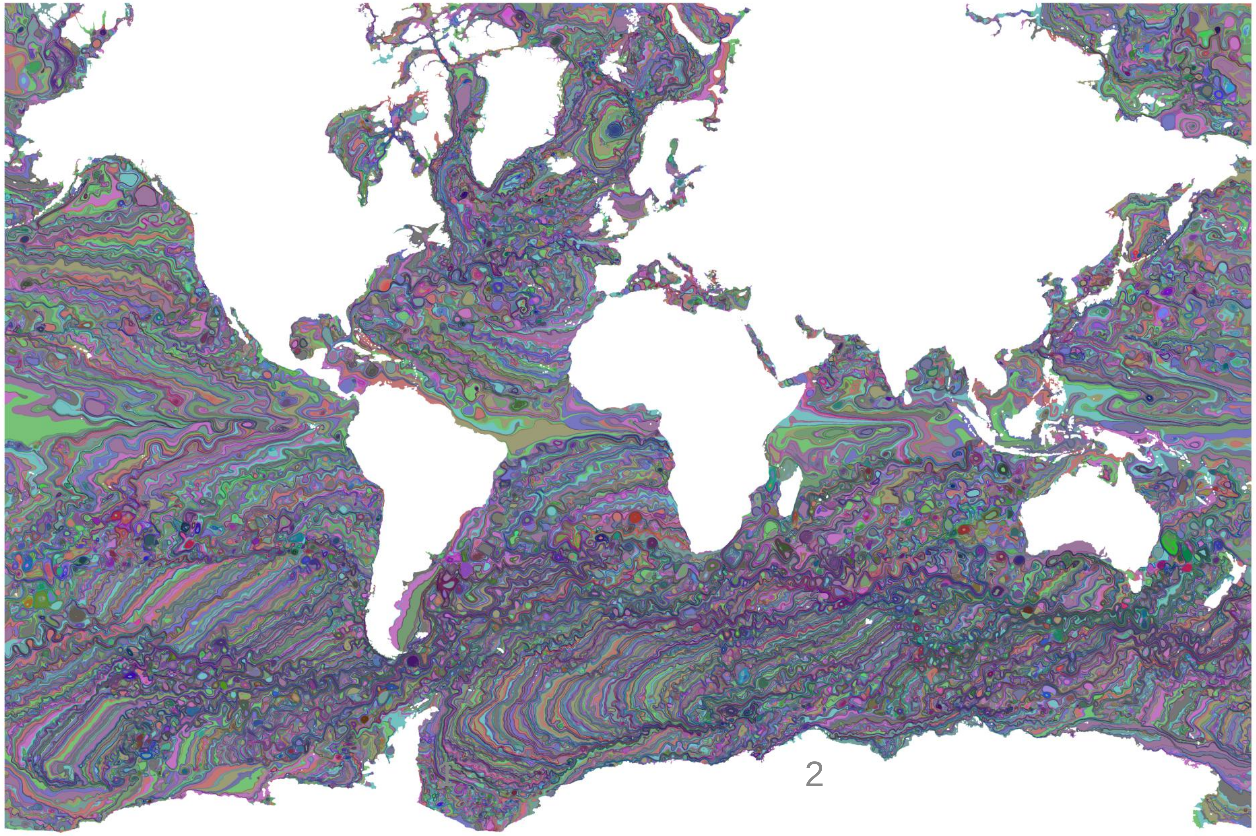
Community Atmosphere Climate Model

Input: 21 parameter settings

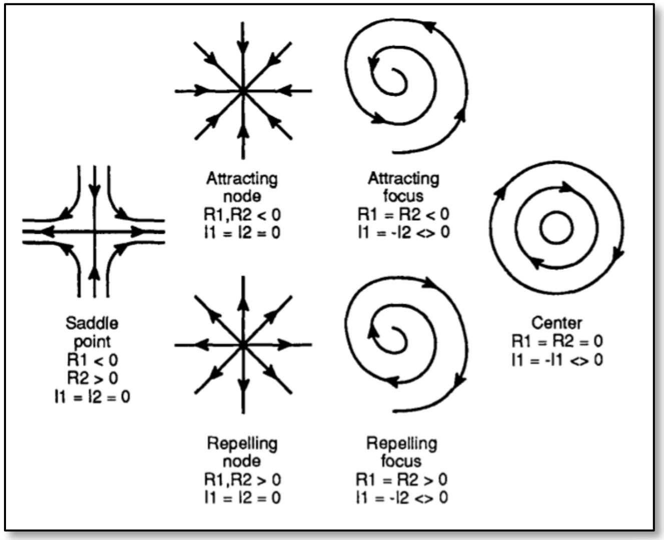
Output: Net long wave flux (thermal radiation)

Discrete Analysis of Vector Field Structures is Enabling New Robust Analysis of Climate Data

How many eddies in the ocean transport energy of a global climate model?

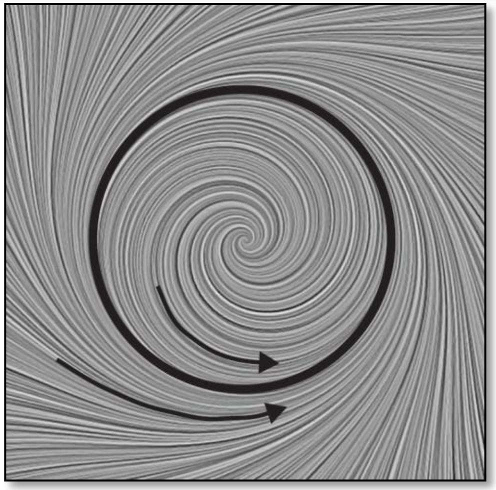


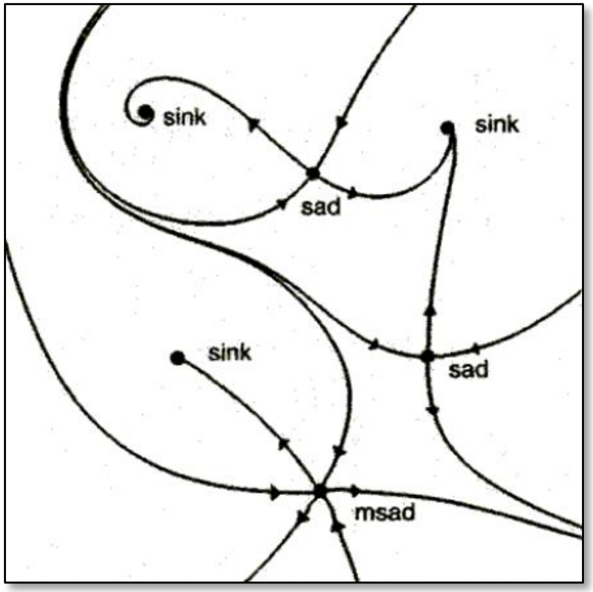
Encoding Must Explicitly Capture Flow's Topological Features



J. L. Helman and L. Hesselink. Representation and Display of Vector Field Topology in Fluid Flow Data Sets. IEEE Computer, 1989.

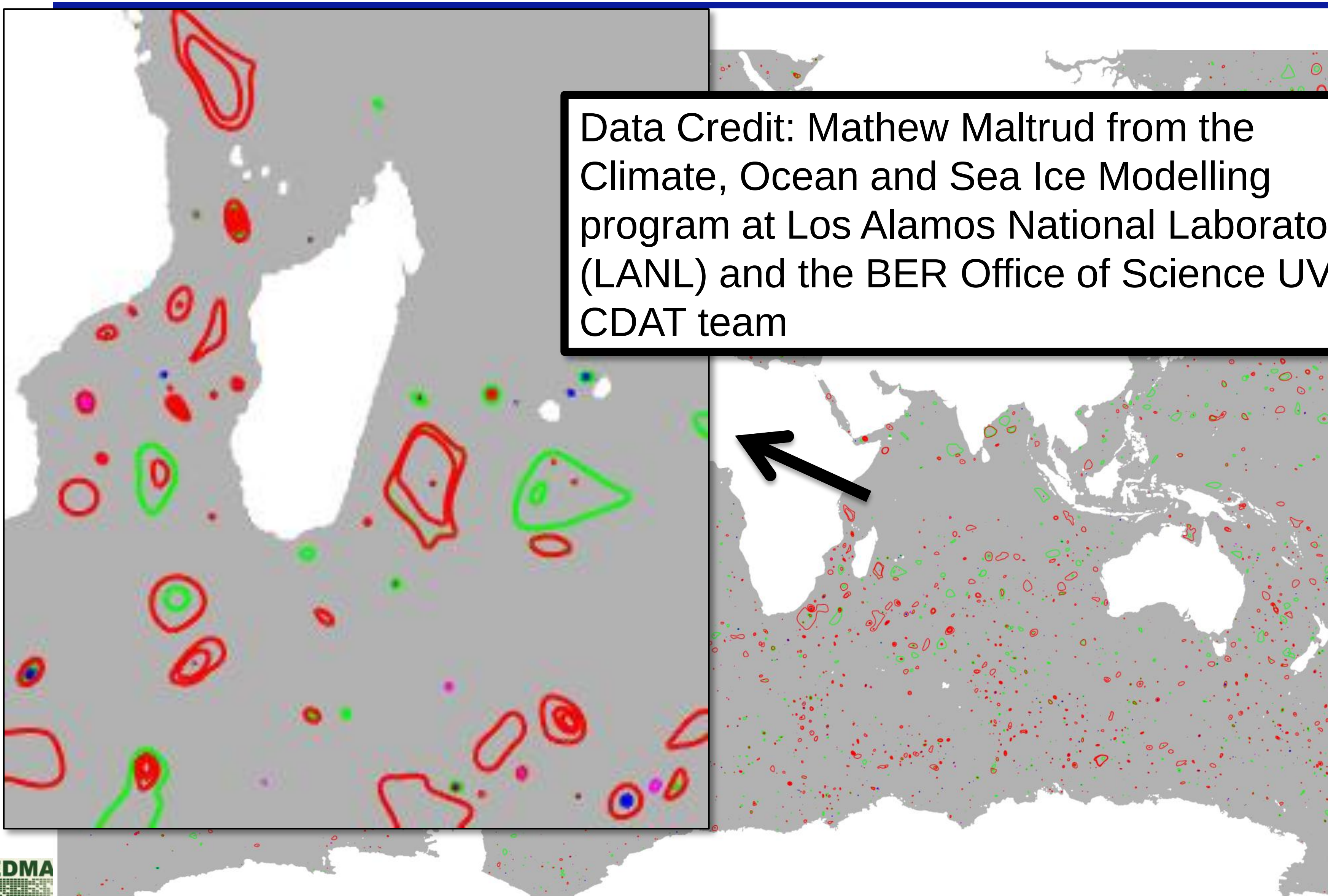
H. Theisel, T. Weinkauf, H.-C. Hege, H.-P. Seidel. Grid-independent Detection of Closed Stream Lines in 2D Vector Fields. VMV 2004.





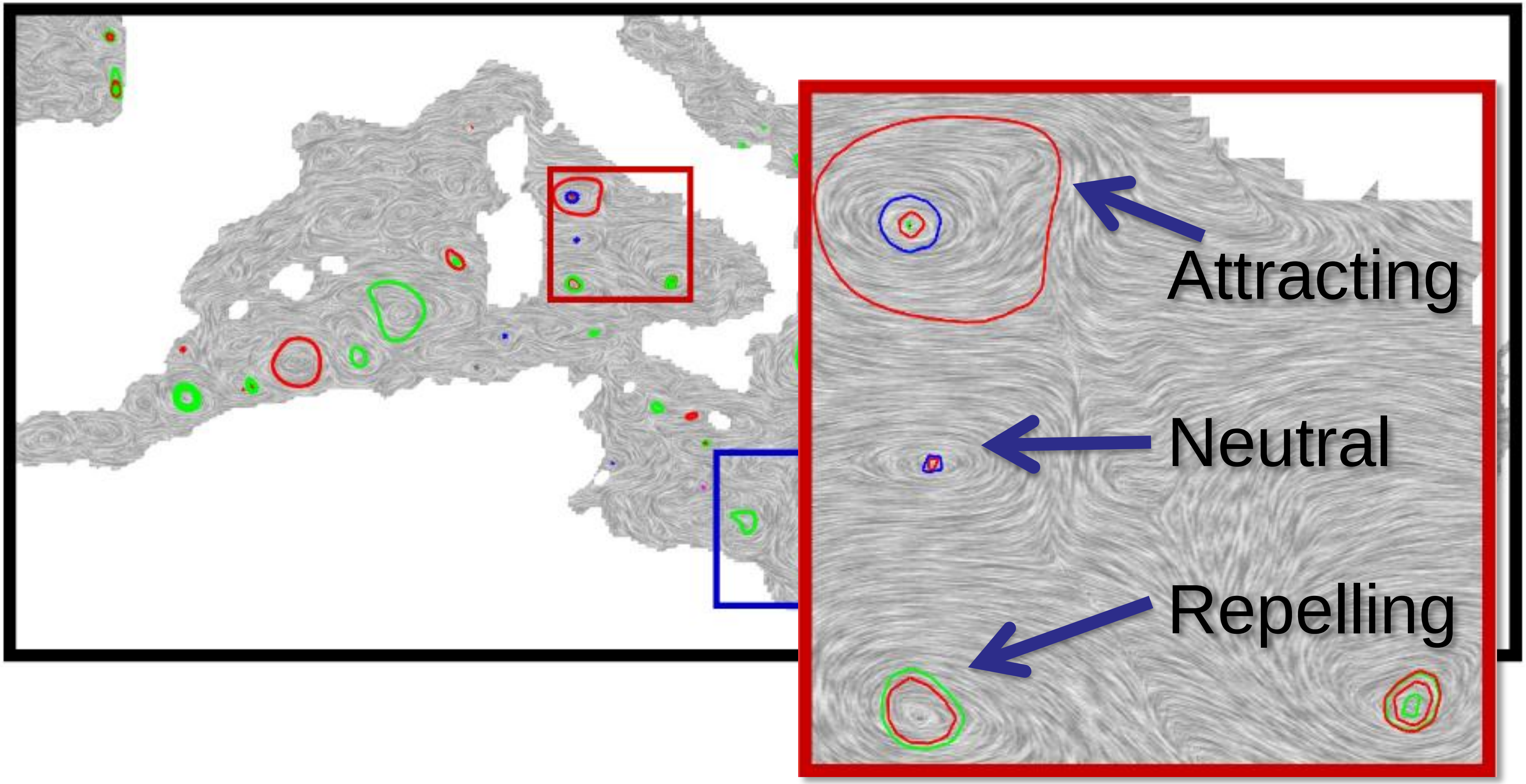
G. Scheuermann, X. Tricoche. Topological Methods in Flow Visualization. Visualization Handbook, 2004

Detecting Features Robustly Is Critical

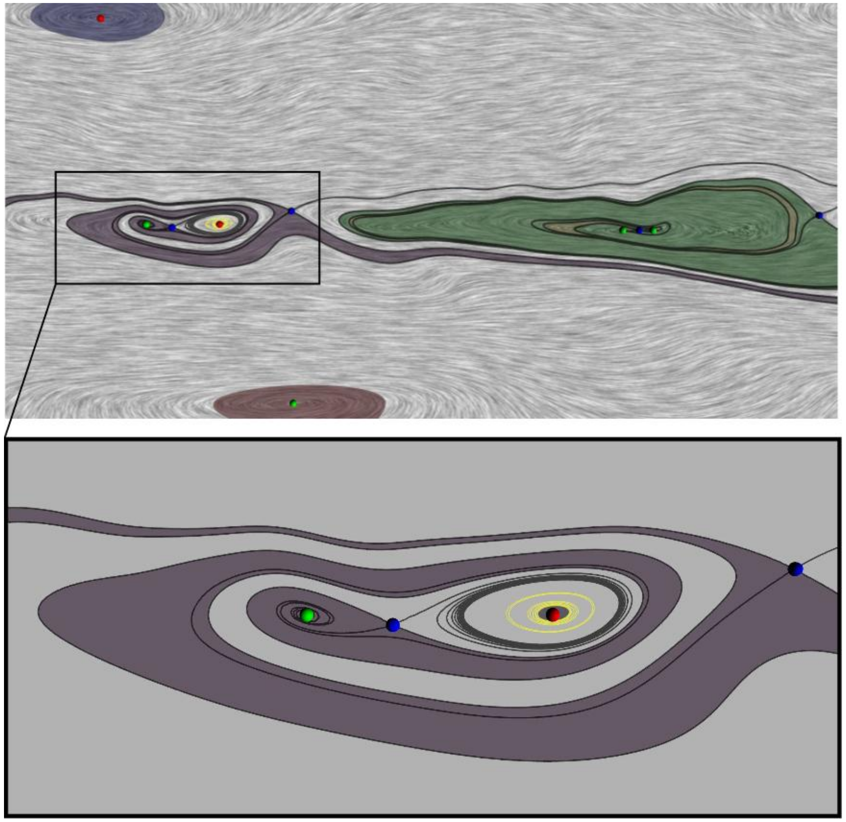


Data Credit: Mathew Maltrud from the Climate, Ocean and Sea Ice Modelling program at Los Alamos National Laboratory (LANL) and the BER Office of Science UV-CDAT team

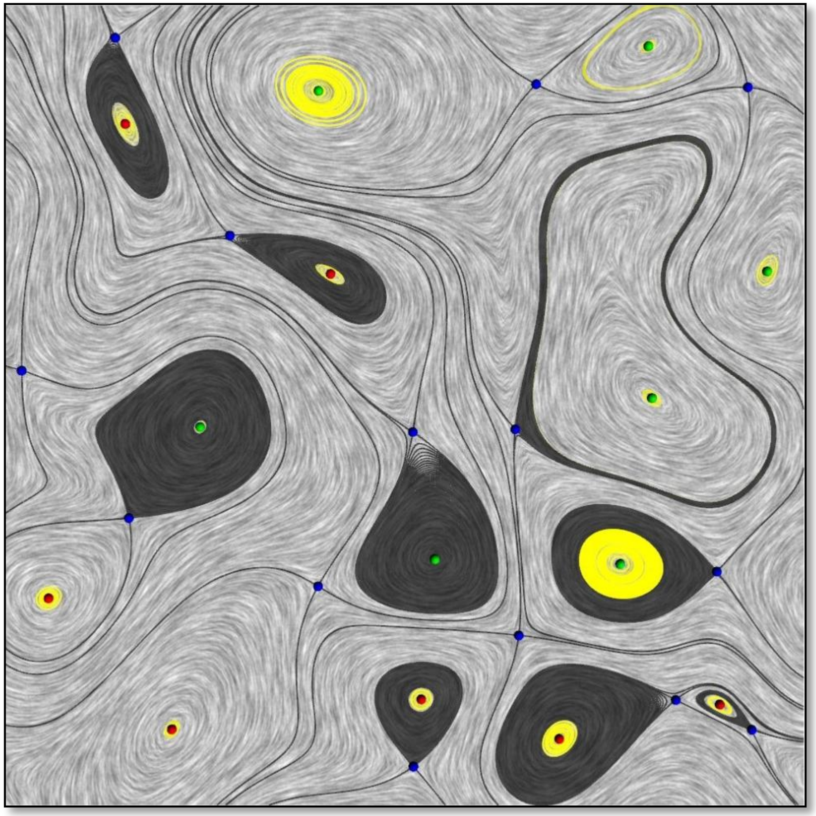
Classification of Closed Orbits



Quantized Features: Topological Decompositions



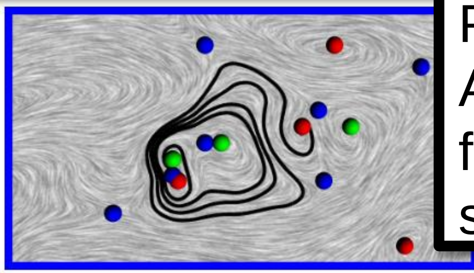
Stable Manifolds



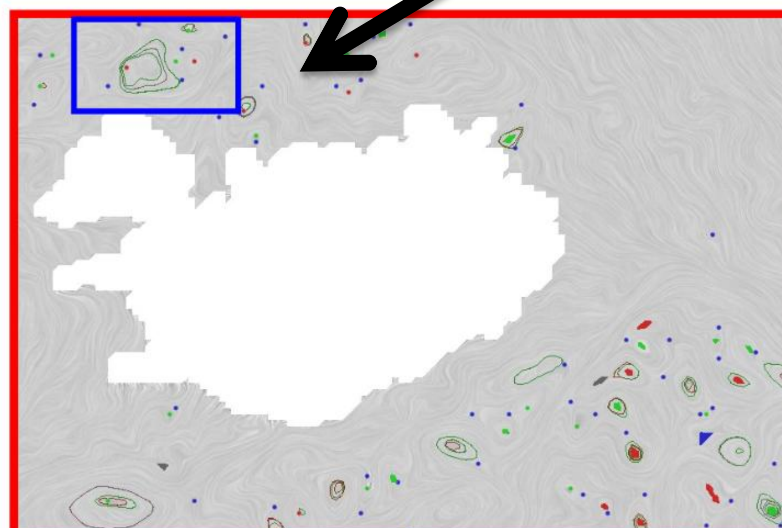
Topological Skeletons

Comparison to Other Discretizations

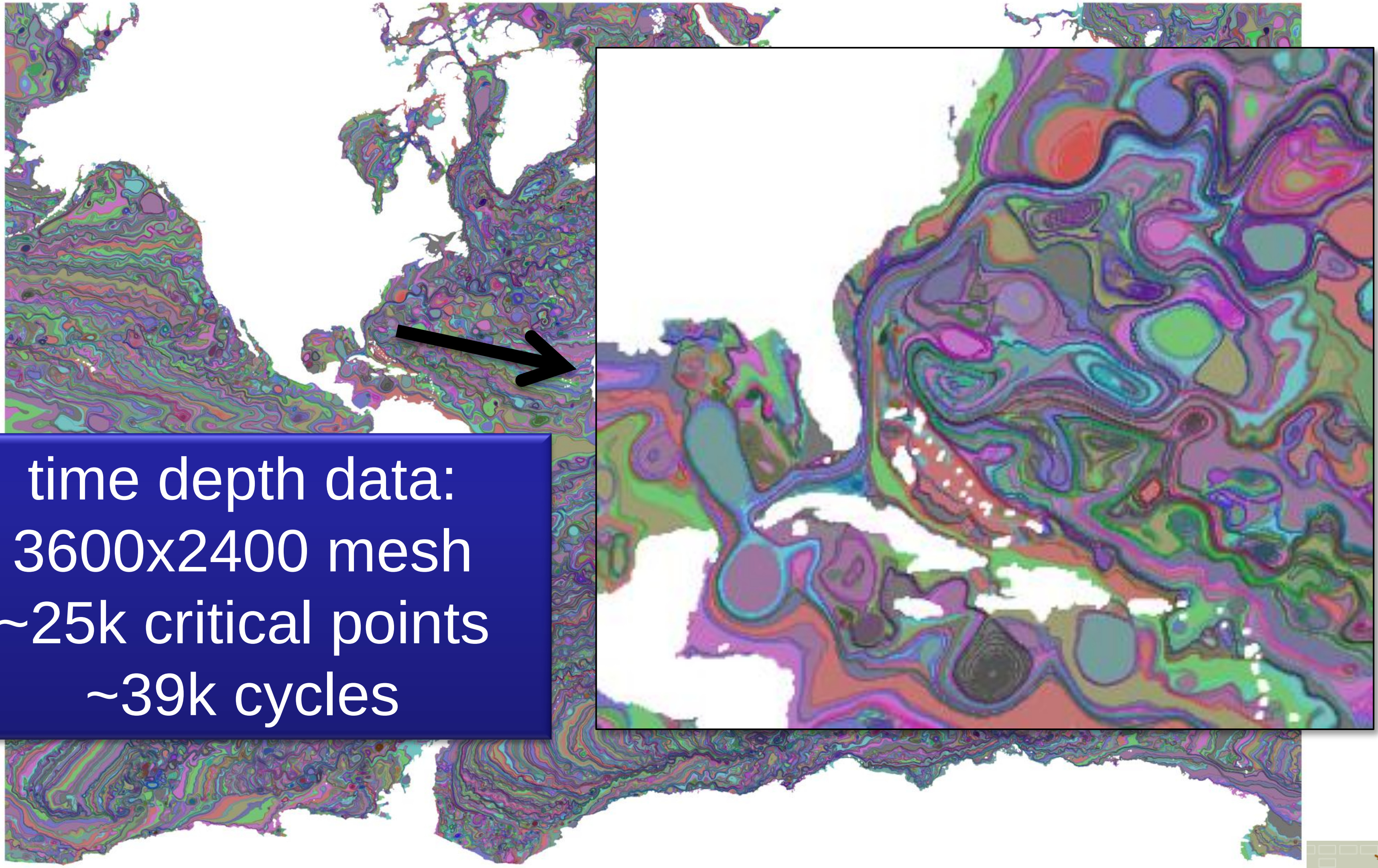
Morse Sets
G. Chen, Q. Deng, A. Szymczak, R. S. Laramee, E. Zhang. Morse set classification and hierarchical refinement using the Conley index. IEEE TVCG, 2011.



PC Morse Sets
A. Szymczak. Stable Morse Decompositions for piecewise constant vector fields on surfaces. CGF 2011.



Stable Manifolds of Ocean Data



time depth data:
3600x2400 mesh
~25k critical points
~39k cycles

CG Enabled Fundamental Advances in Scientific Data Analysis and Visualization

- **Tight cycle of :**
  - **basic research,**
  - **software deployment**
  - **user support**

- **Plenty of Open Problems:**
  - **<u>Combinatorial Methods</u> for:**
    - **Vector Fields**
    - **Tensor Fields**
    - **Dynamic models**
  - **<u>Efficient</u> approaches for high dimensional data**
  - **<u>Simple</u> data structure and algorithms (fast in practice and easier to validate)**