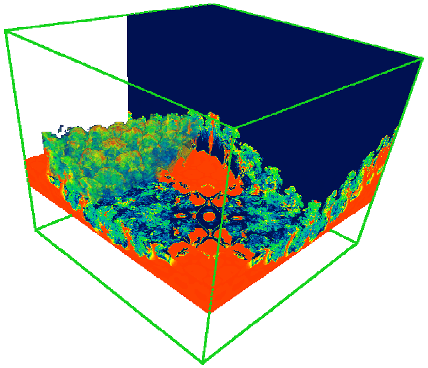
Robust Topology-Based Analysis of Large Scale Data

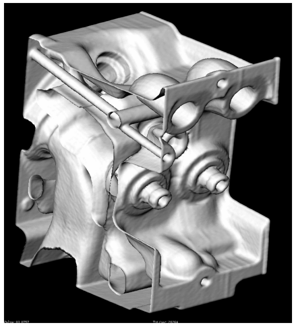
Valerio Pascucci

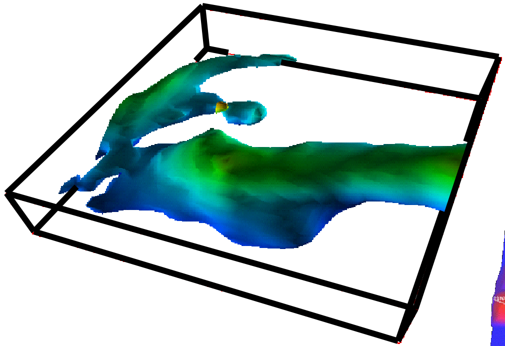
CASC, LLNL

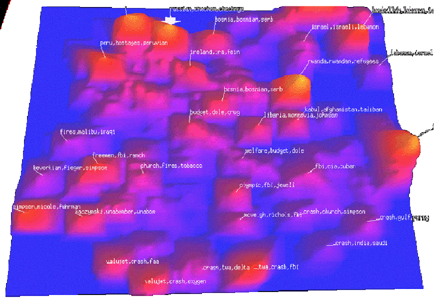
CS, UCDAVIS

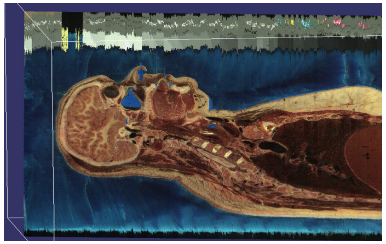
We develop tools for efficient and reliable data exploration and understanding.



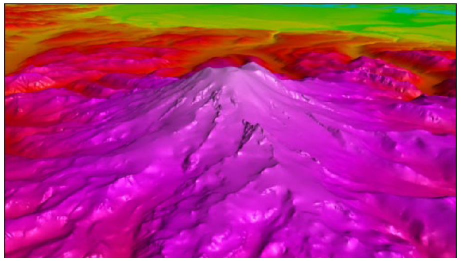


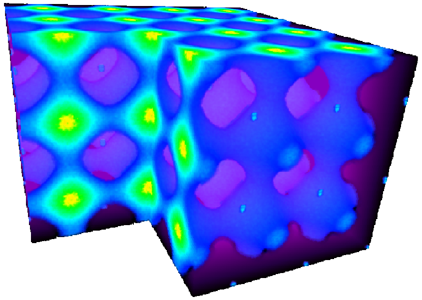


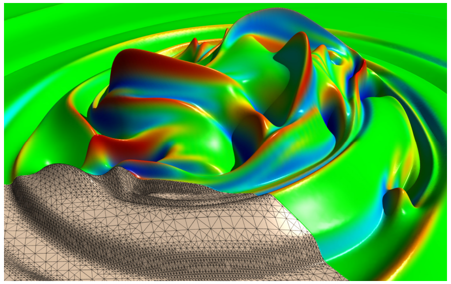


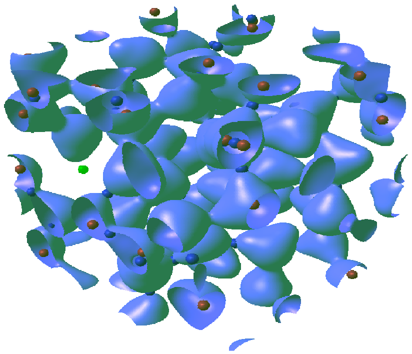


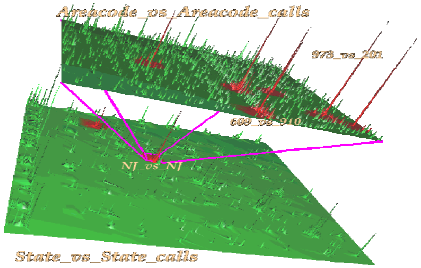
Real functions are ubiquitous in the representation of scientific information.





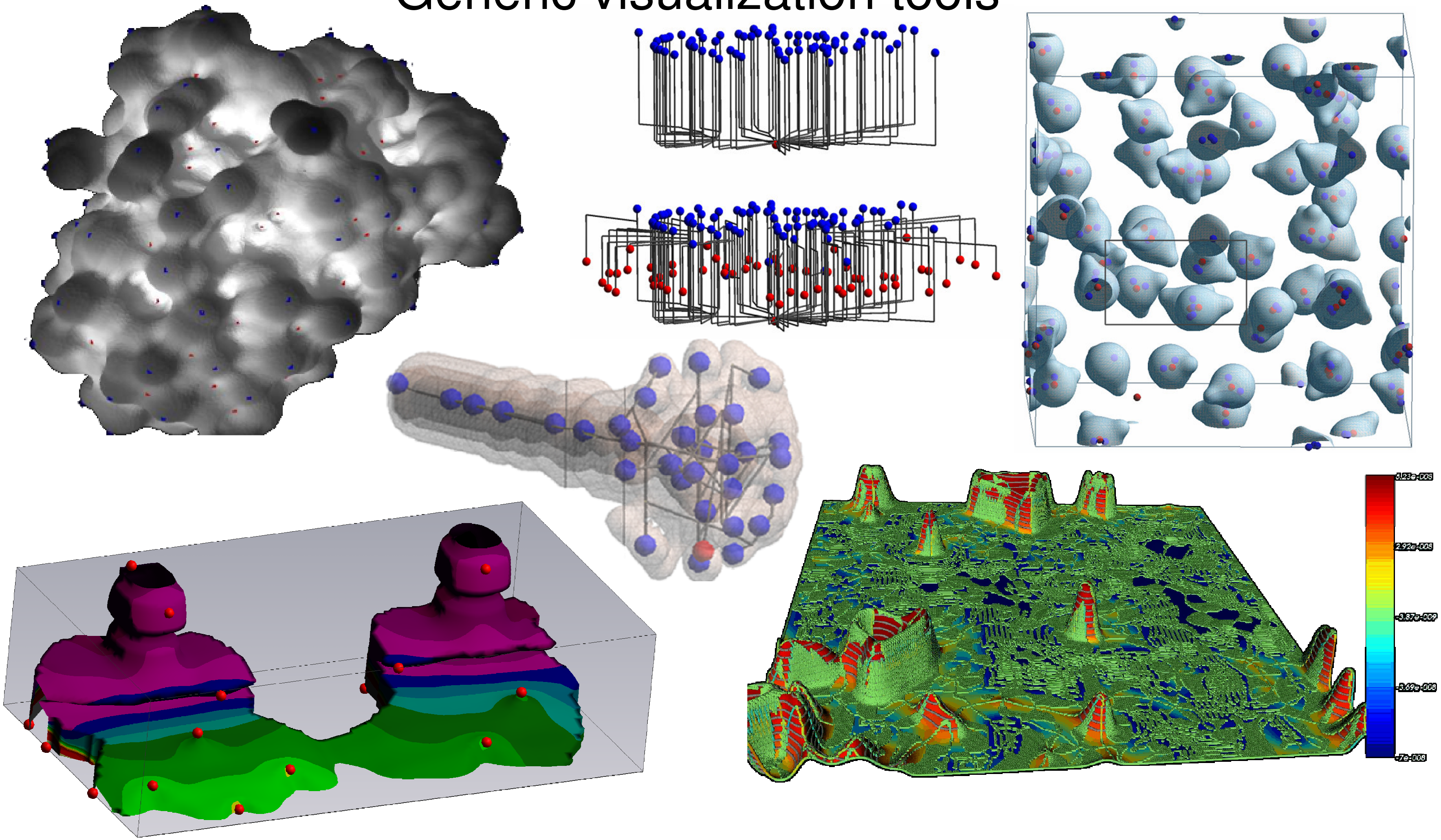






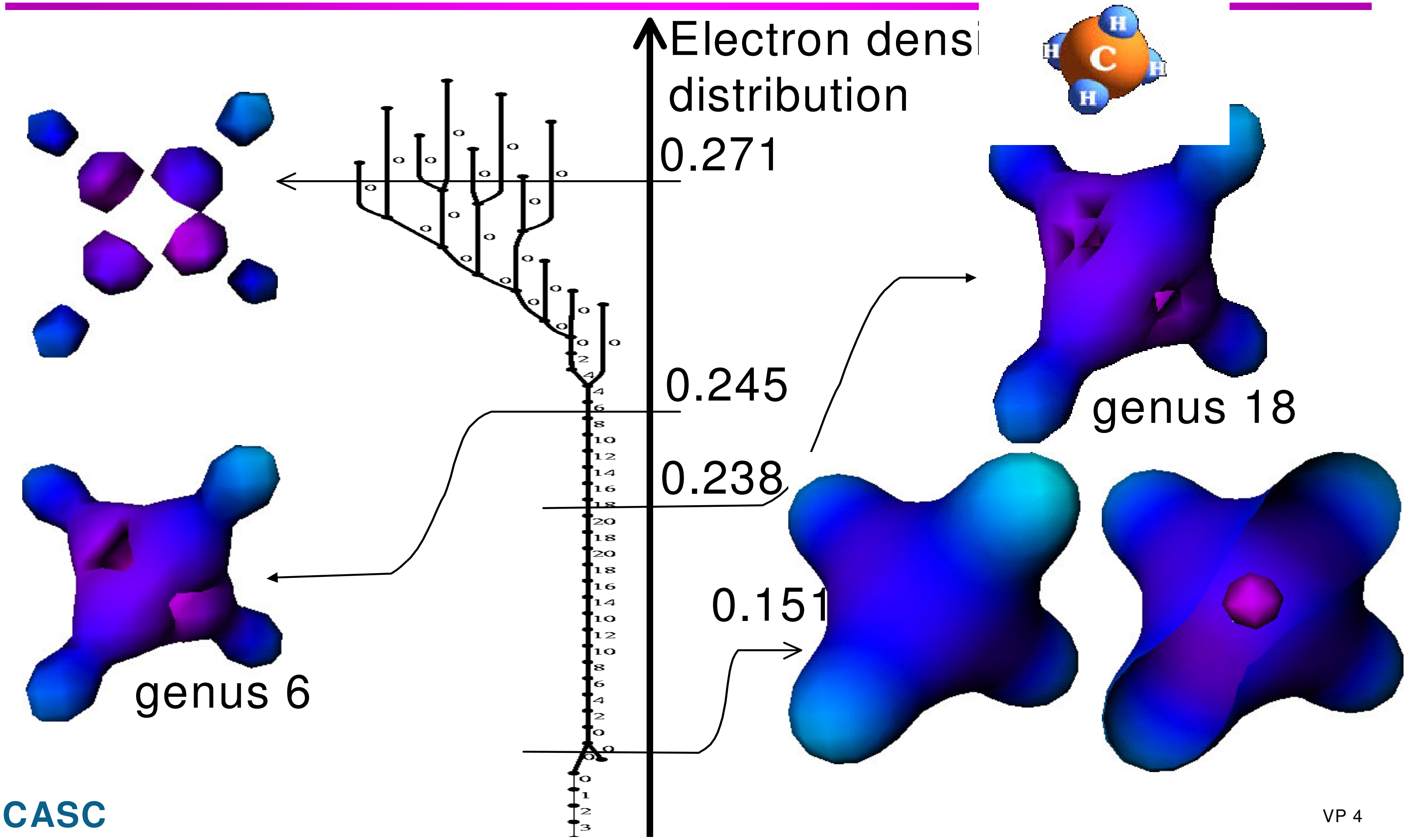
We aim at a robust framework for data visualization, analysis and illustration.

Generic visualization tools



We aim at a robust framework for data visualization, analysis and illustration.

Illustration

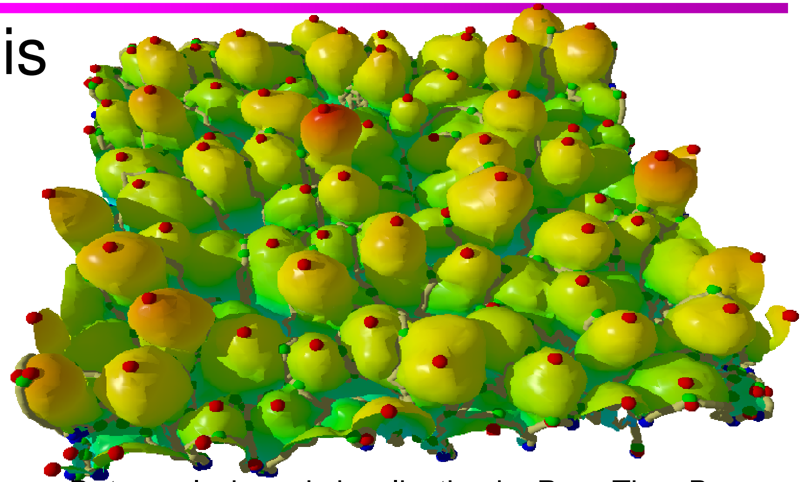


CASC
VP 4

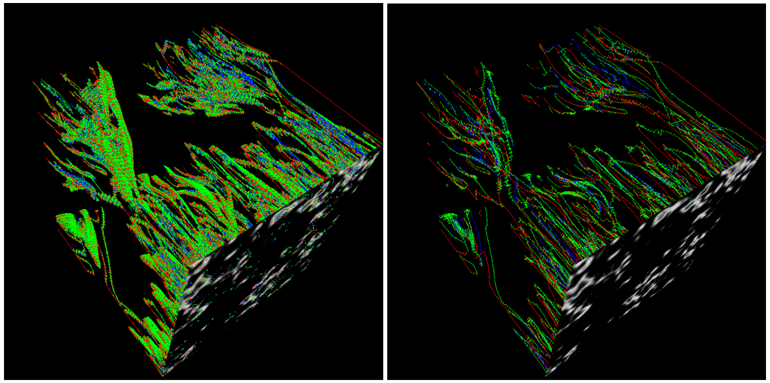
We aim at a robust framework for data visualization, analysis and illustration.

Data Analysis

Bubbles and spikes in turbulent mixing of Rayleigh-Taylor instability. →



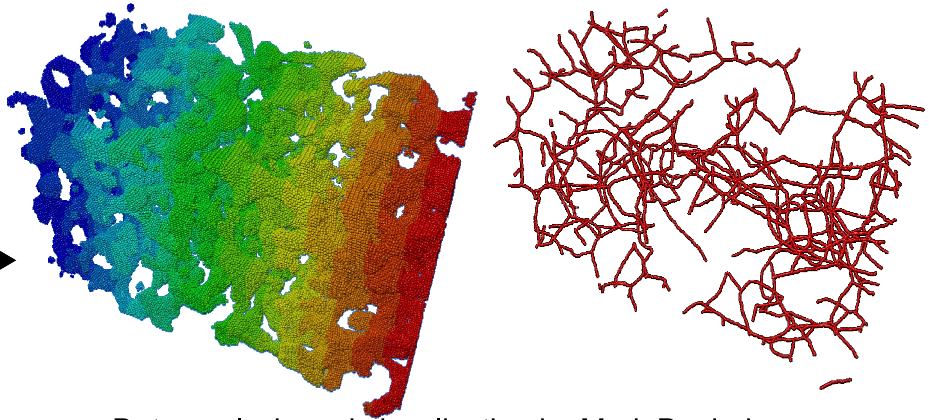
Data analysis and visualization by Peer-Timo Bremer



Data analysis and visualization by Ajith Mascarenhas

← Multi-scale time tracking of combustion particles.

Topological analysis of the channel structures in porous media. →



Data analysis and visualization by Mark Duchaineau

We develop a theory for a framework of “practically good” topological algorithms.

- **Use fundamental characterizations of the representation model:**

  e.g. characterize the behavior of real functions using Morse Theory.

- **Intrinsically robust computations:**

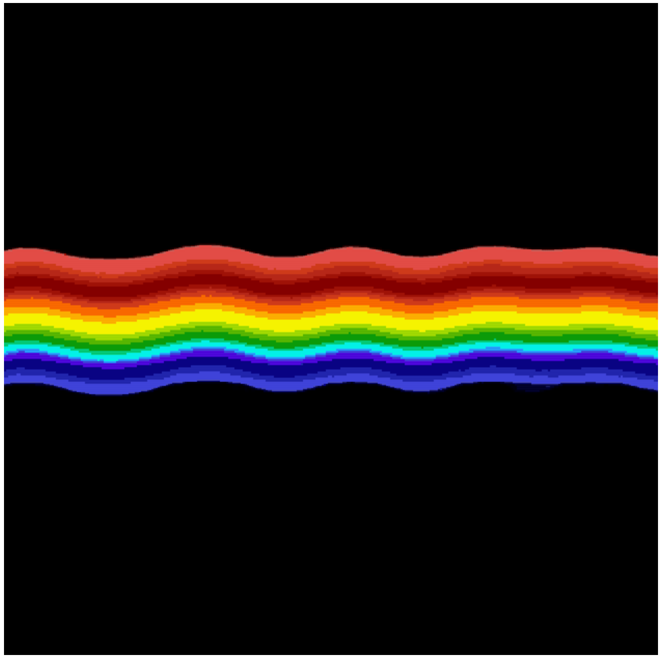
  theory developed on computer based representations.

- **Comprehensive analysis:**

  guaranteed extraction of all the features present in the data

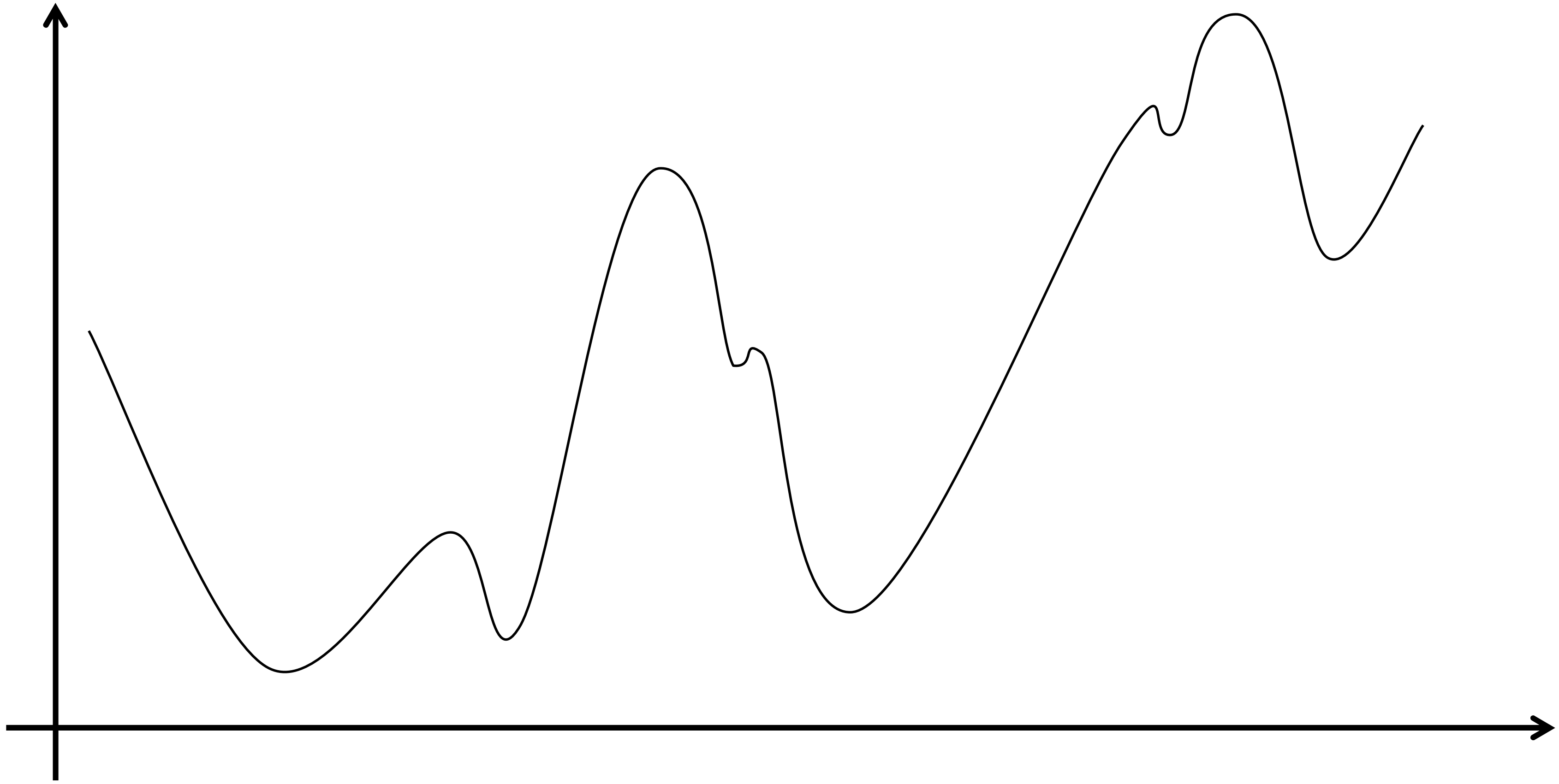
- **Multi-scale representation model:**

  flexible and scalable data analysis and exploration



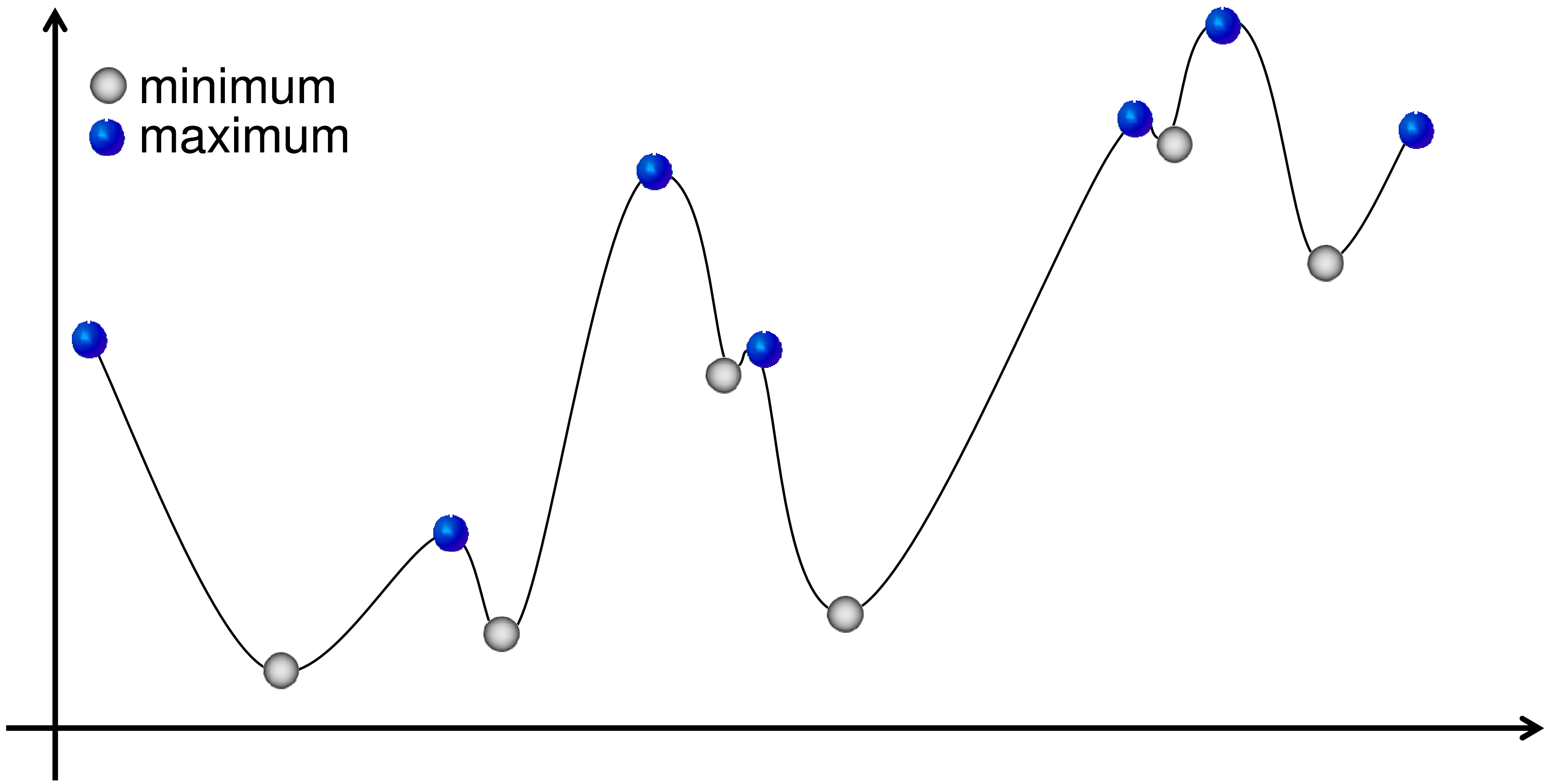
We take a formal approach to the analysis of real functions based on Morse theory.

- **In 1D this is the simple and natural approach for studying the trends in a real function.**



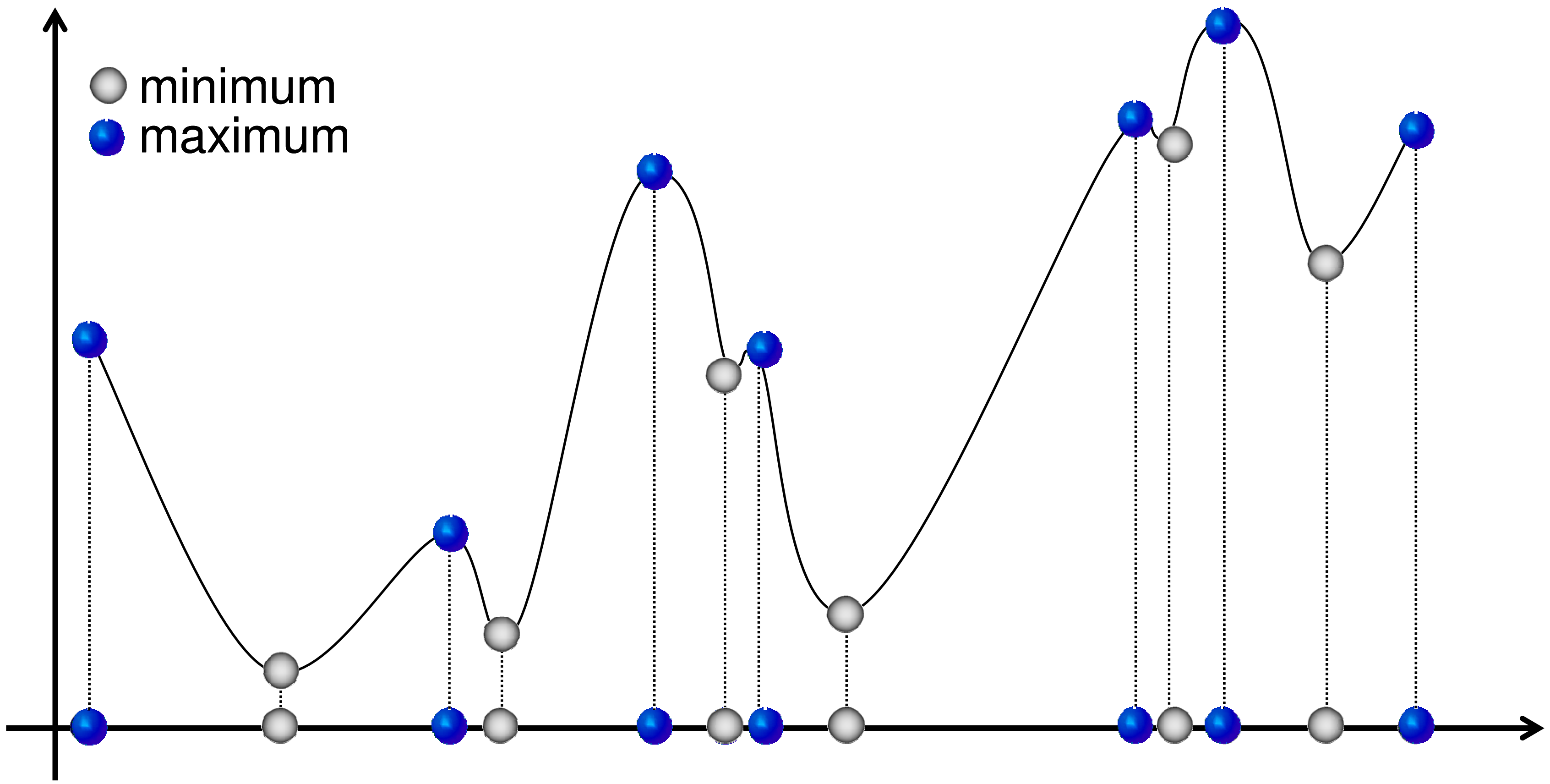
We take a formal approach to the analysis of real functions based on Morse theory.

- **Identify critical points (zero derivative).**



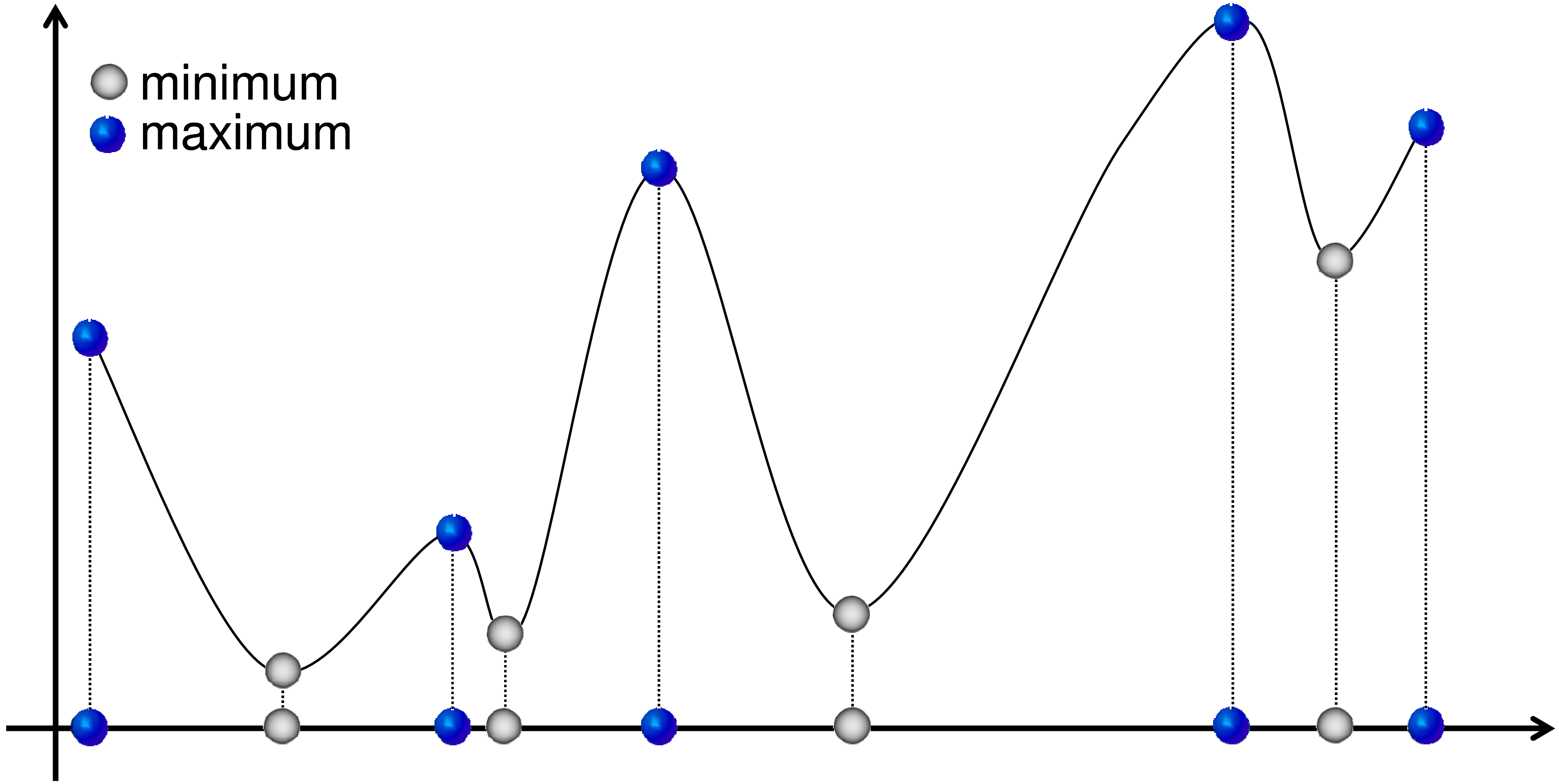
We take a formal approach to the analysis of real functions based on Morse theory.

- **Partition the domain into monotonic regions.**



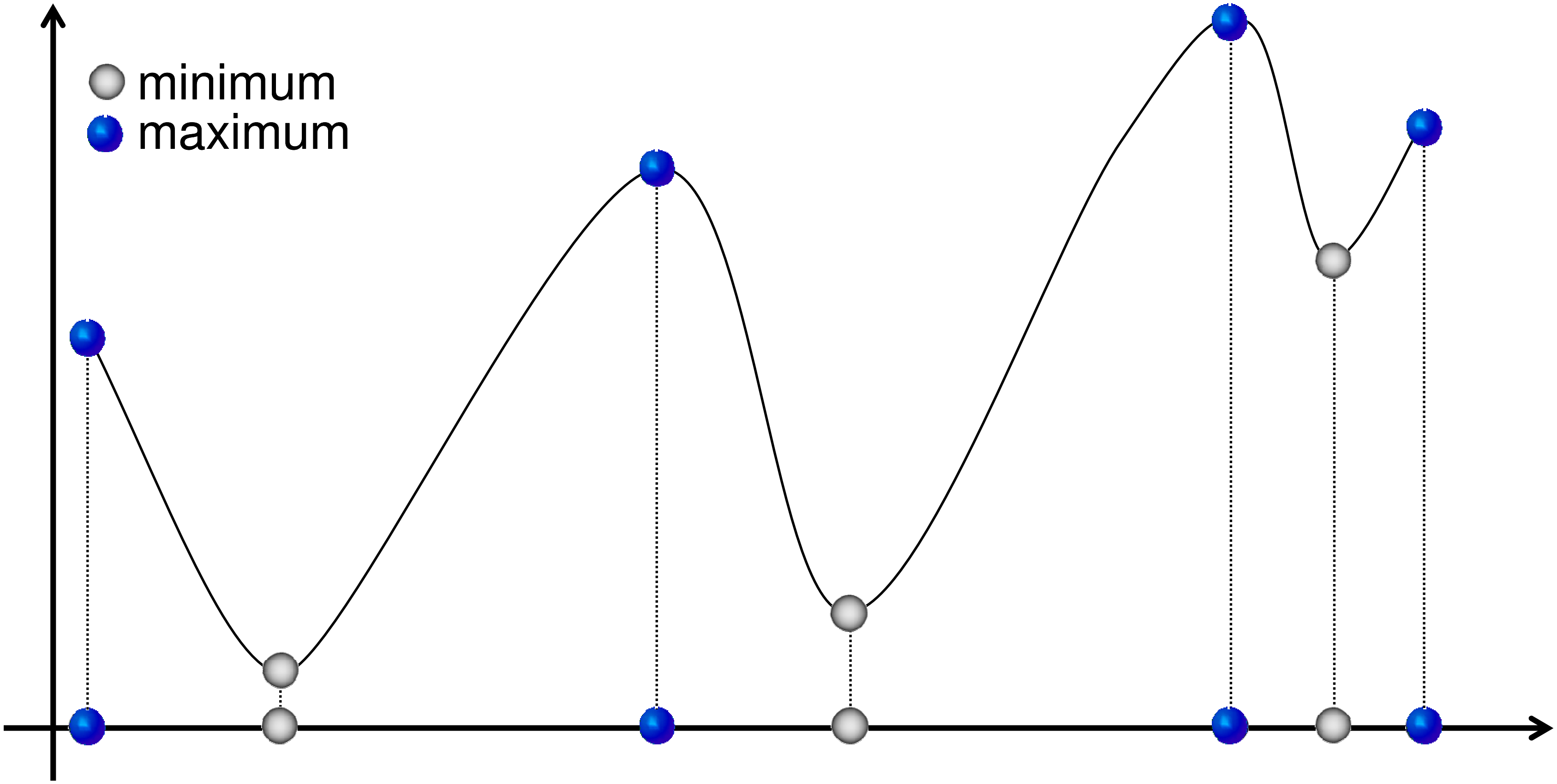
We take a formal approach to the analysis of real functions based on Morse theory.

- **Remove small features that are irrelevant (noise).**



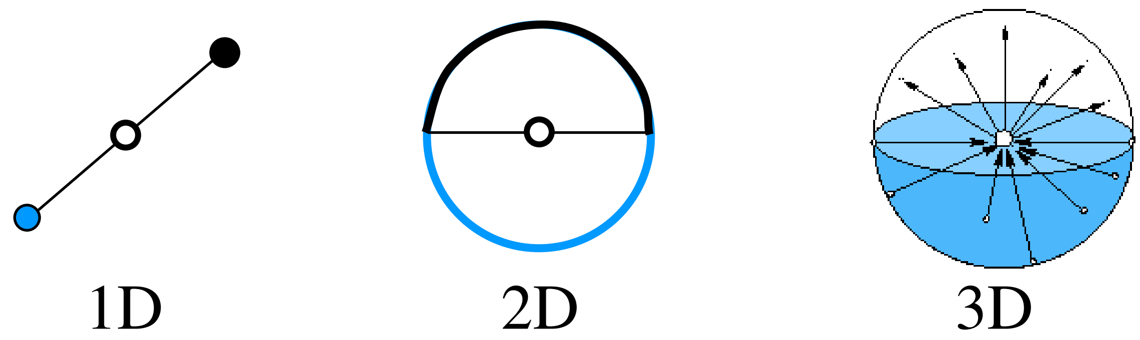
We take a formal approach to the analysis of real functions based on Morse theory.

- **Simplify features that are below the "scale" of interest.**



We define regular points based on purely combinatorial properties.

	$f(x): D \rightarrow \Re$	$F(x): S \rightarrow \Re$
	Classical definitions	Simulation of Differentiability
	D smooth manifold	S simplicial complex
	f infinitely differentiable	$F(x)$ PL-extension of $f(x_i)$
regular point	$\nabla f(x) \neq 0$	$LowerLink(x) = B^{n-1}$



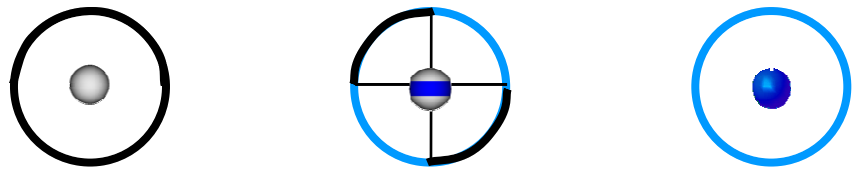
Simple and isolated critical points are classified with the Morse lemma.

$$\nabla f(p) = 0 \quad \Rightarrow \quad f(x)\Big|_p = f(p) + \sum_{i=1}^{n-k} x_i^2 - \sum_{i=n-k+1}^{n} x_i^2$$

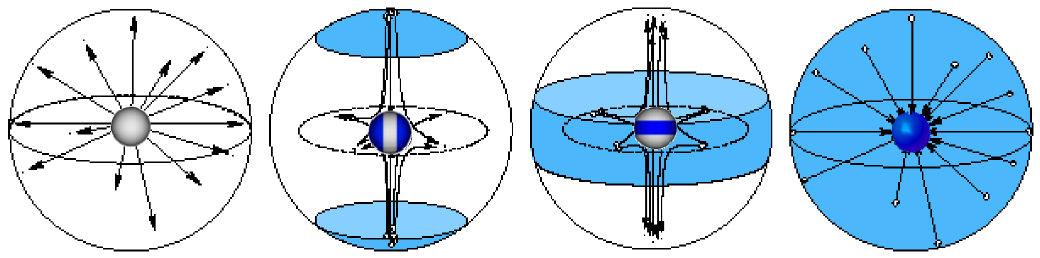
There are *n+1* types of critical points

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

Minimum   Maximum



Minimum   Saddle   Maximum

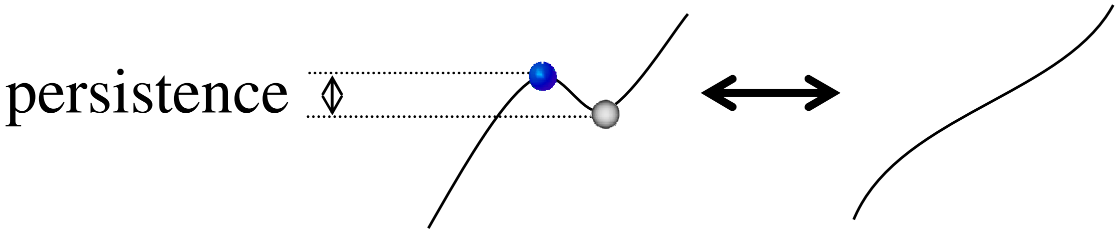


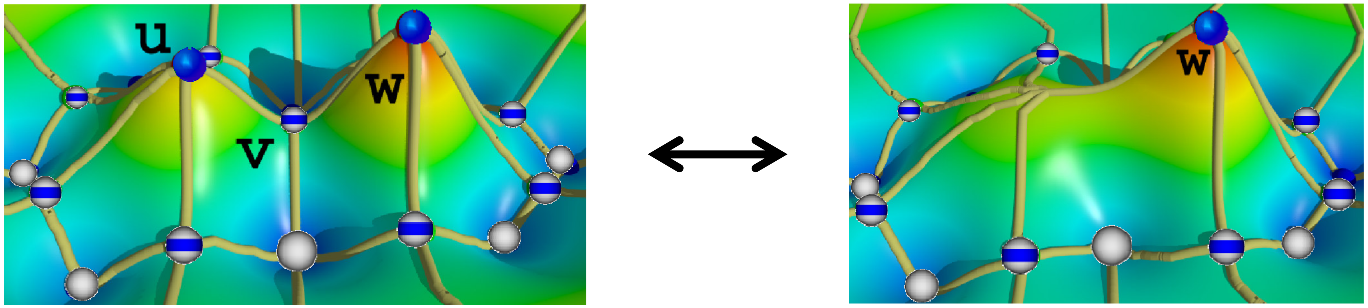
Minimum   1-saddle   2-saddle   Maximum

Simple and isolated critical points can be simplified based on the index lemma.

Index Lemma

Critical points can be created or destroyed in pairs with index that differs by one.





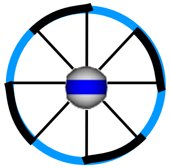
Minimum 1-saddle 2-saddle Maximum

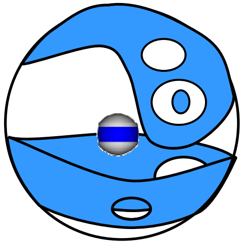
It is easy to attain isolated critical points but hard to keep them non-degenerated.

$$F(v_i) < F(v_j) \quad \Leftrightarrow \quad f(v_i) < f(v_j) \quad or \quad i < j$$

$$\cancel{F(v_i) = F(v_j)}$$

$$\begin{vmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{vmatrix} \neq 0$$





Computationally the approach cannot be extend to dimension higher than four.

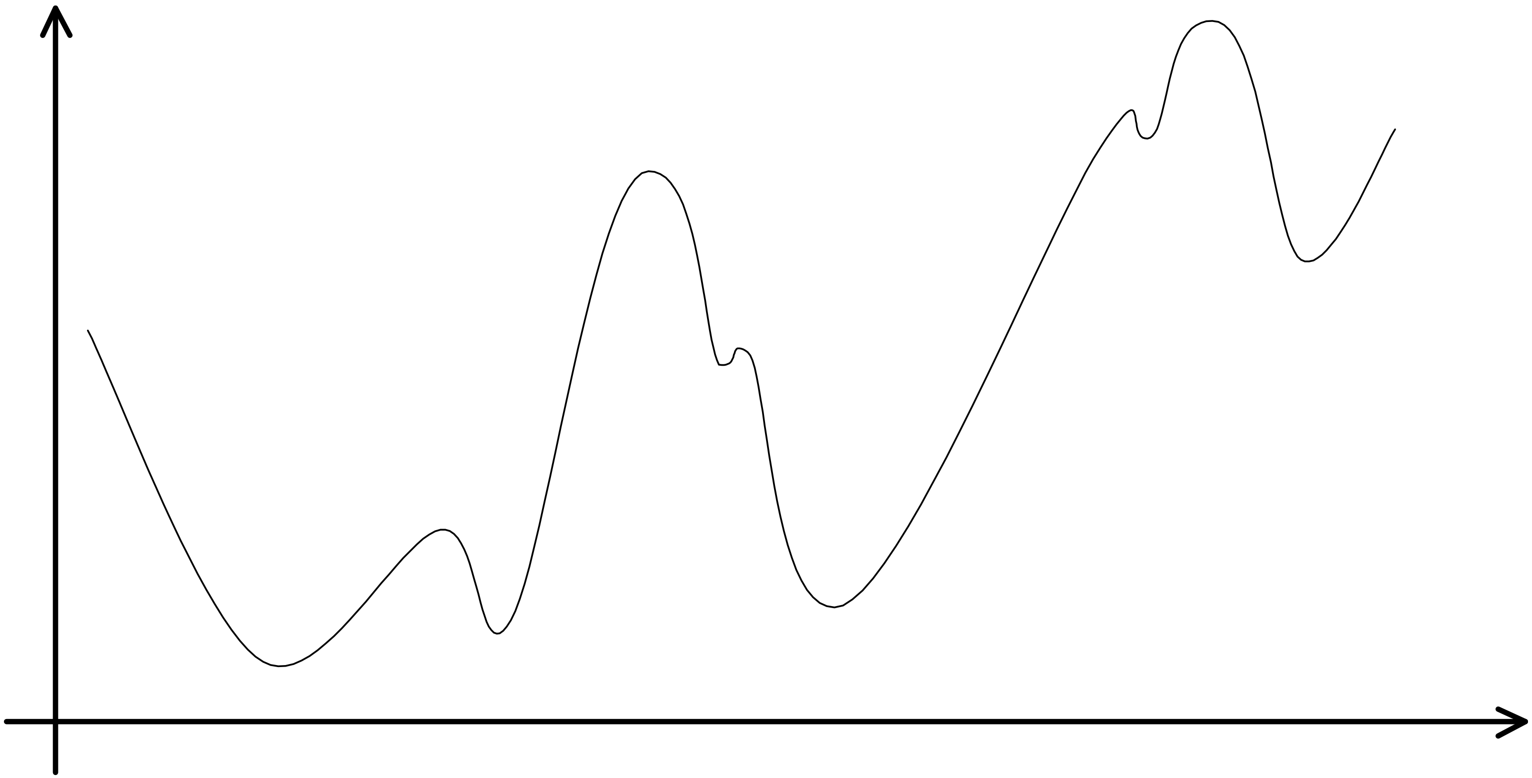
Insolubility of the Problem of Homeomorphism.

A. A. Markov. 1958

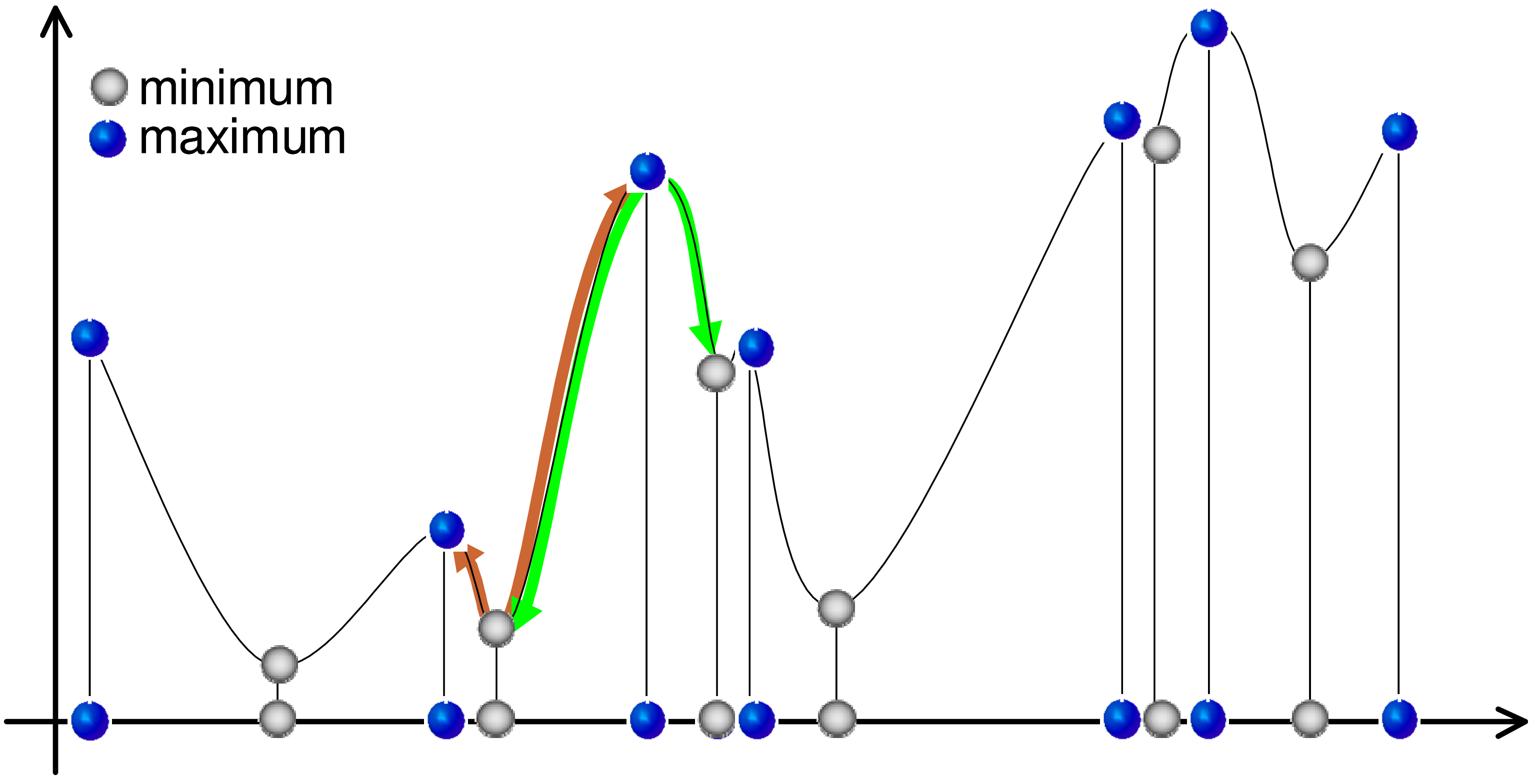
(translation by Afra Zomorodian).

For every natural number $n > 3$,
one can create an n-manifold M^n,
such that the problem of homeomorphism
of manifolds to M^n is undecidable.

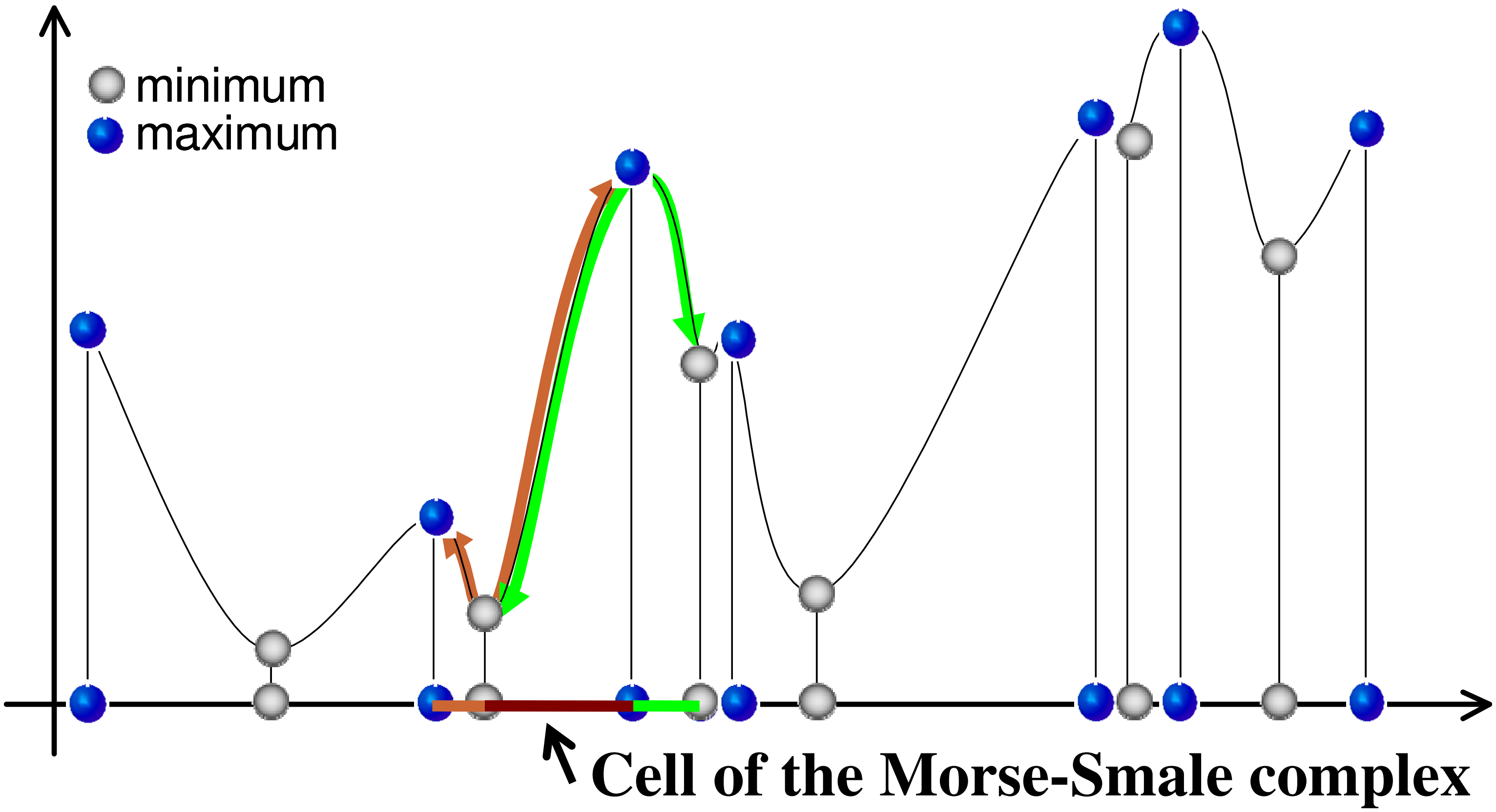
We use ascending/descending manifolds to characterize the cells of the complex .



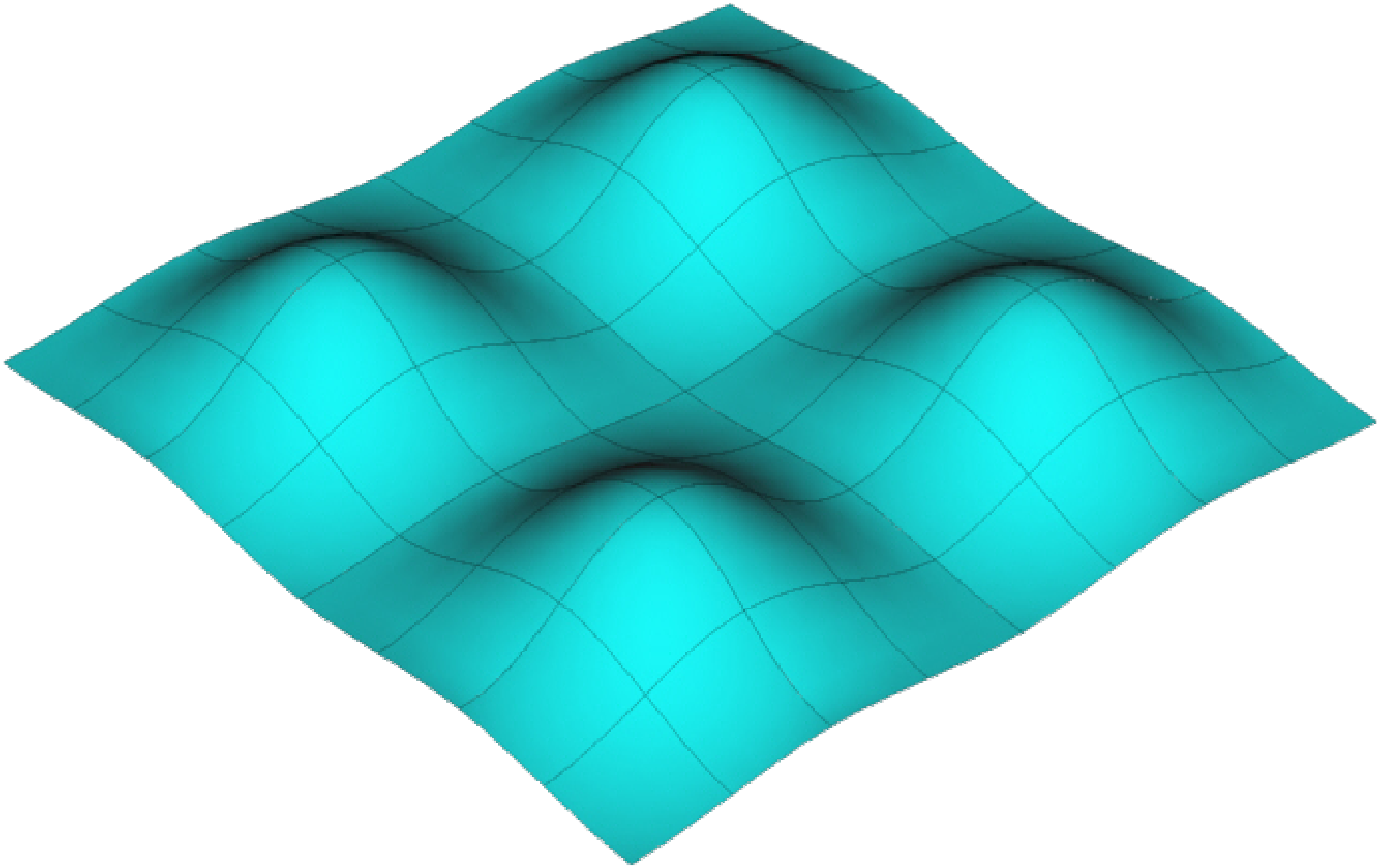
We use ascending/descending manifolds to characterize the cells of the complex .



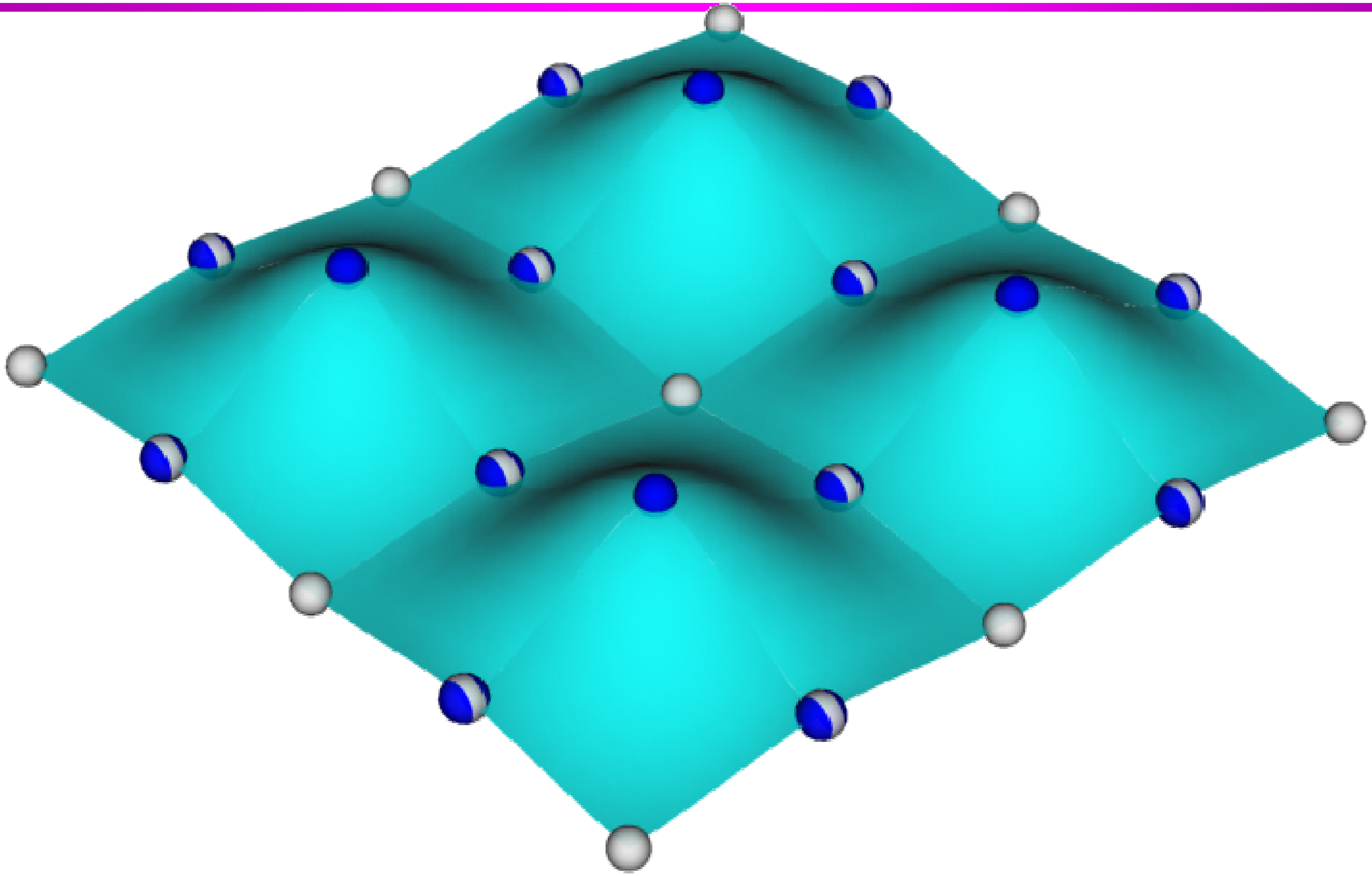
We use ascending/descending manifolds to characterize the cells of the complex .



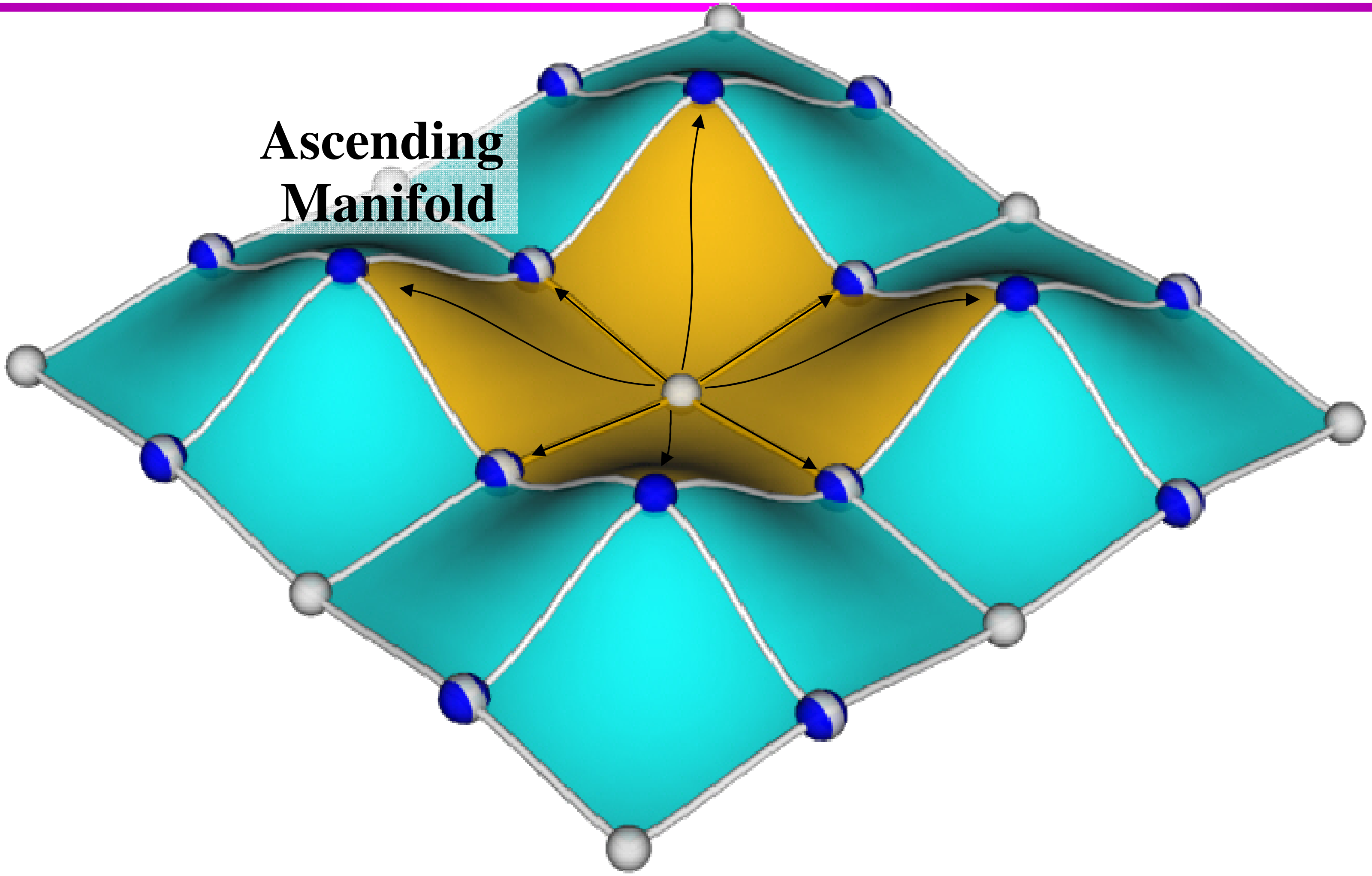
Consider a simple 2D function.



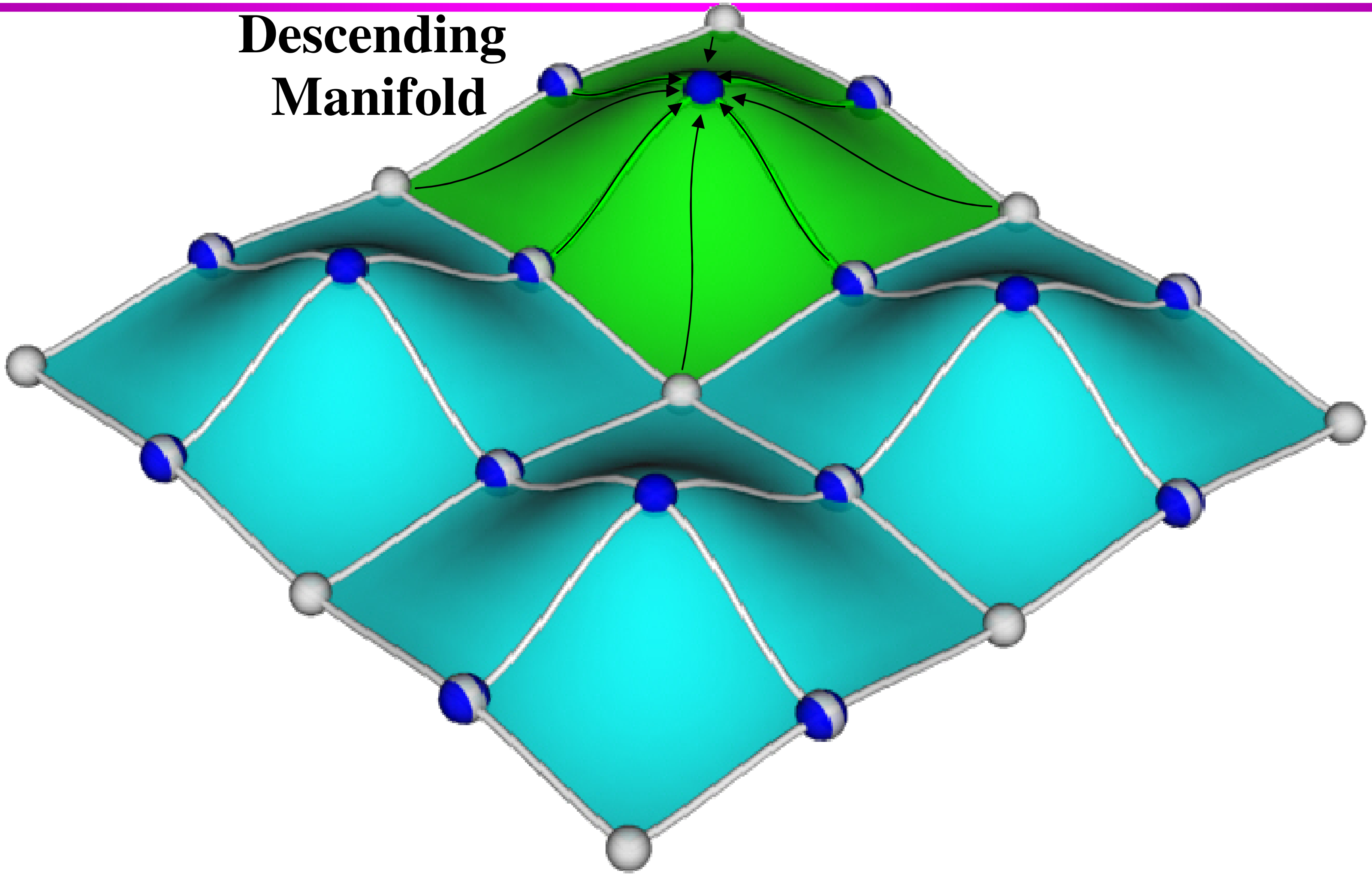
We use ascending/descending manifolds to characterize the cells of the complex.



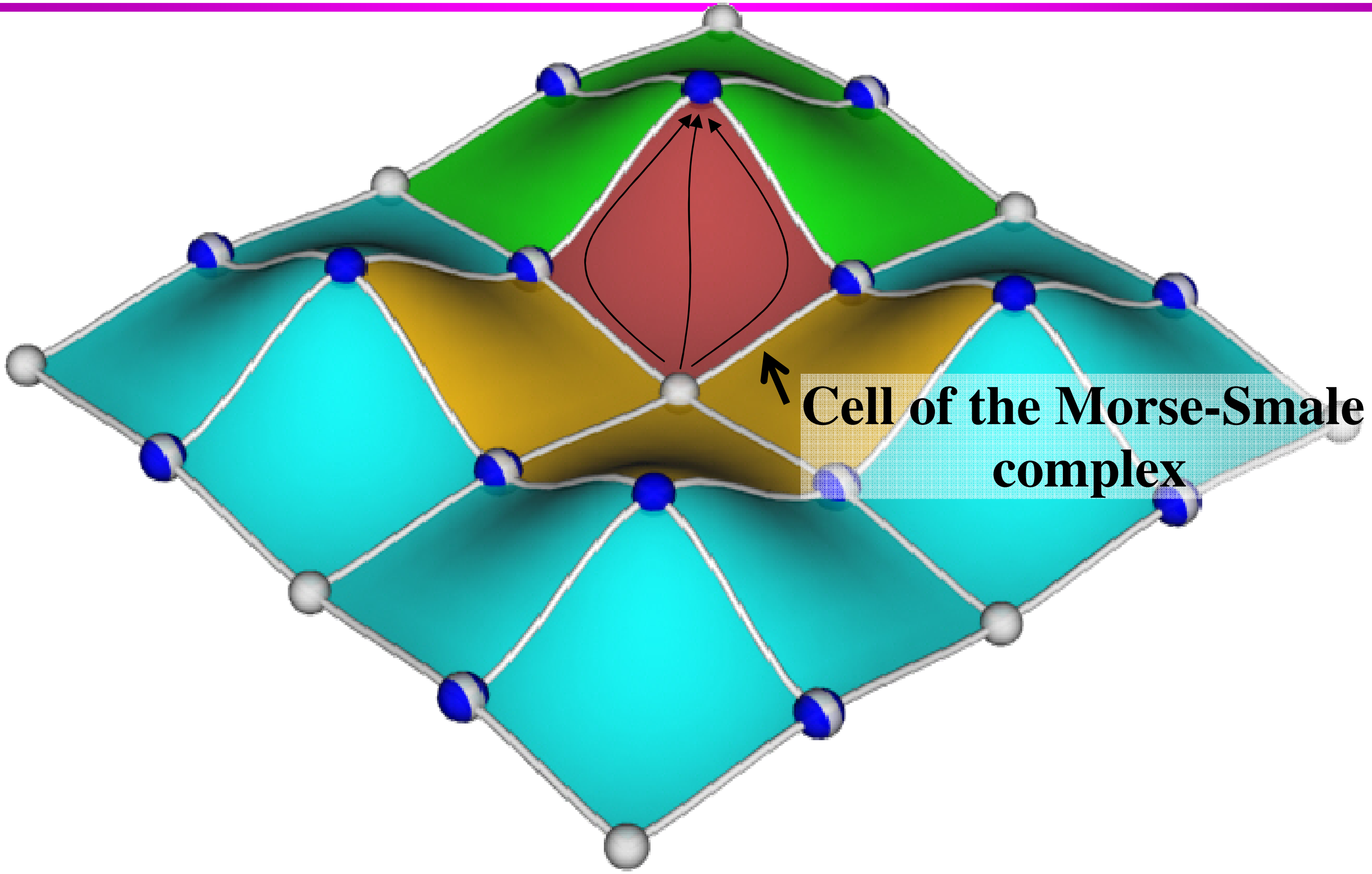
We use ascending/descending manifolds to characterize the cells of the complex.



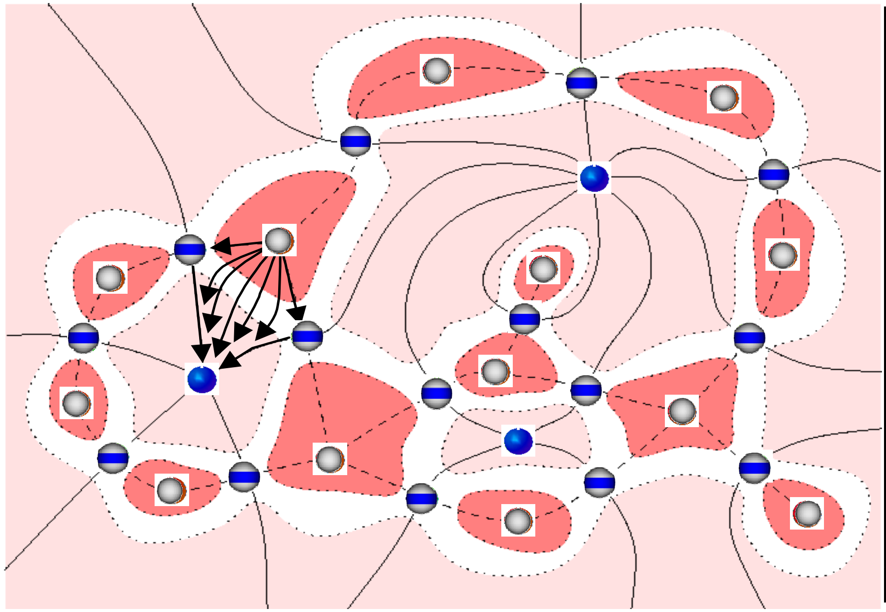
We use ascending/descending manifolds to characterize the cells of the complex.

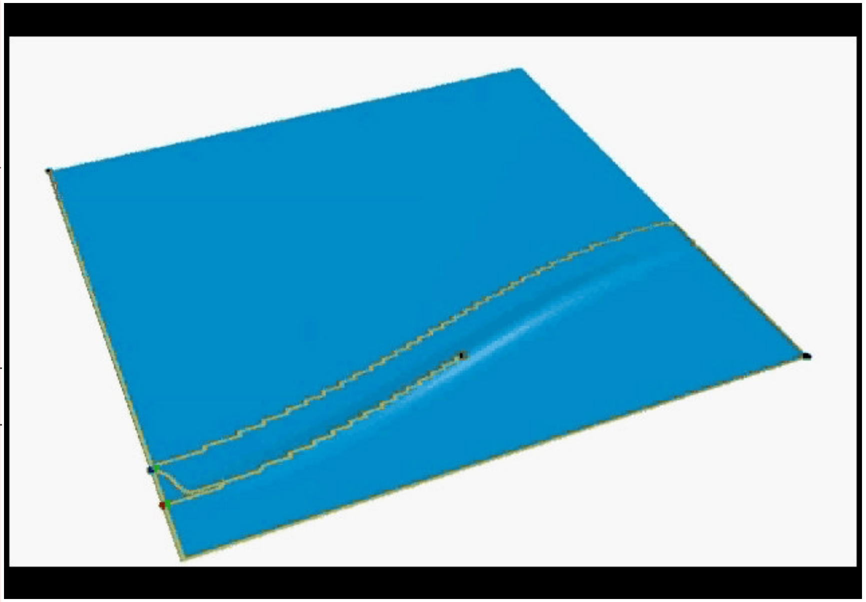


We use ascending/descending manifolds to characterize the cells of the complex.



The Morse-Smale complex decomposes 2-manifolds of any genus into quads.

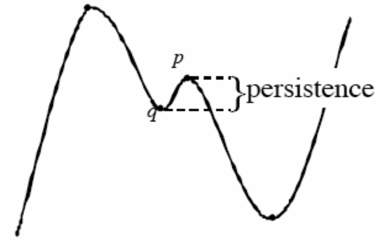




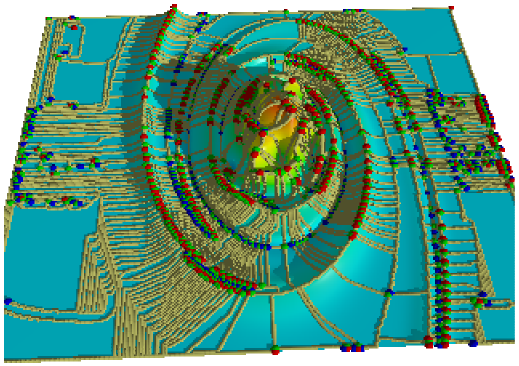
A *d*-cell contains gradient lines connecting critical points that differ in index by *d*.

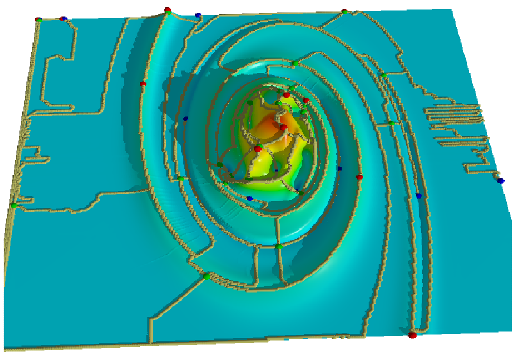
Morse-theoretical approaches have a natural multi-scale structure.

1D

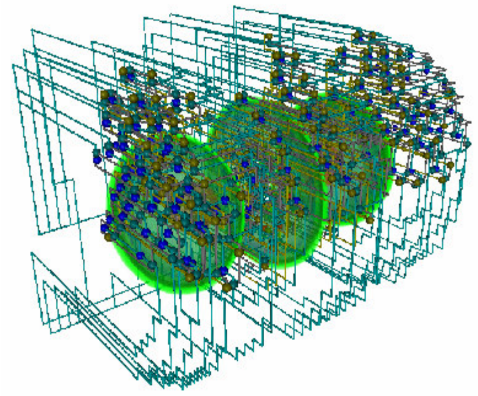


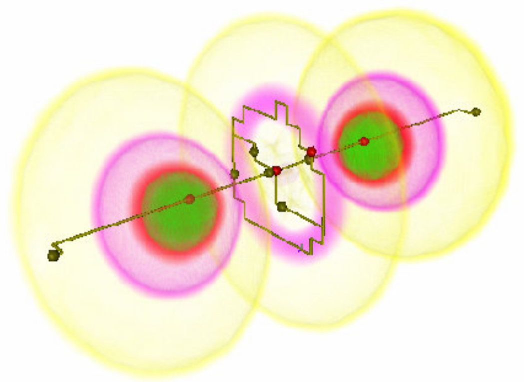
2D

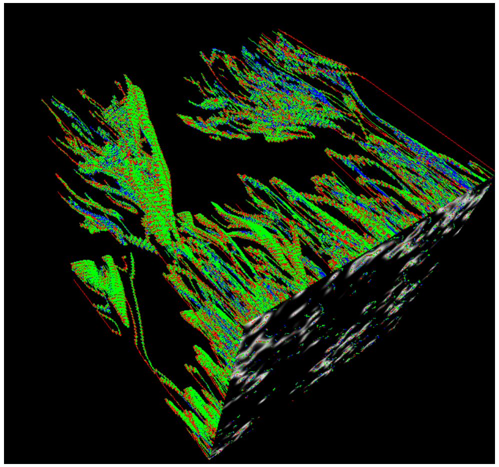




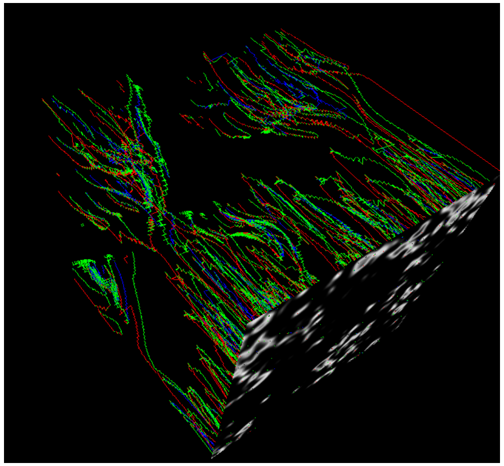
3D





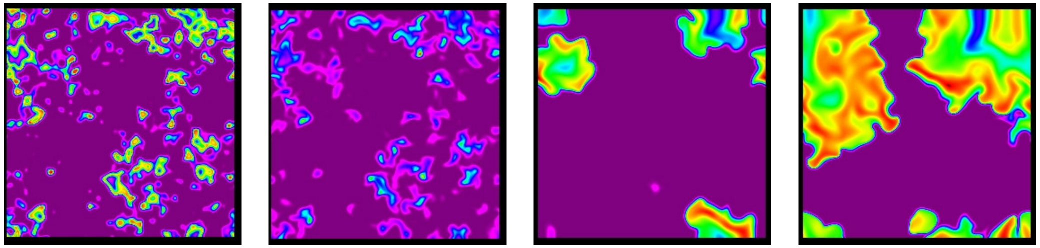


time ↓



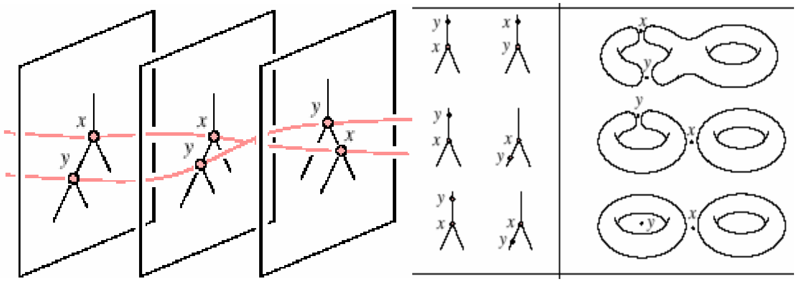
We are developing new algorithms for robust tracking of topological features.

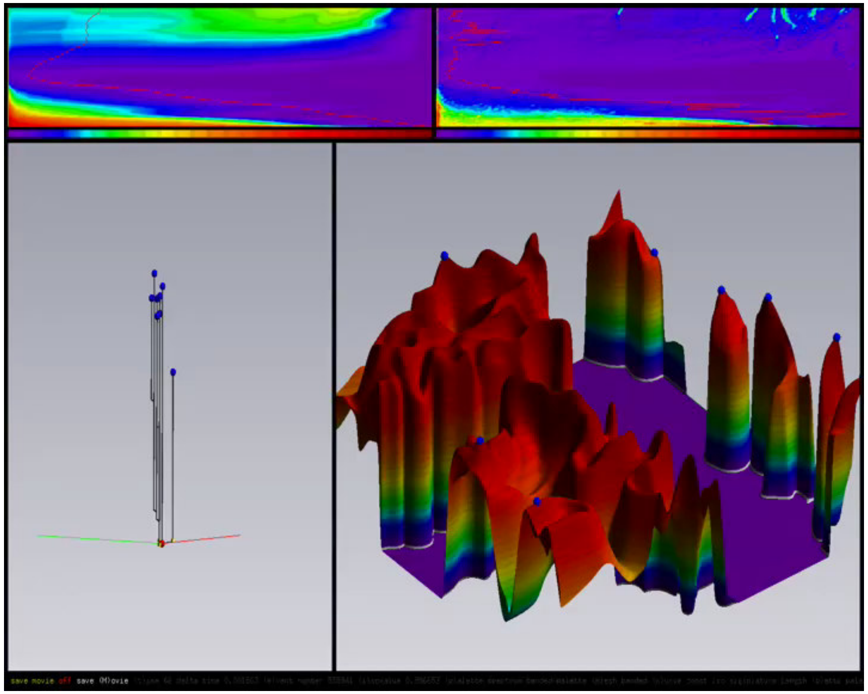
- Reeb-graphs help us understand scalar fields
  at moments in time



(1) (2) (3) (4)

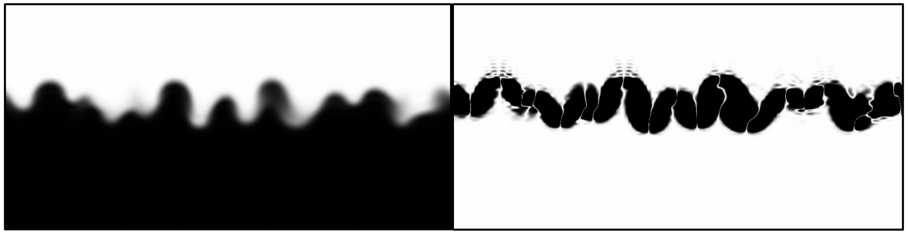
- Time-varying data results in complex changes in Reeb graph structure as the level sets evolve



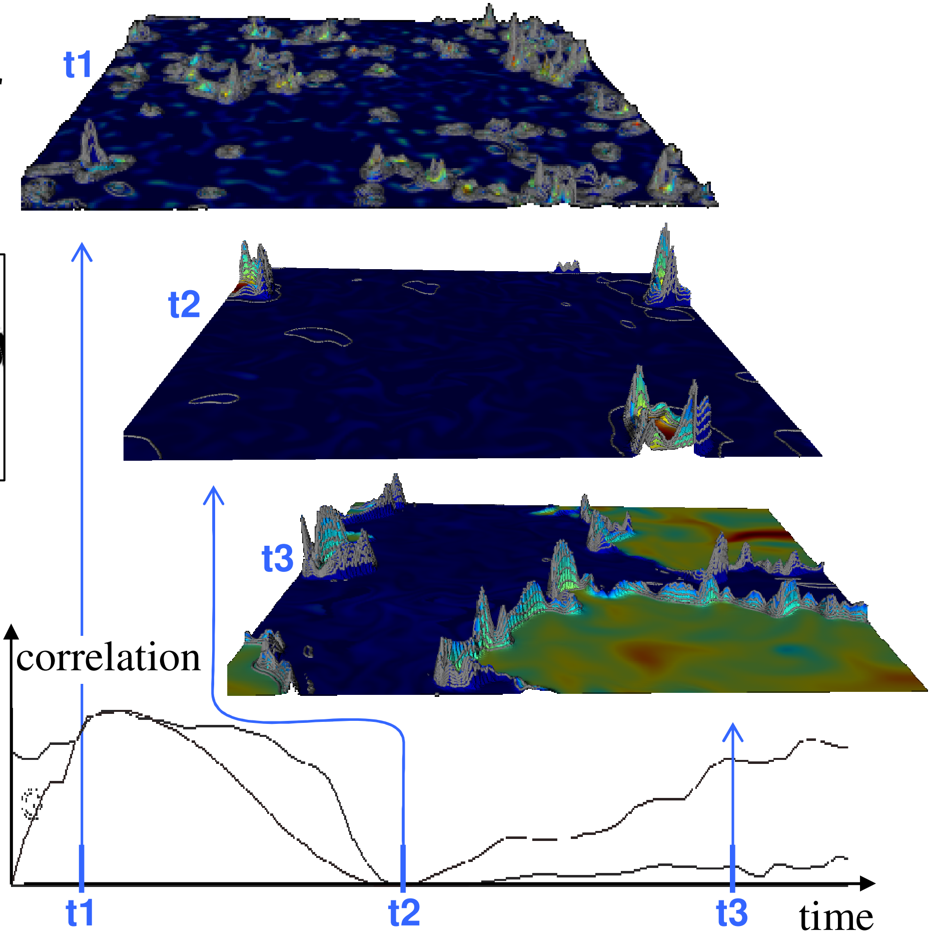


We use Jacobi sets to develop new comparison metrics for scalar fields.

The Jacobi set contains the points where the gradients of two fields are parallel.

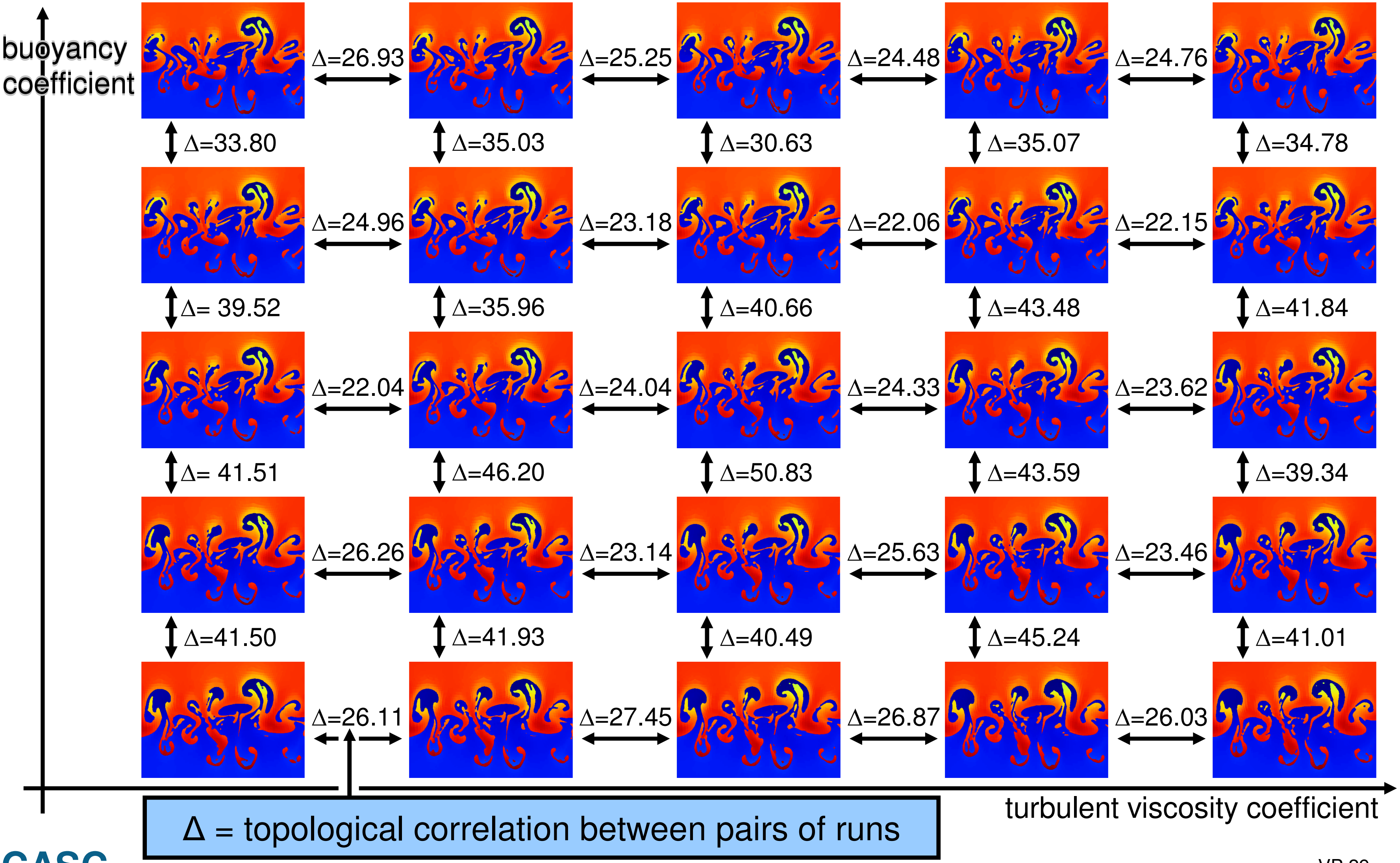


Jacobi set of density and pressure gradients highlights regions of vorticity generation. The critical points lie on the Jacobi set, relating the Morse complex to the Jacobi set.

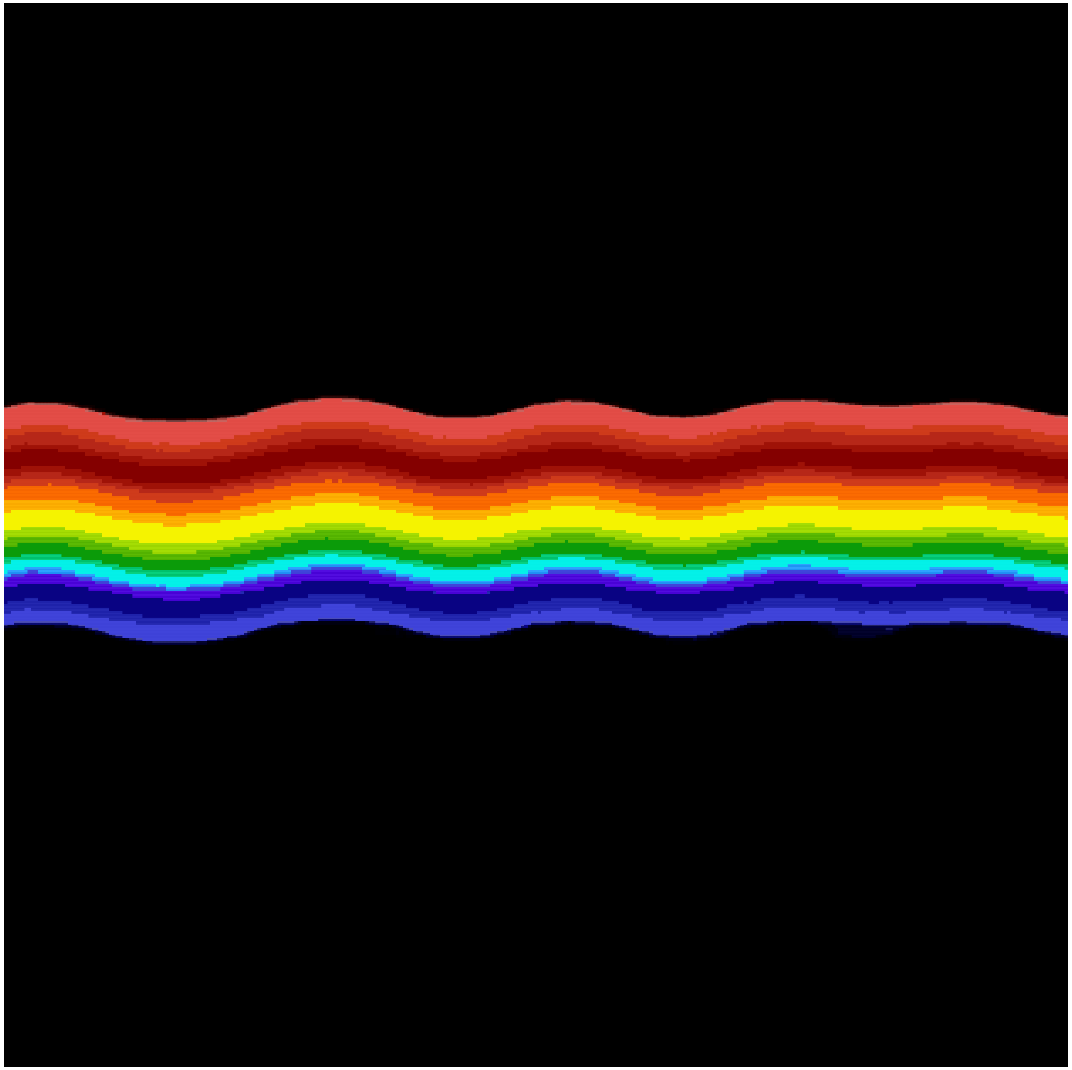


We have developed new metrics for scalar field comparison [6]

We are developing topology based shape characterization for V&V applications.



Rayleigh-Taylor instability occurs in super-novae, fusion, and other phenomena



Mixed region is shown at initial time

Heavy fluid is above mid-plane, light fluid is below

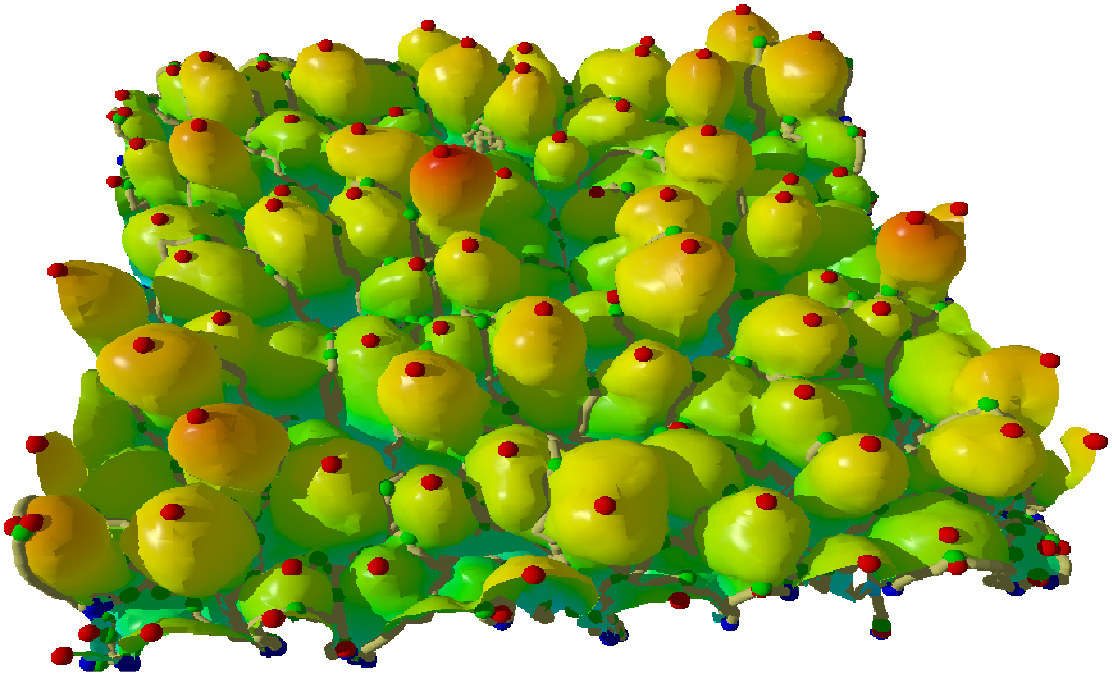
The mixing region lies between the upper envelope surface (red) and the lower envelope surface (blue)

We analyze Rayleigh-Taylor instability simulation data at high resolution

- ALC run: 1152 x 1152 x 1152
  - ~ 40 G / dump
  - 759 dumps, about 25 TB
- BlueGene/L run: 3072 x 3072 x Z
  - Z depends on width of mixing layer
  - Over 40 TB
- Bubble-like structures are observed in laboratory and simulations:
  - no prevalent formal definition of bubbles
- Bubble dynamics are thought to be one way to characterize the mixing process
- Previous bubble models
  - Kartoon *et. al*., single resolution model for early time
  - Bubbles are multi-scale structures

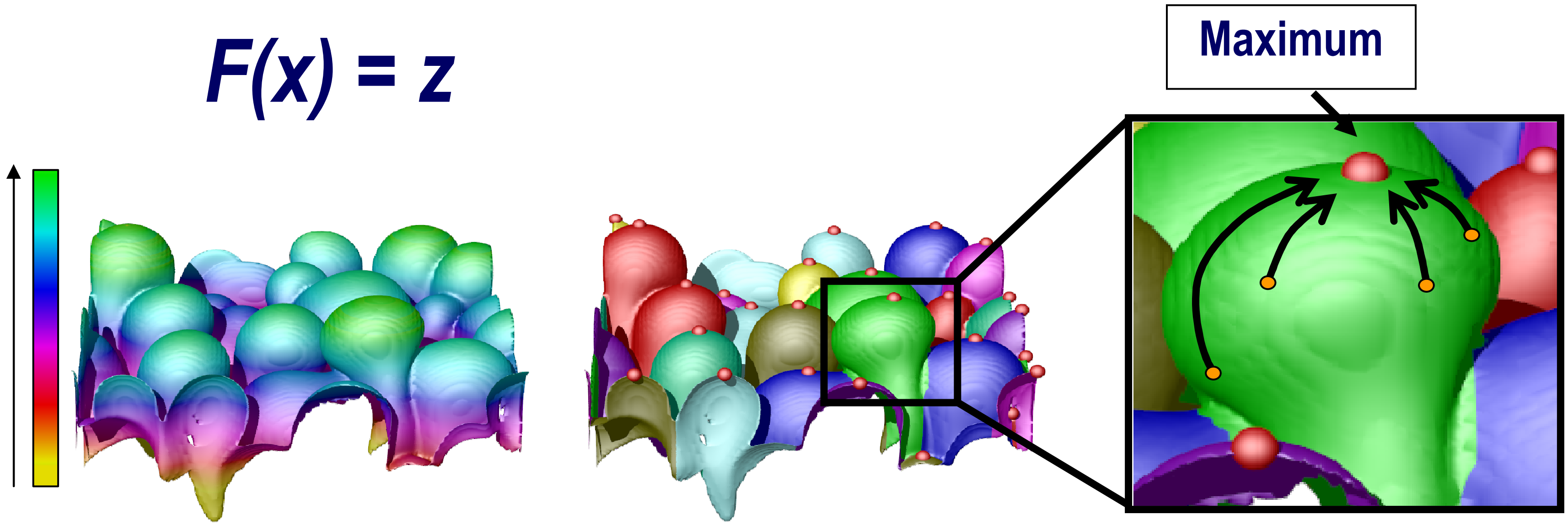
We propose Morse analysis of the upper envelope and interactive visualization

- The envelope surfaces are already utilized in the CFD community for analysis
- The structures in the surface provide a bound on the activity of bubbles in the flow



We compute the Morse-Smale complex of the envelope surface

F(x) = z



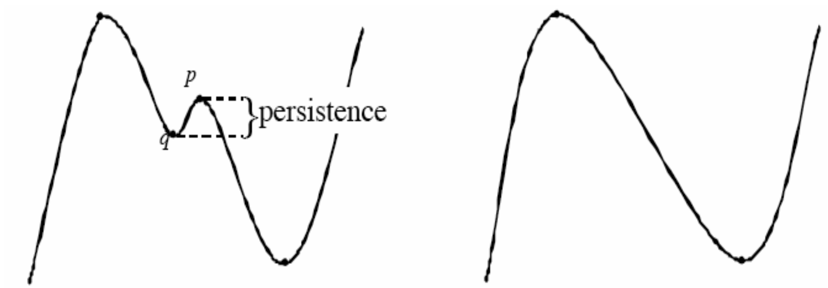
F(x) on the surface is aligned against the direction of gravity which drives the flow.

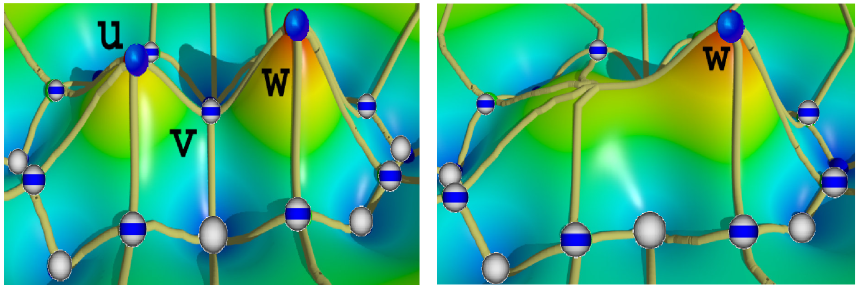
Morse complex cells drawn in distinct colors.

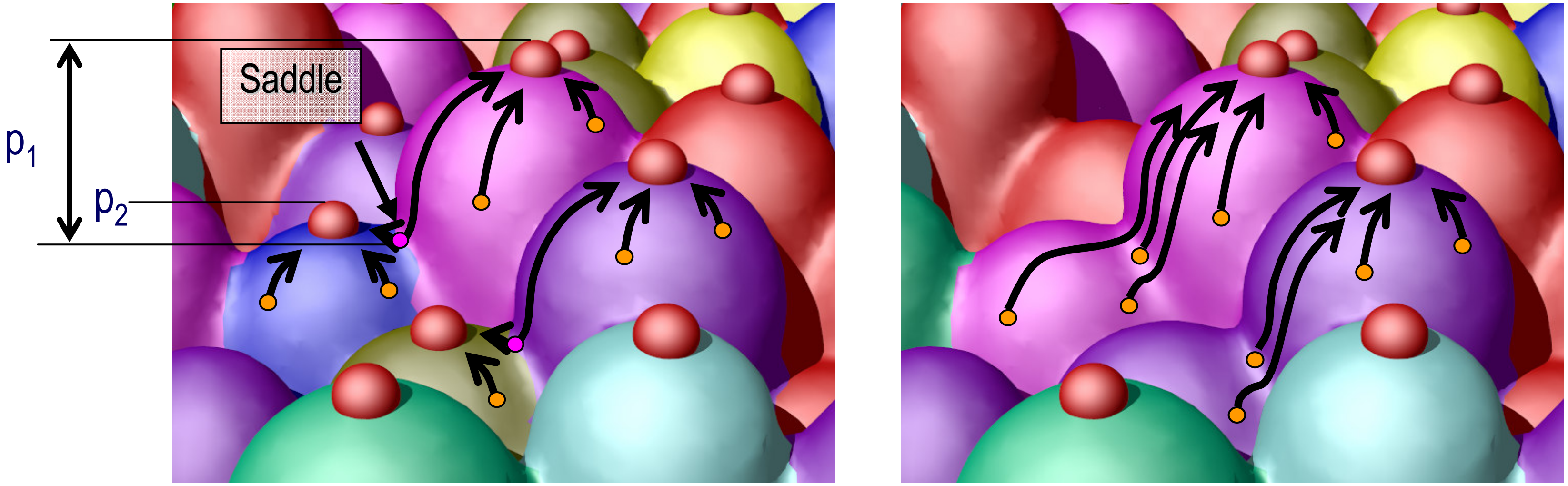
In each Morse complex cell all steepest ascending lines converge to one maximum.

Hierarchies are generated by simplifying topology

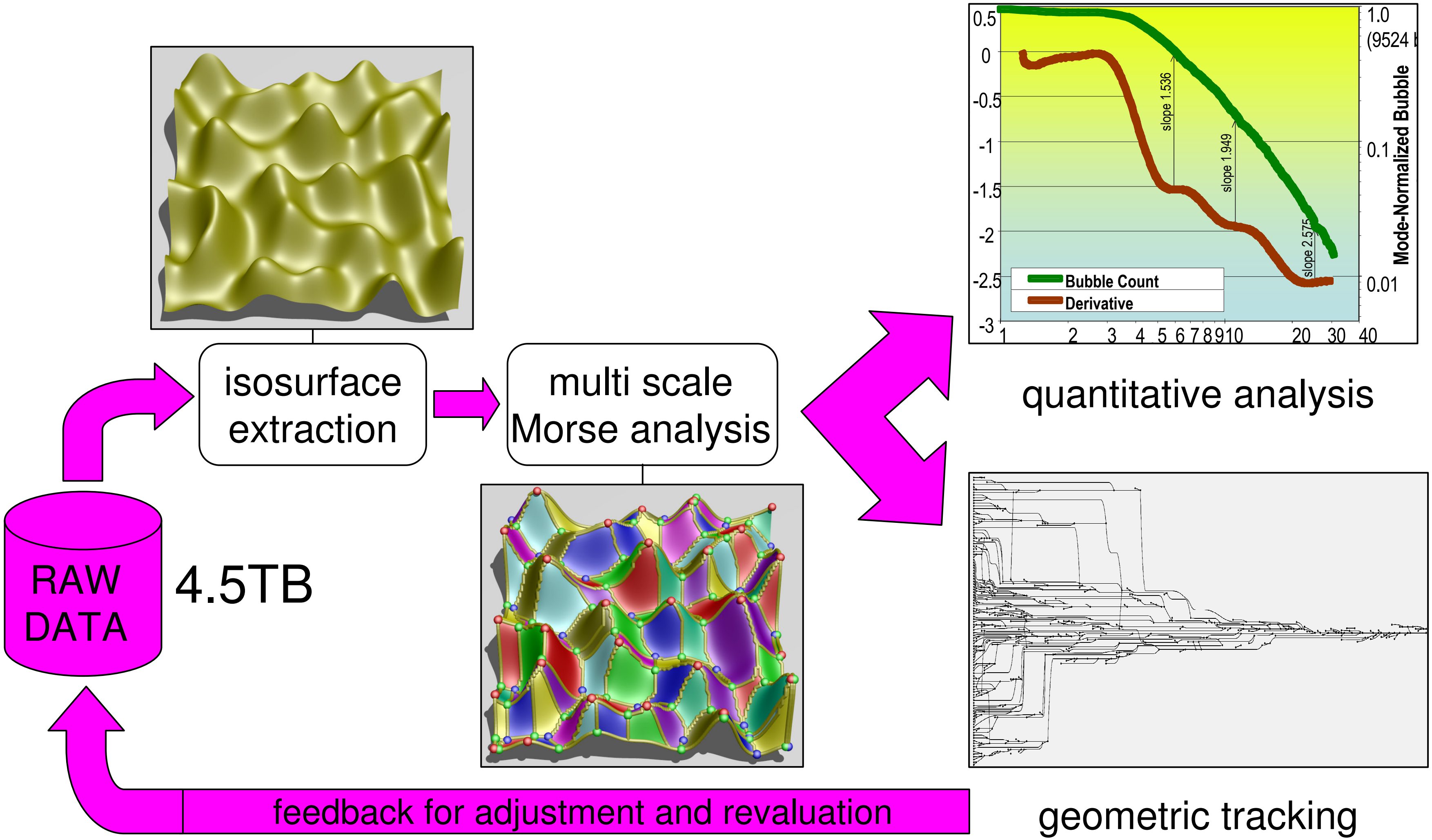
- **For example, the "persistence" may be varied, annihilating pairs of critical points and producing a smoothed surface**



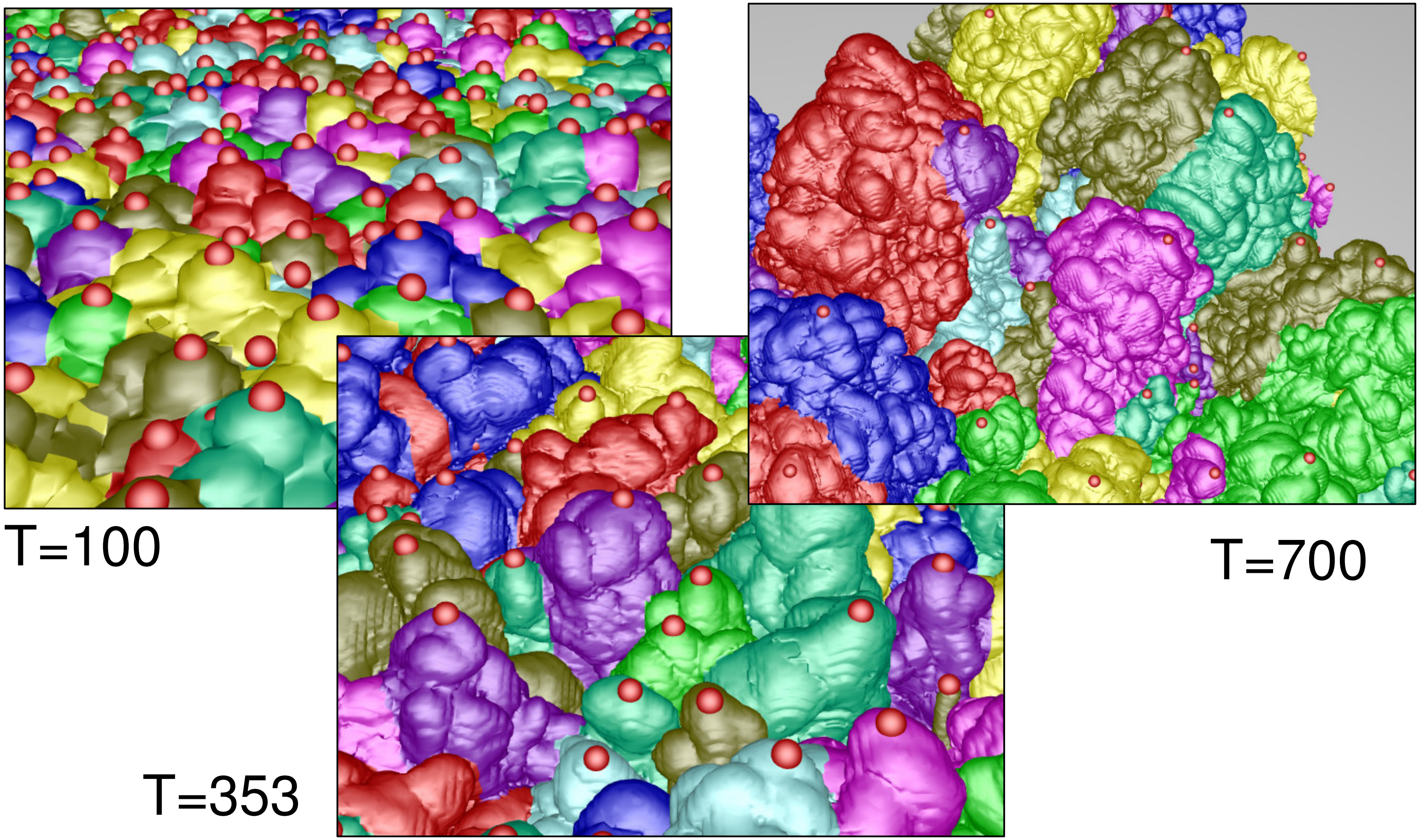




Our workflow utilizes streaming data management and analysis tools



We segment the envelope surface according to the Morse-Smale complex

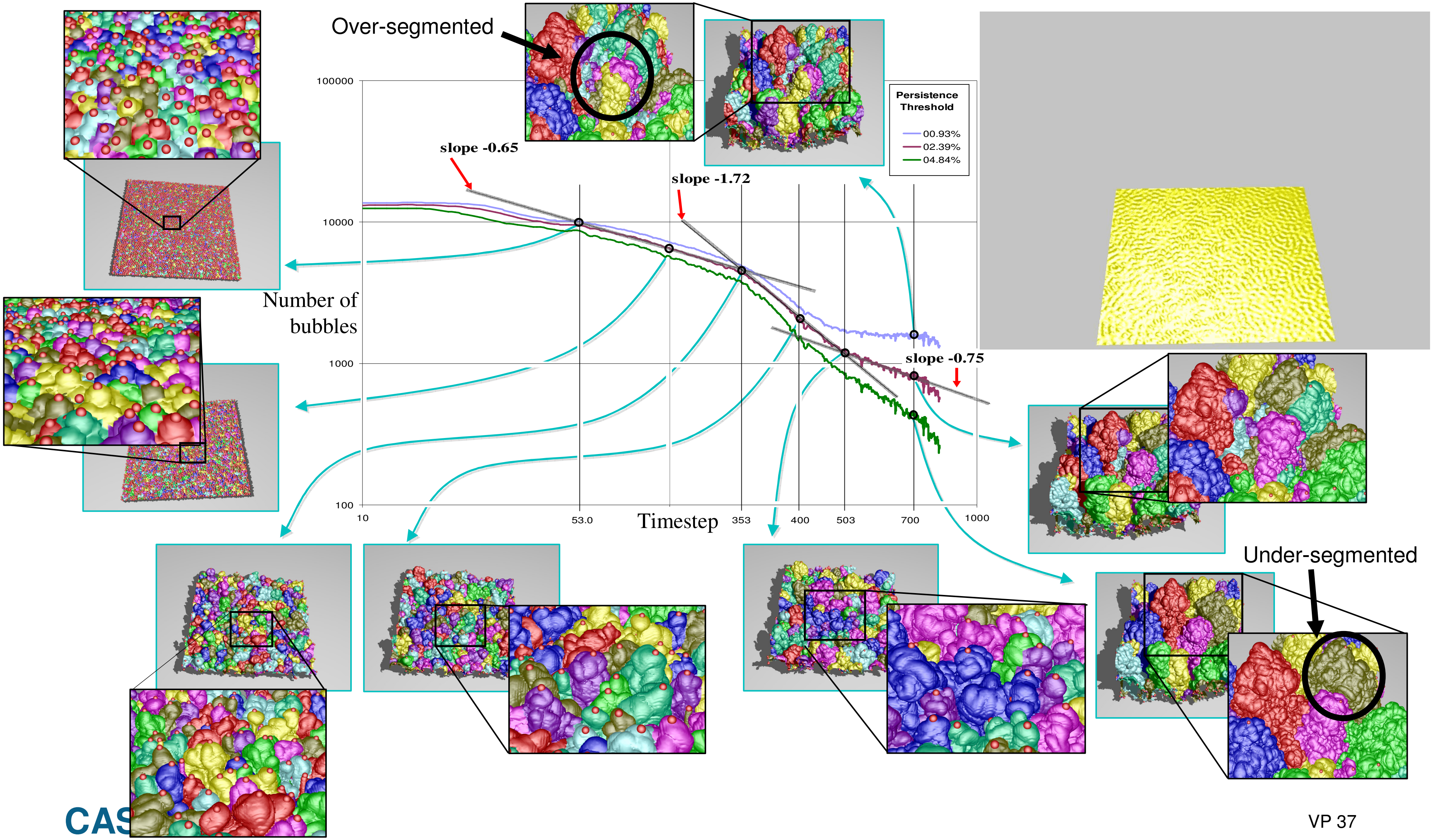


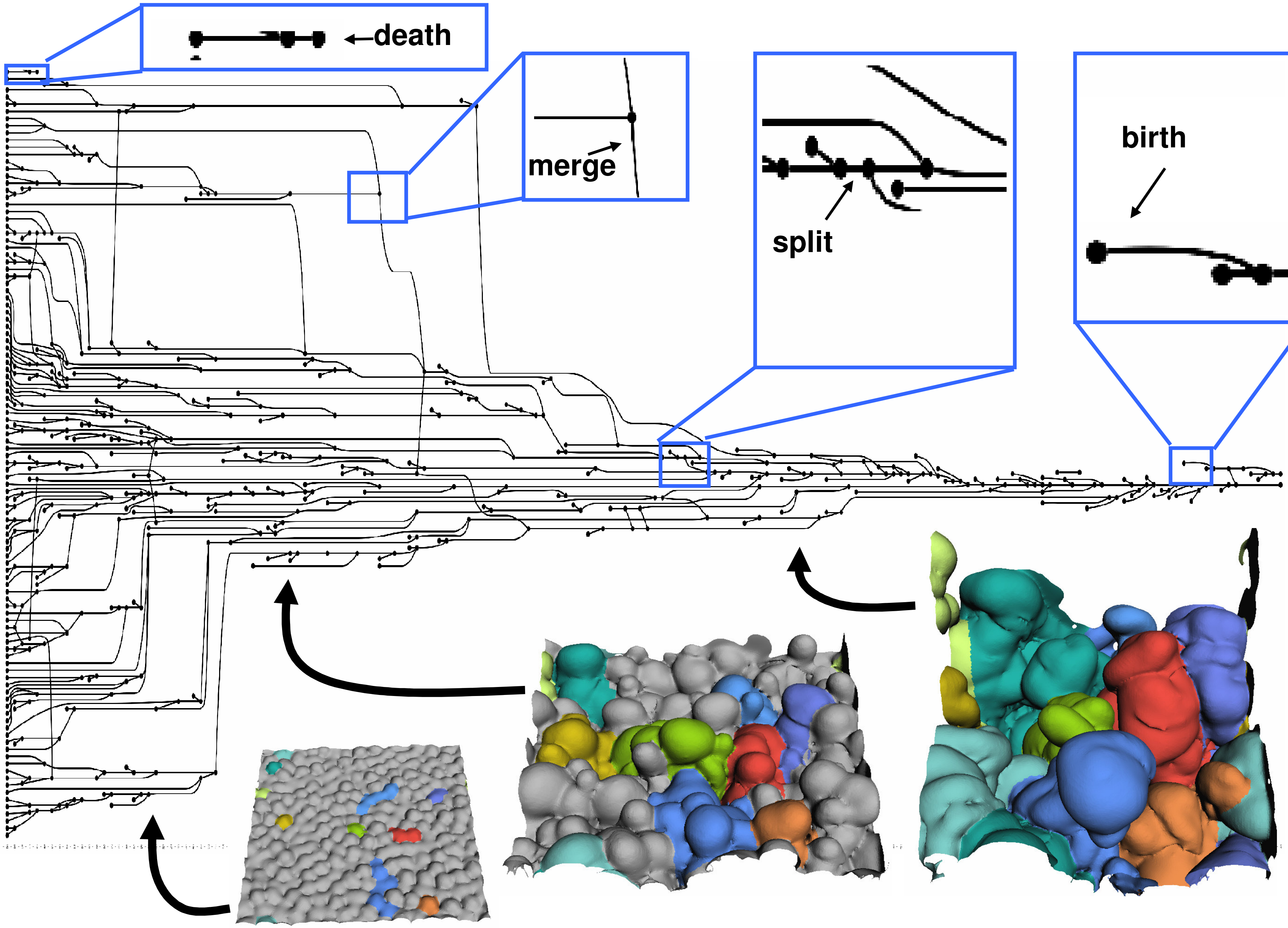
T=100

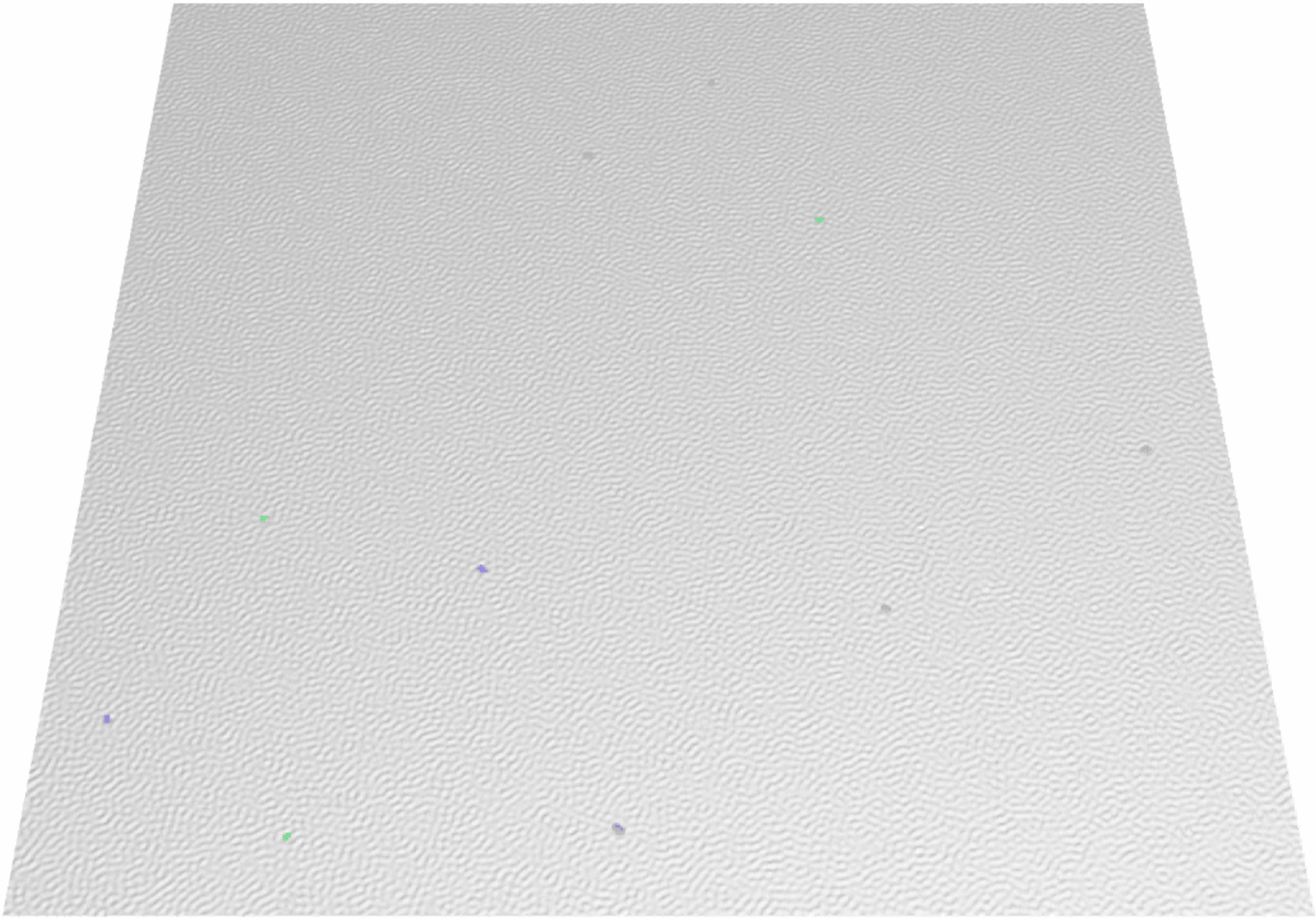
T=353

T=700

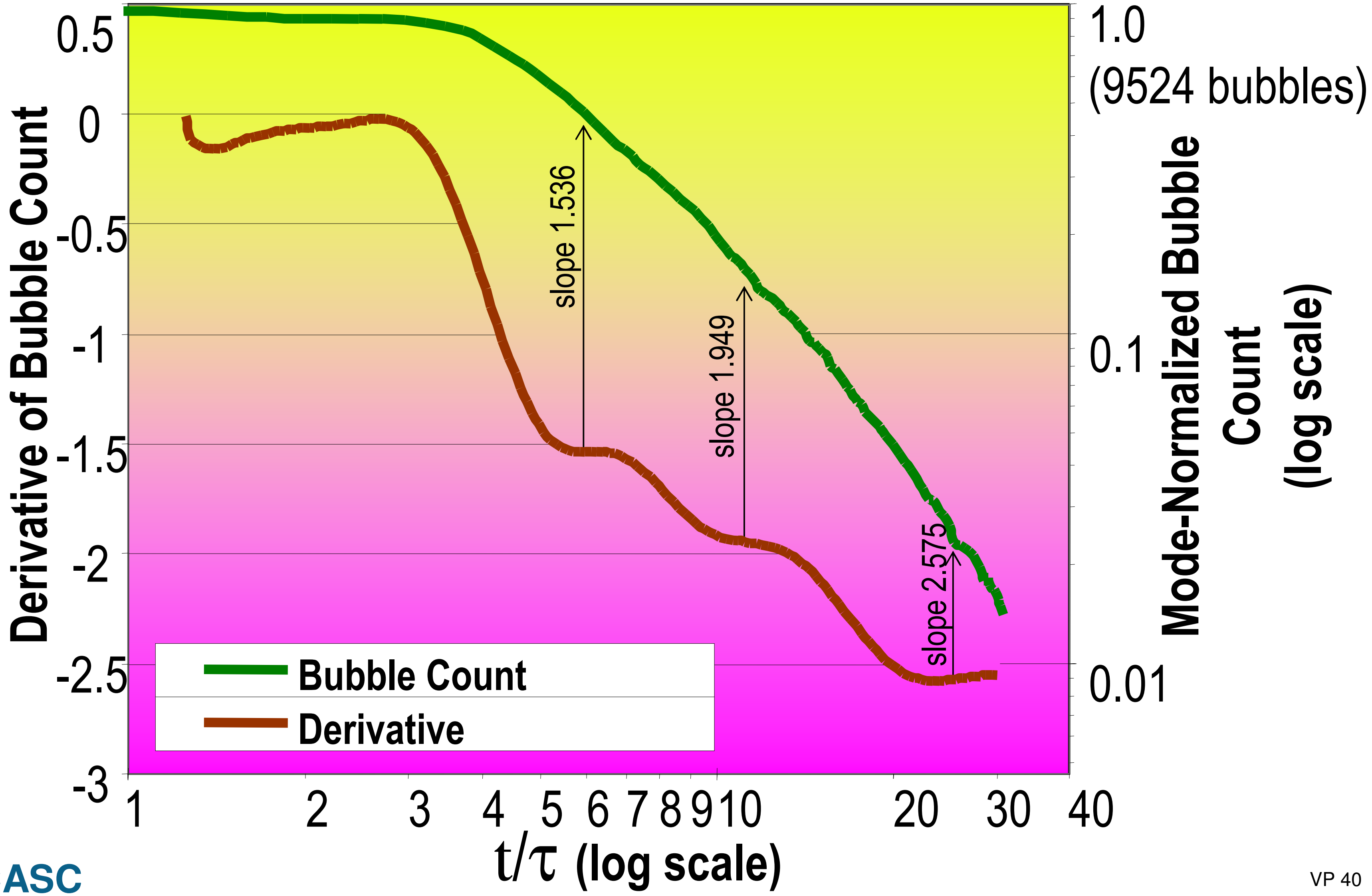
We are using Morse theory for quantitative analysis of Rayleigh-Taylor turbulence.



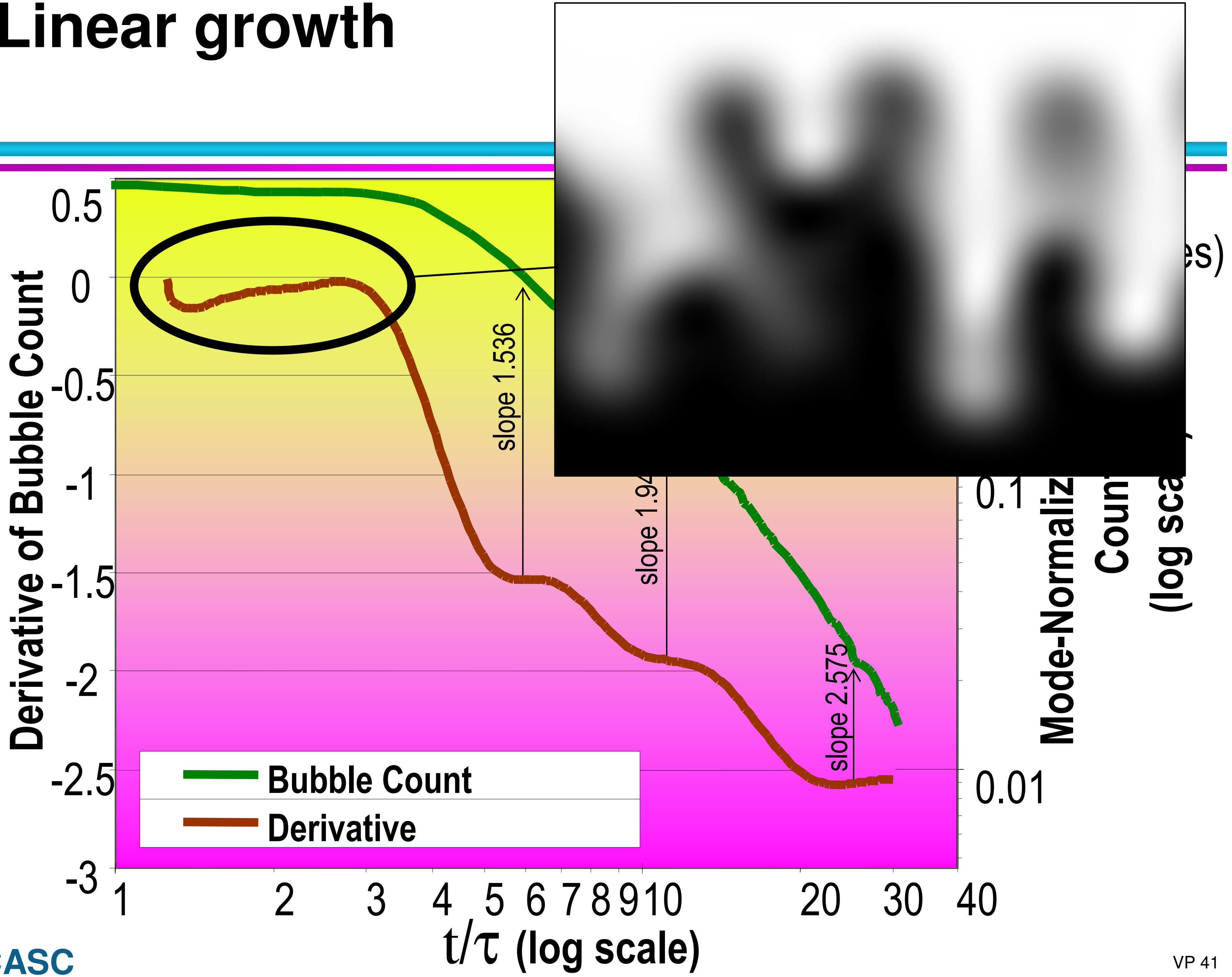

death
merge
split
birth



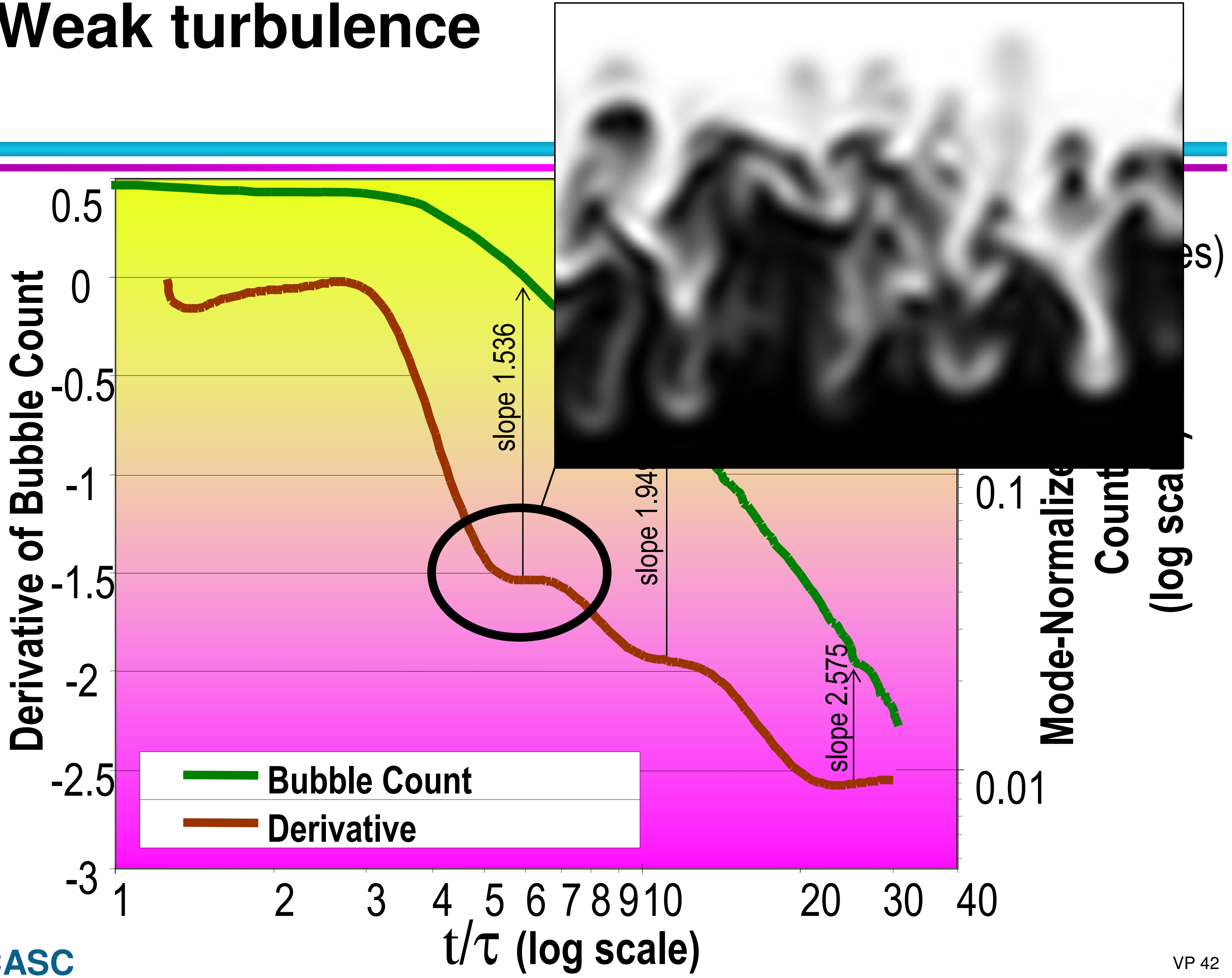
Adapting persistence produces quantitatively accurate bubble counts over time



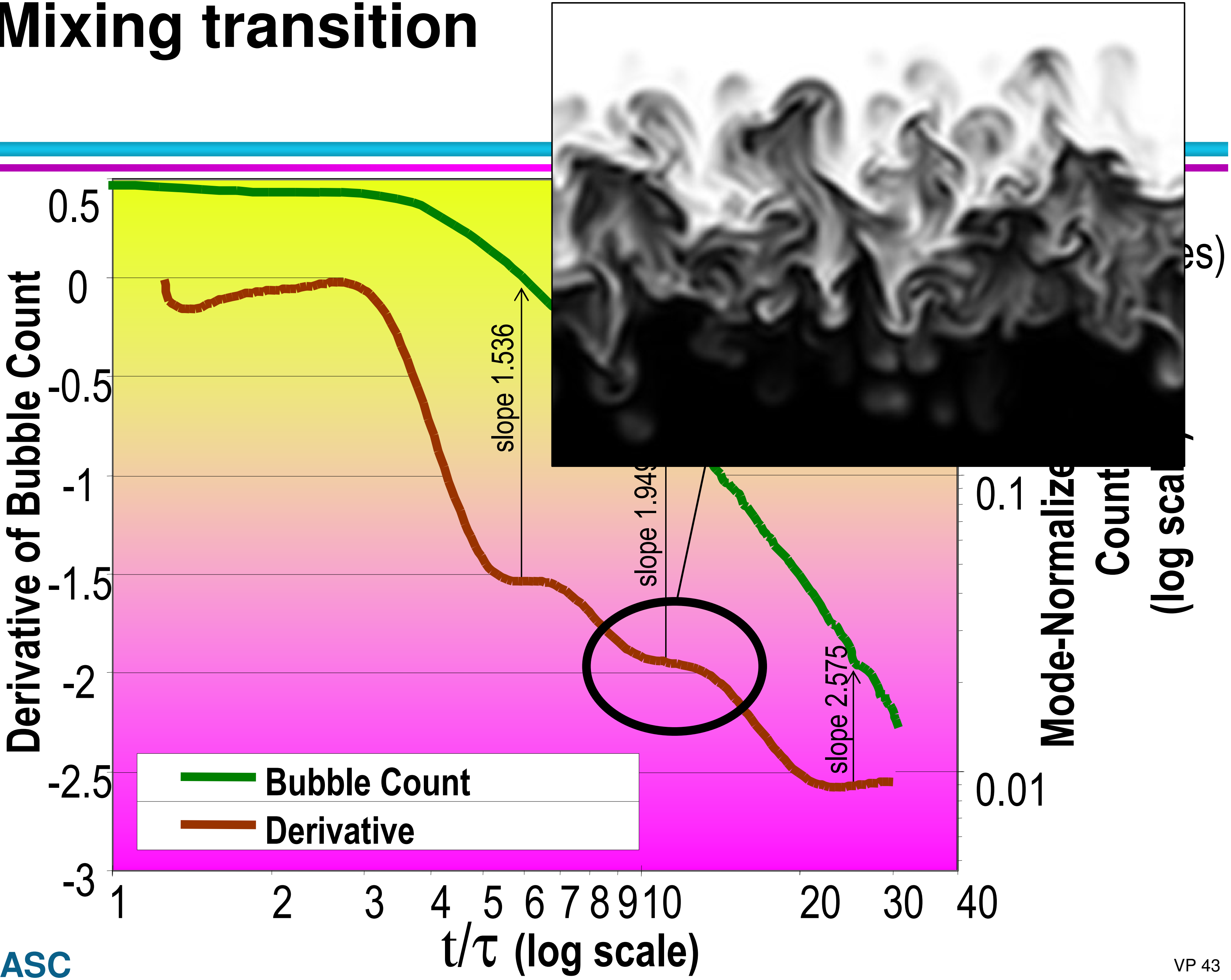
Linear growth



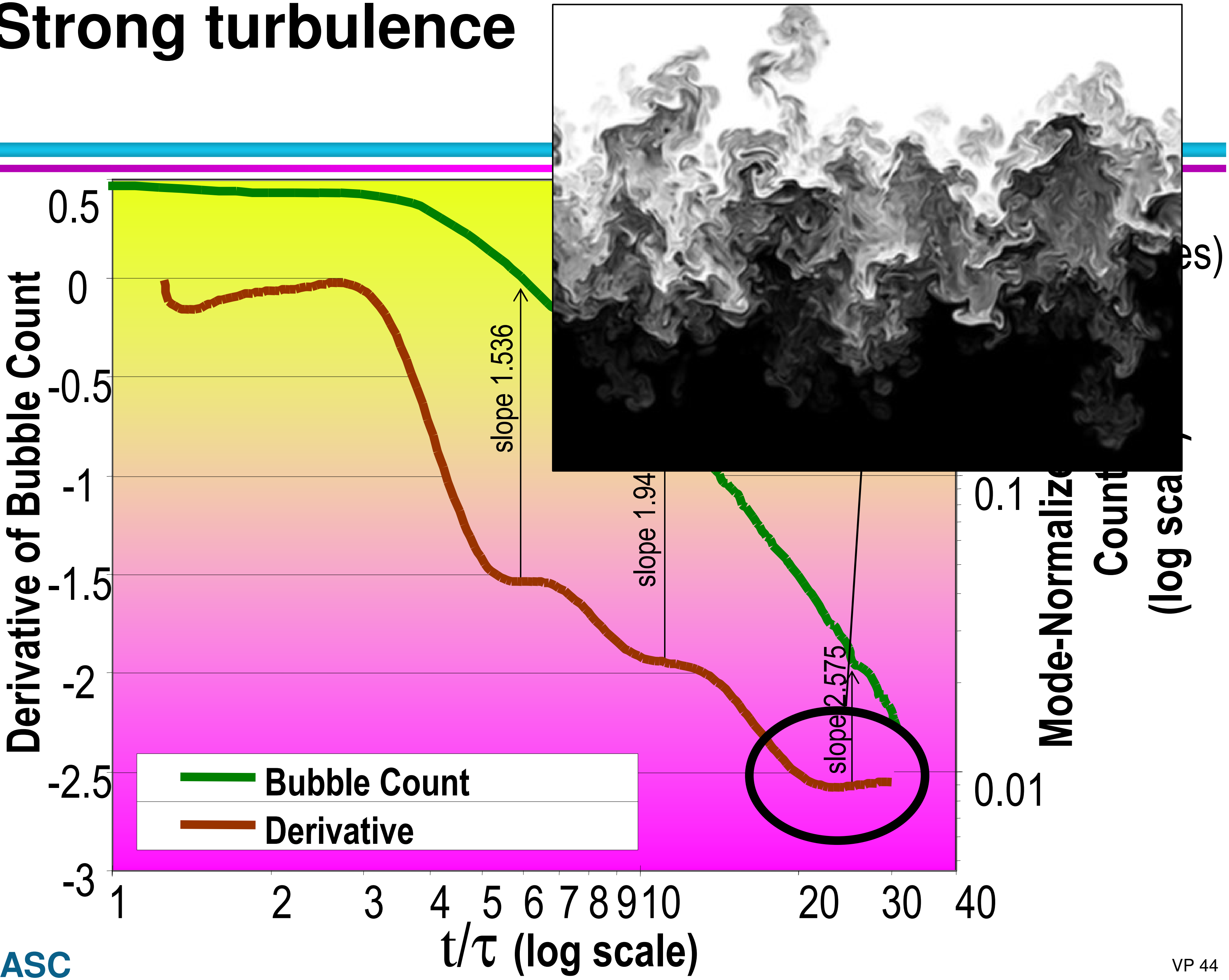
Weak turbulence



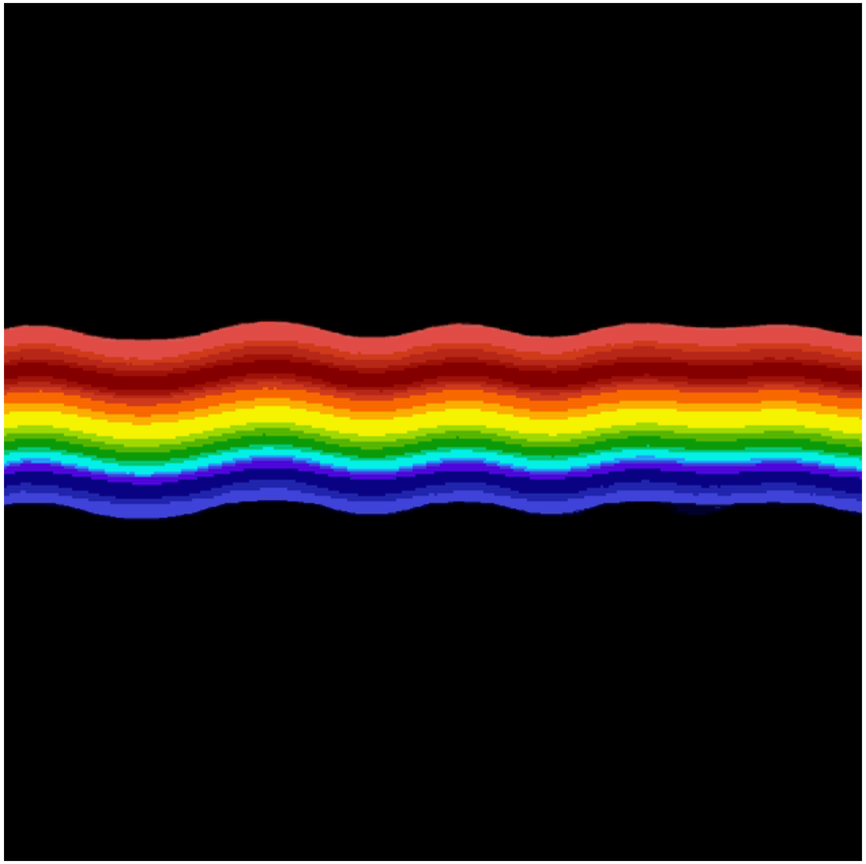
Mixing transition



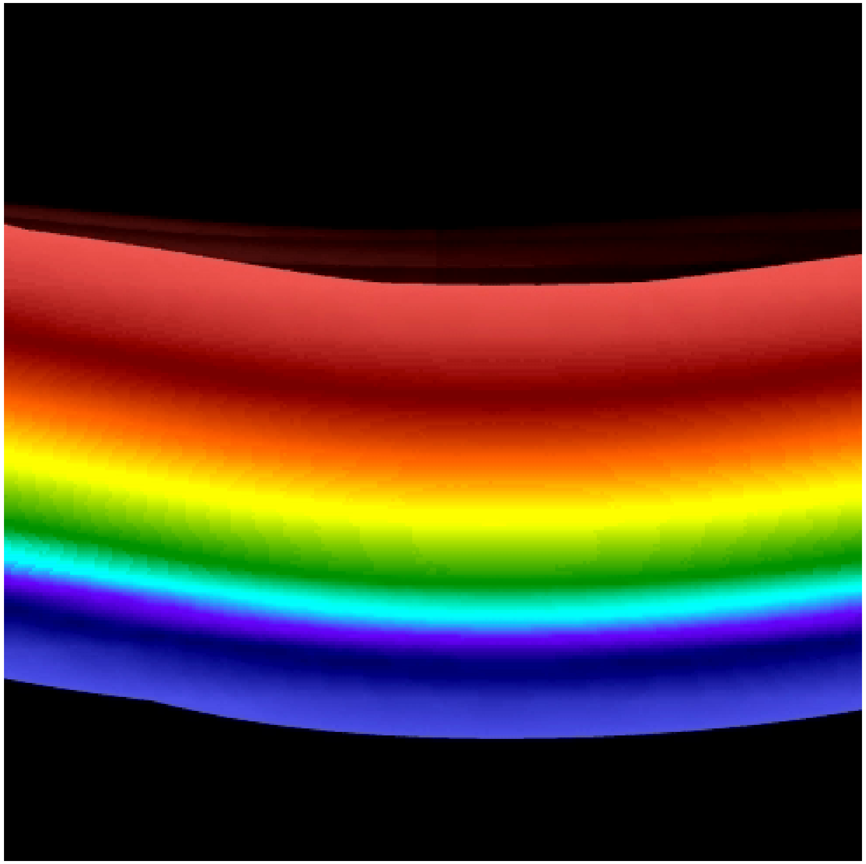
Strong turbulence



Analysis and comparison of turbulent mixing simulations of different type.

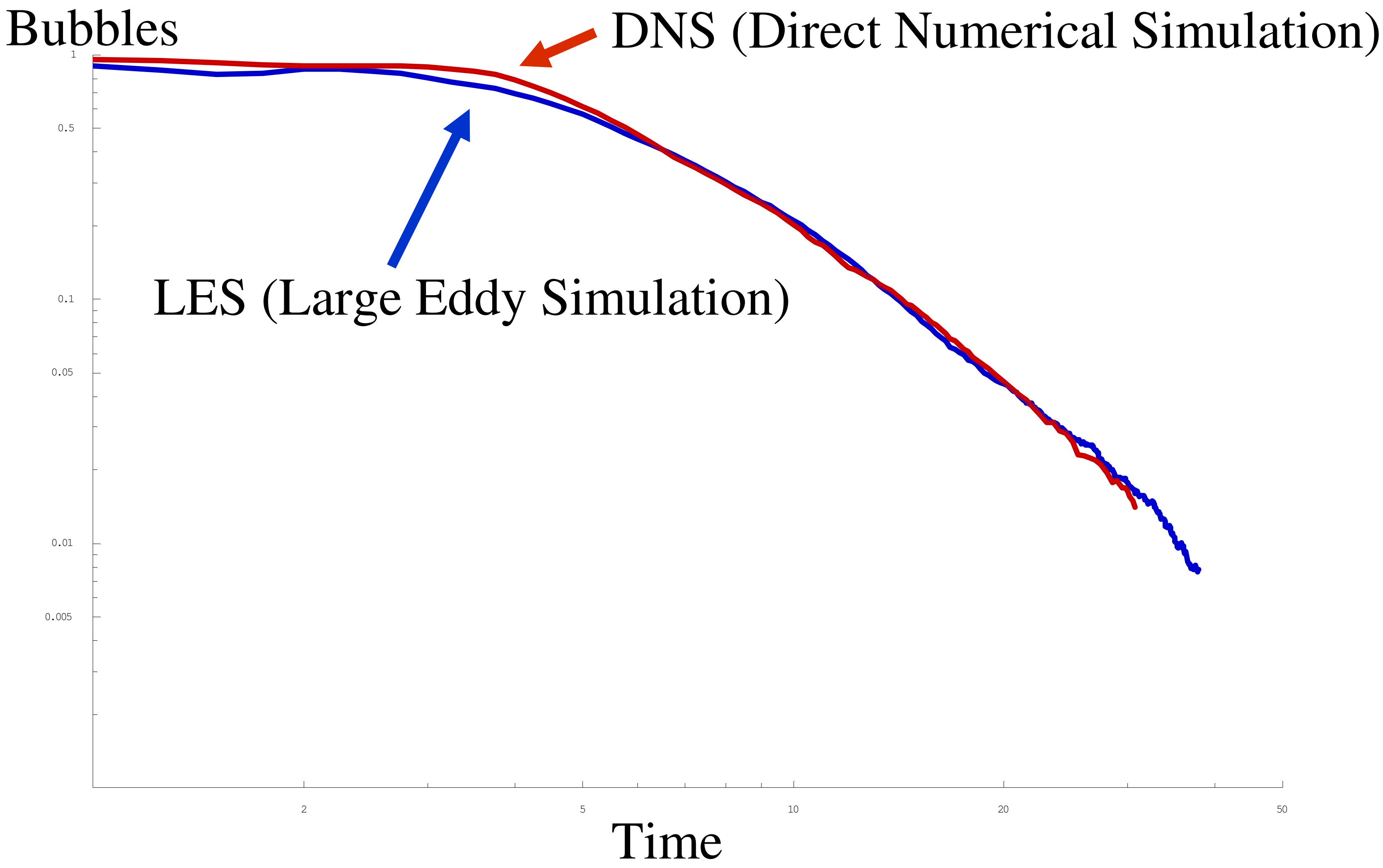


Large Eddy Simulation



Direct Numerical Simulation

Feature based comparison and validation of a DNS with a LES.



Conclusions

- Development of robust and reliable algorithms based on sound theory of model that can be represent on a computer.
- Design of proper Morse functions for application specific segmentation.
- Extension to general vector and tensor fields.
- Extension to discrete data (e.g. ontology graphs).

UCRL-PRES-226524

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.