VisWeek 09
VIS • INFOVIS • VAST

Topology-Based Data Analysis in Practice

Peer-Timo Bremer
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory

Lawrence Livermore National Laboratory LLNL-PRES-413967

A Significant Portion of Data Analysis is Feature-Driven

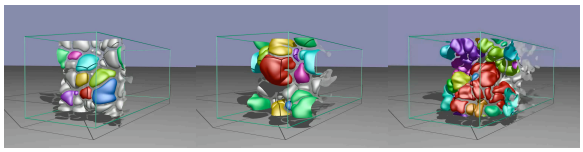
- Practical data analysis relies on the definition of features leading to the common workflow of:
  - Mathematically define features;
  - Choose a set of selection parameters, e.g. iso-values, thresholds, etc.;
  - Extract the corresponding collection of features; and
  - Analyze feature properties , e.g. count, distributions, conditional statistics etc..
- Some of the challenges related to feature based analysis are:
  - Arbitrary/Unintuitive parameter selection;
  - Lack of stability under changes in parameter selection; and
  - High computational cost of repeated evaluation.

Topology-Based Feature Definition and Analysis Addresses These Challenges For Two Broad Classes of Features

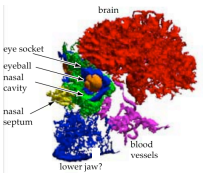
- Topology provides a general framework to define features:
  - Robustly: stable under noise, adapted to piece-wise linear and/or constant data;
  - Reliably: strong guarantees wrt. feature dimension and global structure;
  - Hierarchically: inherits natural feature-based hierarchies;
  - Efficiently: compute/encode entire feature families; and
  - Flexibly: features are sub-sets of more general structures allowing a small set of tools to apply to different applications.
- Topological theory forms a flexible language to define and analyze threshold- and gradient-based features in scalar fields

Threshold-Based Features

- Features defined through contours are called *threshold-based*:
  - Scalar classifiers: E.g. Regions with high fuel consumption

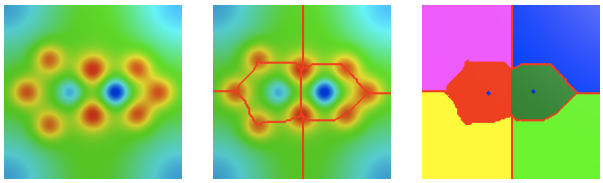


  - Flexible iso-surfaces: Regions delimited by locally varying iso-values



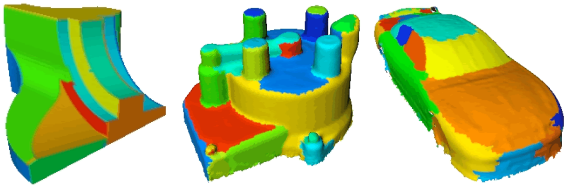
Gradient-Based Features

- Features defined as regions of similar (gradient) flow:
  - Image watershed



Gradient-Based Features

- Features defined as regions of similar (gradient) flow:
  - Image watershed
  - Feature based surface segmentation: e.g. curvature



http://imaging.utk.edu/people/former/sun/segmentation/segmentation.htm

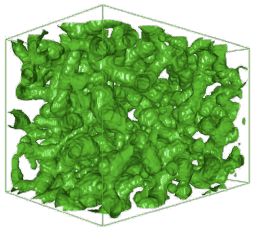
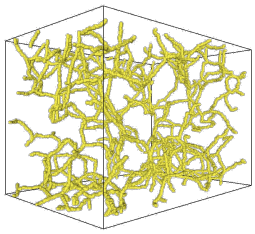
Gradient-Based Features

- Features defined as regions of similar (gradient) flow:
  - Image watershed
  - Feature based surface segmentation
  - Edge detection


http://www.cs.lth.se/home/Calle_Lejdfors/pyip/

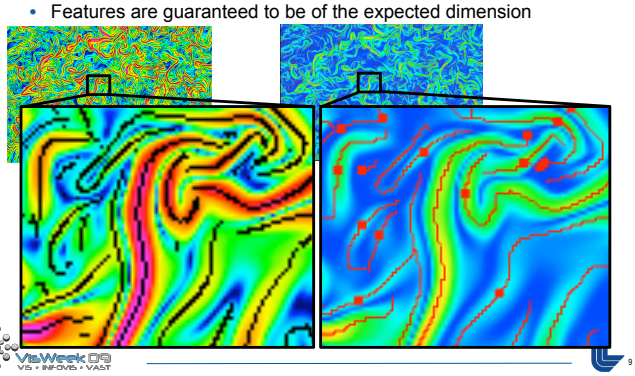
Gradient-Based Features

- Features defined as regions of similar (gradient) flow:
  - Image watershed
  - Feature based surface segmentation
  - Edge detection
  - Surface skeletons

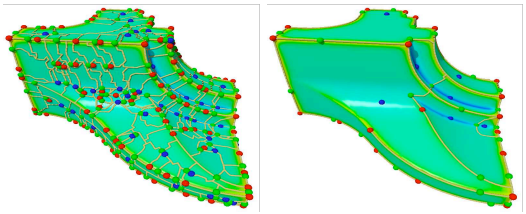
Why Use Topology to Define/Extract Well-Known Features ?

- Structural consistency:
  - Features are guaranteed to be of the expected dimension



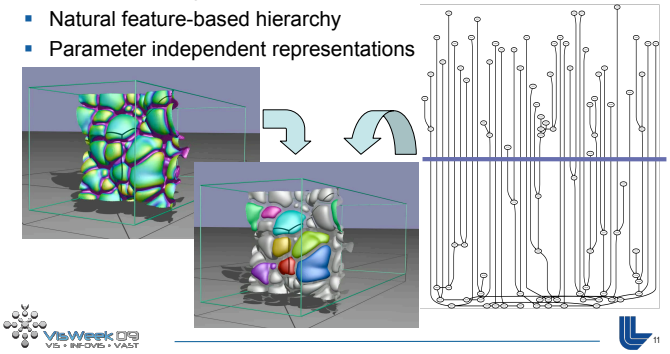
Why Use Topology to Define/Extract Well-Known Features ?

- Structural consistency:
  - Features are guaranteed to be of the expected dimension
- Natural feature-based hierarchy:



Why Use Topology to Define/Extract (Well-Known) Features ?

- Structural consistency:
  - Features are guaranteed to be of the expected dimension
- Natural feature-based hierarchy
- Parameter independent representations



Why Use Topology to Define/Extract (Well-Known) Features ?

- Structural consistency:
  - Features are guaranteed to be of the expected dimension
- Natural feature-based hierarchy
- Parameter independent representations
- Small set of general tools

Threshold-based	*Gradient-based*
⇩	⇩
Contour-tree / Reeb graph	**Morse-Smale complex**

Recap: Six Different Topological Structures

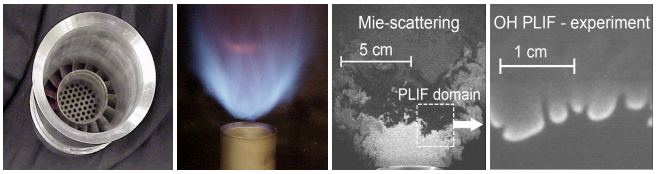
- Join tree: Maxima/Super-level sets and their merging behavior
- Split tree: Minima/Sub-level sets and their merging behavior
- Contour tree: Connected sets of contours and their evolution
- Reeb graph: Contours on non-genus 0 domains
- Morse complex: Gradient basins (watershed)
- Morse-Smale complex: Regions of similar flow behavior

Topological Data Analysis Proceeds in Four Stages

- Which topological element corresponds to features of interest:
  - What feature type are we interested in (thresholds, gradients, etc.) ?
- Which topological (sub-)structure is required for the analysis:
  - Which property do we analyze (e.g. count, extend, shape, etc.) ?
- What is the appropriate simplification metric:
  - When do feature merge (e.g. as they touch, similar function value) ?
  - What is the appropriate metric (e.g. persistence, volume) ?
- What global sub-selection is needed:
  - Determine appropriate parameters (e.g. noise-to-signal rations).
  - Define global sub-selection (e.g. disregard small features).

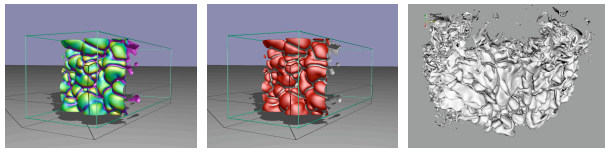
Analysis of Low-Swirl Turbulent Combustion

- Low-swirl injectors are an emerging technology that has the potential to stabilize a lifted, lean hydrogen air flame.
- Such flames can reduce emissions dramatically and enable new technology in stationary power generation.
- However, hydrogen burns in a complex cellular mode that defies standard diagnostic techniques.



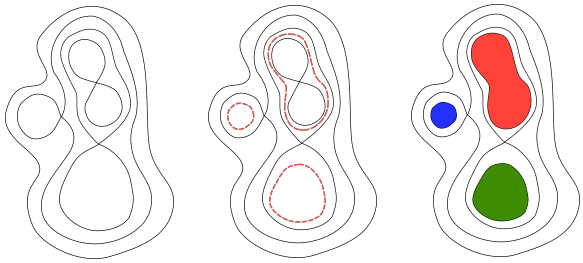
Detailed Simulations Augment Laser-Based Diagnostics

- Numerical simulations of laboratory-scale burners help to interpret the complex, time-dependent evolution of burning cells:
  - How many cells exist and any given time interval ?
  - What are their chemical and physical properties (size, temp., etc.) ?
  - What is their temporal evolution ?
  - How do the flame characteristics depend on the turbulence level ?
- Burning cells are defined as regions of high ($\geq t$) fuel consumption on an isotherm (flame surface) or in 3D space



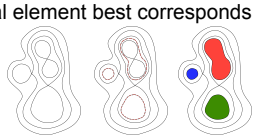
Select Topological Element Corresponding to Burning Cells

- Which topological element best corresponds to burning cells at threshold t ?



Select Topological Element Corresponding to Burning Cells

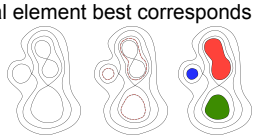
- Which topological element best corresponds to burning cells at threshold t ?



⇨ Components of super-level sets

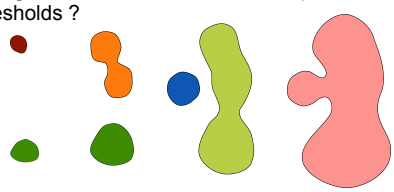
Select Topological Structure to Encode Burning Cells

- Which topological element best corresponds to burning cells at threshold *t* ?



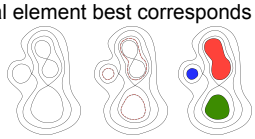
⇨ Components of super-level sets

- Which topological structure best encodes super-level sets at various thresholds ?



Select Topological Structure to Encode Burning Cells

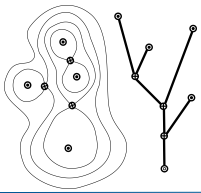
- Which topological element best corresponds to burning cells at threshold *t* ?



⇨ Components of super-level sets

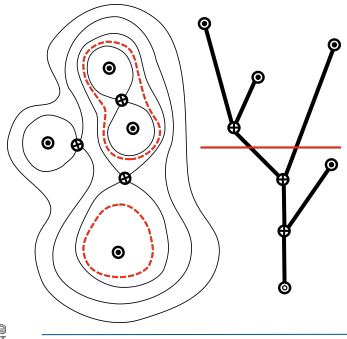
- Which topological structure best encodes super-level sets at various thresholds ?

⇨ Join tree



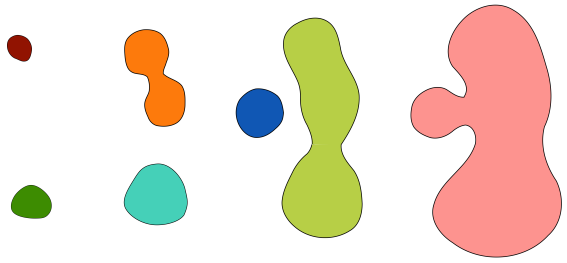
Select Topological Structure to Encode Burning Cells

- Which properties do we need for the analysis ?
  - Cell count: Number of intersecting branches.



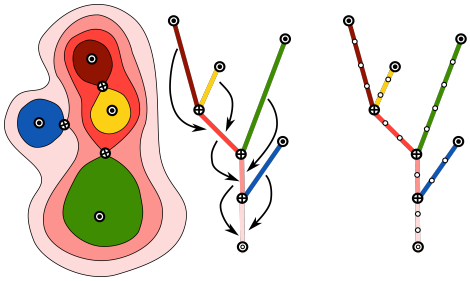
Select Topological Structure to Encode Burning Cells

- Which properties do we need for the analysis ?
  - Cell count: Number of intersecting branches.
  - Cell properties (size, avg. temp, etc.):



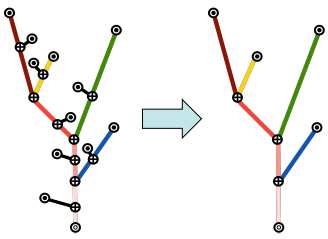
Select Topological Structure to Encode Burning Cells

- Which properties do we need for the analysis ?
  - Cell count: Number of intersecting branches.
  - Cell properties (size, avg. temp, etc.): Branch-based segmentation.



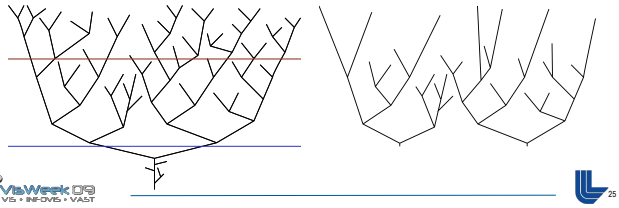
Combustion Analysis Using Topology

- Which properties do we need for the analysis ?
  - Cell count: Number of intersecting branches.
  - Cell properties (size, avg. temp, etc.): Branch-based segmentation.
- What is the appropriate simplification metric ?
  - Persistence to reduce instabilities in the threshold



Combustion Analysis Using Topology

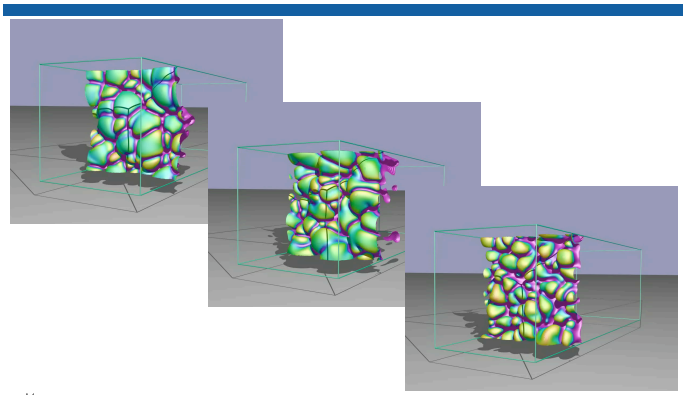
- Which properties do we need for the analysis ?
  - Cell count: Number of intersecting branches.
  - Cell properties (size, avg. temp, etc.): Branch-based segmentation.
- What is the appropriate simplification metric ?
  - Potentially persistence to reduce instabilities in the threshold
- Additional global sub-selection ?
  - Potentially simplify the trees around the *interesting* parameter range



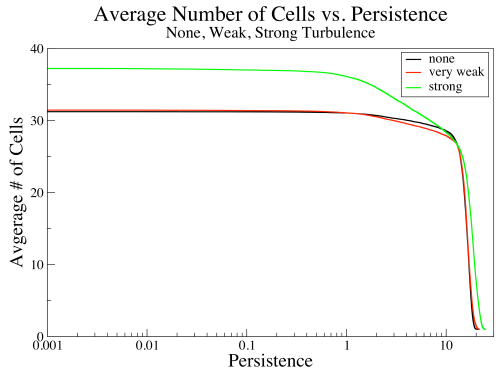
Segmented Merge Trees Provide Unprecedented Analysis Capabilities to Study Turbulent Combustion

- Efficient encoding of all possible burning cell segmentations
  - Freedom to choose thresholds and simplification post-analysis
- Instantaneous access to any pre-computed statistics
  - Ability to process multiple time-series on low-end hardware
  - Feasibility of extensive parameter studies
- Encoding is computed in a single computational pass over the data
  - Efficient and embarrassingly parallel pre-processing well suited to common parallel resources
- Data compression on the order of two magnitudes
  - Reducing data sizes from around 4TB to 2GB statistical data

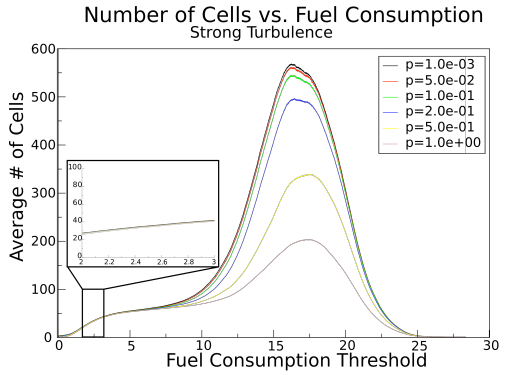
Topological Analysis To Understand the Influence of Turbulence on Lean, Pre-Mixed Combustion



Parameter Studies Show That Simplification is Unnecessary

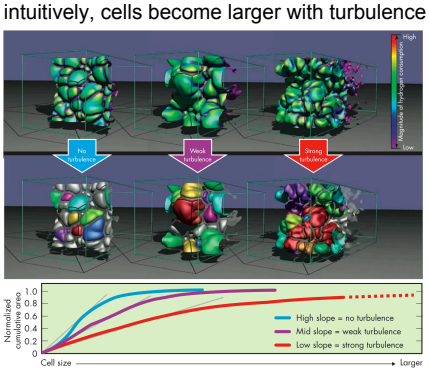


Determining Stable Threshold for Analysis

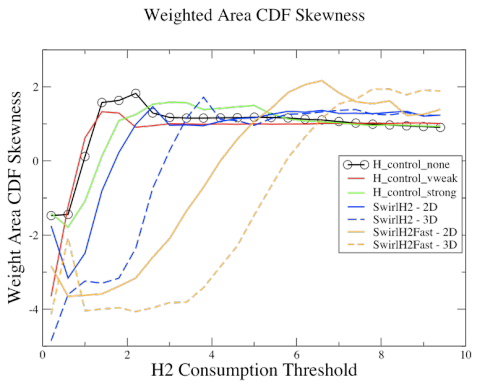


Topological Analysis Provides New Scientific Insights

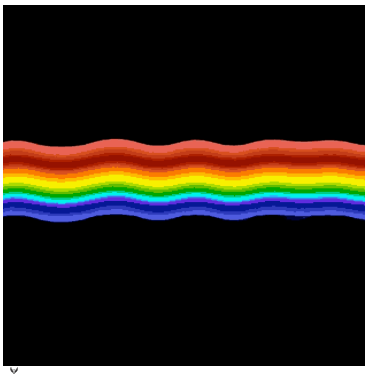
- Counter intuitively, cells become larger with turbulence



Novel Framework Makes New Types of Analysis Feasible

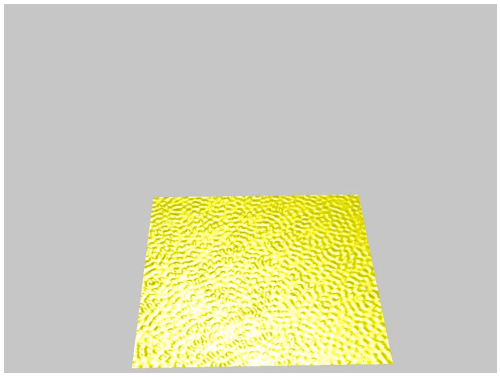


Analysis of Hydrodynamic Instabilities



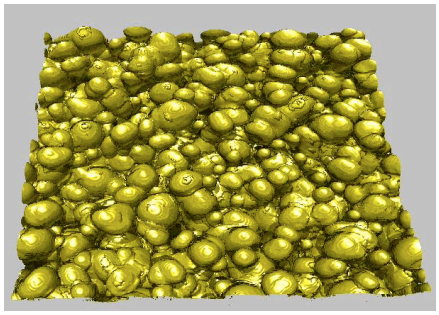
- Miranda simulation.
- LES: 1152^3, about 25 TB
- DNS: 3072^2 x varying Z, over 40 TB
- Goal:
  - Understand bubble dynamics
- Challenge:
  - No prevalent definition of a bubble
  - Bubbles are multi-scale with the scale highly dependent on the mixing phase

Analyze Bubble Structure of Upper Envelope Surface



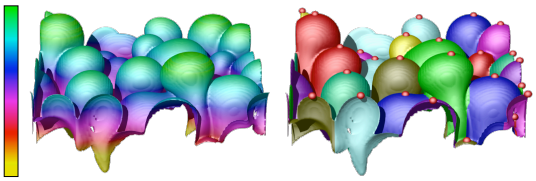
Select Topological Structure to Analyze Bubbles

- How to mathematically define the intuitive notion of bubbles?
  - Hint: We have yet to define the function to analyze



Select Topological Structure to Analyze Bubbles

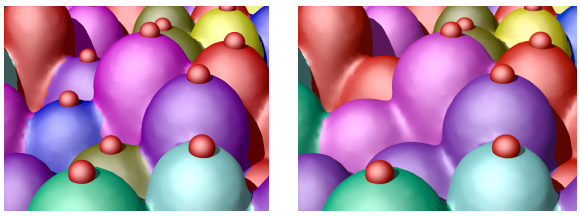
- How to mathematically define the intuitive notion of bubbles?
  - Hint: We have yet to define the function to analyze
  - Use gravity as function



⇨ Morse complex

What is The Appropriate Simplification Metric ?

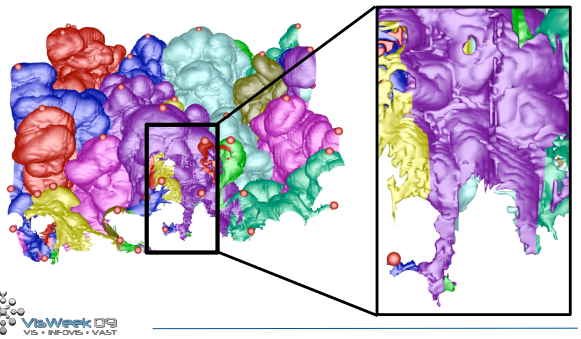
- When do two bubbles should be considered the same ?
  - If the surface needs to only minimally change to merge them



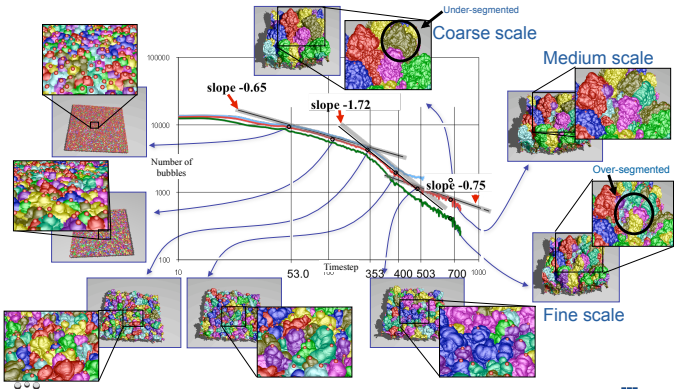
⇨ Persistence

Global Sub-Selection Necessary ?

- Surface artifacts create inverted maxima
  - Artificially reduce their persistence depending on surface normal

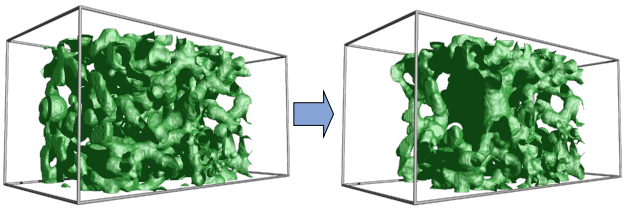


Topological Analysis Reveals The Four Distinct Stages of The Mixing Process



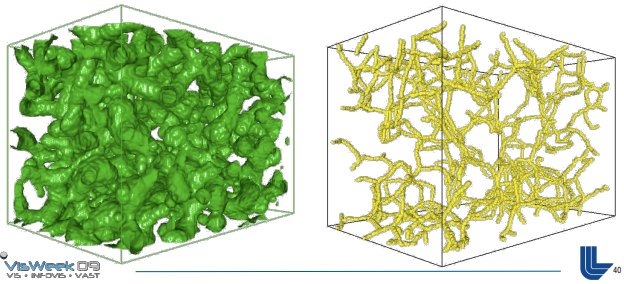
Quantitative Analysis of the Impact of a Micrometeoroid in a Porous Medium

- How does the structure change?
- How does the density profile change?
- What is the loss in porosity of the material?



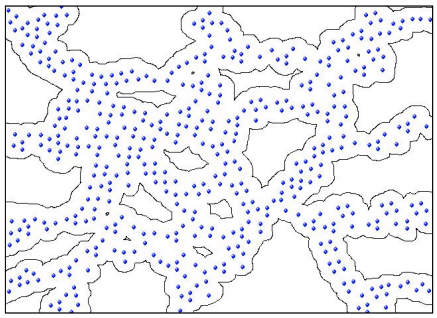
Quantitative Analysis of the Impact of a Micrometeoroid in a Porous Medium

- How does the structure change?
- How does the density profile change?
- What is the loss in porosity of the material?
- The material is best described by its filament structure



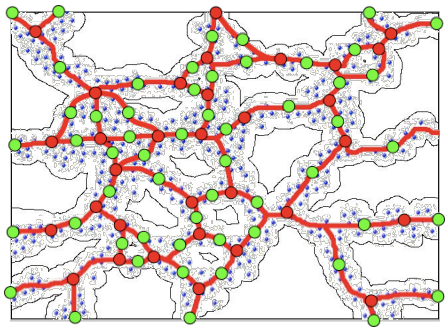
Select Topological Structure to Extract / Analyze Filaments

- How can we define / analyze the skeleton structure ?



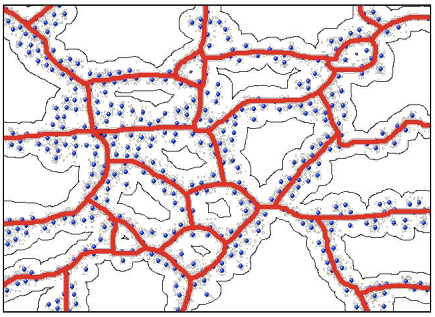
Select Topological Structure to Extract / Analyze Filaments

- How can we define / analyze the skeleton structure ?



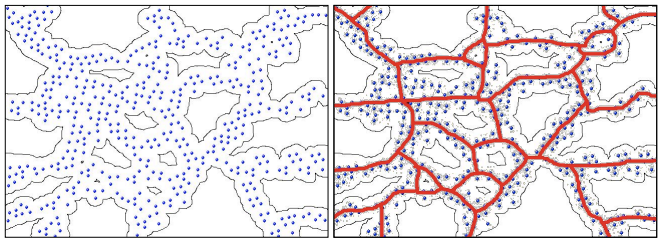
Select Topological Structure to Extract / Analyze Filaments

- How can we define / analyze the skeleton structure ?



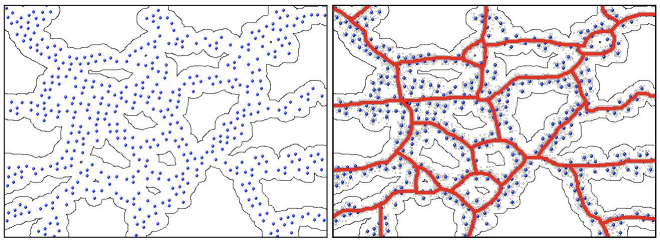
Select Topological Structure to Extract / Analyze Filaments

- How can we define / analyze the skeleton structure ?
  - Lines of “maximal” atomic density are ascending one-manifolds of the Morse-Smale complex



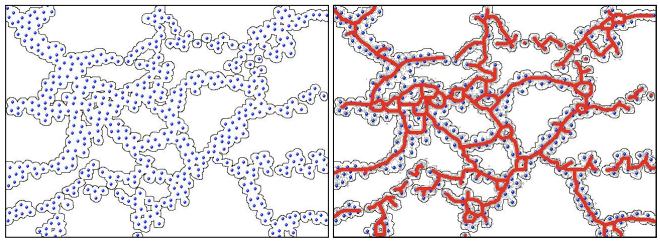
How to Pick Appropriate Threshold ?

- The saddle-maxima skeleton is strongly dependent on the atomic density threshold



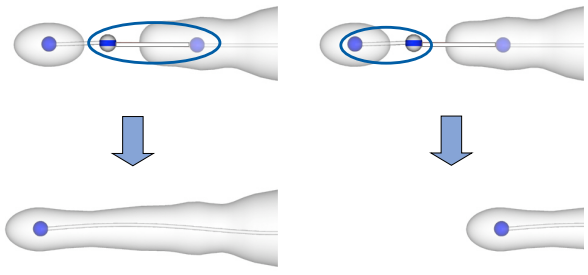
How to Pick Appropriate Threshold ?

- The saddle-maxima skeleton is strongly dependent on the atomic density threshold

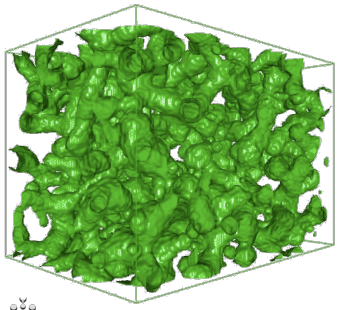


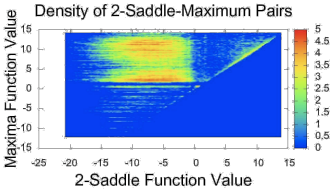
Topological Simplification Enables Local Density Thresholds

- Saddle-Maxima cancellations simulate local changes in threshold

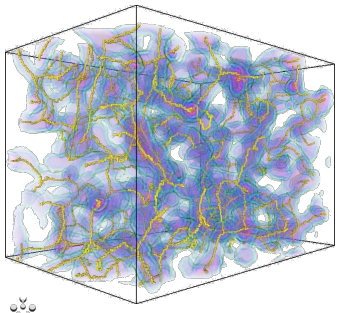


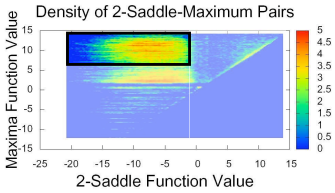
Histograms of All Possible Local Thresholds Provide Stable Thresholds for Analysis



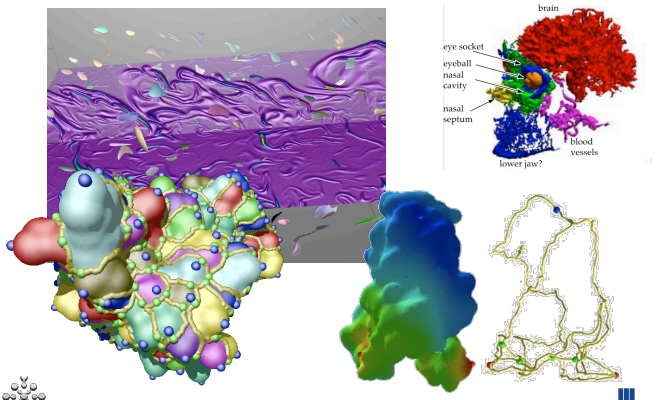


Histograms of All Possible Local Thresholds Provide Stable Thresholds for Analysis

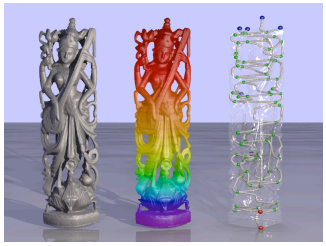


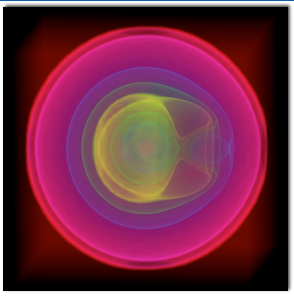


Other Examples of Topological Analysis



Other Examples of Topological Analysis





The Topological Framework Provides Several Advantages

- Strong theoretical foundation to precisely define a broad set of features.
- Robust and efficient combinatorial algorithms to extract features.
- Reliable results independent of noise and/or no-smooth functions.
- Flexible feature-based hierarchies amenable to different metrics
- Compact representations of parameter dependent feature families leading to
  - Powerful and flexible analysis
  - Interactive visualization
  - Data compression

General analysis framework easily adaptable to different applications.

References

- *Understanding the Structure of the Turbulent Mixing Layer in Hydrodynamic Instabilities*, Laney et al., IEEE Transactions on Visualization and Computer Graphics Vol. 12, No. 5, pp. 1053-1060, 2006.
- *Robust On-line Computation of Reeb Graphs: Simplicity and Speed*, Pascucci et al., ACM Transactions on graphics, pp. 58.1-58.9, 2007.
- *Turbulence effects on cellular burning structures in lean premixed hydrogen flames*, M. Day, J. Bell, P.-T. Bremer, V. Pascucci, V. Beckner, and M. Lijewski,, Combustion and Flame, 156pp.1035-1045, 2009.
- *Analyzing and Tracking Burning Structures in Lean Premixed Hydrogen Flames*, P.-T. Bremer, G. Weber, V. Pascucci, M. Day, and J. Bell,, IEEE Transactions on Visualization and Computer Graphics, 2009, to appear.
- *Topologically Clean Distance Fields*, A. Gyulassy, M. Duchaineau, V. Natarajan, V. Pascucci, E.Bringa, A. Higginbotham, and B. Hamann,, IEEE Trans. Comp. Graph. Vis., 13(6), pp.1432--1439, 2007.

References

- *Feature Tracking Using Reeb graphs*, G. Weber, P.-T. Bremer, J. Bell, M. Day, and V. Pascucci,, Proc. 3rd of TopoInVis Workshop , Springer, 2010.
- *Topological feature extraction for comparison of terascale combustion simulation data*, A. Mascarenhas, R. W. Grout, P.-T. Bremer, E. R. Hawkes, V. Pascucci, and J.H. Chen,, Proceedings of the 3rd TopoInVis Workshop, Springer, 2010.
- *Segmenting Molecular Surfaces*, V. Nataranjan, Y. Wang, P.-T. Bremer, V. Pascucci, and B. Hamann,, Computer Aided Geometric Design, 6(23), pp. 495-509, 2006.
- *Simplifying Flexible Isosurfaces Using Local Geometric Measures*, H. Carr, J. Snoeyink, and M. van de Panne,, pp. IEEE Visualization '04, 497--504, 2004.
- *Topology-Controlled Volume Rendering*, G. H. Weber, S. E. Dillard, H. Carr, V. Pascucci, and B. Hamann,, IEEE Trans. Vis. Comp. Graph., 12(2), pp.330-341, 2007.
- *Topological Framework for the Interactive Exploration of Large Scale Turbulent Combustion*, P.-T. Bremer, G. H. Weber, J. Tierny, V. Pascucci, M. Day, and J. B. Bell, A, pp. Proc. {IEEE} International Conference on e-Science, to appear, 2009.