# Topological Data Analysis at the Extreme Scale


## Visualization Group

Gunther H. Weber GHWeber@lbl.gov  Dmitriy Morozov Dmorozov@lbl.gov

BERKELEY LAB

CRD COMPUTATIONAL RESEARCH DIVISION


## Motivation

### Challenges

- Complexity of simulation results already exceeds data analysis capabilities
- Gap between simulation complexity and data analysis capabilities likely to grow
- Aggressive improvements in data analysis necessary to derive new insights from future simulations

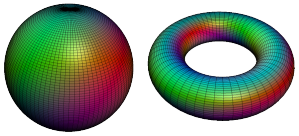
### Topological Methods at the Extreme Scale

- Topological methods provide general feature definitions applicable for wide range of applications
- Develop algorithms implementing topological analysis methods on massively parallel architectures
- Determine required computational resources for topology-based analysis of extreme scale simulation; identify scaling bottlenecks and eliminate them
- Perform *in situ* topological data analysis
- Use topological methods for multidimensional and multivariate feature mining
- **Goal:** Provide practical means of applying topology-based data analysis to hero run results

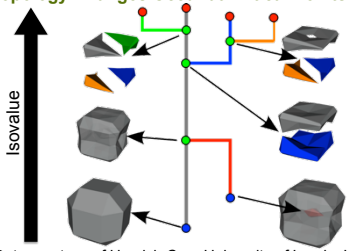
## Background

### Topology of Surfaces

- Properties that remain invariant under elastic deformation
- Topology of compact surface, e.g., defined by:
  - Number of connected components
  - Genus (number of holes)



### Isosurface Topology Changes Occur at Critical Points

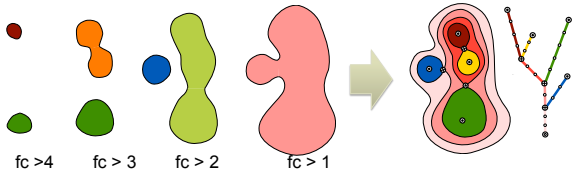



(Data courtesy of Hamish Carr, University of Leeds, UK)

## Acknowledgements


This work is supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 through the "Scientific Data Management and Analysis at Extreme Scale" (DE-FOA-0000256) program.

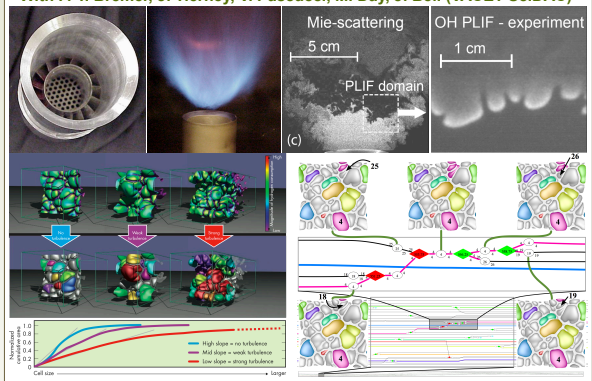

## Merge Trees




- Merge tree tracks all connected regions in threshold-based segmentation (when are they created, when do they merge)
- Efficient calculation of derived quantities (size, average temperature) for multiple classification thresholds

## Application 1: Analysis of Low-Swirl Turbulent Combustion

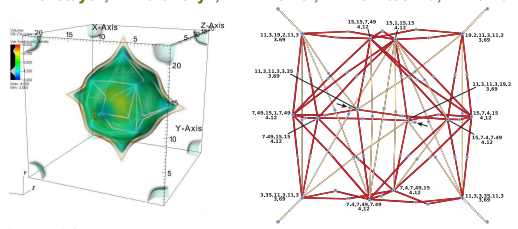
**With P.-T. Bremer, J. Tierney, V. Pascucci, M. Day, J. Bell (VACET SciDAC)**




- Low swirl burners produce stable lifted flame
  - Burns more fuel efficient, produces fewer emissions
  - Does not interact with burner
- Scientists characterize the combustion process via burning cells defined as regions of high fuel consumption
  - How many cells exist at a given time ?
  - What are their sizes and other properties, e.g. average temperature ?
  - How do they evolve over time?
- No pre-set the fuel consumption threshold exist, analyzing data with various thresholds provides important information
- Topological data analysis provides effective means to compute properties for families of threshold-based segmentations
- Tracking burning regions provides insight into their evolution

## Application 2: Analysis of Transformation Paths in Chemical Systems

**With K. Beketayev, M. Haranczyk, P.-T. Bremer, M. Hlawitschka, B. Hamann**



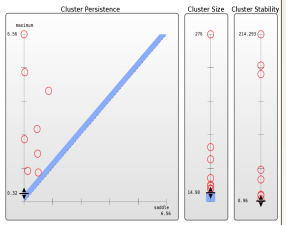

**Energy function of $CH_4$ molecule in LTA zeolite and lowest energy paths connecting neighboring minima.**

- Transformation in chemical systems of fundamental interest
- Energy as function of coordinates of system components
- Lack of effective high-dimensional visualization techniques limits analysis to one or two coordinates at a time
- Comprehensive understanding requires showing relationship between all coordinates at the same time
- Combine concepts from topological analysis, multidimensional scaling and graph layout
- Enable energy function analysis for wide range of molecular structures
- Provide chemists with important tool for understanding complex reactions

## Application 3: Topological Analysis of High-dimensional Density Clustering

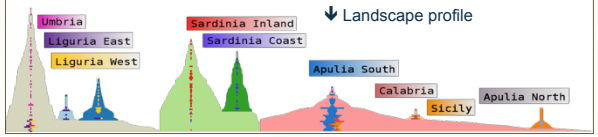
**With P. Oesterling, C. Heine, G. Scheuermann**

- **Input**: Points in n-dimensional space
- **Shown example:** Eight dimensions = Percentages of eight fatty acids in Italian olive oils.
- Interested in clusters of high density
- Use merge tree to represent density function and identify clusters
- Present results as landscape profile with hills showing clusters
- Link to views showing simplification parameters
- Couple with parallel coordinate views and dimension reduction approaches




↑ Simplification parameters (persistence, cluster size and cluster stabilty)

↓ Landscape profile





U.S. DEPARTMENT OF ENERGY | Office of Science