Lawrence Berkeley National Laboratory
Lawrence Berkeley National Laboratory

Title
Topology-based Visualization and Analysis of High-dimensional Data and Time-varying Data at the Extreme Scale

Permalink
https://escholarship.org/uc/item/5672m5d5

Author
Weber, Gunther H.

Publication Date
2012-04-09

eScholarship.org Powered by the California Digital Library
University of California

Topology-based Visualization and Analysis of High-dimensional Data and Time-varying Data at the Extreme Scale

Project Members: Gunther H. Weber (LBNL), Dmitriy Morozov (LBNL)

Collaborators (projects on poster): Kenes Beketayev (UC Davis), John Bell (LBNL), Peer-Timo Bremer (LLNL), Marc Day (LBNL), Bernd Hamann (UC Davis), Christian Heine (ETH Zürich), Maciej Haranczyk (LBNL), Mario Hlawitschka (U of Leipzig), Valerio Pascucci (U of Utah), Patrick Oesterling (U of Leipzig), Gerik Scheuermann (U of Leipzig)

Exascale/Big Data 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
- *Important Approach:* General methods to extract features and analyze them

Research Agenda

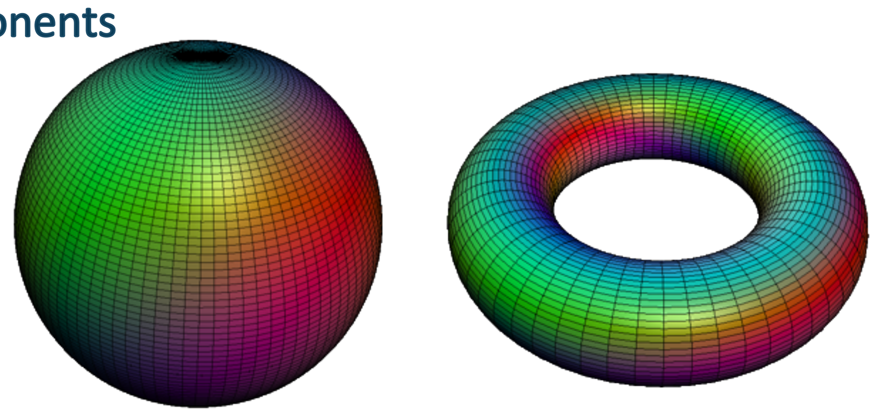
- Topological methods provide general feature definitions applicable to wide application range
- Extend and improve these techniques:
  - Implement topological methods at the extreme scale
  - Perform in-situ topological data analysis
  - Enable multidimensional and multivariate feature mining using topological methods

Topological Methods at the Extreme Scale

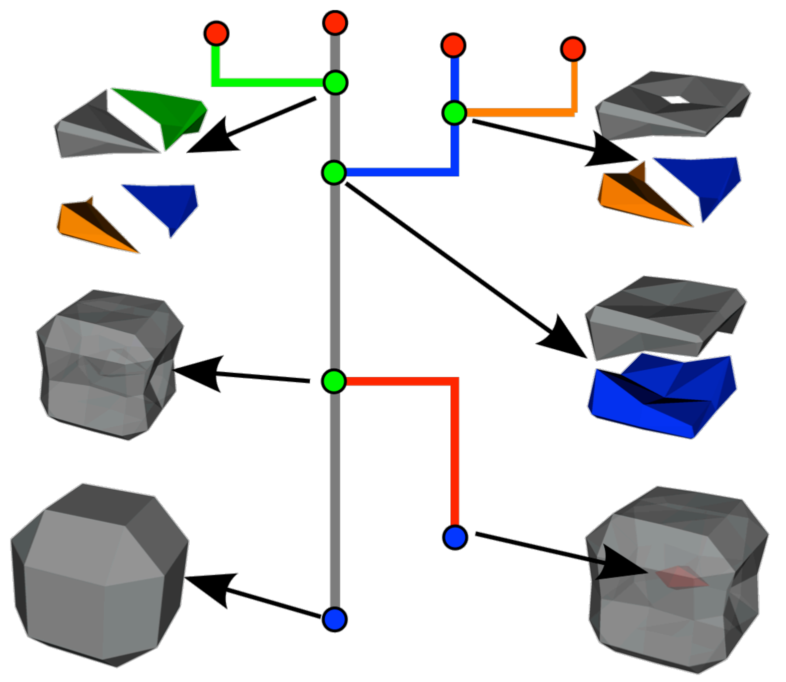
- Develop algorithms that implement current (3D) topological analysis methods on massively parallel architectures
- Determine required computational resources for topological analysis of extreme scale simulation
- Identify bottlenecks in scaling and address them
- **Goal:** Provide practical means of applying topological data analysis to hero run results

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)



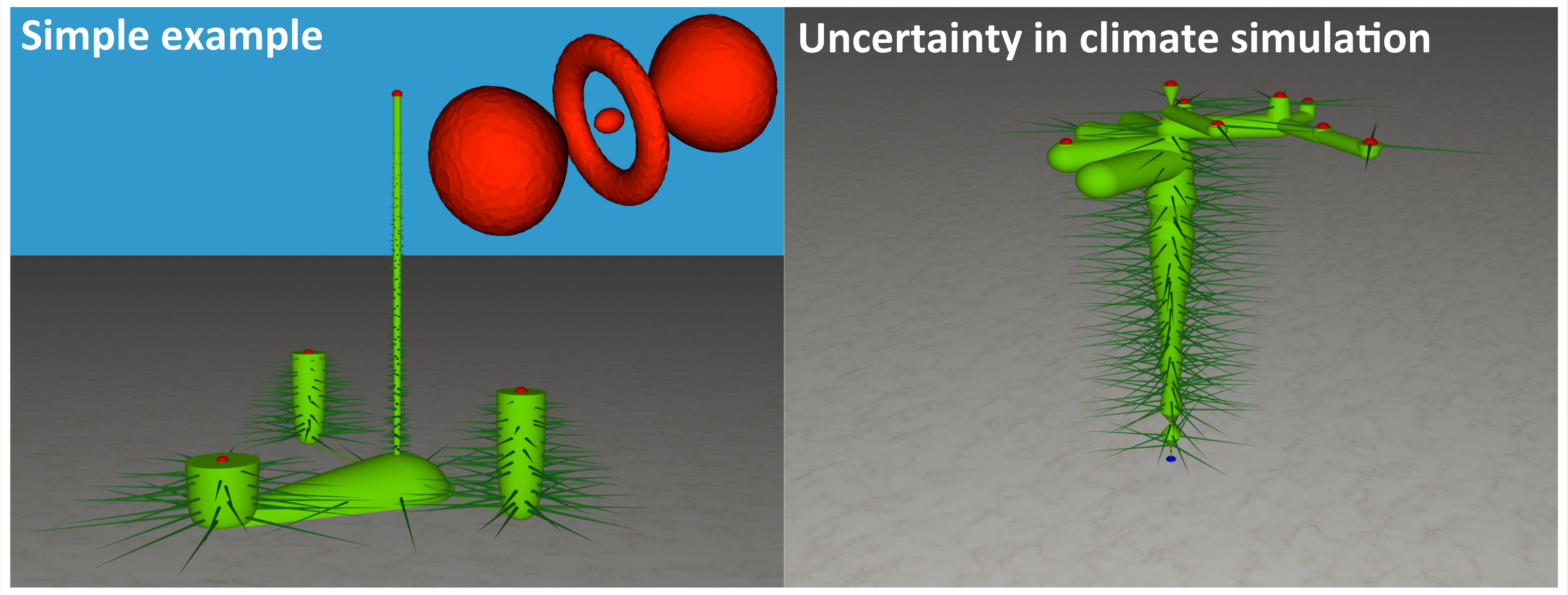
Contour Tree Encodes Isosurface Topology Changes



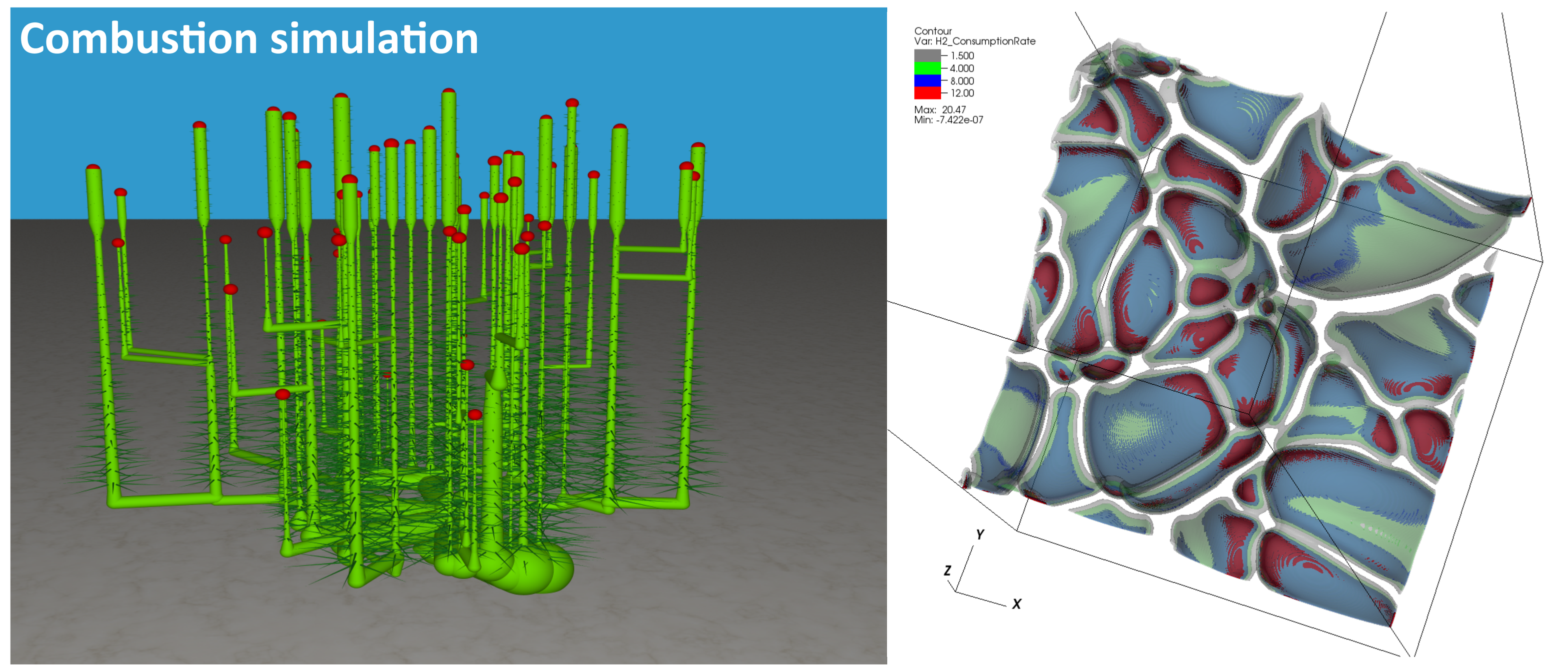
(Data courtesy of Hamish Carr, University of Leeds, UK)

Presenting Structural and Quantitative Information

- Isosurface extraction is versatile visualization and analysis method
- Common approaches provide two complementary types of information
- Structural information, e.g., number of connected components ➔ Contour tree
- Quantitative information, e.g., surface area, volume ➔ Contour spectrum
- **Goal:** Combine in single visualization ➔ Topological Cacti



- **Simple example:** Spatial probability distribution of the electron in an hydrogen atom, residing in a strong magnetic field. Branch width corresponds to cube root of volume and spike length to the square root surface area.
- **Uncertainty in climate simulation:** Average longwave energy output (FLUT) in the winter months of a 21 dimensional climate simulation ensemble comprising 1197 climate simulation. The width corresponds to cube root of vertex count–"volume" in parameter space. Spike length shows mean value of FLUT.



- **Combustion simulation:** Fuel consumption rate in a combustion simulation (one 4.3G time step from a coarse AMR simulation). We are interested in how the size of burning regions correlates with temperature variance. Thus, we map the cube root of volume – approximated as the number of vertices – to branch width and temperature variance to spike length.

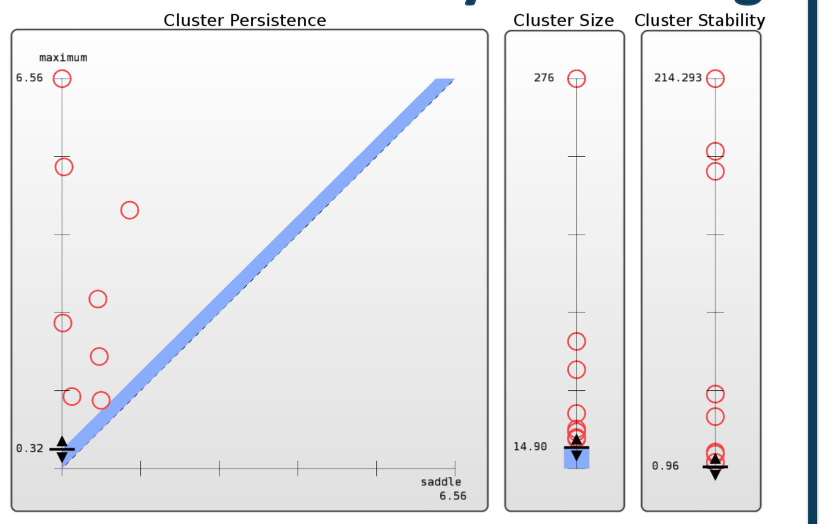
(Work with P.-T. Bremer and V. Pascucci, Combustion data: John Bell, Marc Day)

Multidimensional and Multivariate Feature Mining

- Utilize topological structures to define features in high-dimensional data
- Derive quantities (such as size distribution) from features
- Define visualization methods based on topological structures (graphs) and derived quantities
- **Goals:**
  - Show applicability of topological analysis to high-dimensional data
  - Identify commonalities in different areas ➔ general feature definitions

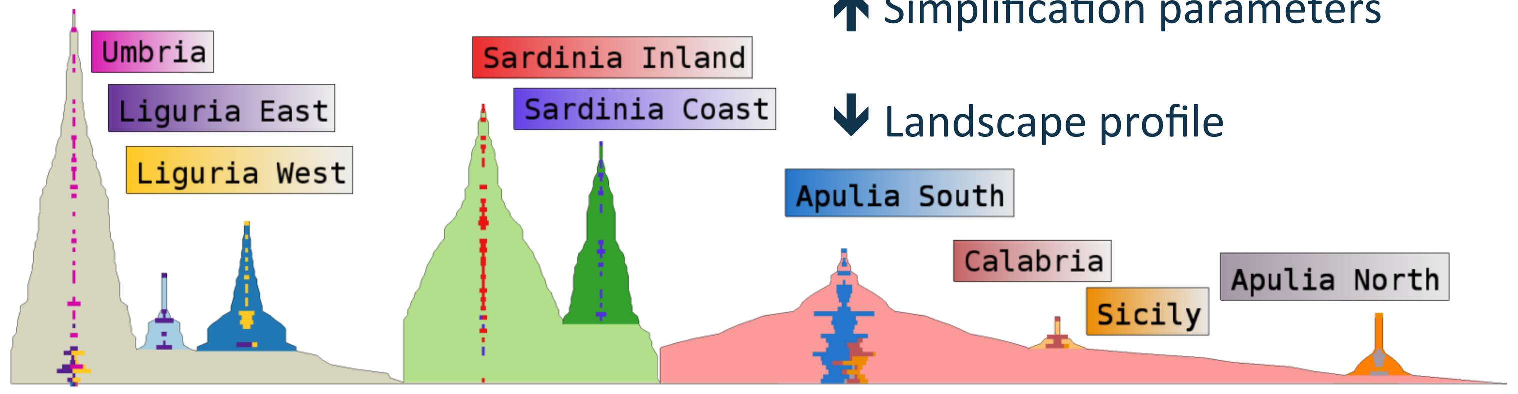
Topological Analysis of High-dimensional Density Clustering

- **Input:** Points in n-dimensional space
- Interested in clusters of high density
- Use merge tree to represent density function and identify clusters
- Present results as landscape profile
- Link to views showing simplification parameters
- Couple with parallel coordinate views and dimension reduction approaches



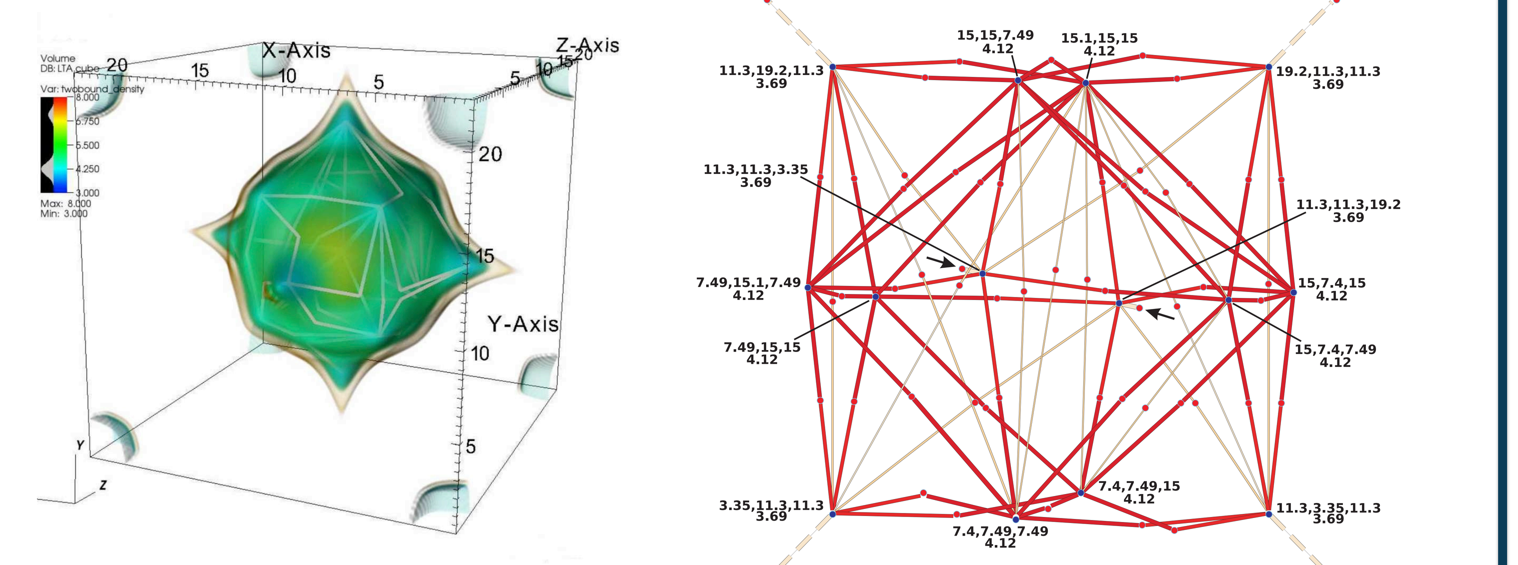
🡑 Simplification parameters

🡓 Landscape profile



(Work with Patrick Oesterling, Christian Heine and Gerik Scheuermann)

Analysis of Transformation Paths in Chemical Systems



- Transformation in chemical systems of fundamental interest
- Consider 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

(Work with K. Beketayev, M. Haranczyk, P.-T. Bremer, M. Hlawitschka, B. Hamann)

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.
- We thank the members of the
  - Visualization and Analytics Center for Enabling Technologies (VACET)
  - Center for Computational Sciences and Engineering (CCSE)
  - LBNL Scientific Computing Group

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.