Data Intensive Analysis Techniques and Tools

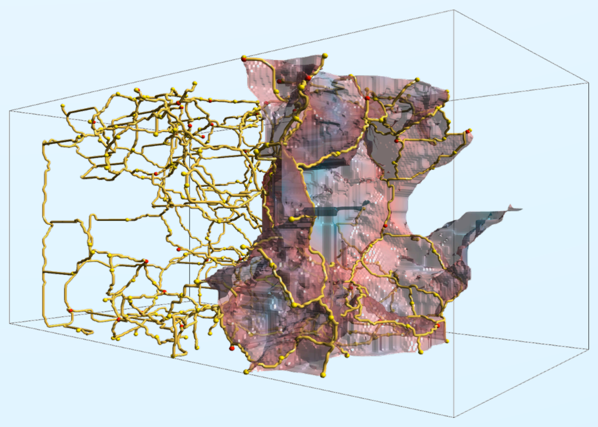
W. Bethel (LBNL) P.-T. Bremer (LLNL/Utah),), H. Childs (LBNL), A. Gyulassy (Uath), K. Joy (UCD)
V. Pascucci (Utah-PNNL), T. Peterka (ANL), R. Ross (ANL), O. Rübel (LBNL)

TOPOLOGIKA: Topological Techniques for Scientific Feature Extraction

Extracting features of interest in a flexible, efficient, and automatic manner is a crucial component to analyze high resolution simulations. In TOPOLOGIKA, we are developing a theoretical framework, algorithms, and software packages to define and extract a wide variety of features based on topological concepts. The resulting concepts are broadly applicable to areas ranging from combustion to climate and from material science to biology.

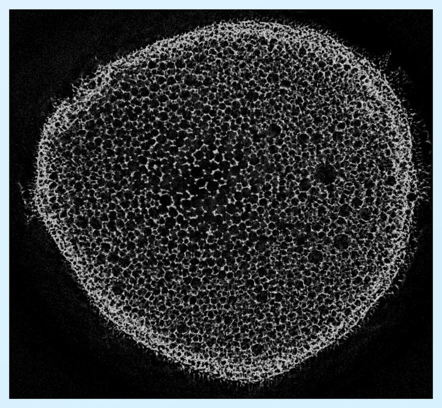
Key features of our approach are:

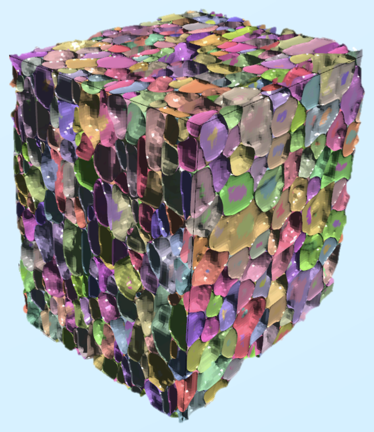
- A unified mathematical framework to express a broad range of different features in a wide range of applications.
- Scalable and efficient algorithms to extract features streaming, in-situ, in-transit, or out-of-core.
- Theoretical guarantees on the robustness and correctness as well as multi-resolution techniques to address noise and support multi-scale analysis



Screenshot from the **MSCEER** tool, showing interface surface of an impact into a porous medium as represented by steepest ascending surfaces between outside and inside. Skeleton of porous structure represented as steepest ascending lines in the density field.

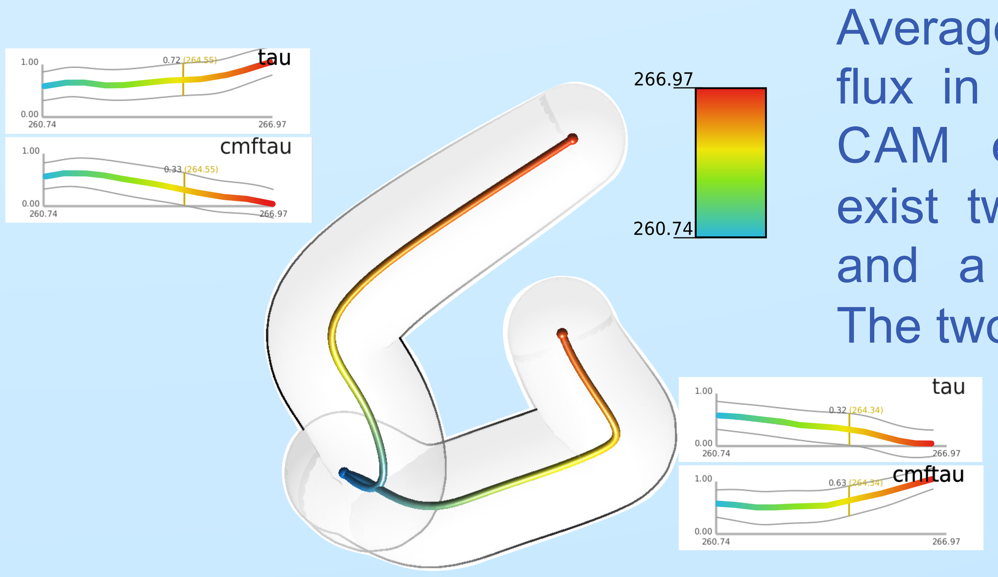
Foam pockets extracted in **MSCEER** as three-dimensional basins around density minima. Extracted from an industrial CT scan.





High-Dimensional Analysis

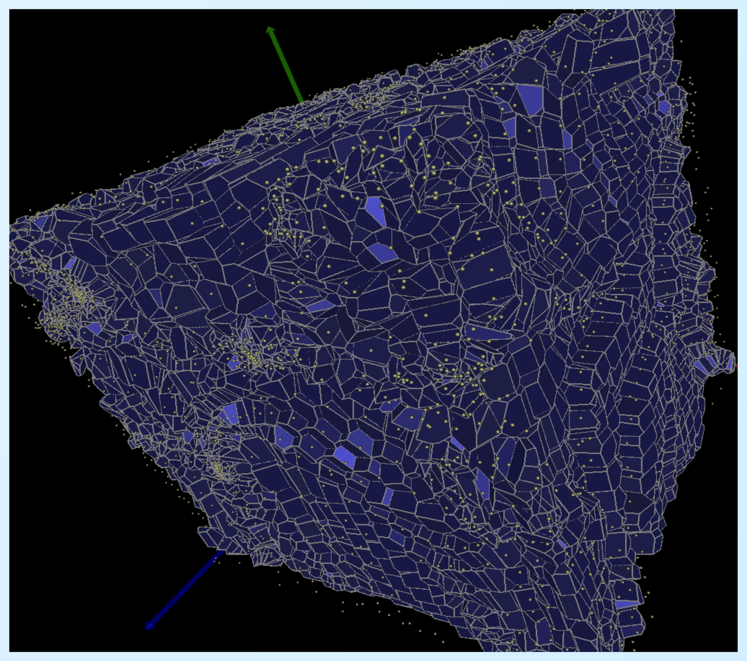
Many applications are interested in high-dimensional and non-spatial data. Traditional examples are optimization metrics, multi-dimensional PDFs, or simulation ensembles. The challenge is that few existing techniques apply to higher dimensions and the results are often unintuitive. Extending topological concepts to high dimensions provides us with reliable means to extract and analyze features as well as new visual metaphors to illustrate the results.



Average upward clear sky flux in a 21 dimensional CAM ensemble. There exist two strong maxima and a global minimum. The two crystals show an Inverse relation of tau and cmftau which govern deep and shallow convection.

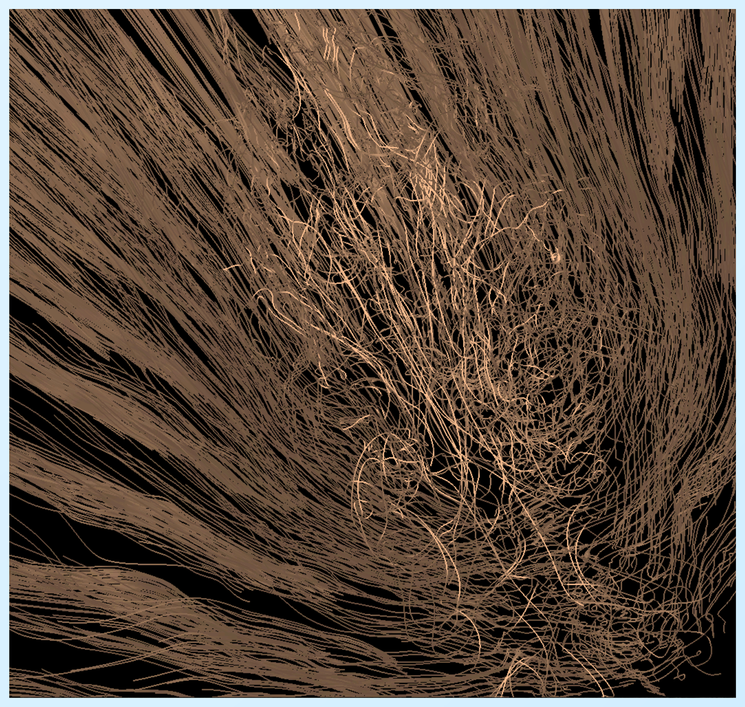
Parallel Voronoi Tessalation

Voronoi tessalations allow scientists to derive a dense mesh representation from a sparse N-body particle simulation. This is especially useful, for example, for cosmological data like the one shown, here as it self-adapts to widely varying particle distributions.



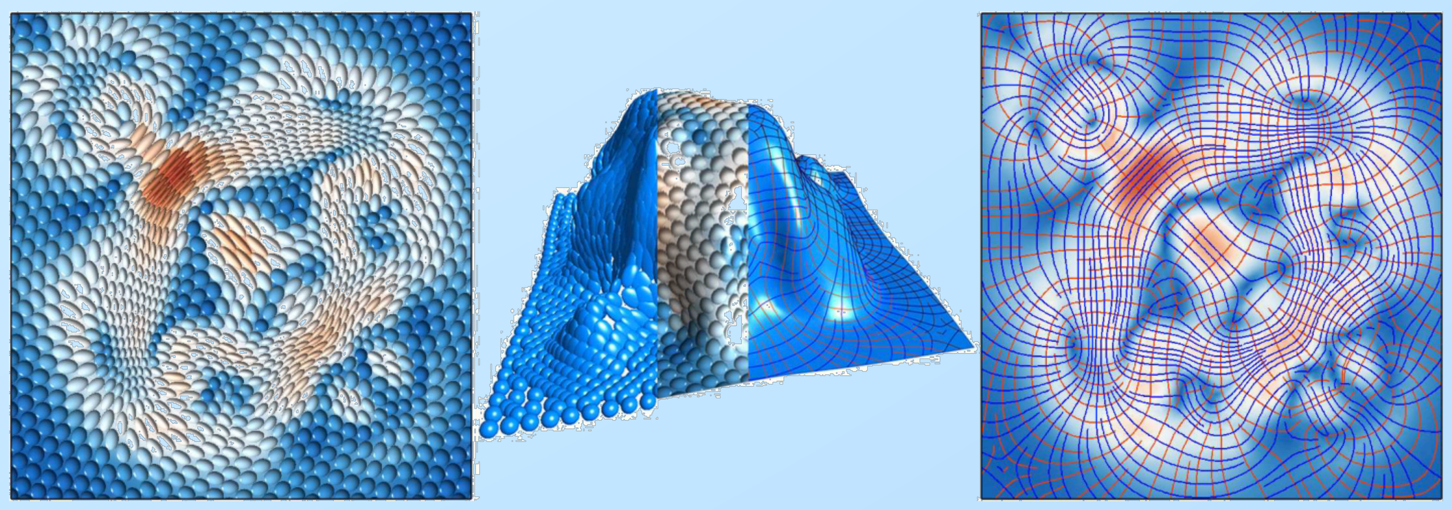
Parallel Particle Tracing

Particle tracing for flow visualization is used to generate streamlines and pathlines. It is difficult to parallelize efficiently because of widely varying communication. We use a mix of in core and out of core processing of a time series coupled with a multiblock assignment, early particle termination, and adjustable synchronization communication.



Metric Tensor Enhancement for Flow Surface Analysis

Flow surfaces are an effective data analysis technique that generalizes the computation of integral lines and provide new powerful insights in vector field structures. Basic flow surface analysis focuses on displacement information, neglecting the more complex notion of deformation in the flows, such as shearing and stretching properties that are central to the field of continuum mechanics. New use of a metric tensor field allows encoding local surface deformations as induced by the velocity gradient of the underlying flow field. The properties of the resulting approach enhance the analysis of the flow surface generated providing new fundamental structural insights to the users.

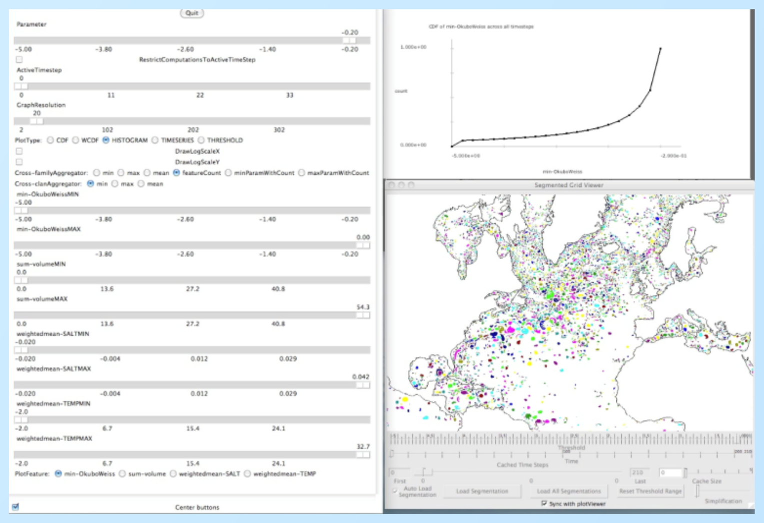


TALASS: Feature-Based Analysis and Tracking

Given today's high fidelity simulations the trend is to concentrate on understanding how individual feature, e.g. vortices, ignition kernel, or eddies, behave and are formed. In particular, one might be interested in their spatial distributions or dimensions or how they correlate with physical parameters, i.e. temperature, or chemical species. However, extracting features with traditional techniques is time-consuming and resource intensive preventing a thorough exploration of the solution space.

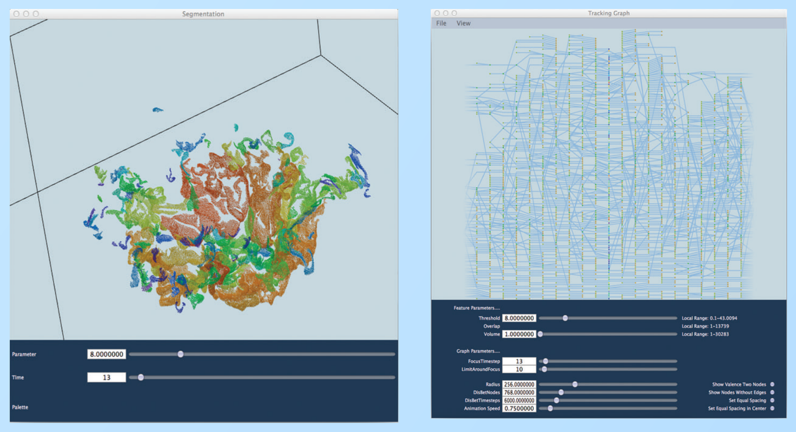
Using a single massively parallel pre-computation we allow a flexible and interactive analysis of a wide variety of features that includes:

- Changing feature parameters, e.g. iso-values or scale parameters, interactively
- Creating feature based statistics such as conditional PDFs
- Exploring feature evolution through time-tracking



Screen capture of the TALASS tool showing interactively extracted eddies from an ocean simulation. (Left) Subselection interface allowing to filter features by attributes. (Top right) Global distribution of Okubo-Weiss parameter for all features. (Bottom right) Three-dimensional Feature display

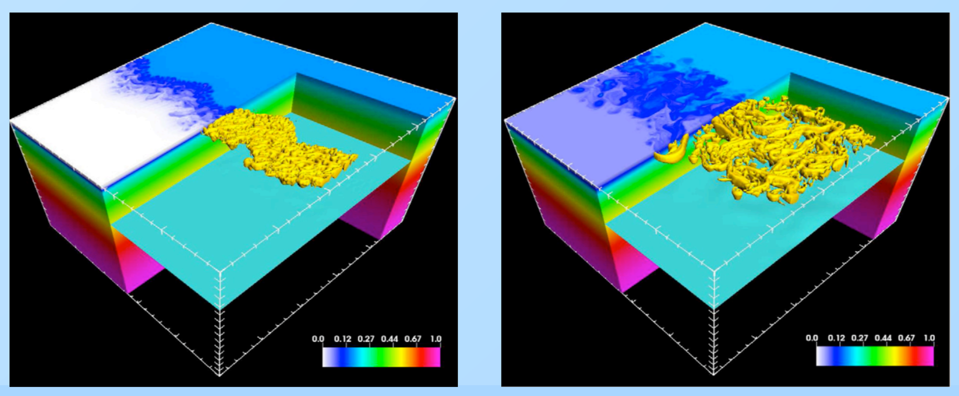
Interactive tracking of burning cells in a combustion simulation. (Left) Feature display; (Right) Interactively computed tracking graph of all features including subselection interface and parameter selection



Parallel Computation of Finite Time Lyanopuv Exponents

Finite Time Lyanopuv Exponents (FTLE) constitute an essential data analysis tool that measures the extent that a local flow separates or remains compact. It is used to find coherent structures in vector fields. It is a compute intensive technique and its parallel computation is essential to tackle the large scale data problem of modern science.

The example below shows the use of the parallel FTLE capability in VisIt to study a Nek5000 simulation of oil dispersion in the ocean. The findings helped to design a 300 drifter experiment in the Gulf of Mexico that occurred this summer.



SDAV
Scalable Data Management, Analysis, and Visualization

U.S. DEPARTMENT OF ENERGY | Office of Science