# PSA 2013 Workshop: Topological Data Analysis and Visualization for Large-Scale and High-Dimensional Science Discovery


Bei Wang[1], Dan Maljovec[1], Diego Mandelli[2]
Additional Organizers: Valerio Pascucci[1], Peer-Timo Bremer[3]

[1]Scientific Computing and Imaging Institute (SCI), University of Utah
[2]Idaho National Laboratory
[3]Lawrence Livermore National Laboratory


September 22, 2013

# Workshop Overview...

# Workshop Overview...

- Part 1: Large Data Analysis and Visualization for Science Discovery
  - Data processing
  - Topological analysis and visualization
  - Applications: case studies

# Workshop Overview...

- Part 1: Large Data Analysis and Visualization for Science Discovery
  - Data processing
  - Topological analysis and visualization
  - Applications: case studies
- Part 2: Analysis and Visualization of High-Dim Data for Sensitivity Analysis
  - Integrating topology with statistics
  - Demo

# Workshop Overview...

- Part 1: Large Data Analysis and Visualization for Science Discovery
  - Data processing
  - Topological analysis and visualization
  - Applications: case studies
- Part 2: Analysis and Visualization of High-Dim Data for Sensitivity Analysis
  - Integrating topology with statistics
  - Demo
- Part 3: Perspective from a Nuclear Scientist

# Workshop Overview...

- Part 1: Large Data Analysis and Visualization for Science Discovery
    - Data processing
    - Topological analysis and visualization
    - Applications: case studies
- Part 2: Analysis and Visualization of High-Dim Data for Sensitivity Analysis
    - Integrating topology with statistics
    - Demo
- Part 3: Perspective from a Nuclear Scientist
- Part 4: Discussions

# Workshop Overview...

- Part 1: Large Data Analysis and Visualization for Science Discovery
  - Data processing
  - Topological analysis and visualization
  - Applications: case studies
- Part 2: Analysis and Visualization of High-Dim Data for Sensitivity Analysis
  - Integrating topology with statistics
  - Demo
- Part 3: Perspective from a Nuclear Scientist
- Part 4: Discussions
- Part 1 (1.5 hours), Part 2 (1.5 hours), Part 3 (0.5 hours), Part 4 (0.5 hours)

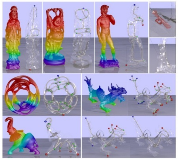
# Center for Extreme Data Management Analysis and Visualization (CEDMAV)

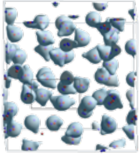
Research summary from members from our research group:
Valerio Pascucci (SCI, CEDMAV director, pascucci@sci.utah.edu)
Peer-Timo Bremer (SCI & LLNL, bremer5@llnl.gov)
Bei Wang (SCI, beiwang@sci.utah.edu)
Attila Gyulassy (SCI, jediati@sci.utah.edu)
Brian Summa (SCI, bsumma@sci.utah.edu)
Many SCI faculties, research scientists, students and collaborators...

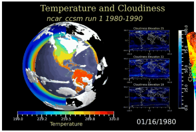
Image/Video Courtesy of Valerio Pascucci

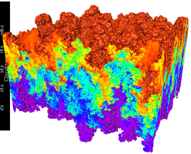
# A Data Analysis and Visualization Center Can be a Catalyst for a Virtuous Cycle of Collaborative Activities

- Tight cycle of: basic research, software deployment and user support
- Coordination among multiple projects: unified techniques for several applications
- Strong University-Lab-Industry collaboration
- Focused technical approach:
  - performance tools for fast data access
  - general purpose data exploration
  - error bounded quantitative analysis
  - feature extraction and tracking



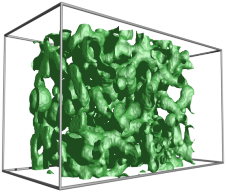


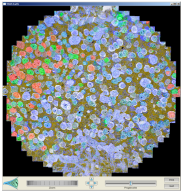


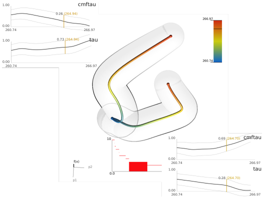



# A Data Analysis and Visualization Center Can be a Catalyst for a Virtuous Cycle of Collaborative Activities

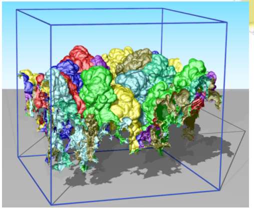
- Interdisciplinary collaboration with domain scientists (from math to physics):
    - motivating the work
    - formal theoretical approaches
    - feedback to specific disciplines

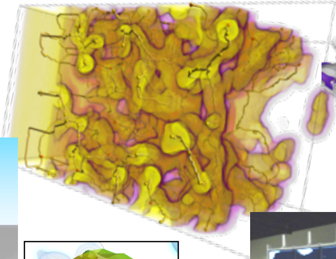


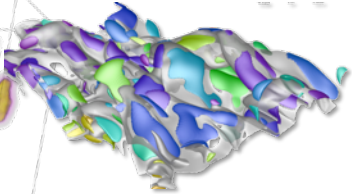


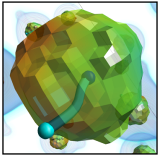


# Part 1: Large Data Analysis and Visualization for Science Discovery

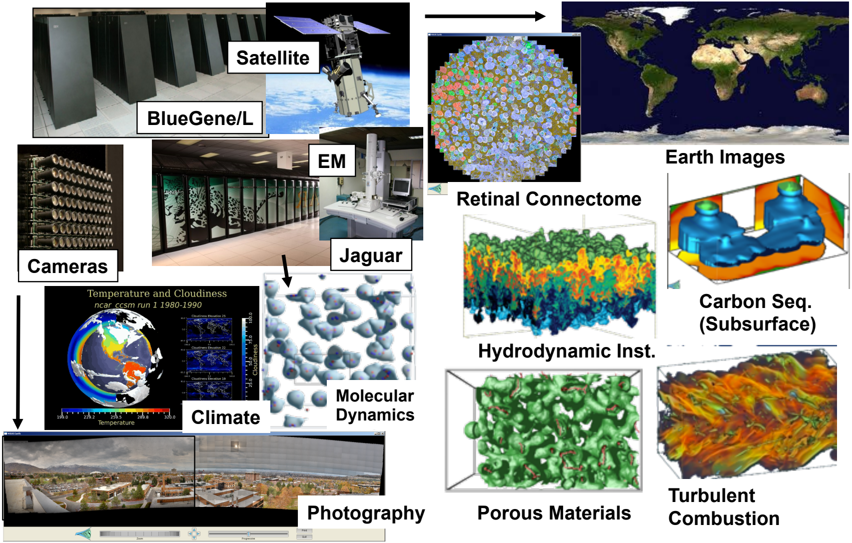






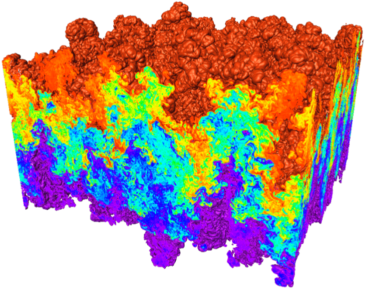


# Massive Simulation and Sensing Devices Generate Great Challenges and Opportunities



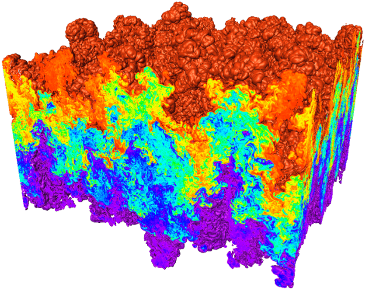
# Traditional Data Analysis Tools are Often Ineffective for Massive Models

- Massive models are challenging, e.g. Rayleigh Taylor instability (instability of an interface between two fluids of different densities that occurs when one of the fluids is accelerated into the other)
  - Sheer volume of info
  - Complexity of the info represented
  - Complexity of presentation



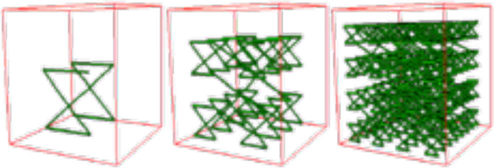
# Furthermore...

- Tools do not scale with the data sizes
- Difficult to capture multiple scales
- Numerical methods unstable and sensitive to noise
- Need proper abstractions and metaphors to convey information reliably and efficiently
- Data Management, Analysis and Visualization are needed in a Unified Environment!



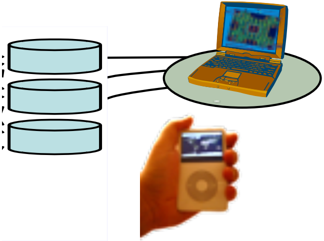
# A Cyber-infrastructure Requires Efficient Data Management and Processing

- Advanced data storage techniques
  - Data re-organization
  - Compression
- Advanced algorithmic techniques
  - Streaming
  - Progressive multi-resolution
  - Out of core computations



# A Cyber-infrastructure Requires Efficient Data Management and Processing

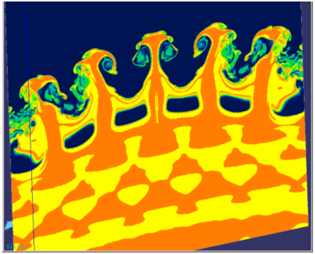
- Scalability across a wide range of running conditions:
  - From laptop, to office desktop, to cluster of PC, to BG/L
  - Memory, to disk, to remote data access

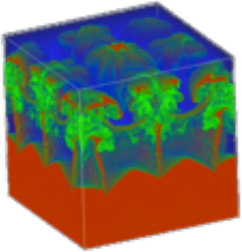


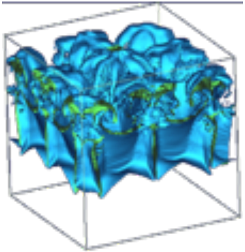
# We Redesigned the Data Management and Visualization Pipeline with New Principles

Basic core techniques:

- Slicing
- Volume rendering
- Iso-surfaces

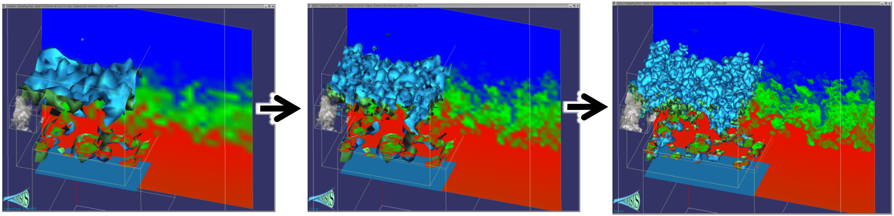




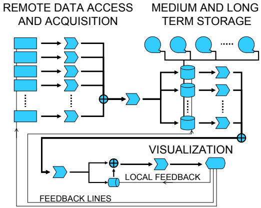


# We Redesigned the Data Management and Visualization Pipeline with New Principles

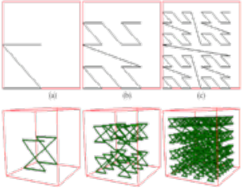
- Cache-oblivious out-of-core processing optimizing access locality for any size of data blocks
- Coarse-to-fine construction of multi-resolution models
- Pipelines of progressive algorithms
- Remote data streaming



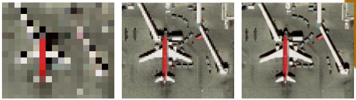
# We Consider the Three Main Components Defining a Computing Infrastructure




**Processing Network**
(Data Access Path)

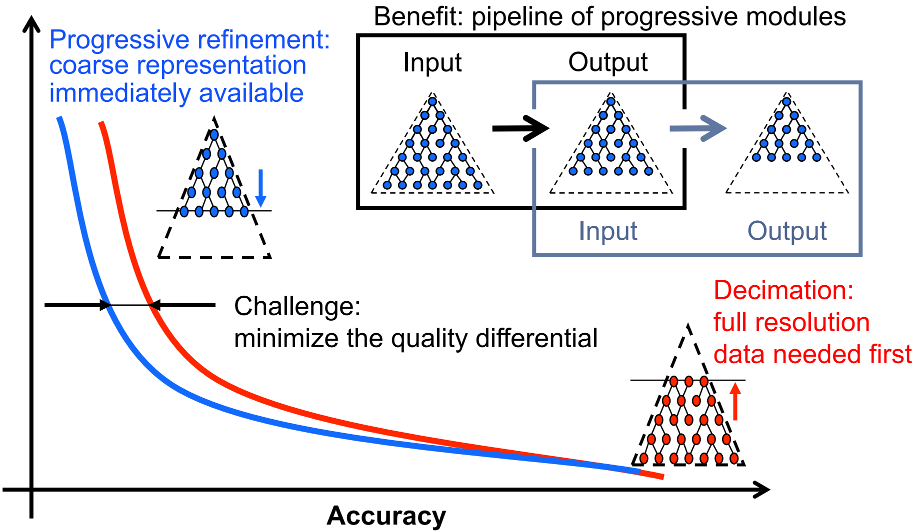



**Data Layout**
(Cache Oblivious)



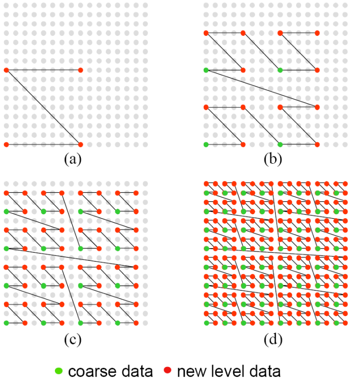
**Algorithm Design**
(Progressive Processing)

# The use of top-down and bottom-up processes have a strong impact on the data stream



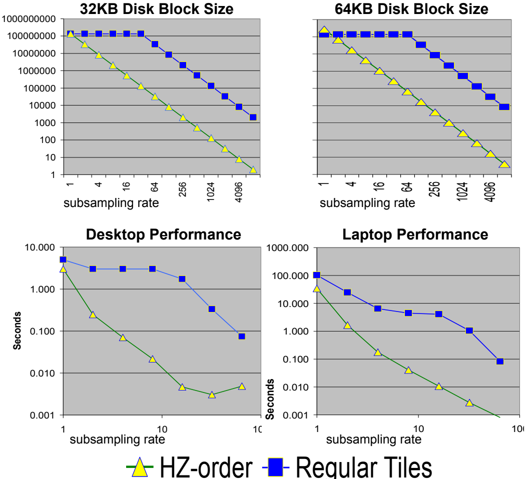
# We Introduced Multi-resolution Cache Oblivious Layouts for Image Data

- Z-order curve used to define a hierarchical sub-sampling over a grid
- Improve access locality:
  - Interleaving hierarchical levels
  - Maintaining geometric proximity
- Data layout is independent of the traversal of the data

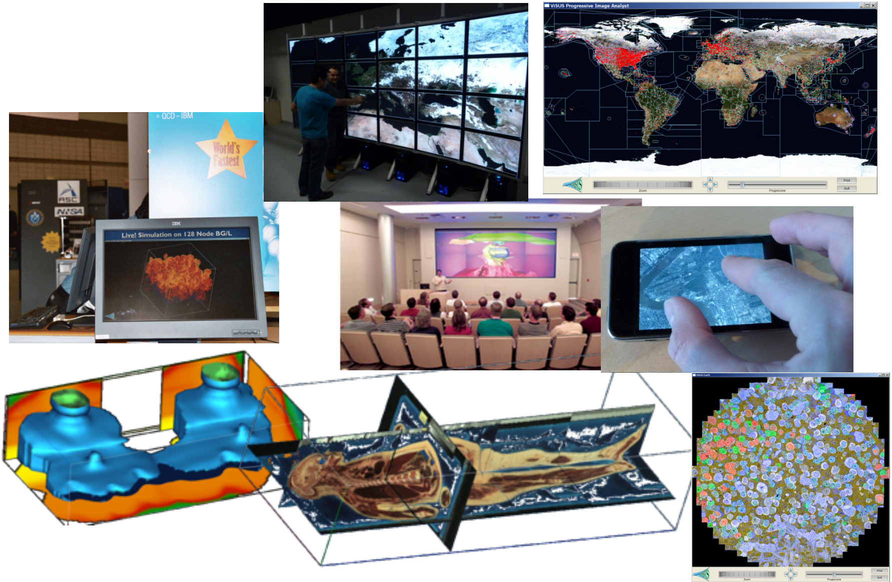


# Cache-Oblivious Data Layouts Scale Well Across Different Storage Blocking Factors

- Formal analysis predicts performance and scalability
- Performance improved by orders of magnitude
- Independence of architecture and storage characteristics



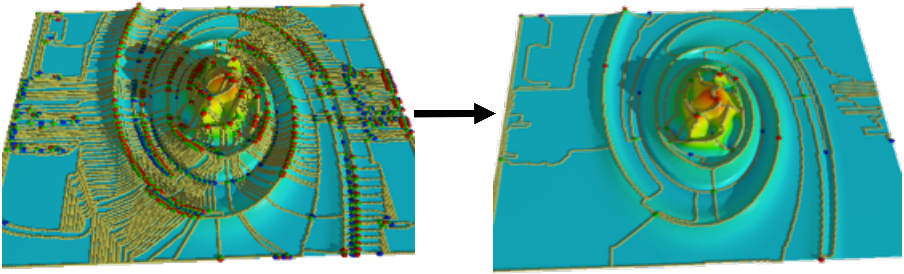
# We Demonstrated Performance and Scalability in a Variety of Applications



Brief Introduction to Topological Data Analysis...

# Topology is an Effective Language to Describe Abstractions of Features from Raw Data

Hierarchical topology of a 2D Miranda vorticity field



# We Adopt Robust Topological Methods to Abstract Features from Raw Data

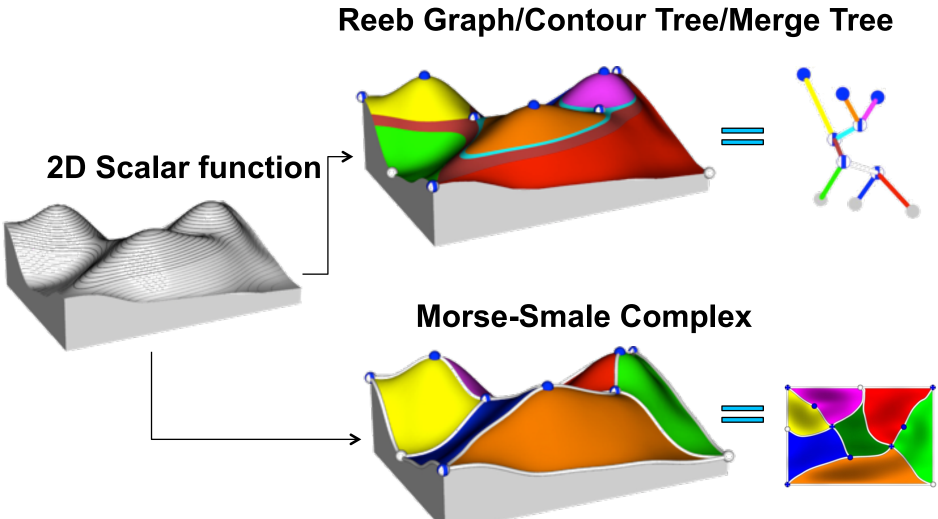
- Provably robust computation
- Provably complete feature extraction and quantification
- Hierarchical structures used to capture multiple scales
- Error-bounded approximations associated with each scale
- Formal definition associated with each analysis
- Streaming techniques to achieve scalable performance

# A really old joke...

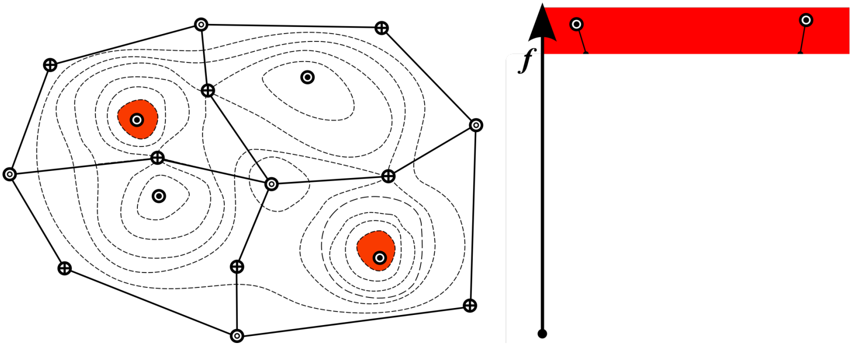
Who thinks the coffee mug and a donut is the same?


# Key development in topological data analysis (TDA)

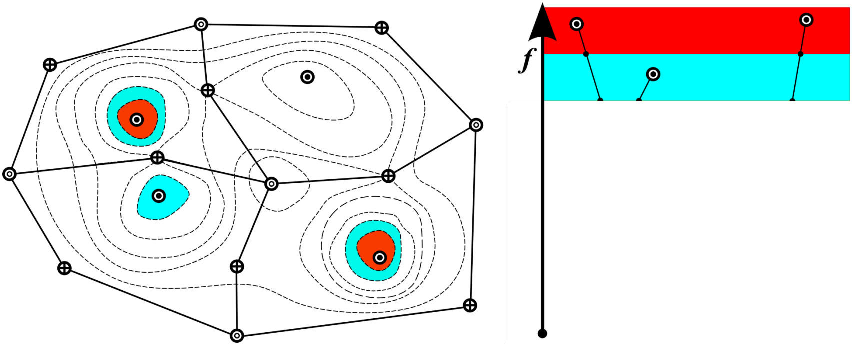
1. Abstraction of the data: topological structures and their combinatorial representations
2. Separate features from noise: persistent homology



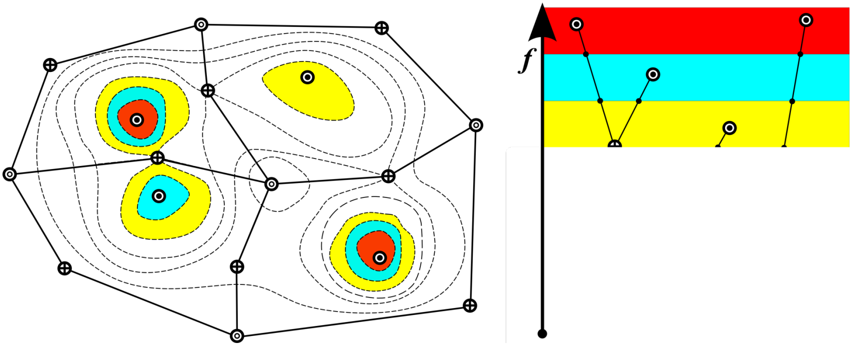
# Contour tree



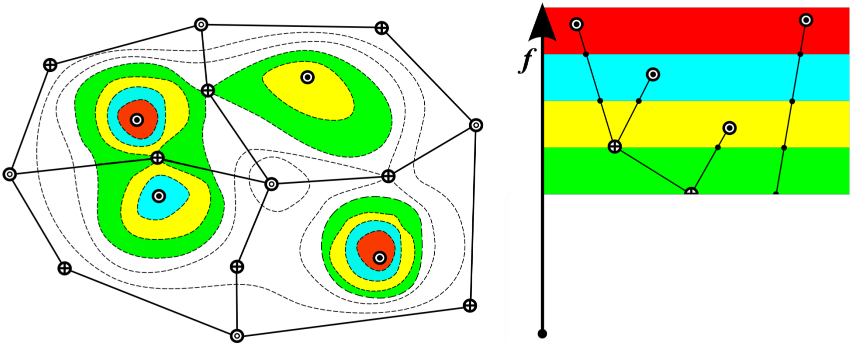
# Contour tree



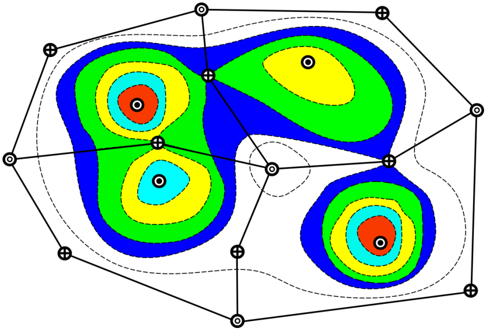
# Contour tree

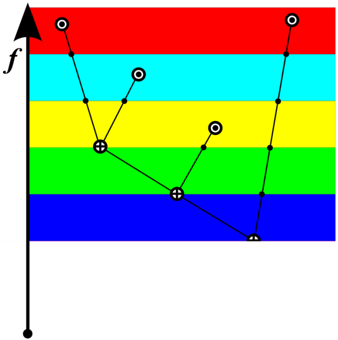


# Contour tree

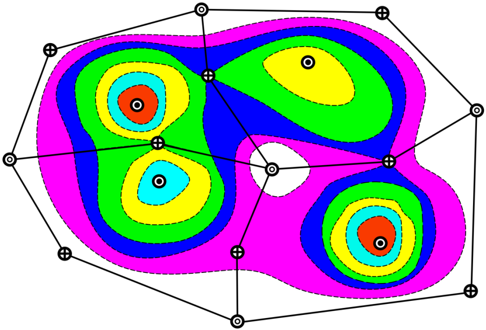


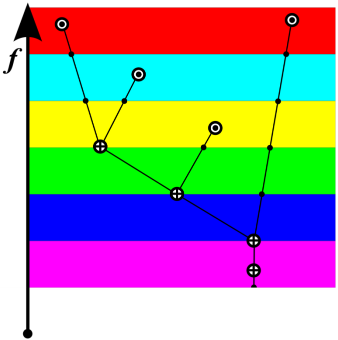
# Contour tree



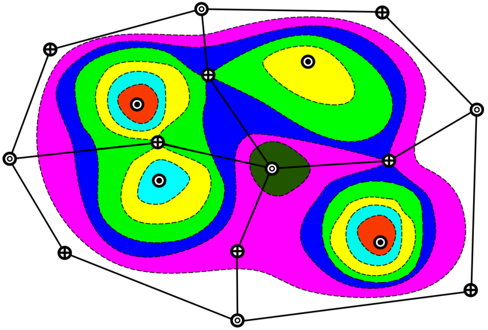


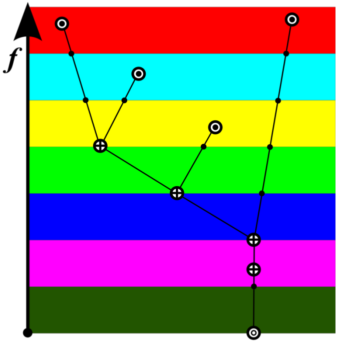
# Contour tree





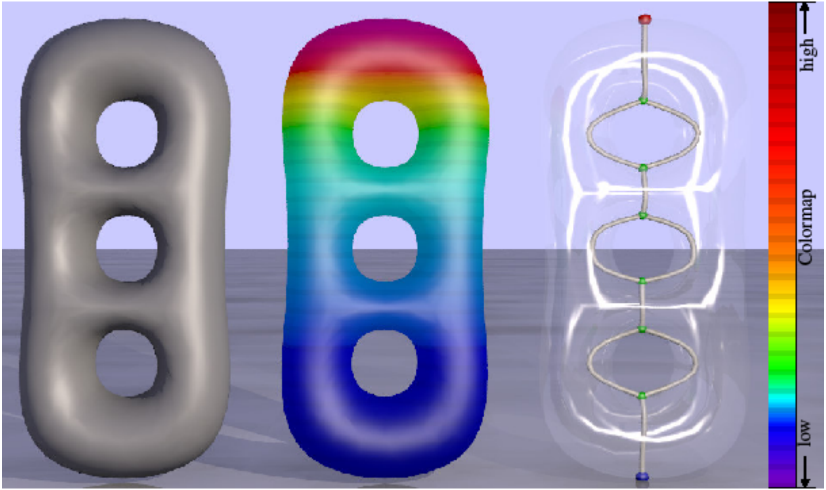
# Contour tree





# Reeb graph

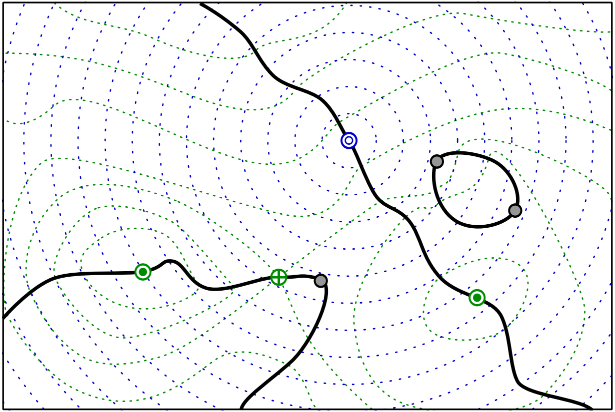
Graph obtained by continuos contraction of all the contours in a scalar field, where each contour is collapsed to a distinct point.




[K. Cole-McLaughlin, H. Edelsbrunner, J. Harer, V. Natarajan and V. Pascucci. Loops in Reeb Graphs of 2-Manifolds. 2004]

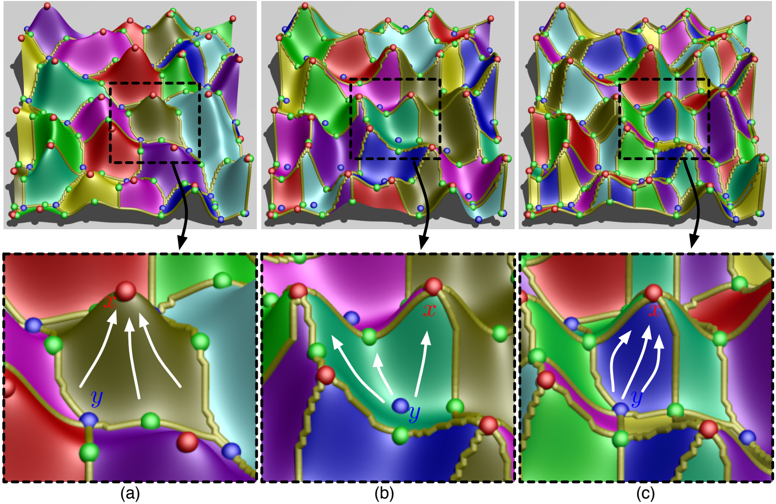
# Jacobi Set

[H. Edelsbrunner and J. Harer. Jacobi sets of multiple Morse functions. 2002]



# Morse-Smale Complex

Partition data into monotonic regions based on gradient flow




Descending Manifolds  Ascending Manifolds  Morse-Smale Complex

# Morse-Smale complex

[P.-T Bremer, H. Edelsbrunner, B. Hamann and V. Pascucci. A Multi-resolution Data Structure for Two-dimensional Morse-Smale Functions. 2003]

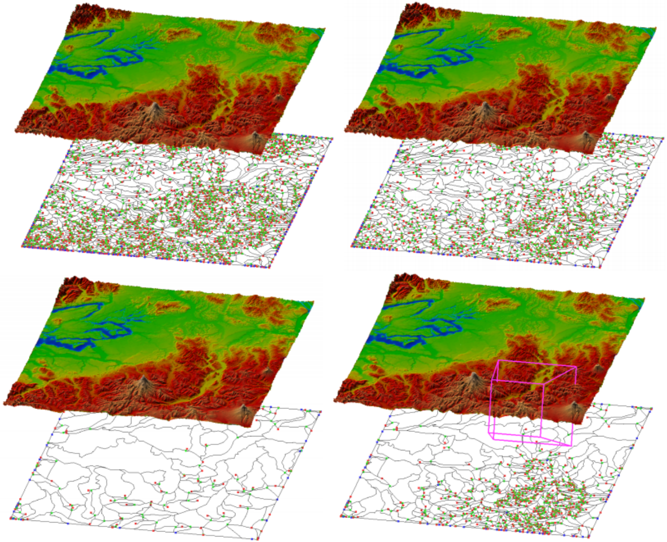
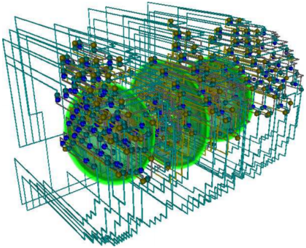


Figure 11: (Upper-left) Puget Sound data after topological noise removal. (Upper-right) Data at persistence of 1.2% of the maximum height. (Lower-left) Data at persistence 20% of the maximum height. (Lower-right) View-dependent re□nement (purple: view frustum).

# Morse-Smale complex

[A. Gyulassy, V. Natarajan, V. Pascucci, P.-T. Bremer, B. Hamann. Topology-based Simplication for Feature Extraction from 3D Scalar Fields, 2005]



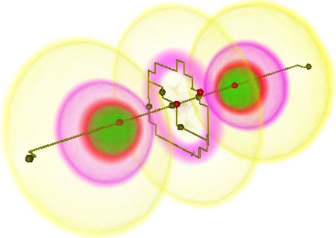
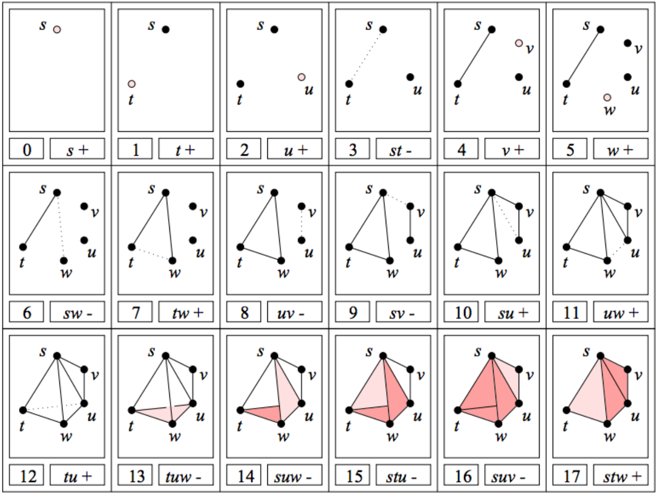


Figure: Topology simplication applied on electron density data for a hydrogen atom: the input has a large number of critical points, several of which are identied as being insignicant and removed by repeated application of two atomic operations. Features are identied by the surviving critical points and enhanced in a volume rendered image by an automatically designed transfer function

# Persistent homology

[H. Edelsbrunner, D. Letscher and A. Zomorodian. Topological persistence and simplification. 2002] [A. Zomorodian, G. Carlsson. Computing Persistent Homology. 2004] Persistence diagram v.s. barcodes and persistence modules.



# Persistent homology

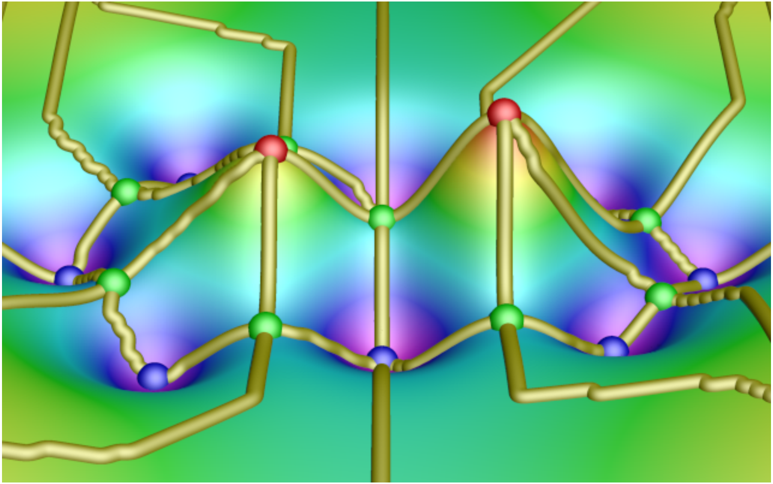
When data is corrupted by noise, how can we tell features from noise?

"The eye, or the brain, performs the marvelous task of taking the sense data of individual points and assembling them into a coherent image of a continuumit infers the continuous from the discrete."

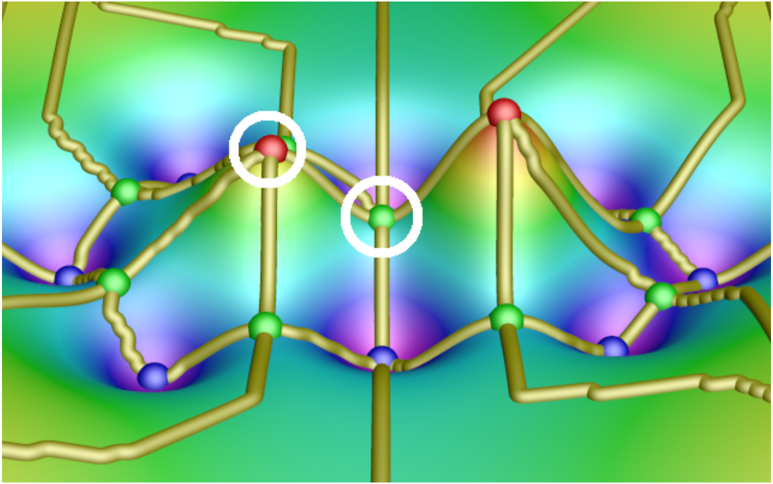
Figure: The Seine at La Grande Jatte by Georges Seurat

[S. Weinberger. What is persistent homology? 2011]

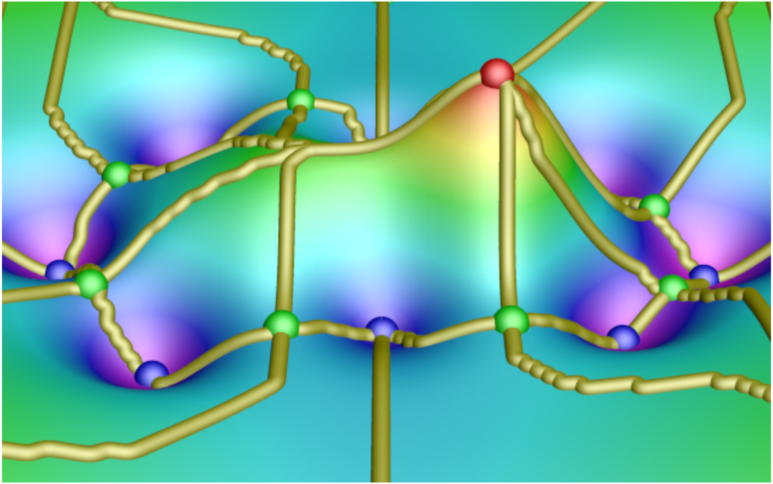
Simplifying topological features



Simplifying topological features

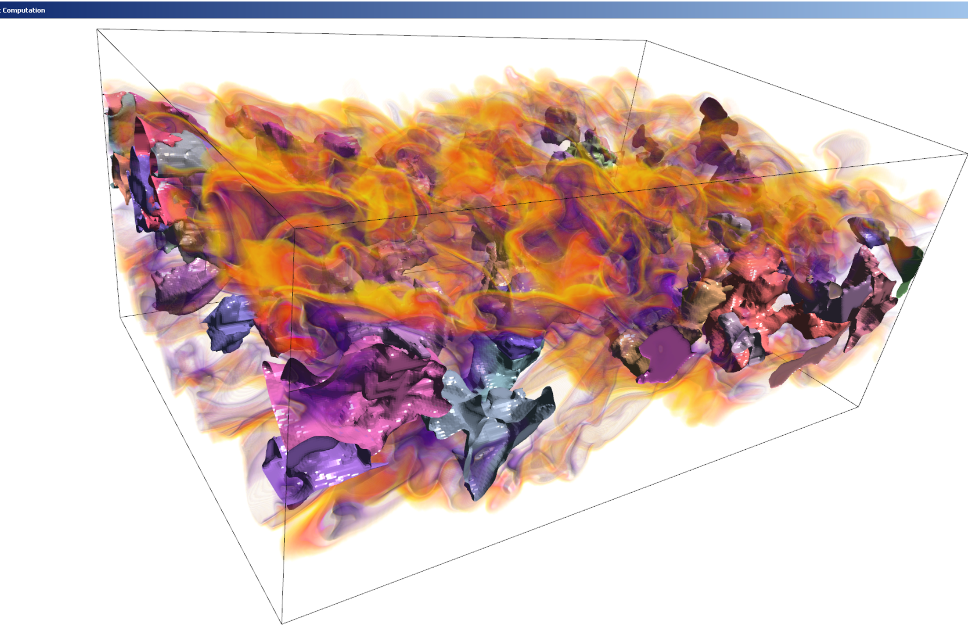


Simplifying topological features

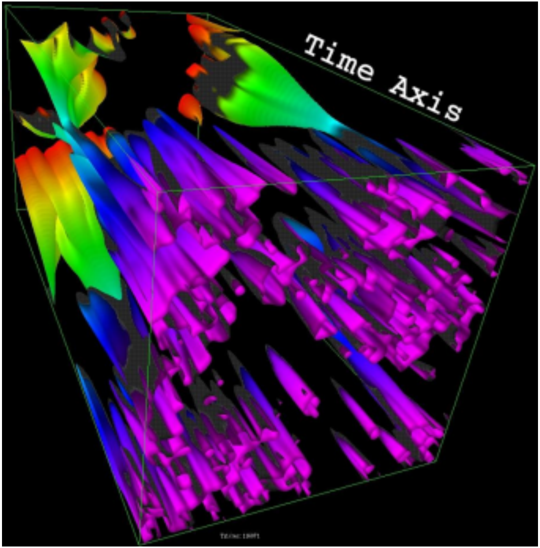


Why is Topo-In-Vis cool for large data science discovery?
Some Application Stories...

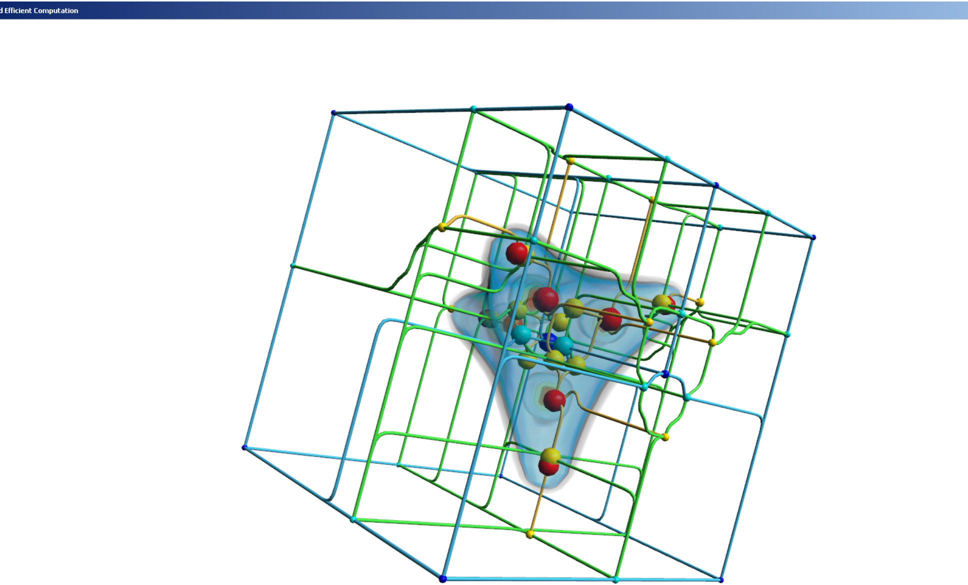
# Combustion simulation



# Tracking 2D Combustion

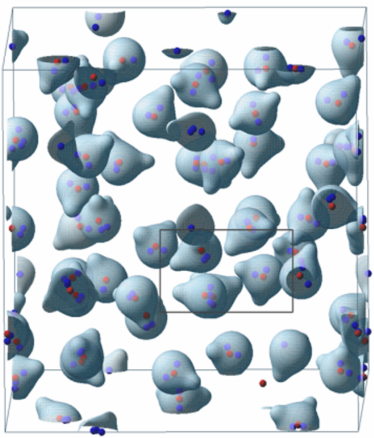


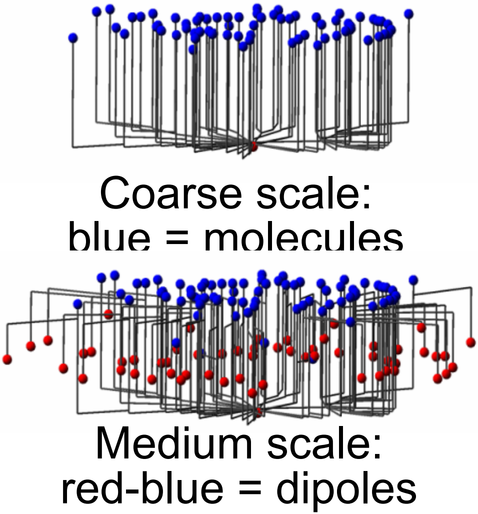
# Chemical compound: C4H4



# Molecular dynamics

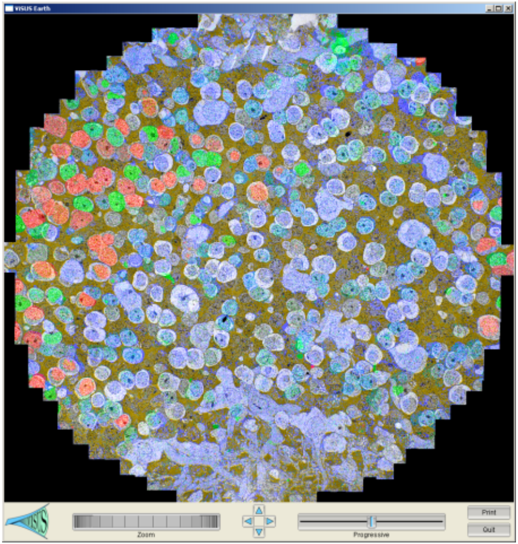
Molecular dynamics simulation (left) with abstract graph representation of its features at two scales (right)





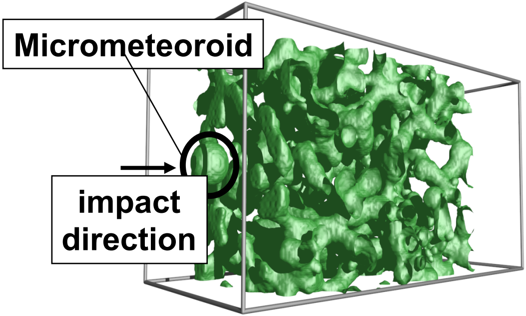
# Retinal connectome

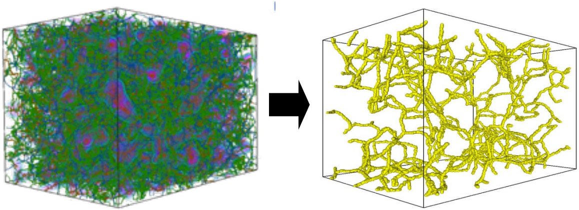
A connectome is a comprehensive map of neural connections in the brain [wiki]



# Case Study A: Material science

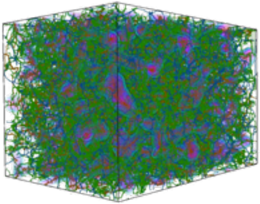
Quantitative Analysis of the Impact of a Micrometeoroid in a Porous Medium; reconstructing the structure of porous medium

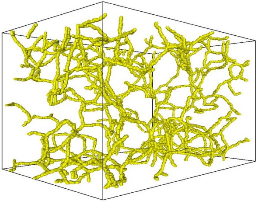


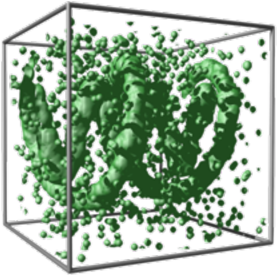
# Case study A: Topological Reconstruction

The Topological Reconstruction Method is Validated with a Controlled Test Shape



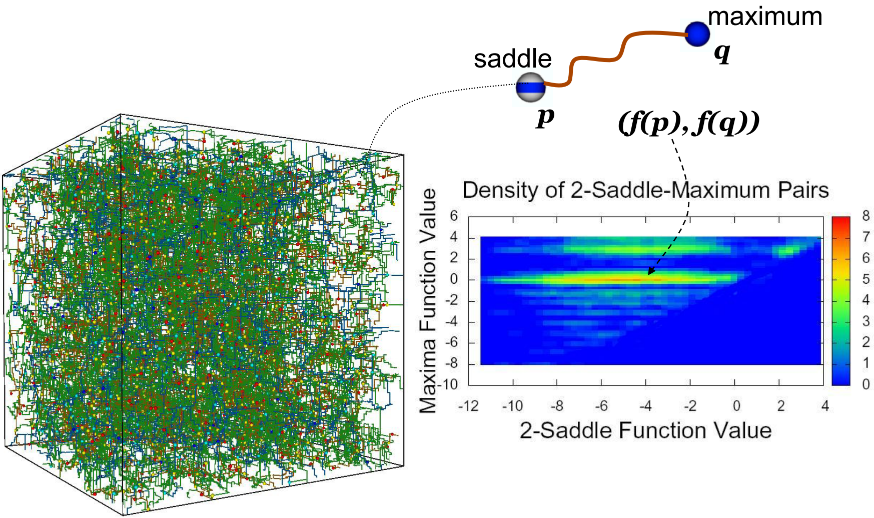


Preparation: we develop control test data to validate the approach



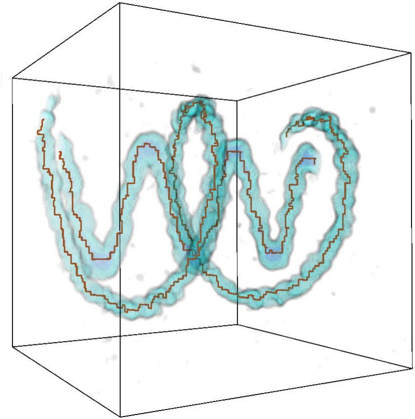
# Case study A: Control Data (dist. of topological features)

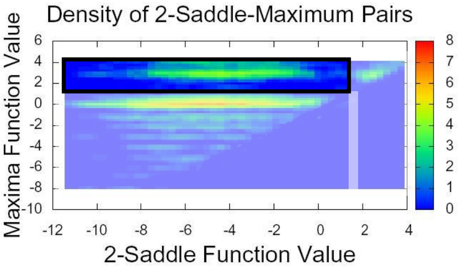
We Report the Distribution of Topological Features in the Full Resolution Data



# Case study A: Control Data (Hierarchical MSC)

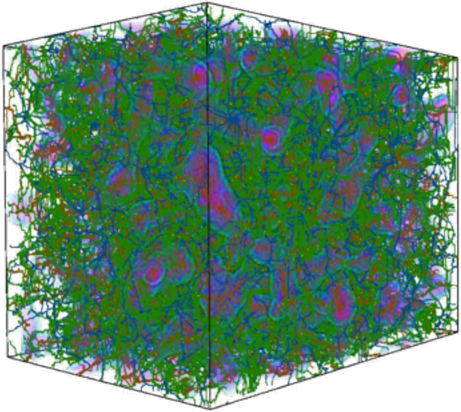
The Hierarchical Morse-Smale Complex Has Very Good Reconstruction Properties

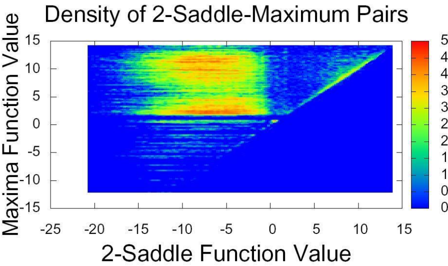




# Case study A: Porous Medium (dist. of topological features)

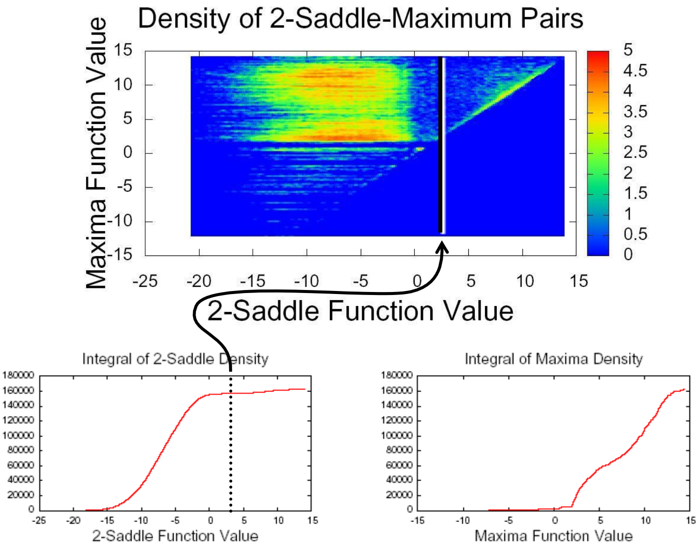
We Compute the Complete Morse-Smale Complex for the Porous Medium





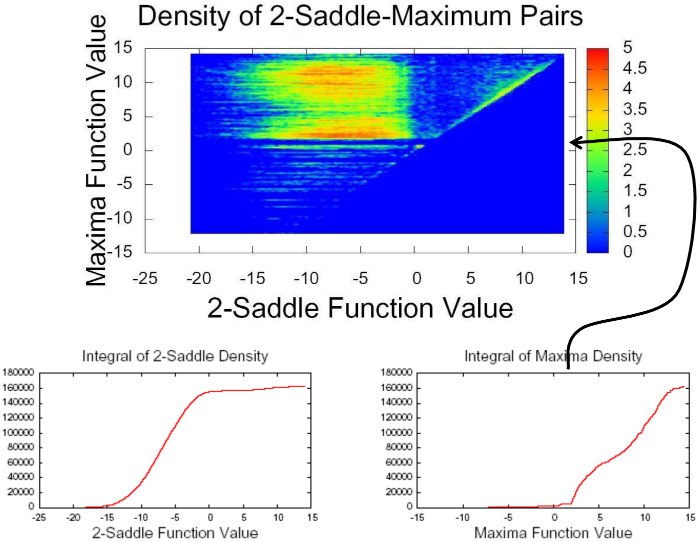
# Case study A: Porous Medium

Need to Find Proper Threshold Values and Characterize the Stability of the Solution



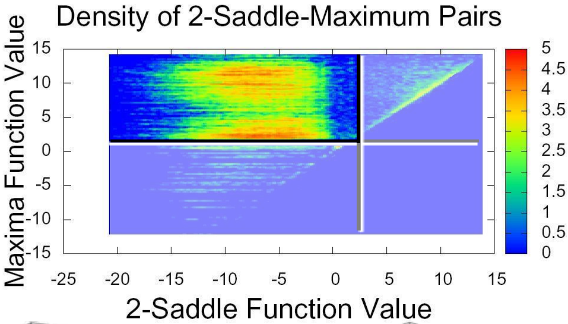
# Case study A: Porous Medium

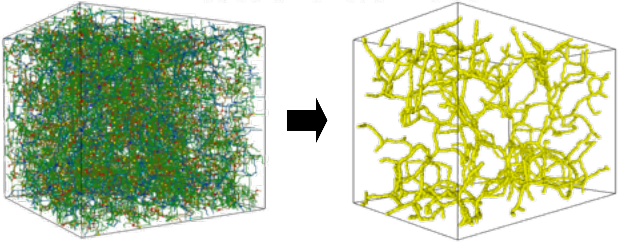
Need to Find Proper Threshold Values and Characterize the Stability of the Solution



# Case study A: Porous Medium

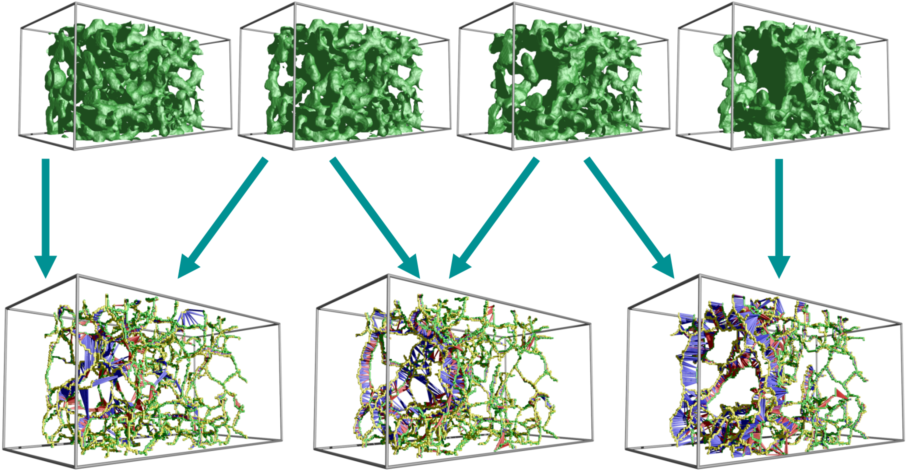
We Obtain a Robust Reconstruction of the Filament Structures in the Material



# Case study A: Porous Medium

We Track the Evolution of the Filament Structure of the Material Under Impact

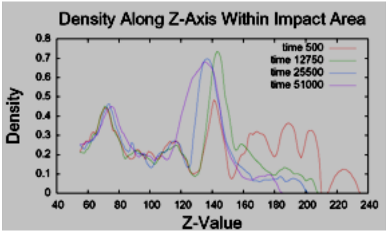


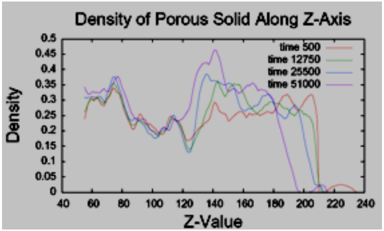
**Time comparison of the reconstructions**

# Case study A: Porous Medium

The Extracted Structures Allow to Quantify the Change in Porosity of the Material

## Density profiles



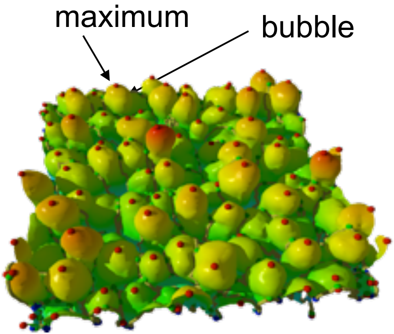




## Decay in porosity of the material

| Metric | t=500 | t=12750 | t=25500 | t=51000 |
|---|---|---|---|---|
| # Cycles | 762 | 340 | 372 | 256 |
| Total Length | 34756 | 24316 | 23798 | 18912 |

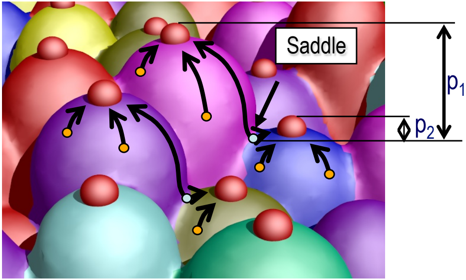
# Case study B: feature definition - Bubble Tracking

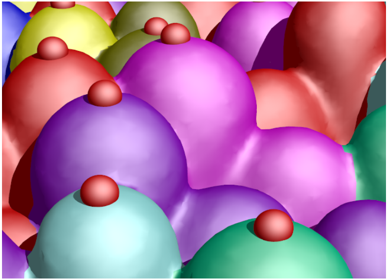
Analyze high-resolution Rayleigh Taylor instability simulations



# Case study B: persistence simplification

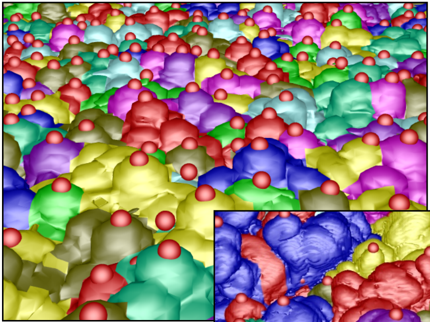
Analyze high-resolution Rayleigh Taylor instability simulations



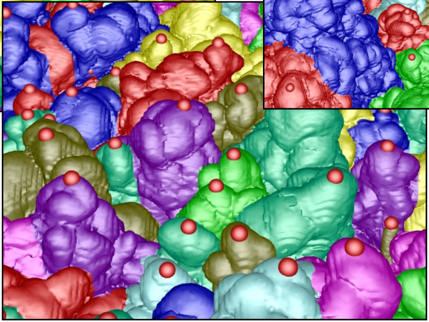



# Case study B: robust segmentation

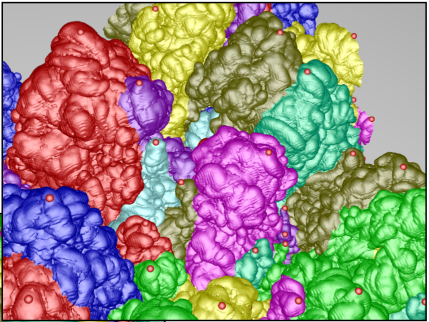
The segmentation method is robust from early mixing to late turbulence



T=100

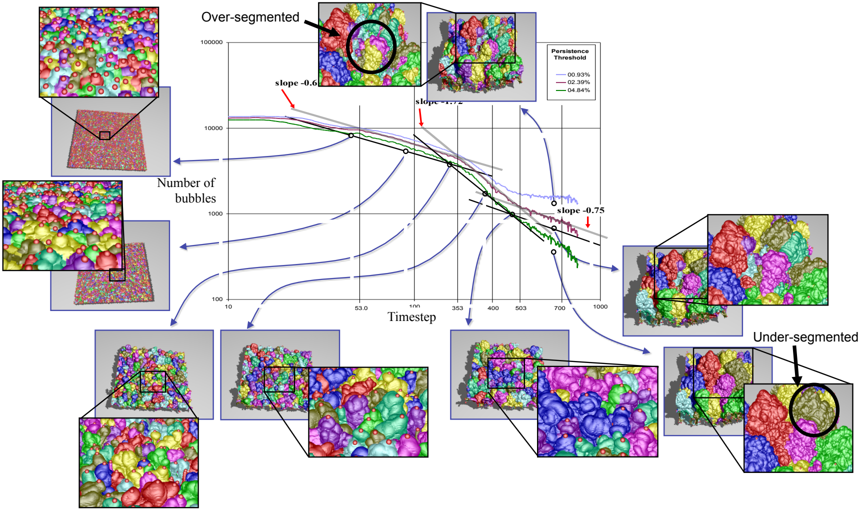


T=353



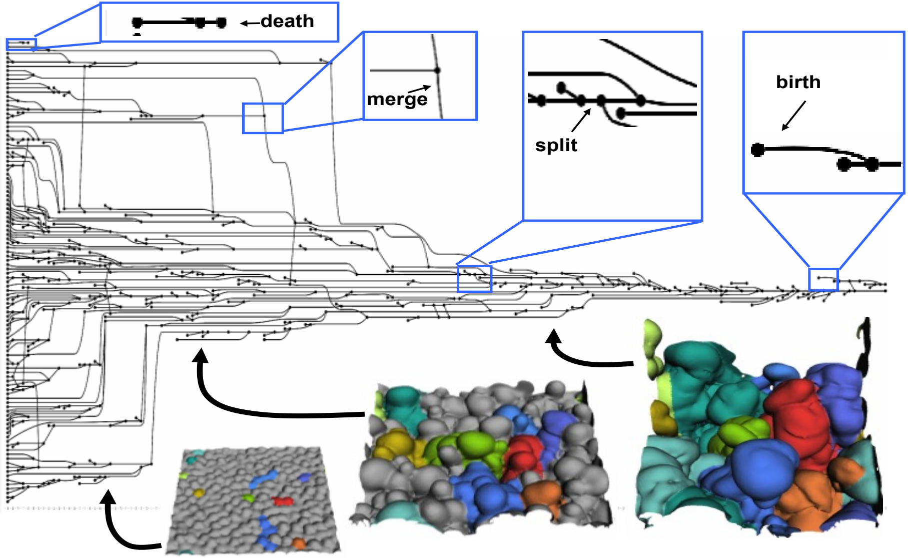
T=700

## We Evaluated Our Quantitative Analysis at Multiple Scales



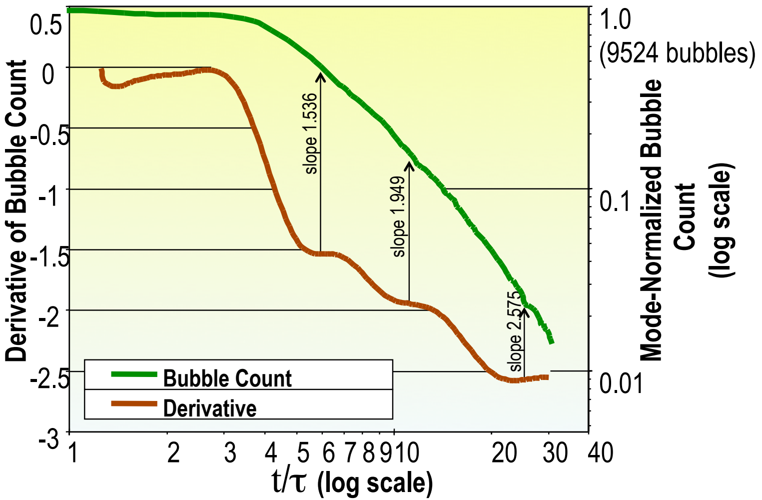
# Case study B: event characterization

We characterize events that occur in the mixing process

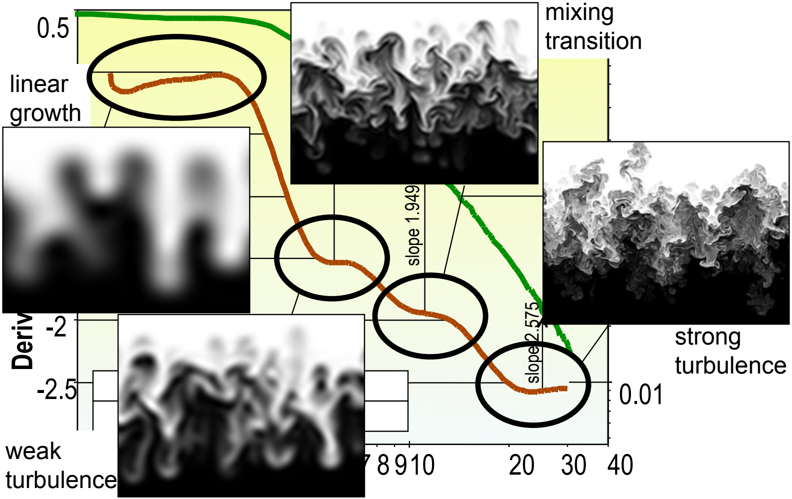


# Case study B: Exciting Result

First Time Scientists Can Quantify Robustly Mixing Rates by Bubble Count

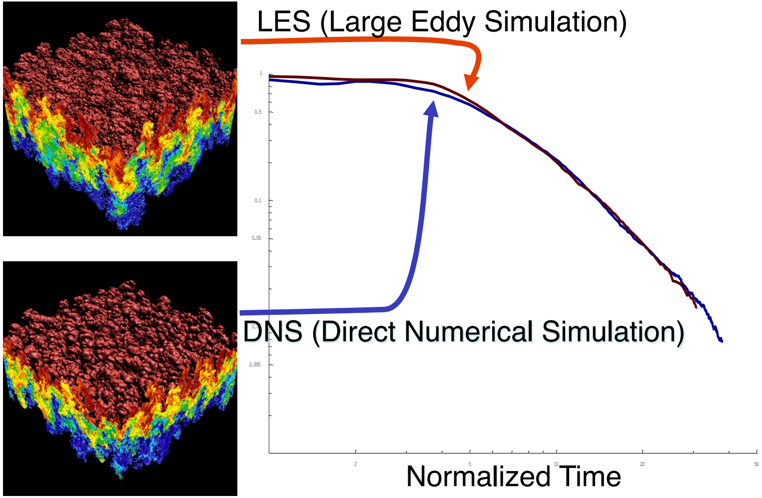


We Provide the First Quantification of Known Stages of the Mixing Process



# Case study B: Exciting Result

We Provided the First Feature-Based Validation of a LES with Respect to a DNS

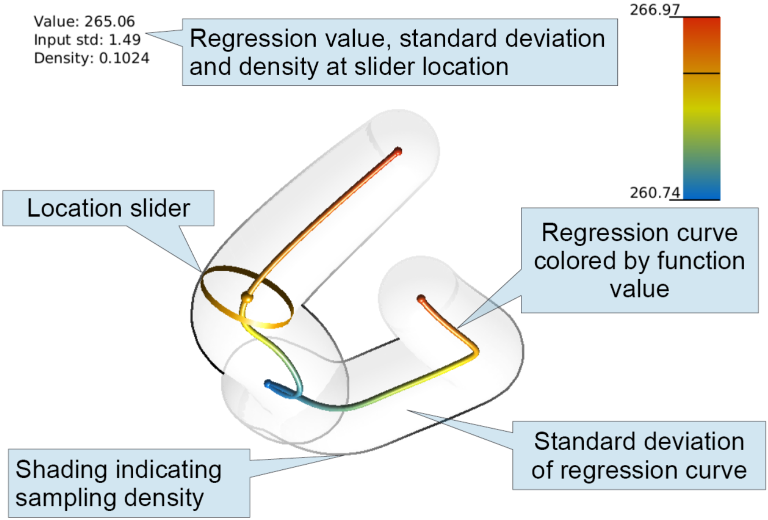


## Tracking Bubbles in a Rayleigh-Taylor Instability (video)

Coming up next:
What about hight dimensional? Data analysis and visualization is not seperable...

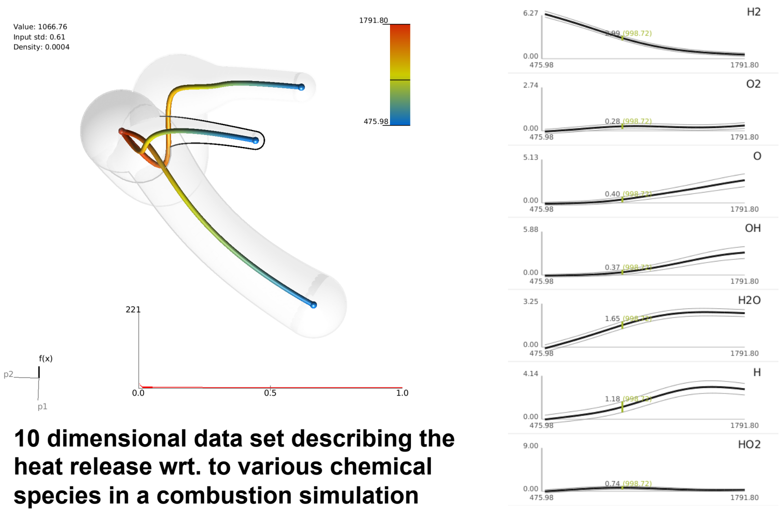
# High dimensional scalar function

[S. Gerber, P.-T. Bremer, V. Pascucci, R. Whitaker. Visual Exploration of High Dimensional Scalar Functions. 2010]



# High dimensional scalar function

[S. Gerber, P.-T. Bremer, V. Pascucci, R. Whitaker. Visual Exploration of High Dimensional Scalar Functions. 2010]




**10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation**

What are some of the cool open problems?

# For both analysis and visualization...

- Robustness of topological structures
- Scalability, approximation
- High-dimensional data
- Integration with statistics and manifold learning
- Usability

Break!

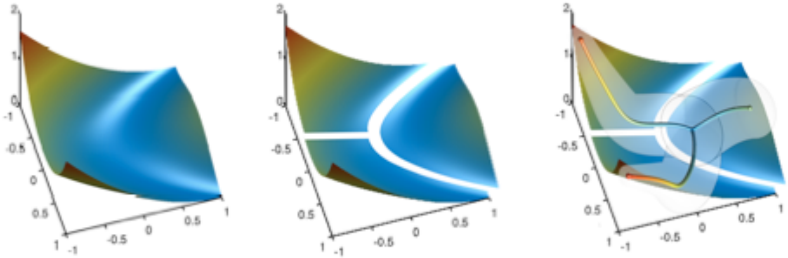
## Part 2:

## Exploration of High Dimensional Functions for Sensitivity Analysis

Joint work: Samuel Gerber, Ross Whitaker,
Dan Maljovec, Bei Wang,
Diego Mandelli, Peer-Timo Bremer, Valerio Pascucci

# Key ideas

- Domain decomposition using Morse-Smale approximation
- Geometric summaries of each crystal using regression
- Dimension reduction to embed regression curves

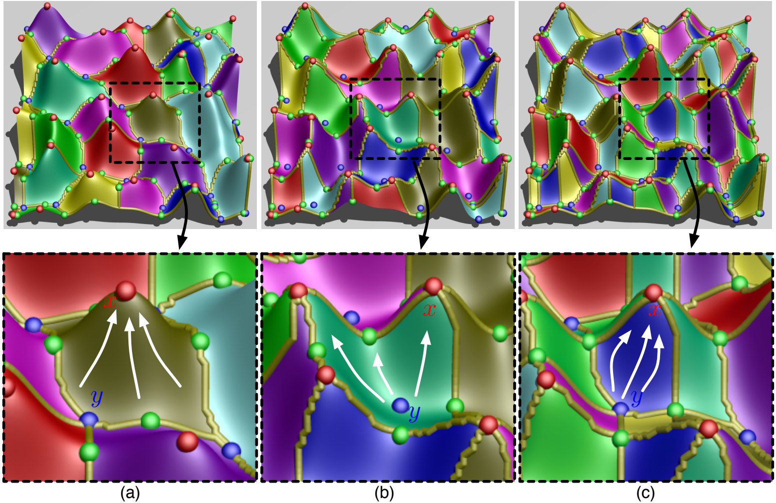



(a) (b) (c)

Later, more machine learning capabilities

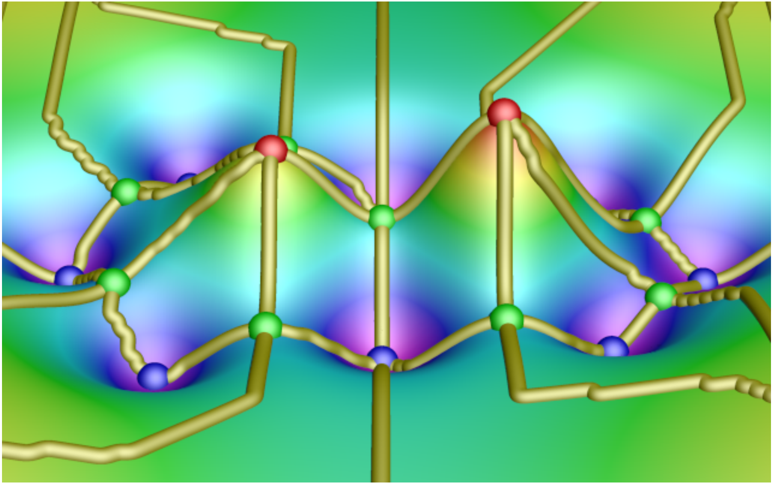
# Morse-Smale Complex

Partition data into monotonic regions based on gradient flow

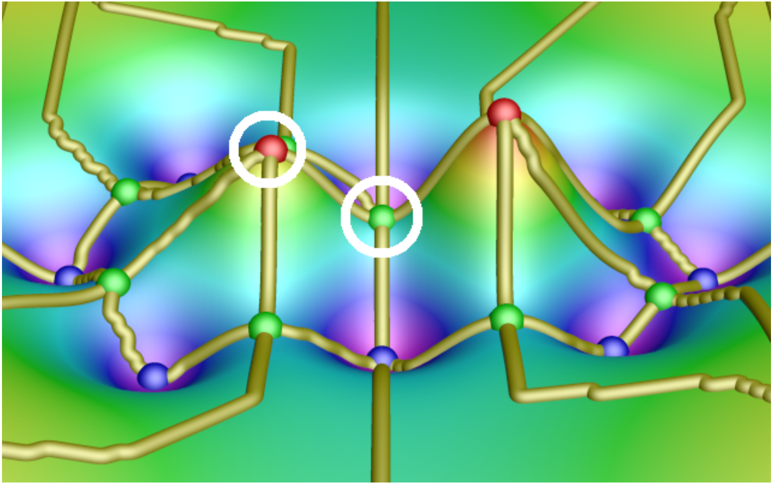



Descending Manifolds    Ascending Manifolds    Morse-Smale Complex

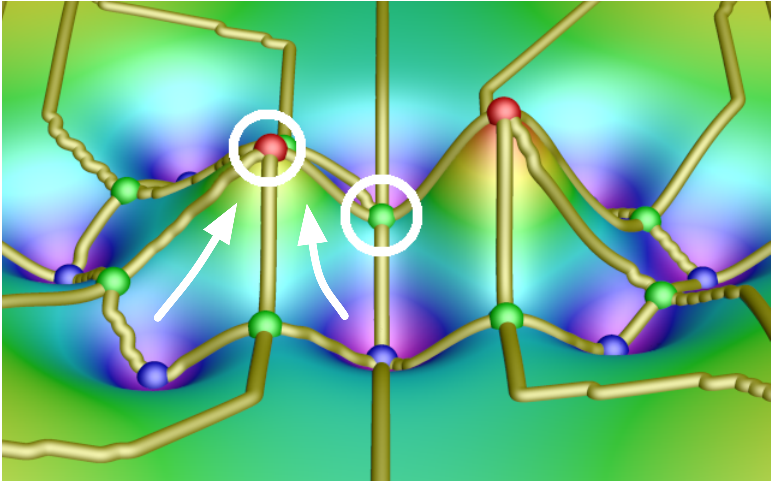
# Persistence Simplification 2D Example



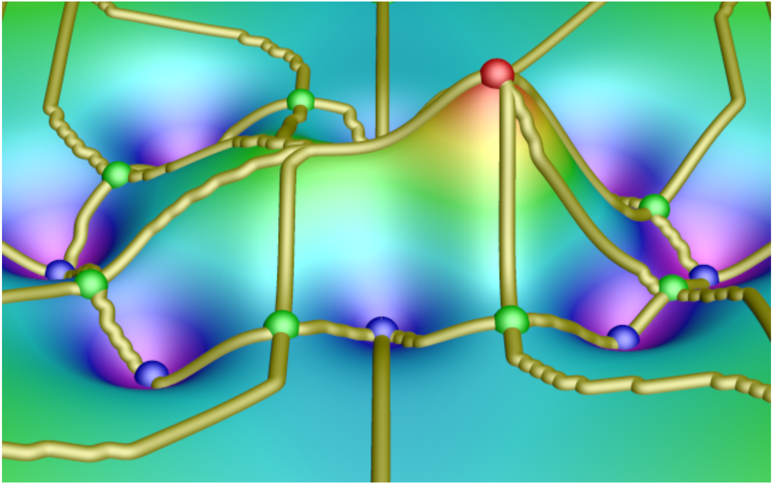
# Persistence Simplification 2D Example



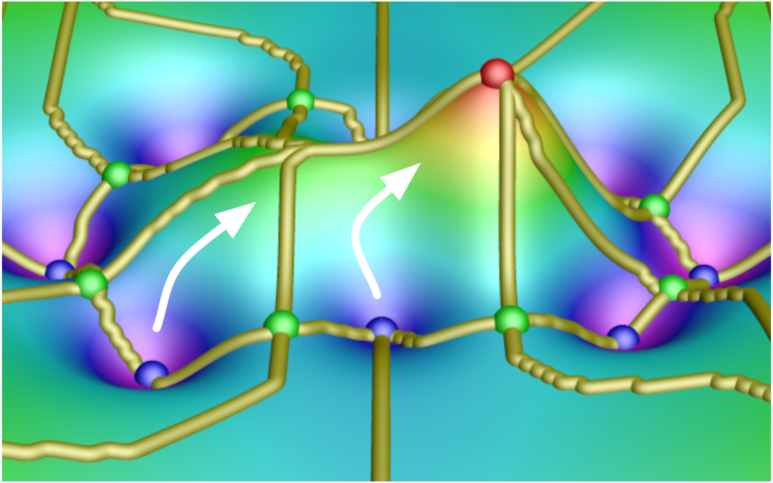
# Persistence Simplification 2D Example



# Persistence Simplification 2D Example

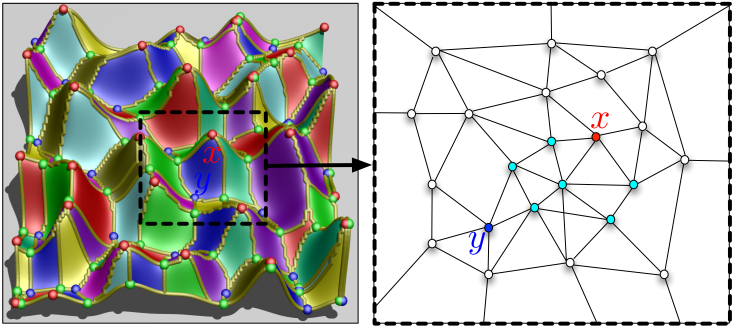


# Persistence Simplification 2D Example

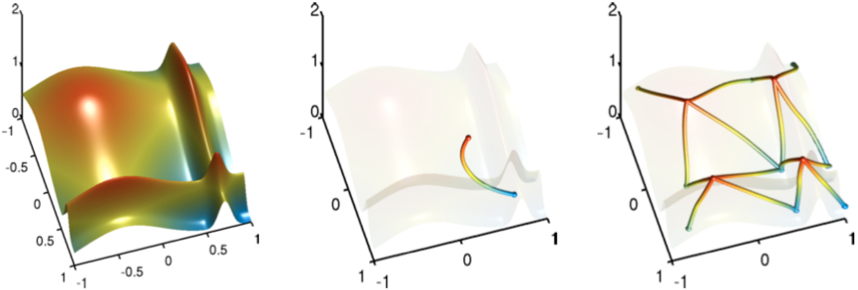


# Approximating MSC in high dimensions

Based on KNN graph and gradient approximations

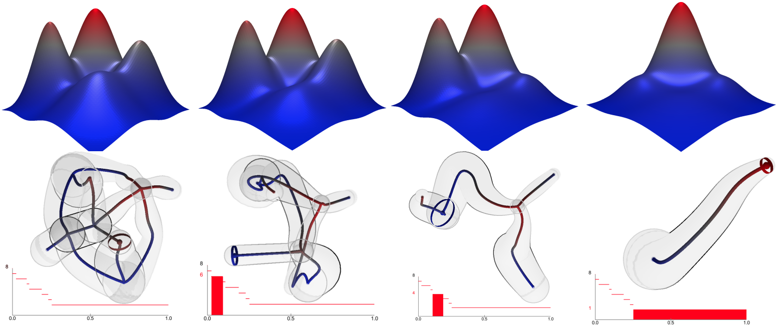


# A simple example



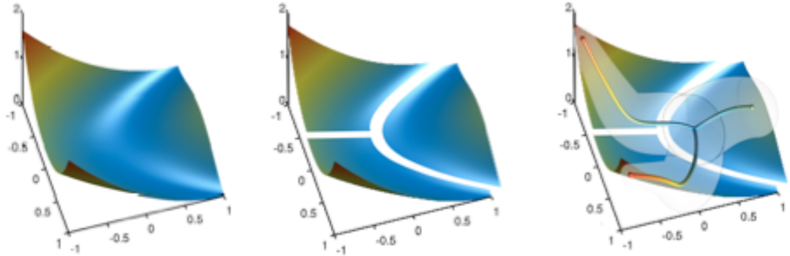

(a) (b) (c)

# Multi-Level Persistence Simplification



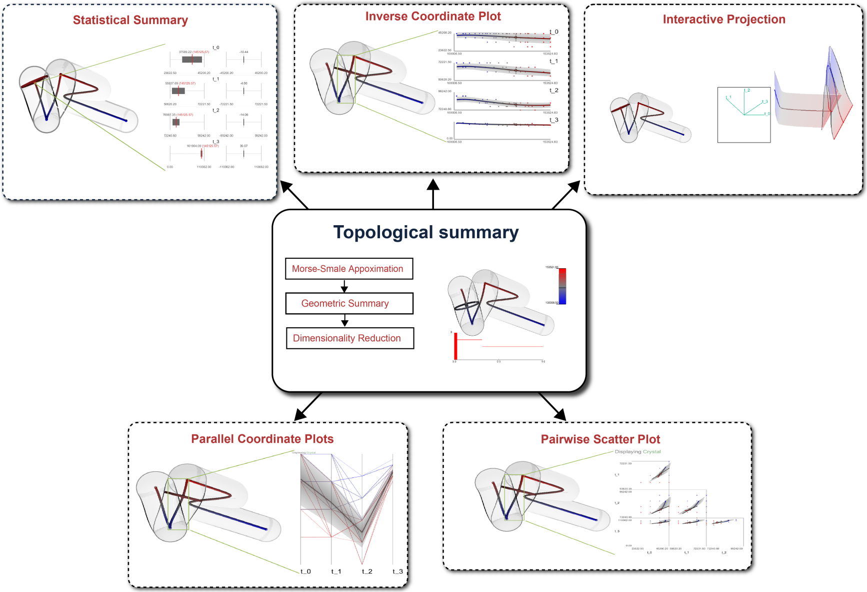
# Key ideas: Revisited

- Domain decomposition using Morse-Smale approximation
- Geometric summaries of each crystal using regression
- Dimension reduction to embed regression curves

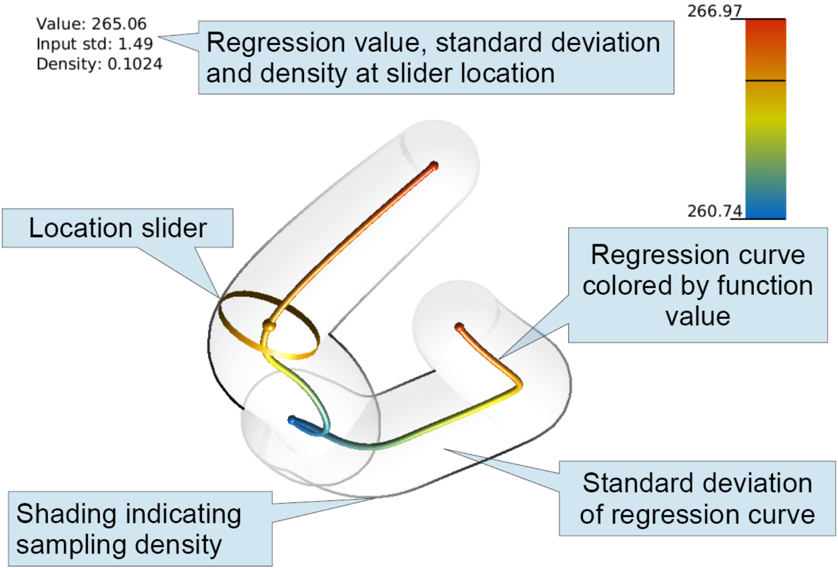


(a) (b) (c)

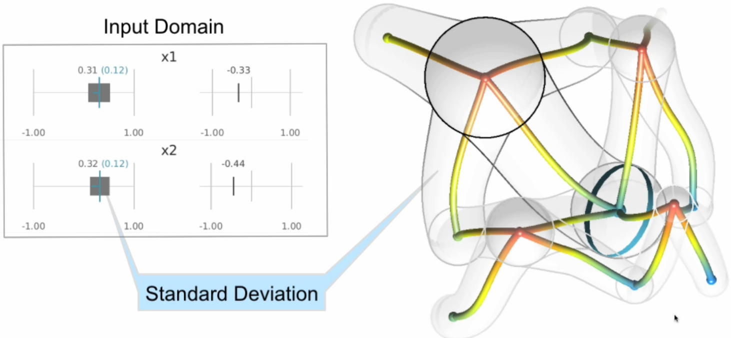
# Integrated Views



# Integrated presentation of statistics and topology



# Integrated presentation of statistics and topology




The set of regression curves provides a platform to visualize further information, such as standard deviation and sampling density. The color corresponds to the function value.

## Combustion Dataset

# Combustion

**Data:** Combustion simulation of Jet flames

**Sample:** 700K samples of chemical composition and temperature extracted point-wise from the simulation

**Input:** Composition of **10** chemical species, i.e. H2 and CO (fuel), O2 (Oxidizer)

**Output:** Temperature (heat released)

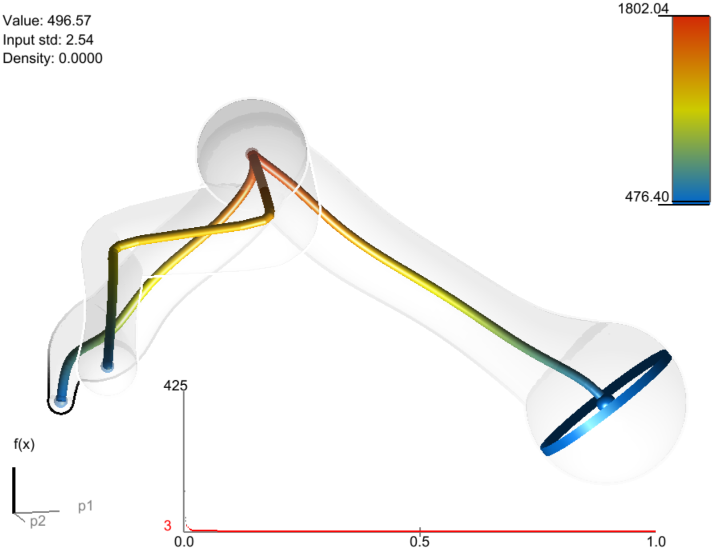
**Key:** Understand extinction and re-ignition phenomena

# Combustion: Input Detail

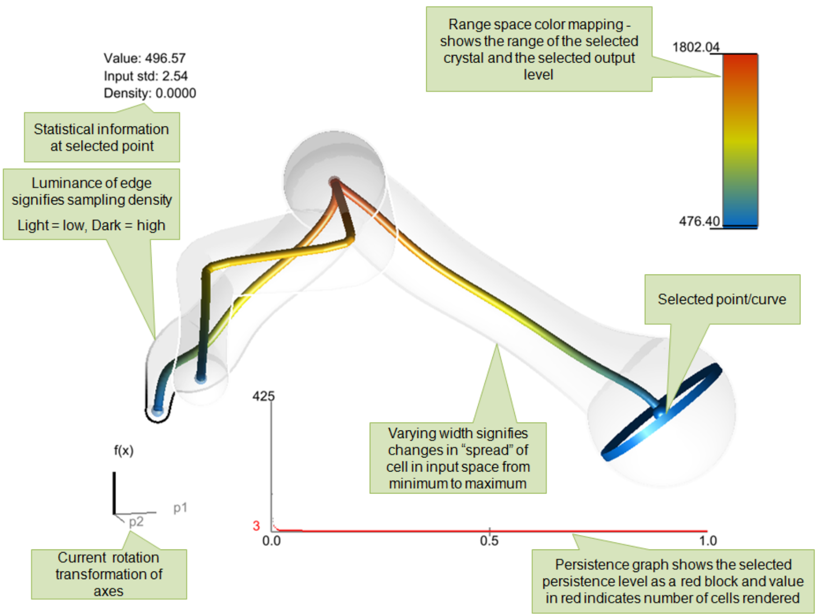
Chemical species involved in combustion simulation:

- $O_2$ (Oxygen gas / Oxidizer)
- O (Oxygen)
- OH (Hydroxide)
- $H_2O$ (water)
- H (Hydrogen)
- $HO_2$
- CO (Carbon monoxide)
- $CO_2$ (Carbon dioxide)
- HCO

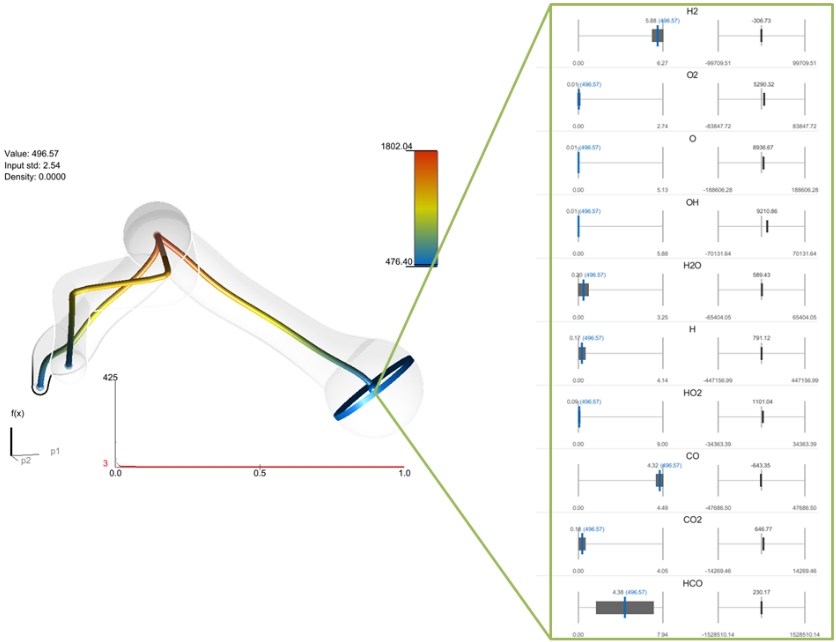
# Interface: Topological Summary



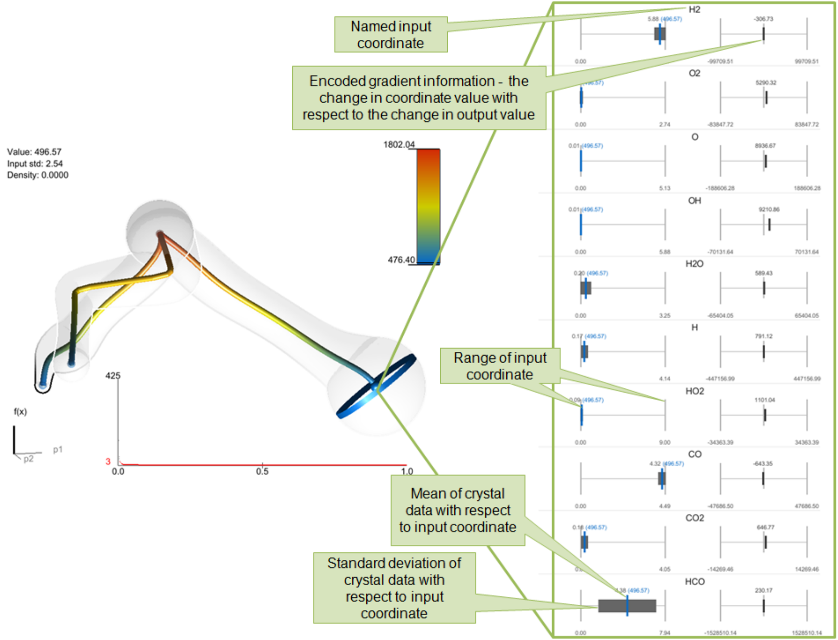
# Interface: Topological Summary



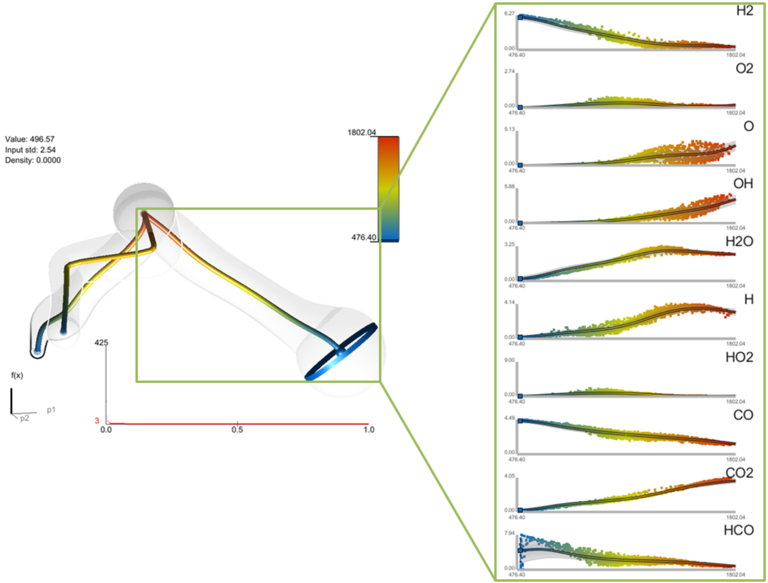
# Interface: Statistical Summary



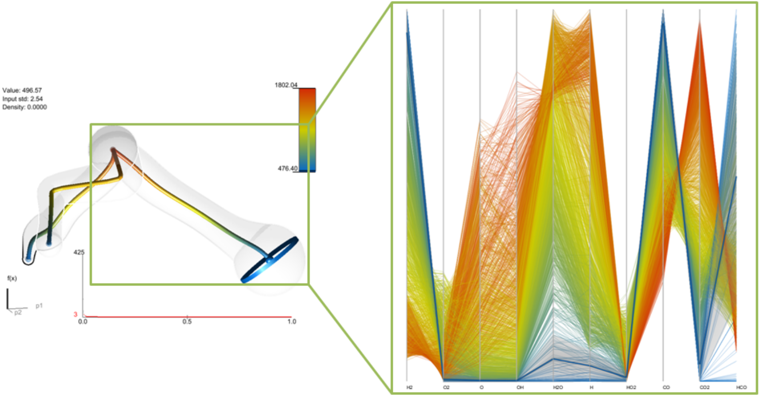
# Interface: Statistical Summary



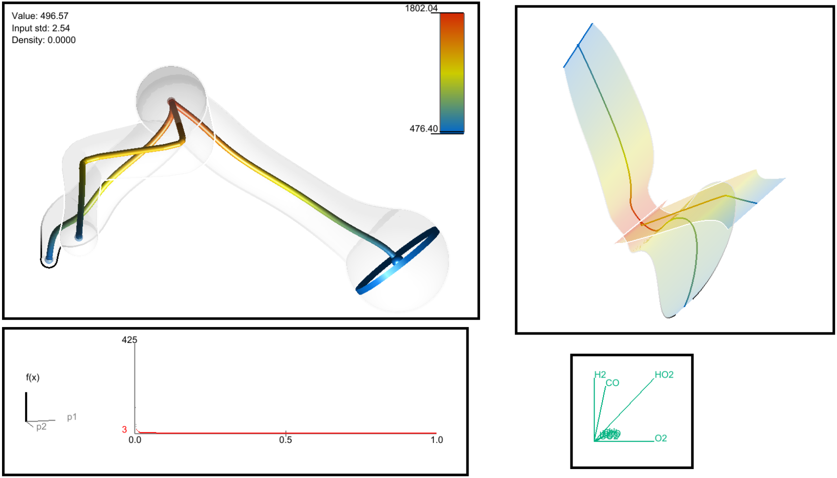
# Visual Interface: Inverse Coordinate Plots



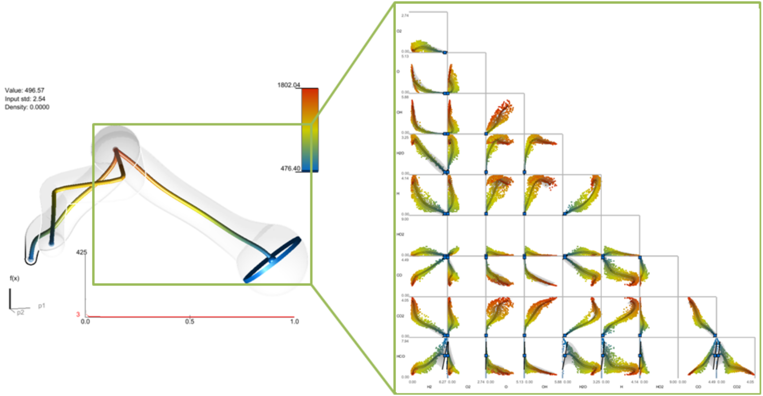
# Visual Interface: Parallel Coordinates Plots



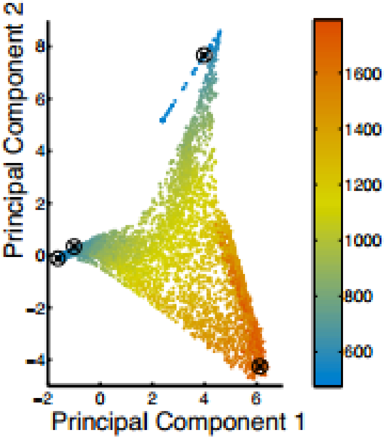
User manipulate how each axis is projected



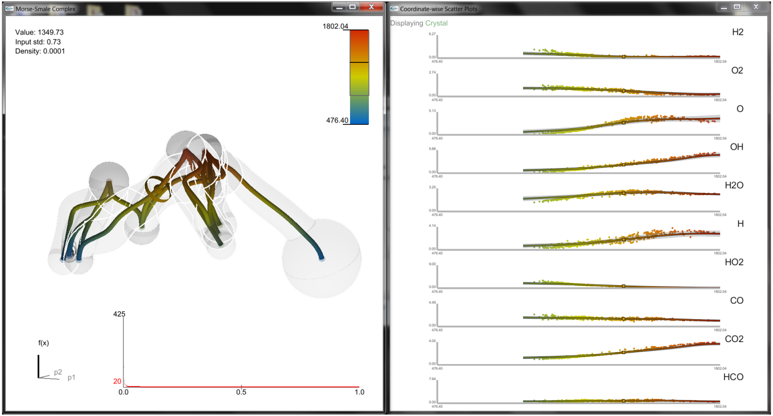
# Visual Interface: Pairwise Scatter Plots



# Combustion: Using PCA



# Combustion: Full Resolution

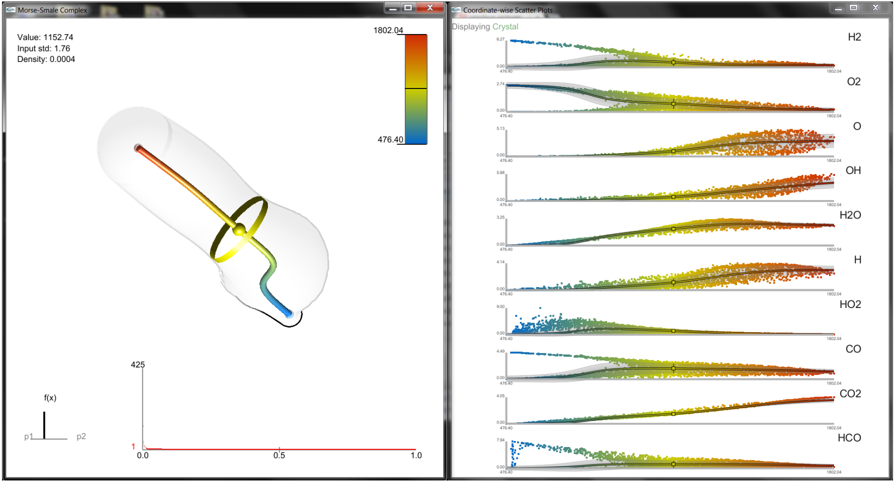


# Combustion: Example 1

1 crystal (1 min, 1 max)

min: high level of oxidizer, lack of fuel, no combustion

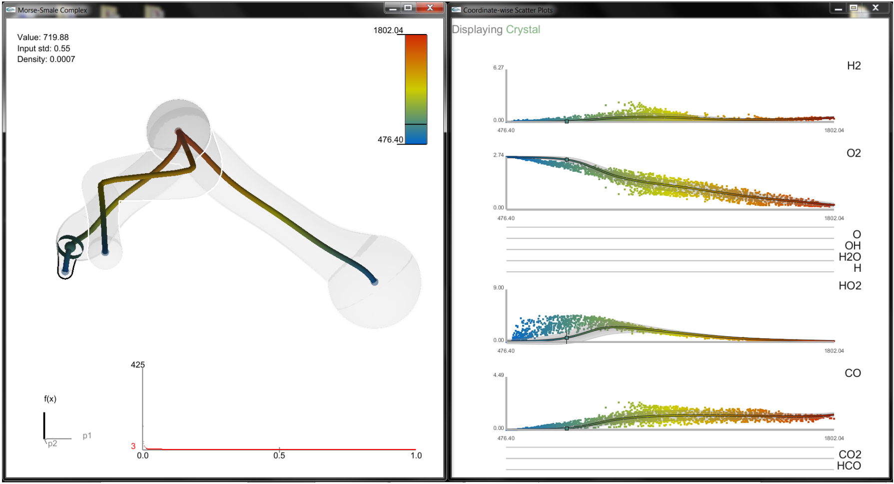
max: peak corresponds to combustion, many chemical reactions occur which is reflected in the high std of the peak



# Combustion: Example 2 , Crystal (c)

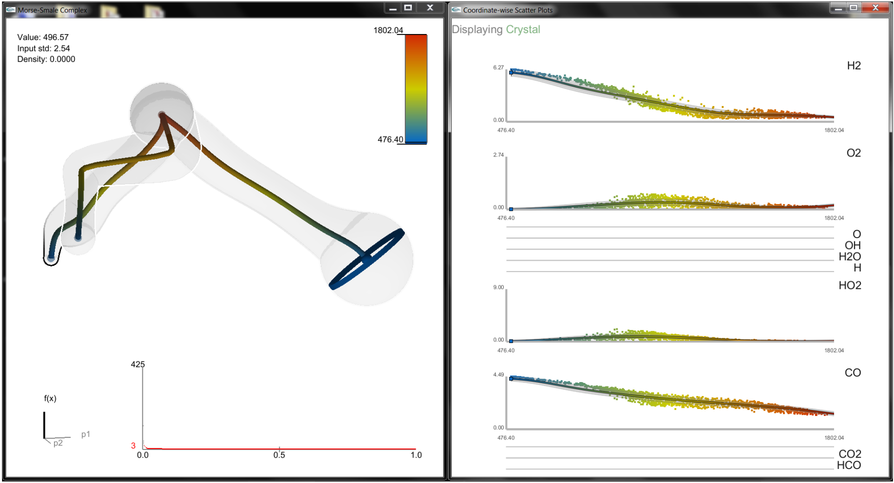
3 crystals (3 min and 1 max), 4 distinct modes of combustion; 3 minima have distinct chemical compositions.

Min (c): **pure oxidizer** (O2). Lack of fuel. No chem. reaction.



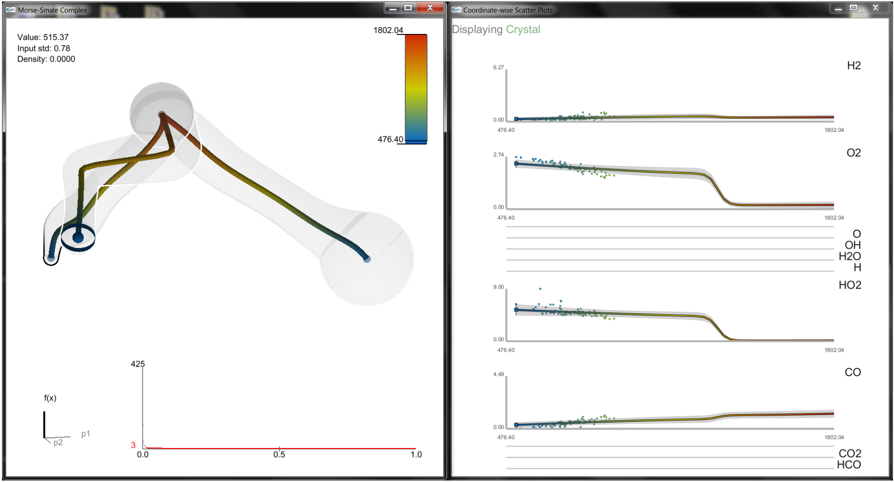
# Combustion: Example 2, Crystal(b)

Min (b): **pure fuel** ($H2$ and CO). Lack of oxidizer.
No chem. reaction.



# Combustion: Example 2, Crystal(a)

Min (a): **extinction**. Fuel and oxidizer is highly turbulent and blows the flame out, resulting large amount of HO2.



# Combustion Dataset: Live Demo

## Climate Dataset

# Climate

**Data:** Community Atmosphere Climate Model. Understand uncertainty in climate simulation by using an ensemble of simulations for various input parameters.

**Sample:** 593 runs of Community Atmosphere Climate Model

**Input:** **21** parameters setting, describe various aspects of physics

**Output:** thermal radiation (net long wave flux, leaving the planet)

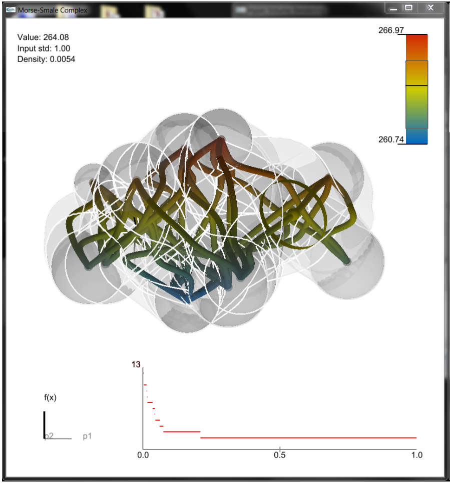
**Key:** How radiation (total upwards long wave flux) influenced by input parameters

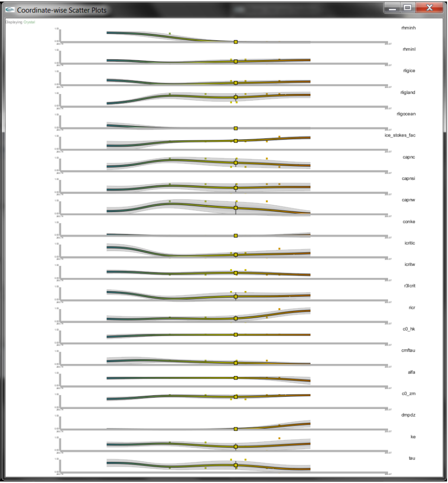
# Climate: Input Detail

For example,

- tau: deep convection ($> 500$ hPa). Convection: thermal driven upwelling of warm, moist air.
- cftau: shallow convection ($< 500$ hPa).
- tau, cftau: both are related to **cloud formation**, there imbalance leads to fewer clouds and high thermal radiation

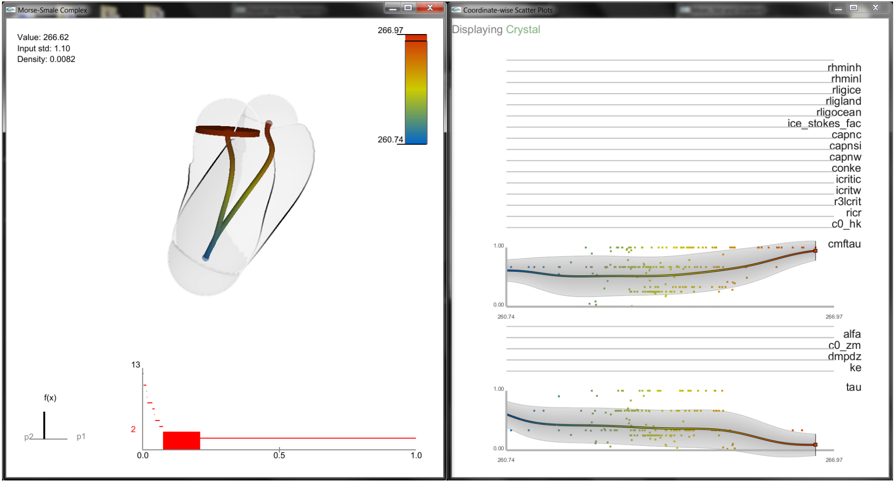
# Climate: Full Resolution



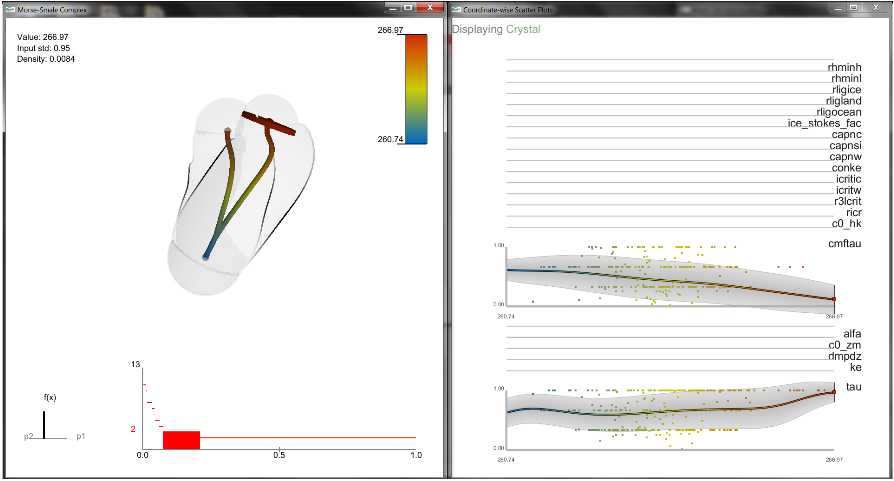
# Climate: Example, Crystal (a)

2 crystals (2 max and 1 min)
max (a): high radiation, small tau, large cmftau, unbalanced.



# Climate: Example, Crystal (b)

max (b): high radiation, large tau, small cmftau, unbalanced.
This is not apparent in standard statistical approach.



## Climate Dataset: Live Demo

# Crime Dataset

# Crime

**Data:** Communities and crimes

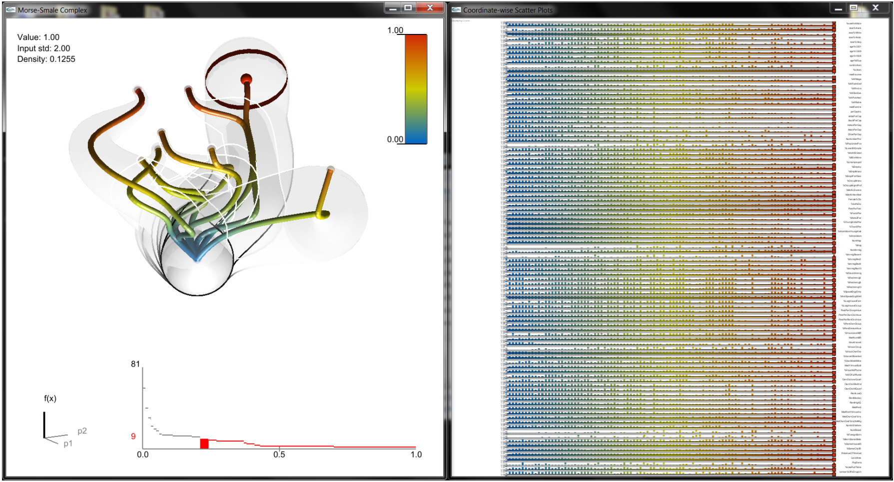
**Sample:** 1990 FBI uniform crime report, 1993 data points

**Input:** 100 social and economic variables of communities across the US, i.e. median income, unemployment rate, etc.

**Output:** Per capita crimes

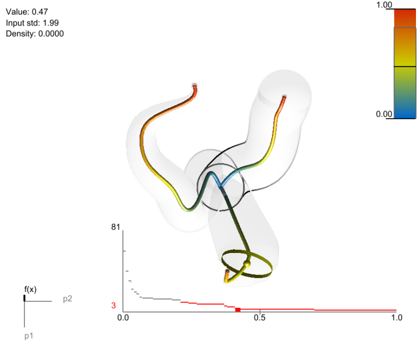
**Key:** Understand how social and economical factors affect crime rate

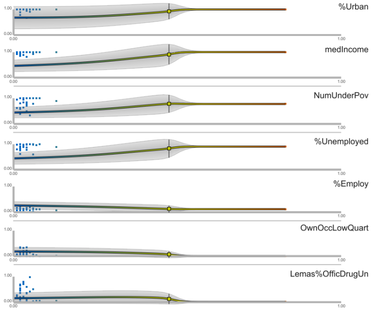
# Crime: Full Resolution



# Crime: Example, Crystal (a)

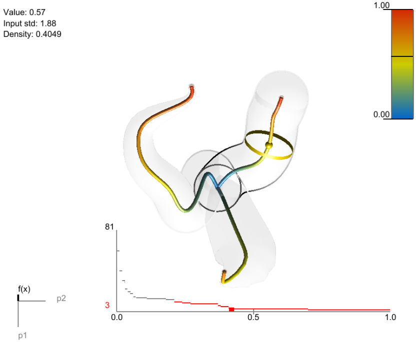
3 Crystals. Multiple peaks indicate different factors leading to high crime rate. Max (a): urban, high median income (MedIncome), and high unemployment rates, a large gap between rich and poor.

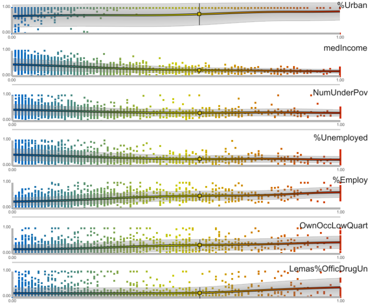


# Crime: Example, Crystal (b)

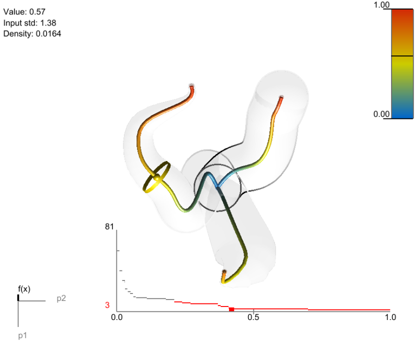
Max (b): urban, high percentage of officers assigned to drug cases

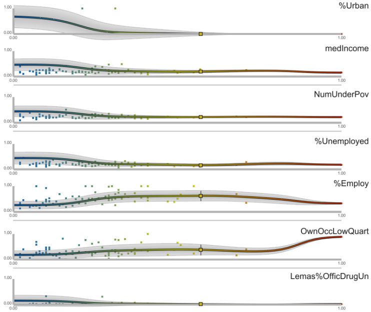


# Crime: Example, Crystal (c)

Max (c): rural, low Urban percentage, high percentage of low income housing occupancy, low employment



## Crime Dataset: Live Demo

# Concrete Dataset

# Concrete

**Data:** Concrete compressive strength

**Sample:** 1030 samples of different concrete cores tested for strength

**Input:** **8** chemical components, i.e. cement, water, fly ash, etc.

**Output:** (compressive) strength

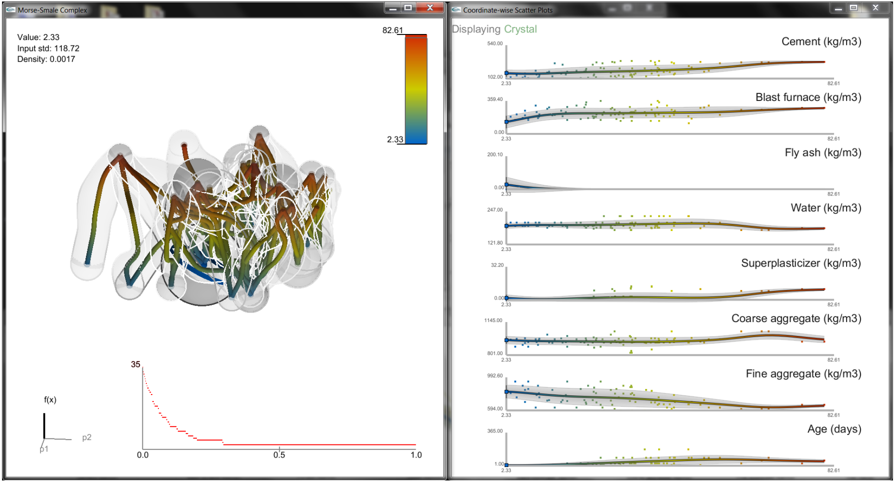
**Key:** Examine the effect of different cement mixtures on compressive strength of the resulting concrete

# Concrete: Input Details

Chemical species involved in concrete formation:

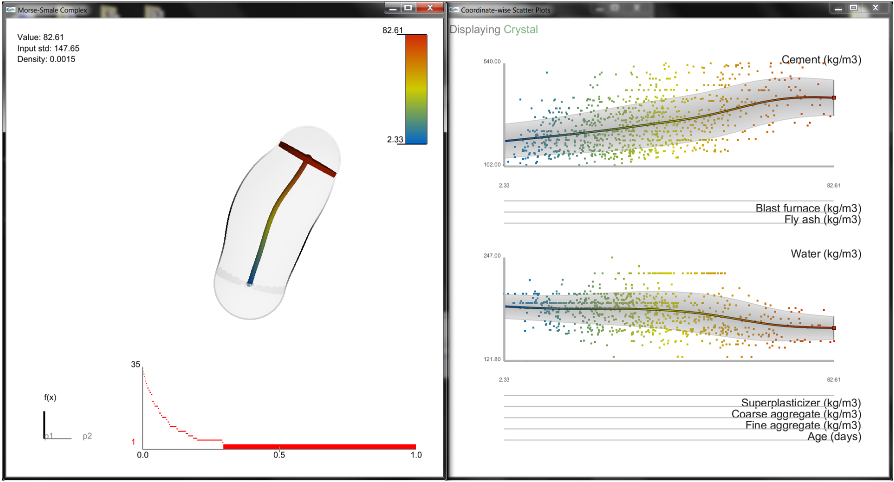
- Cement
- Blast furnace slag (BFS)
- Fly ash
- Water
- Superplasticizer
- Coarse aggregate (CA)
- Fine aggregate (FA)
- Age

# Concrete: Full Resolution



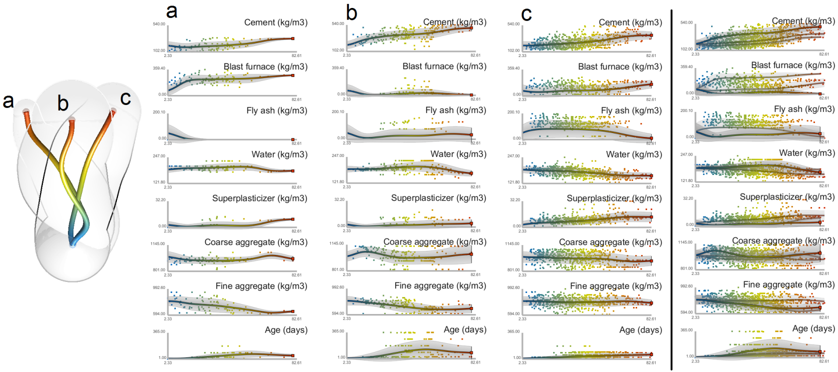
# Concrete: Example 1

1 crystal (singe max, single min)
Cement/water ratio, the higher, the stronger



# Concrete: Example 2

3 crystals: different mixtures could lead to similar strength
Minima differ in their settings of Fly ash, BFS, CA/FA ratio



## Concrete Dataset: Live Demo

## Nuclear 6D Dataset (INL)

# Nuclear 6D

**Data:** extracted from a $\mathbf{VR}_2^+$ nuclear reactor simulator

**Sample:** an ensemble of 10000 simulation trials where a SCRAM is simulated due to a failure in the system. A SCRAM event is when the control rods of the reactor are inserted into the core in order to prevent overheating of the reactor core.

**Input:** **6** parameters:

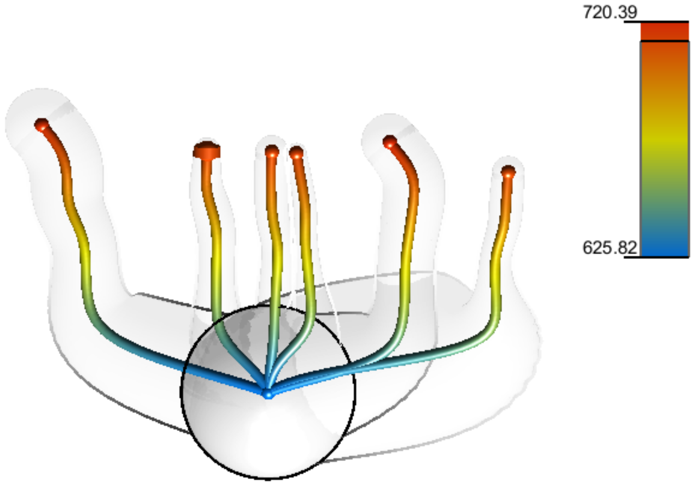
- **PumpTripPre** - min pressure in the heat exchange pump causing the SCRAM to trip
- **PumpStopTime** - relaxation time of pump's phase-out
- **PumpPow** - end power of the pump
- **SCRAMtemp** - max temp. causing the SCRAM to trip
- **CRinject** - control rod position at the end of SCRAM
- **CRtime** - relaxation time of the control rod system

**Output:** peak coolant temperature (PCT), measured in Kelvin

**Key:** what combination of conditions (in the form of input parameters) can cause potential reactor failure (i.e. nuclear meltdown witnessed by PCT exceeding a threshold value).

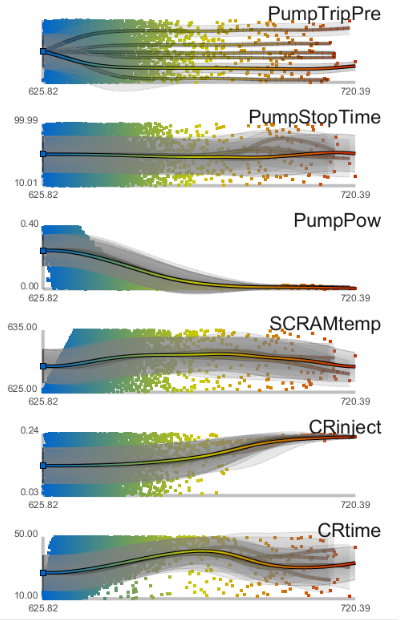
# Nuclear 6D: interface

6 crystal (singe min, six max)



# Nuclear 6D: inverse coordinate plots

All crystals combined:



# Nuclear 6D Dataset: Live Demo

## Nuclear 4D Dataset (INL)

# Nuclear 4D

**Data:** analysis of recovery from an aircraft crash into nuclear reactor. The reactor decay heat is released to the atmosphere through four cooling towers. During a simulation, the plant is operating at 100% power when an airplane crashes into the plant, destroying three of the four towers. A recovery crew then arrives at the site and attempts to reestablish the capability of the reactor by restoring the damaged towers one by one.

**Sample:** 610 simulations has been generated, and among which 132 cases are considered system failures when the reactor reaches a maximum temperature of 1000K before the end of simulation.

**Input:** **4** parameters, time for the crew to arrive at the plant $t_0$, and the time for them to recover the first, second and third tower ($t_1$, $t_2$ and $t_3$)

**Output:** e.g. maximum temperature reached in the simulation (MT)

**Key:** understand how these input variables impact system dynamics, help domain scientists to make decisions regarding repair strategies and evacuation plans.

# Visual interface highlighting clustering structure

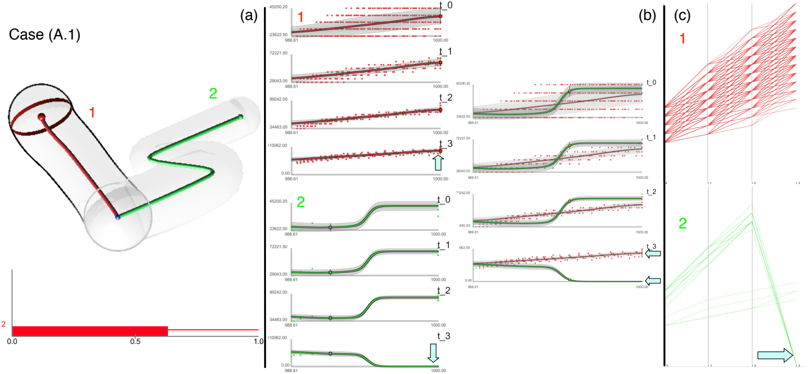



Figure: (a) The topological summary visual interface. (b) Inverse coordinate plots for both crystals individually and combined. (c) Parallel coordinate plots.

## Nuclear 4D Dataset: Live Demo

## Nuclear 9D Dataset (INL)

# Nuclear 9D

**Data:** analysis of recovery attempts of a Loss of Offsite Power event followed by loss of diesel generators resulting in Station BlackOut (SBO).

**Sample:** 19996 simulations generated, among which 6597 failed, i.e. reactor temperature breached threshold resulting in reactor core damage. 13399 trials were successfully able to keep the temperature below the threshold temperature while either diesel generator power or offsite power were restored, or the firewater system is aligned allowing cooling to the core via the firewater system.

**Output:** Maximum clad temperature reached in the simulation (MT)

**or:** Reactor Power

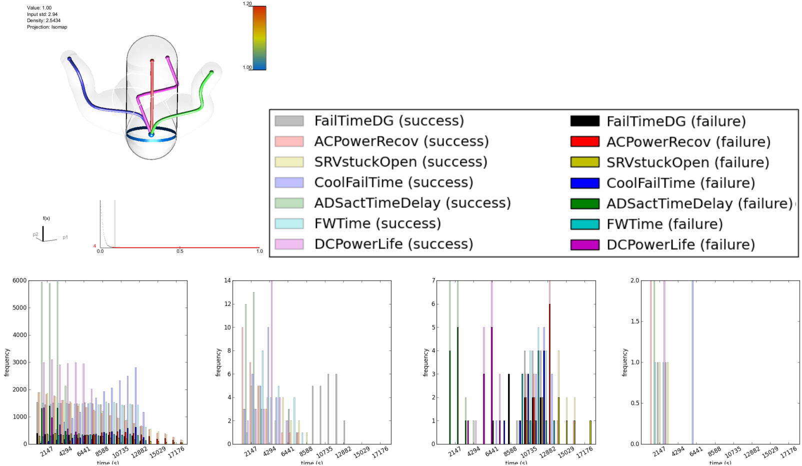
**Key:** (a) understand how these input variables impact system dynamics, help domain scientists to make decisions regarding probability of success/failure with regard to different stochastic variables. (b) the impact of increased reactor power on safety of the nuclear plant, in terms of time required for various recovery procedures.

# Nuclear 9D

**Input:** **9** parameters (each dimension is normalized to have a zero mean and a standard deviation of one):

- **FailureTimeDG** - Failure time of the diesel generators
- **ACPowerRecoveryTime** - Minimum time to recovery either offsite AC Power or power from the diesel generators
- **SRVstuckOpenTime** - Time when one safety relief valve gets stuck in the open position
- **cladFailureTemp** - Threshold temperature representing system failure
- **CoolingFailToRunTime** - The time when both the High Pressure Core Injection (HPCI) cooling system and the Reactor Core Isolation Cooling (RCIC) fail to run
- **ReactorPower** - Percent of upscaling of the raw material used in the reactor core
- **ADSactivationTimeDelay** - Time delay between triggering of an HCTL event and the time it takes to activate the ADS.
- **FWTime** - Time to align the firewater system
- **TotalBatteryLife** - Time where the secondary cooling system's DC power source fails

# Visual interface highlighting clustering structure



# Nuclear 9D Dataset: Live Demo

## Material Science Dataset (PNNL)

# Material Science Dataset: Live Demo

Break!

## Part 3: Perspective from a Nuclear Scientist

# Part 4: Discussions

# Some references to be found at:

- http://www.sci.utah.edu/∼beiwang/
- http://www.pascucci.org/
- http://cedmav.sci.utah.edu/research-projects/high-d-data-anal-and-vis.html

Thank you!

beiwang@sci.utah.edu
maljovec@cs.utah.edu
diego.mandelli@inl.gov