SAND2011-5754C

Data analysis and visualization of petascale combustion science simulation data

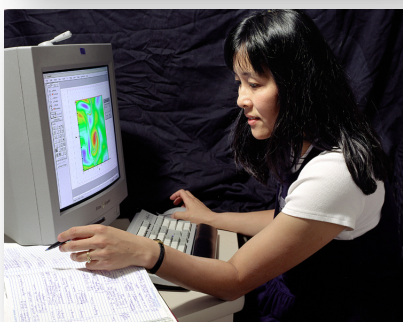
J. Bennett, J. Chen, H. Kolla, A. Mascarenhas, H. Yu @ **Sandia National Laboratories**
A. Gyulassy, V. Krishnamoorthy, S. Liu, V. Pascucci @ **University of Utah**
P.-T. Bremer, C. Correa @ **Lawrence Livermore National Laboratory**
K.-L. Ma, A. Tikhonova, J. Wei @ **University of California, Davis**
R. Grout @ **National Renewable Energy Laboratory**
E. Hawkes @ **University of New South Wales**
C. Wang @ **Michigan Technical Institute**
A. Gruber @ **SINTEF**

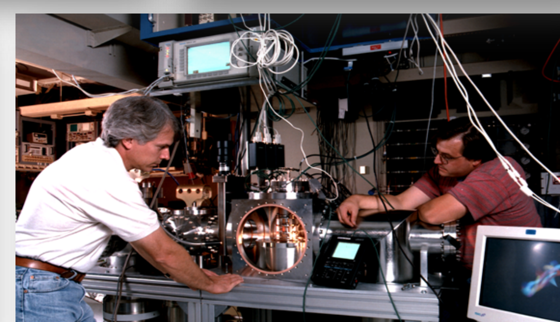
CScADS Workshop on Scientific Data & Analytics for Extreme Scale Computing
July 26, 2011

Sandia National Laboratories is a multi program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Combustion Research Facility at Sandia National Laboratories

A DOE user facility dedicated to energy science and technology for the twenty-first century





Combustion and energy security

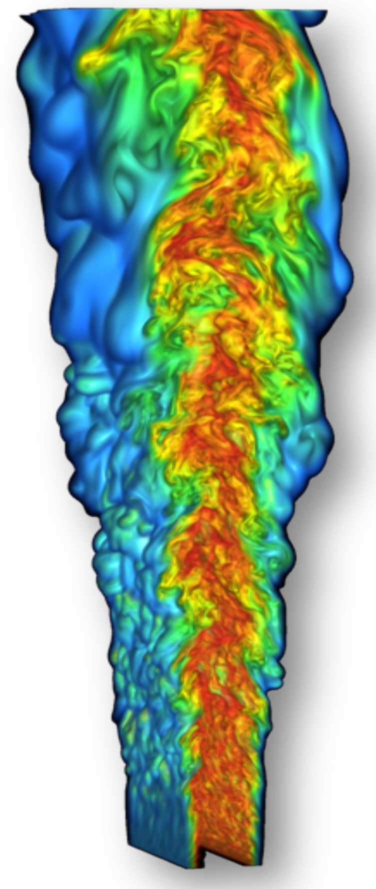
- Combustion accounts for a majority of energy used in the United States

- Computer simulations provide tools for design of efficient clean burning devices

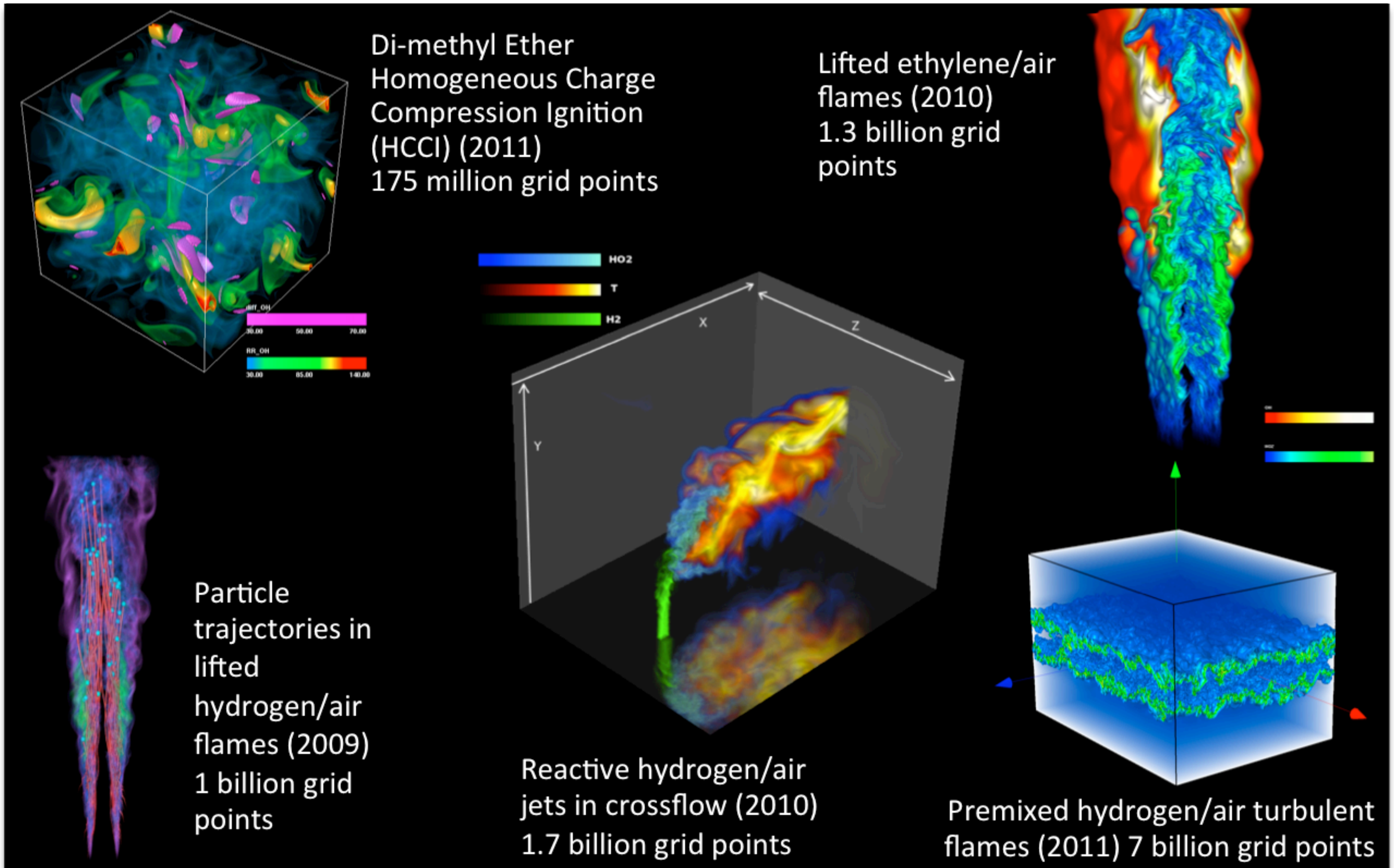
- Sound scientific understanding is necessary to develop predictive, validated multi-scale models

DNS simulations are used to study fundamental turbulence-chemistry interactions

- Turbulence
  - Entrains, advects, strains and wrinkles a flame creating more area for burning
  - Causes molecular mixing of reactants
- Chemical reactions
  - Are enhanced with mixing to a limit → extinction
  - Create heat release
- Heat release, dilatation
  - Reduce turbulence intensity through density, and property changes



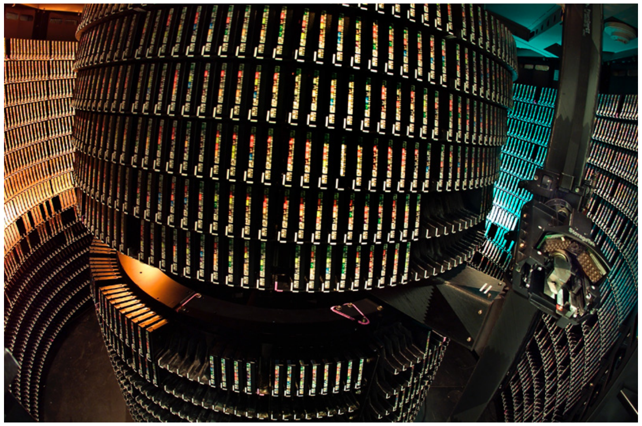
Simulation benchmark data for model development and validation with S3D

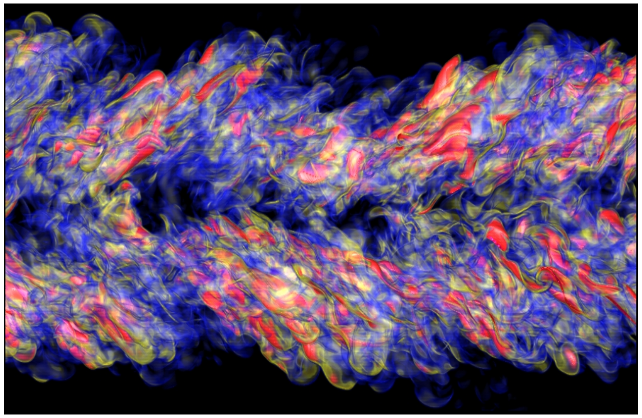


Data analysis is complicated by multiple factors

- Data size
  - Billions of grid points per time step
  - Hundreds of time steps written to disk
- Data complexity
  - Multivariate
  - Turbulence is a complex phenomenon
  - Length scales: microns to centimeters
  - Temporal scales: nanoseconds to milliseconds

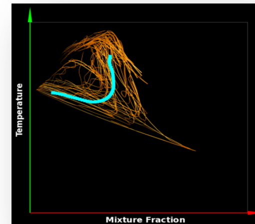
HPSS storage facility at NERSC

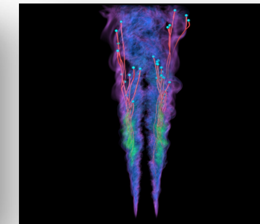


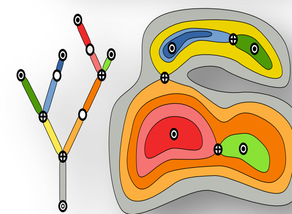


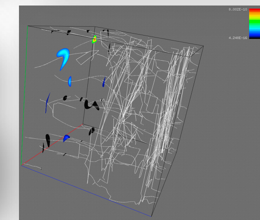
We want to analyze the data in a variety of ways

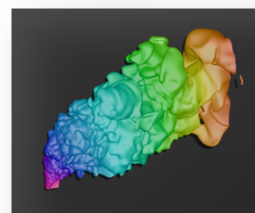
- Feature-based techniques
  - Particle trajectory clustering
  - Features defined by function thresholds
    - Threshold values not necessarily known *a priori*
    - Want to track features across time steps
    - Compute conditional statistics
      - Summaries of small-scale intermittent structures
      - Based on position within a large-scale structure
- Visualization
- DNS as a sub-grid model for LES/RANS

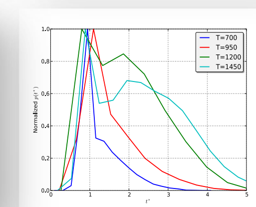


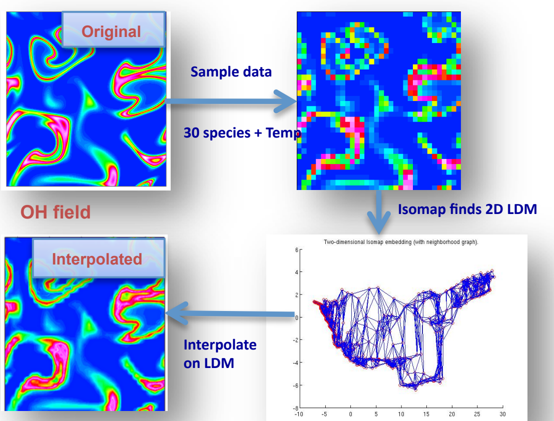


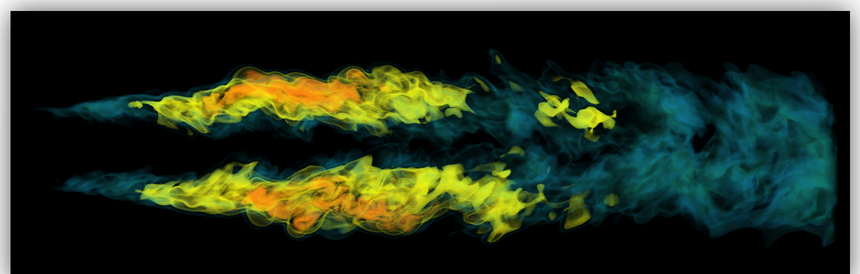






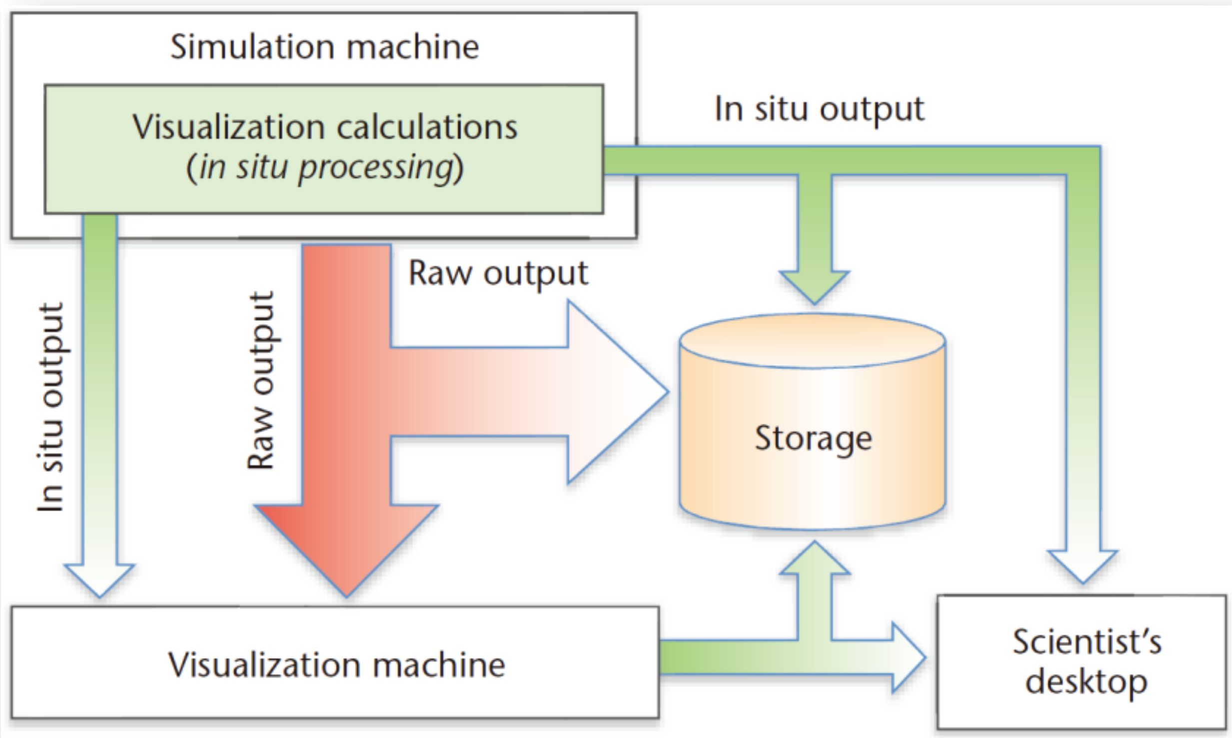






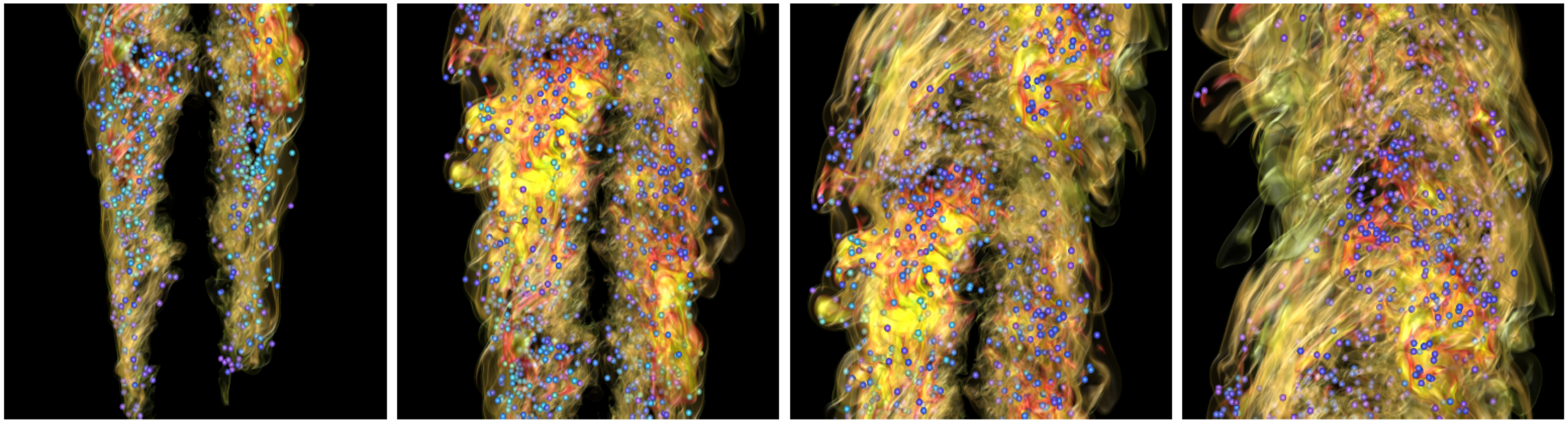
In situ analysis challenges

- Share data structures between simulation and analysis codes to optimize memory usage
- Develop scalable analysis methods constrained to the simulation data partitioning
- Develop analysis modules that can be run at a very small fraction of simulation time



In situ visualization

- Developed an interface to ease the integration of simulation and visualization modules & optimize memory usage
- Highly scalable parallel volume rendering, particle rendering, and image compositing
- Visualization time is less than 1% of simulation time when performed every 10^{th} time step *(based on the experimental results with 15,360 cores, 1620x1280x320 volume size, and 1024x1024 image size on JaguarPF at ORNL)*

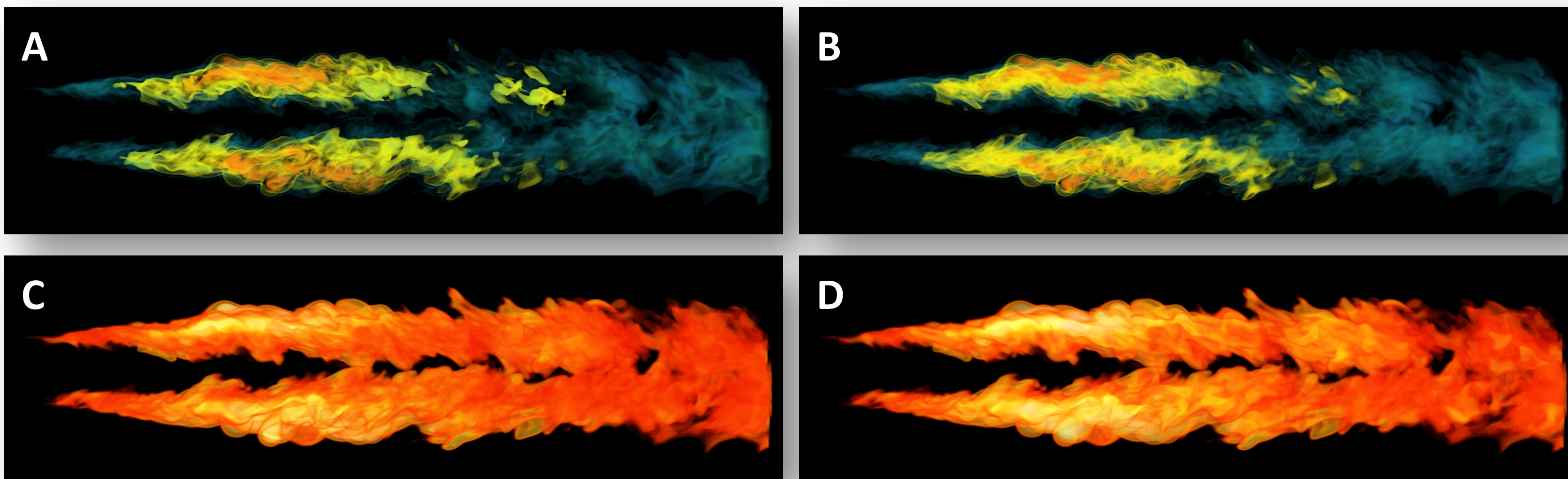


Selected zoomed-in views of mix rendering of volume and particle data (volume variable CH2O and particle variable HO2)

In-Situ Visualization for Large-Scale Combustion (H. Yu, C. Wang, R. Grout, J. Chen, K.-L. Ma. IEEE Computer Graphics and Applications. vol. 30 no. 3, pp. 45-57, 2010)

In situ visualization

- Developed a parallel algorithm to efficiently construct a compact representation, named Ray Attenuation Functions (RAF), of original volume data
  - Size of RAF can be four orders of magnitude less than original volume data
  - Allow to explore transfer function space in post-processing without accessing original volume data
- Construction time is around 1.5% of simulation time if construction is performed every 10th time step *(based on the experimental results with 6,480 cores, 1215x960x240 volume size, and 1024x1024 image size on JaguarPF at ORNL)*

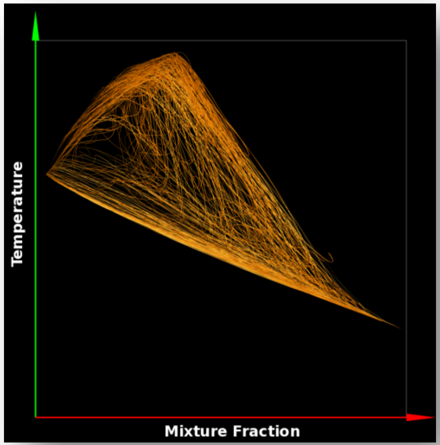


(A) and (C) are rendered images using RAF. (B) and (D) are ground truth volume rendered images using original data.

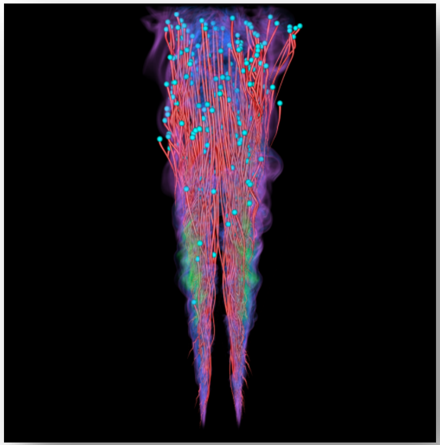
A Preview and Exploratory Technique for Large-Scale Scientific Simulations (A. Tikhonova, H. Yu, C. Correa, J. Chen, K.-L. Ma. In Proceedings of Eurographics Parallel Graphics and Visualization Symposium, April 2011)

Dual space analysis of turbulent combustion particle data

- Millions of particles are tracked to capture dynamic flame behaviour
  - Trajectories track particle positions in physical space
  - Time series curves characterize the evolution of particle thermo-chemical states in phase space
- Goal: determine relationship between time series curves and trajectories
- Challenges:
  - Categorizing and understanding the large amount of particle histories is non-trivial
  - Substantial clutter introduced by heavily interweaving of dense lines



Time series curves in phase space

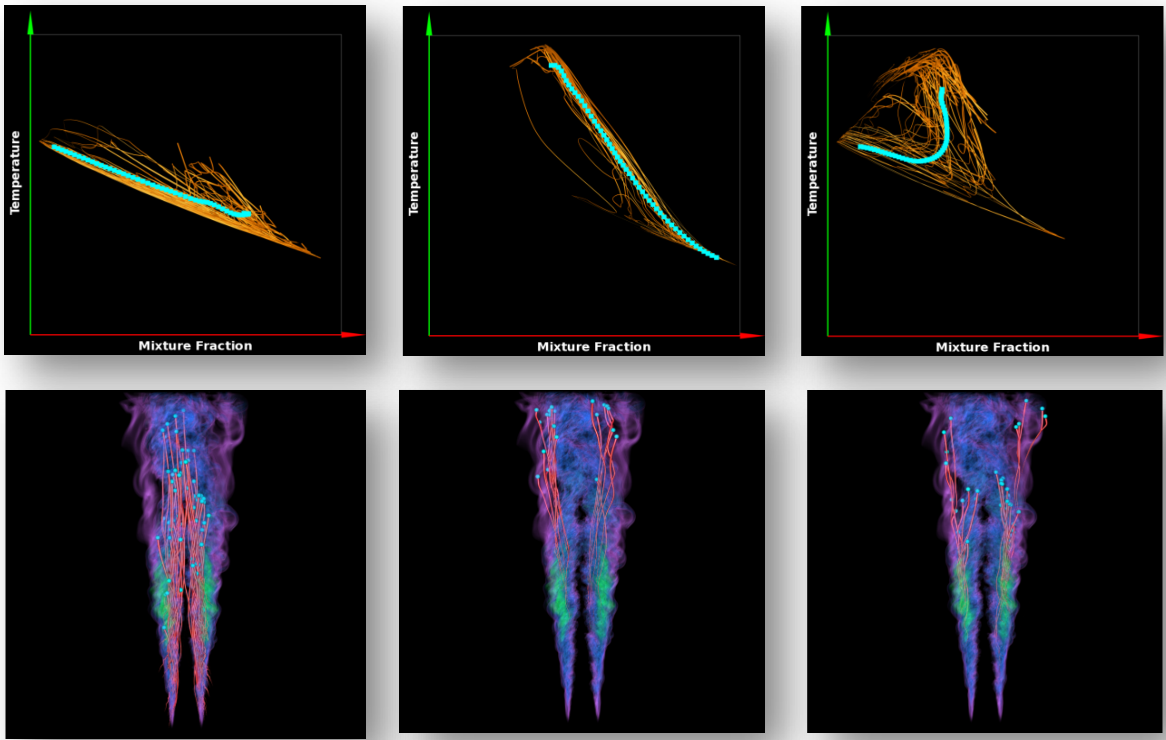


Trajectories in physical space

Dual Space Analysis of Turbulent Combustion Particle Data (J. Wei, H. Yu, R. Grout, J. Chen, K.-L. Ma, IEEE Pacific Visualization Symposium 2011)

Dual space analysis of turbulent combustion particle data

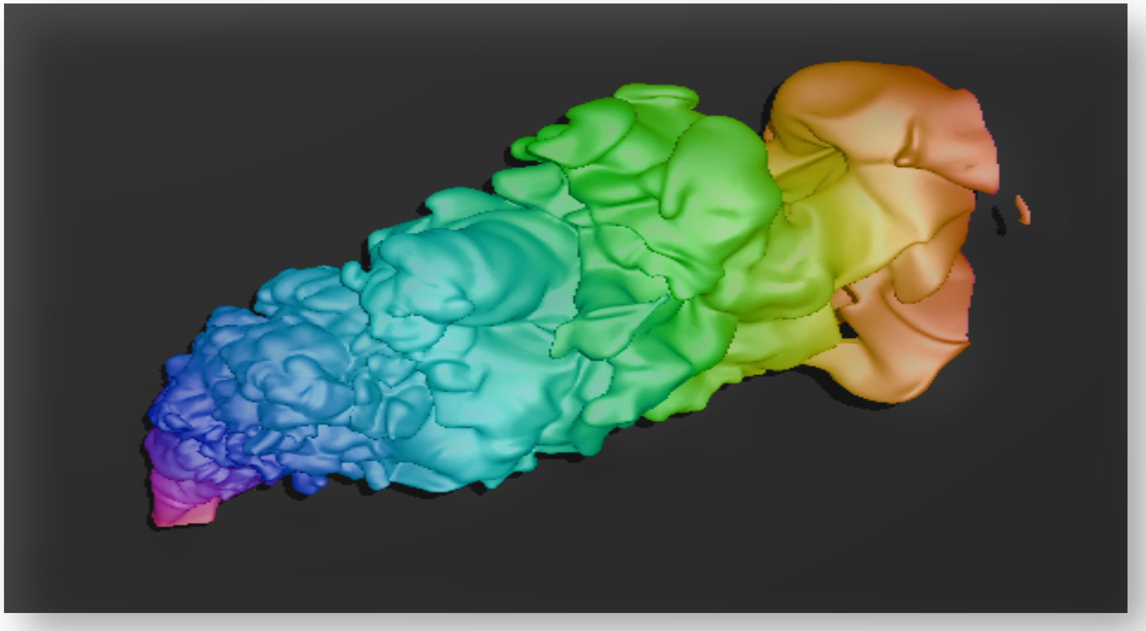
- Automatic clustering initializes the classification
- Semi-supervised clustering refines the classification
- Visualization renders classified particle movement simultaneously in the phase and physical spaces



Dual Space Analysis of Turbulent Combustion Particle Data (J. Wei, H. Yu, R. Grout, J. Chen, K.-L. Ma, IEEE Pacific Visualization Symposium 2011)

Conditional analysis using global shape characteristics

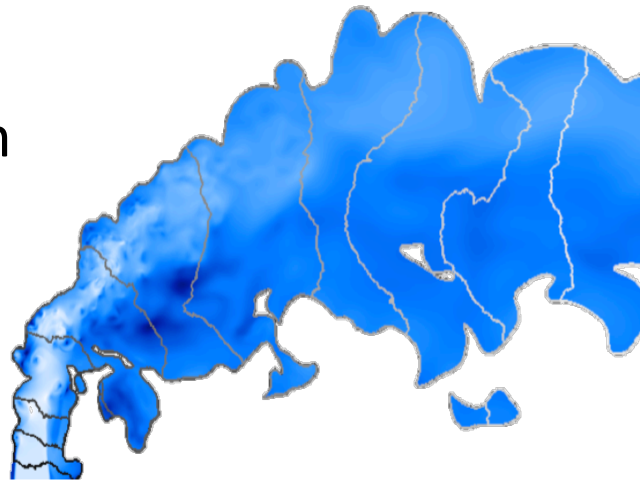
A jet-based coordinate system allows for the aggregation of averages conditioned on bulk flame position.



Heat Release and Turbulence Statistics From a Direct Numerical Simulation of a Reacting Jet in Cross-Flow Parameterized in a Jet Natural Coordinate System Developed From Scalar Quantities, R. Grout, P.-T. Bremer, A. Gruber, J. Bennett, J. Chen, A. Gyulassy, 13th International Conference on Numerical Combustion, Apr. 2011)

A direct numerical simulation study of turbulence and flame structure in a round jet in cross-flow (R. Grout, A. Gruber, H. Kolla, J. Bennett, P.-T. Bremer, A. Gyulassy, and J. Chen, submitted to J. Fluid Mech. 2011)

Conditional analysis using global shape characteristics



- Reduce mesh resolution using Qslim
- Solve Laplace's equation on the simplified jet surface
- Tetrahedralize jet interior using TetGen
- Extend surface parameterization to jet interior by solving Laplace's equation again
- Compute parameter values for all grid points inside the jet

Heat Release and Turbulence Statistics From a Direct Numerical Simulation of a Reacting Jet in Cross-Flow Parameterized in a Jet Natural Coordinate System Developed From Scalar Quantities, R. Grout, P.-T. Bremer, A. Gruber, J. Bennett, J. Chen, A. Gyulassy, 13th International Conference on Numerical Combustion, Apr. 2011)

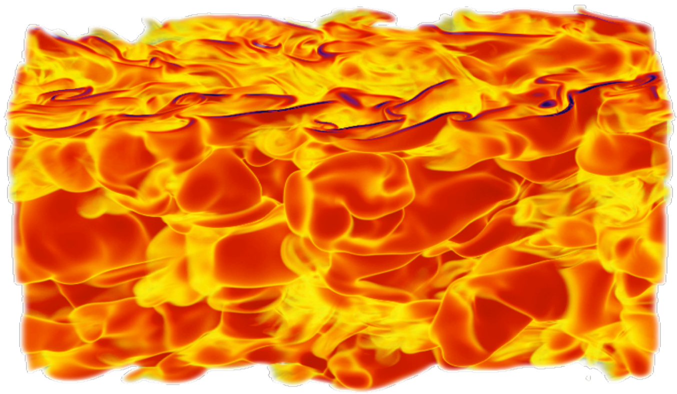
A direct numerical simulation study of turbulence and flame structure in a round jet in cross-flow (R. Grout, A. Gruber, H. Kolla, J. Bennett, P.-T. Bremer, A. Gyulassy, and J. Chen, submitted to J. Fluid Mech. 2011)

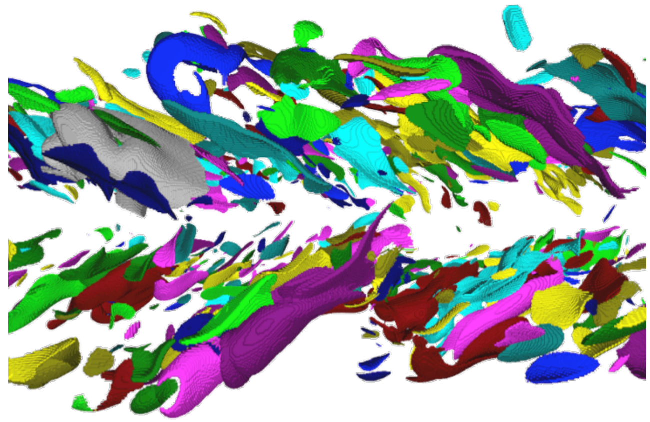
Conditional analysis using local shape characteristics

- Scalar dissipation rate χ indicates the rate of molecular mixing

- Goal: Understand relationship between the thickness and the mean temperature within features

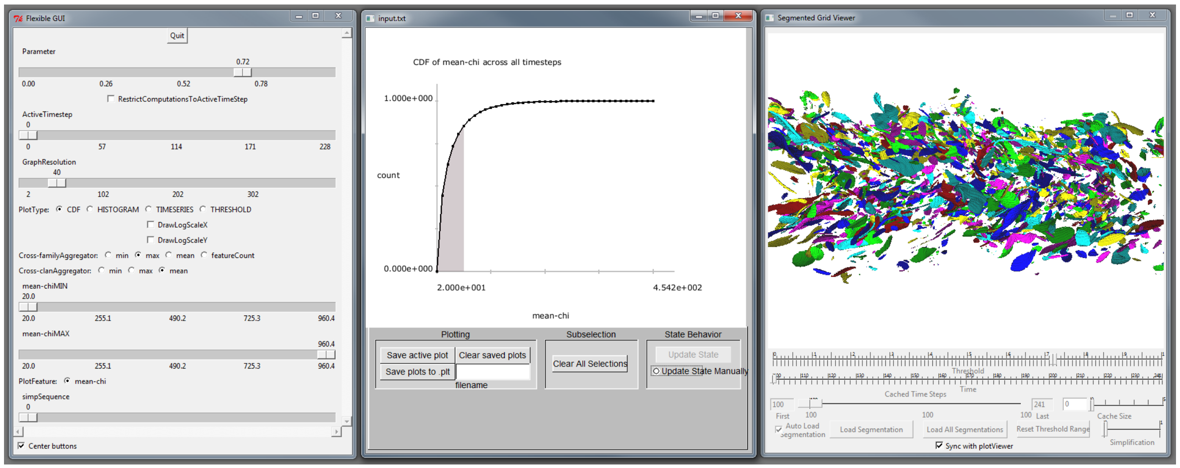
- Challenges:
  - Scalar dissipation structures are defined by locally varying isovalues of a function
  - Sub-selection based on other criteria is important
  - Visual feedback of the effect of parameter choices is desired
  - Large data complicates matters





Feature-Based Statistical Analysis of Combustion Simulation Data (J. Bennett, V. Krishnamoorthy, S. Liu, R. Grout, E. Hawkes, J. Chen, J. Shepherd, V. Pascucci, P.-T. Bremer, accepted to IEEE Visualization 2011)

Integrated analysis & visualization framework for computing statistical summaries

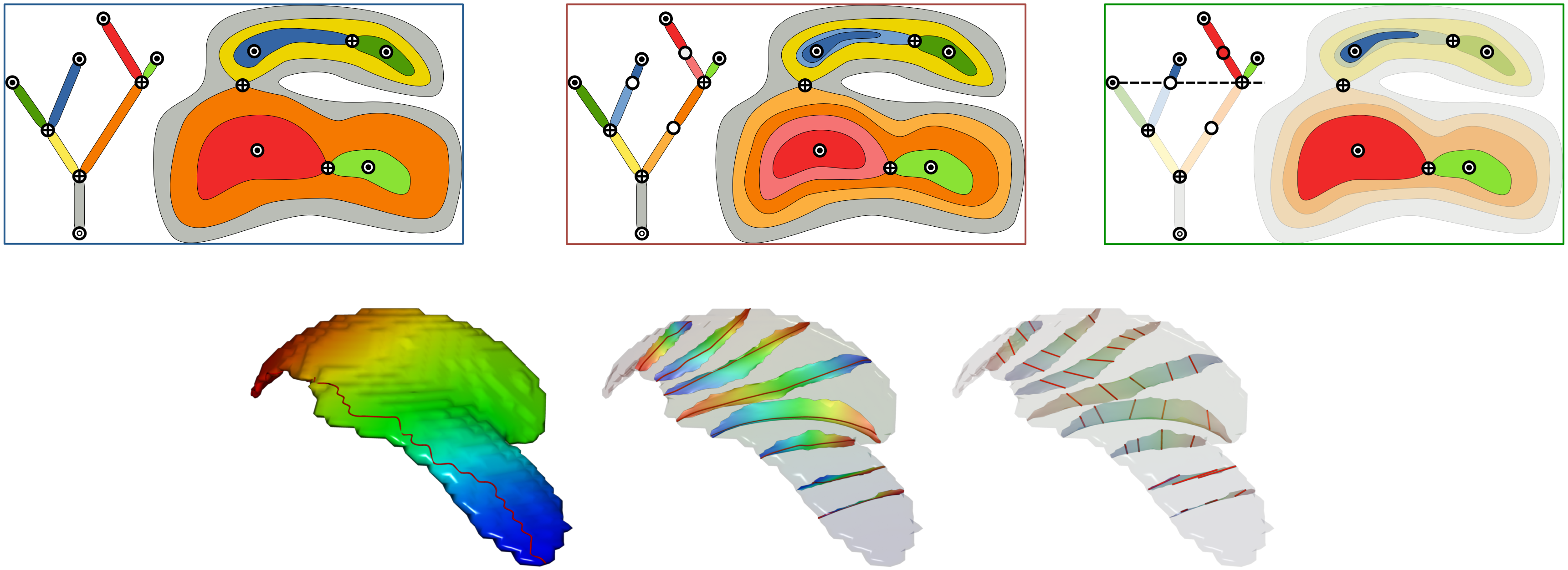


- On the fly aggregation of feature-based spatial and temporal statistics
- A system for interactive creation of spatial and temporal statistical summaries
- An interactive linked view display of statistics and features with picking and highlighting

Feature-Based Statistical Analysis of Combustion Simulation Data (J. Bennett, V. Krishnamoorthy, S. Liu, R. Grout, E. Hawkes, J. Chen, J. Shepherd, V. Pascucci, P.-T. Bremer, accepted to IEEE Visualization 2011)

Sandia National Laboratories

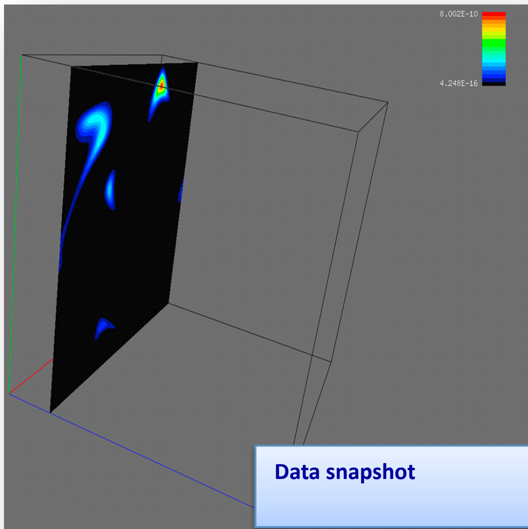
Augmented feature families using merge trees

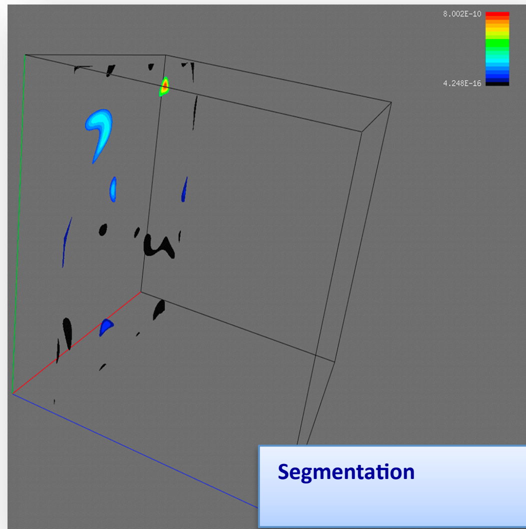


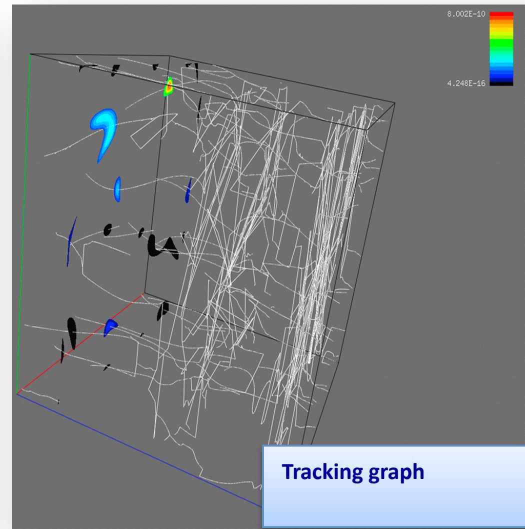
By pre-computing statistics of interest we create a compact representation of the simulation data that can be explored interactively on commodity desktops without affecting the quality of the resulting analysis.

Feature-Based Statistical Analysis of Combustion Simulation Data (J. Bennett, V. Krishnamoorthy, S. Liu, R. Grout, E. Hawkes, J. Chen, J. Shepherd, V. Pascucci, P.-T. Bremer, accepted to IEEE Visualization 2011)

Tracking features in space and time



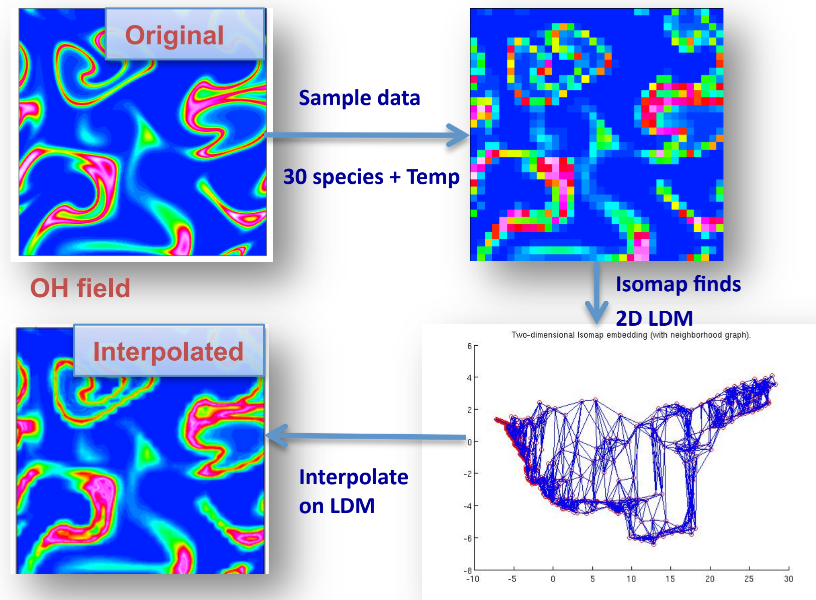




Direct Numerical Simulation of Auto-ignition in Stratified Turbulent Dimethyl Ether/Air Mixtures (G. Bansal, A. Mascarenhas, J. Chen, et al. to be submitted to Combustion and Flame.

Towards sub-grid modeling

- Detect intrinsic low-dimensional manifolds (LDM) in turbulent combustion data using an *Isomap* based technique.
- Enable high-fidelity interpolation on LDM to recover samples thus enabling data-reduction.
- LDM can potentially serve as novel sub-grid model for use in LES.

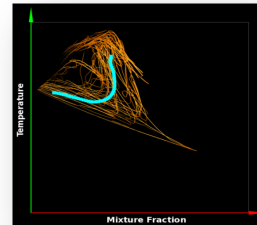


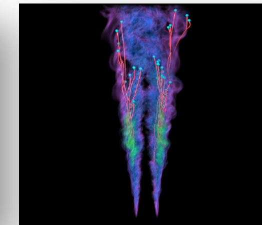
- Established that high-dimensional (~30D) HCCI combustion data of Di-Methyl Ether can be embedded in low dimensions (< 4D)

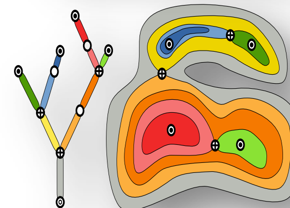
Identification of Intrinsic Low Dimensional Manifolds in Turbulent Combustion using an Isomap based technique (G. Bansal, A. Mascarenhas, J. Chen. Accepted in "7th US National Technical Meeting of the Combustion Institute", March 2011. Extended journal version in preparation)

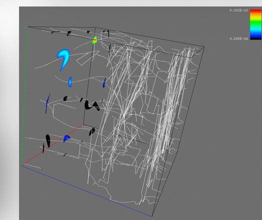
Future work

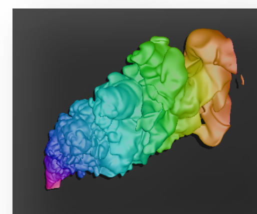
- Move analysis *in situ* when possible!!!
- Research required to:
  - Get analysis algorithms to scale
  - Work with the simulation data partitioning constraints
  - Develop analysis modules that can be run at a very small fraction of simulation time

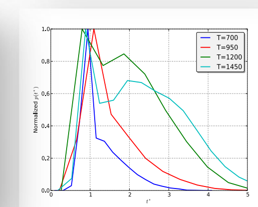


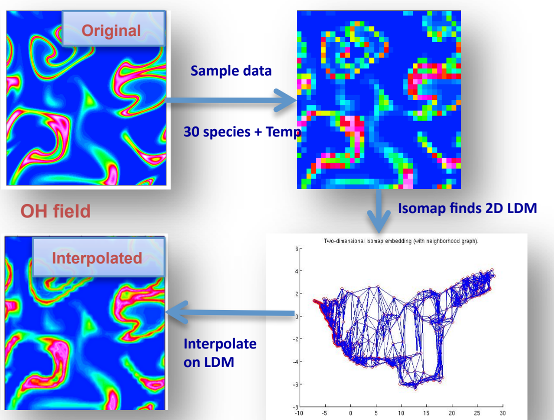


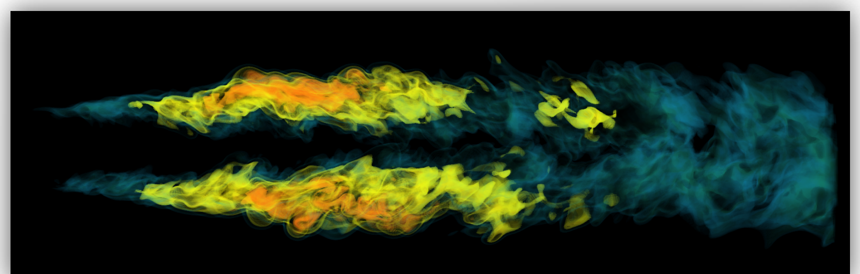












Questions?

Contact:

Janine Bennett

Sandia National Laboratories

jcbenne@sandia.gov

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.