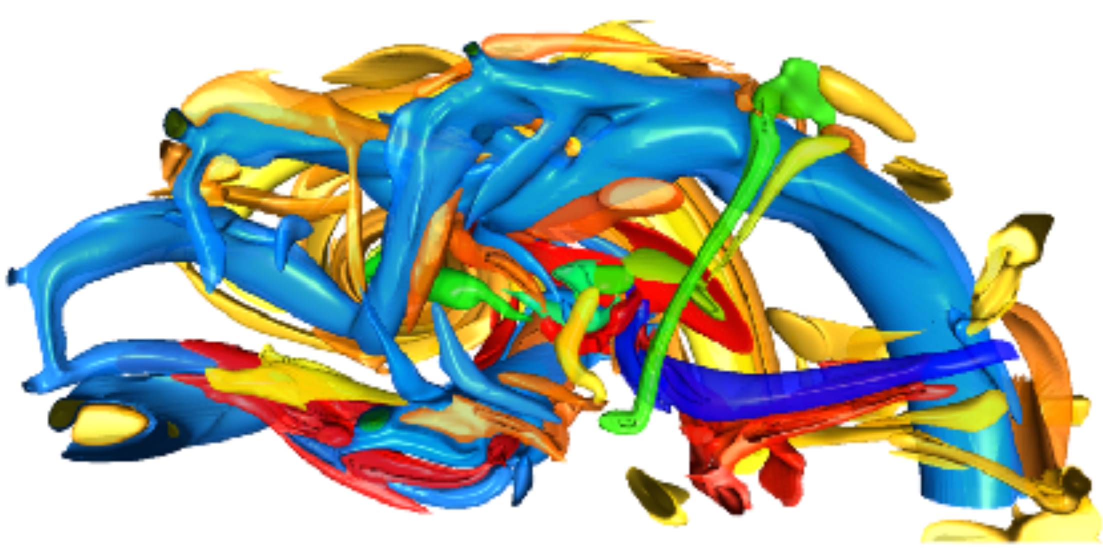
Vector Field Techniques for Large-Scale Data

Christoph Garth

Computational Topology Group
University of Kaiserslautern

Tutorial on
Recent Advancements of Feature-based Flow Visualization and Analysis

IEEE VIS 2016



Outline

- **Part I**
  Parallel Algorithms for Integral Curves
- **Part II**
  In Situ Techniques
- **Part III**
  Features in Large Ensembles

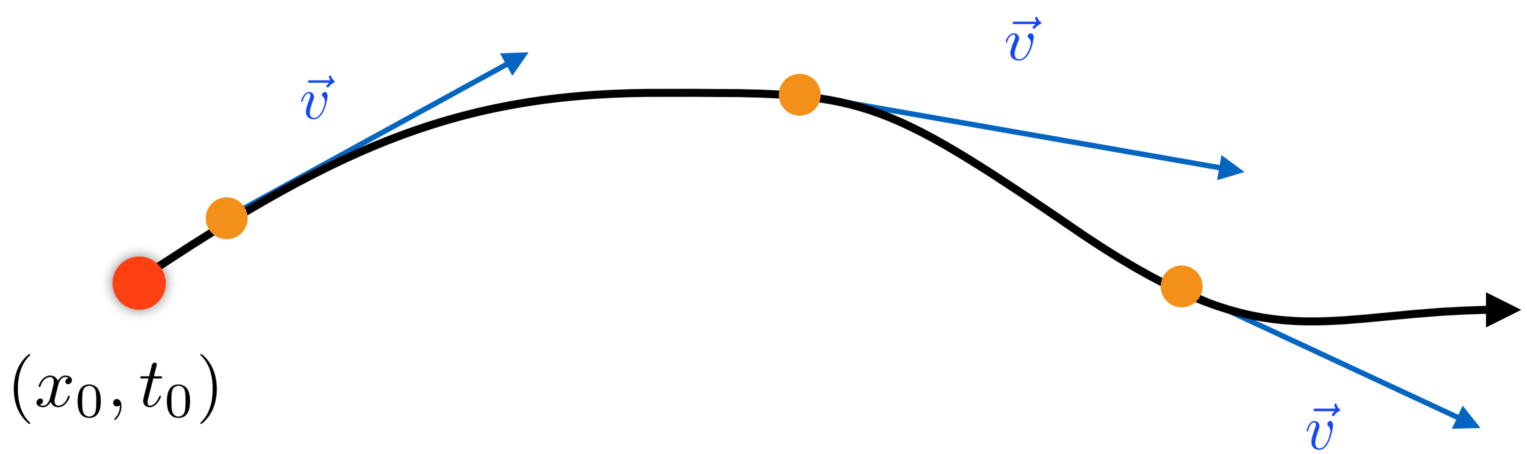
Part I. Parallel Algorithms for Integral Curves

Integral Curves

Integral curves (or trajectories, orbits):
solutions of ordinary differential equation with vector field right hand side.

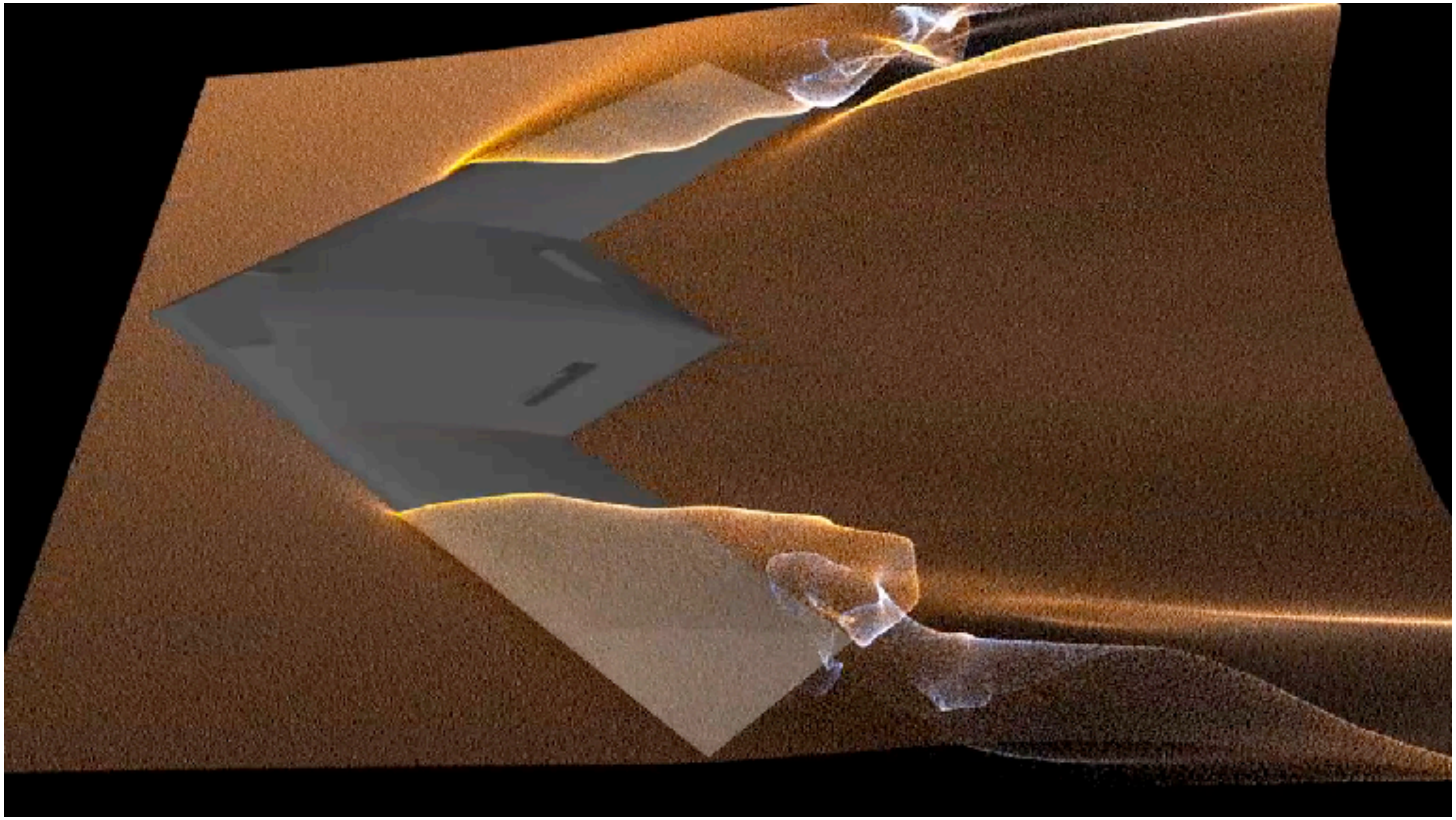
$$\dot{s}(t) = \vec{v}(t, x)$$
$$s(t_0) = x_0$$

I.e. curves that are at every point tangent to the vector field.



Many Integral Curves

Typical (naïve) flow visualization: 1M particles moving with a flow.



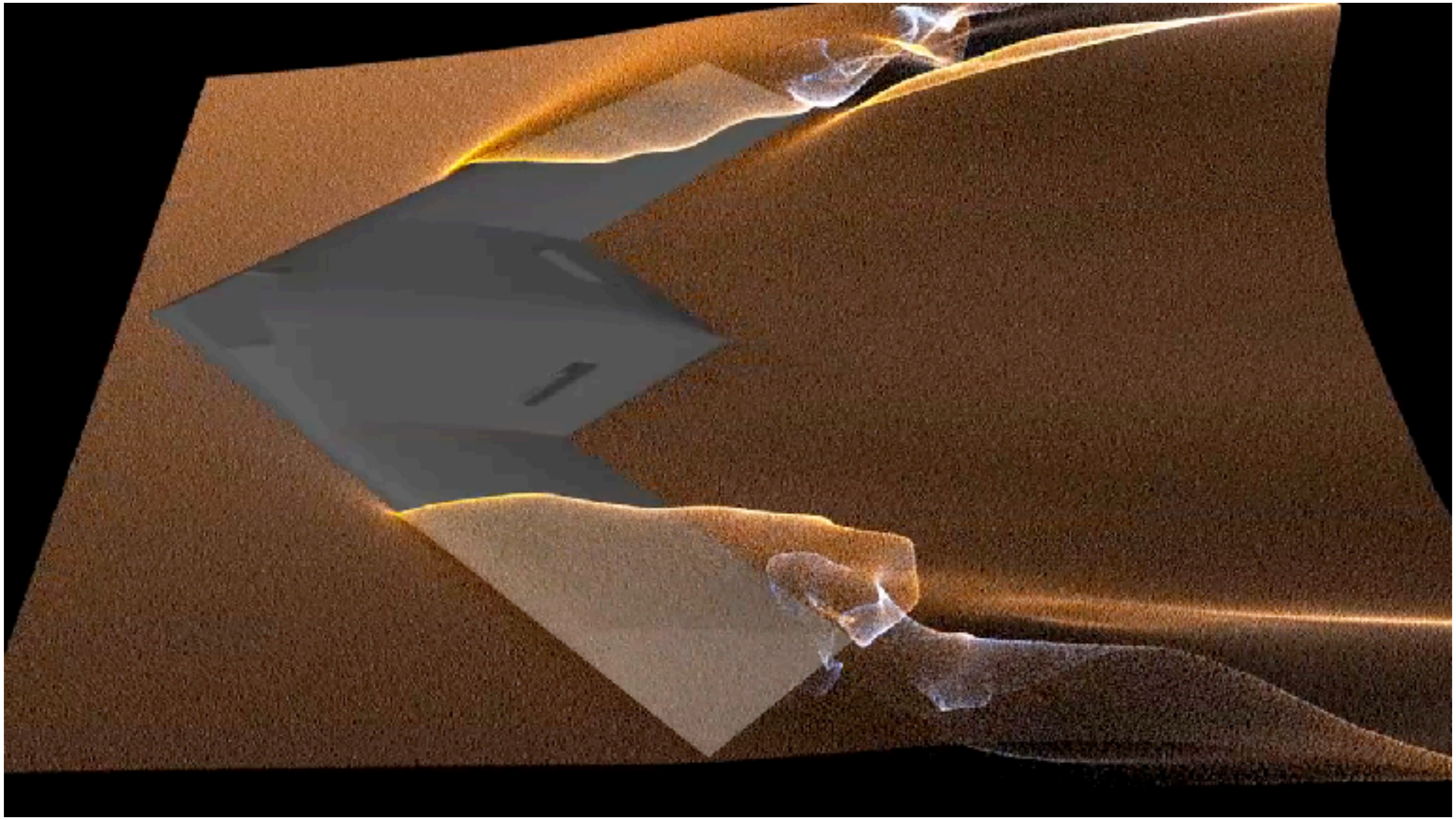
(UAV data courtesy M. Rütten, DLR Göttingen)

Real-time due to GPU parallelization (all data resides in GPU memory).

C. Garth, K. I. Joy: *Fast, Memory-Efficient Cell Location in Unstructured Grids for Visualization*. IEEE TVCG 16(6):1541-1550.

Many Integral Curves

Typical (naïve) flow visualization: 1M particles moving with a flow.



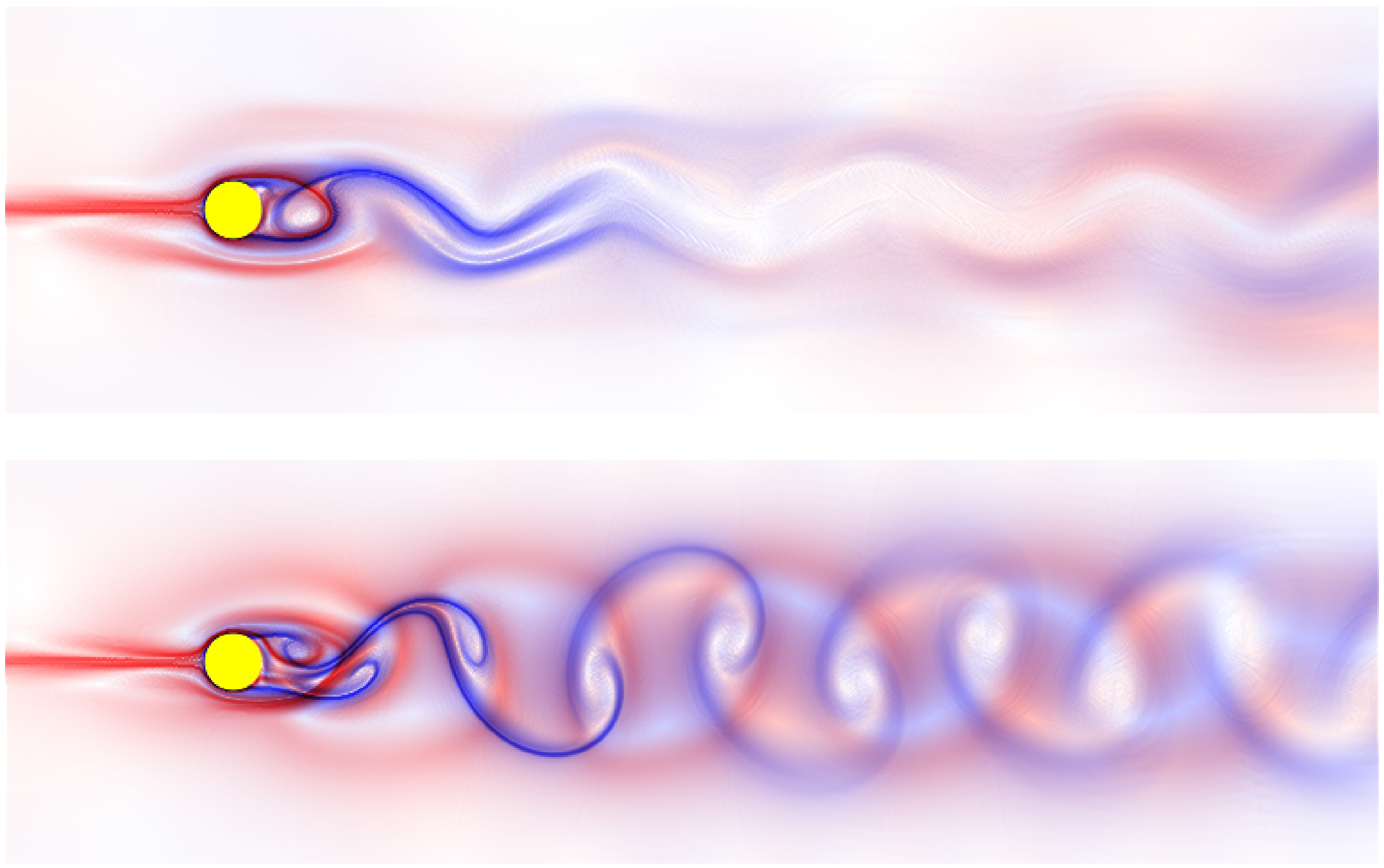
(UAV data courtesy M. Rütten, DLR Göttingen)

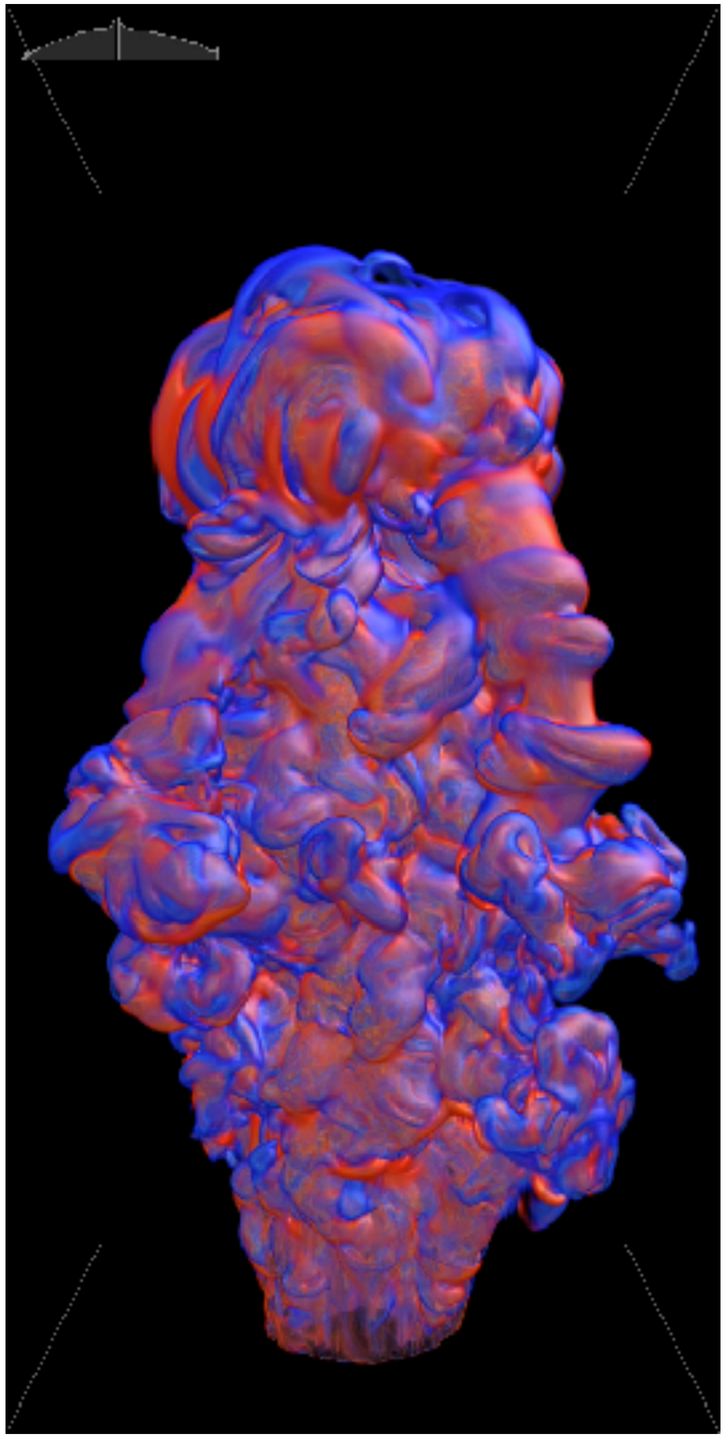
Real-time due to GPU parallelization (all data resides in GPU memory).

C. Garth, K. I. Joy: *Fast, Memory-Efficient Cell Location in Unstructured Grids for Visualization*. IEEE TVCG 16(6):1541-1550.

Integration-Based Vector Field Features

Interation-based Features:
e.g. Lagrangian Coherent Structures
(Haller, 2005; Shadden, 2009)



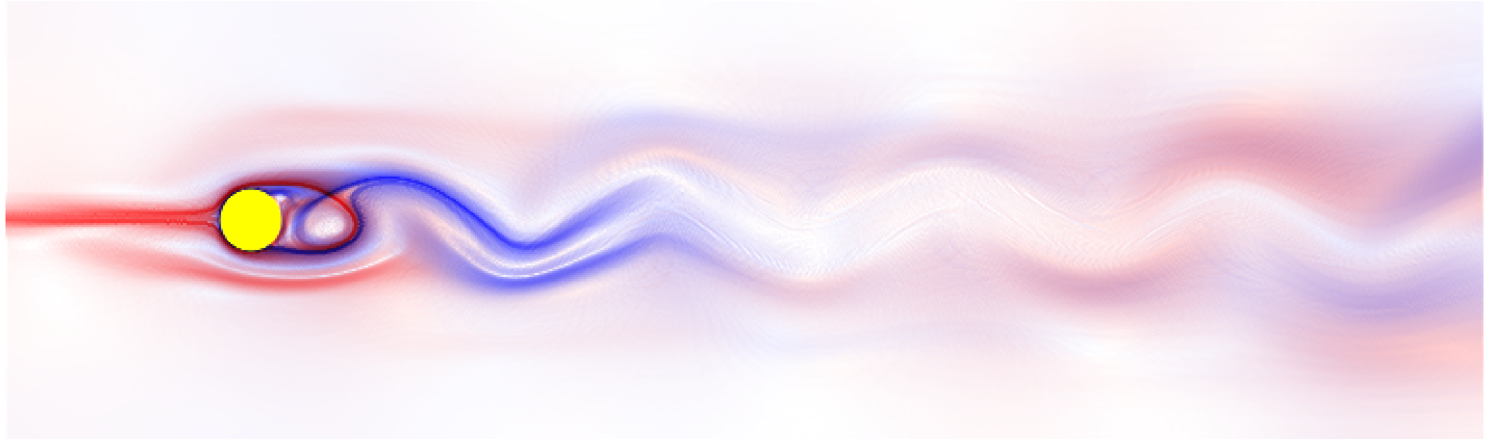


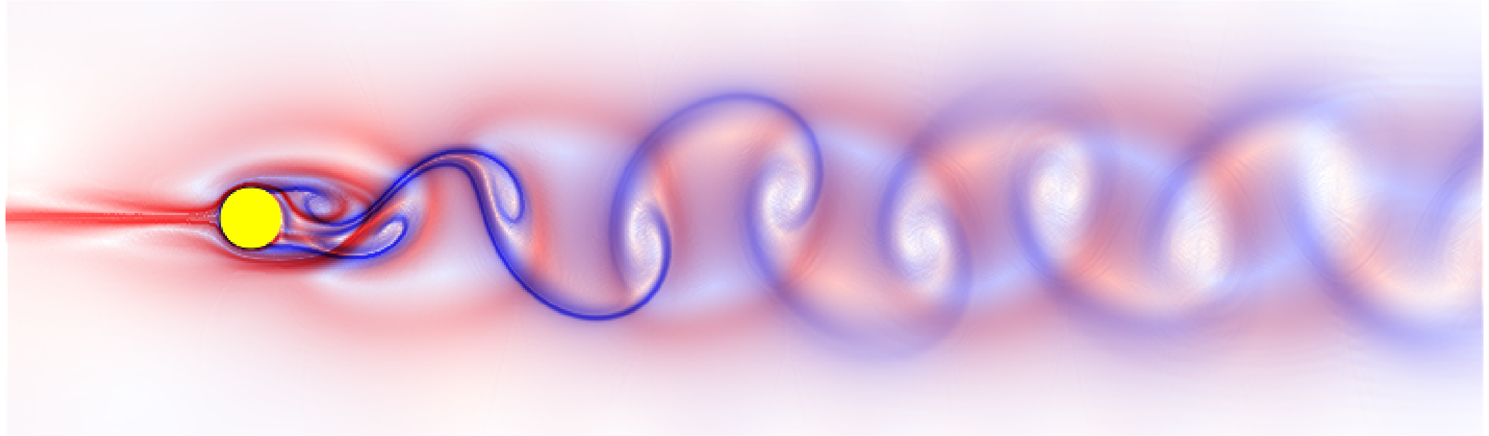
~ 50M pathlines over
~150MB vector field data

~ 100 billion pathlines over
~1 TB vector field data

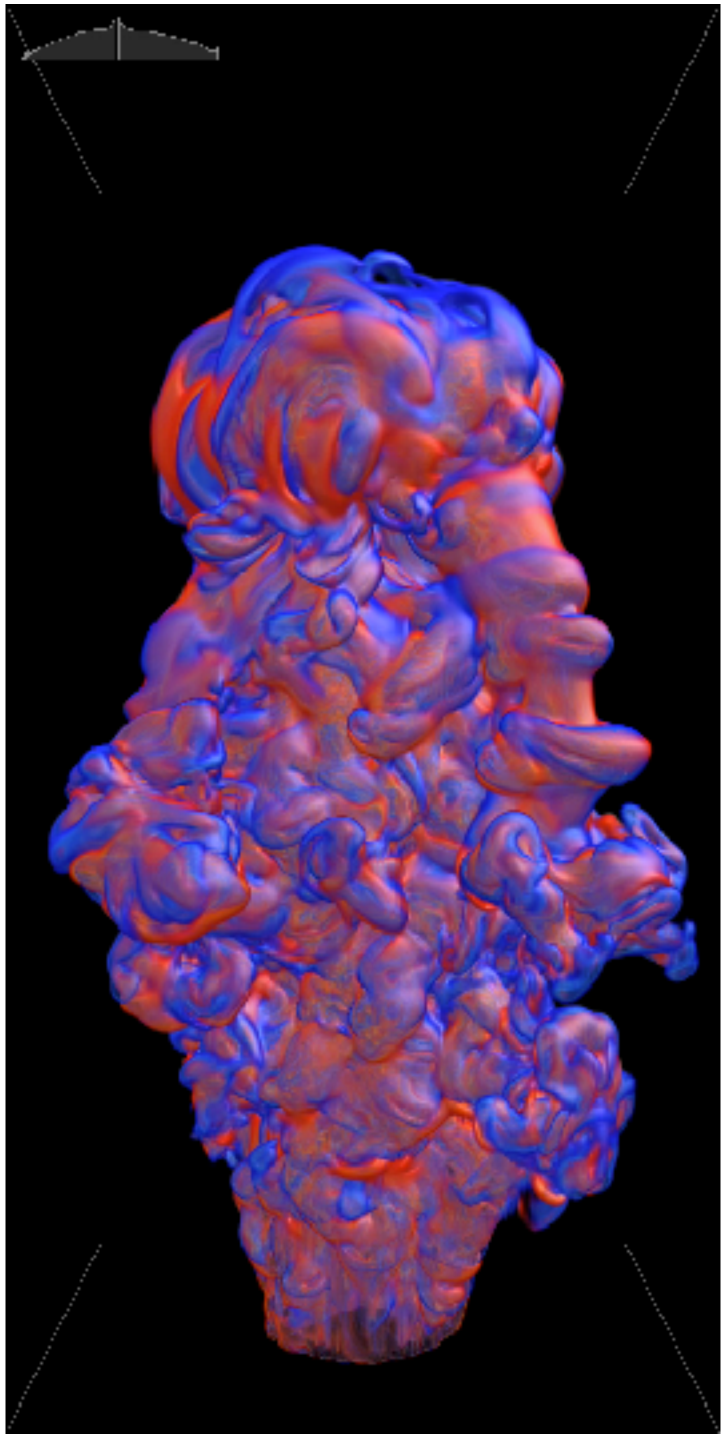
Integration-Based Vector Field Features

Interation-based Features:
e.g. Lagrangian Coherent Structures
(Haller, 2005; Shadden, 2009)





~ 50M pathlines over
~150MB vector field data



~ 100 billion pathlines over
~1 TB vector field data

Integration-Based Vector Field Features

Pathline Attributes:
pathlines with relevant attributes define features.

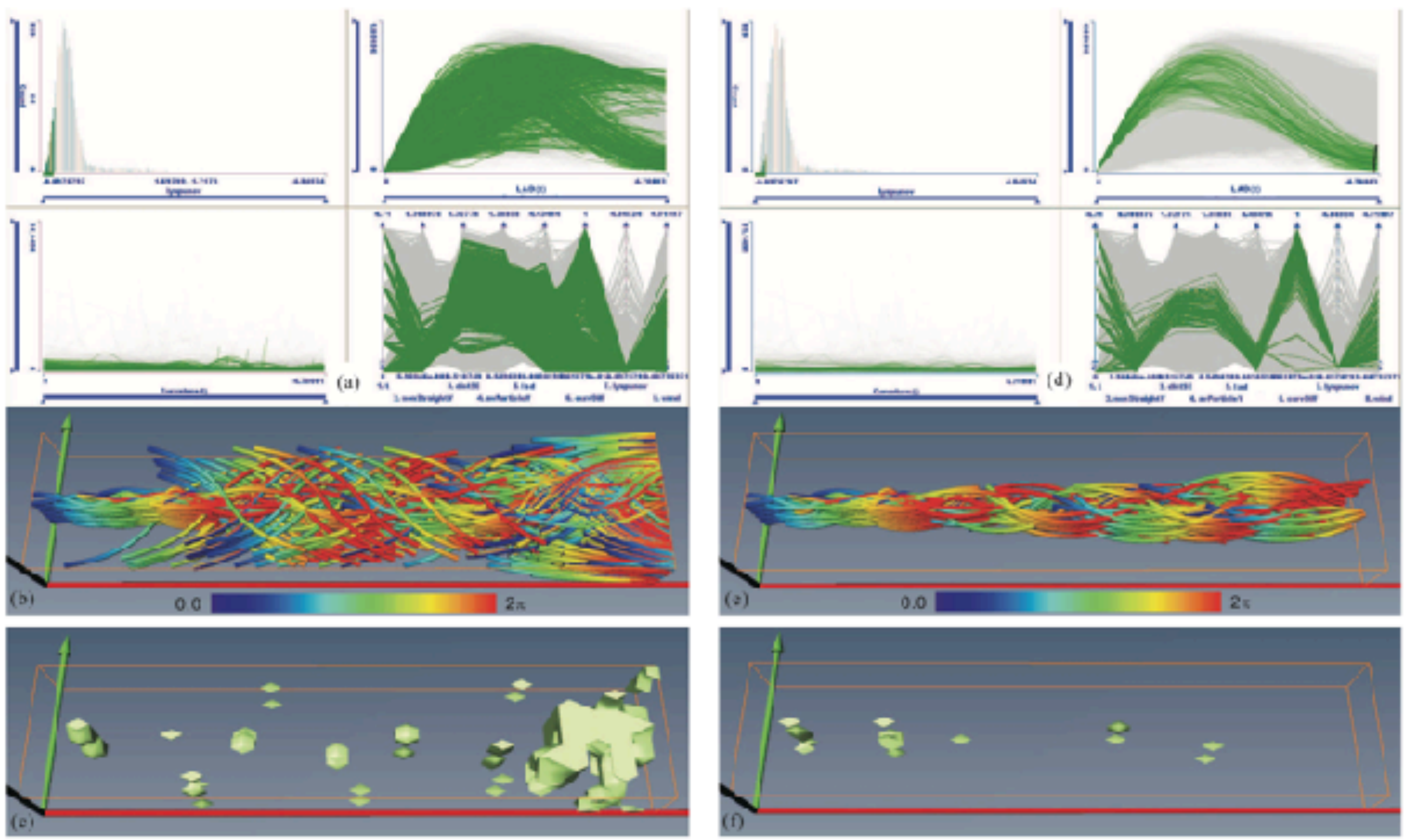
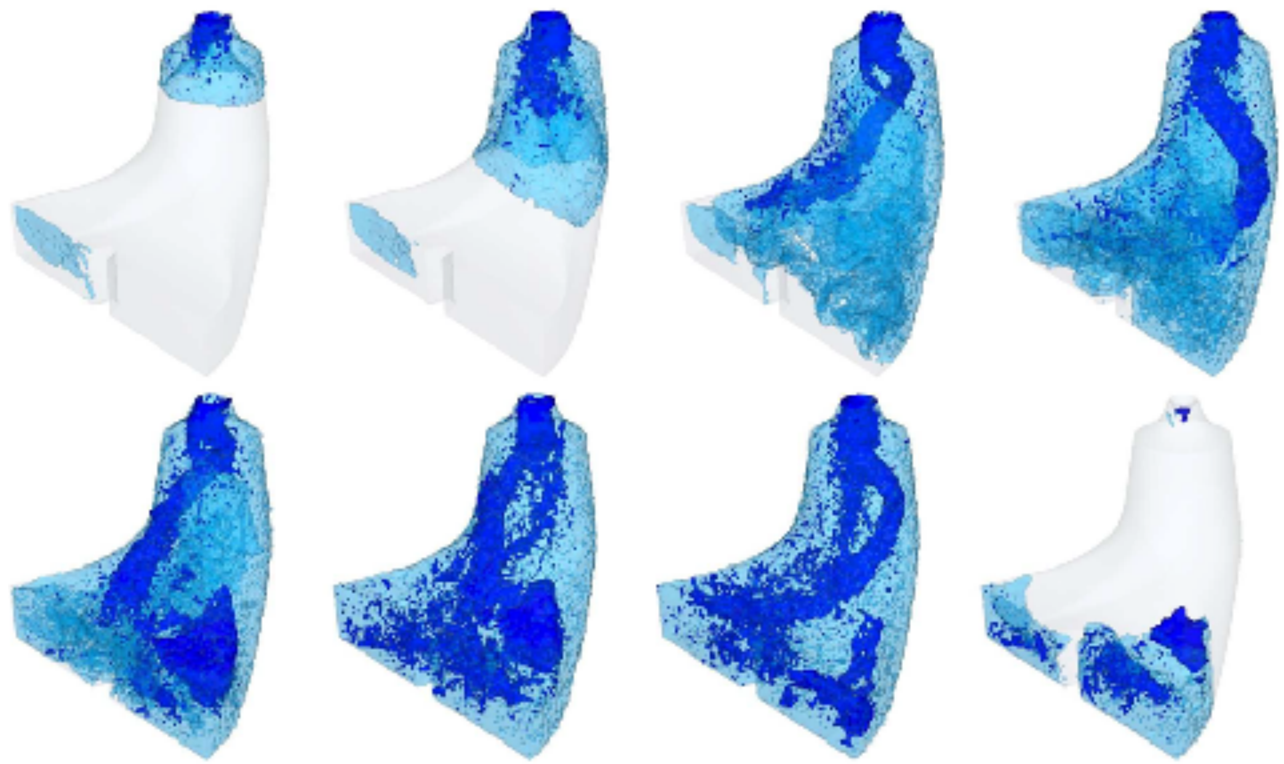


image from K. Shi, H. Theisel, H. Hauser, T. Weinkauf, K. Matkovic, H.-C. Hege, H.-P. Seidel:
Path Line Attributes - an Information Visualization Approach to Analyzing the Dynamic Behavior of 3D Time-Dependent Flow Fields.
In Topology-Based Methods in Visualization II, Springer, 2009.

Integration-Based Vector Field Features

Pathline Predicates: integral curves in combination with Boolean predicates and set operations



pathline predicates over 250,000 pathlines

Turbine dataset courtesy R. Peikert, ETH Zürich

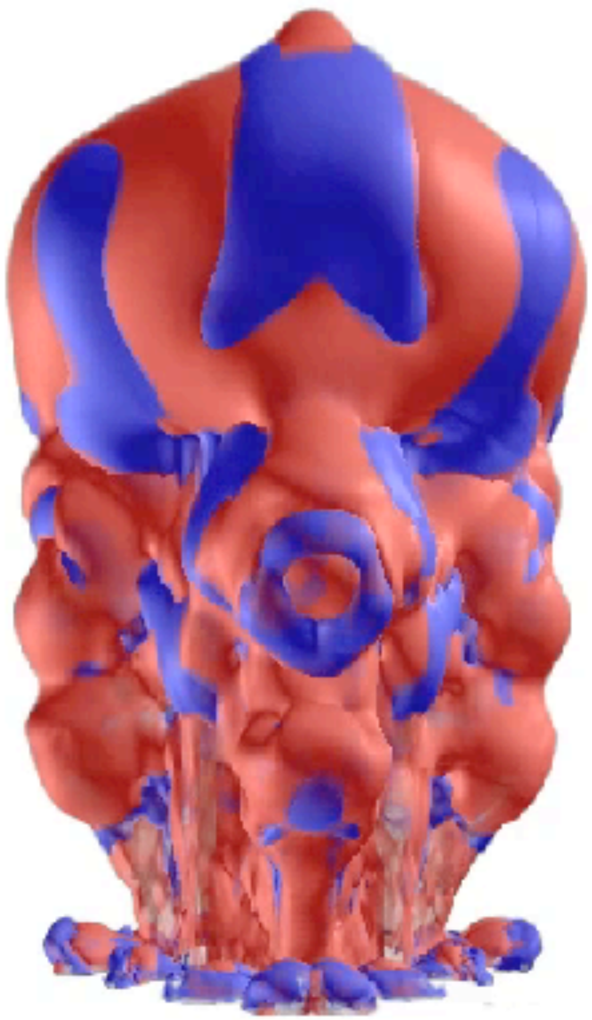
Definition and extraction of features as relevant sets of curves.

T. Salzbrunn, C. Garth, G. Scheuermann, J. Meyer. Pathline predicates and unsteady flow structures. The Visual Computer 24(12):1039–1051, 2008

GPU-based Feature Calculation

Computing many pathlines is a massively parallel problem, and ideally suited to GPUs.

Example: progressive visualization of LCS in datasets up to main memory size.

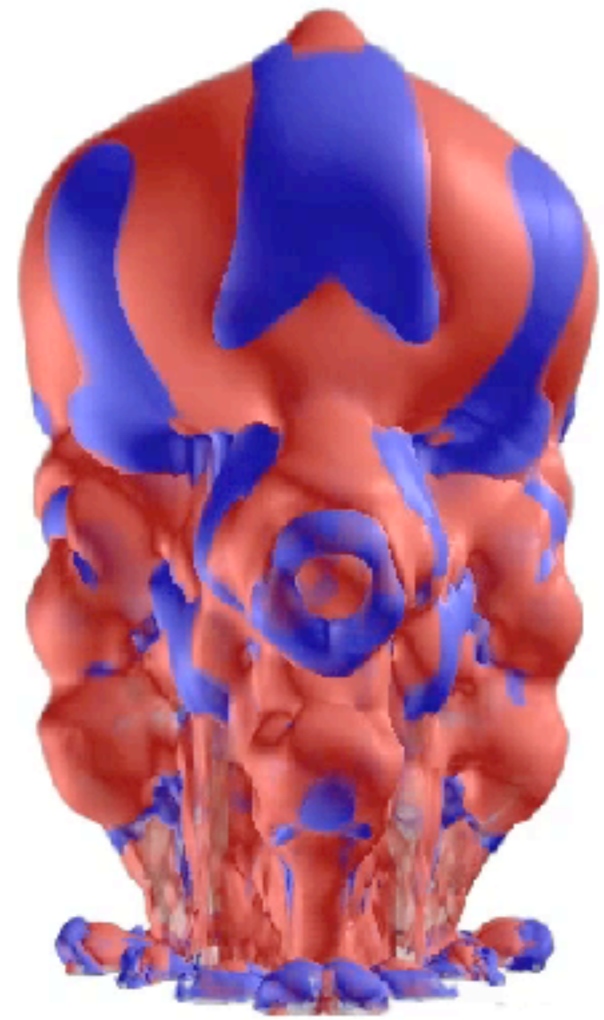


S. Barakat, C. Garth, X. Tricoche: Adaptive Refinement of the Flow Map Using Sparse Samples. IEEE TVCG 19(12):2753-2762, 2013.

GPU-based Feature Calculation

Computing many pathlines is a massively parallel problem, and ideally suited to GPUs.

Example: progressive visualization of LCS in datasets up to main memory size.



S. Barakat, C. Garth, X. Tricoche: Adaptive Refinement of the Flow Map Using Sparse Samples. IEEE TVCG 19(12):2753-2762, 2013.

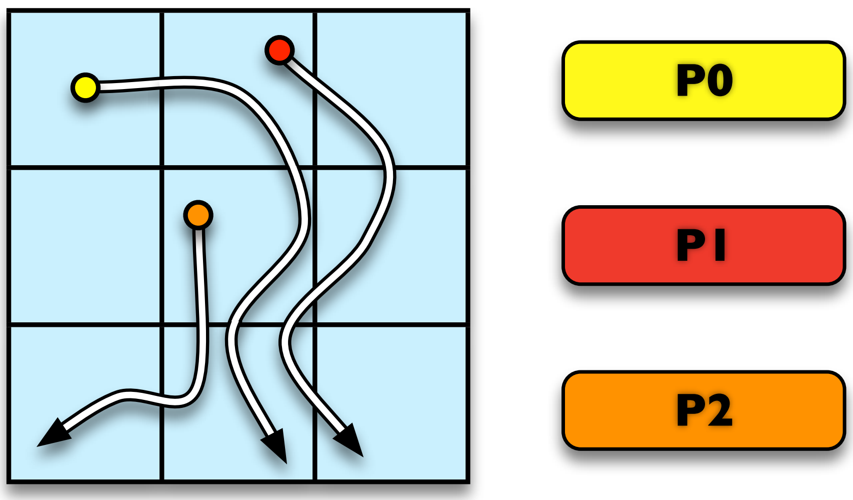
Scalable Integral Curve Algorithms

Integral Curves are conceptually simple to compute; parallelization is straightforward for **in-core** data.

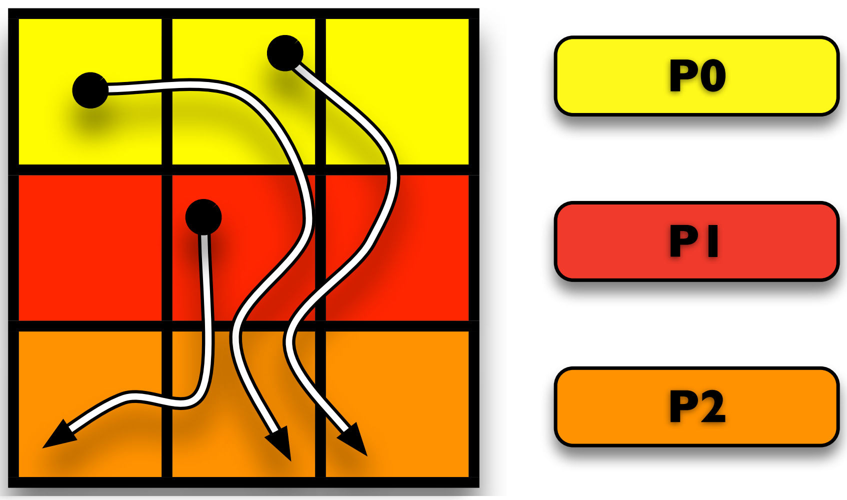
Large-scale (possibly time-varying) vector field data:

Data in blocks, all blocks do not fit in core of single machine; particles need correct data blocks for further propagation.

<u>Distributed</u> approach (cluster / supercomputer):



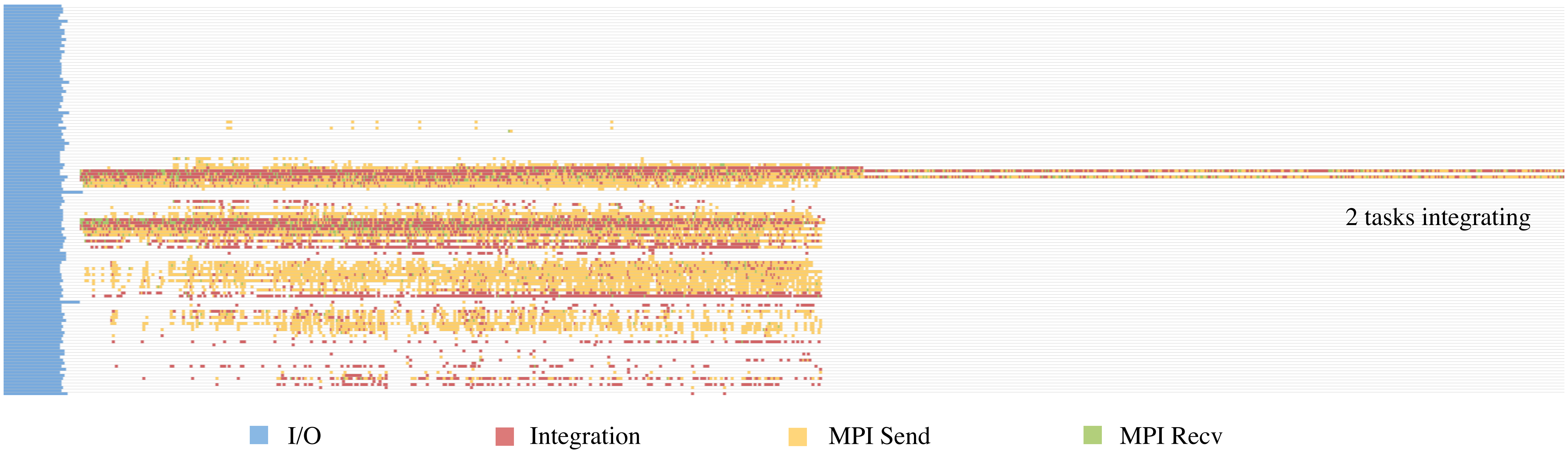
parallelize-over-seeds
distribute work



parallelize-over-blocks
distribute data

Scalable Distributed Integral Curve Algorithms

Degenerate behavior appears in both cases, leading to bad performance and scalability.



Distribution of data: Gantt chart of processor activity

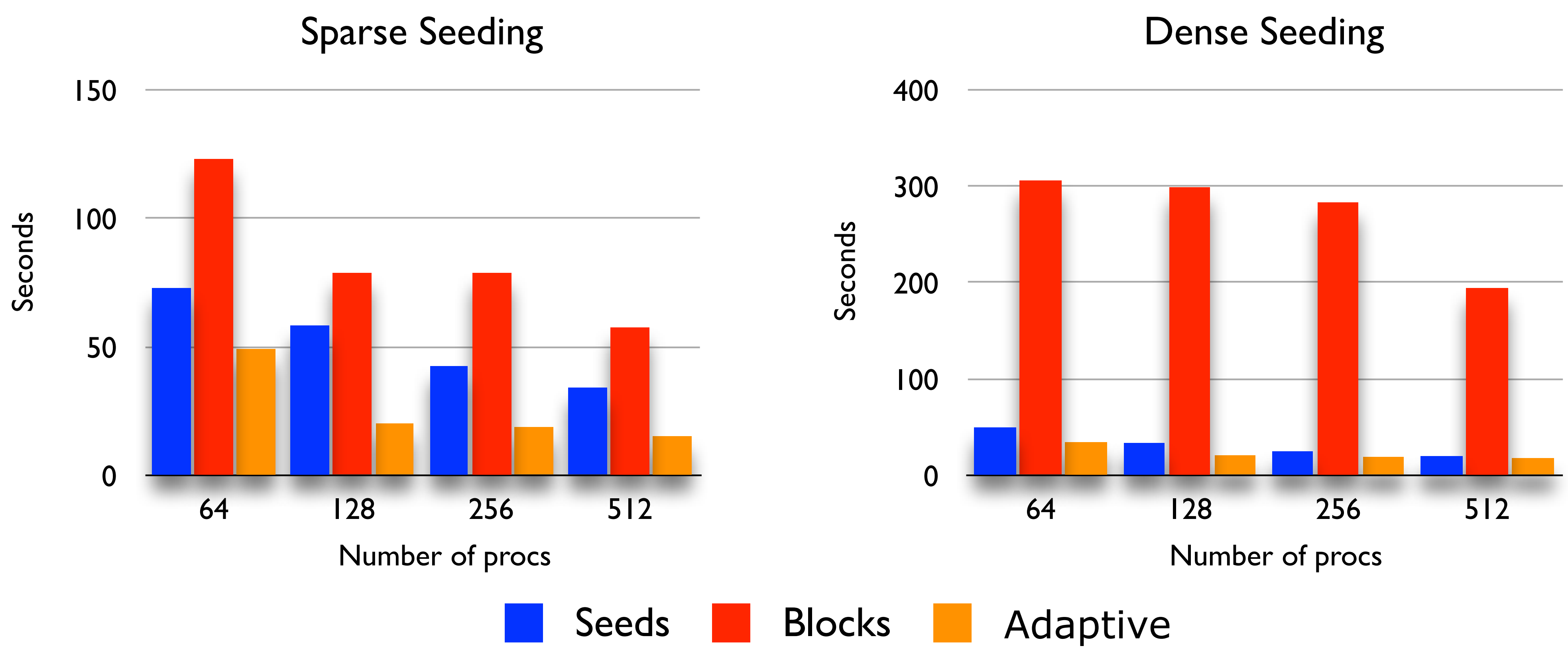
Load characteristics depend strongly on user input and data.

D. Pugmire, H. Childs, C. Garth, S. Ahern, G. H. Weber. *Scalable Computation of Streamlines on Very Large Datasets.* In Proc. ACM Supercomputing (SC'09), 2009.

Scalable Distributed Integral Curve Algorithms

Adaptive load balancing:
distribute work and data, redistribute when needed

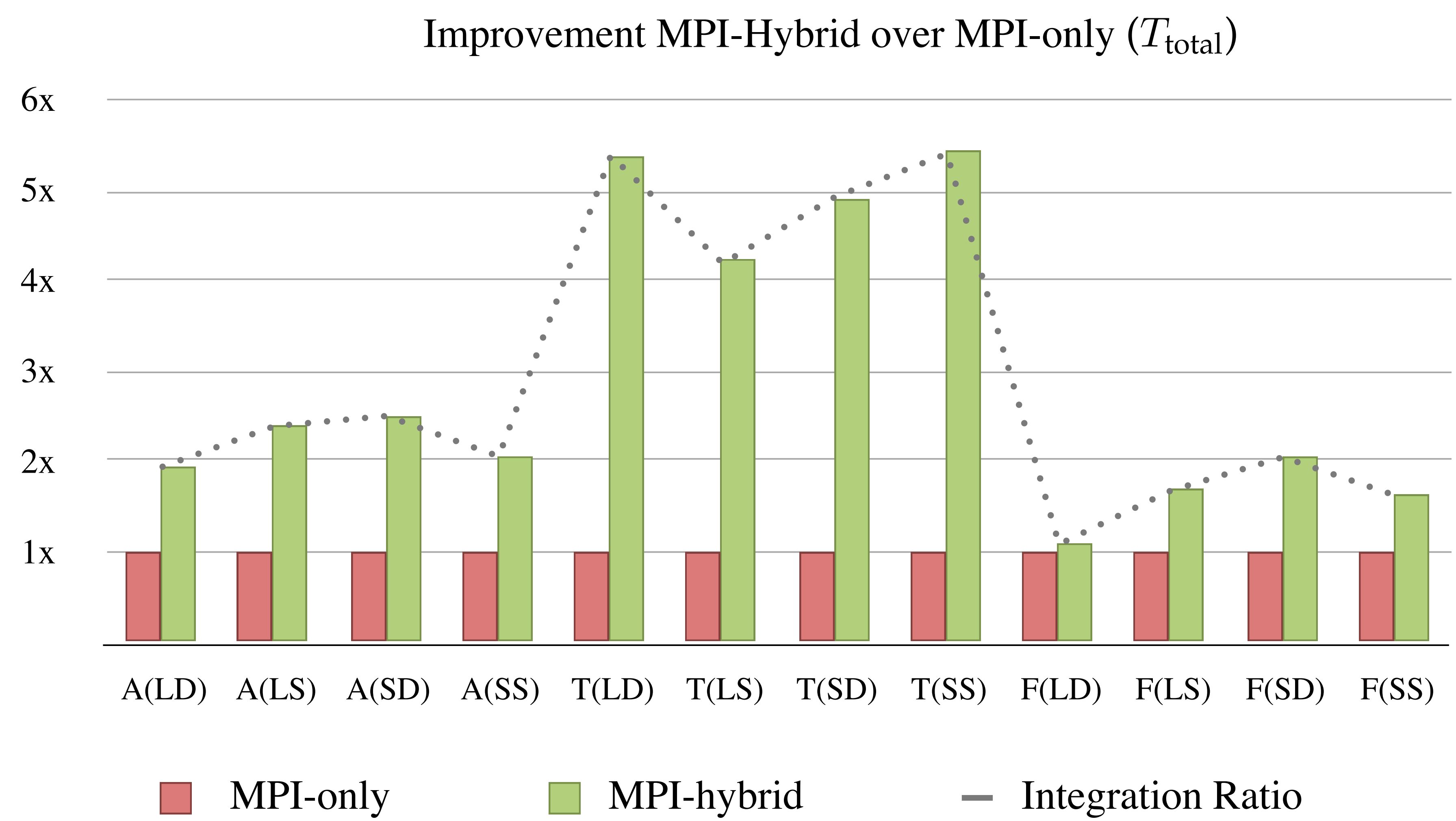
- Master/Slave approach yields better scalability & performance



D. Pugmire, H. Childs, C. Garth, S. Ahern, G. H. Weber. Scalable Computation of Streamlines on Very Large Datasets. In Proc. ACM Supercomputing (SC'09), 2009.

Scalable Distributed Integral Curve Algorithms

Improvements using hybrid parallelism (MPI + threads)



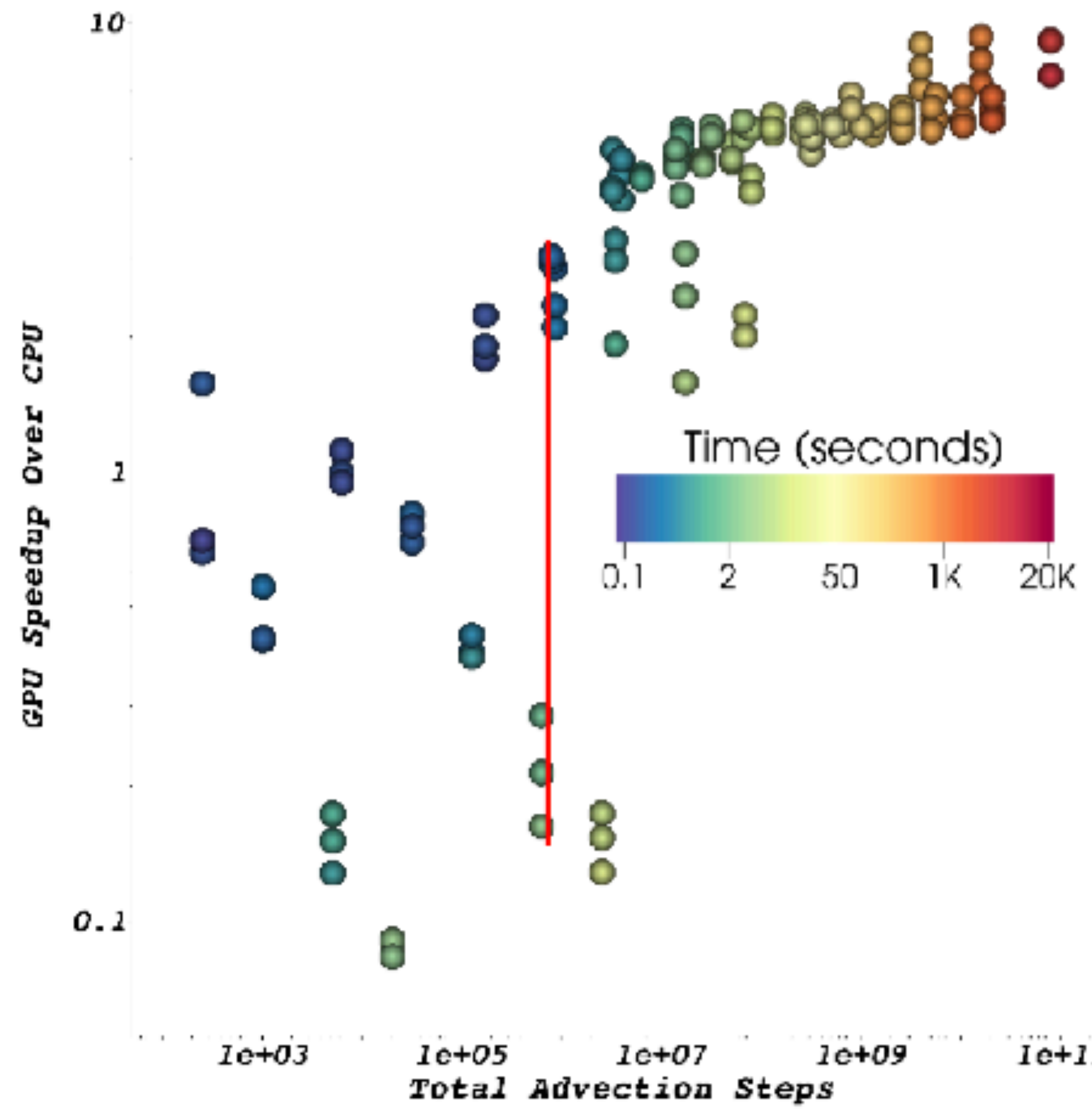
Benchmarks on Jaguar (Cray XT5), Oak Ridge National Laboratory

D. Camp, C. Garth, H. Childs, D. Pugmire, K. I. Joy. *Streamline Integration Using MPI-Hybrid P arallelism on a Large Multicore Architecture*. IEEE TVCG 17(11):1702-1713, 2011.

Scalable Integral Curve Algorithms

GPUs can also be leveraged to accelerate computation in a distributed setting.

Speedup depends strongly on problem characteristics.



D. Camp, H. Krishnan, D. Pugmire, C. Garth, I. Johnson, E. W. Bethel, K. I. Joy, H. Childs: *GPU acceleration of particle advection workloads in a parallel, distributed memory setting.* In Proc. EGPGV, 2013.

Scalable Distributed Integral Curve Algorithms

Scalable integral curve algorithms available in the **VisIt visualization software** (https://wci.llnl.gov/codes/visit/)

Alternative approaches and improvements:

- **work stealing** for load balancing with parallelize-over-seeds
- **optimized communication** patterns with parallelize-over-blocks
- **out-of-core techniques** using prefetching

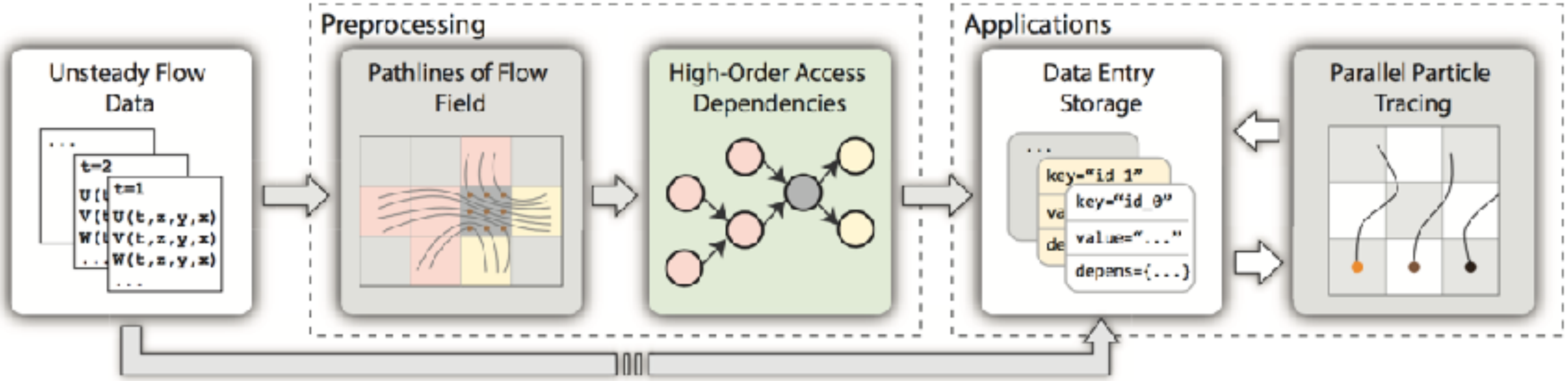
C. Müller, D. Camp, B. Hentschel, C. Garth: *Distributed parallel particle advection using work requesting.* In Proc. Large-Scale Data Analysis & Visualization 2013.

T. Peterka, W. Kendall, D. Goodell, B. Nouanesengsey, H.-W. Shen, J. Huang, K. Moreland, R. Thakur, R. Ross: *Performance of Communication Patterns for Extreme-Scale Analysis and Visualization.* Journal of Physics: Conference Series SciDAC 2010, 2010

Out-of-Core Techniques

Precomputing information about the behavior of integral curves allows better data access strategies and scheduling.

Example: higher-order access dependencies records multiple iterations of traversed blocks



[image from J. Zhang, H. Guo, X. Yuan: *Efficient Unsteady Flow Visualization with High-Order Access Dependencies.* In Proc. IEEE Pacific Visualization Symposium (PacificVis), 2016]

Implementations / Further Reading

Easy to use implementations of many methods advanced / experimental are readily available in the **OSUFLOW** library:

GitHub:
https://github.com/GRAVITYLab/OSUFlow

VTK integration:
http://www.sdav-scidac.org/highlights/29-highlights/visualization/67-osuflow-vtk-integration.html

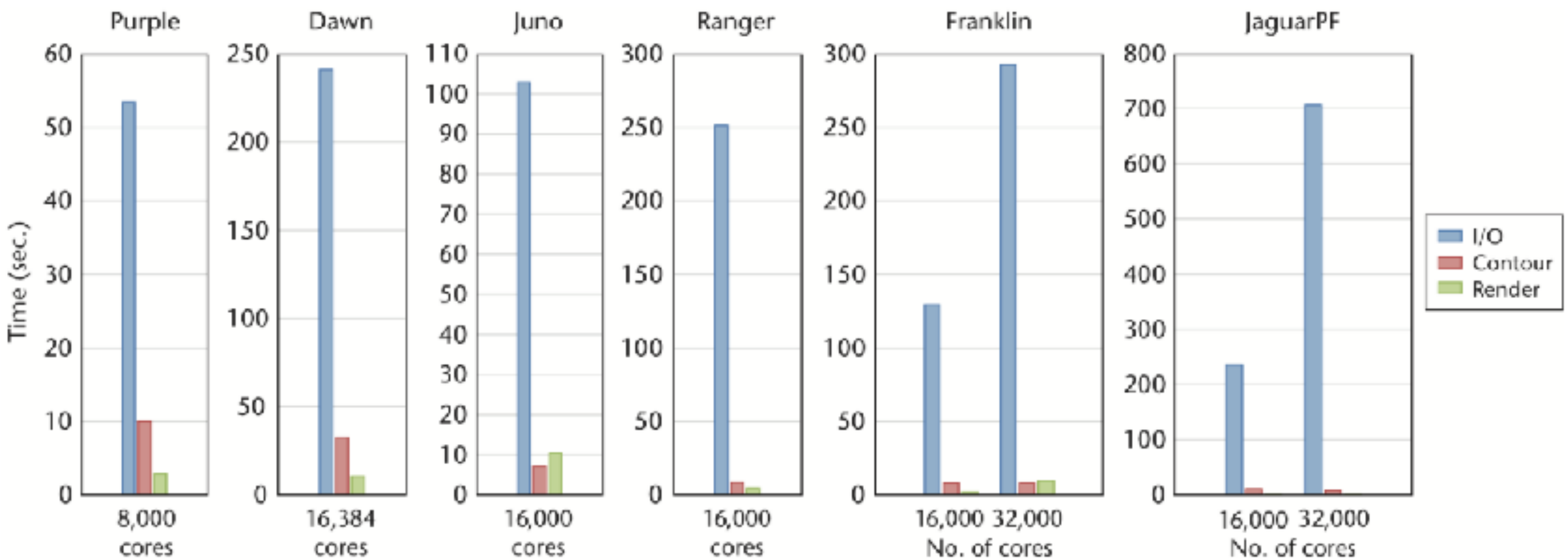
Good survey paper:

Peterka, T., Ross, R., Nouanesengsey, B., Lee, T.-Y., Shen, H.-W., Kendall, W., Huang, J.
A Study of Parallel Particle Tracing for Steady-State and Time-Varying Flow Fields.
Proceedings IPDPS'11, Anchorage AK, May 2011

Part II. In Situ Techniques for Vector Fields

In Situ Visualization

Re-thinking visualization as **post-processing**: for many practical problems, storing simulation output has become too expensive.



[from Childs et al.: *Extreme scaling of production visualization software on diverse architectures.* IEEE CG&A 30(3):22-31, 2010]

I/O bandwidth does not scale with CPU & memory performance.

In Situ Visualization

In situ visualization:

- compute visualization <u>during</u> simulation, while data is <u>in memory</u> and readily accessible
- store only visualization results (images), or
  intermediate results (features or even integral curves)

Drawbacks:

- interactive exploration is not possible
- what should be precomputed?

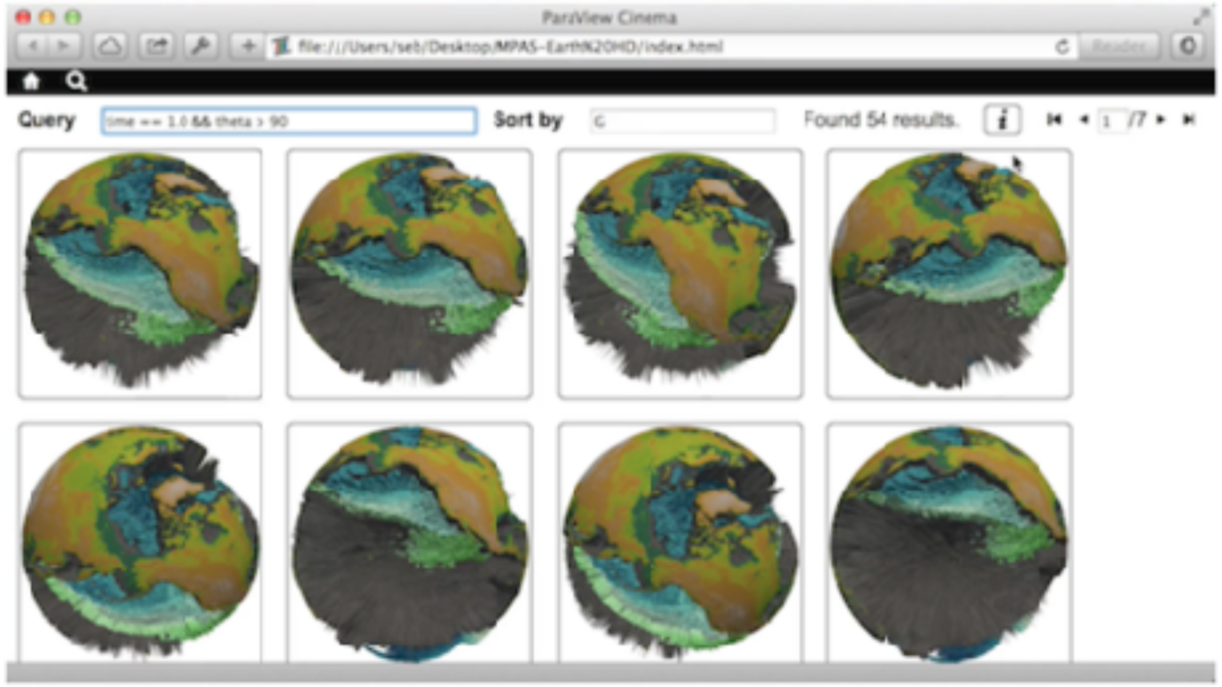
Some flexibility can be preserved by clever choice of images,
or selection of features / integral curves can be automated
(e.g. streamline selection → Jun's talk).

In Situ Visualization

Example: **Paraview Cinema** (Ahrens et al. 2014)

- pre-specify all visualization parameters and store only images

- preserve flexibility by storing many different combinations of visualizations and viewpoints



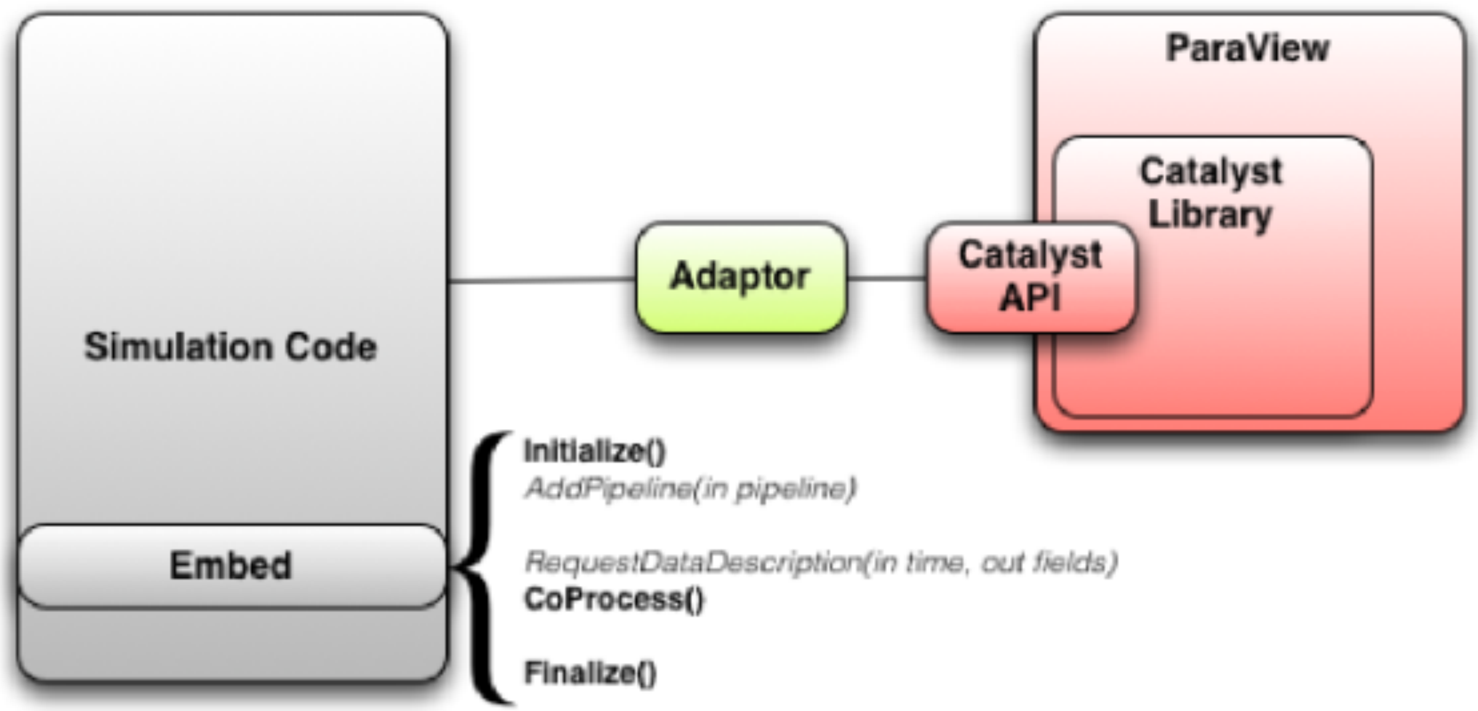
from Ahrens et al.: In Situ MPAS-Ocean Image-based Visualization. In Proc ACM Supercomputing (SC'14)

see also: https://blog.kitware.com/paraview-cinema-an-image-based-approach-to-extreme-scale-data-analysis/

In Situ Visualization

Ready-to-roll solution:
Paraview Cinema + Paraview Catalyst

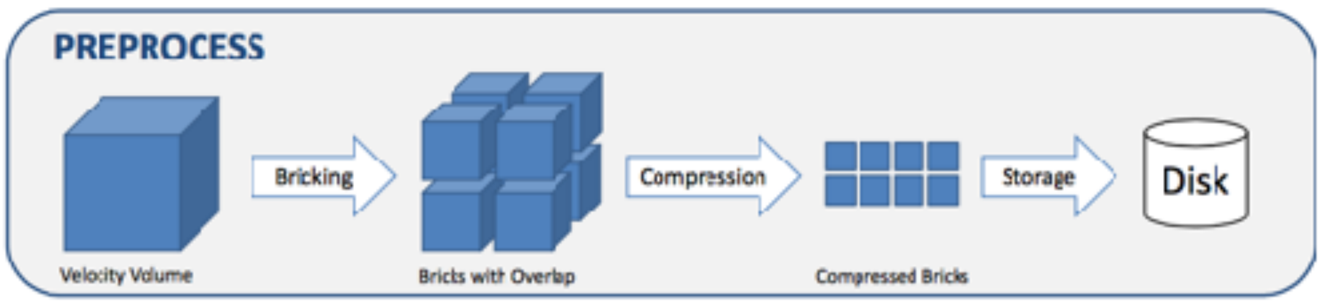
Computes visualization during simulation.



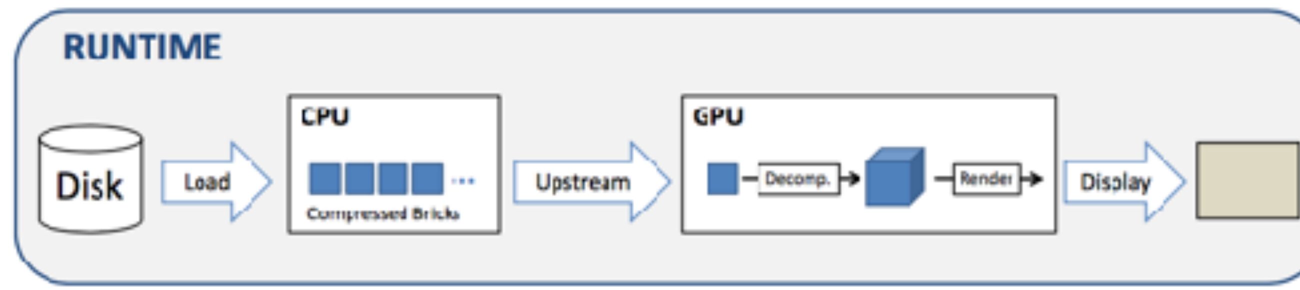
Part of newer ParaView releases → http://paraview.org

In Situ Compression

Compression of vector field data with features in mind:
e.g. turbulent vortex structures [Treib et al. 2012]



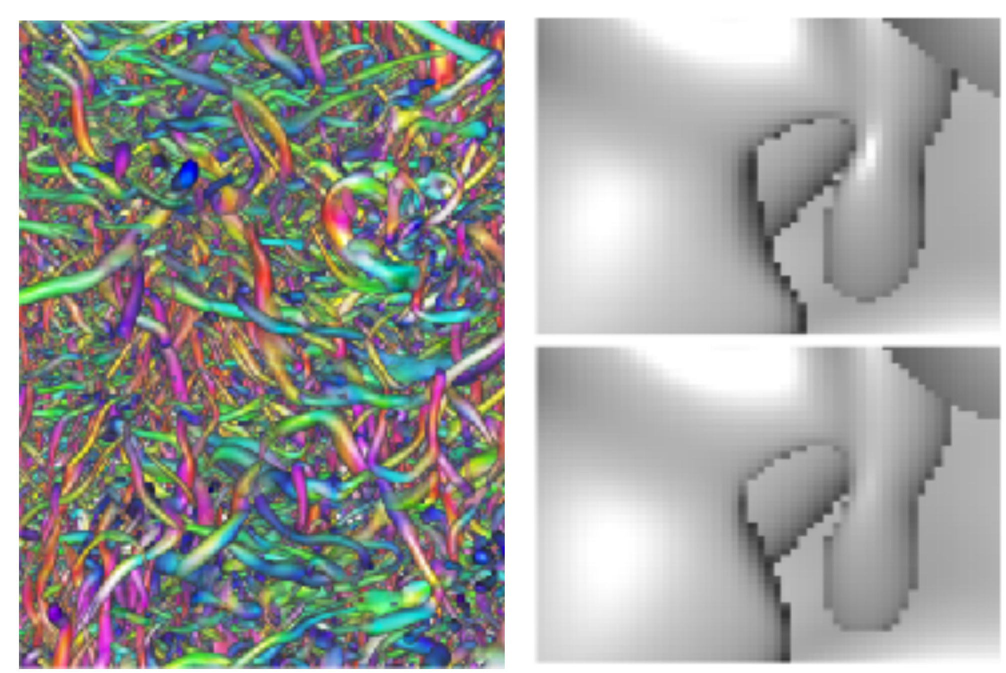
in situ processing



interactive post-processing

[images from Treib et al.: Turbulence Visualization at the Terascale on Desktop PCs. IEEE TVCG 18(12):2169–77, 2012]

Wavelet-based compression with ~3 bits per voxel.

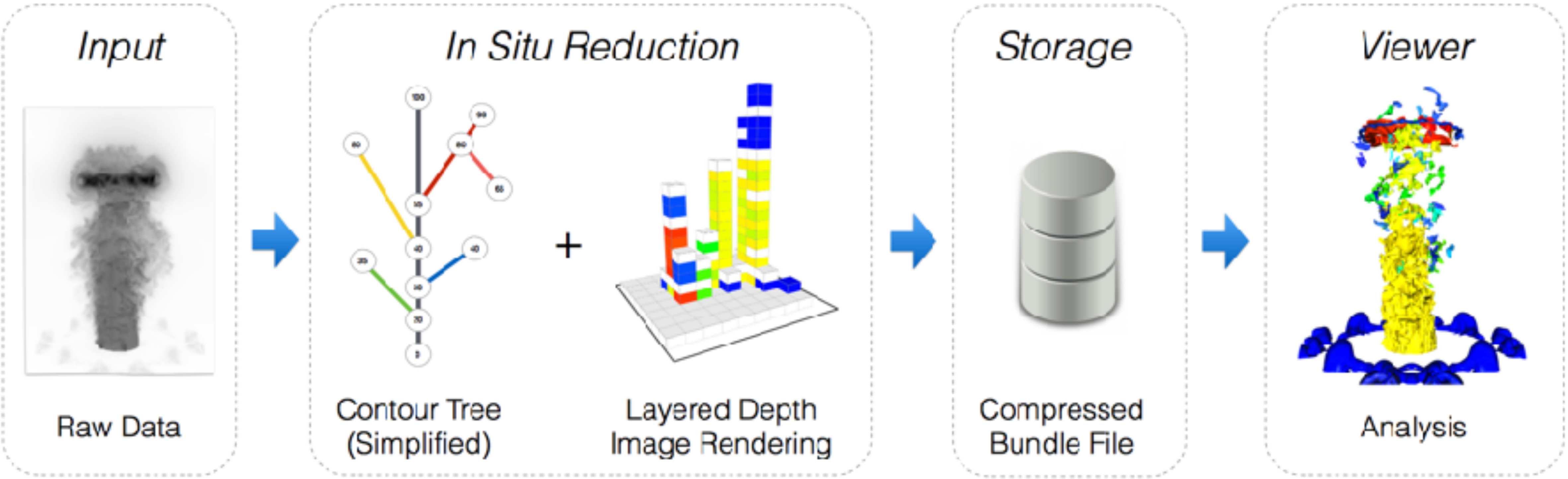


original

3 bpv

In Situ Feature Extraction

Feature-based visualization: combine **image-based approach** with **topological feature extraction.**

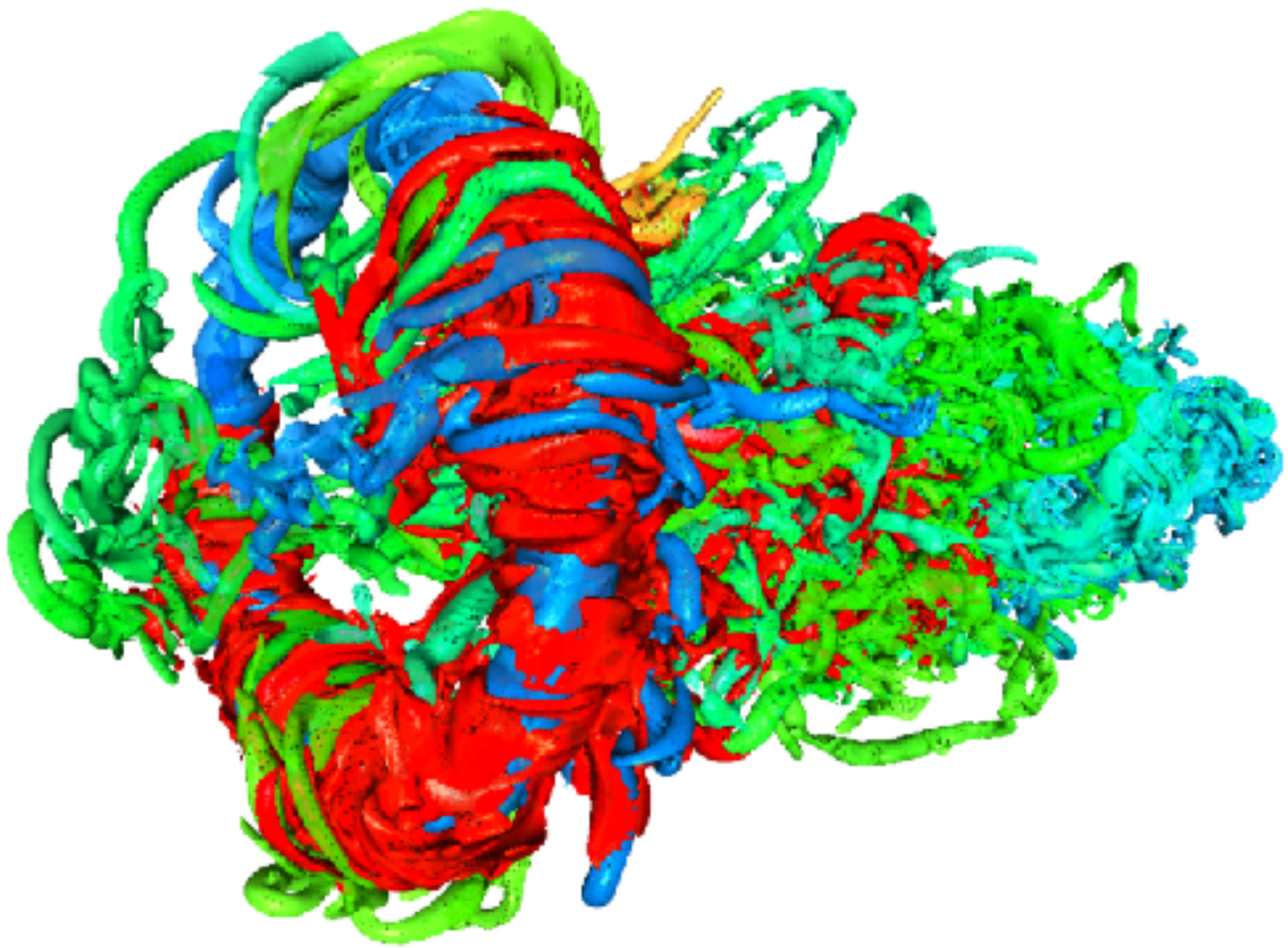


Raw data is converted into a topological representation, simplified, and rendered into a **layered depth image**.

T. Biedert, C. Garth: Contour Tree Depth Images For Large Data Visualization. In Proc. EGPGV, 2015.

In Situ Feature Extraction

The depth image stores only a fixed number of the most important topological features, after simplification.

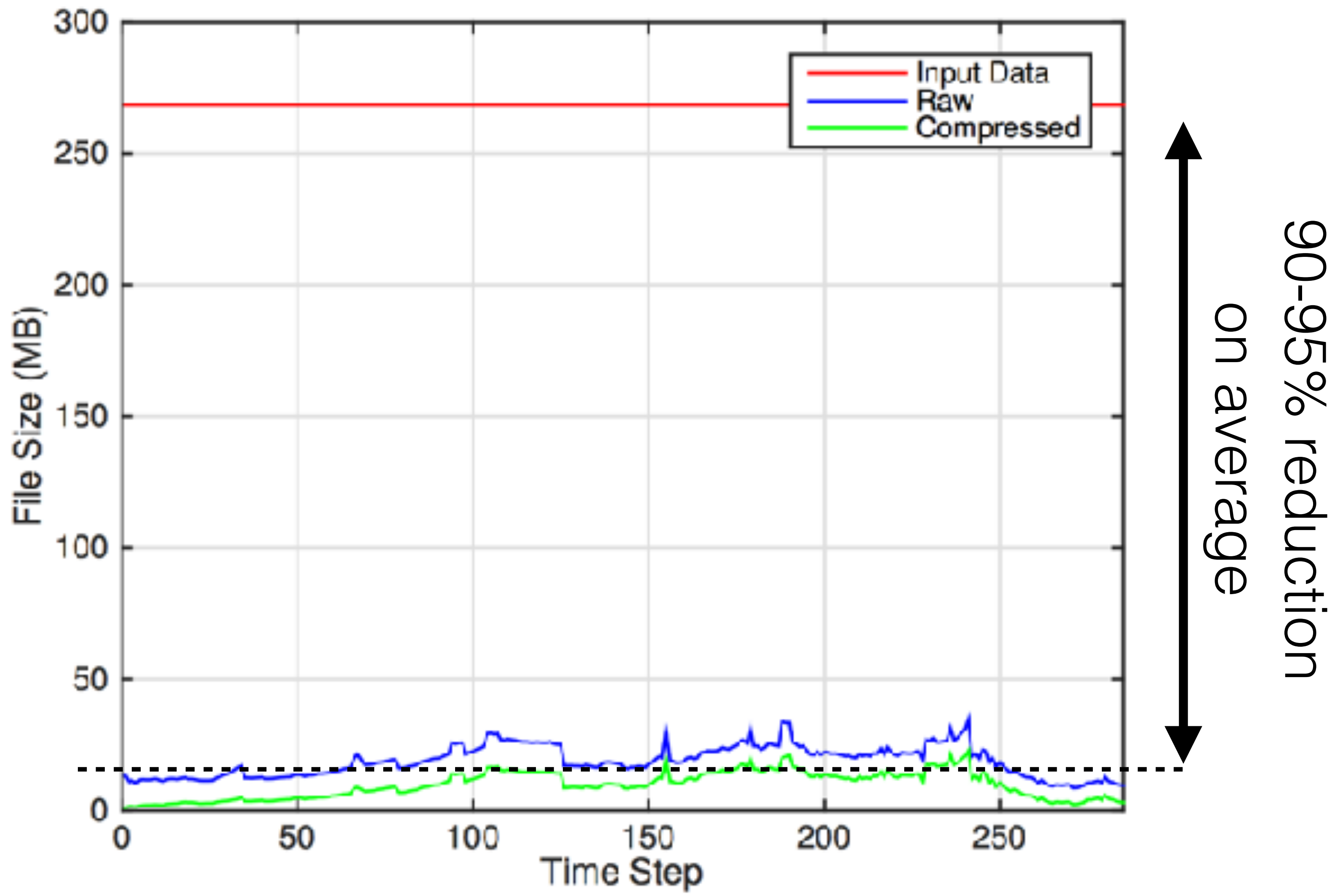


Vortices in a jet flow characterized via λ_2–criterion
(**100** branches stored, **1.5M** branches before simplification)

T. Biedert, C. Garth: Contour Tree Depth Images For Large Data Visualization. In Proc. EGPGV, 2015.

In Situ Feature Extraction

Large storage savings are achievable.

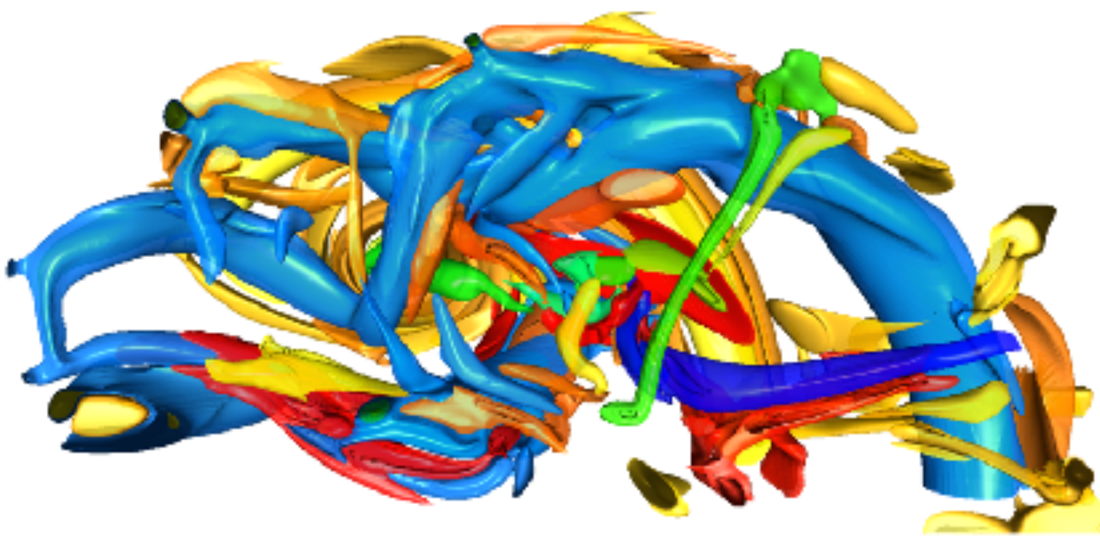


Data can be copied to and visualized on a standard PC.

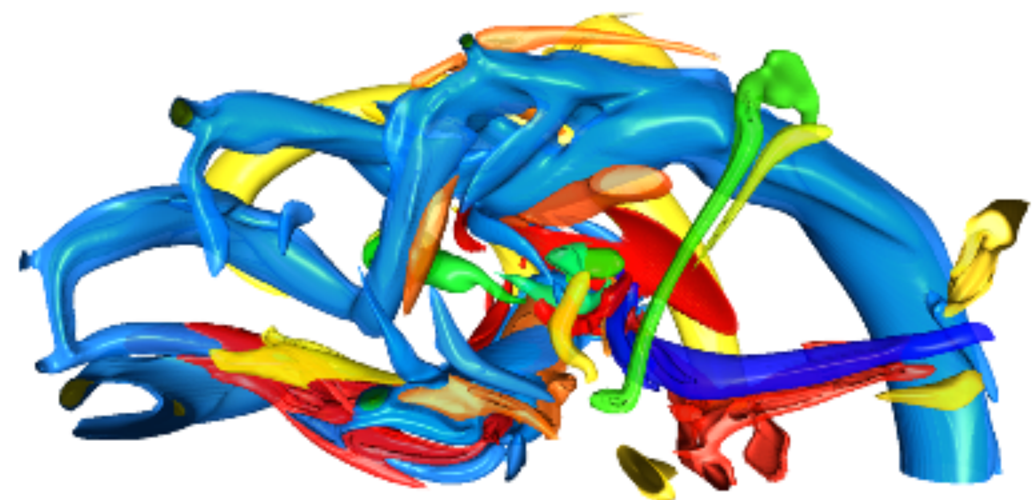
T. Biedert, C. Garth: Contour Tree Depth Images For Large Data Visualization. In Proc. EGPGV, 2015.

In Situ Feature Extraction

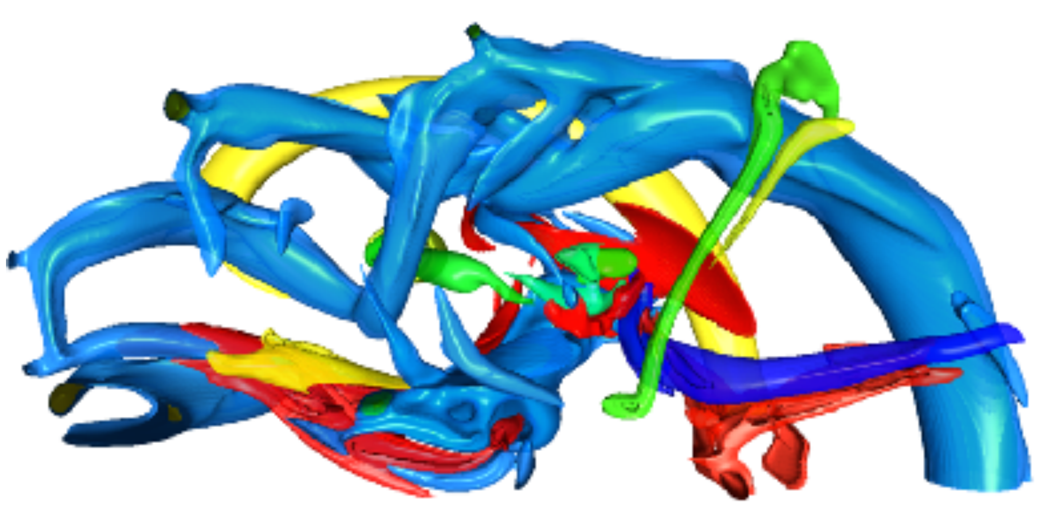
Flexibility for further filtering and simplification is preserved (multiresolution visualization).



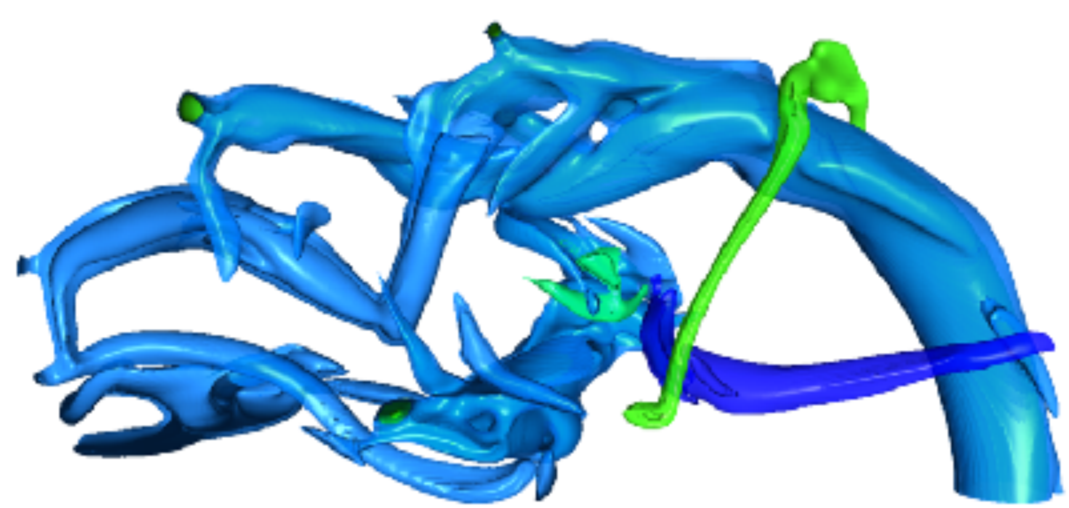
100 features



50 features



20 features



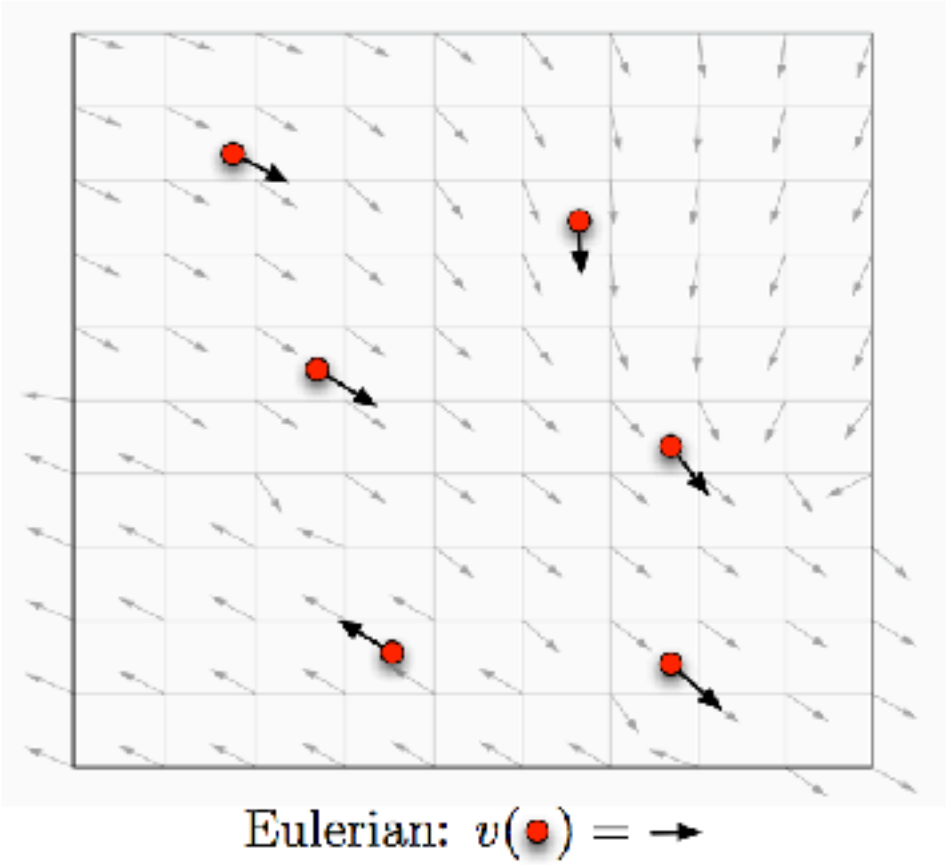
10 features

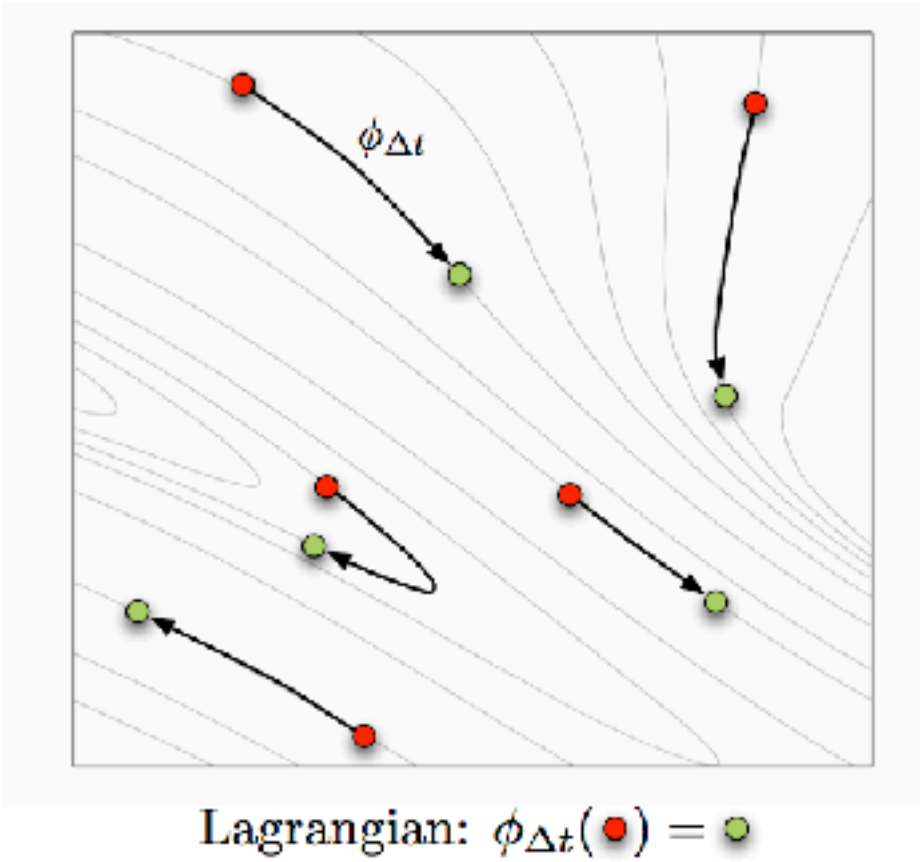
T. Biedert, C. Garth: Contour Tree Depth Images For Large Data Visualization. In Proc. EGPGV, 2015.

Lagrangian Representation

Can exploration be preserved? Is it feasible to precompute and store integral curves in situ?

Lagrangian representation of vector fields (“flow map”).



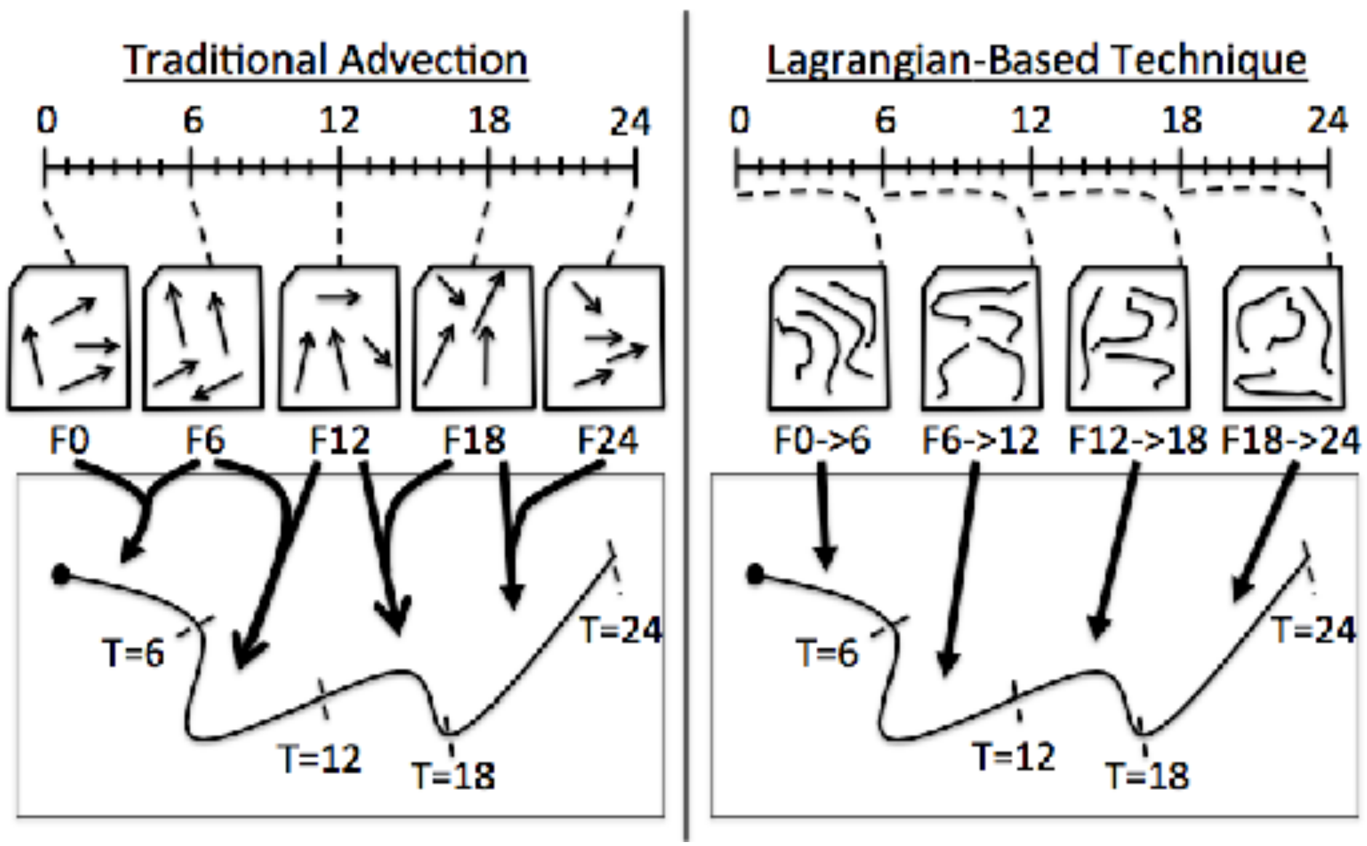


A. Agranovsky, C. Garth, K. I. Joy: *Extracting Flow Structures using Sparse Particles*. In Proc. Vision, Modeling, Visualization, 2011.

Lagrangian Representation

Example: representation of flow maps (computed in situ) using a representative set of sparse particles.

Multi-resolution reconstruction of LCS (FTLE) using Moving-Least Squares (MLS)

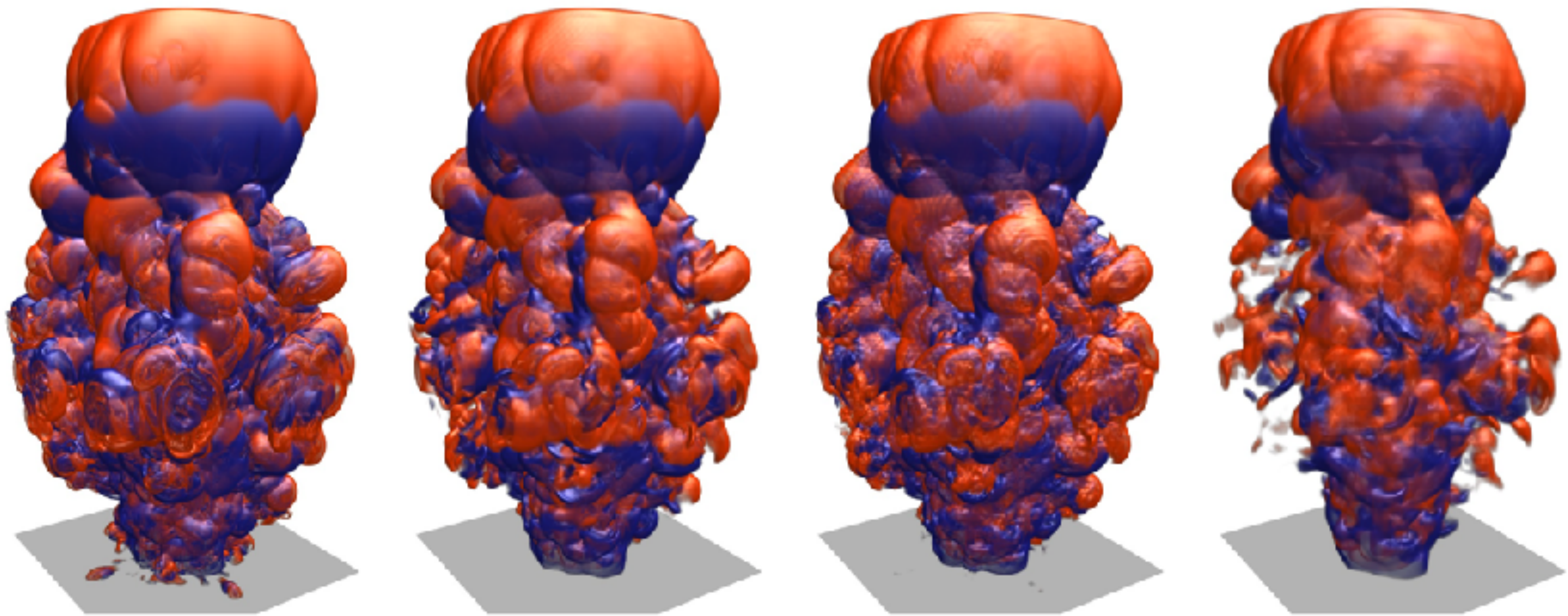


A. Agranovsky, C. Garth, K. I. Joy: *Extracting Flow Structures using Sparse Particles*. In Proc. Vision, Modeling, Visualization, 2011.

Lagrangian Representation

Example: representation of flow maps (computed in situ) using a representative set of sparse particles.

Multi-resolution reconstruction of LCS (FTLE) using Moving-Least Squares (MLS)

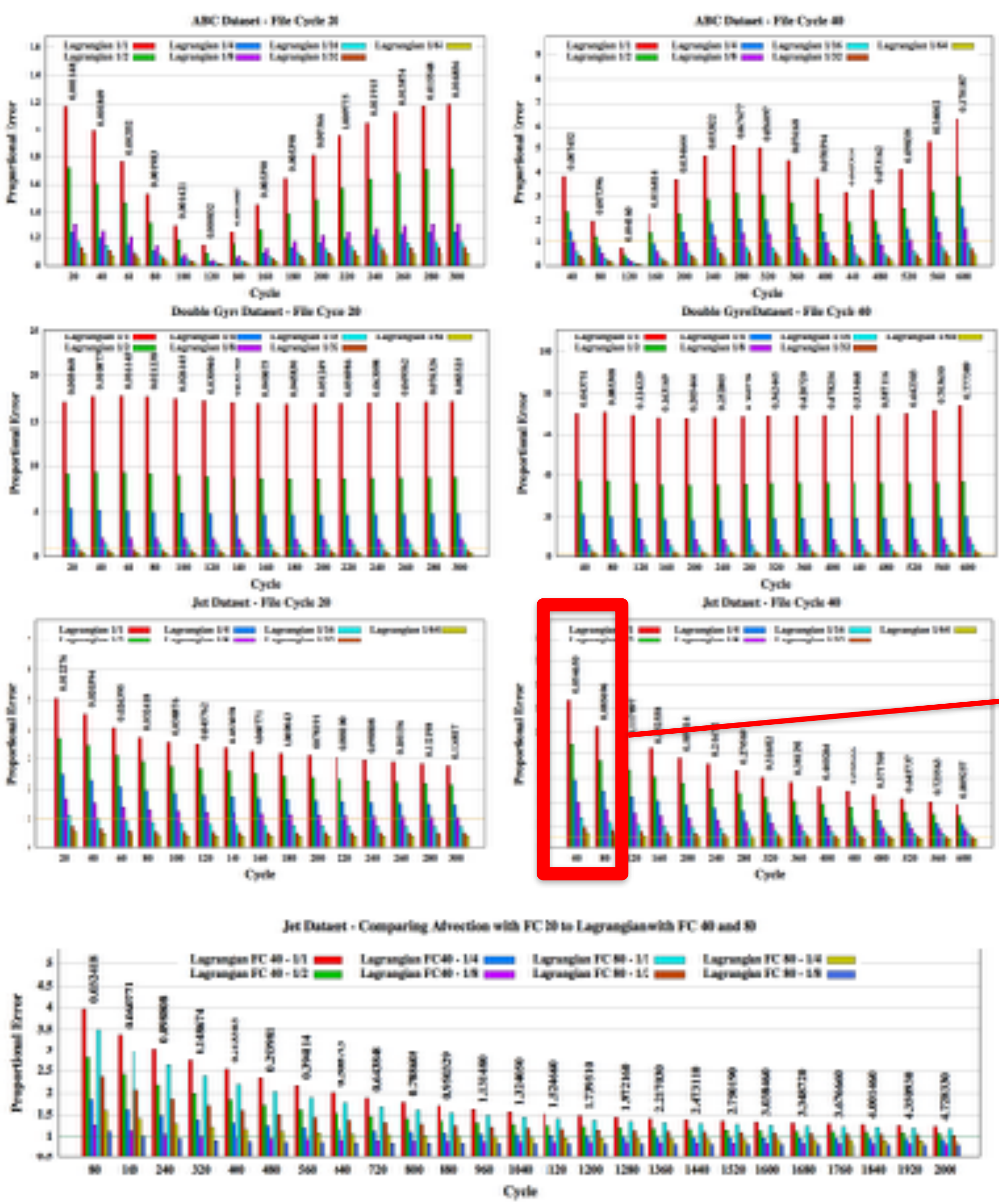


FD, 100% samples — MLS, 1.5% samples — FD, 1.5% samples — MLS, 0.2% samples

A. Agranovsky, C. Garth, K. I. Joy: *Extracting Flow Structures using Sparse Particles*. In Proc. Vision, Modeling, Visualization, 2011.

Lagrangian Representation

A great property of Lagrangian representation:



Strong data reduction, but also, strongly improved error/cost control

Storage Reduction	0%	50%	75%	88%	93%	96%	98%
Accuracy Improvement	12X	9X	5.5X	4X	2.5X	1.8X	1.2X

A. Agranovsky, D. Camp, C. Garth, E. W. Bethel, K. I. Joy, H. Childs: *Improved Post Hoc Flow Analysis Via Lagrangian Representations.* In Proc. Large-Data Analysis and Visualization, 2014

Open question: best storage for Lagrangian representation?

Part III. Features in Large Ensembles

Ensemble Simulation

Computational capacity can not only be used to increase resolution, but also to study parameter effects

Result: an "ensemble" of multiple simulations ("members")

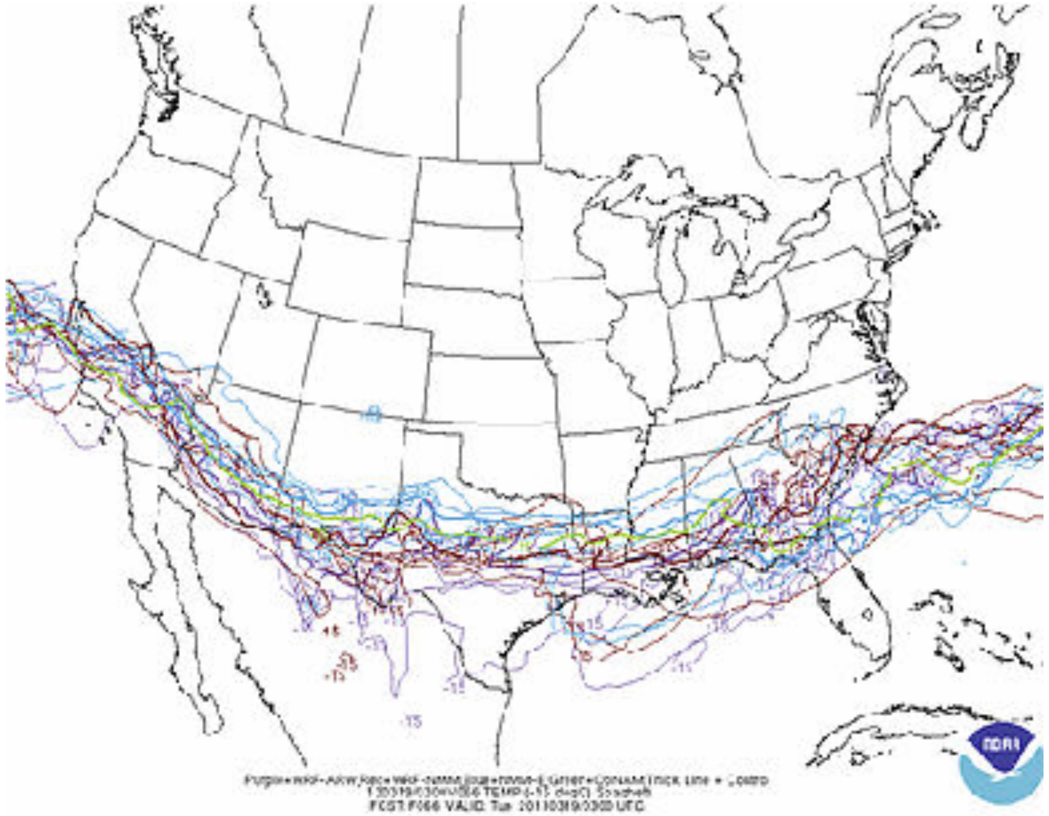


Image: Wikipedia

Question: how well do the members agree?

Features in Flow Visualization

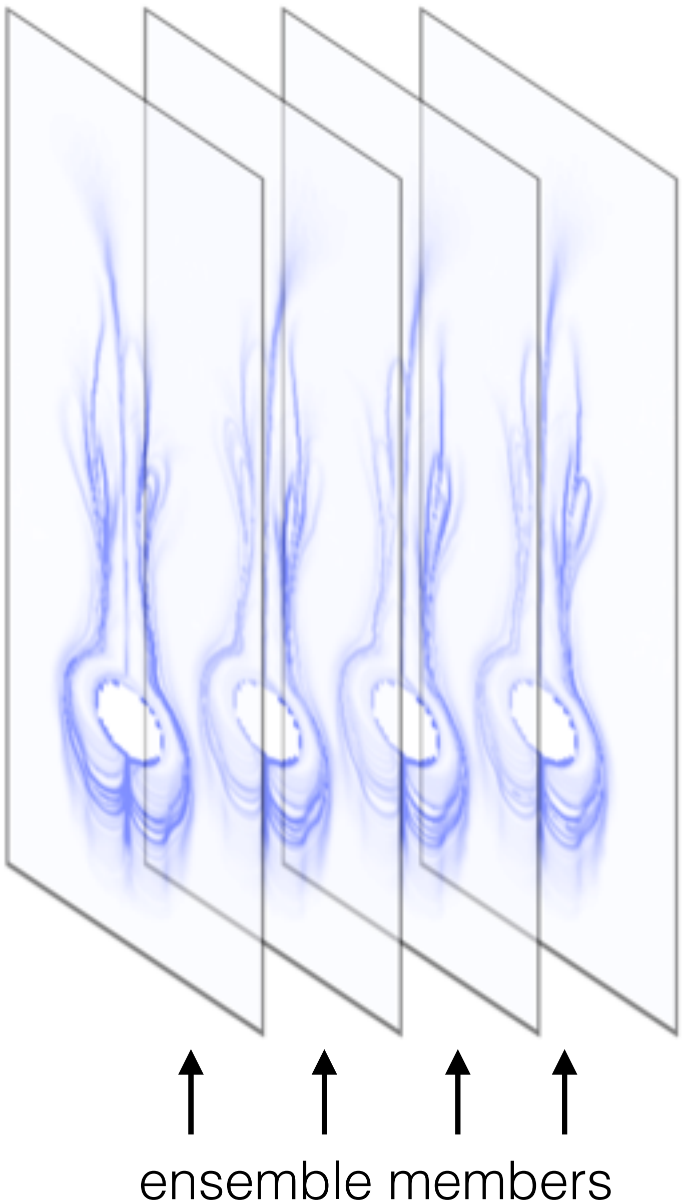
How can integral curve-derived features be identified and understood in **ensemble vector fields**?

Visualization problem:
how well do ensemble members agree with respect to features?

Consider the **flow map**

$$\phi^{\Delta t}(t,x) \ := \ x + \int_t^{t+\Delta t} v(\tau,x)d\tau$$

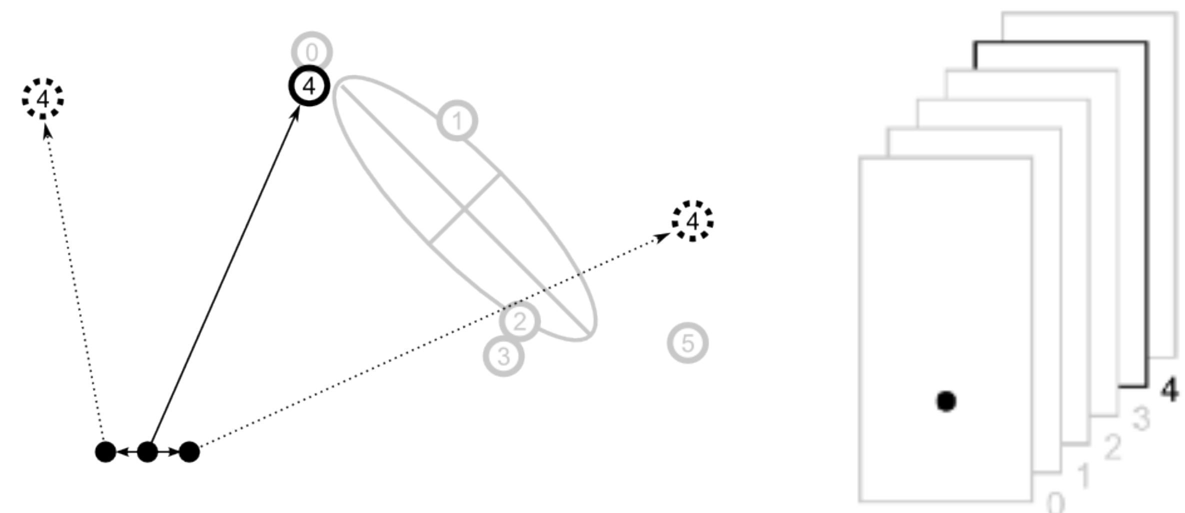
in each ensemble member.



Features from Flow Map Ensembles

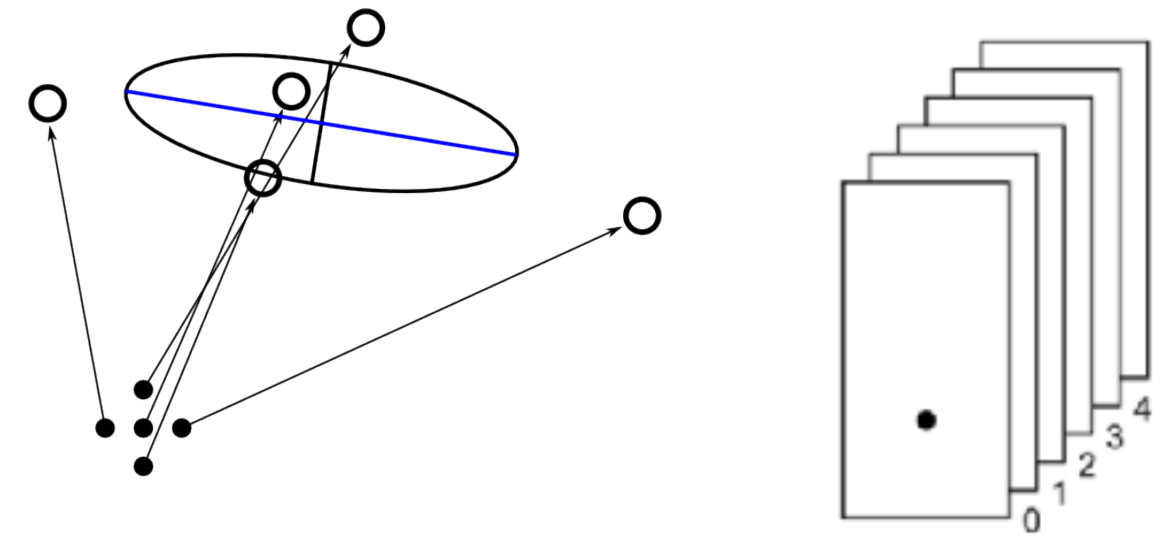
Compare for the flow map

individual variance
(variation across a small neighborhood around a point)



against

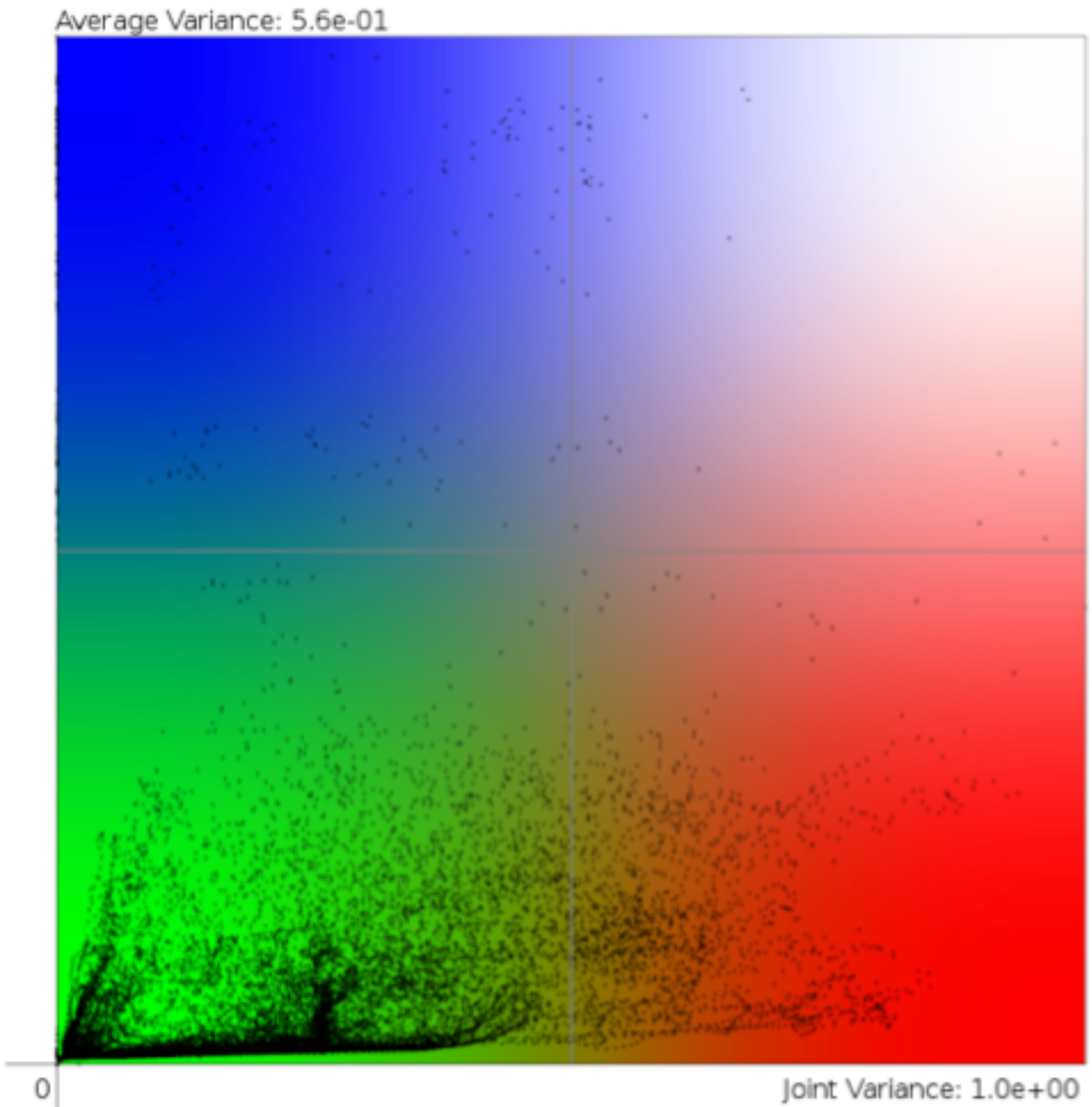
joint variance
(variation across the ensemble at a particular point in the domain)



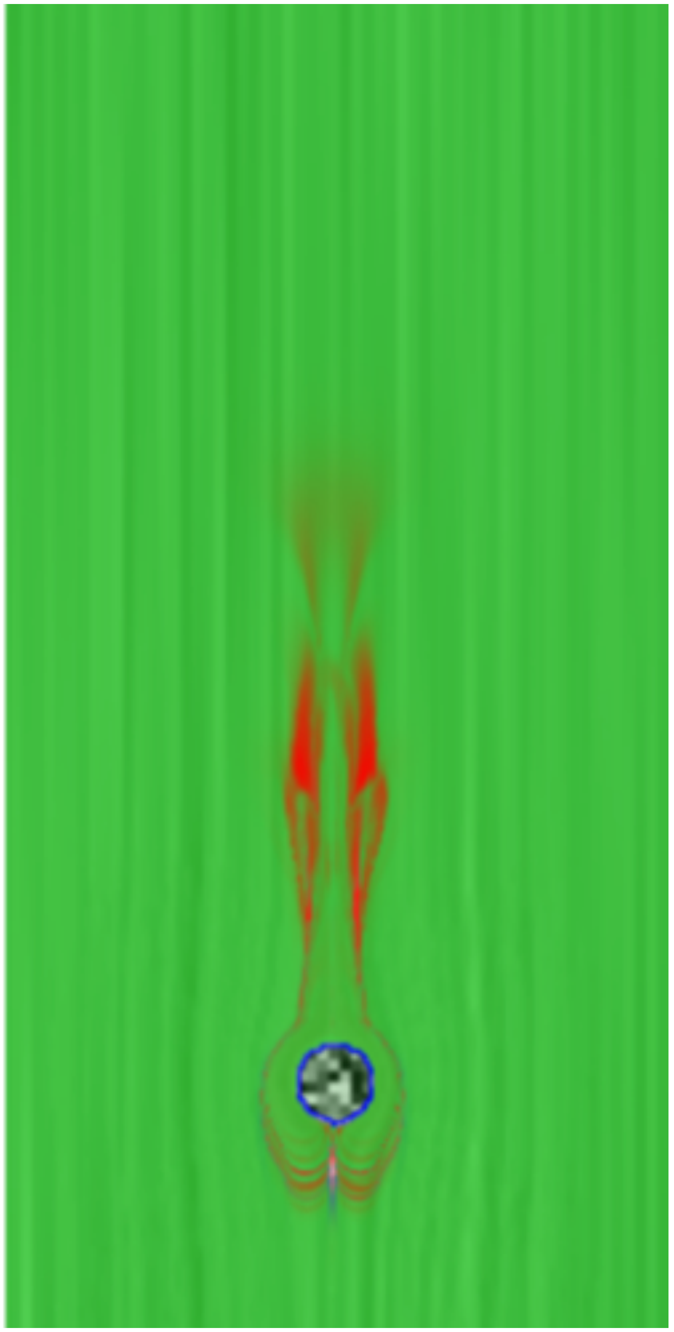
to understand agreement / disagreement with respect to flow map features.

→ **Classification Space (2D)**

Features from Flow Map Ensembles



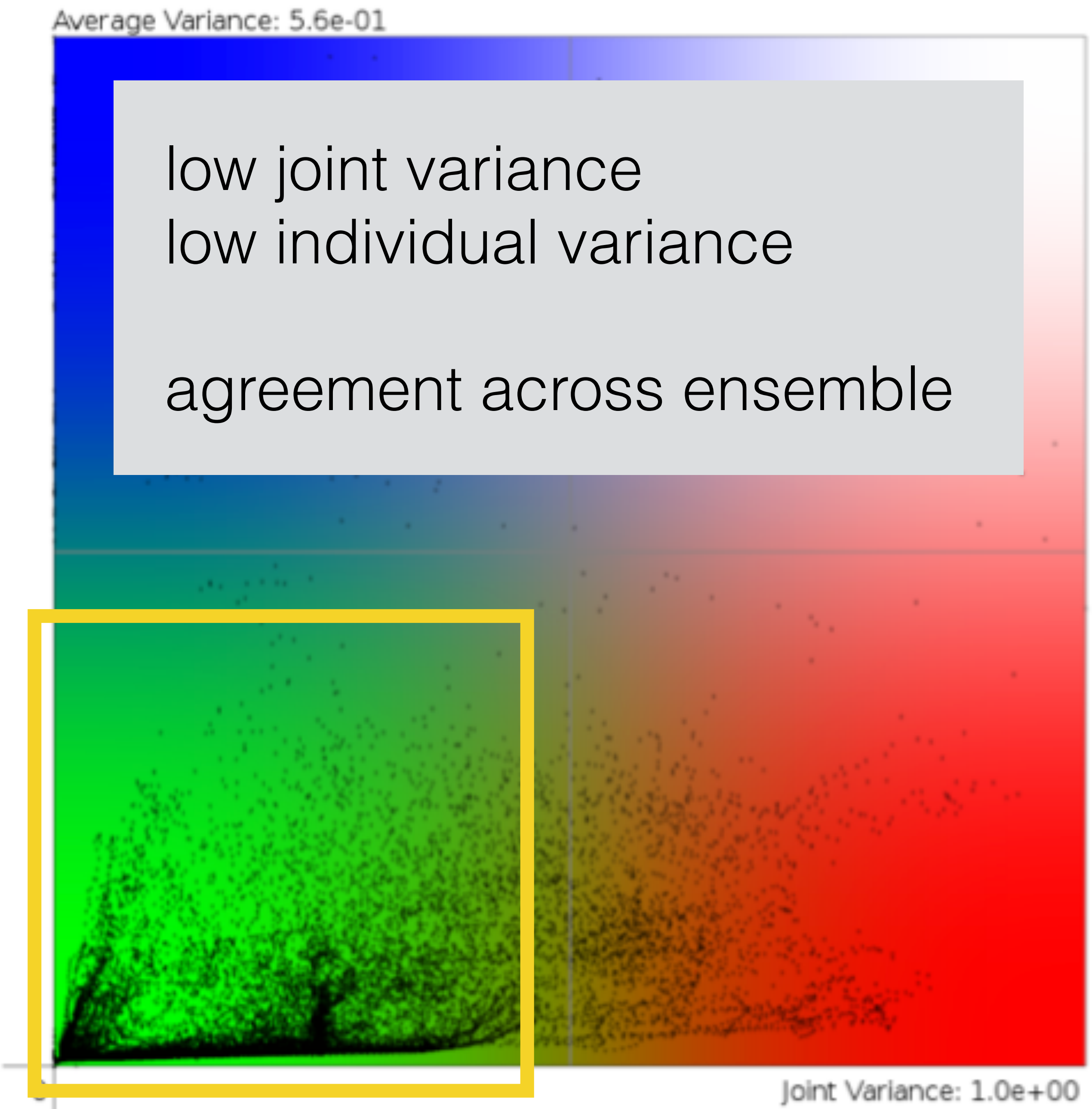
classification space



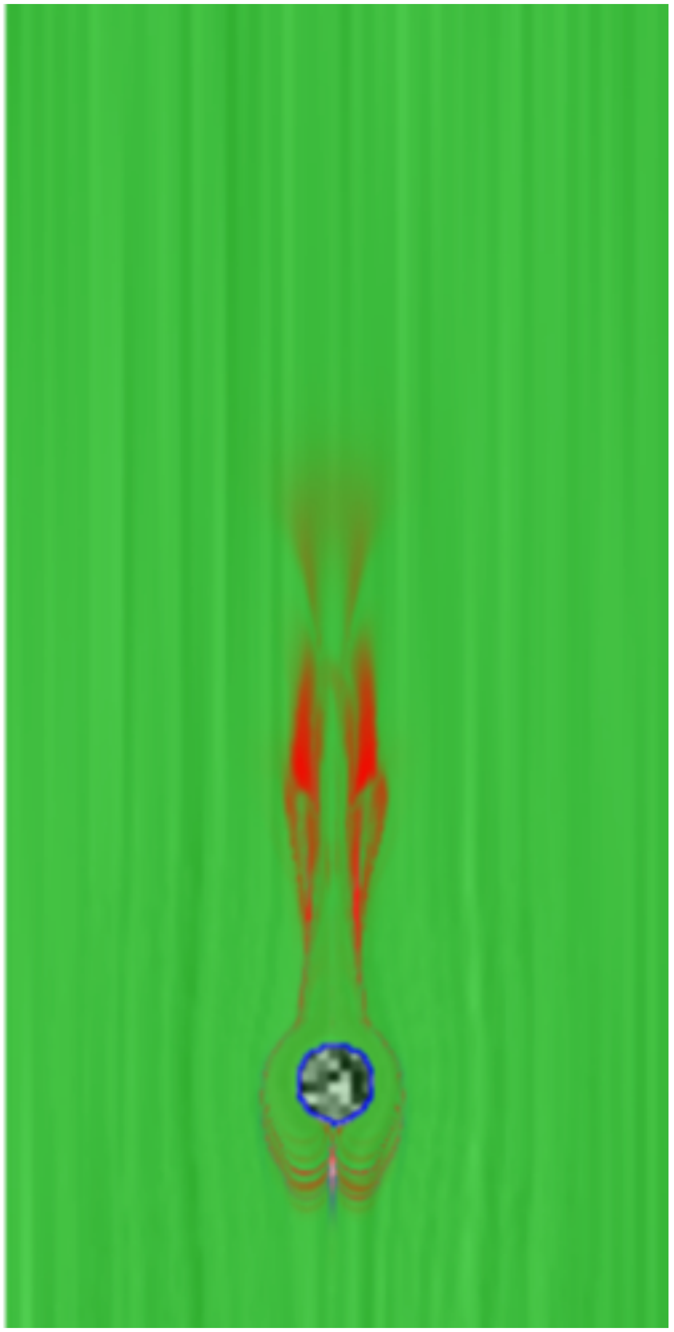
flow map domain

M. Hummel, C. Garth, H. Obermaier, K. I. Joy. *Comparative Visual Analysis of Lagrangian Transport in CFD Ensembles.* IEEE TVCG 19(12):2743–52, 2013 (best paper at IEEE SciVis 2013)

Features from Flow Map Ensembles



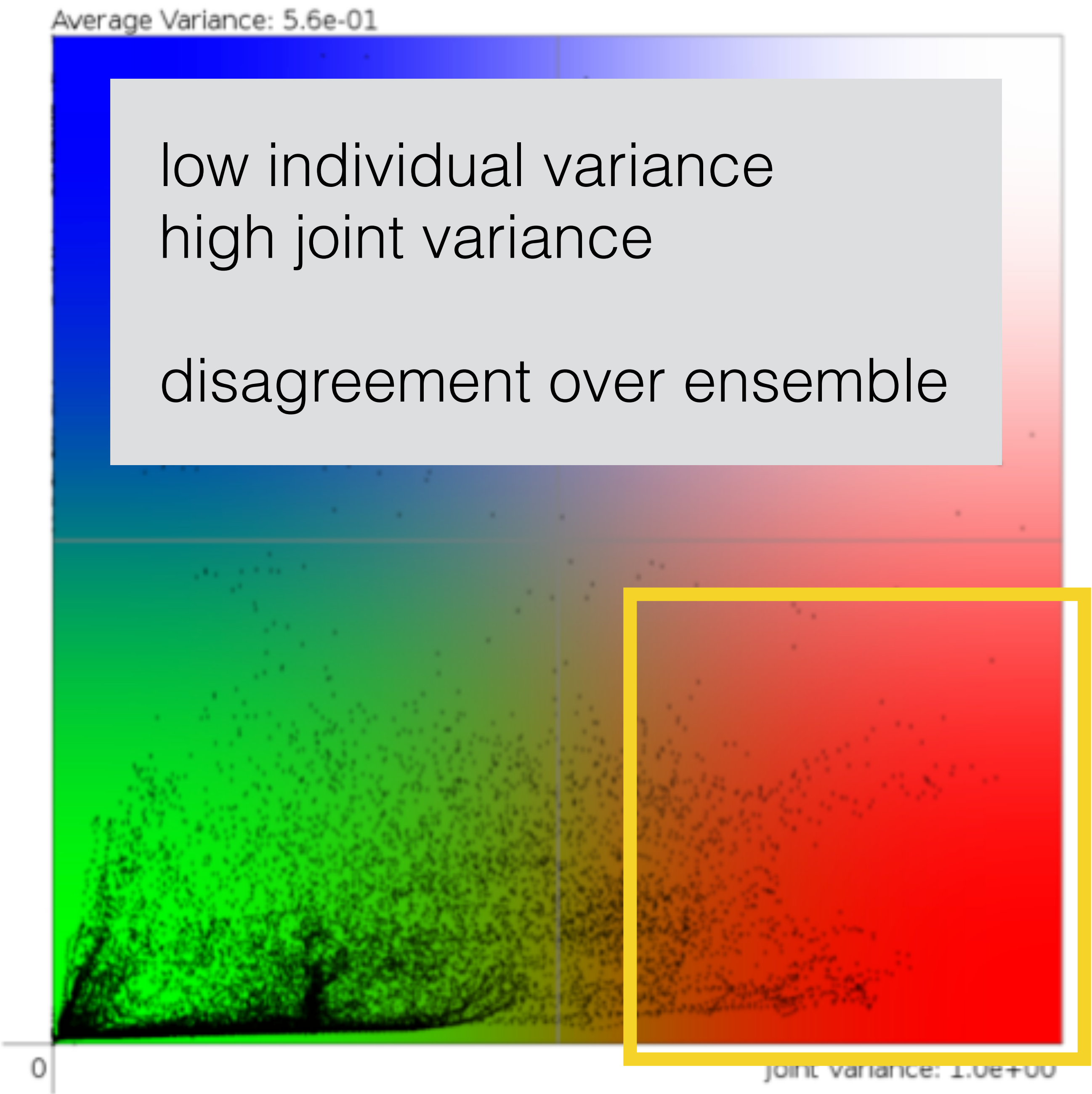
classification space



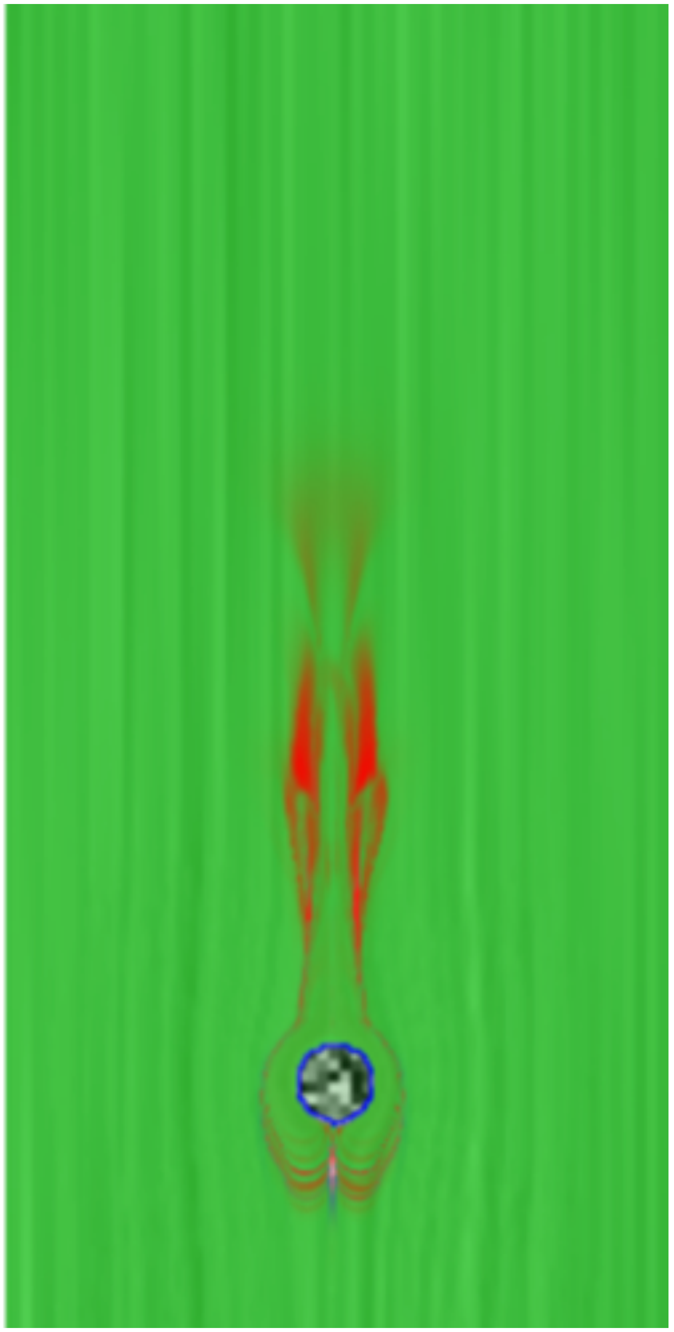
flow map domain

M. Hummel, C. Garth, H. Obermaier, K. I. Joy. *Comparative Visual Analysis of Lagrangian Transport in CFD Ensembles.* IEEE TVCG 19(12):2743–52, 2013 (best paper at IEEE SciVis 2013)

Features from Flow Map Ensembles



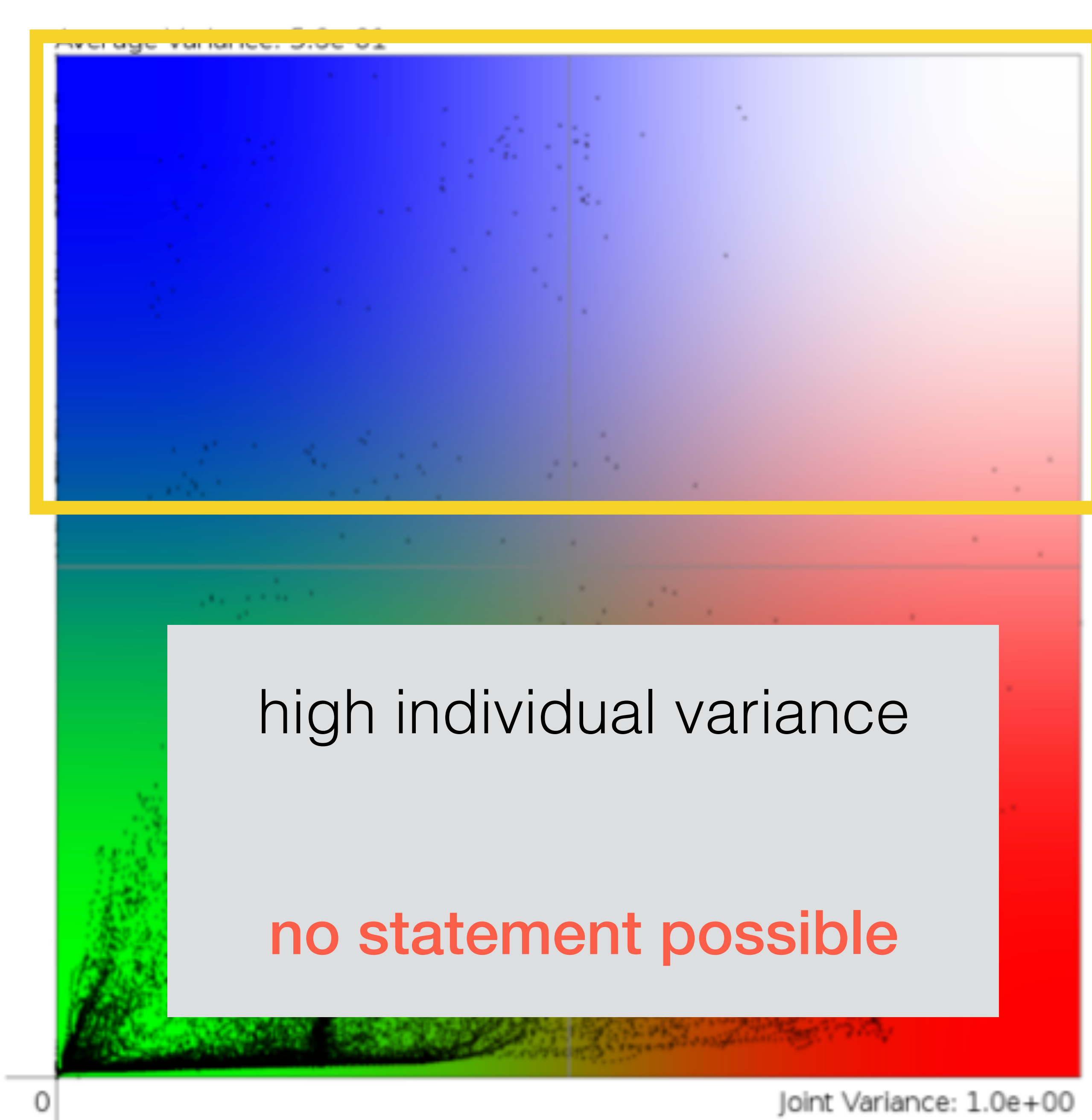
classification space



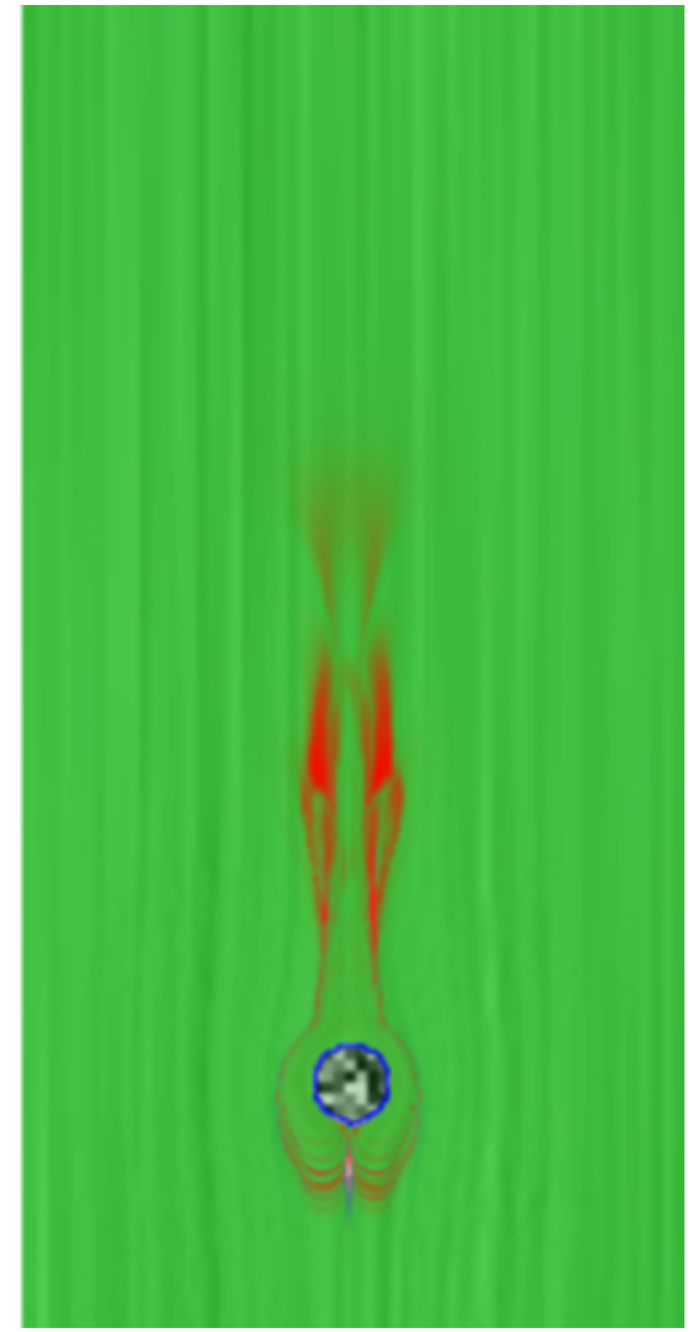
flow map domain

M. Hummel, C. Garth, H. Obermaier, K. I. Joy. *Comparative Visual Analysis of Lagrangian Transport in CFD Ensembles.* IEEE TVCG 19(12):2743–52, 2013 (best paper at IEEE SciVis 2013)

Features from Flow Map Ensembles



classification space

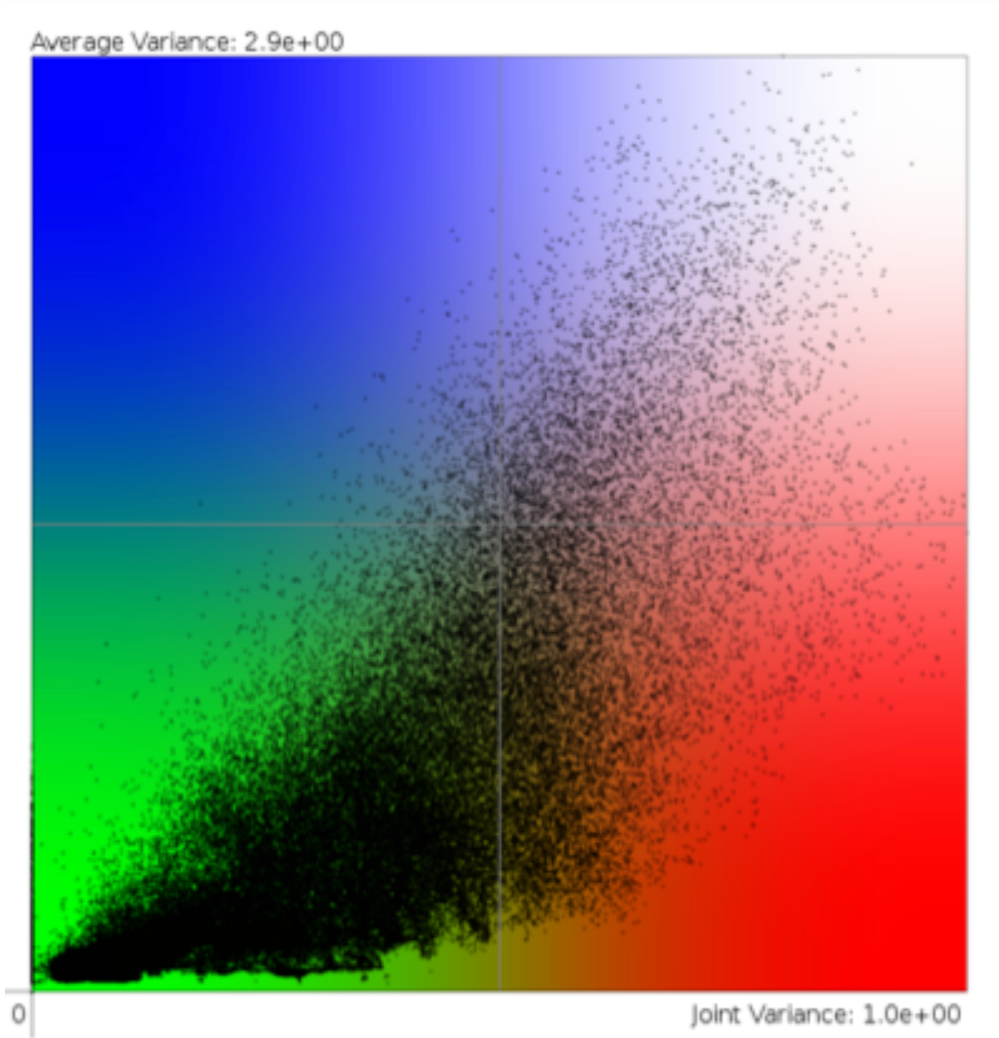


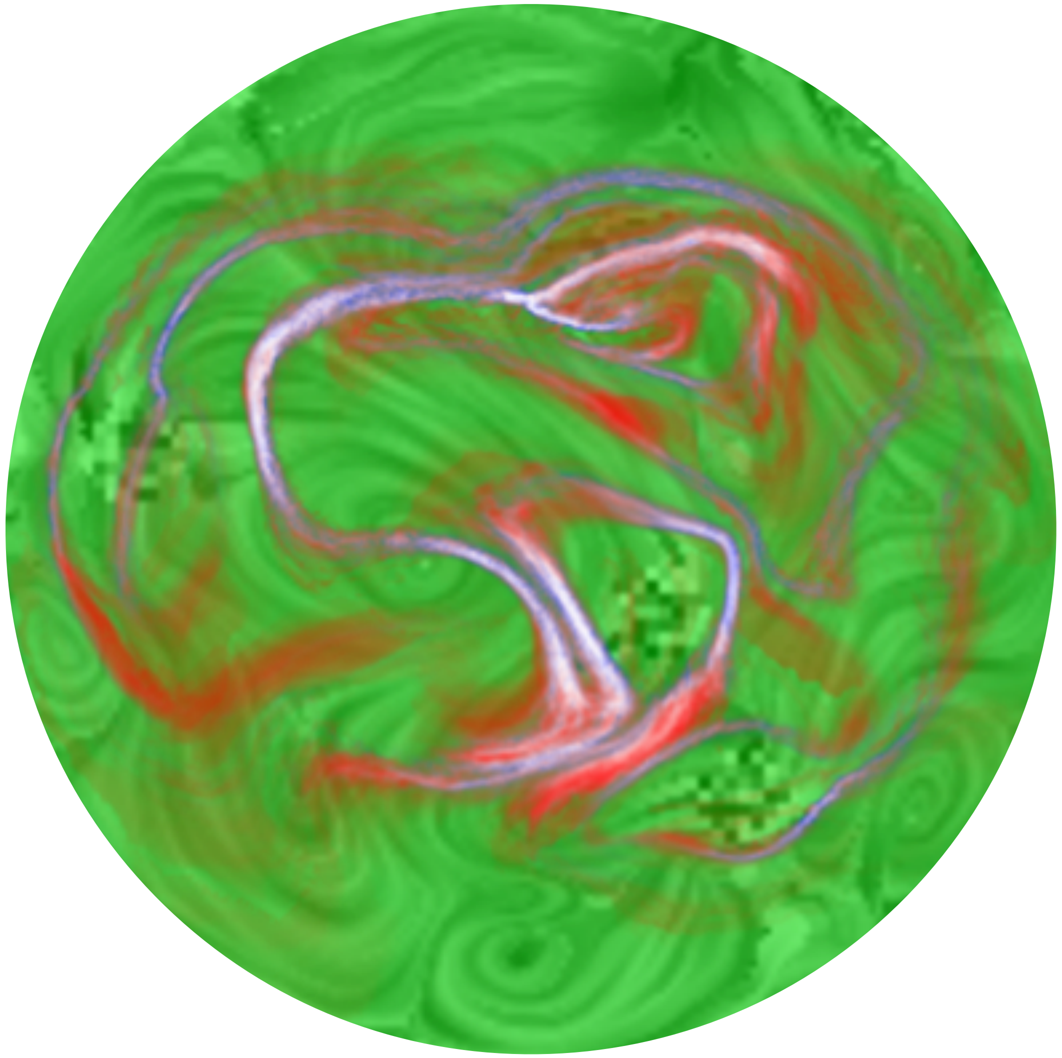
flow map domain

M. Hummel, C. Garth, H. Obermaier, K. I. Joy. *Comparative Visual Analysis of Lagrangian Transport in CFD Ensembles.* IEEE TVCG 19(12):2743–52, 2013 (best paper at IEEE SciVis 2013)

Features from Flow Map Ensembles

Example: Stirred mixing



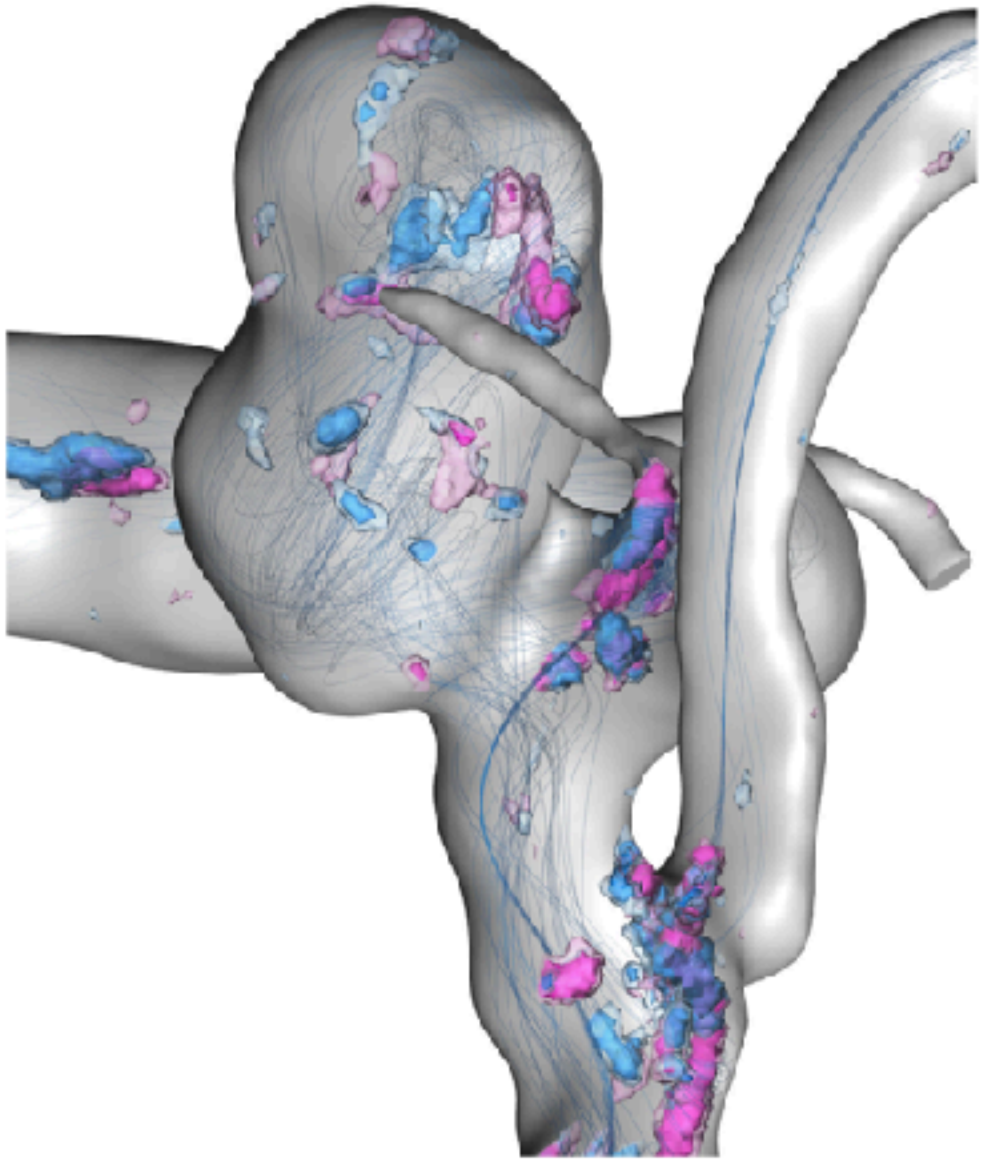


(data courtesy H. Obermaier, UC Davis)

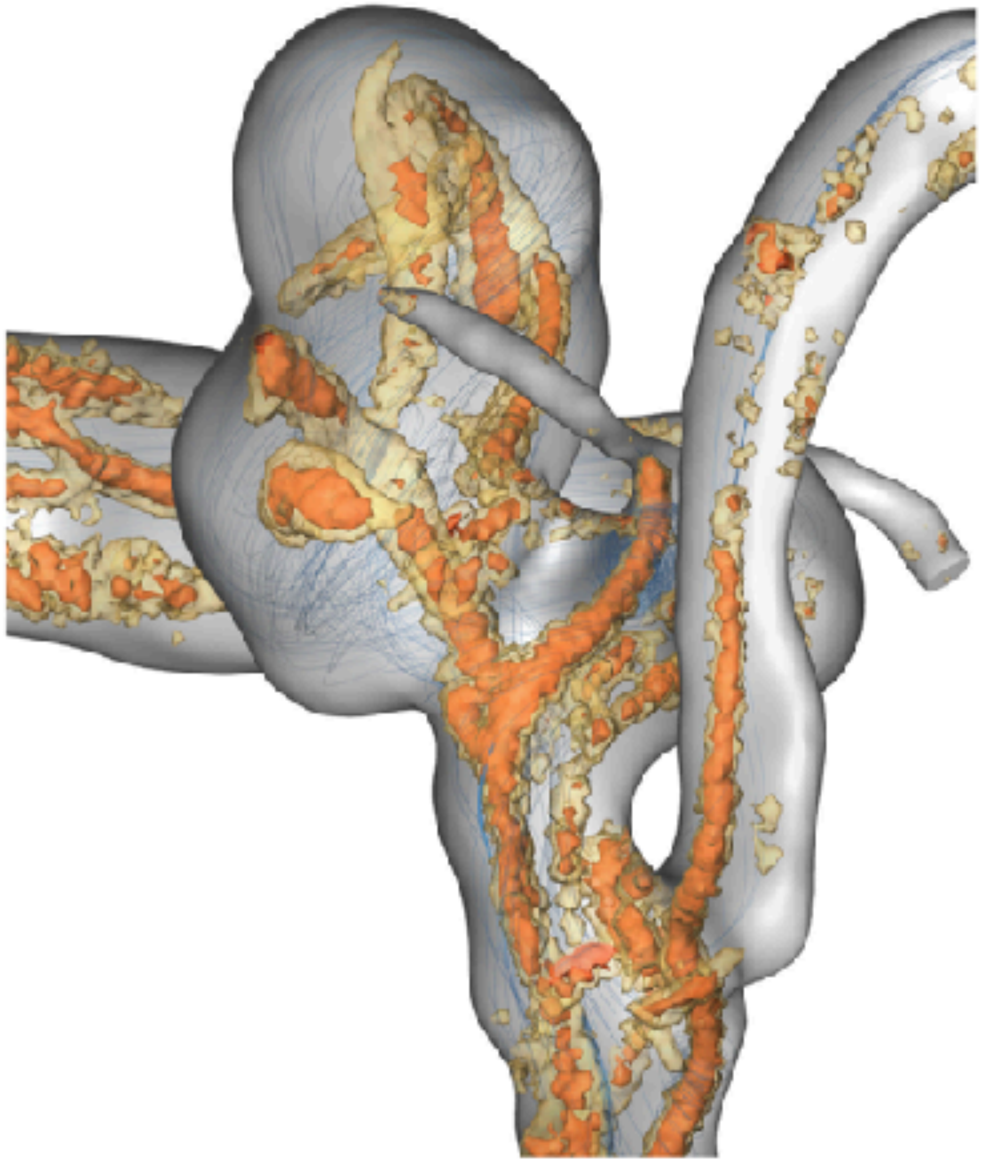
M. Hummel, C. Garth, H. Obermaier, K. I. Joy. *Comparative Visual Analysis of Lagrangian Transport in CFD Ensembles.* IEEE TVCG 19(12):2743–52, 2013 (best paper at IEEE SciVis 2013)

Features in Vector Field Ensembles

Flow features from stochastic / uncertain descriptions:



critical points



cores of swirling motion

Image from C. Petz, K. Pöthkow, H.-C. Hege: Probabilistic Local Features in Uncertain Vector Fields with Spatial Correlation. Probabilistic Local Features in Uncertain Vector Fields with Spatial Correlation. Computer Graphics Forum 31(3):1045–54, 2012.

Features in Vector Field Ensembles

Curve Boxplots: “median integral curves” as features

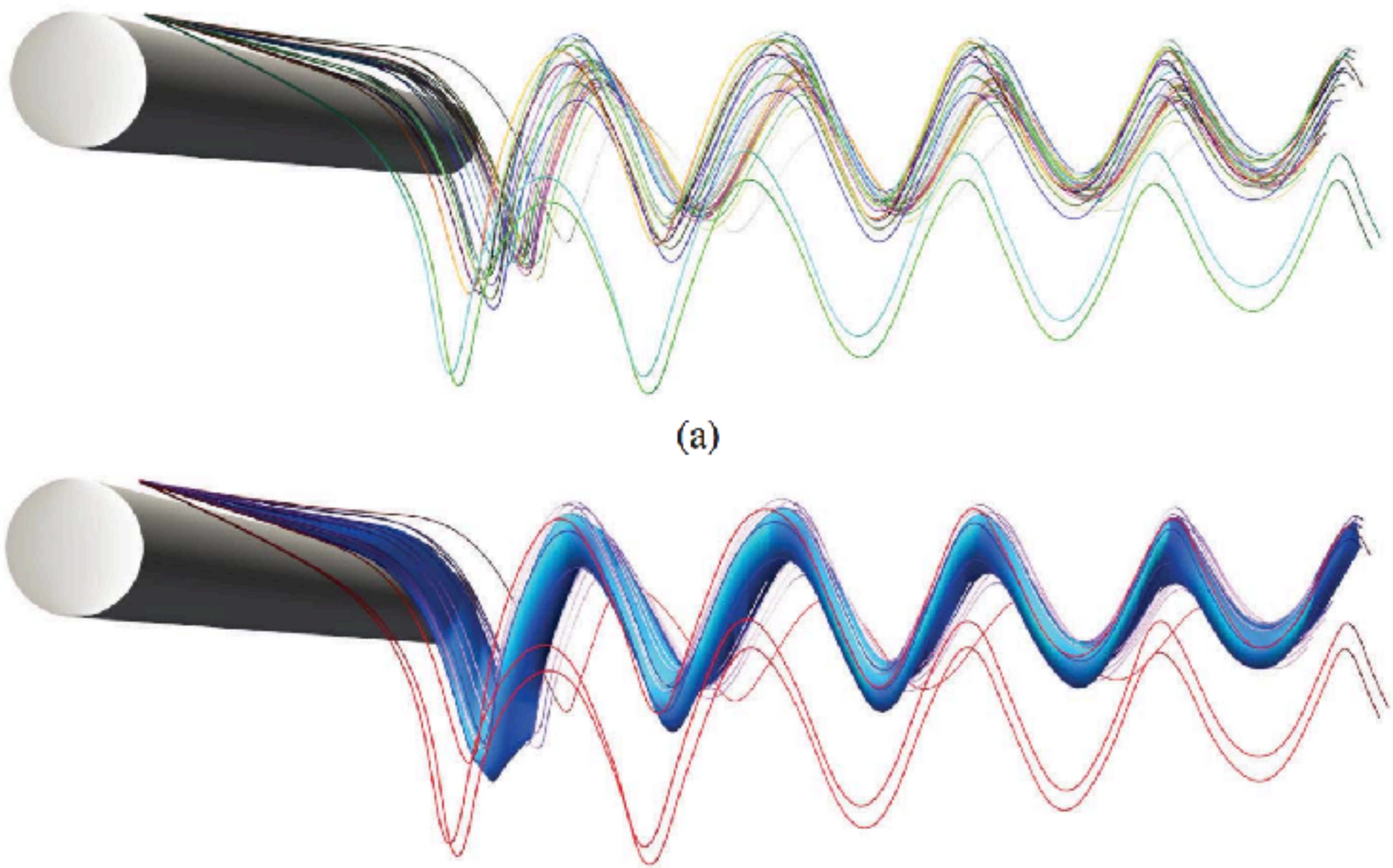
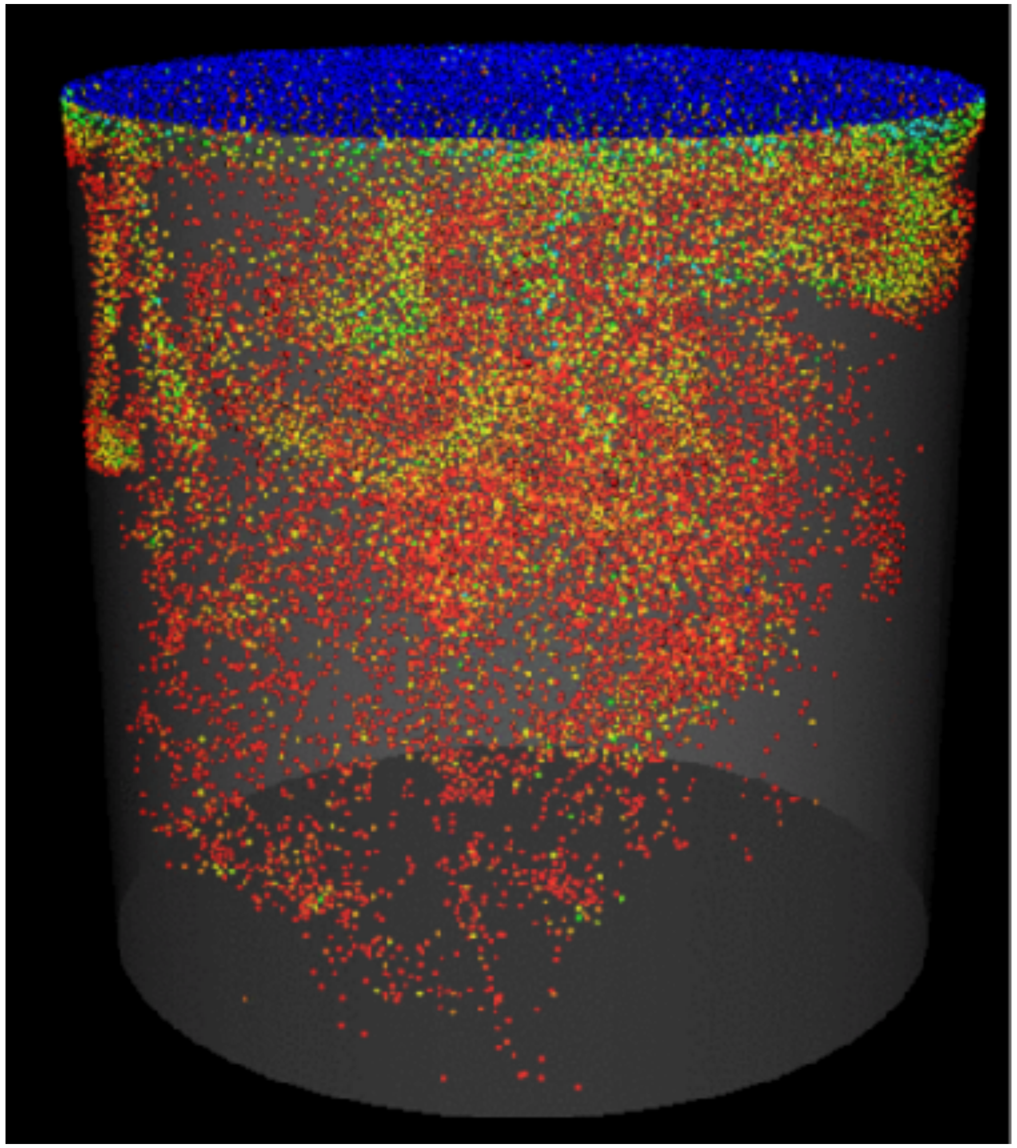


Image from M. Mirzargar, R. T. Whitaker, R. M. Kirby:
Curve Boxplot: Generalization of Boxplot for Ensembles of Curves. IEEE TVCG 20(12):2654–63, 2014.

Large Ensembles + Large Data

Shameless plug:
IEEE Scientific Visualization Contest 2016
Visualizing large ensembles of large particle simulations.

- 3 resolution levels
- 50 members per level
- time-varying simulation
- feature: concentration fingers



Six clever and unique solutions!
Wednesday, 2:00pm, Key 1+2+5

Conclusion

Conclusion

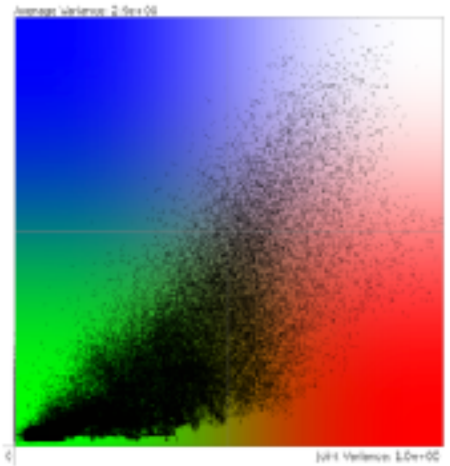
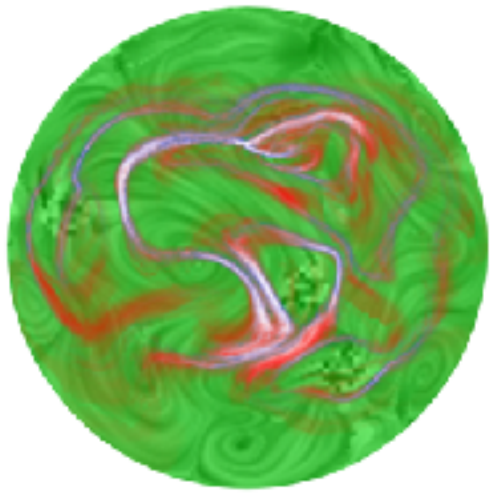
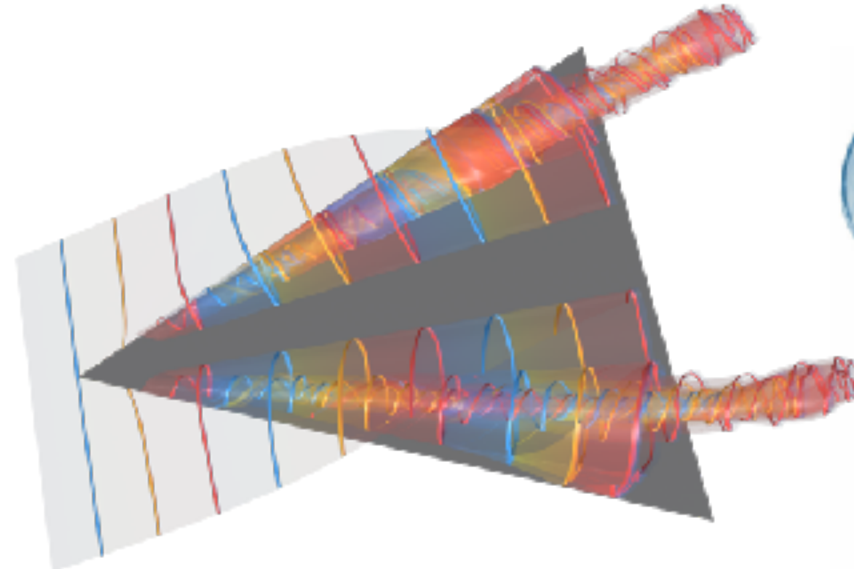
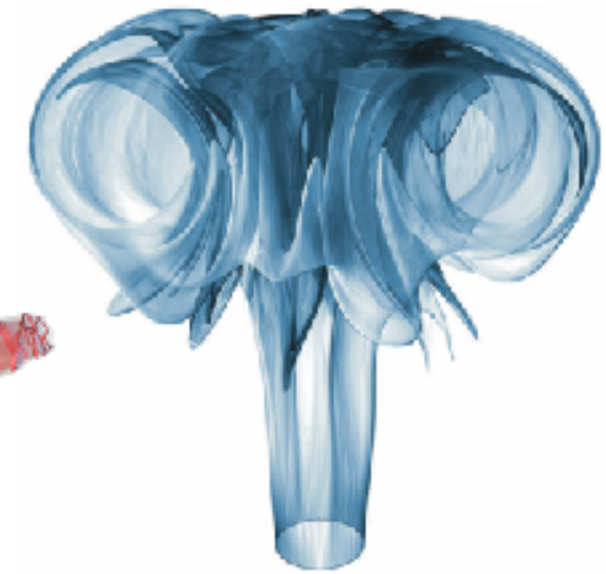
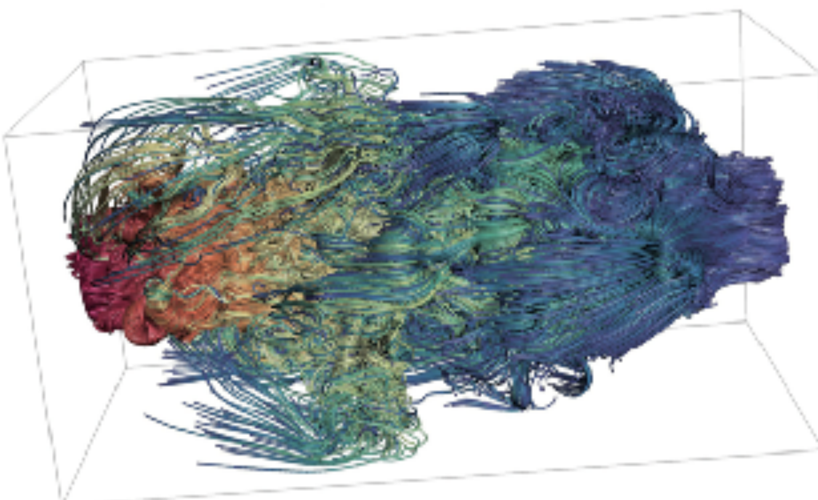
Large-data vector field visualization is still a widely open research problem. Some interesting techniques have been proposed:

- **computation:**
  parallel computation of integral curves for feature extraction

- **data reduction & browsing:**
  in situ extraction and intelligent storage of features

- **uncertainty modeling:**
  features in vector field ensembles

Still many open questions!

Thank you!

Questions?