VAPOR and Petascale Visualization

Alan Norton
National Center for Atmospheric Research
Boulder, CO USA
Presentation at CSCaDS August 5, 2009

This work is funded in part through a U.S. National Science Foundation, Information Technology Research program grant

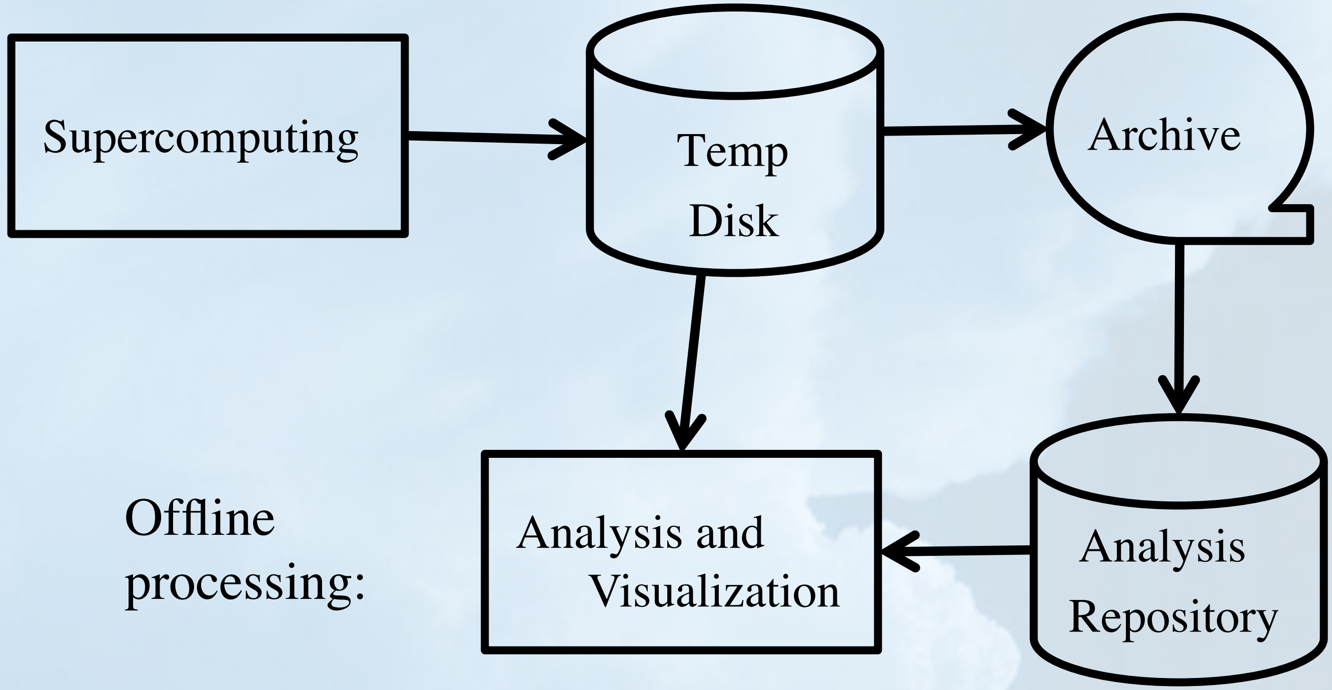
Outline

- The supercomputing/analysis workflow is broken!
- VAPOR and its approach to handling massive data
- VAPOR visualization and analysis capabilities
  - Seven techniques useful for *interactive* data understanding
- Demo:
  - Understanding high-resolution hurricane dynamics
- VAPOR's second generation data model for petascale applications
- Research efforts for petascale visualization

Alan's Big Adventure Part II

Alan Norton (vapor@ucar.edu)

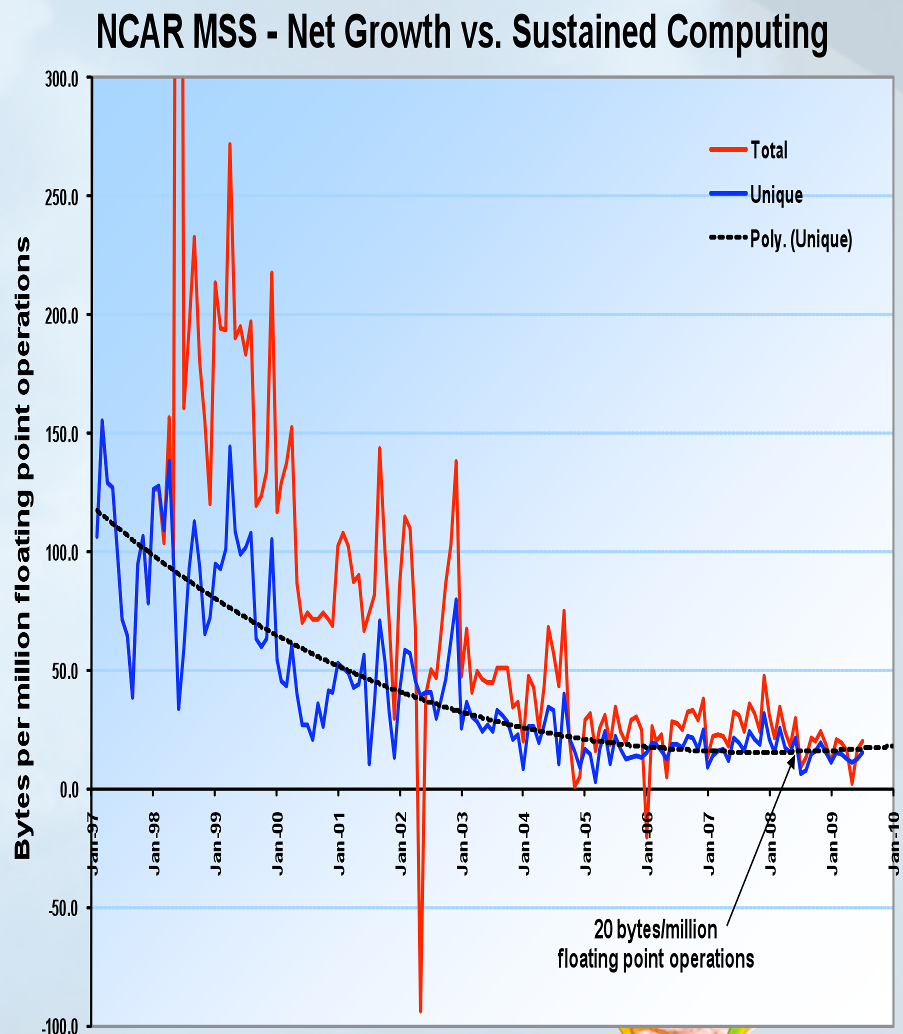
Typical Analysis/Vis Workflow



Offline processing:

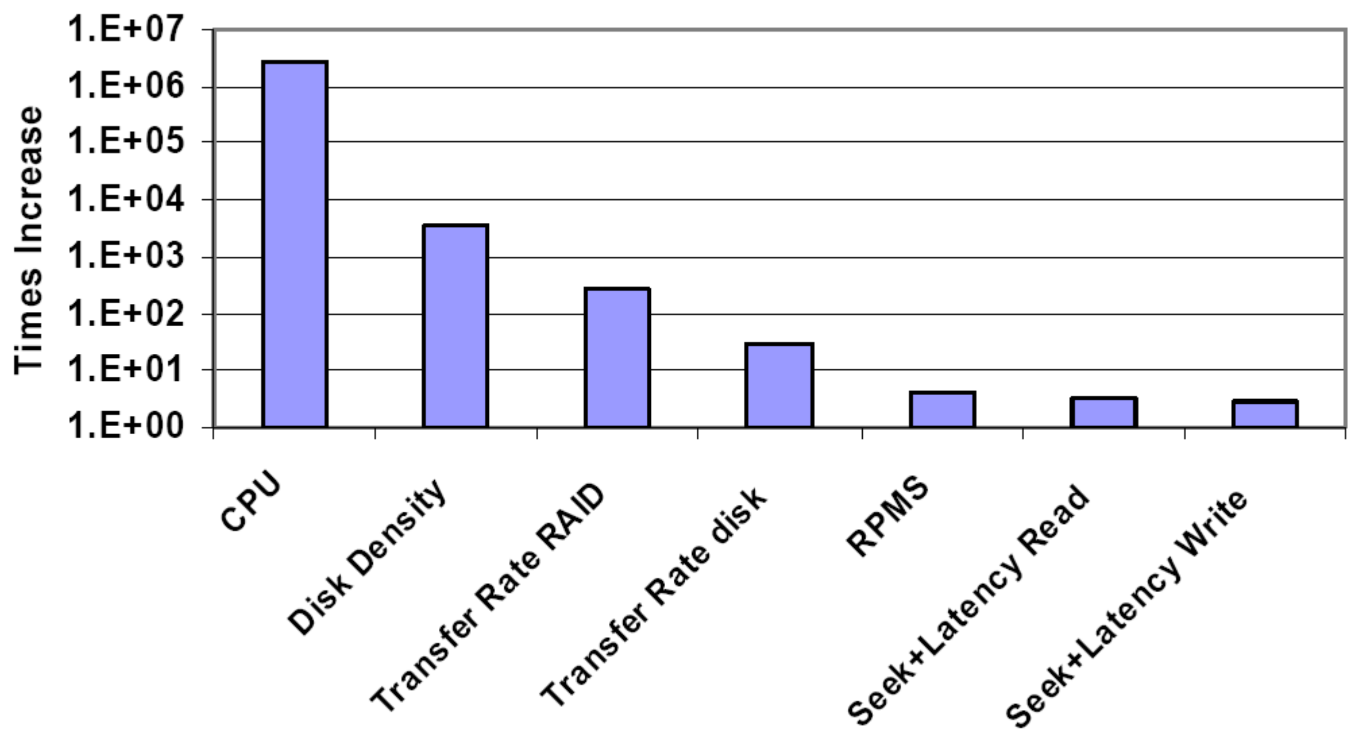
Archival is not keeping up

- Supercomputer sustained computation rate is doubling every 12-15 months
- Archive storage capacity is doubling every 25-26 months
- Fraction of data saved for analysis halves every ~2 years.



Visualization and Analysis are limited by I/O

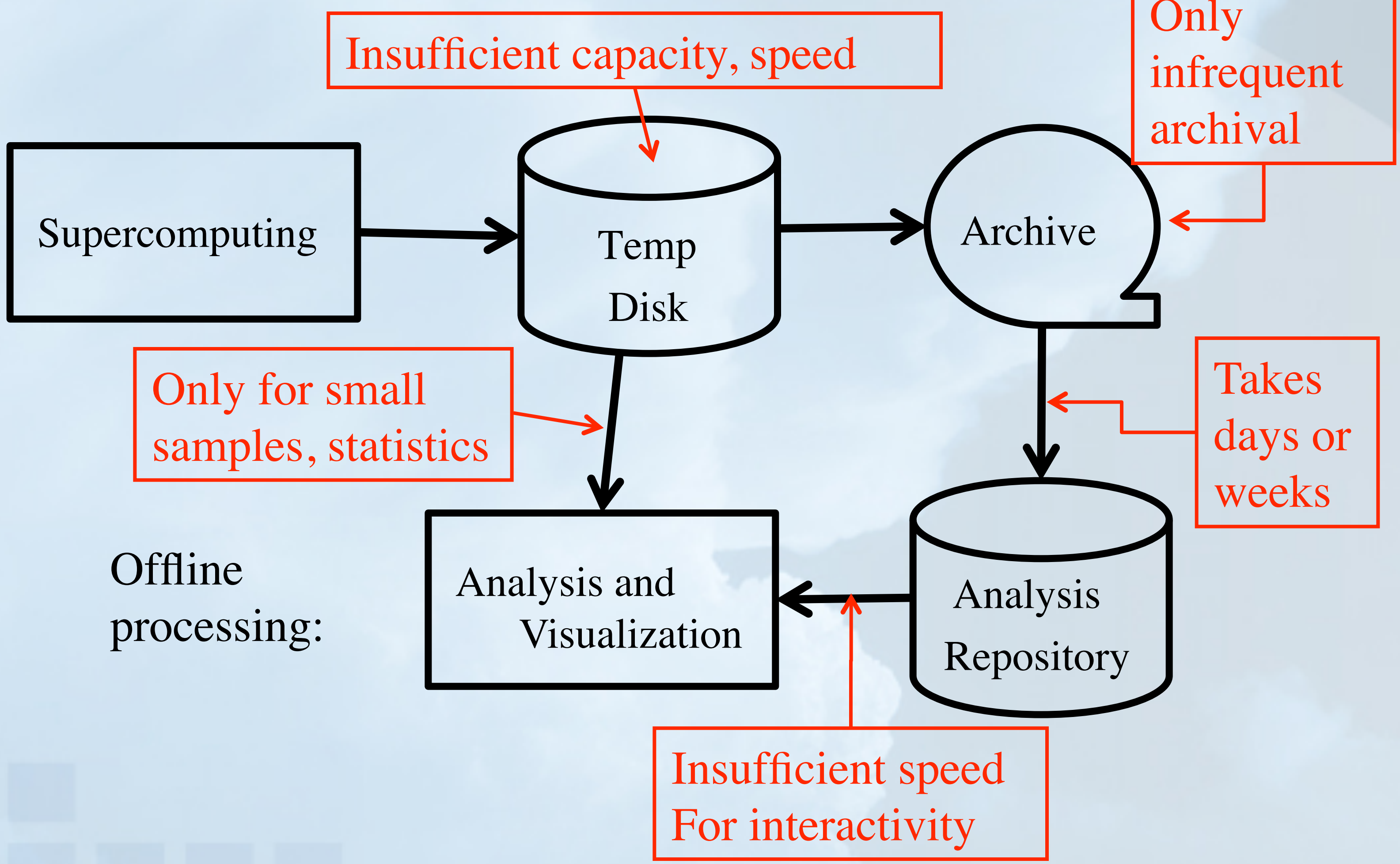
Performance improvements for I/O 1977-2005, compared with computation rate improvements



Source: DARPA HPCS I/O presentation

Alan Norton (vapor@ucar.edu)

Problems with Petascale Analysis/Vis Workflow



Implications for Visualization and Analysis of Petascale computations

Two serious problems:

- Smaller portion of data is available for analysis because of limited storage and archive capacity.
- Analysis and visualization of the available data becomes non-interactive due to limited IO rates

Result: Loss of scientific productivity

[Numerical] models that can currently be run on typical supercomputing platforms produce data in amounts that make storage expensive, movement cumbersome, visualization difficult, and detailed analysis impossible. The result is a significantly reduced scientific return from the nation's largest computational efforts.

Mark Rast
University of Colorado, LASP

What can be done to maximize value from supercomputing?

- Save more intelligently
  - Save only the most significant events
  - Save decimated or compressed data
  - Feature identification
  - Feature tracking
  - Use machine learning
  - Backup and rerun
- Improve interactivity of analysis and visualization
  - Exploit GPU's
  - Multi-resolution access
  - Provide tools customized to scientific needs

VAPOR project overview

The **VAPOR** project is intended to address the problem of datasets that are becoming too big to analyze and visualize interactively

- **VAPOR** is the **V**isualization and **A**nalysis **P**latform for **O**ceanic, atmospheric and solar **R**esearch
- **Goal**: Enable scientists to *interactively* analyze and visualize massive datasets resulting from fluid dynamics simulation
- **Domain focus**: 2D and 3D, gridded, time-varying turbulence datasets, especially earth-science simulation output.
- **Essential features:**
  - Multi-resolution data representation for accelerated data access
  - Exploits GPU for accelerated rendering
  - Interactive user interface for scientific visual data exploration

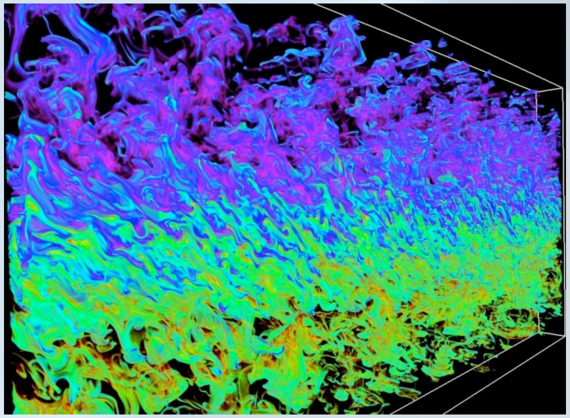
Wavelet transforms for 3D multiresolution data representation

- Some wavelet properties:
  - Permit hierarchical data representation
  - Invertible and lossless (subject to floating point round off errors)
  - Numerically efficient (O(n))
    - forward and inverse transform
  - No additional storage cost

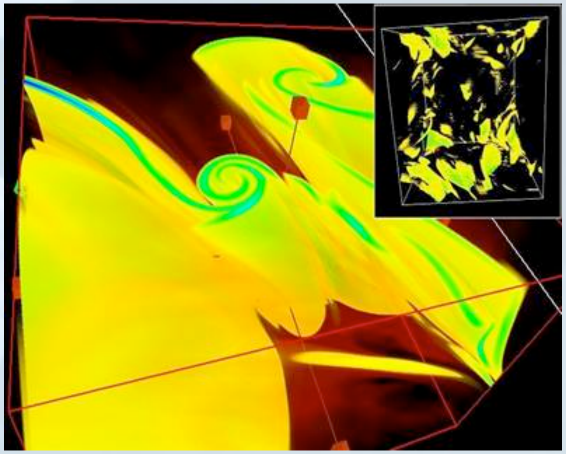
VAPOR capabilities (newest version 1.5)

Emphasis is on ease-of-use for fluid dynamics research

- All tools perform interactively, exploiting multi-resolution representation
- GPU-accelerated interactive graphics
  - Volume rendering
  - Isosurfaces
- Flow integration
  - Interactive streamlines and path lines
  - Field line advection
- Data probing and contour planes
- Support for WRF and terrain-following grids
- Geo-referenced image support
- Bidirectional integration with IDL® for analysis



Smyth, salt sheet boundary simulation



Mininni, Current roll

Alan Norton (vapor@ucar.edu)

How VAPOR differs from other visualization platforms

- Multi-resolution data representation
  - To enable interactive display and analysis of terabyte datasets
- Coupled with analysis toolkit (IDL®)
- Intended to be used by scientists, not visualization engineers
  - Requirements defined by a steering committee of scientists
- Narrow focus: turbulence simulation on gridded domains
- *Not* built on existing visualization libraries (e.g. VTK)
- Emphasis on desktop/laptop platforms; no parallel implementation

Interaction Techniques for understanding turbulence data with VAPOR

NCAR

Interactive feedback is key to visual data understanding

1. Multi-resolution data browsing
   – Enables interactive navigation of very large data
2. Visual color and transparency editing
   – Interactively control color and opacity to identify features of interest
3. Export/import data to/from analysis toolkit
   – Currently supporting IDL®
4. Use planar probe for visual flow seed positioning
   – Local data values guide seed placement
5. Animate flow with image-based flow visualization
6. Track evolving structures with field line advection
   – Animate field lines to understand time-evolution of structures
7. Use the GPU for interactive rendering
   – Accelerate volume rendering, isosurfaces on Cartesian or spherical grids

CISL

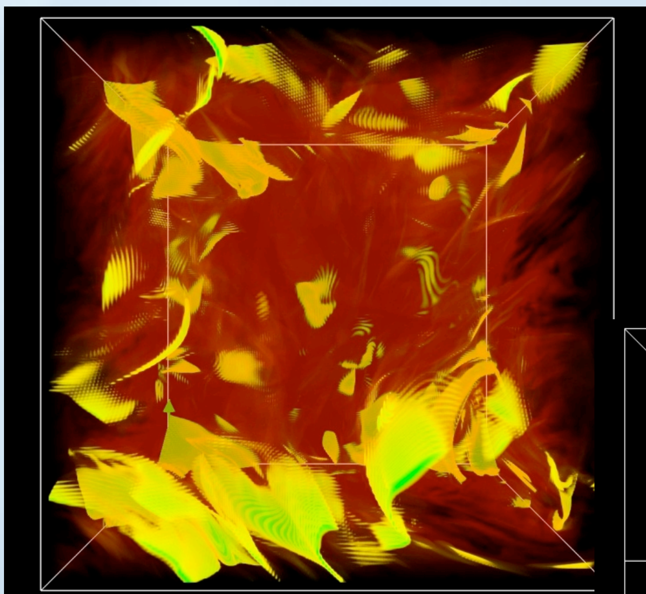
Alan Norton (vapor@ucar.edu)

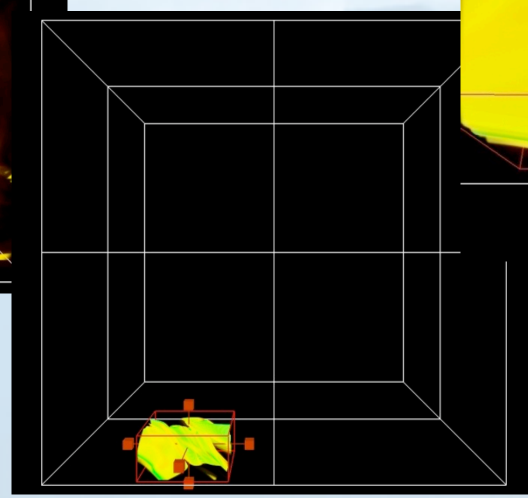
VAPOR Visualization & Analysis Platform

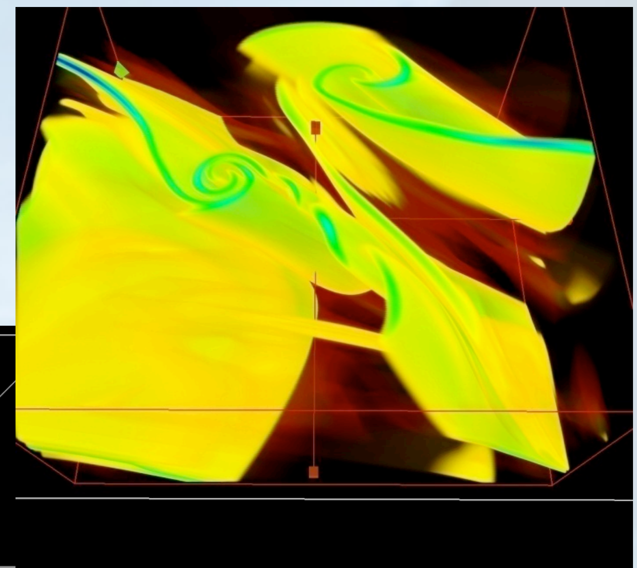
Interaction Technique 1: Multiresolution data browsing

Enabled by wavelet data representation

- Interactively visualize full data at low resolution
- Zoom in, increase resolution for detailed understanding





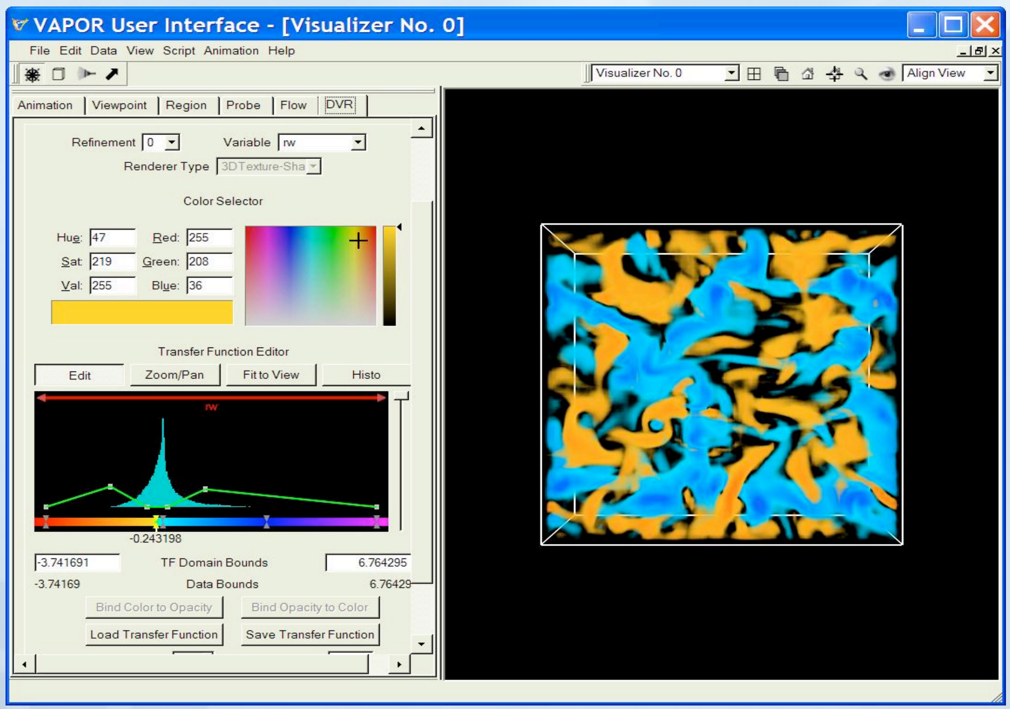


P. Mininni, current roll

Interaction Technique 2: Visual color/transparency editing

Design developed with Mark Rast

- Drag control points to define opacity and color mapping
- Histogram used to guide placement
- Continuous visual feedback in 3D scene



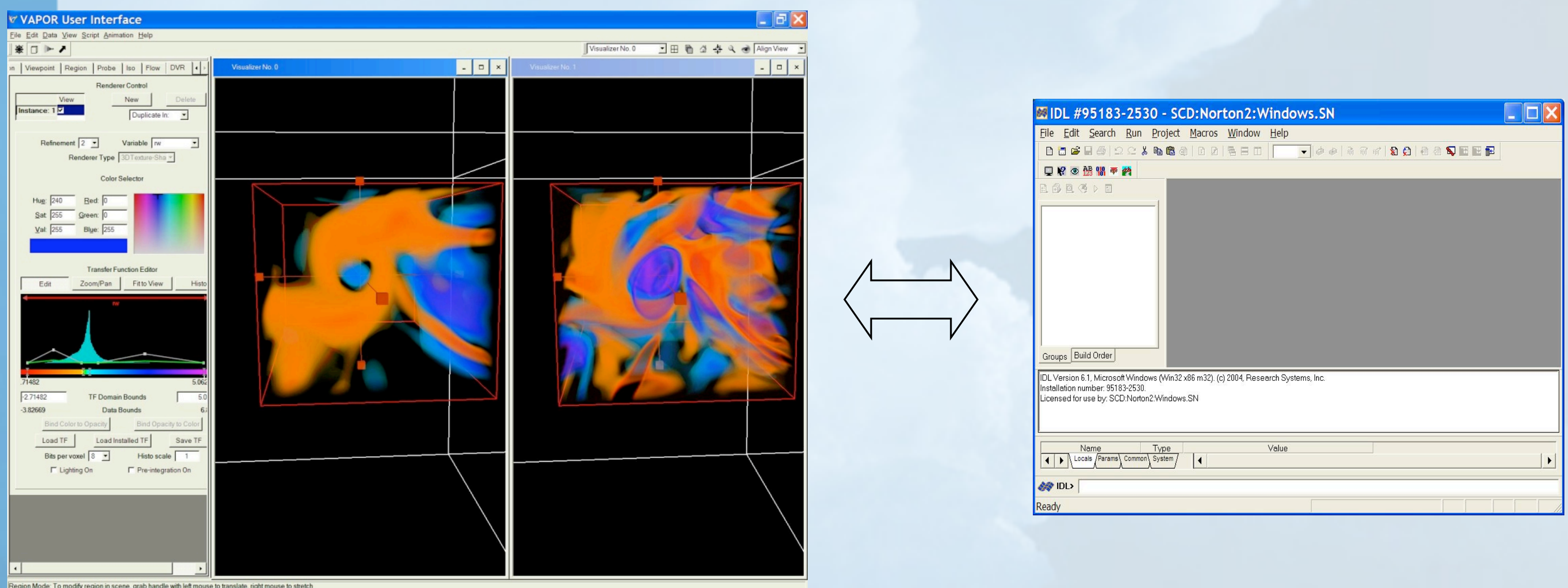
Alan Norton (vapor@ucar.edu)

Interaction Technique 3: Export/import data to/from analysis toolkit

NCAR

Currently using IDL®

- User specifies region to export to IDL session
- IDL performs operations on specified region in IDL
- Results imported as new variables in VAPOR

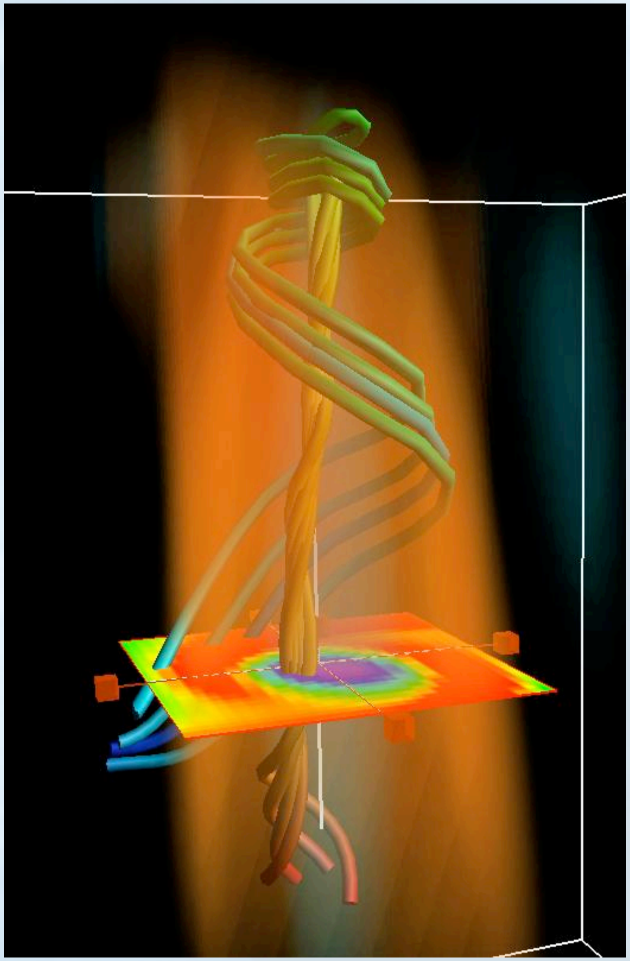


Interaction Technique 4: Use planar probe for visual flow seed placement

NCAR

Useful to place flow seeds based on local data values

- Planar probe provides cursor for precise placement in 3D
- Field lines are immediately reconstructed as seeds are specified

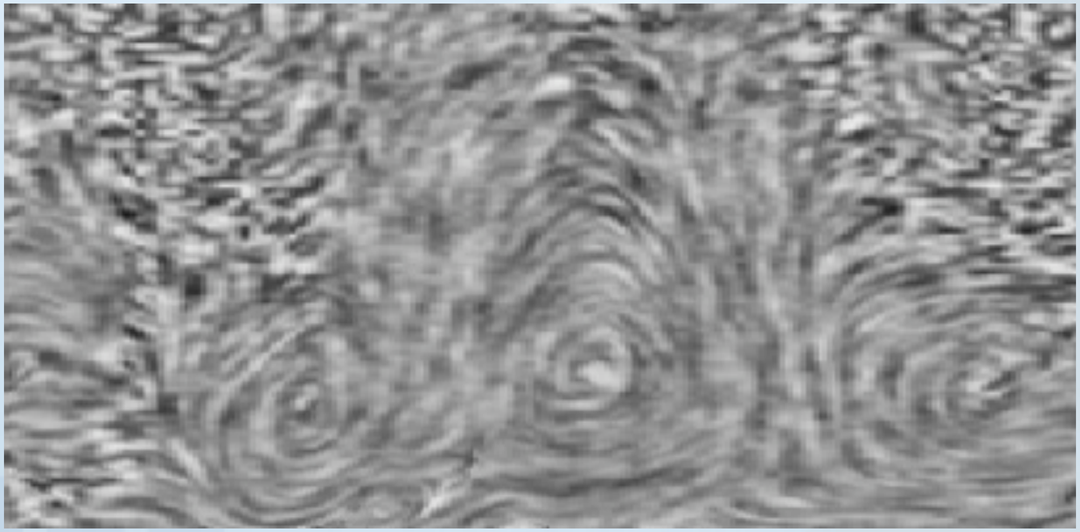


Data provided by Mark Rast

Interaction Technique 5: Visualize flow in animated planar sections

NCAR

- Implements "Image Based Flow Visualization" technique of Jarke Van Wyck
- Spot noise pattern advected in planar projection of velocity field, results blended into successive images.
- Vortices in cross-section of hurricane eye-wall:

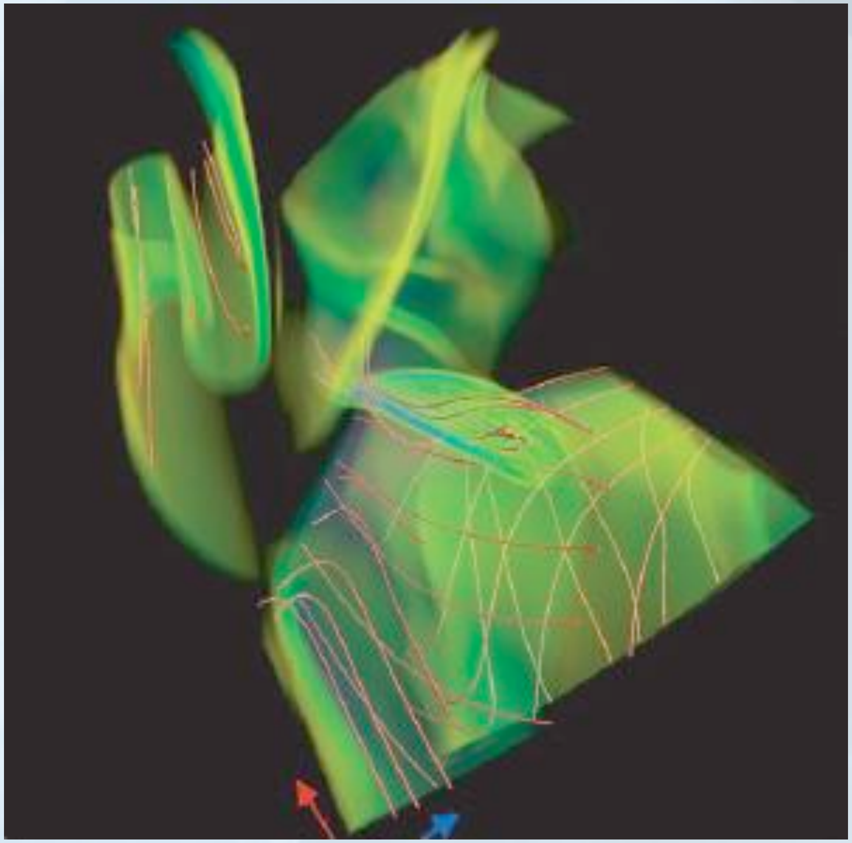


Data provided byYongsheng Chen

Interaction Technique 6:
Track evolving structures with field line advection

Animates field lines in velocity field

- Useful for tracking evolution of geometric structures (e.g. vorticity field lines in tornado)
- Based on algorithm proposed by Aake Nordlund

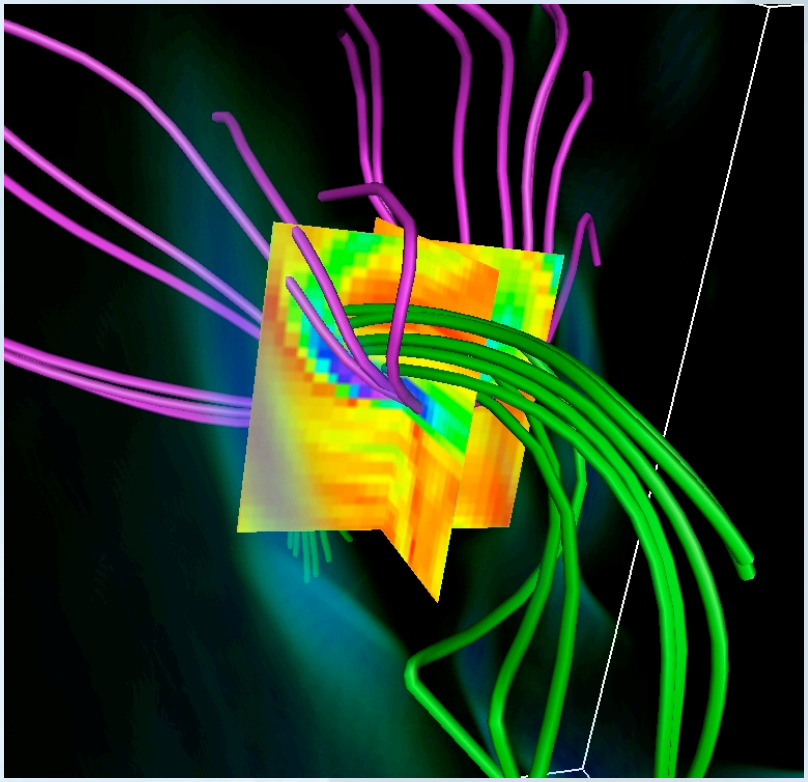


Data provided by P. Mininni

Interaction Technique 6: Track evolving structures with field line advection

Animates field lines in velocity field

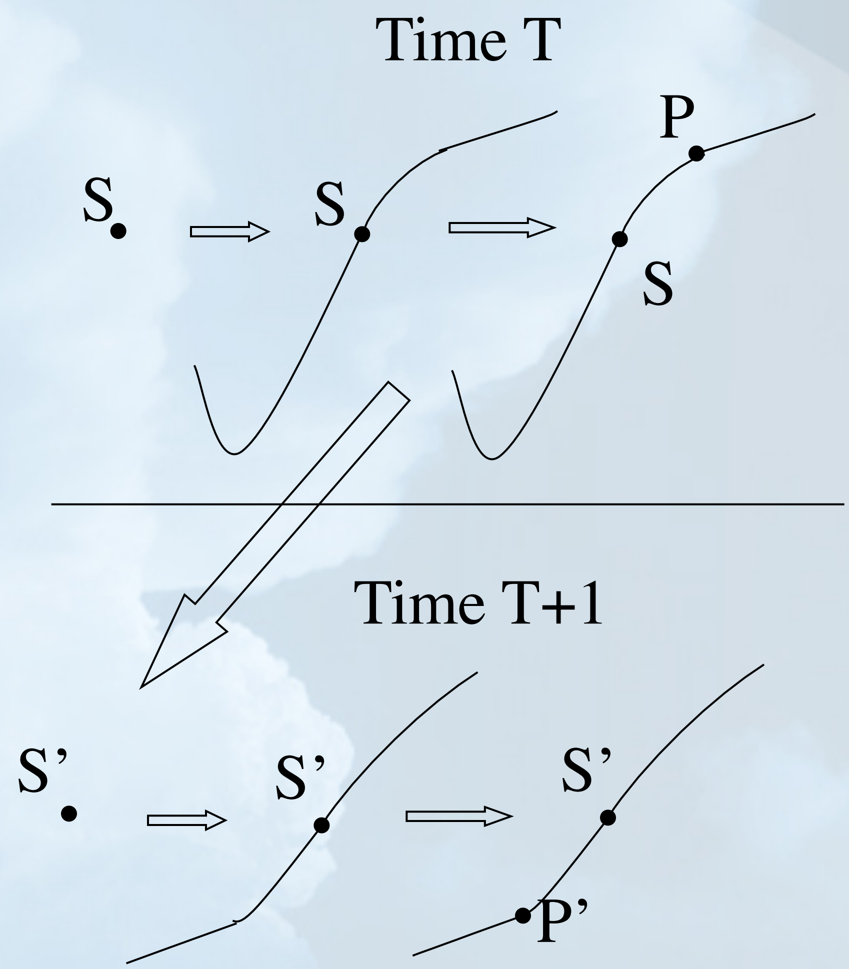
- Useful for tracking evolution of geometric structures (e.g. vorticity field lines in tornado)
- Based on algorithm proposed by Aake Nordlund



Data provided by P. Mininni

Field Line Advection: how it works

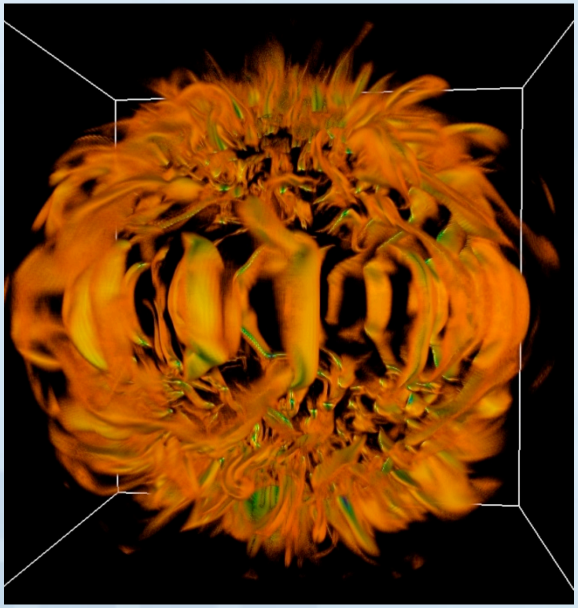
- Enables animation of magnetic field lines in a velocity field
- Algorithm proposed by Aake Nordlund;
  - Start with seed point S at an initial time step
  1. Construct field line through S
  2. Find point P along field line having maximal field strength
  3. Time-advect P to subsequent time step, resulting in new seed point S’
  - Repeat steps 1-3 for each additional time step



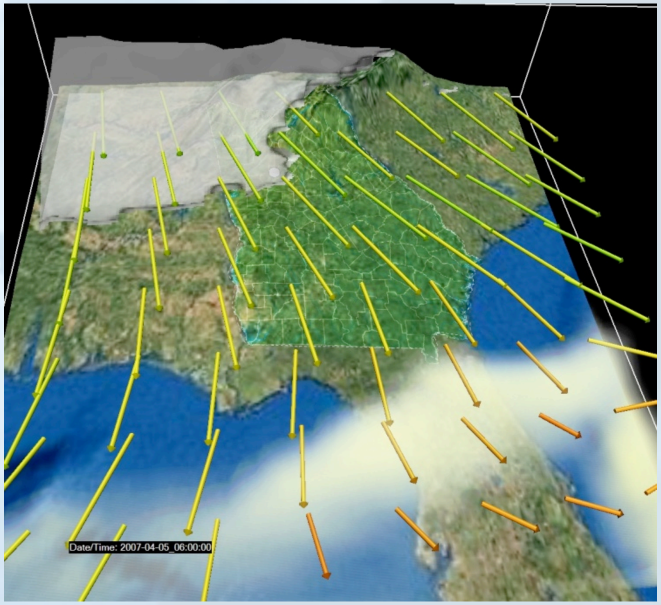
Alan Norton (vapor@ucar.edu)

Interaction Technique 7: Use the GPU for interactive rendering

- Exploit modern GPU's for accelerated rendering
  - GPU's are SIMD clusters, efficiently traverse data arrays
  - Support for Cartesian, spherical, terrain-following grids



B. Brown, Solar MHD simulation

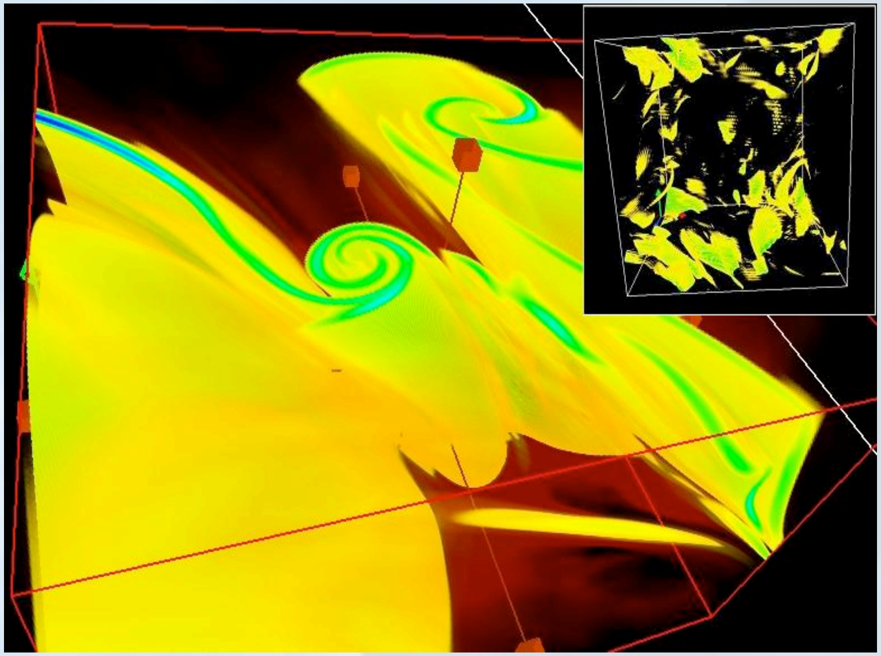


T. Prabhakaran, April 2007 cold event in WRF

MHD exploration DEMO

Small scale structures in MHD turbulence with high Reynolds number

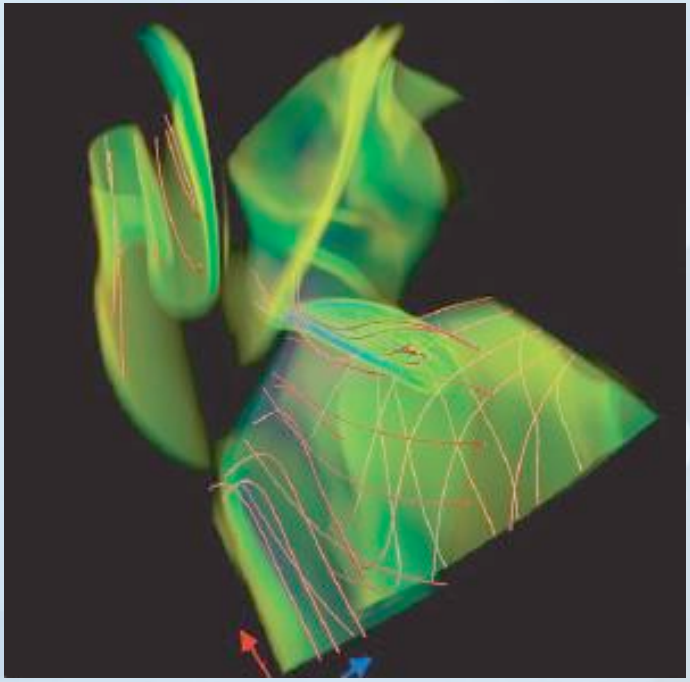
- Data from Pablo Mininni, NCAR
- 1536x1536x1536 volume, 16 variables (216 GB per timestep)
- Scientific goal: understand MHD flow dynamics at high resolution and high Reynolds no.
- Analysis and visualization performed with VAPOR and IDL
- Resulted in discovery of intertwining current sheets (“current rolls”)



Use field line advection to track structural changes

NCAR

- Subsequent work investigated evolution of current roll using field line advection
- FLA follows magnetic field lines, tracking structural changes



P. Mininni et al., NJP 2008

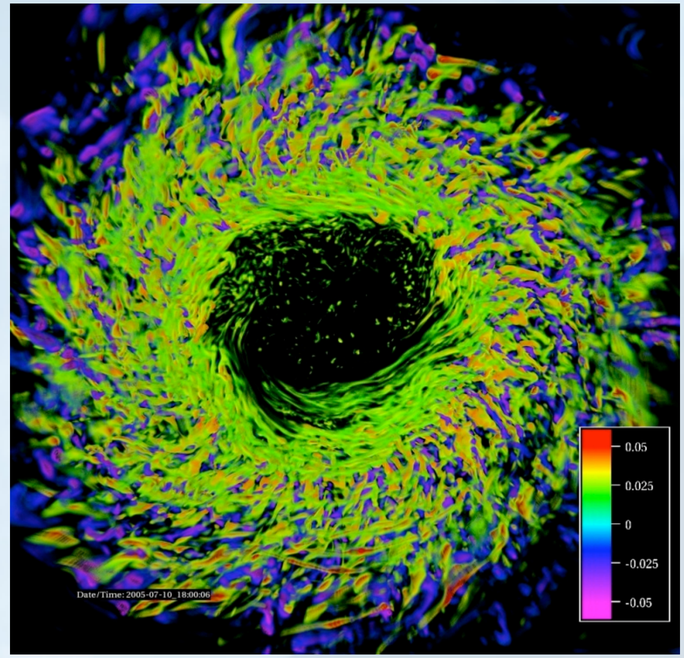
CISL

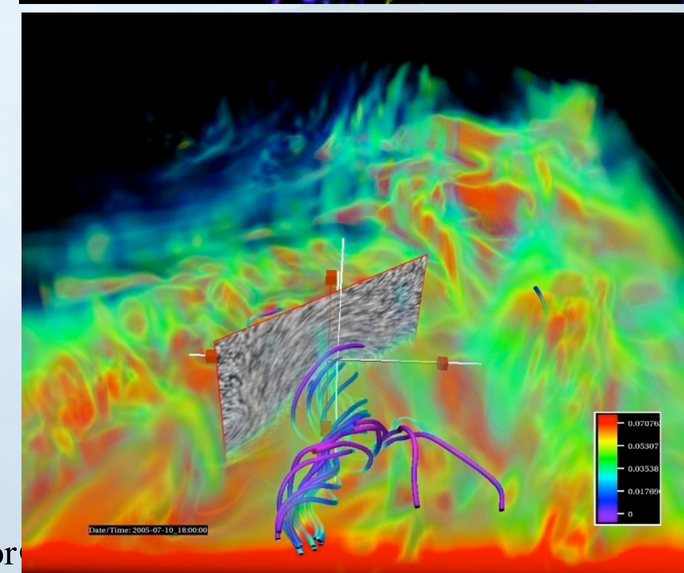
VAPOR Visualization & Analysis Platform

High-res hurricane analysis demo

Visualization was used to understand the nature of increased turbulence along eye-wall

- Unsteady flow shows overall wind dynamics
- VAPOR's IBFV tool is used to identify horizontally oriented transient vortex tubes near the ocean surface
- Using VAPOR's field line advection, these vortices can be tracked and animated over time

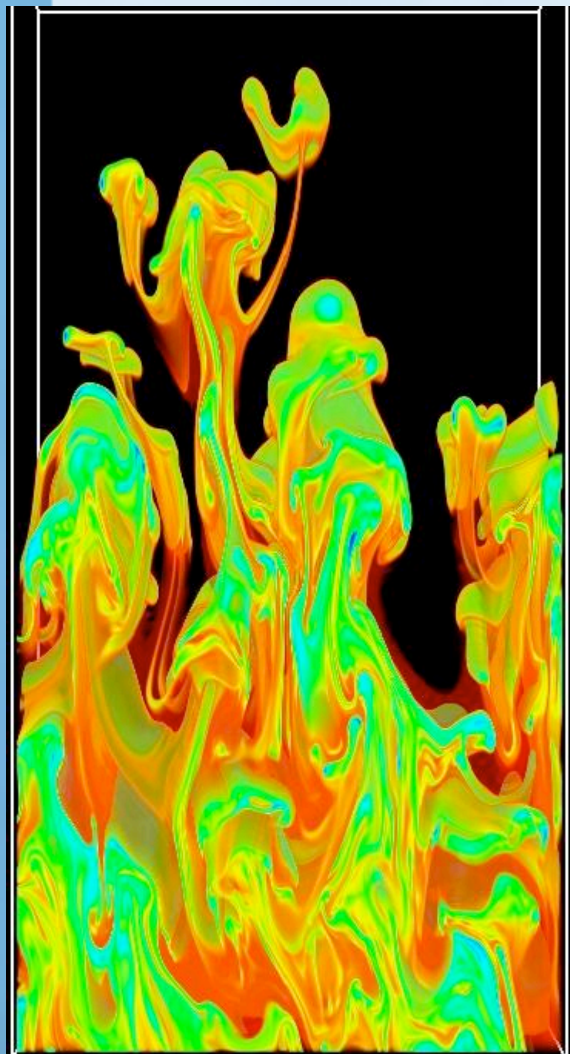




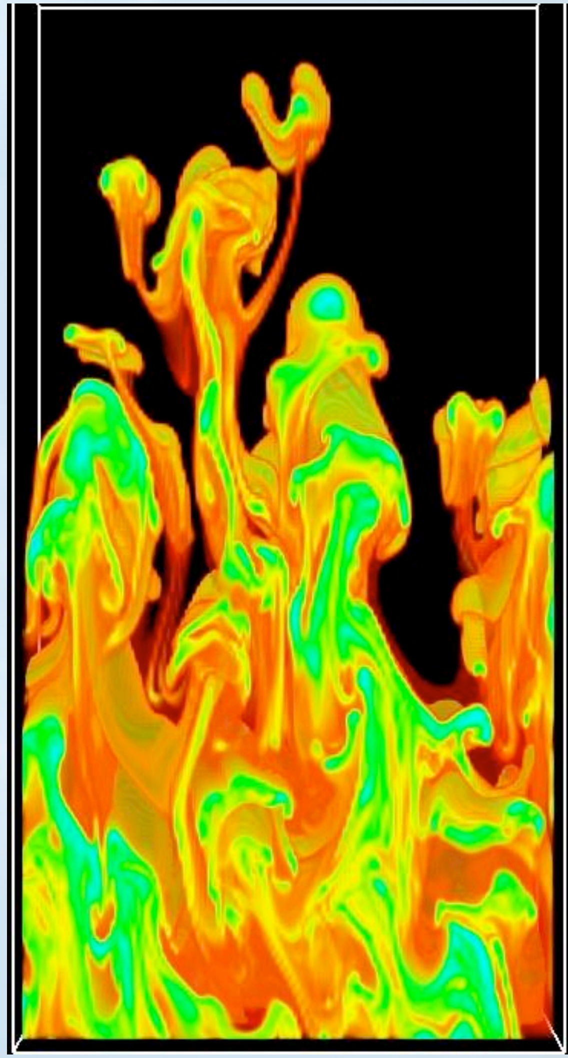
VAPOR plans for Petascale computation

- VAPOR's existing multi-scale model is useful in terascale, but not sufficient for petascale apps
- We are planning an improved data model in VAPOR
  - Want a reduction factor of 100 or more in I/O for visualization and analysis
  - Use wavelet compression/decompression using wavelet families that efficiently compress turbulence data
  - Data access model must be designed to work efficiently in the petascale analysis/visualization workflow
- J. Clyne: Analyzing design choices, e.g.:
  - Choice of wavelet
  - Data blocking (needed for efficient access)
  - Coefficient prioritization
  - Boundary extension method

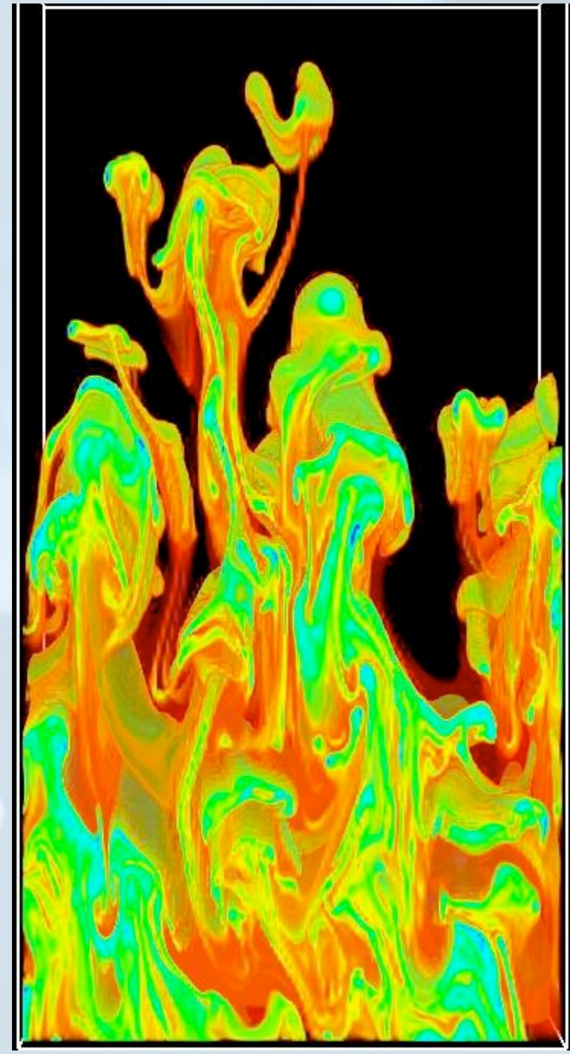
64- way volume compression of salt density in Smyth's salt infusion results



original



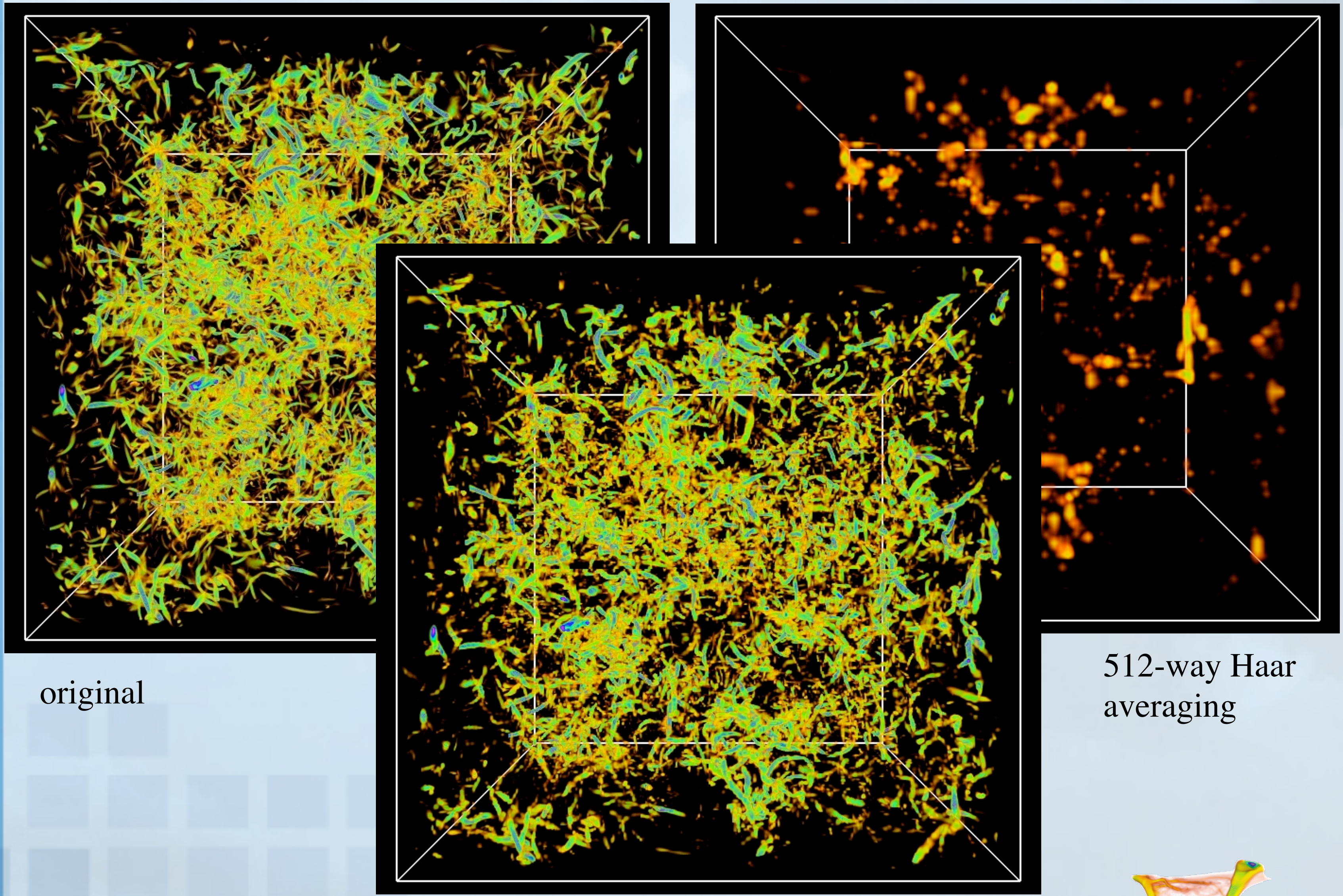
64-fold averaged with Haar



64-fold compress biorth spline

John Clyne (vapor@ucar.edu)

512-way compression of Mininni's MHD data



original

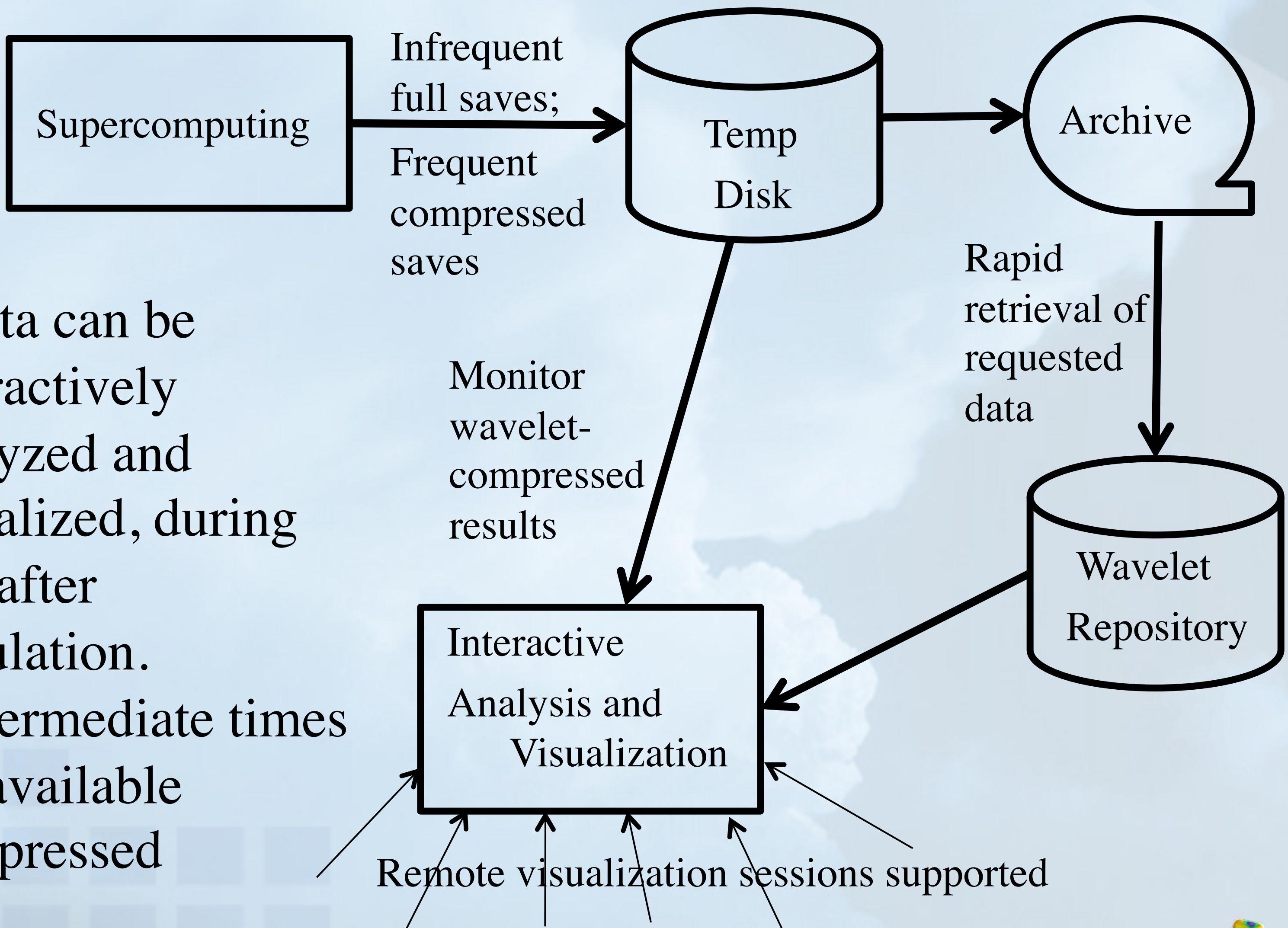
512-way Haar averaging

512-way compressed using biortho spline wavelets

Next generation plans and requirements

- We are preparing a second generation of multi-scale infrastructure for an upcoming release of VAPOR.
- Requirements include:
  - Efficient coefficient database structure must enable prioritized access and must exploit locality
  - Optimal wavelet compression performance can depend on type of data, choice of wavelet
  - Must tolerate coded values (e.g. "fill_value") that can interfere with compression
  - Wavelet processing must operate efficiently with petascale apps

Desired workflow using wavelet compression



- Data can be interactively analyzed and visualized, during and after simulation.
- Intermediate times are available compressed

Research efforts for improving analysis in petascale computation

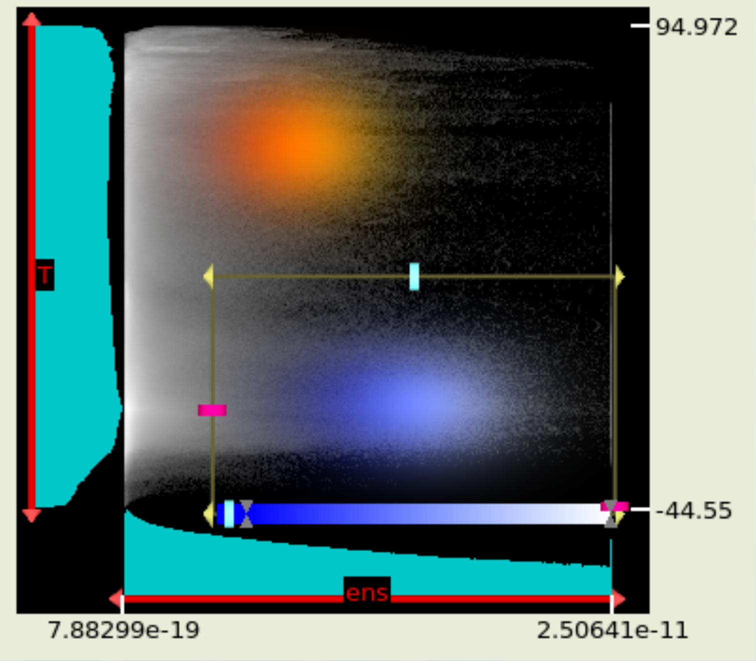
Two experimental VAPOR extensions for facilitating visual understanding of petascale computation.

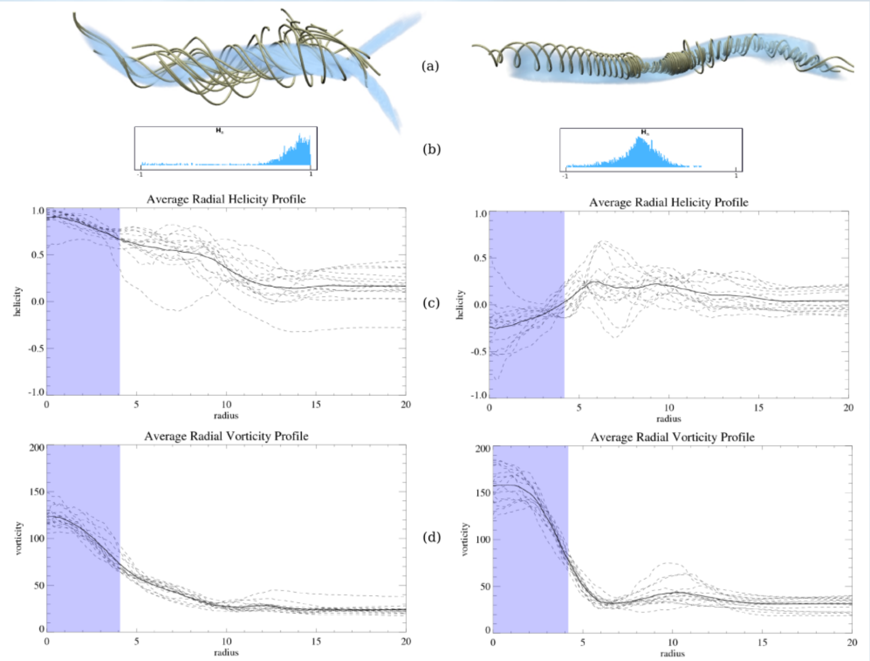
General need: Enable interactive analysis when I/O constraints prohibit exhaustive search or visualization of full data sets

- Feature identification (K. Gruchalla, M. Rast)
  - Interactively search for features of interest in massive data volumes, based on statistical properties
- Feature tracking (P. Mininni)
  - When an important feature has been discovered at a specific time, extract only the data needed to advect this feature (forward and backward in time), without examining the full volume at all time steps.

Feature Identification (K. Gruchalla)

- Use a multidimensional transfer function to identify and select structures of interest
- Connected component analysis isolates these structures.
- Local statistical measures used to classify dynamics

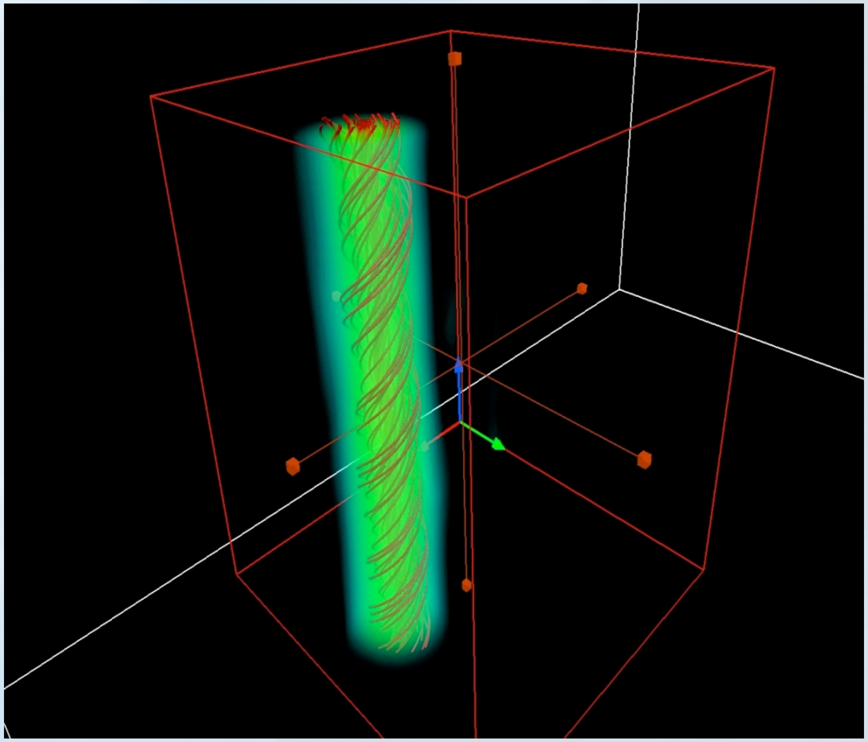




Alan Norton (vapor@ucar.edu)

Region advection for Feature tracking (P. Mininni)

- Problem: Simulation output can only be infrequently saved. Retrieval and examination of full data at multiple time steps is too time-consuming
- Proposed approach:
  - Identify feature of interest in small region
  - Determine motion of feature, based on appropriate mathematical model
  - Apply field line advection algorithm to advect feature
  - Retrieve data at other time-steps, restricted to the time-varying volume associated with the advected region.



VAPOR plans

VAPOR's steering committee and other users help prioritize features, with releases every 6-9 months

- Vapor 1.5 is being prepared for release

Some high priority features for an upcoming release:

- Next-generation data model for petascale:
  - Compression support
  - Prioritized data access
  - Parallel conversion
- Built-in expression calculator
- Animation control
- Extensible architecture
- Direct import of data

Summary

- VAPOR is designed to enable interactive visualization and analysis of massive datasets by exploiting the wavelet multi-scale representation.
- VAPOR supports a variety of useful interactive techniques for investigating and visualizing data, based on needs expressed by scientific users.
- We are developing improvements to VAPOR to enable interactive access to petabyte datasets, and to support anticipated petascale workflows.

VAPOR Availability

- Version 1.5.0 preparing for release (momentarily!)
  - Version 1.4.2 available on Website
- Runs on Linux, Windows, Mac
- System requirements:
  - a modern (nVidia or ATI) graphics card (available for about $200)
  - ~1GB of memory
- Software dependencies
  - IDL® http://www.ittvis.com/ (only for interactive analysis)
- Executables, documentation available (free) at http://www.vapor.ucar.edu/
- Source code, feature requests, etc. at http://sourceforge.net/projects/vapor

Acknowledgements

- Steering Committee
  - Nic Brummell - CU
  - Yuhong Fan - NCAR, HAO
  - Aimé Fournier – NCAR, IMAGe
  - Pablo Mininni, NCAR, IMAGe
  - Aake Nordlund, University of Copenhagen
  - Helene Politano - Observatoire de la Cote d'Azur
  - Yannick Ponty - Observatoire de la Cote d'Azur
  - Annick Pouquet - NCAR, ESSL
  - Mark Rast - CU
  - Duane Rosenberg - NCAR, IMAGe
  - Matthias Rempel - NCAR, HAO
  - Geoff Vasil, CU
  - Thara Phrabhakaran, IITM
  - Gerry Creager, TAMU
  - Leigh Orf, Central Mich. U.
- WRF consultation
  - Wei Wang – NCAR, MMM
  - Cindy Bruyere –NCAR, MMM
  - Yongsheng Chen-NCAR,MMM
  - Wei Huang – NCAR, CISL
- Developers
  - John Clyne – NCAR, CISL
  - Alan Norton – NCAR, CISL
  - Kenny Gruchalla – CU
  - Victor Snyder – CSM
  - Rick Brownrigg – NCAR, CISL
  - Pam Gillman – NCAR, CISL
- Research Collaborators
  - Kwan-Liu Ma, U.C. Davis
  - Hiroshi Akiba, U.C. Davis
  - Han-Wei Shen, Ohio State
  - Liya Li, Ohio State
- Systems Support
  - Joey Mendoza, NCAR, CISL

Questions?

Alan Norton (vapor@ucar.edu)