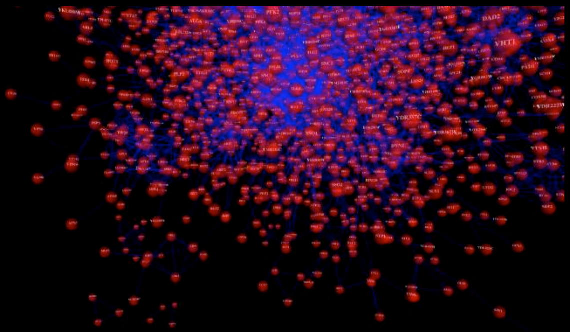
Visualization at TACC

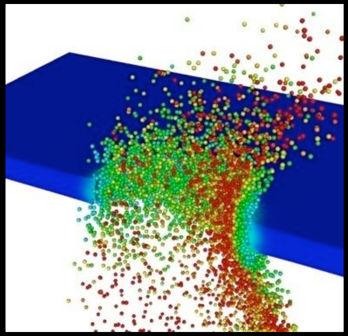
Kelly Gaither
Director, Data & Information Analysis
Research Scientist
Texas Advanced Computing Center

July 29, 2010

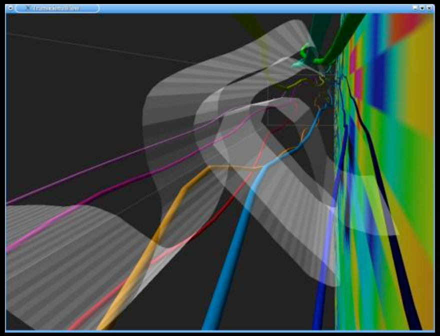
Visualization at TACC



Bioinformatics

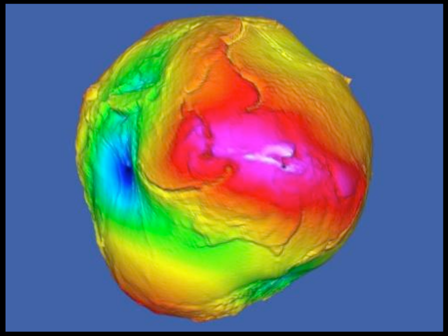


Orbital Debris

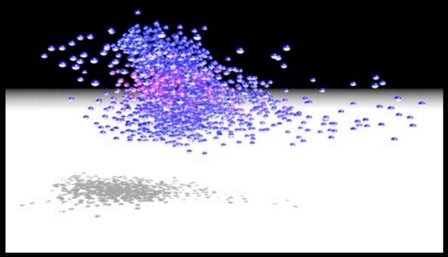


Turbulent Flow

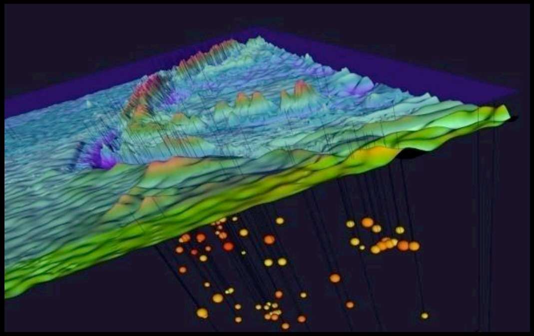
CT Models



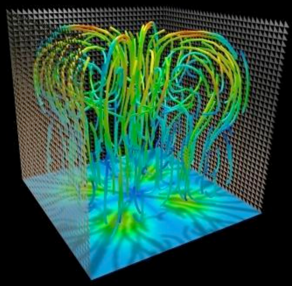
Gravity Map



Quantum Chemistry



GeoSciences



Natural Convection

TACC Visualization Group

- Provides resources/services to a growing local and national user community to enable scientific discovery and insight.
- Researches and develops tools/techniques for the next generation of problems facing the user community.
- Trains the next generation of scientists to visually analyze datasets of all sizes.

TACC Visualization Personnel

- 10 Full Time Staff
- 2 Undergraduate Students
- 1 PhD Student
- Areas of Expertise
  - Large Scale GPU Clusters
  - Large Scale Tiled Displays
  - Remote & Collaborative Visualization
  - Large Data Visualization
  - Data Mining & Feature Detection

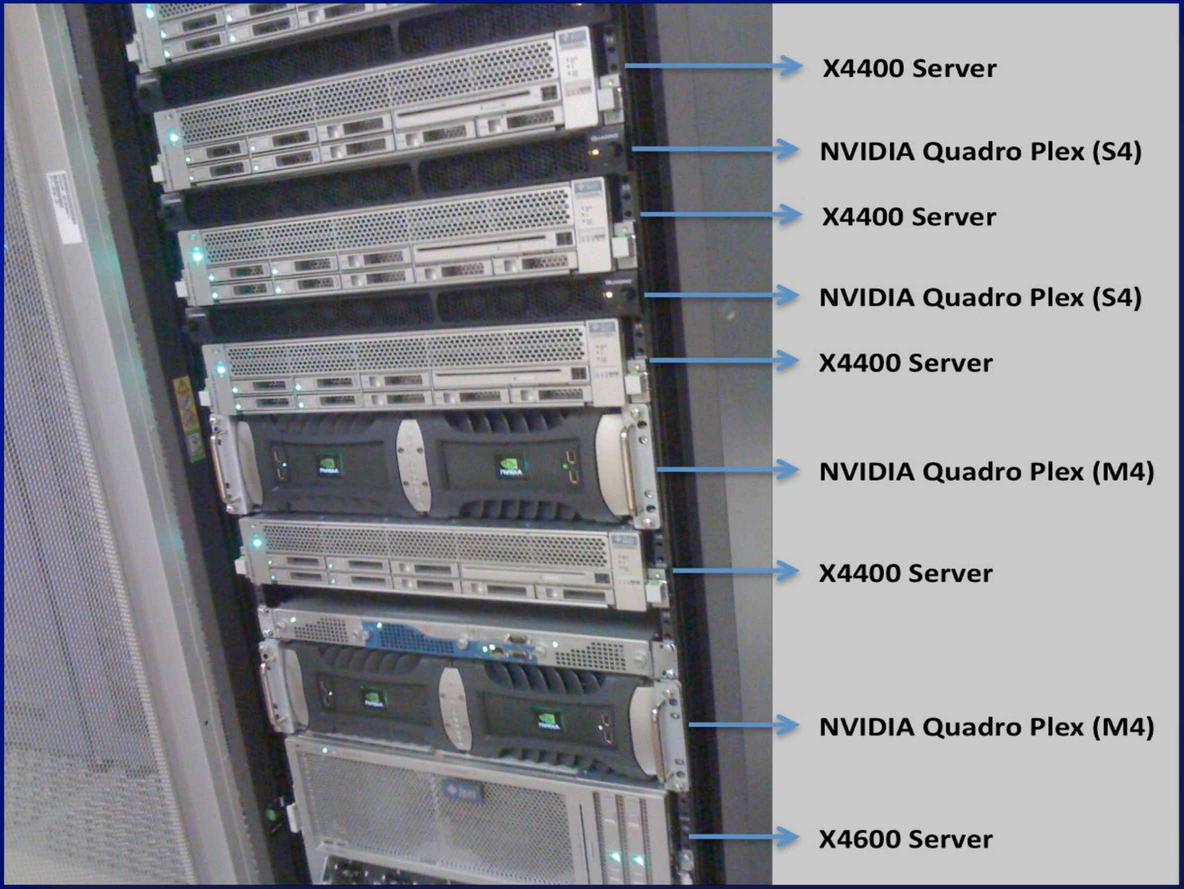
Large-Scale GPU Clusters

Spur

Remote, Interactive Visualization System Directly Connected to Ranger

128 cores, 1 TB aggregate memory, 32 GPUs

- spur.tacc.utexas.edu
- 1 fat memory node
  - Sun Fire X4600 server
  - 8 AMD Opteron dual-core CPUs @ 3 GHz
  - 256 GB memory
  - 4 NVIDIA FX5600 GPUs
- 7 other nodes
  - Sun Fire X4440 server
  - 4 AMD Opteron quad-core CPUs @ 2.3 GHz
  - 128 GB memory
  - 4 NVIDIA FX5600 GPUs



Longhorn

First NSF XD Visualization Resource

256 Nodes, 2048 Cores, 512 GPUs, 14.5 TB Memory

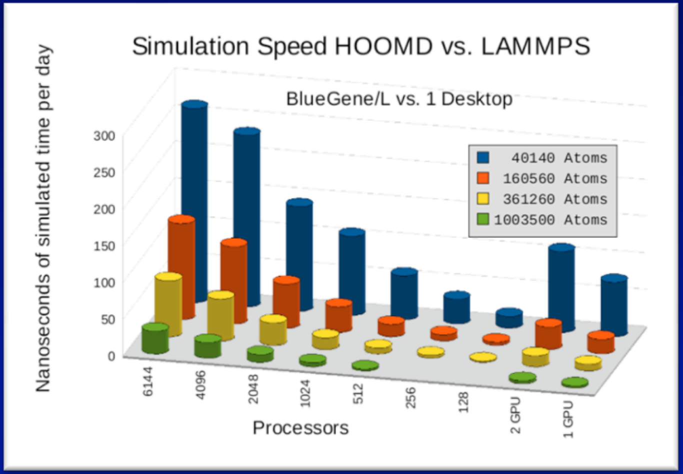
- 256 Dell Dual Socket, Quad Core Intel Nehalem Nodes
  - 240 with 48 GB shared memory/node (6 GB/core)
  - 16 with 144 GB shared memory/node (18 GB/core)
  - 73 GB Local Disk
  - 2 Nvidia GPUs/Node (FX 5800 – 4GB RAM)
- ~14.5 TB aggregate memory
- QDR InfiniBand Interconnect
- Direct Connection to Ranger's Lustre Parallel File System
- 10G Connection to 210 TB Local Lustre Parallel File System
- Jobs launched through SGE

Kelly Gaither (PI), Valerio Pascucci, Chuck Hansen, David Ebert, John Clyne (Co-PI), Hank Childs

GPU Aware MPI (GAMPI)

Paul Navratil, Don Fussell, Calvin Lin

- Developing a GPU-aware extension of MPI that provides access to clusters with GPUs.
- Minimal changes to existing MPI programs.
- Enables efficient communication and data movement across GPUs hosted in a distributed cluster.
- Preserves CPU-based MPI syntax so existing CPU-based codes will function without modification.
- Exposes an advanced user interface for micro-control of scheduling on the GPUs.



Benjamin G. Levine, David N. LeBard, Axel Kohlmeyer, and Michael L. Klein. Accelerating ab initio molecular dynamics and data analysis via high-performance architectures (http://cyberchem.ncsa.illinois.edu/projects/levine.html), 2010

TACC

Large-Scale Distributed GPU-Based Visualization Framework

Byungil Jeong, Greg Abram, Greg Johnson, Paul Navratil

- Enables efficient overlapping of GPU operations with I/O operations and sort-last image compositing to achieve high throughput, in-core rendering.
- Achieves an approximate speedup for 4.5x over CPU-based visualization methods on a 2048^3 scalar volume.

To Be Presented at NVIDIA GPU Tech Conference, September 2010

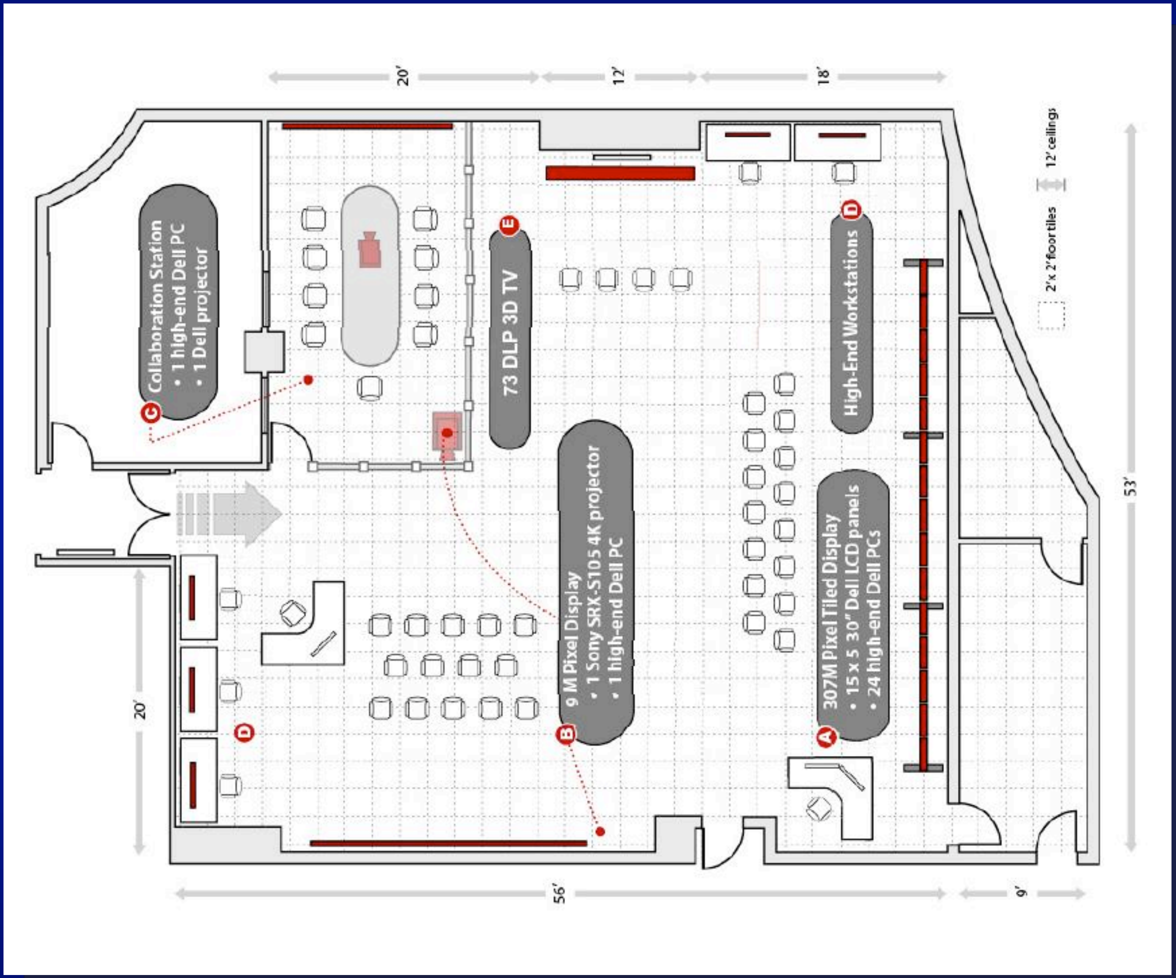
TACC

Large-Scale Tiled Displays

ACES Visualization Laboratory

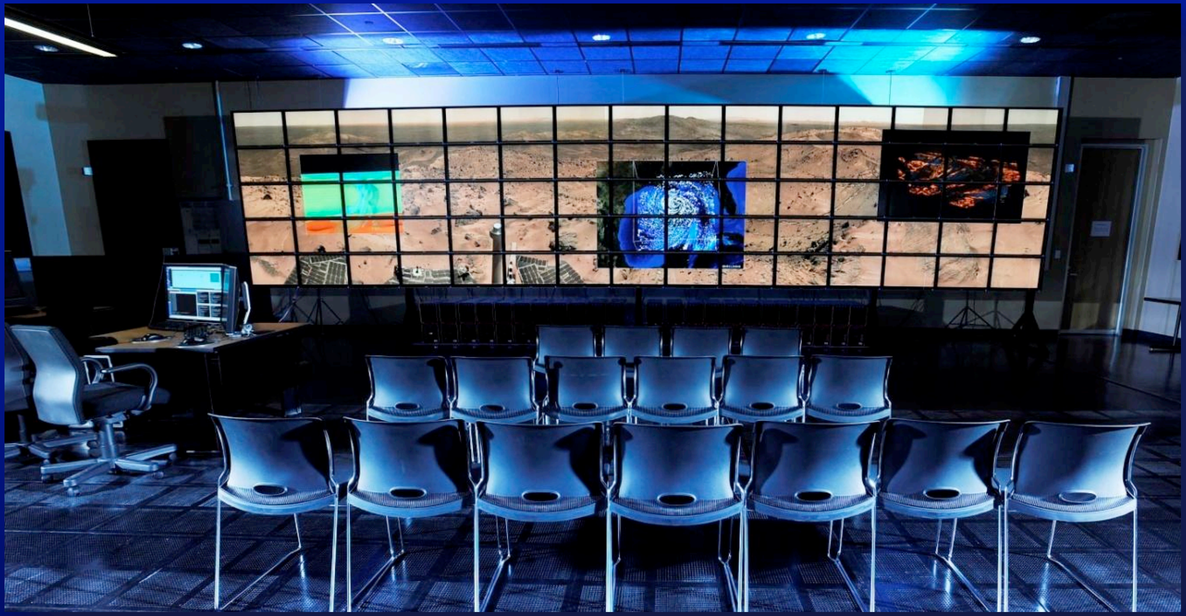
Campus Presence for Collaborative Visualization

- Multi-user space with reserveable resources.
- Seamless environments from laptops to large-scale displays.
- Provides large pixel count displays and a collaboration room.
- Reconfigurable, flexible environment that can be used in a variety of ways.



Stallion

- 15x5 tiled display of Dell 30-inch flat panel monitors
- 307M pixel resolution, 4.7:1 aspect ratio
- 100 processing cores with over 36GB of graphics memory and 108GB of system memory
- 6TB shared file system



Visualization of Southwest Power Grid (ERCOT and EEI) streamed from Pasadena, California to Stallion

Brandt Westing

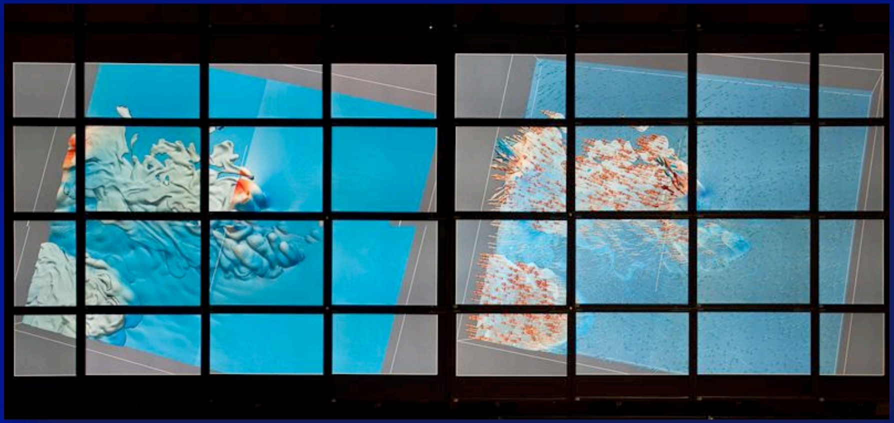
Downtown Austin @ 1 Gigapixel (77263 x 14225)

HDR Photograph taken by Ricardo Mileschi, Austin TX

Live Demonstration at SC09 Using Colt

Using SAGE for Remote Visualization of Large-Scale Data for Ultra-High Resolution Display Environments

Sungwon Nam, Byungil Jeong, Luc Renambot, Andrew Johnson, Kelly Gaither, Jason Leigh

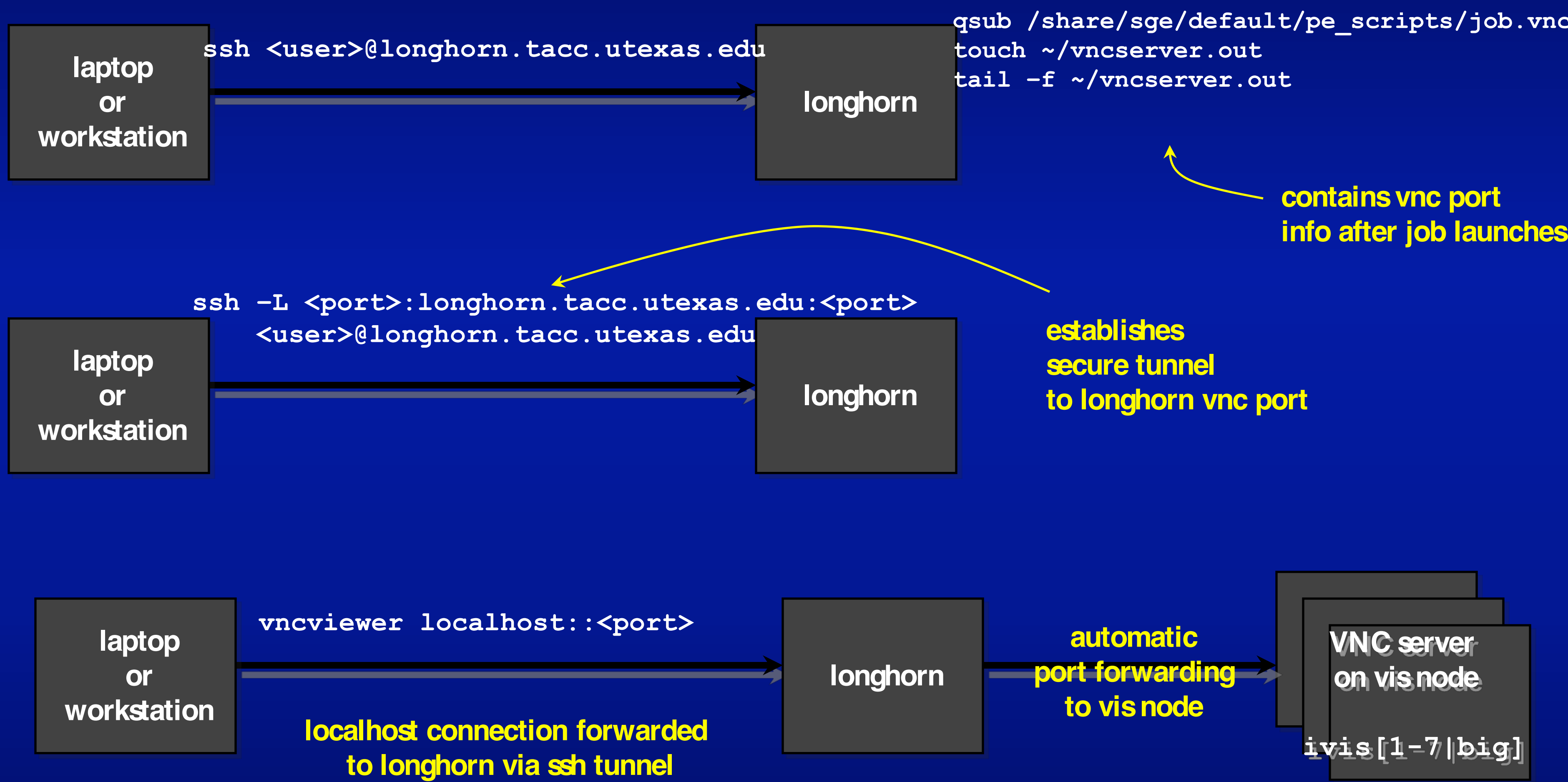


100 million pixel tiled display at EVL displaying two ParaView sessions. On the left, two local rendering nodes stream to the display. On the right, four remote rendering nodes stream from TACC.

- Integrates ParaView and SAGE.
- Enables ParaView and VTK-based applications to stream high-resolution visualizations from remote rendering servers to clients ranging from laptops to scalable tiled displays.

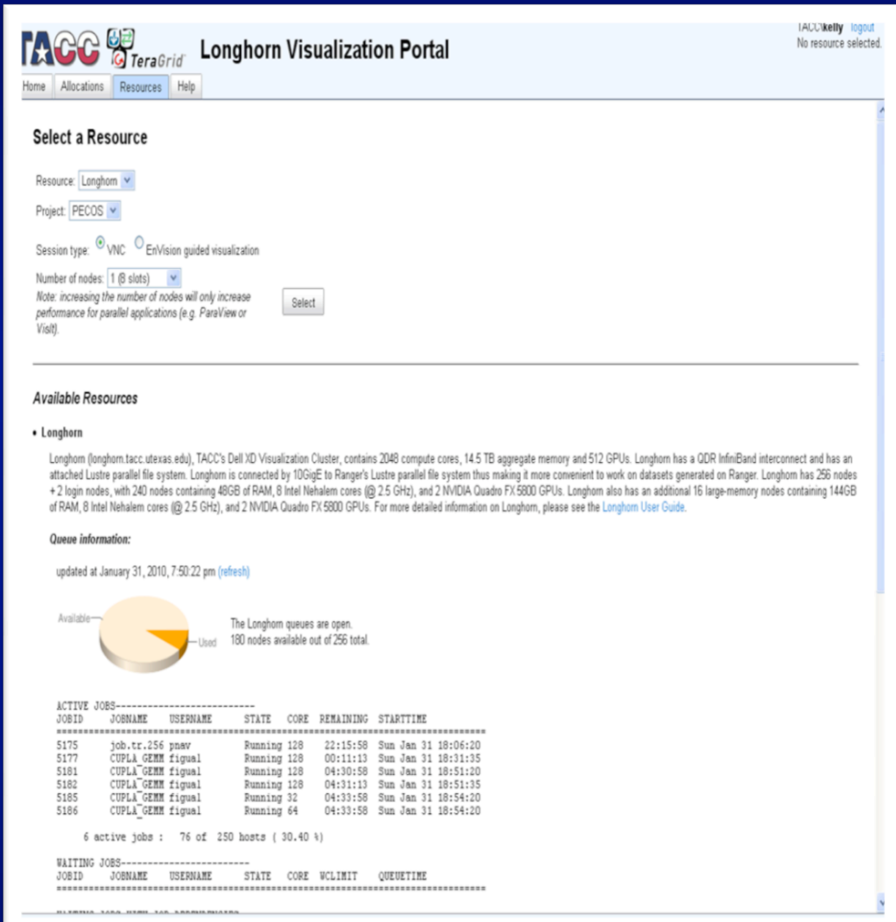
Remote and Collaborative Visualization

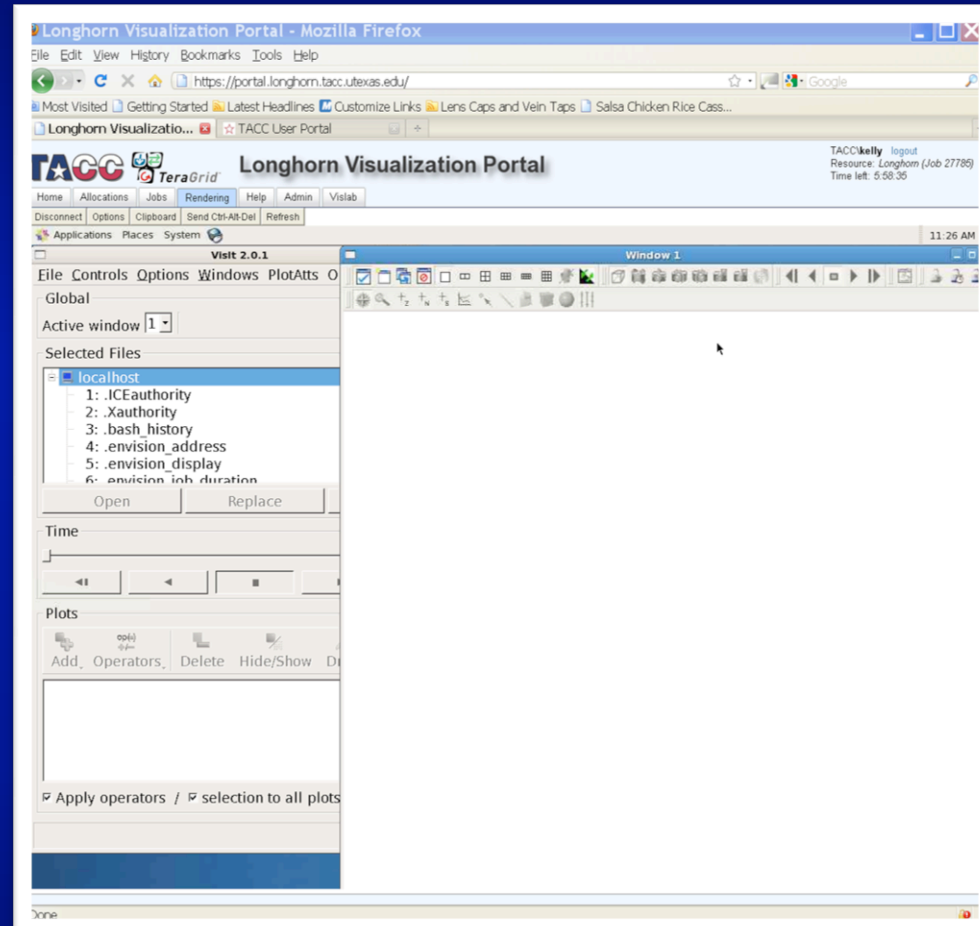
Connecting to Longhorn/Spur Using VNC



Longhorn Visualization Portal

portal.longhorn.tacc.utexas.edu



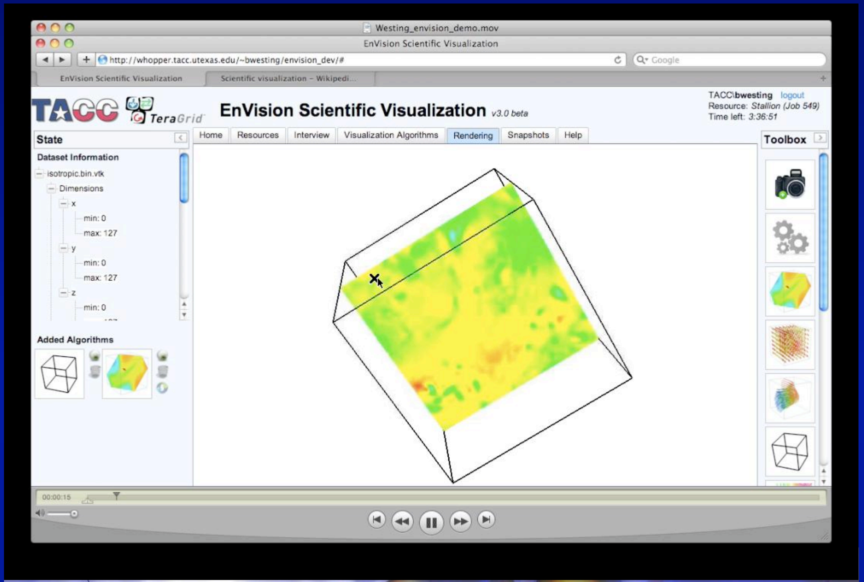


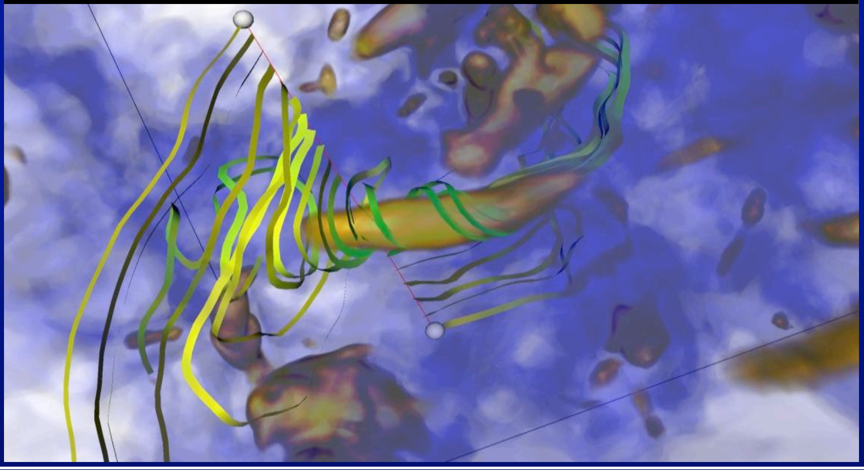
>3000 jobs submitted through the portal

EnVision

Greg Johnson, Brandt Westing

- Web-based visualization software that allows researchers to develop interactive visualizations intuitively.
- Currently integrated into the Longhorn Visualization Portal but can run independently.
- Began collaborations with ParaView team.



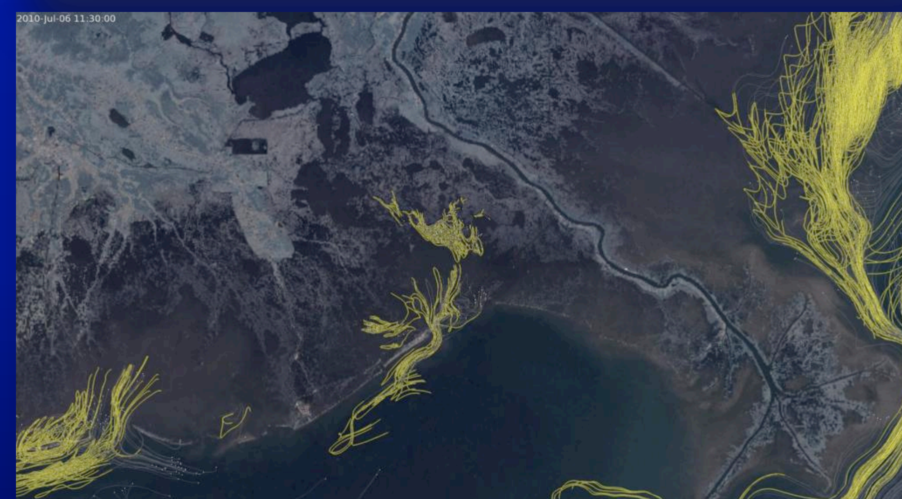


Large Data Visualization

Visualizing Oil Spill

Adam Kubach, Karla Vega, Clint Dawson

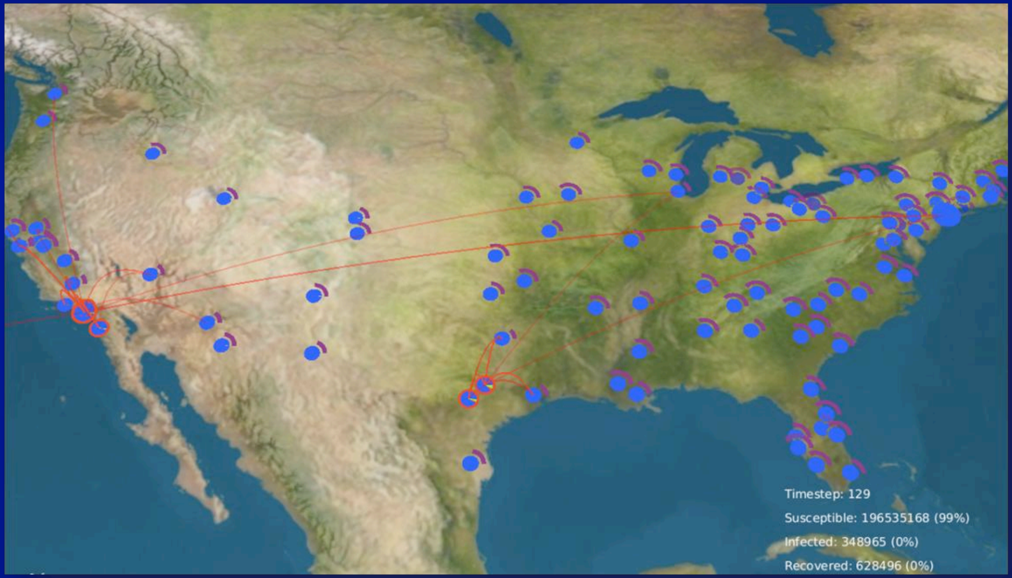
- Visualization focused on the overlay of particle movement and satellite or aerial imaging data.
- The particles in the visualization represent the oil spill and their position is either hypothetical or reflect the position of the oil on the surface.
- The data has been visualized using Longhorn and MINERVA, which is an open source geospatial software. The data is generated daily and is approximately 100 GB in size.



H1N1 Flu Outbreak Simulation

Greg Johnson, Brandt Westing, TACC; Ned Dimitrov, Lauren Meyers, UT Comp. Bio

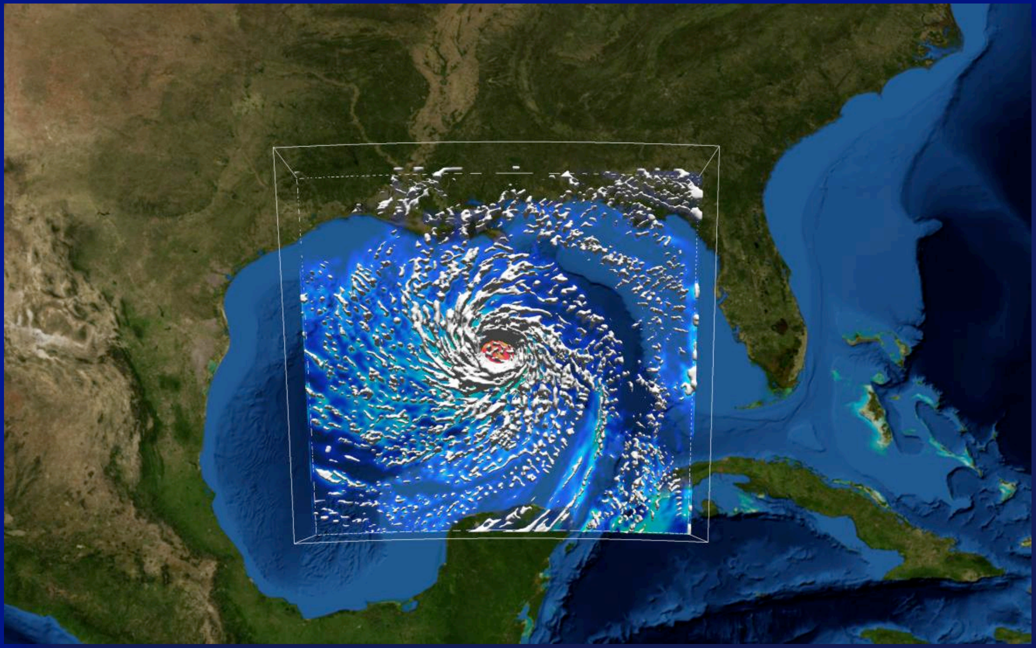
- Visualization of a swine flu epidemic spreading throughout North America.
- Epidemic begins in Mexico City.
- Visualization classifies individuals into three groups: susceptible (blue), infected (red), and recovered (green). Available antivirals are shown in purple.
- Cities and transportation links are highlighted in red to indicate large numbers of infected individuals and infectious travelers.



Visualization of Hurricane Ike

Greg Johnson, Romy Schneider, John Cazes, Karl Schulz, Bill Barth, TACC; Frank Marks, NOAA; Fuqing Zheng, University of Pennsylvania; Yonghui Weng, Texas A&M.

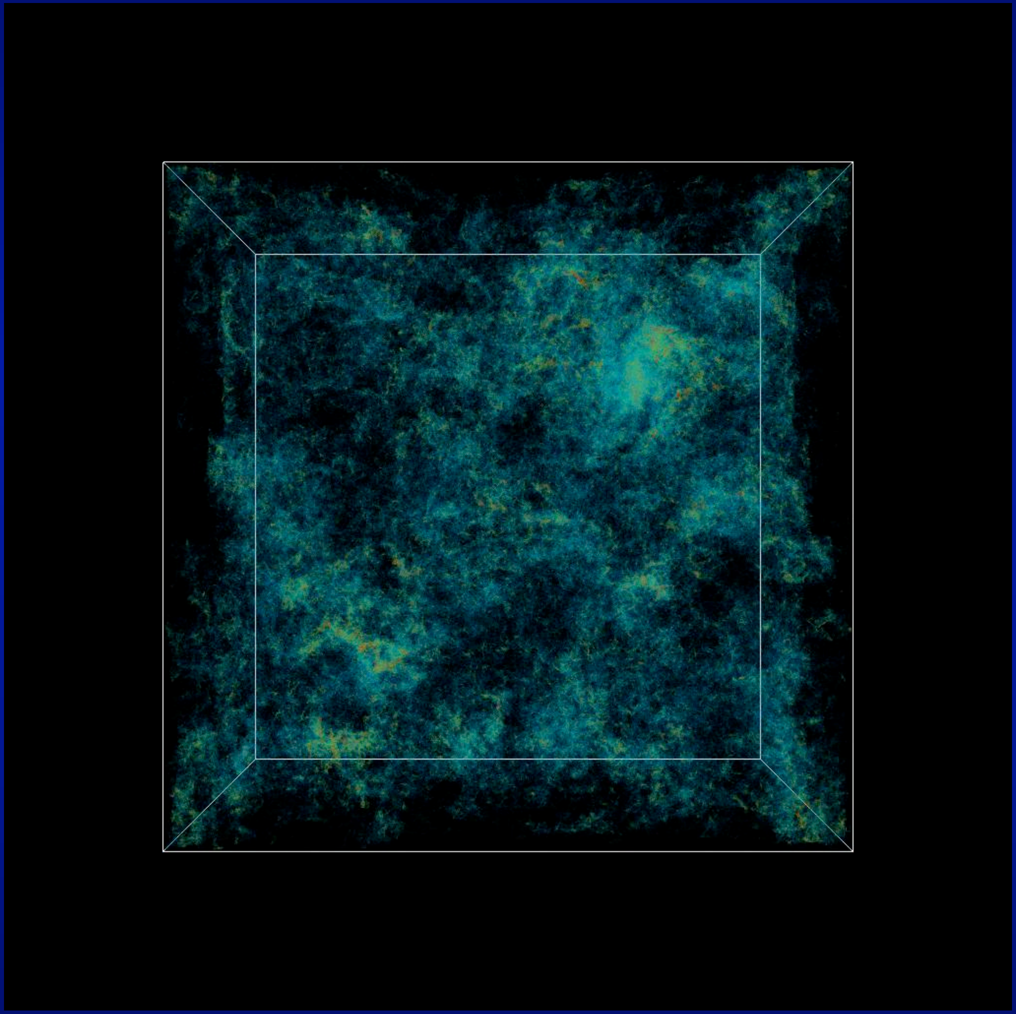
- Throughout the 2008 hurricane season, the TACC was an active participant in a NOAA research effort to develop next-generation hurricane models.
- Using up to 40,000 processing cores at once, researchers simulated both global and regional weather models and received on-demand access to Ranger.
- Visualization of Hurricane Ike shows the storm developing in the gulf and making landfall on the Texas coast.



Visualization of Large Scale Turbulent Flow

Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

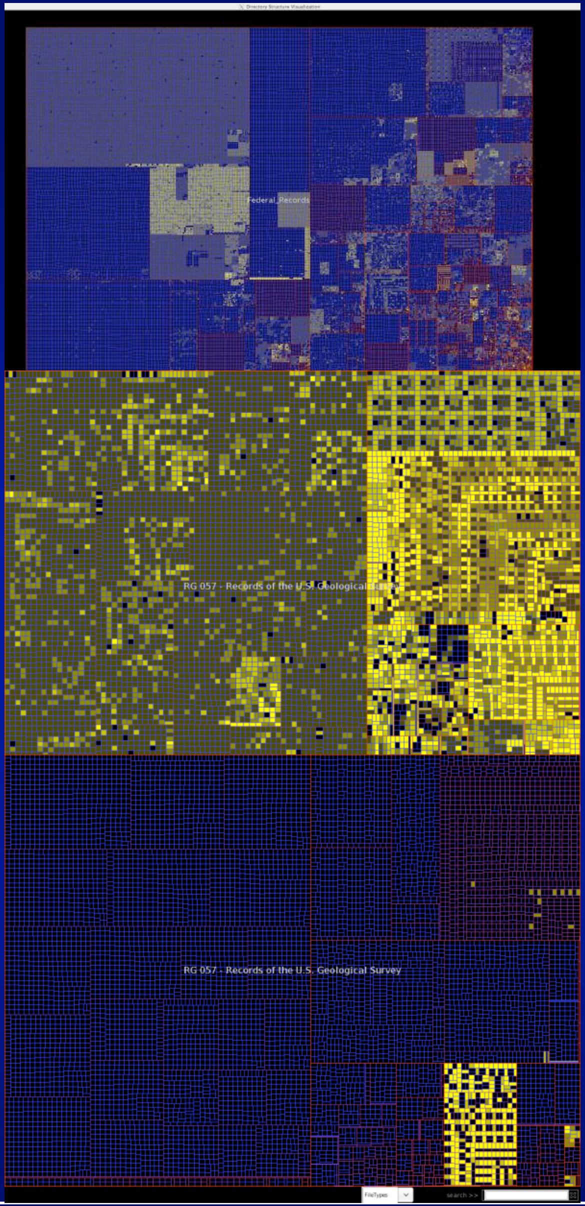
- Remote interactive visualization of 17 time-steps of the largest turbulent flow simulation computed to date.
- First time this had been visualized interactively at this scale (4096^3).



Data Mining & Feature Detection

National Archives and Records Administration (NARA) Records Visualization

Maria Esteva, Weijia Xu

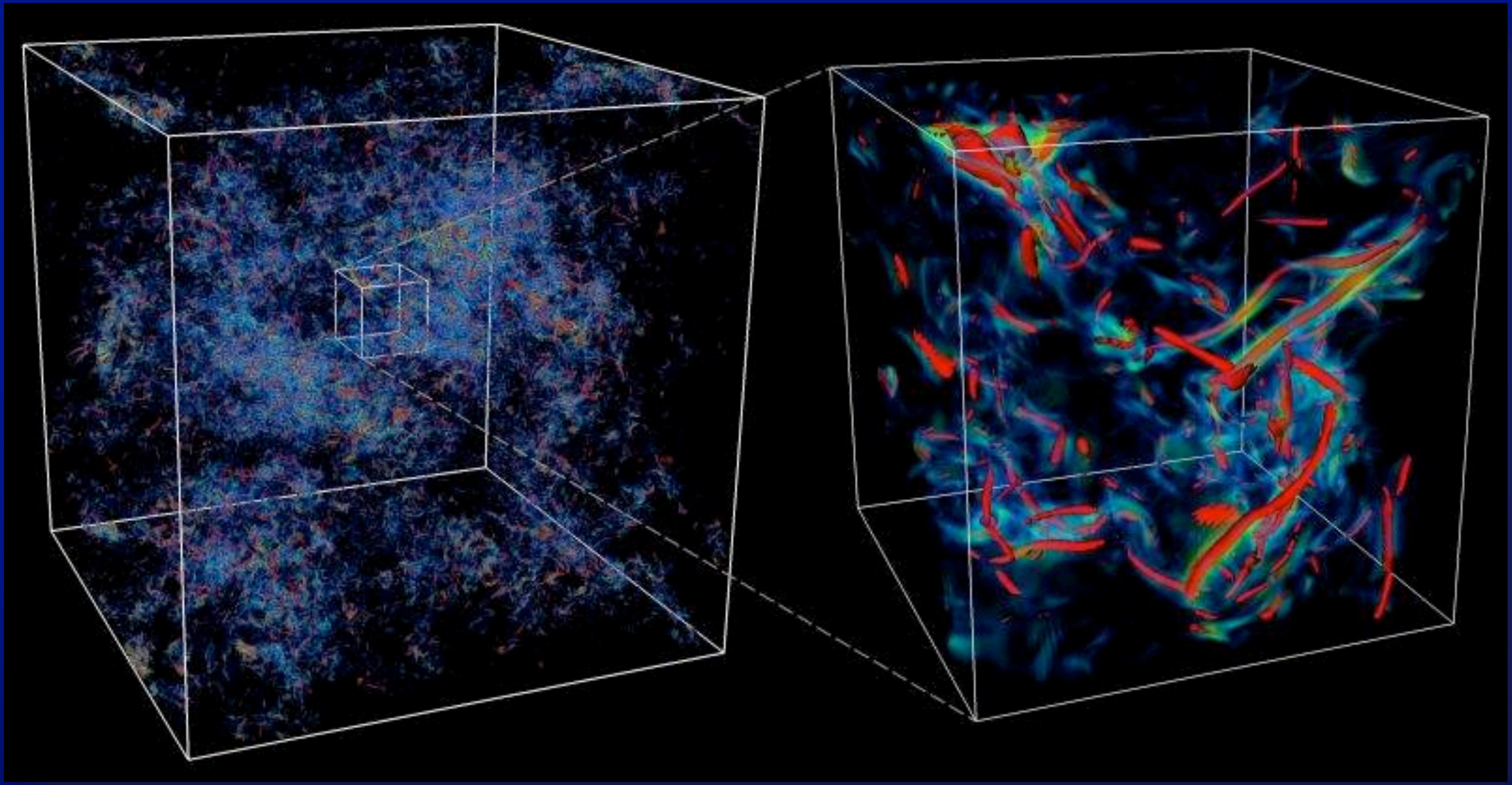


- NARA is charged with ensuring continuous access to government records.
- Digital archival collections are diverse in nature, presenting multiple media types organized in diverse arrangements that serve the purposes of the many authors, software, and hardware involved in their creation.
- To preserve and provide access to electronic records collections, archivists need first to conduct a series of analysis with purposes of discovering their structure and content, and to make decisions about their long-term preservation needs.
- This research examines information visualization for archival analysis and long-term preservation planning of terabyte size collections.

Visualization and Analysis of Large Scale Turbulent Flow

Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

- Remote interactive visualization of 17 time-steps of DNS turbulence in a box computed by P.K. Yeung.

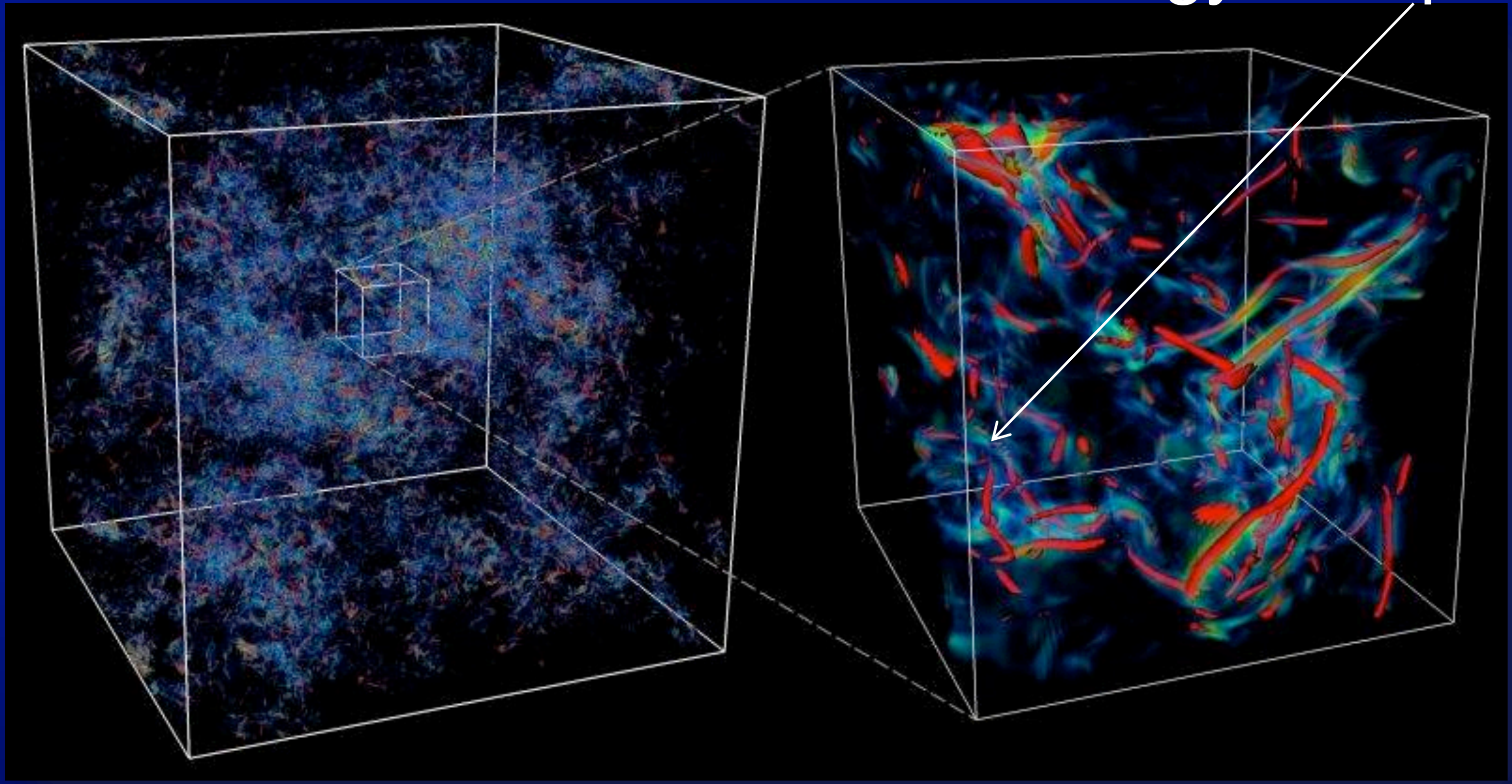


Visualization and Analysis of Large Scale Turbulent Flow

Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

- Remote interactive visualization of 17 time-steps of DNS turbulence in a box computed by P.K. Yeung.

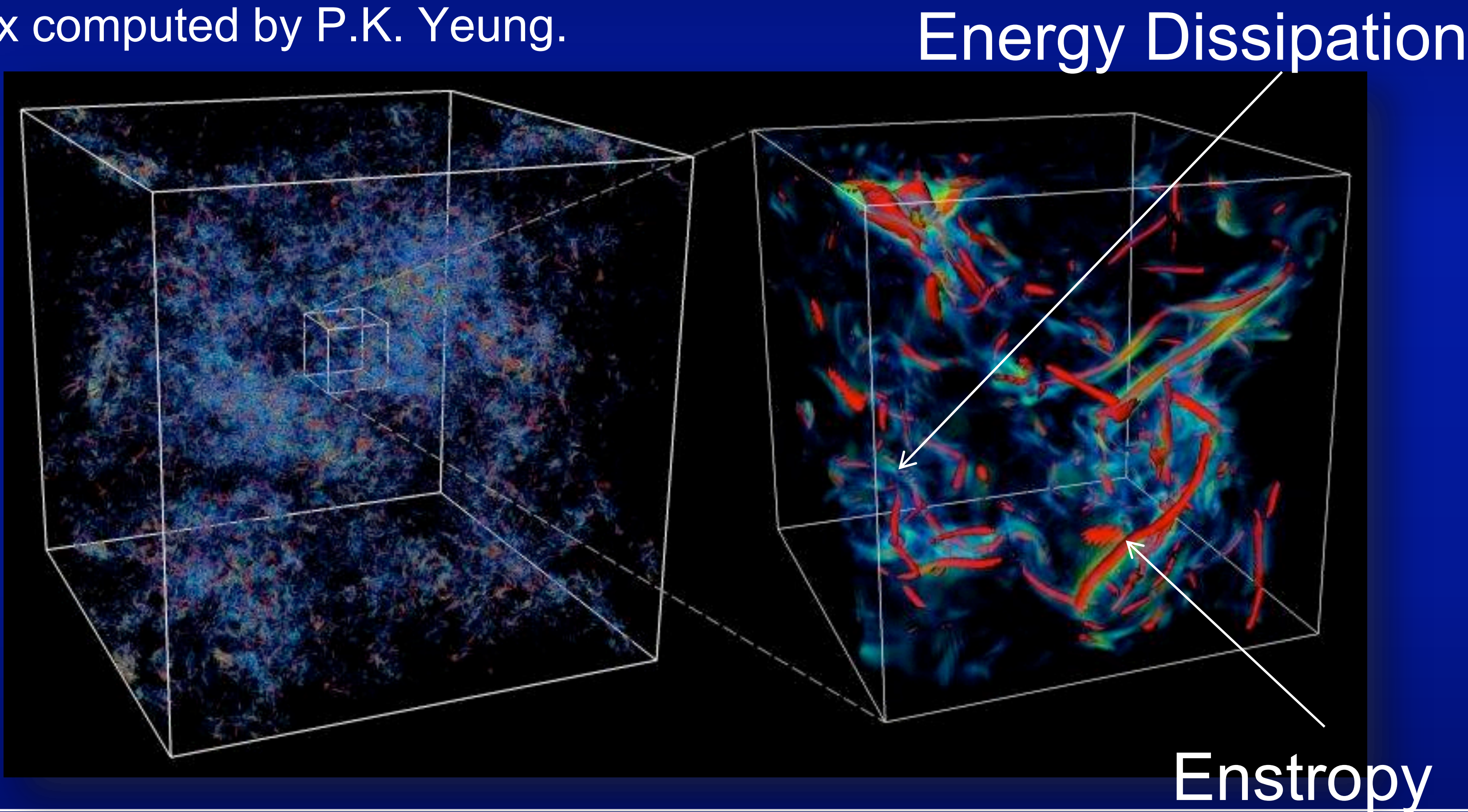
Energy Dissipation



Visualization and Analysis of Large Scale Turbulent Flow

Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

- Remote interactive visualization of 17 time-steps of DNS turbulence in a box computed by P.K. Yeung.



Visualization and Analysis of Large Scale Turbulent Flow

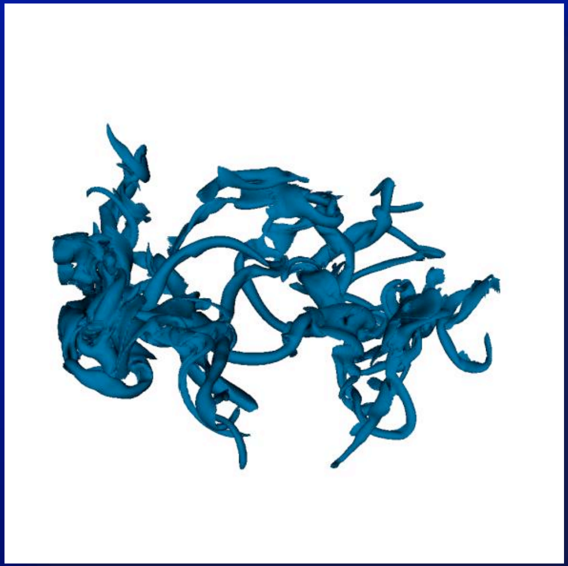
Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

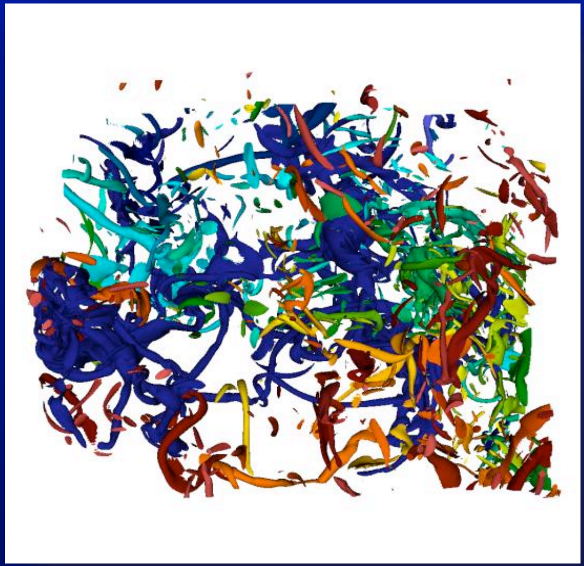
- Central Science Questions:
  - Can we identify intermittency in the field (very intense, but localized events in space)?
  - What is the shape of these tubes of enstrophy?
  - Can we characterize what is happening with dissipation in the areas surrounding these tubes of enstrophy?
  - How do these tubes of enstrophy behave over time?

Visualization and Analysis of Large Scale Turbulent Flow

Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

- Step 1: For each timestep, compute the isovolumes created in a range of isovalues.
- Step 2: For each timestep, create connected components of these isovolumes that fall within a specified threshold

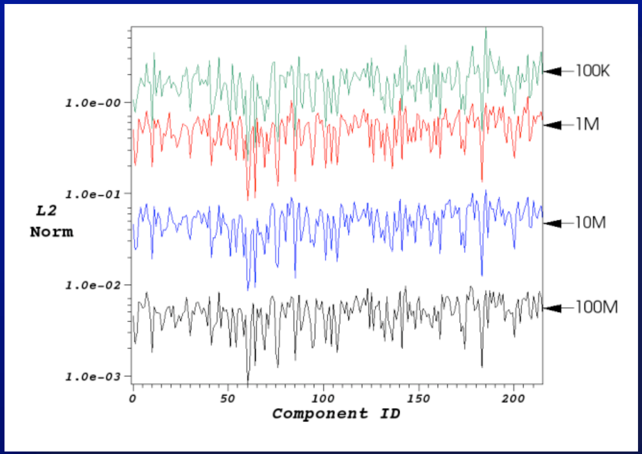




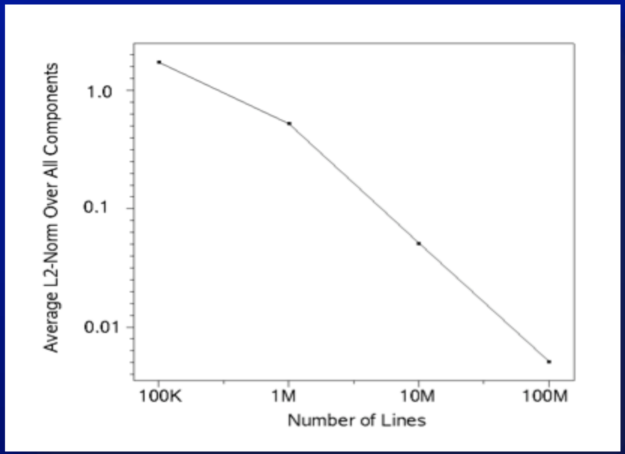
Visualization and Analysis of Large Scale Turbulent Flow

Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

- Step 3: For each connected component within a specified threshold, compute the chord length distribution
  - Construct a set of uniform density, random lines
  - Calculate the intersection of these lines with the shape
  - Re-distribute the segments
  - Calculate the chord length
  - Collect Results



L2-norm between chord length distributions for each component at a single time slice of the data set.

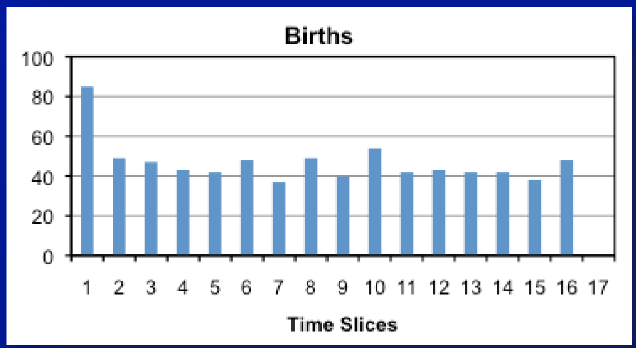


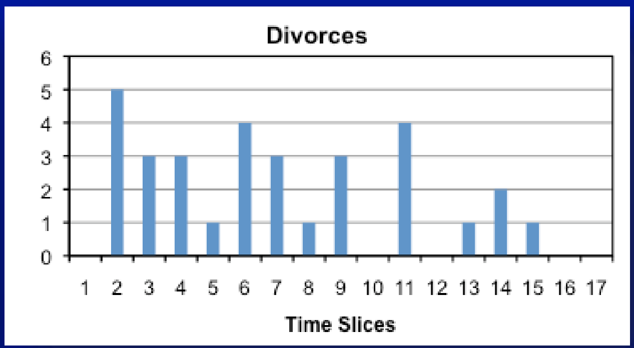
Average difference in L2-norm over all components at a single time slice of the data set.

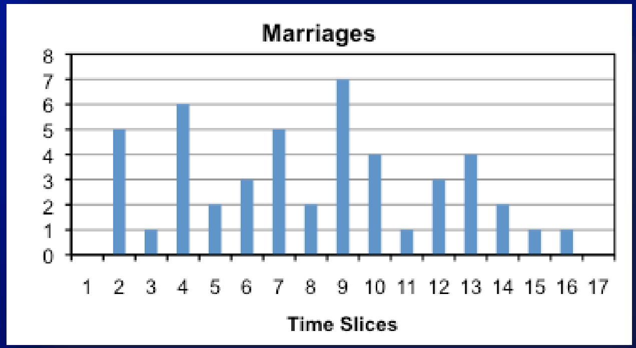
Visualization and Analysis of Large Scale Turbulent Flow

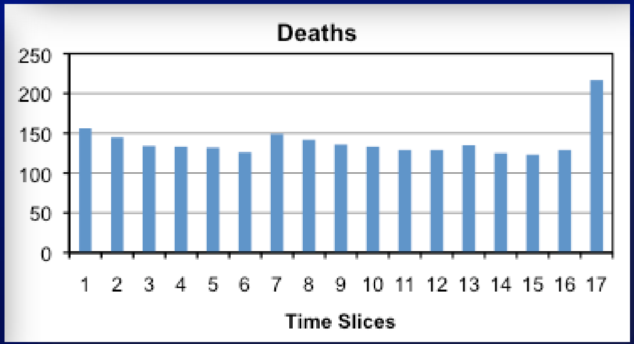
Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

- Step 4: Using the connected components and the chord length distribution, determine behavior over time
  - Births, Marriages, Divorces, Deaths





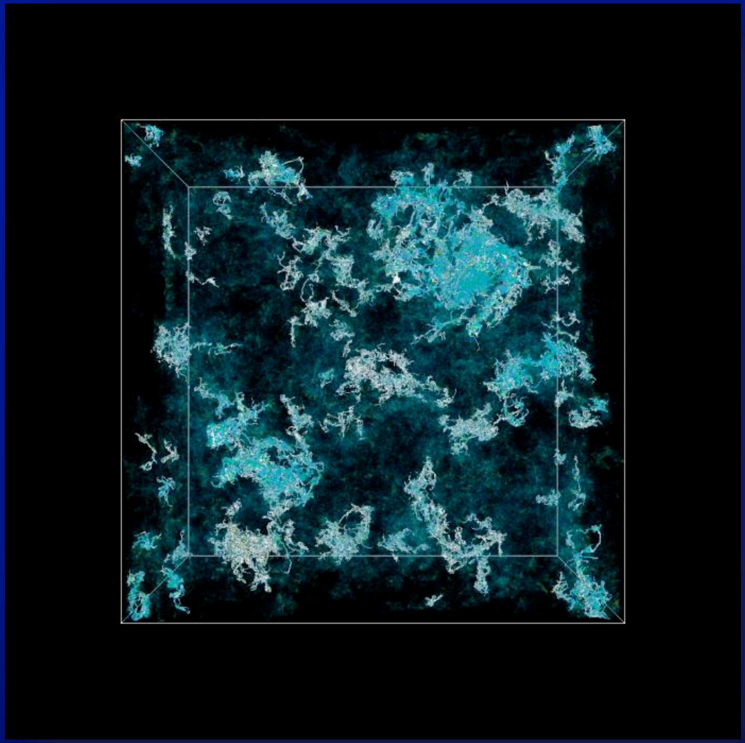


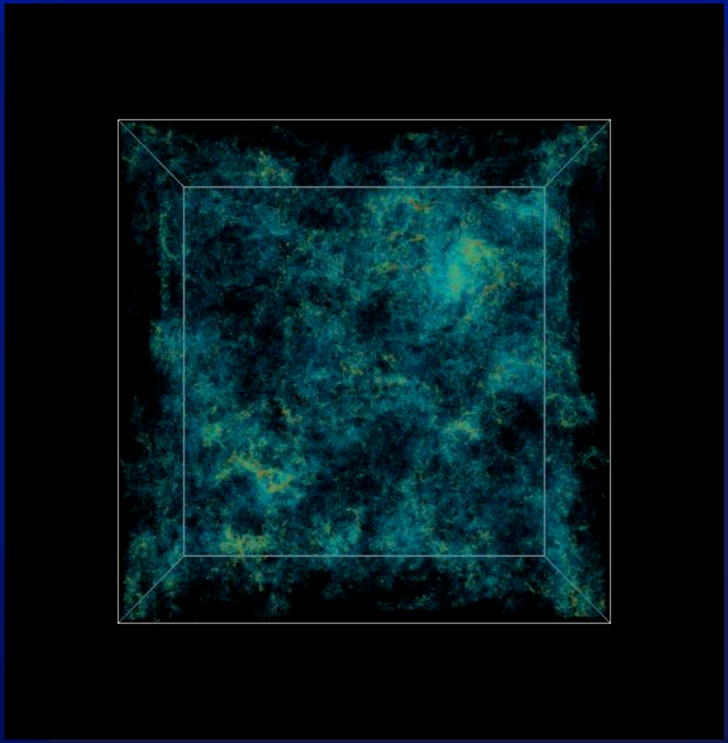


Visualization and Analysis of Large Scale Turbulent Flow

Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

- Step 5: Isolate and Track Connected Components, Births, Marriages, Divorces, and Deaths





Visualization and Analysis of Large Scale Turbulent Flow

Kelly Gaither, Hank Childs, Greg Johnson, Karl Schulz, Cyrus Harrison, Diego Donzis, Texas A&M; P.K. Yeung, Georgia Tech

- Work to Date:
  - Visualized and analyzed the largest turbulent flow simulation to date – 4096^3 (64B Cells * 17 Time Steps = 1 Trillion Cells)
  - Developed shape characterization methods using chord length distributions
  - Developed novel applications of chord length distributions to inform and compare shape
  - Developed automatic feature detection, extraction and classification methods

Questions?

kelly@tacc.utexas.edu