
# Challenges of a Terascale Turbulence Simulation

*Michael E. Papka*

*Argonne National Laboratory / The University of Chicago*

# *Outline*

- Simulation
- Challenges
  - Data
  - Visualization
  - Communication
- Summary and moving forward

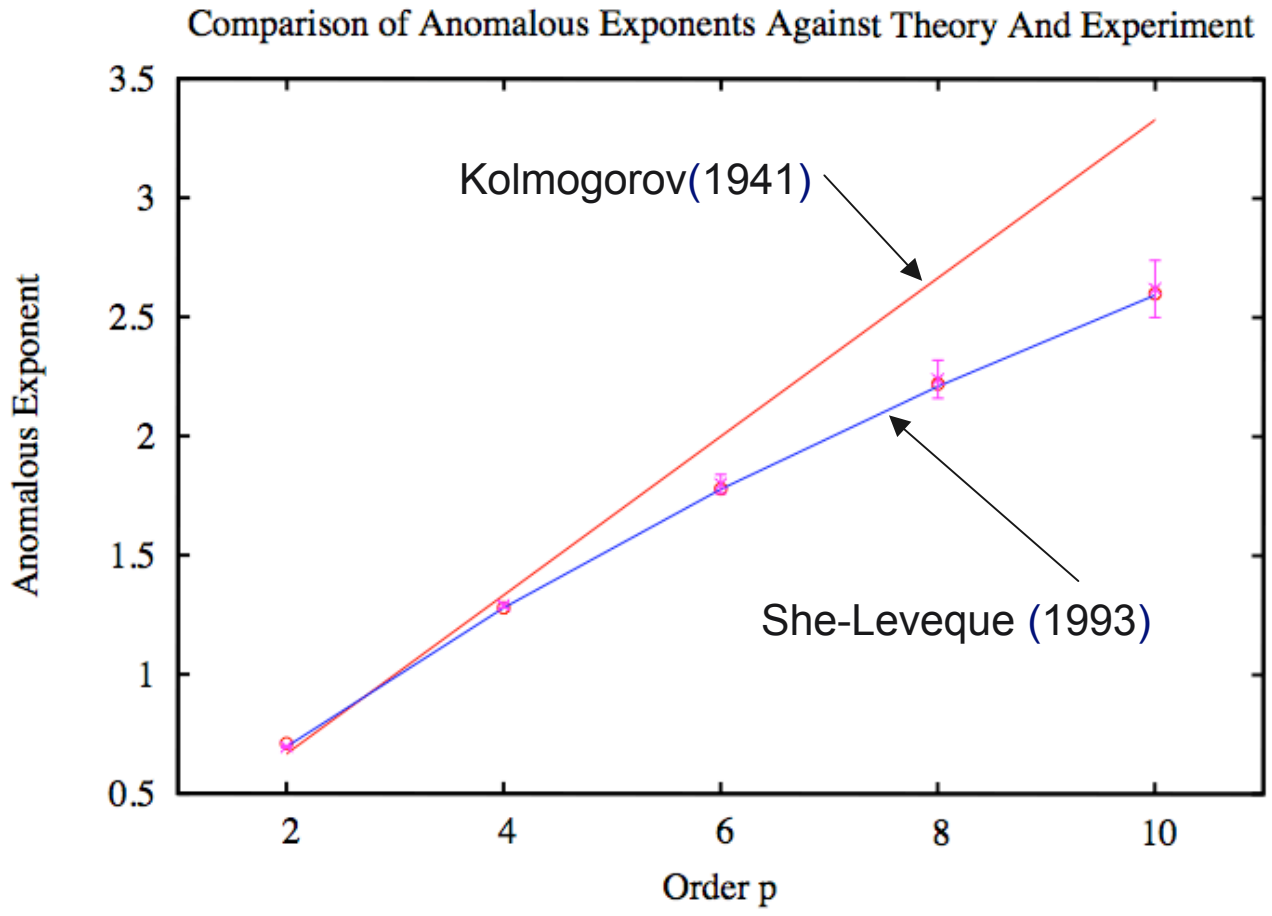
# *Simulation Introduction*

- Simulation done by ASC / Alliance Center Astrophysical Thermonuclear Flashes at The University of Chicago
- Ran on the LLNL BG/L machine during December 2005 - January 2006
- Produced 14TB of analysis data
  - 13.3 TB grid data (1 vector, 3 scalar)
  - 0.7 TB particle data (2 vectors, id)
- Produced 150TB+ of checkpoint/restart files
- Each time-step produced 32K files

# *Simulation Specifics*

- Homogeneous, isotropic compressible turbulence run with Lagrangian tracers on BG/L using FLASH 3
- Computation Size
  - $1856^3$ base grid size ($928^3$ used for analysis files)
  - $256^3$ Lagrangian tracer particles
- Computation Specifics
  - MILES-based approach solving Euler equations using PPM
  - Driven using stochastic driver (Eswaran & Pope, 1988)
  - 3D turbulent RMS Mach number = 0.3 (1D = .17) in steady-state
  - $Re_\lambda$ ~ 500 - 1000
  - Full eddy-turnover time in steady-state
- Roughly one week wall clock on 65,536 processors in CO mode

# Anomalous Scaling Exponents




Circles = Experiment
(Benzi et al, 1993)
Crosses = FLASH

# *Challenges*

- Data
  - Data transfer
  - Data storage
  - Data integrity
- Visualization
  - Real-time analysis
  - Filtering
- Communication
  - Effective communication with scientific team

# *Data Transfer*

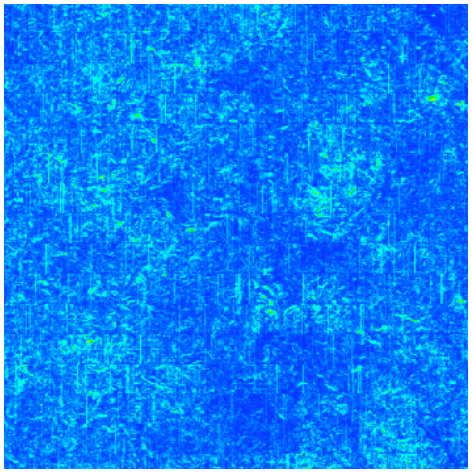
- Tarred directories to transfer as a single 18GB chunk
- GridFTP from LLNL to UC
  - Data got moved to HPSS before all could be transferred so added additional work in transfer
  - Expect scripts to manage transfers from HPSS to scratch space
  - Python scripts to manage transfers from scratch to UC
  - Screen sessions to manage overall effort
- 28 days to get 13.3 TB (grid data) to UC

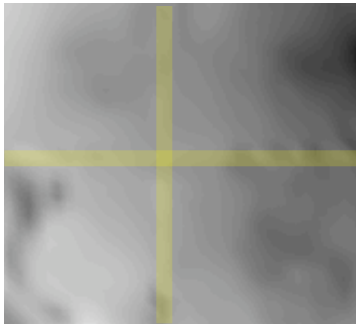
# *Data Storage*

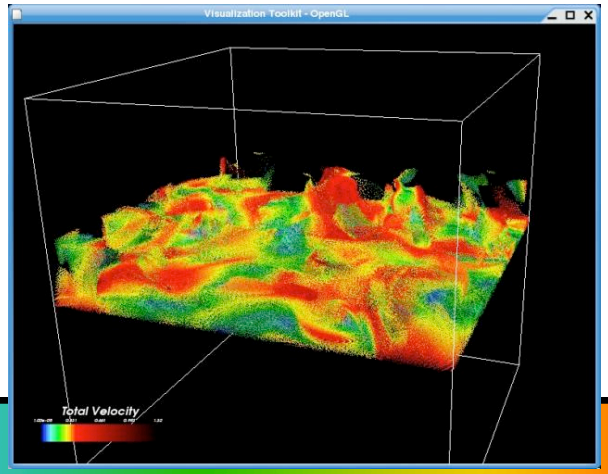
- How do you keep 14TB around?
- Represents 28% of all data the center has on disk
- Spread across multiple different volumes
- Augmented data compounds issue

# *Data Integrity*

- Whose fault is it?
- Enstrophy calculation problems
  - Calculate vorticity from analysis data
- Possible block boundary issues
  - Verify using different tools
  - Ghostcell issues
- Particle ids
  - Corrected midway through run
  - Track reconstruction





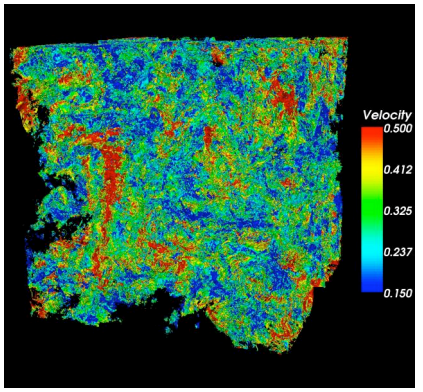


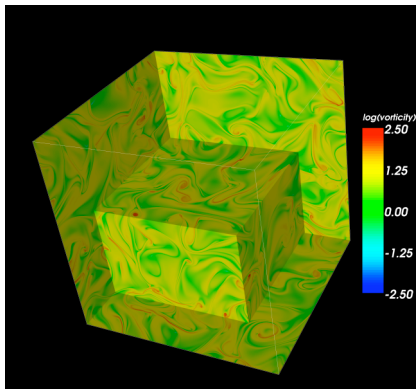
# *Visualization Tools*

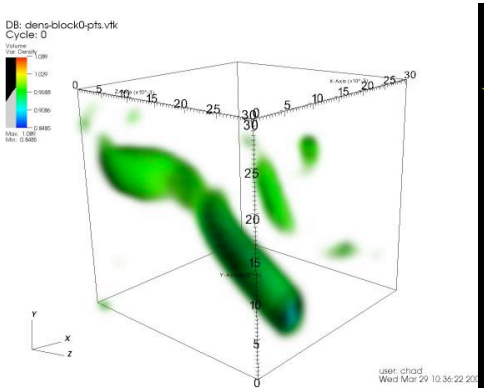
- Community tools
  - ParaView
  - VisIt
  - POVRay
- Group developed tools
  - Volume rendering
  - Particle rendering (built with vtk)

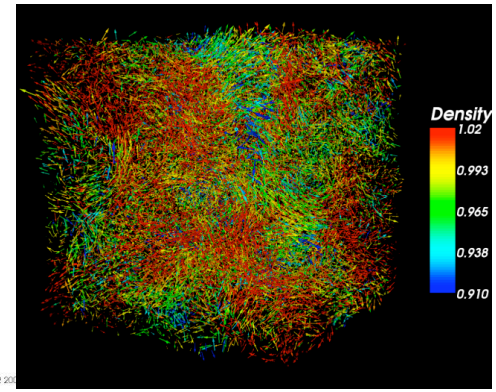
# *ParaView and VisIt*

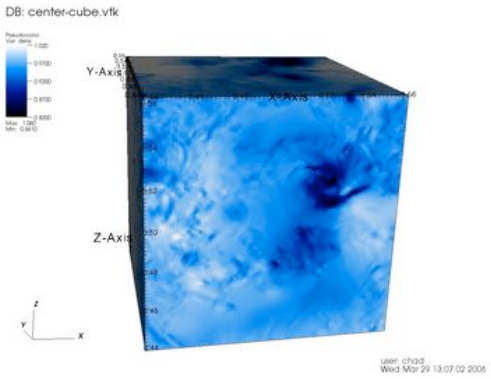
- In use by both visualization team and scientist
- Addresses real-time analysis need
  - Parallel capabilities
  - Hides complexity
  - Supports additional computation



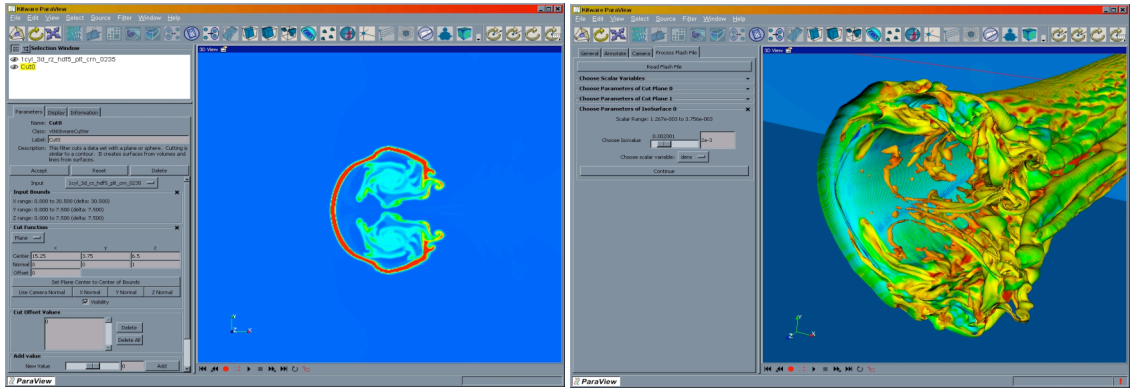



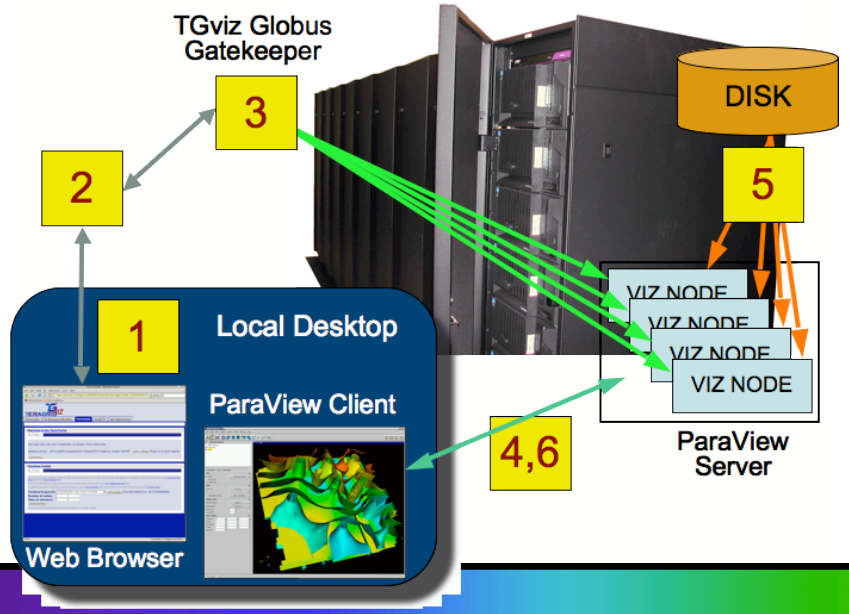



# ParaView Efforts

- Simplified interface

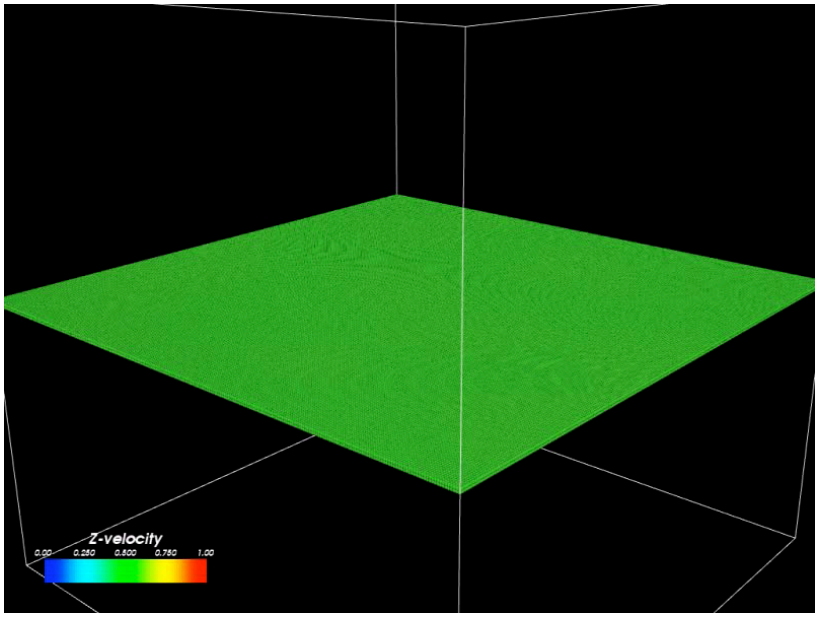


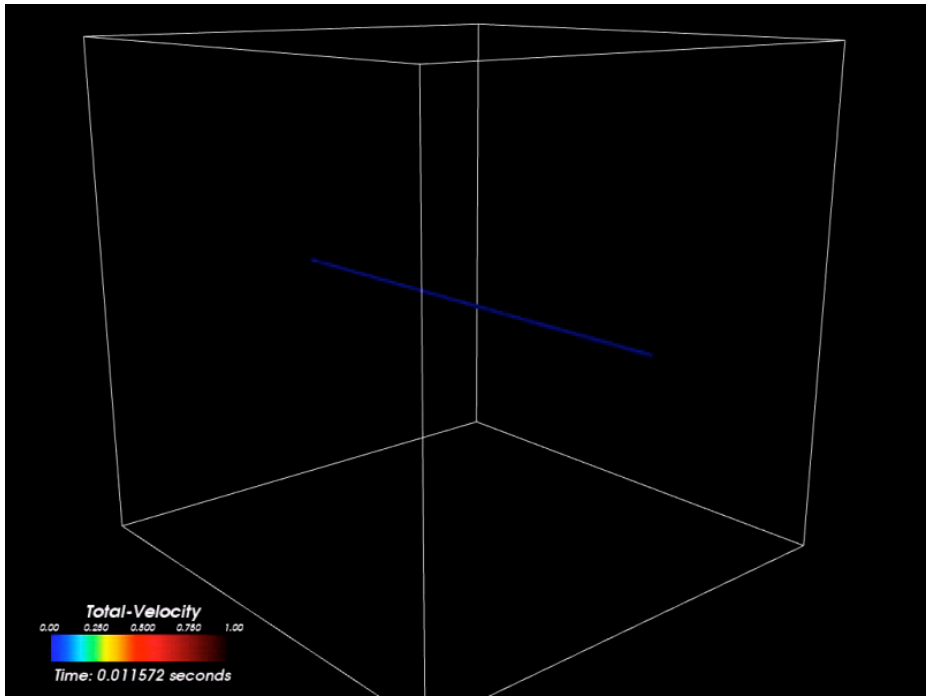
- TeraGrid visualization gateway

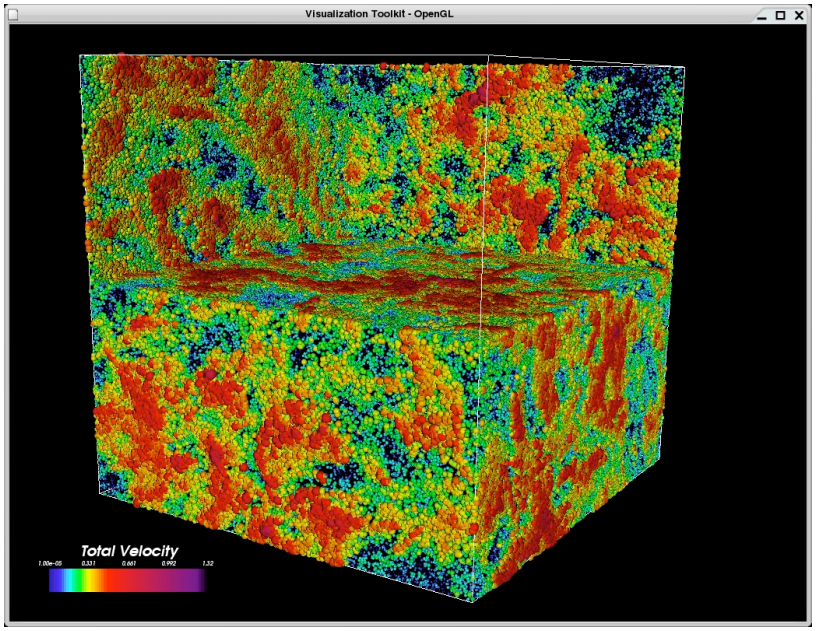


# Particle Rendering

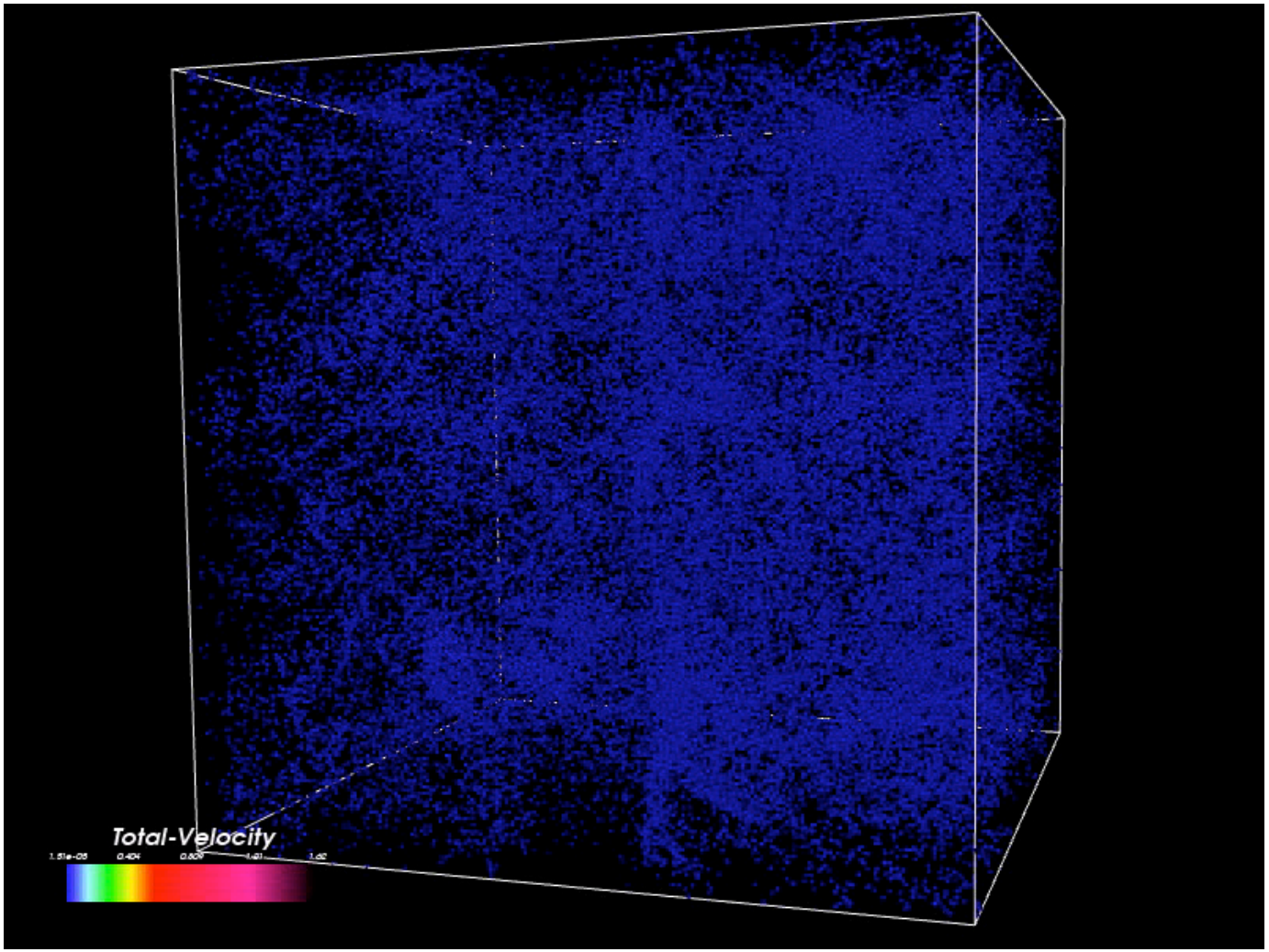
- Filtering of data
  - Geometric extraction
  - Data cuts



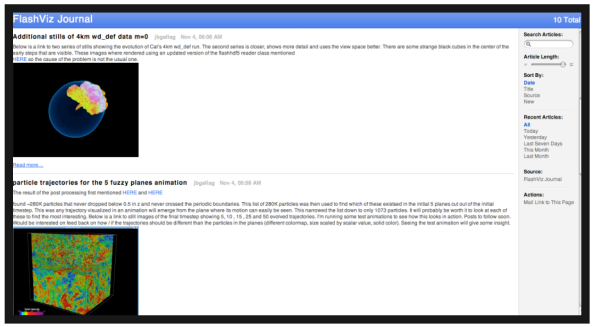



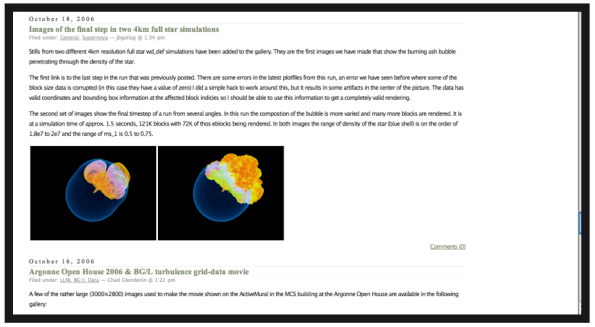
# Particle Visualization with Data Filter

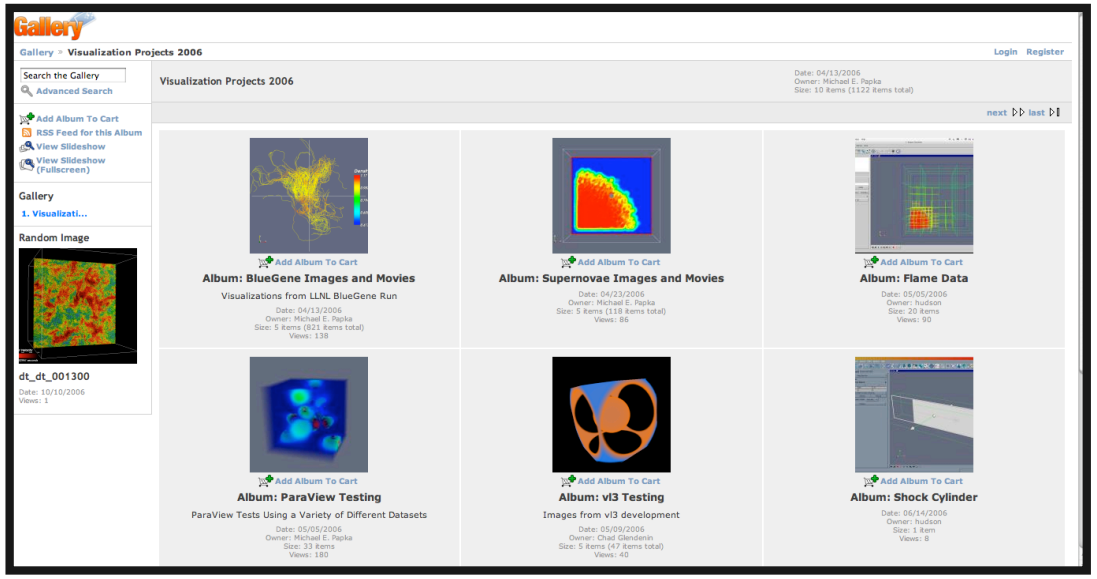


# Communication Challenges

- Keeping scientist updated
- Informing fellow team members
- Organizing results



# *Challenges - Revisited*

- Visualization component is only a fraction of the challenge
  - Usability
  - Simplification
- Data issues dominate the process
  - Location
  - State
- Communication
  - Process moving

# *Moving Forward*

`Turbulence is the most important unsolved problem of classical physics.' - Richard Feynman

- Looking at ways to make data publicly available
- Exploring integration of workflows into the process

# Acknowledgements


- Flash Center: R.Fisher, S. Abarzhi, K. Antypas, S. M. Asida, A. C. Calder, F. Cattaneo, P. Constantin, A. Dubey, I. Foster, J. B. Gallagher, M.K. Ganapathy, C. C. Glendenin, L. Kadanoff, D.Q. Lamb, S. Needham, T. Plewa, L.B. Reid, P. Rich, K. Riley, D. Sheeler, R. Hudson
- TeraGrid: J. Insley, T. Leggett, C. Lueninghoener, J. P. Navarro
- Funding:
  - DOE via the ASC/Alliance Program
  - This work was supported in part by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U.S. Dept. of Energy, under Contract DE-AC02-06CH11357.
  - This work was supported in part by the National Science Foundation under Grant No. NSF OCI 05-04086.