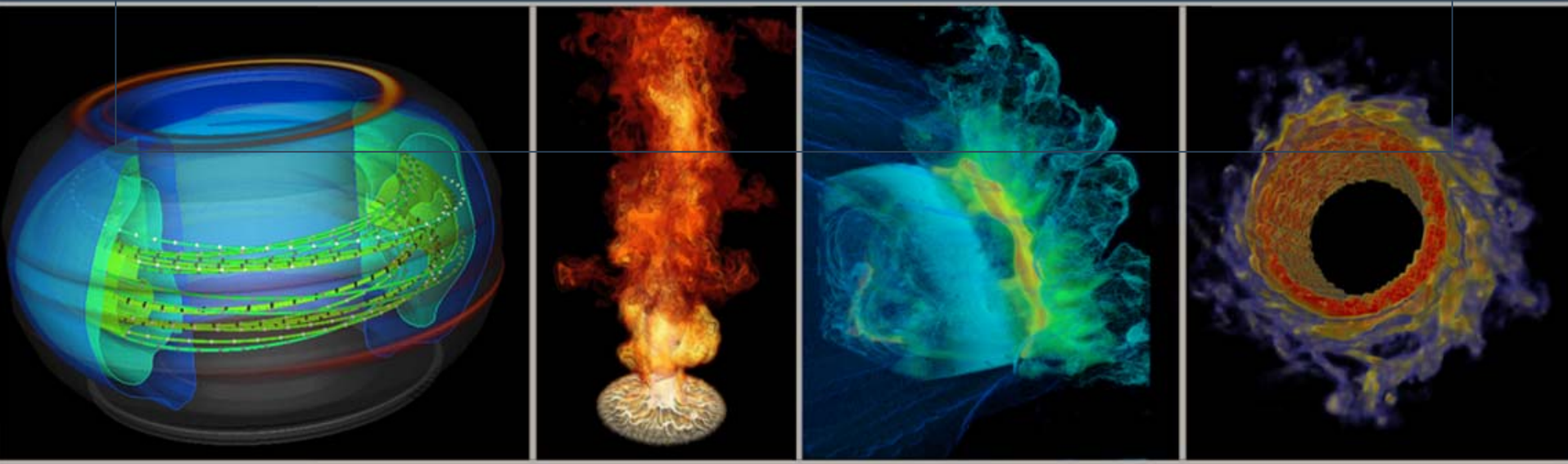
VACET

DOE SciDAC VISUALIZATION AND ANALYTICS CENTER FOR ENABLING TECHNOLOGIES



● IDAV/UC Davis ● LBNL ● LLNL ● ORNL ● SCI/Utah

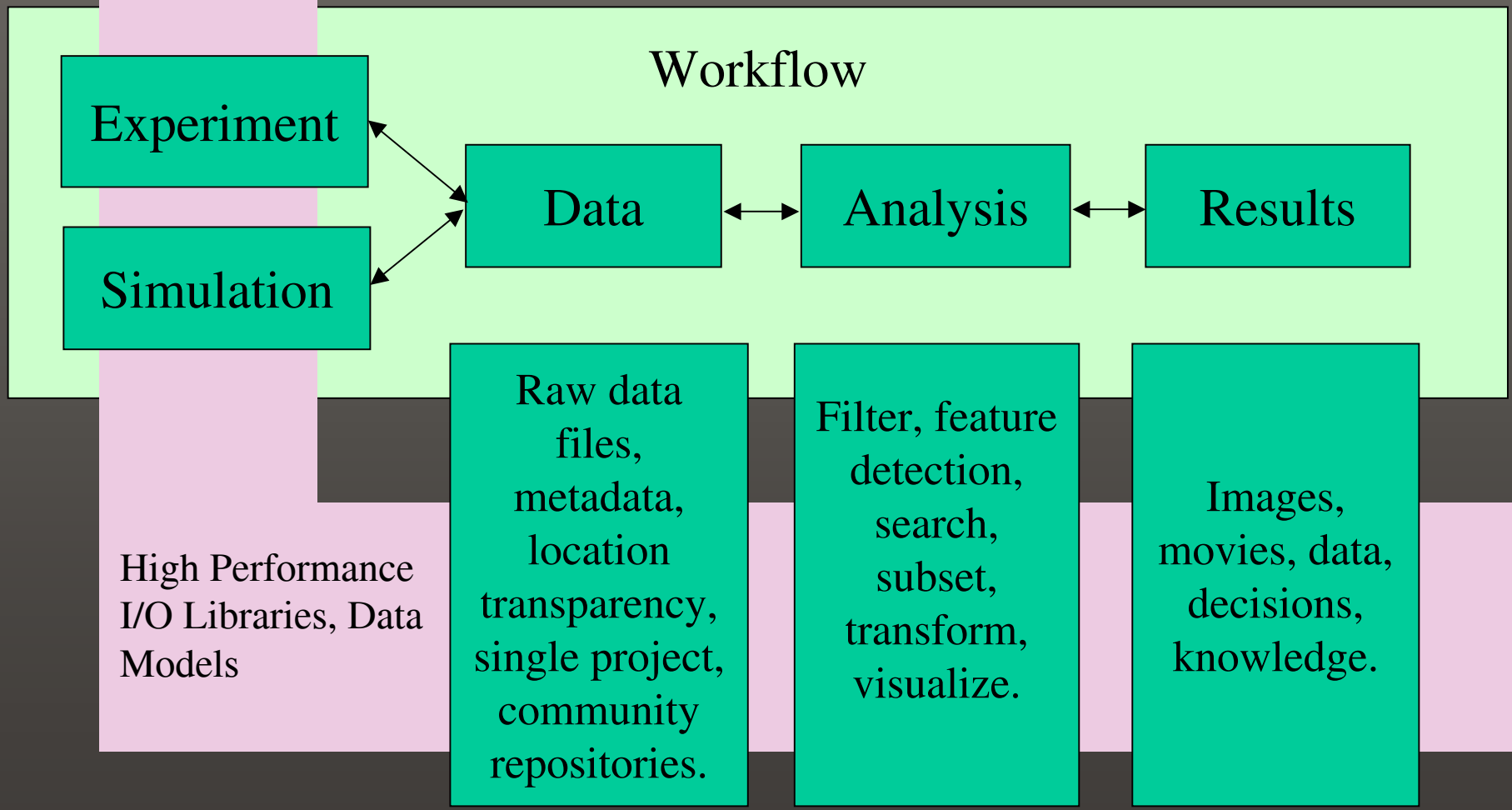
“Visualization Tools”
VACET and COMPASS

E. Wes Bethel
Lawrence Berkeley National Lab
24 Sept 2007

Outline

- Theme: How can we help?
  - Background and historical perspective
  - Data
  - Data and visualization, visual data analysis
  - Analysis and visualization
  - Production deployment
  - Working with VACET, next steps

Background – SDM, Vis, Science



Historical Perspective

- Visualization and SciDAC1
- Visualization and SciDAC2
  - The accelerator story
- VACET mission and objectives
- Strength through leverage
  - Base, SciDAC, ASC
- Our long history with the accelerator modeling community

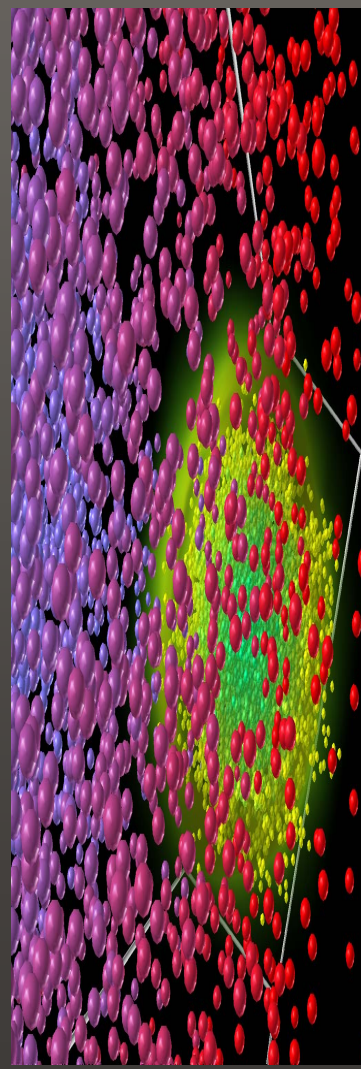
"Visualization Tools"

Outline

- Some historical perspective
- Data – You can run but can't hide!
- Data and visualization, visual data analysis
- Analysis and visualization
- Production deployment
- Working with VACET, next steps

Data – H5part

- What is H5part?
  - API for doing high performance, parallel I/O
  - “Veneer API” atop HDF5, simplifies access to HDF5
  - Parallel I/O, ported to all modern HPC platforms
  - Multiple language interface: C, C++, Fortran, Python
  - Open Source (BSD-like license)
  - Funded under SAP in SciDAC1
  - Didn’t make the SAP cut in COMPASS for SciDAC2

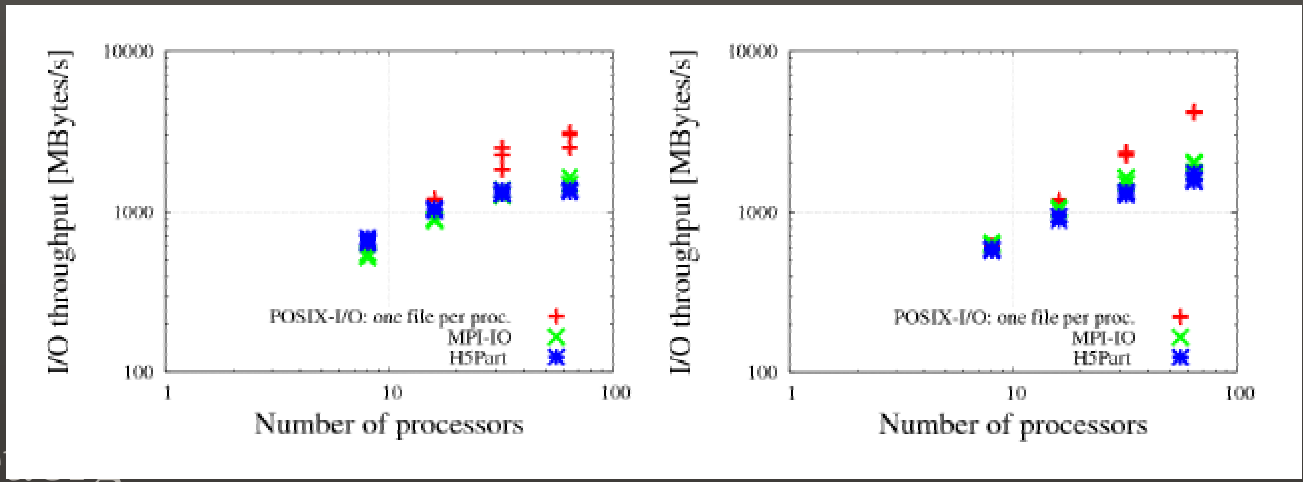


Data – H5part, ctd.

- Why H5part?
  - Single API for accelerator codes to use for I/O
  - Single API for analysis tools to use to access data (e.g., not just for visualization)
  - Promotes interoperability, science efficiency
    - E.g., don't have to do manual data format conversions for analysis.
  - Need for high performance I/O on parallel machines – not an easy task!

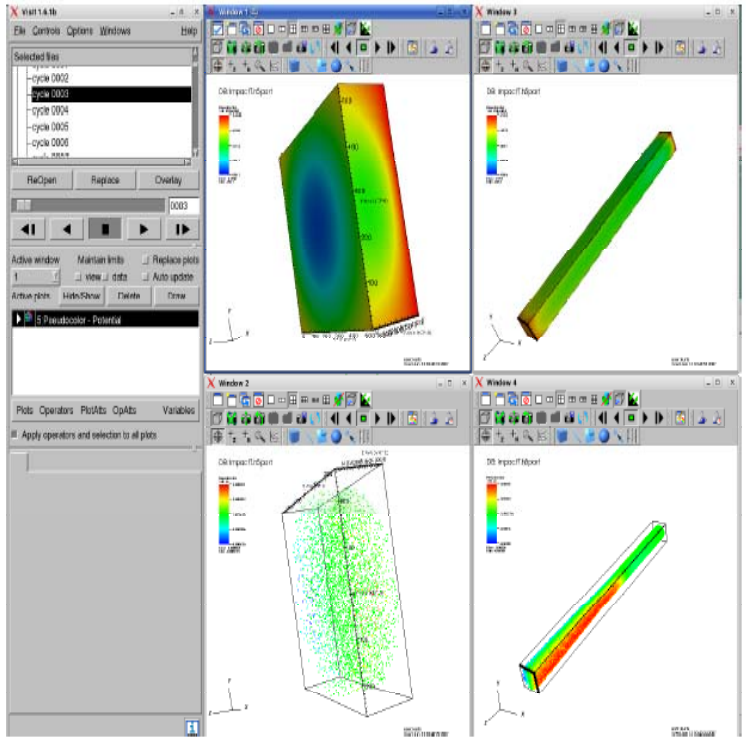
H5part Performance

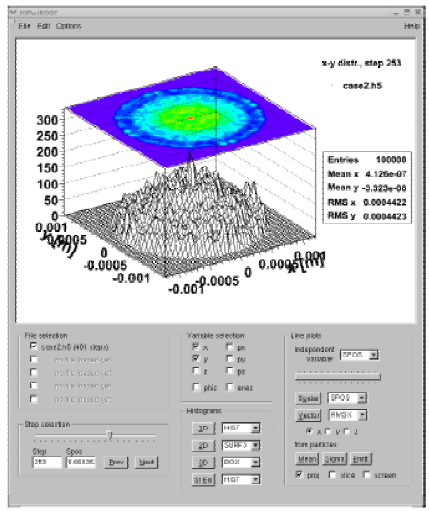
- Recent experiment
  - 8-64 CPUs on Bassi (IBM p5 system)
  - Constant dataset size (left), and constant data load per CPU (right)
  - H5part performance similar to one-file-per PE from MPI: H5part adds no significant overhead.



Data – H5part, ctd.

- H5part and accelerator codes:
  - ImpactT, BeamBeam3D, MaryLie/IMPACT, MAD9p, MC^4, Moldy (MD code)…
  - Next: Osiris (Mori et. al.)
- Loaders:
  - VisIt (more)
  - H5partRoot
  - “Gnuplot”
  - (AVS/Express)





H5part and Similar Efforts

- Tech-X's FMSL/Visual Schema project
  - Subject of J. Cary's presentation after this one
  - Projects have similar goals, but are really orthogonal and complementary.
  - H5part directly addresses the issue of high performance parallel I/O
  - FMSL/VS adds higher level logic to load H5part (and other formats)

H5part Status

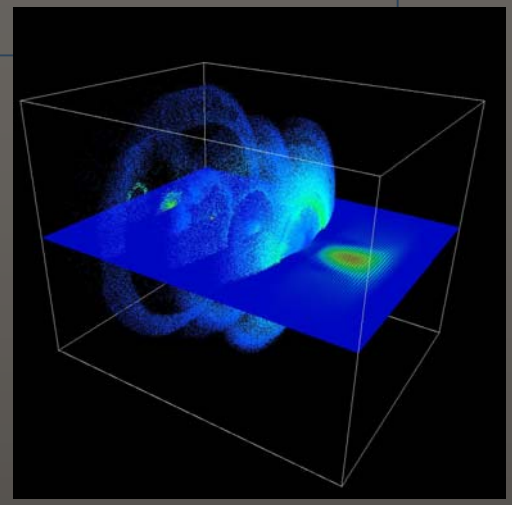
- Production quality – stable.
- In “maintenance phase”
- We can provide help porting codes to use H5part for parallel, high performance I/O.
- Project page:
  - http://vis.lbl.gov/Research/AcceleratorSAPP/
- Contact:
  - Prabhat, prabhat@hpcrd.lbl.gov

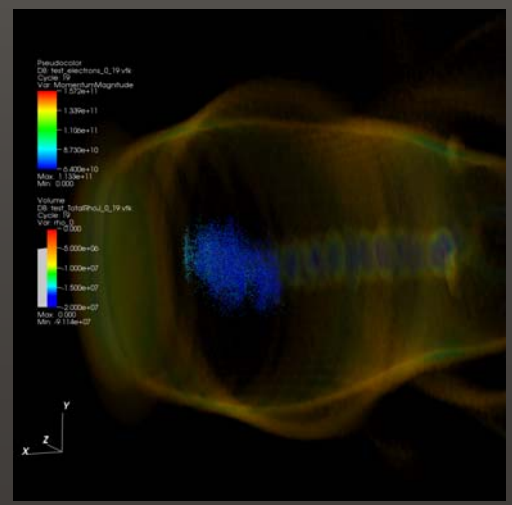
Outline

- Some historical perspective
- Data
- Data and visualization, visual data analysis
- Analysis and visualization
- Production deployment
- Working with VACET, next steps

Data and Visual Data Analysis

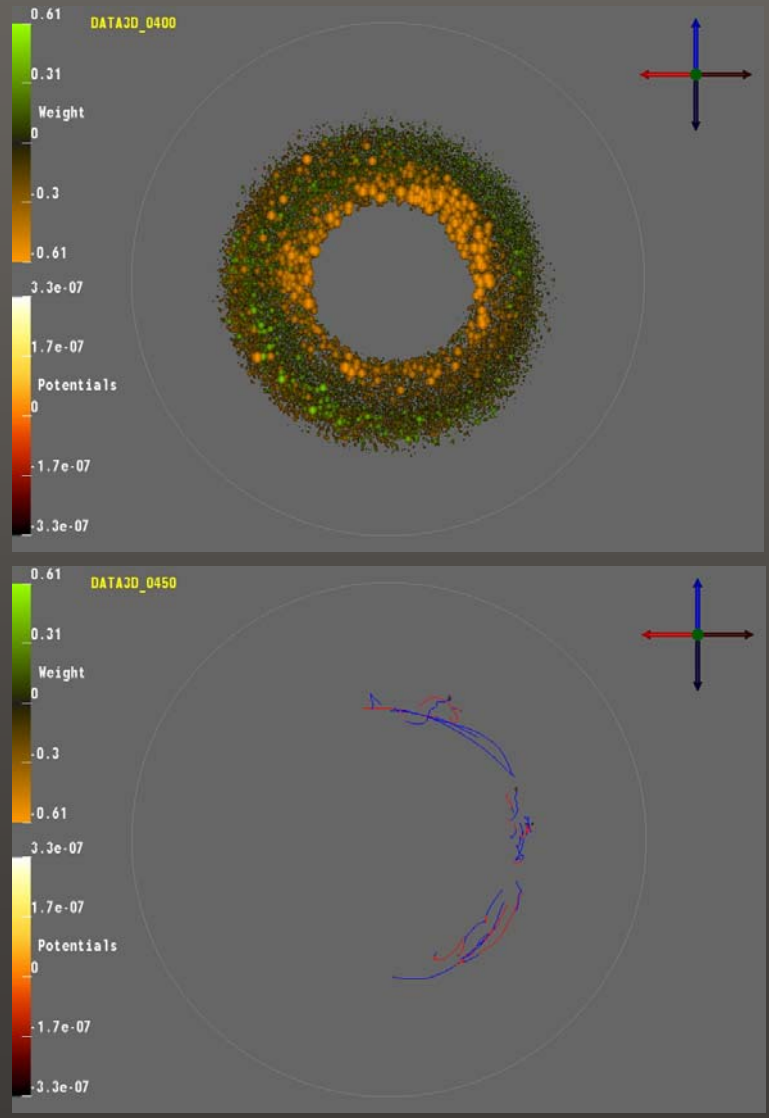
- Premise: focus visualization and analysis processing on “interesting” data.
- Geddes’ laser wakefield work:
  - Top: slice of electric field, particles thresholded by momentum mag.
  - Bottom: volume-rendered plasma density, particles undergoing wakefield acceleration (momentum magnitude threshold filter)





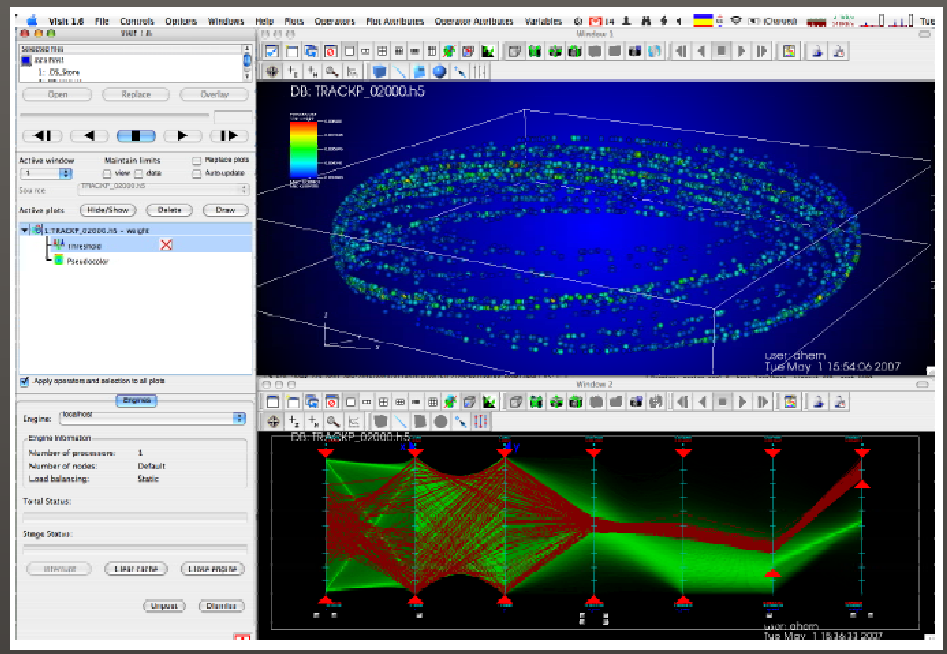
Data and Visual Data Analysis – Fusion

- GTC is a PIC code for modeling microturbulence
  - Top: all particles from a single timestep
  - Bottom: particles that undergo “trapping” at least 20 times.
  - *New VACET project: statistical downsampling to reduce I/O load.*



Data and Visual Data Analysis – Fusion

- Application: fusion microturbulence
- Objective: interactive multidimensional filtering to locate and analyze interesting phenomena.
- Result: capability in production software (VisIt).



Data and Visual Data Analysis – Data

- Problem: want to do better than O(*n*) for data analysis (where *n* is the size of the data).
- Solution(s):
  - Use state-of-the-art index/query. Working with SDM Center to integrate such technology at the data I/O layer in a way that is transparent to simulation code developers.
  - Leverage this capability in visual analysis tools, can provide assistance for use in other types of tools.

Outline

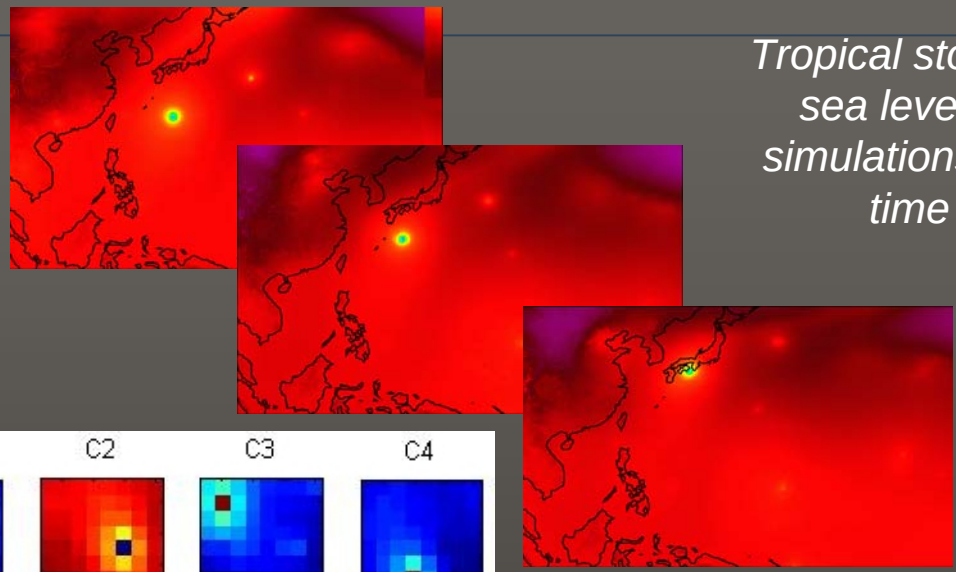
- Some historical perspective
- Data
- Data and visualization, visual data analysis
- Analysis and visualization
- Production deployment
- Working with VACET, next steps

Analysis and Visualization

- Premise: find and analyze “features” in data.
- Methods of location and analysis vary:
  - PCA, ICA, Machine Learning, Support Vector Machines, etc.
  - Topological analysis.
- Impact:
  - Quantitative analysis (rather than “chi-by-eye”)
  - Traction on “big data problems”
  - Basis for comparative analysis (rather than “chi-by-eye”)
- Examples

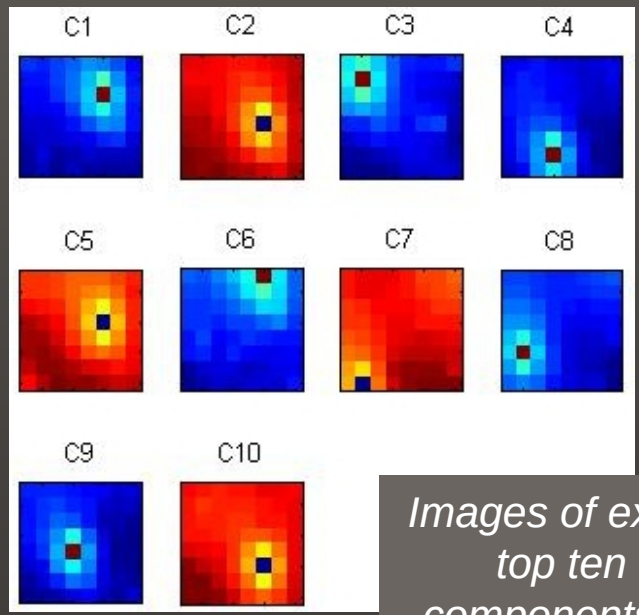
Analysis and Visualization – Climate

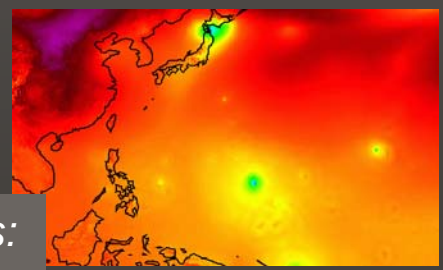
Extracted features can be used as templates for finding similar features.



Tropical storm visible in sea level pressure simulations at multiple time steps.

In this case, the features were variations on rotating low-pressure systems. This was not assumed a priori.

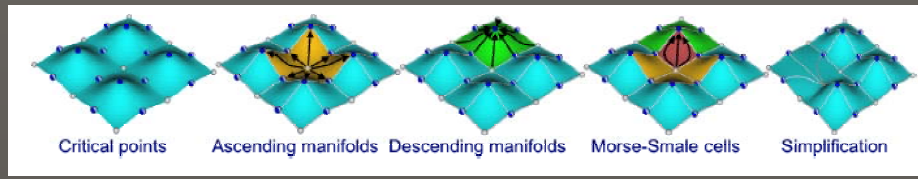


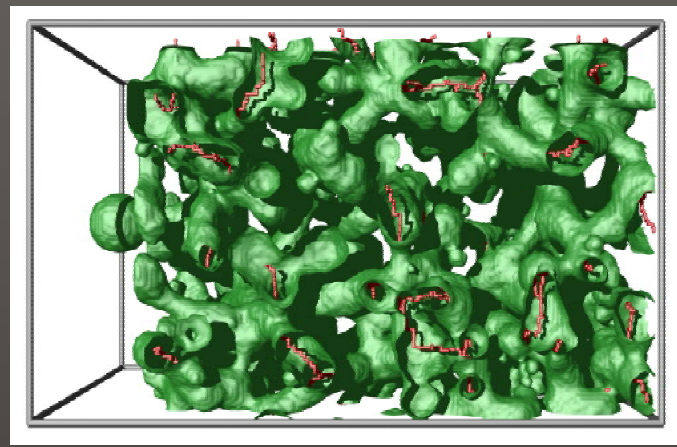


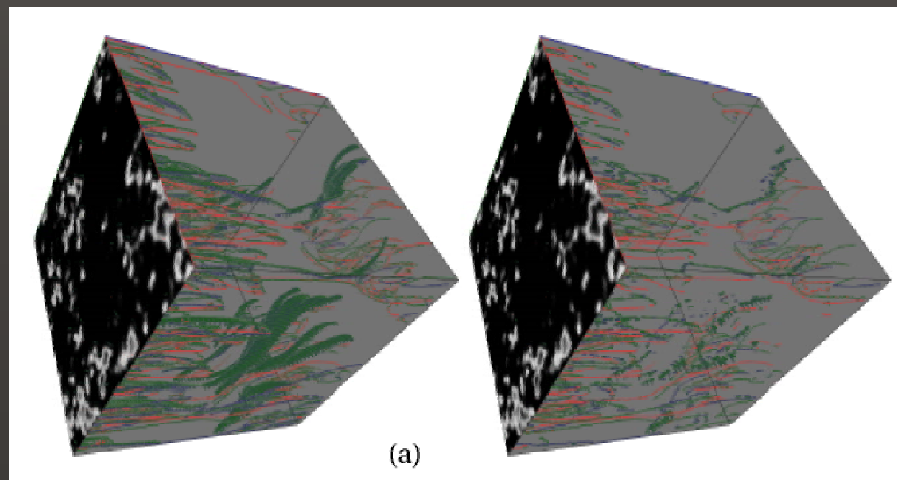
Images of extracted features: top ten independent components were extracted from set of all 8x8 subimages.

Analysis and Visualization – Topology

- Channel structures in porous media : green isosurface separates solid material and empty space; red curves, connecting maxima and 2-saddles, represent channels.
- Combustion kernel feature identification, tracking and analysis. Kernels appear, merge, and extinguish over time. Why? How many?

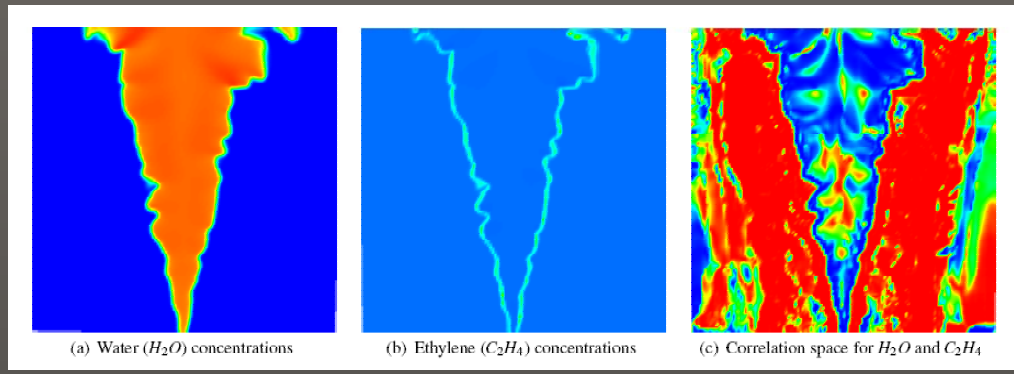




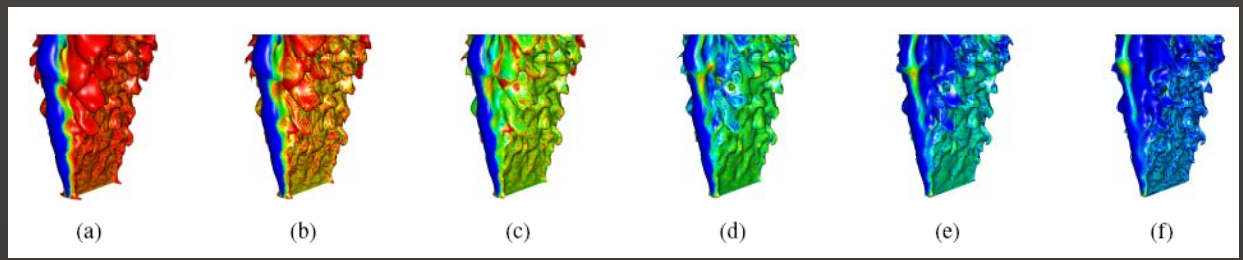


Comparative Analysis

- How are variables related to one another?
- Water and Ethylene concentration (left and middle), correlation field (right)
- Same, but in 3D and mapped onto varying isotherms.



(a) Water (H_2O) concentrations (b) Ethylene (C_2H_4) concentrations (c) Correlation space for H_2O and C_2H_4



(a) (b) (c) (d) (e) (f)

Outline

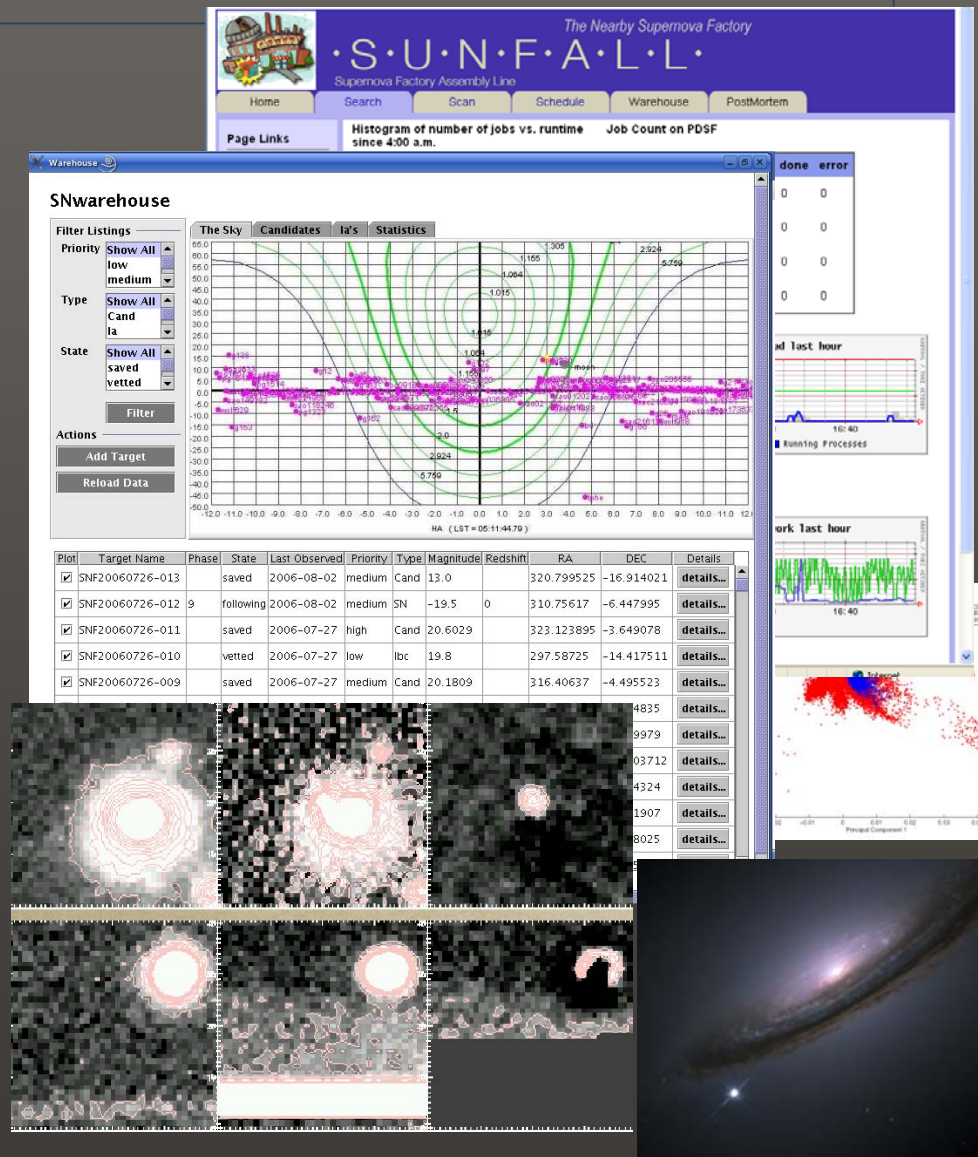
- Some historical perspective
- Data
- Data and visualization, visual data analysis
- Analysis and visualization
- Production deployment
- Working with VACET, next steps

Production Deployment

- How does VACET "deploy" R&D?
  - Production quality software:
    - VisIt – www.llnl.gov/visit
    - SCIRun - http://software.sci.utah.edu/scirun.html
    - Climate: layering new capabilities on CDAT
  - Integrated VACET Software Engineering effort.
- A VACET person "on your team."
- VACET team members at NERSC and LCF.
  - VACET ERCAP allocation at NERSC, discretionary at LCF.

Production Deployment – SUNFALL

- **New supernova data analysis and workflow visualization tools (Sunfall and SNwarehouse) have improved usability and situational awareness, and enabled faster and easier access to data for supernova scientists worldwide**
- **Advanced image processing (Fourier contour analysis) and machine learning techniques running on NERSC platforms have achieved a >40% decrease in human workload in nightly supernova search (>75% FTE)**

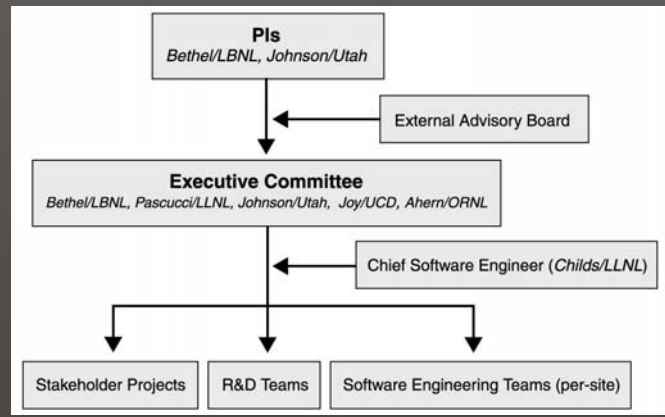


Outline

- Some historical perspective
- Data
- Data and visualization, visual data analysis
- Analysis and visualization
- Production deployment
- Working with VACET, next steps

Working with VACET

- VACET organization – we have a "customer project manager" for each stakeholder project.
- "We" identify a project along with POCs from each team.
  - Identify and prioritize needs
  - Vet with EC
  - R&D plan, goes into our PMP
  - Can draw on resources of the Center to meet objectives
- Early success: APDEC replaces ChomboVis with VisIt as project-wide production visual data analysis app.



Working with VACET, ctd.

- Active accelerator projects:
  - Mori – migrate Osiris from HDF4 to H5part
  - Ryne/Qiang – H5part in ImpactT, workflow and scientific process automation (with SDM Center)
  - Geddes – visualization, query-driven visualization, workflow, data loader for VisIt (with Tech-X)
- We'd like to start up more accelerator projects.
  - Contact info: next (and final) slide.

Epilogue

- This talk barely scratches the surface of all that's happening in VACET. ☺
- Impactful results come from long-term, in-depth multidisciplinary relationships.
  - VACET's M.O. is to have a visual data analysis person be "part of your team."
  - Our ultimate objective is positive impact on your science project.
- VACET is well beyond "vis in isolation."
  - SDM, analysis, architecture, deployment, vis, science focus.
- VACET Contact Information:
  - www.vacet.org
  - Wes Bethel (ewbethel@lbl.gov) and Chris Johnson (crj@sci.utah.edu)