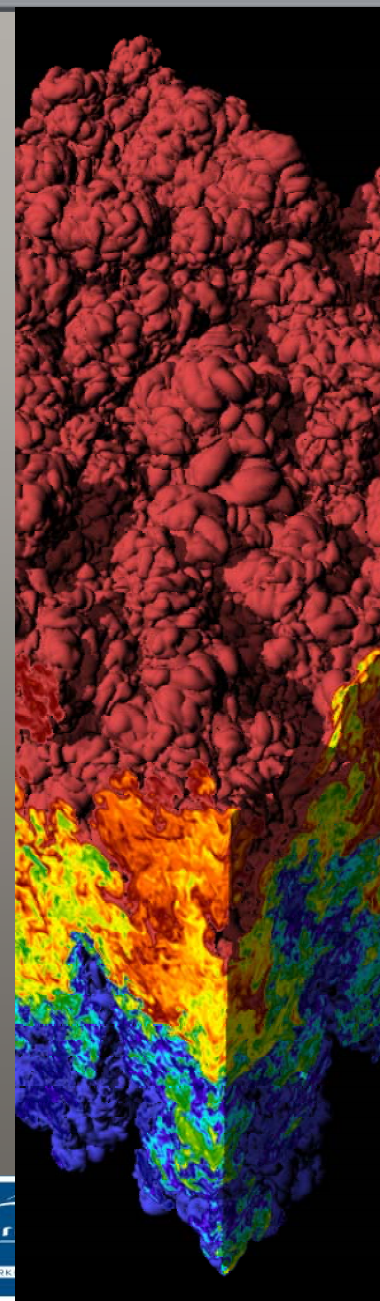
MEET THE SciDAC VISUALIZATION AND ANALYTICS CENTER FOR ENABLING TECHNOLOGIES

E. WES BETHEL (LBNL), CHRIS JOHNSON (UTAH), KEN JOY (UC DAVIS), SEAN AHERN (ORNL), VALERIO PASCUCCI (LLNL), JONATHAN COHEN (LLNL), MARK DUCHAINEAU (LLNL), BERND HAMANN (UC DAVIS), CHARLES HANSEN (UTAH), DAN LANEY (LLNL), PETER LINDSTROM (LLNL), JEREMY MEREDITH (ORNL), GEORGE OSTROUCHOV (ORNL), STEVEN PARKER (UTAH), CLAUDIO SILVA (UTAH), XAVIER TRICOCHE (UTAH), ALLEN SANDERSON (UTAH), HANK CHILDS (LLNL)

www.vacet.org

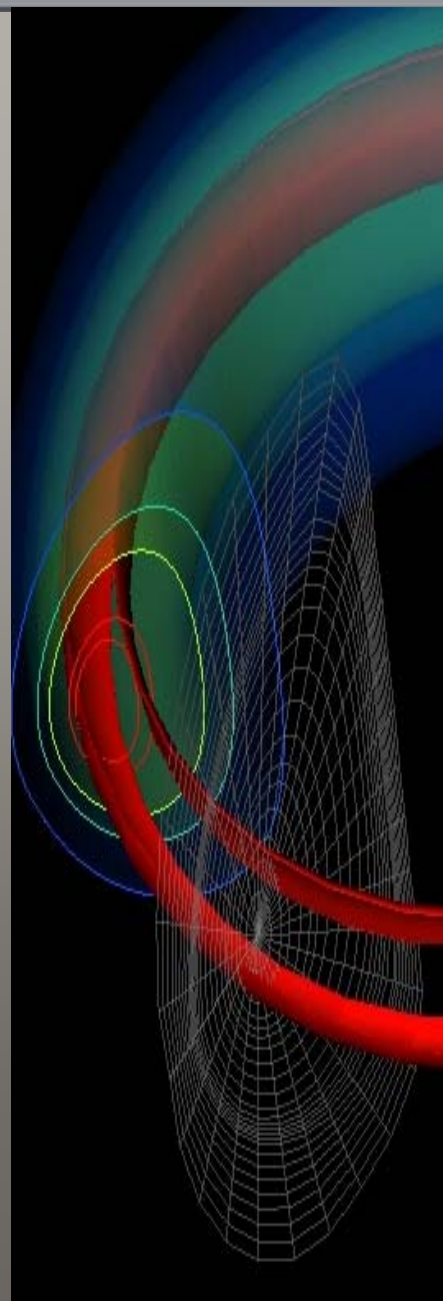
VACET Deploys Technology for Petascale Visual Data Analysis

- Too Much Data
  - Technological advances increase our ability to collect and generate scientific data, resulting in an “information big bang.”
  - The inability to gain understanding from large and complex datasets is a bottleneck in modern science.
  - This situation continues to worsen as computational platforms increase to the Petascale regime.



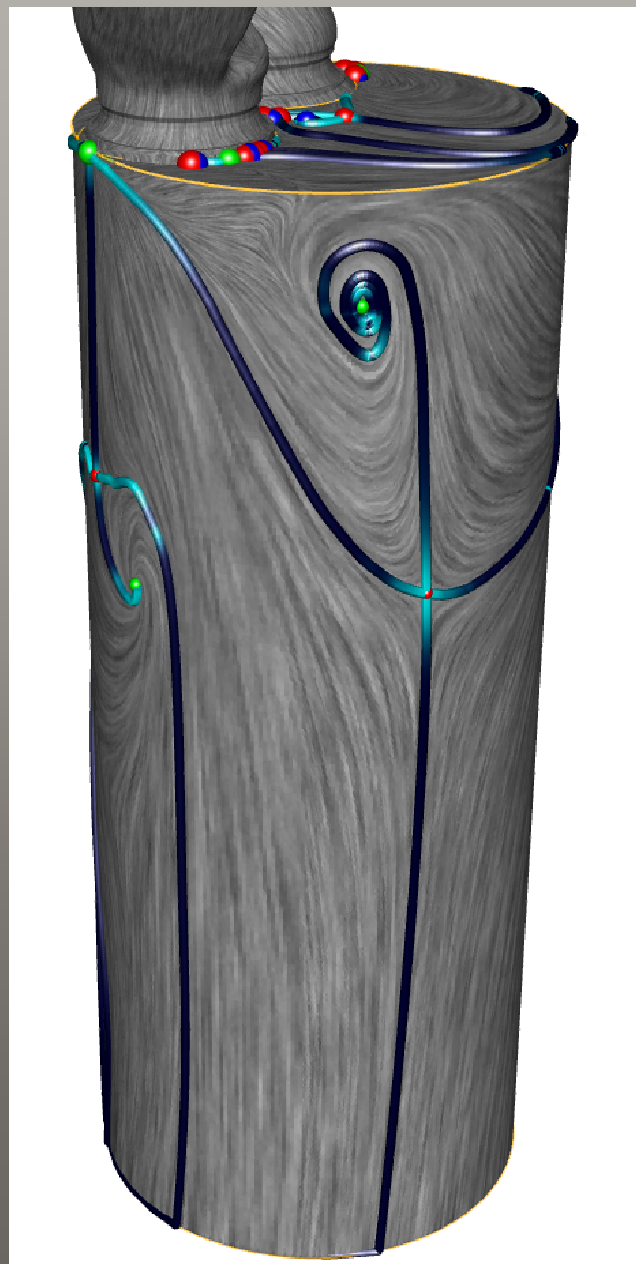
VACET Closes the Gap Between Visualization Research and Best Practice

- Fast and Systematic Deployment
  - To coordinate efforts of DOE's leading visualization research and development programs.
  - To produce and deploy production-quality visualization and analytics software infrastructure capable of addressing petascale data understanding challenges.
- Interdisciplinary Nature of Visualization Needs Strong Coordination/Collaboration
  - To leverage technologies from other SciDAC Centers – data management, mathematics – as part of a broader program (SciDAC).
  - Visualization is the centerpiece of modern scientific data understanding.



VACET Mission Statement

- Meet the Data Understanding Challenge
  - Adapt, extend, create when necessary, and deploy visualization and data understanding technology for SciDAC2 science stakeholders.
- Production-Quality and –Capable Software
  - Provide production-quality visualization and analytics software infrastructure for use at DOE's flagship, open computing facilities.

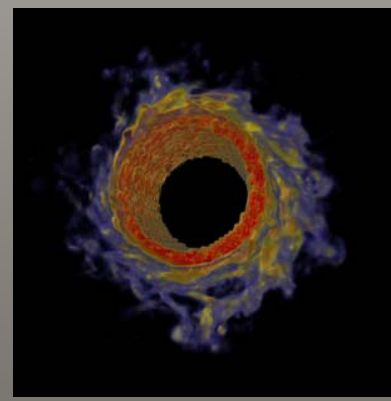


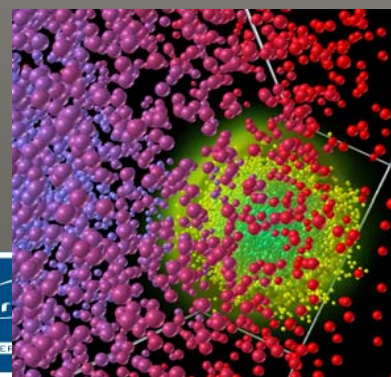
Embeds Support to Science Stakeholders Centralizing Technology Development

- Accelerator Modeling, Astrophysics, Climate, Combustion, Fusion.

Techniques

Applications	Technical Point of Contact	Visualization							Analytics						
		Integration of Basic Tools for Visualization and Analysis	Publication Quality Images, Illustrations, and Movies	Project-Wide Visualization Tools	Flow Techniques	Scalable Tools	Remote Data Access and Streaming Techniques	Collaborative Tools	Query Driven Visualization	Statistical Displays	Feature Detection (topological analysis, semantic range queries)	Temporal specific issues (feature tracking, events, ...)	Multiple Fields	Comparative Visualization and Analytics	Uncertainty Visualization
Fusion	Utah ORNL LBNL	●●	●		●	●		●●	●●●	●●●	●●	●●●	●●●	●●●	●
Combustion	LBNL LLNL ORNL	●	●●●		●●●	●●●	●●●	●●●	●●●	●●	●●●	●●●	●●●	●●	●
Accelerator	LBNL	●	●	●	●	●	●●		●●●	●●	●●●	●●●	●●	●●	●
Astrophysics	ORNL LBNL	●	●●	●	●	●●	●●		●●●	●●	●●	●●	●●●	●●	●
Turbulence	LLNL	●	●●●		●	●●●	●●			●	●●●	●●●	●	●	
Climate	LLNL ORNL	●			●	●●	●●		●●●	●●	●●	●●	●●	●●	●●
Environmental Management	LBNL	●	●●	●●		●●●	●●●	●●							



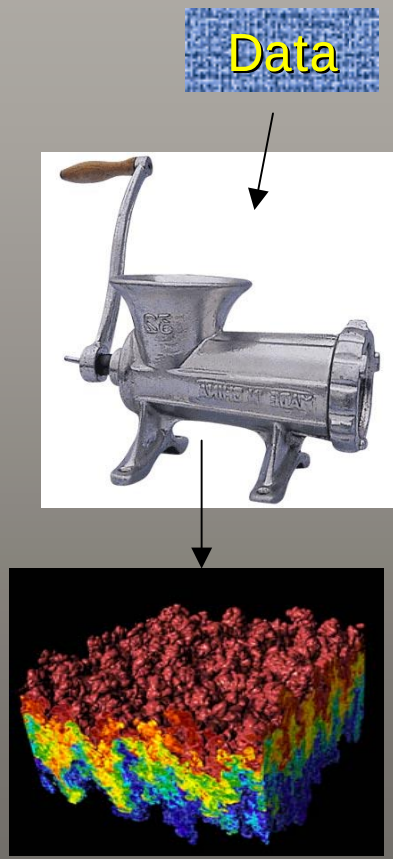


www.vacet.org

SC06 POWERFUL BEYOND IMAGINATION

Petascale Data Understanding

- "Scalable visualization" solutions do not necessarily increase the likelihood of scientific insight.
- More visualization output can cause more problems than it solves.
  - Increased depth complexity.
  - Increased cognitive workload.

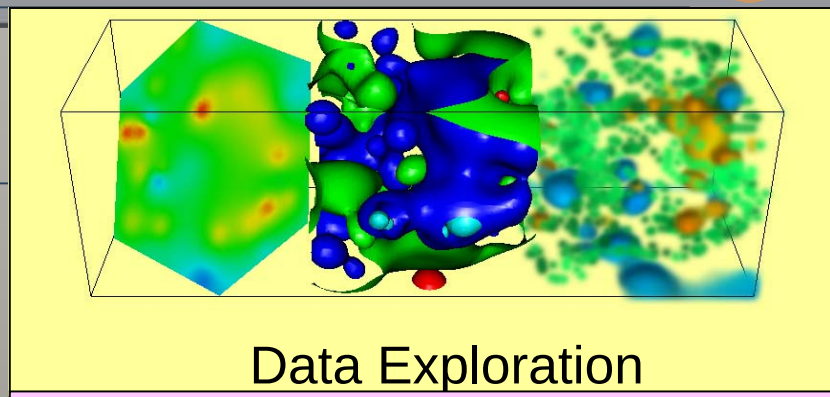


Software Platforms for R&D, Deployment

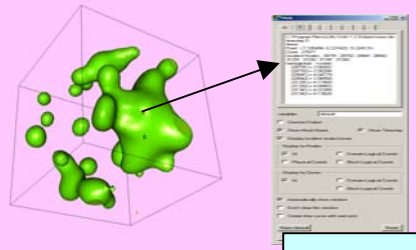
- VisIt and SCIRun
  - Mature, Open Source platforms for visualization and analysis (funded by DOE, NIH, NSF).
  - Extensible to provide new features and capabilities.
  - Customize/encapsulate to create vertically integrated, domain-specific application.
  - Run on modern architectures.

VisIt – www.llnl.gov/visit

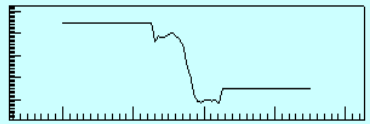
- Full-featured visualization and analysis tool.
- Feature set reflects user requirements.
- Rich set of analysis and visualization capabilities.
- Parallel capable (LLNL BG/L, White)



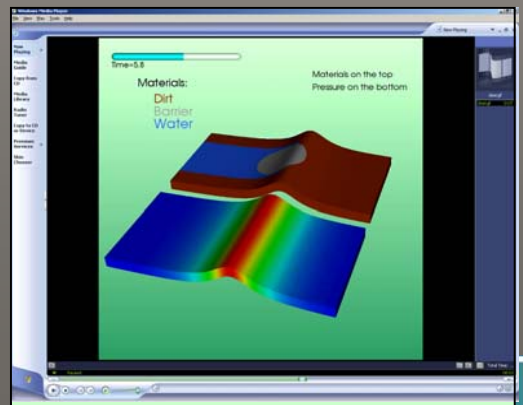
Data Exploration



Code Assessment



Quantitative Analysis



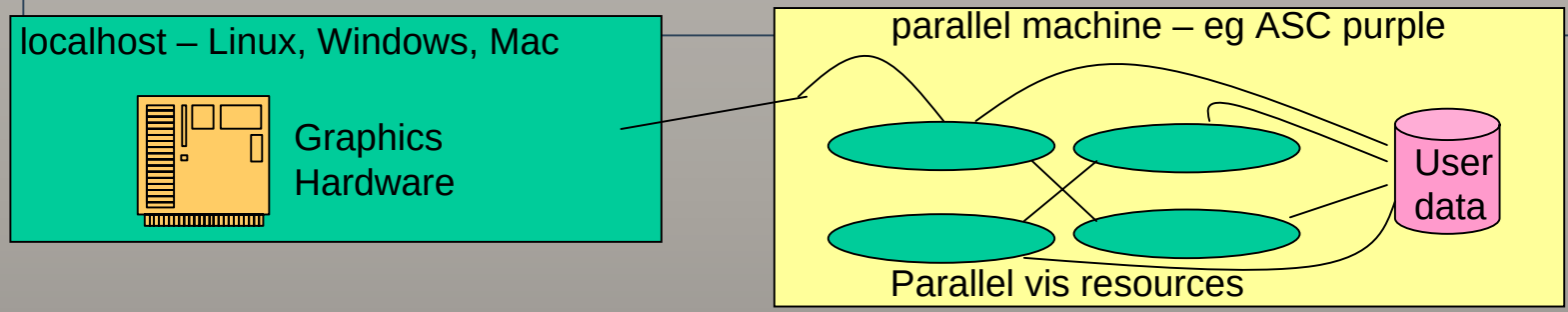
Presentations

Blue Gene/L, 128K procs., 512 vis procs.

ASC white, 8192 procs., 512 vis procs.

R&D 100

VisIt Architecture



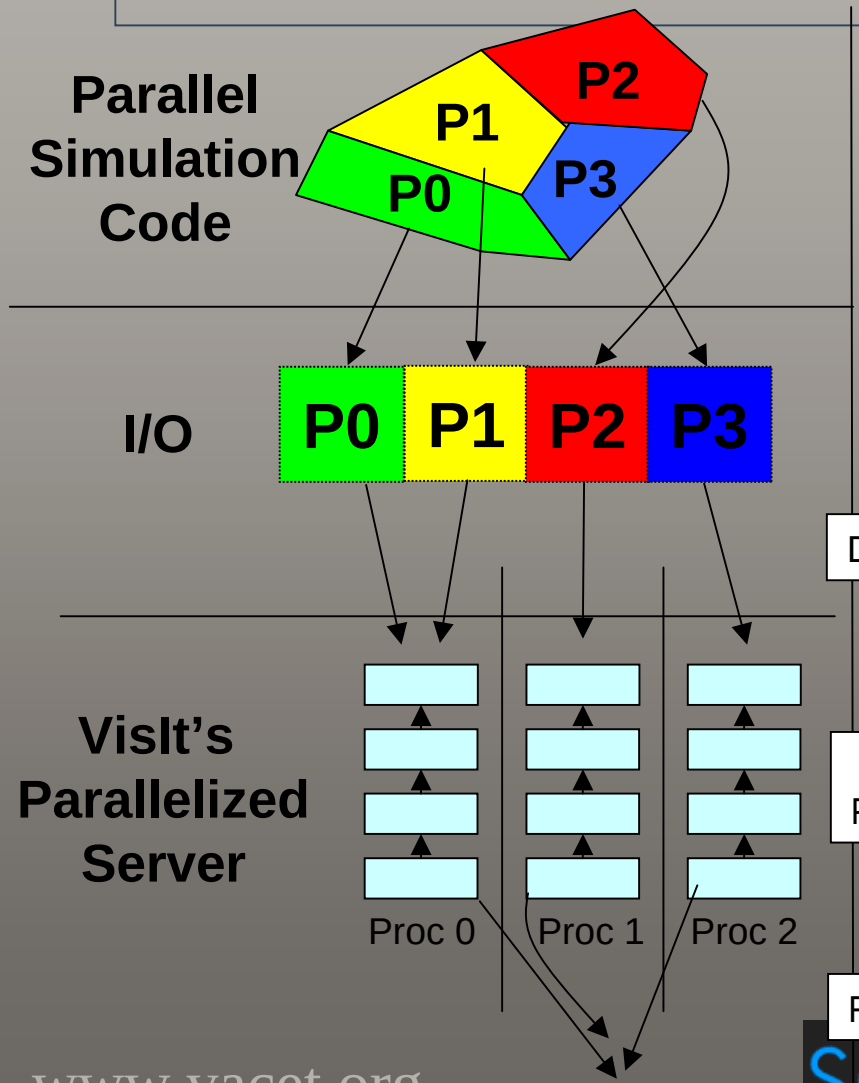
- Good for remote visualization
- Leverages available resources
- Scales well, given adequate resources
- No need to move data
- Additional design considerations:
  - Plugins
  - Multiple UIs: GUI (Qt), CLI (Python), more…

The principal architectural decisions – client/server, VTK, Qt, Python, plugins, etc – were put in place 6 years ago

...

VisIt has been building on that base ever since.

VisIt Back-End Architecture

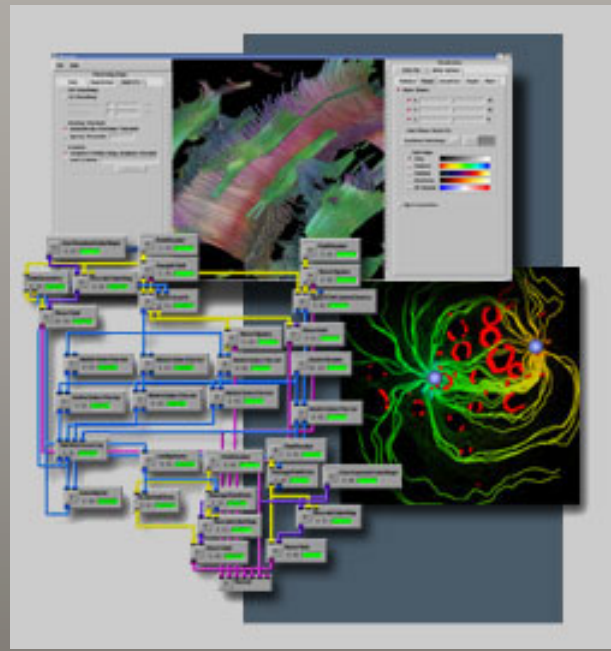


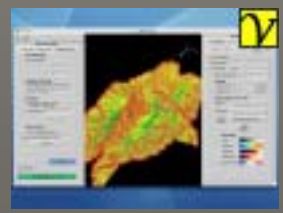
- VisIt: identical data flow networks on each processor.
  - Networks differentiated by portion of data they operate on.
  - Processing happens in parallel
  - Rendering: parallelized if nec
  - "Sc

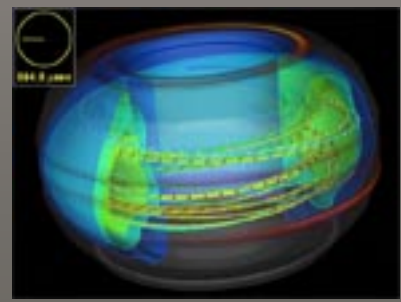
Lots of ignored issues here: partitioning, communication, etc

SCIRun – www.sci.utah.edu

- Problem-solving environment for modeling, simulation and visualization of scientific problems.
- Visualization research platform – visual programming and data flow.
- End-user tool – PowerApps (FEM, Fusion, MRI/Biomedical)







Deployment Strategy

- Production-quality data understanding software infrastructure.
- New capabilities targeted at stakeholder needs delivered in infrastructure.
  - Where practical, also made available as standalone libraries/components for use outside of those infrastructure environments.
- Coordinate VACET-wide software development.
- Leverage existing maintenance and release engineering support.

Feature Identification and Tracking

- Query-driven visualization and analysis
  - Find “interesting data.”
  - Limit visualization and cognitive processing to only “interesting data.”
- Topological analysis
  - Topological features capture important characteristics of data.
  - Basis for high-level comparative analysis.

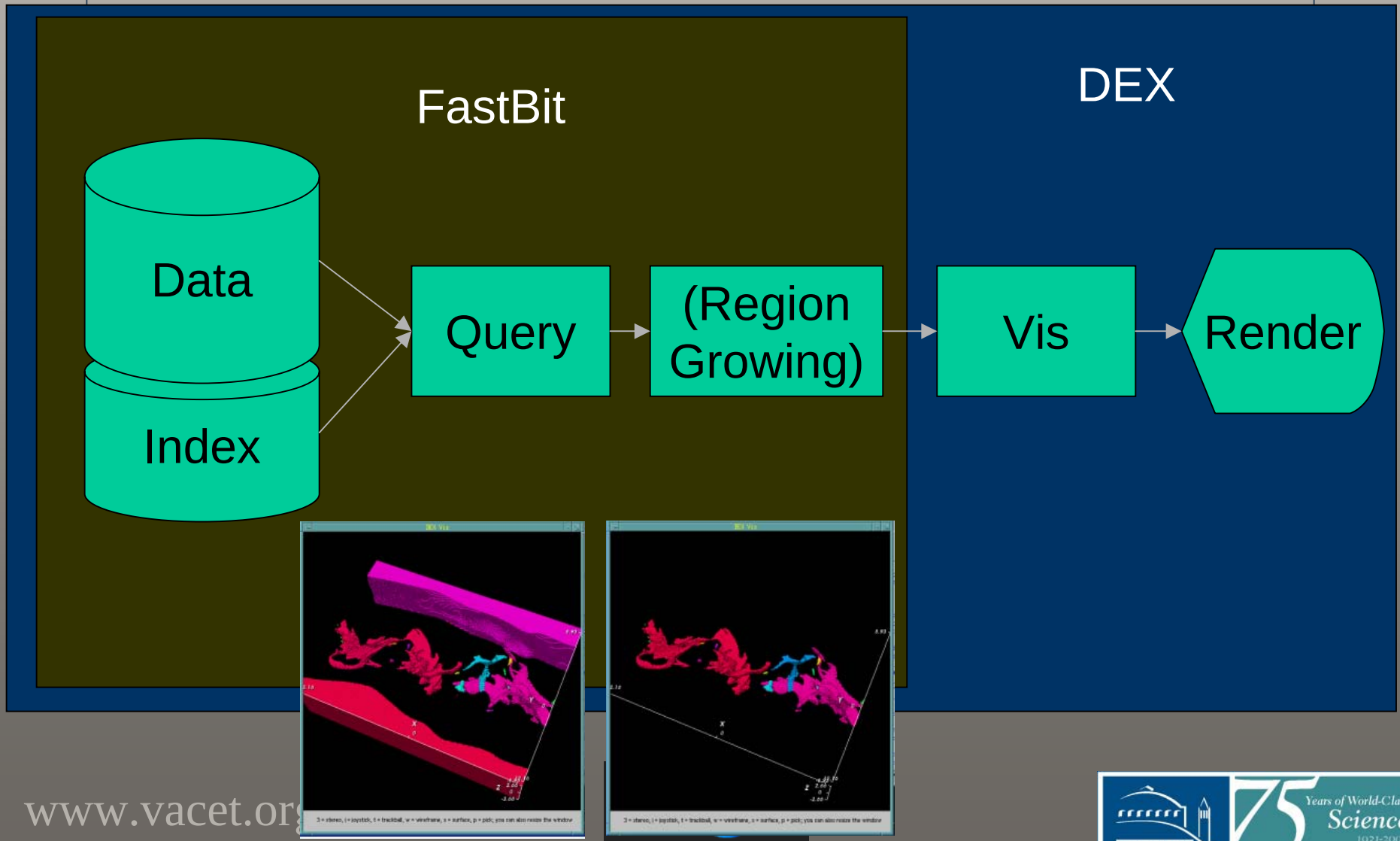
What is Query-Driven Visualization?

- What is Query-Driven Visualization?
  - Find “interesting data” and limit visualization, analysis, machine and cognitive processing to that subset.
- One way to define “interesting” is with compound boolean range queries.
  - E.g., $(CH_4 > 0.1)$ AND $(T_1 < temp < T_2)$
- Quickly locate those data that are “interesting.”
- Pass results along to visualization and analysis pipeline.

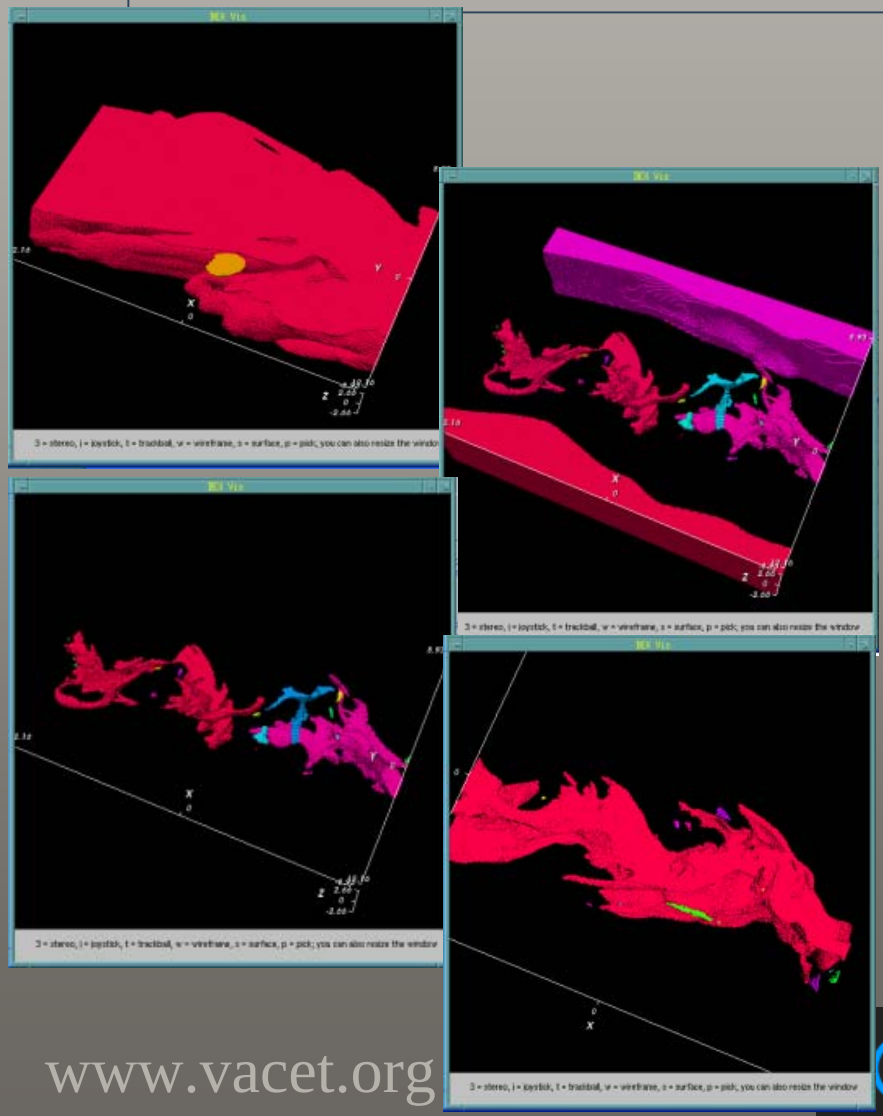
Query-Driven Visualization Architecture

The Canonical Visualization Pipeline

Query-Driven Visualization Architecture



Query-Driven Visualization Example



- $CH_4 > 0.3$
- $Temp < T_1$
- $CH_4 > 0.3$ AND $temp < T_1$
- $CH_4 > 0.3$ AND $temp < T_2$
  - $T_1 < T_2$

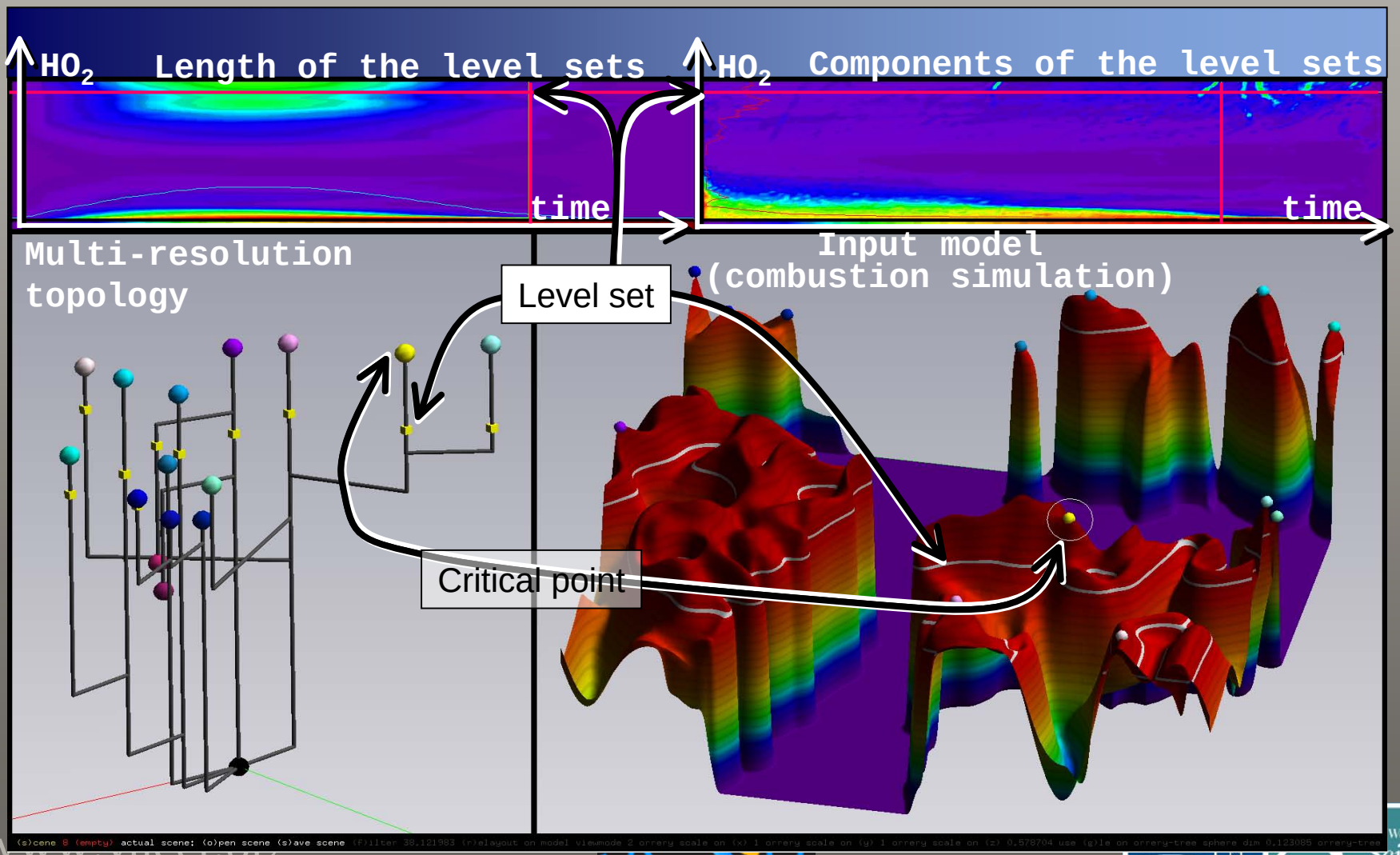
Why Query-Driven Visualization?

- Facilitate scientific understanding.
  - Reduce cognitive processing load.
- Efficient approach for petascale regime.
  - Reduce processing time.
- Traction on difficult visualization and analysis problems:
  - E.g., temporal analysis.
- Observation
  - 80% of code in visualization and analysis deals with scientific data management issues.
  - Leverage state-of-the-art scientific data management technology.
  - Close relationship with the SciDAC Scientific Data Management Center.

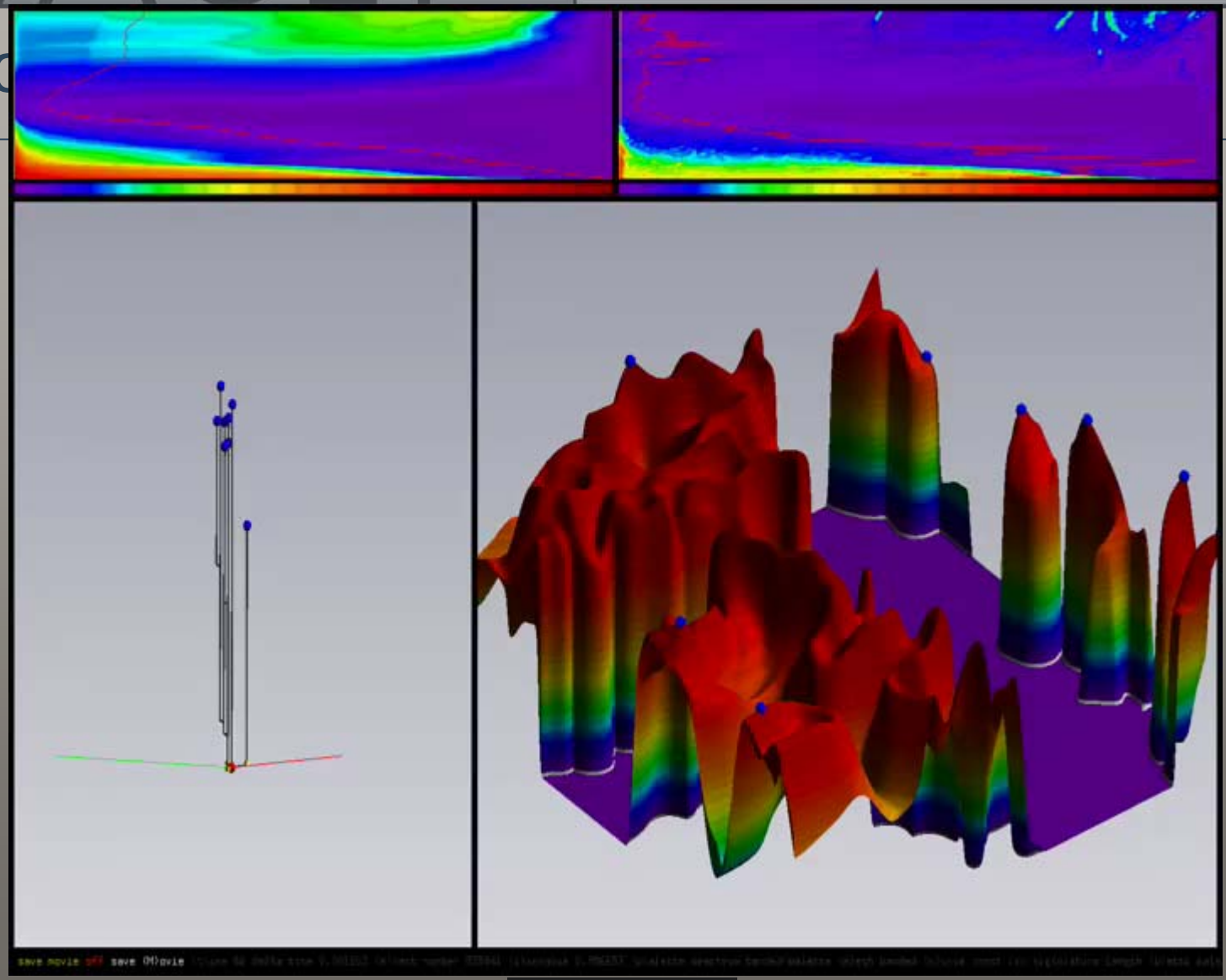
Why Topological Methods?

- Large data results in complex visuals:
  - Nested isosurfaces
  - Occlusion
- Traditional approaches difficult to scale with data sizes.
- Time-varying datasets are treated as sets of independent snapshots
- Not robust to numerical noise.
- Difficult to capture multiple scales.
- Our solution: Topological structures are a multi-scale mesh-independent representation of “shape”.

Topological Analysis (and Linked Views)



To

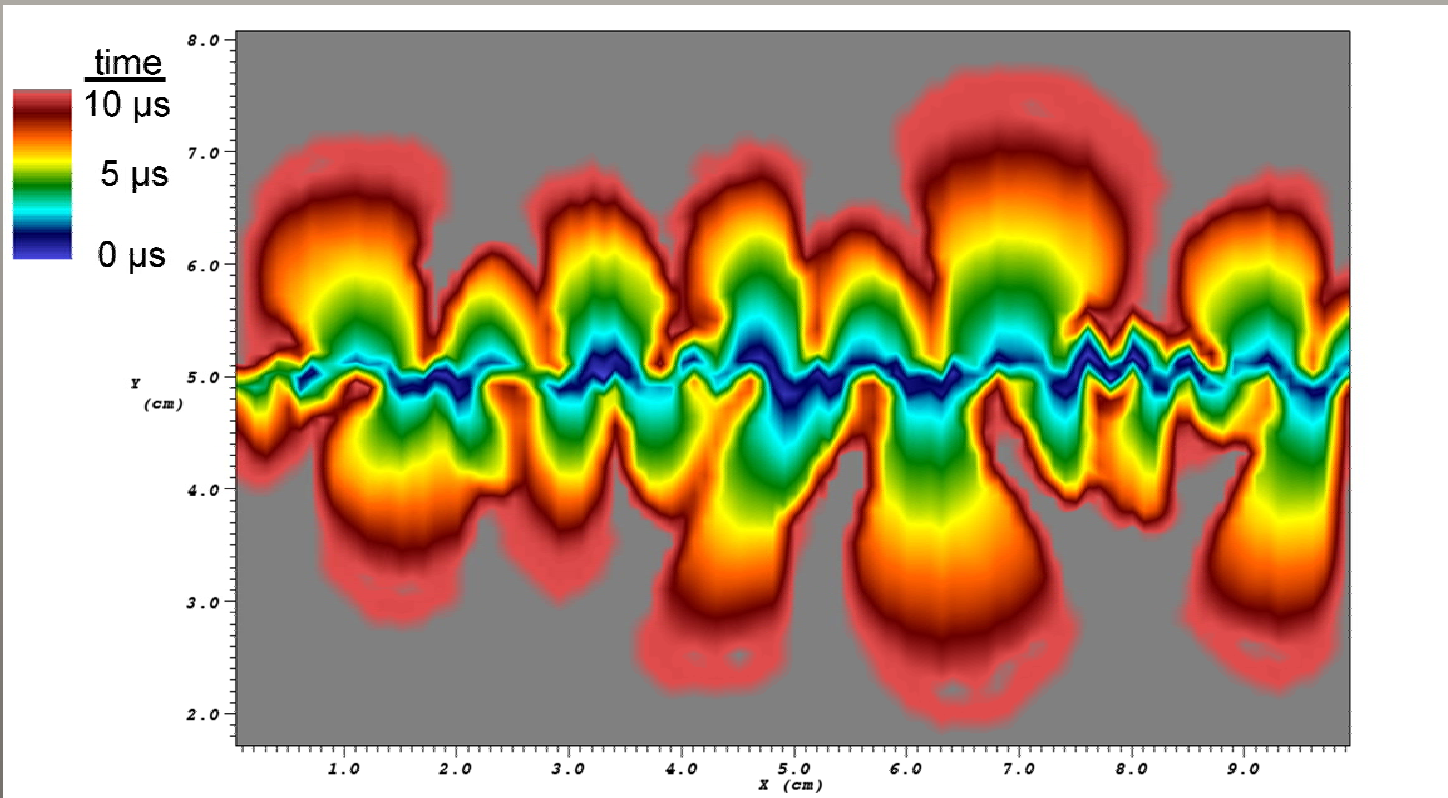


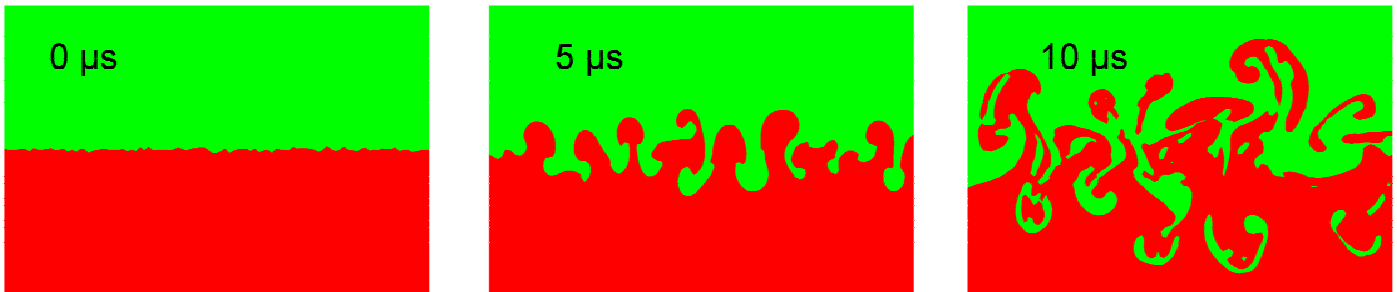
Comparative Visualization and Analysis

- Comparisons are extremely important:
  - Compare one simulation over time
  - Compare simulation to experiment
  - Compare simulation to simulation (x3, x4, … xN?)
  - Compare symmetry conditions (one simulation, one time slice)

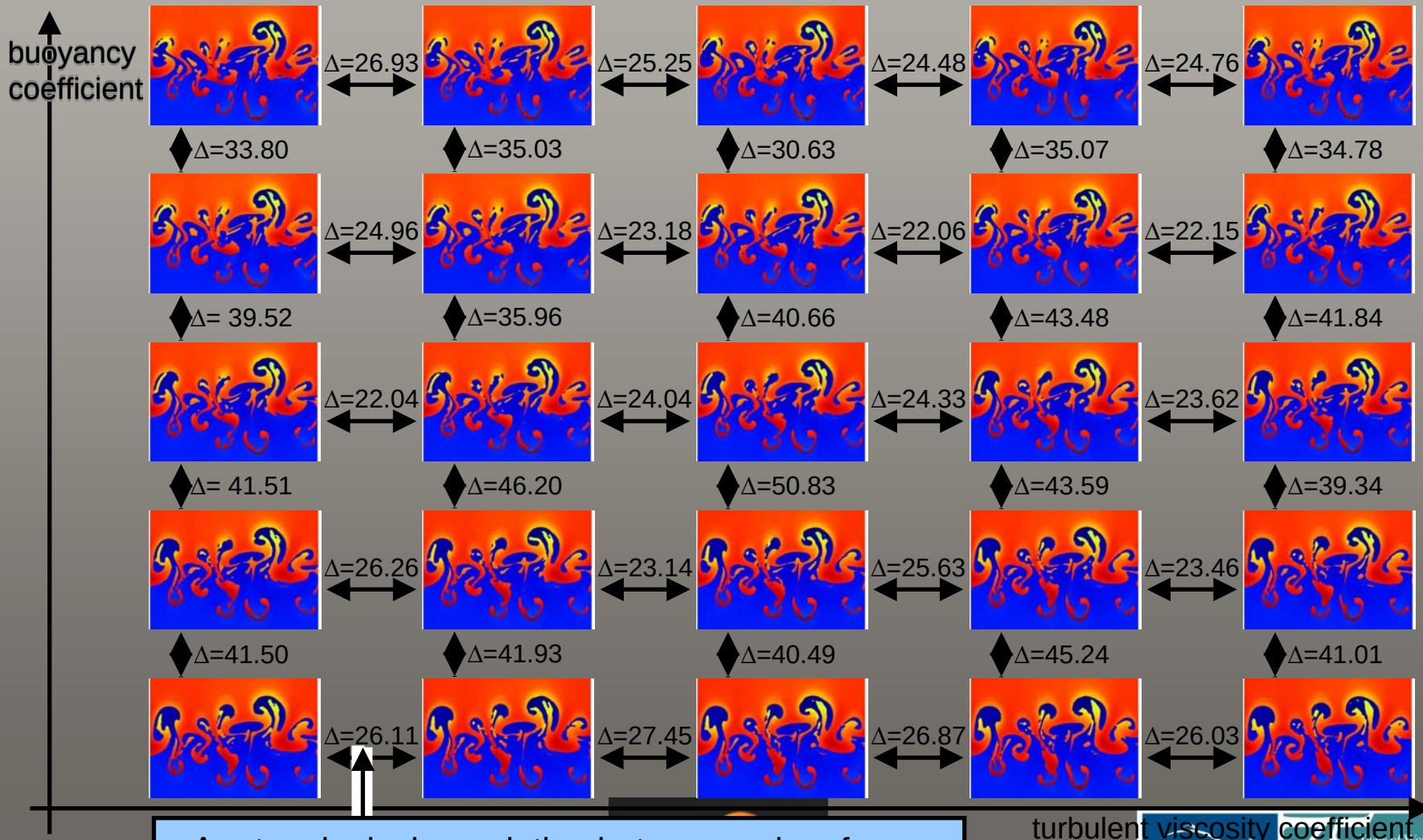
- Goal: use building blocks to accomplish all kinds of comparative visualization and analysis
  i. Image based ← view side by side
  ii. Data level ← (examples to follow)
  iii. Feature level ← (e.g., topological)
  iv. Procedural ← (comparing recipes, sensitivity analysis)

Temporal Analysis



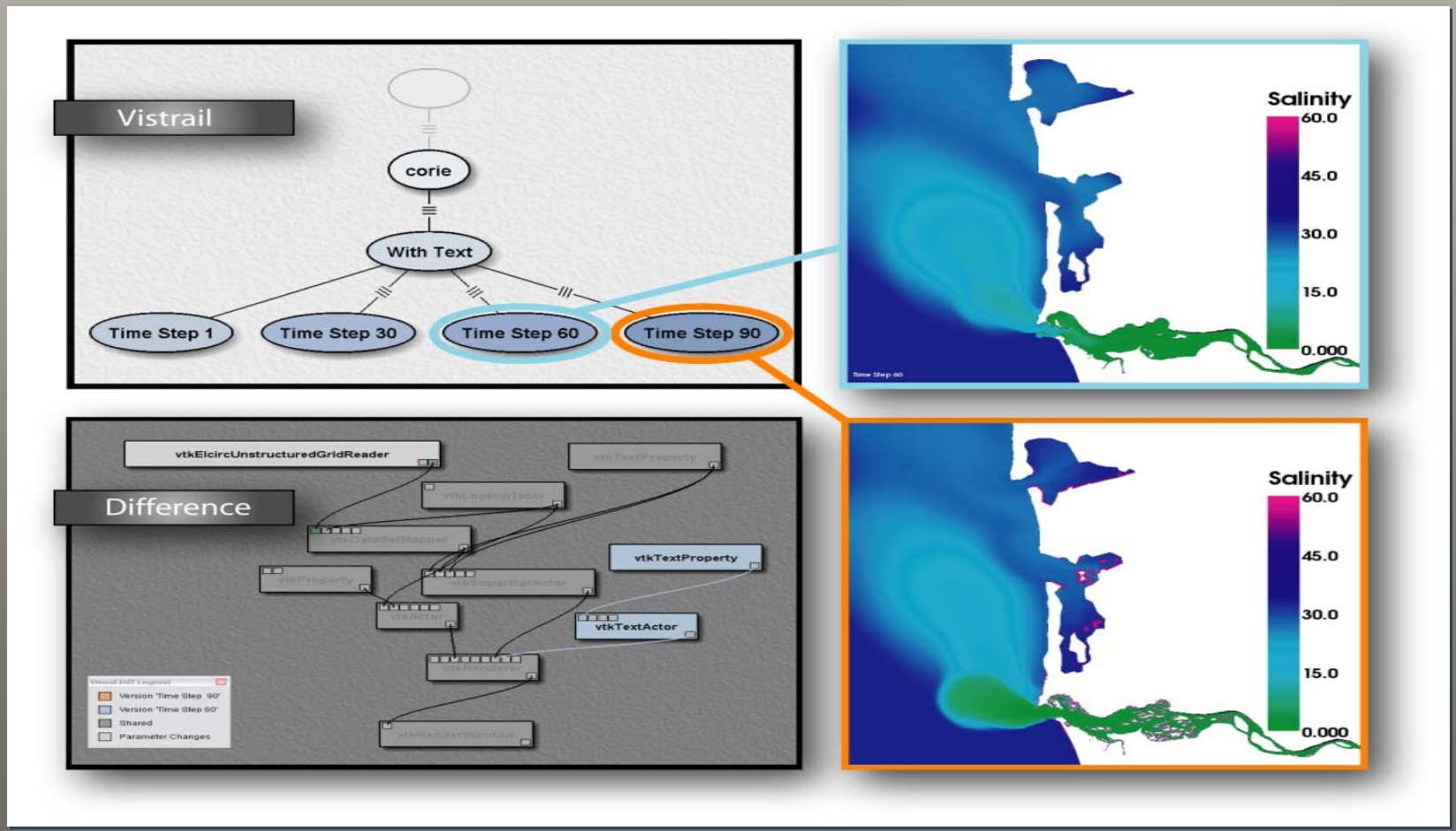


Topological Analysis and V&V



Δ = topological correlation between pairs of runs

Procedural Analysis



Concluding Remarks

- Gaining insight from a firehose of information is a limiting factor in nearly all fields of scientific endeavor.
- VACET's primary objective is facilitating scientific insight.
  - Explosion of data.
  - Visualization/analytics R&D.
  - Effective deployment of production-quality tools.
- This material presents only a subset of VACET's technology and only hints at the scope of our project.

The End