Seeing the Unseeable

E. Wes Bethel[*]and Chris Johnson[†]

Charles Hansen, Claudio Silva, Steven Parker, Allen Sanderson,
Xavier Tricoche, Martin Cole[‡]

Sean Ahern, George Ostrouchov, Dave Pugmire, Jeremy Meredith[§]

Valerio Pascucci, Hank Childs, Peer-Timo Bremer,
Daniel Laney, Ajith Mascarenhas, Kathleen Bonnell[¶]

Ken Joy, Bernd Hamann[‖]

Cecilia Aragon, Gunther Weber, Prabhat[**]

November 2007

1 Abstract

Scientific visualization, which is the transformation of abstract data into readily comprehensible images, plays an integral part of the scientific knowledge discovery process. DOE's SciDAC Visualization and Analytics Center for Enabling Technologies (VACET) is the focal point in SciDAC for production-quality visual data analysis software infrastructure. As a SciDAC Center, VACET's mission encompasses both research and development activities. Research activities consist of conceiving algorithms that address difficult scientific data understanding challenges. Development activities focus on design and implementation of software architectures that serve as the basis for delivering production-quality visual data analysis solutions that are deployed across DOE's computational science ecosystem. We present here an overview of VACET with a focus on accomplishments from several multidisciplinary projects.

2 Introduction

Galileo Galilei (15 February 1564 – 8 January 1642) has been credited with fundamental improvements to early telescope designs that resulted in the first practically usable instrument for observing

---

[*] Lawrence Berkeley National Laboratory

[†] Scientific Computing Institute, University of Utah

[‡] Scientific Computing Institute, University of Utah

[§] Oak Ridge National Laboratory

[¶] Lawrence Livermore National Laboratory

[‖] University of California – Davis

[**] Lawrence Berkeley National Laboratory

the heavens. With his “invention,” Galileo went on to many notable astronomical discoveries: the satellites of Jupiter, sunspots and the rotation of the sun, and proved the Copernican heliocentric model of the solar system (where the sun, rather than the earth, is the center of the solar system). These discoveries, and their subsequent impact on science and society, would not have been possible without the aid of the telescope – a device that serves to transform the “unseeable” into the “seeable.”

Modern scientific visualization plays a similarly significant role in contemporary science. Visualization is the transformation of abstract data, whether it be observed, simulated, or both, into readily comprehensible images (see Figure 1). Like the telescope and other modern instruments, visualization has proven to be an indispensable part of the scientific discovery process in virtually all fields of study. It is largely accepted that the term “scientific visualization” was coined in the landmark 1987 report [2] that offered a glimpse into the important role visualization could play in scientific discovery.

Visualization research and application efforts tend to fall into one of three primary use modalities. The first is “exploration visualization” where you have no idea what you’re looking for. This type of use model is typically the most challenging, since its success relies on interactive, “random-access” exploration of large and complex datasets. Another use model is “analytical visualization” where you know what you are looking for in data. Analytical visualization techniques are often those reduced into optimal practice after being established during exploratory visualization. Finally, “presentation visualization” is where you wish to convey a specific concept to others.

In the present day, the U.S. Department of Energy has a significant investment in many science programs. Some of these programs, carried out under the Scientific Discovery through Advanced Computing (SciDAC) program, aim to study scientific phenomena via simulation on the worlds largest computer systems. These new scientific simulations, which are being carried out on fractional-petascale sized machines today, generate vast amounts of output data. Managing and gaining insight from such data is widely accepted as one of the bottlenecks in contemporary science. As a result, DOE’s SciDAC program includes efforts aimed at addressing data management and knowledge discovery to complement the computational science efforts.

The focus of this article is on how one group of researchers – the DOE SciDAC Visualization and Analytics Center for Enabling Technologies (VACET) – is tackling the daunting task of enabling knowledge discovery through visualization and analytics on some of the world’s most large and complex datasets and on some of the world’s largest computational platforms. As a Center for Enabling Technology, our mission is to provide production-quality visualization and knowledge discovery software infrastructure that runs on large, parallel computer systems at DOE’s open computing facilities to solve challenging visual data exploration and knowledge discovery needs of modern science, particularly the DOE science community.

In this article, we present a small subset of our first-year accomplishments from VACET that span a broad cross-section of our technical portfolio. These accomplishments reflect an effective balance between forward-looking, award-winning research (Section 3) and the software engineering needed to realize effective production-quality visual data analysis software infrastructure (Section 4).

3 Forward-Looking Research

Like many other fields of computer and computational science, visualization has felt the “stress” created by an ever-increasing amount of data produced by simulations and experiments. Specifically, algorithms and implementations that have proven effective when applied to “modest” data size and

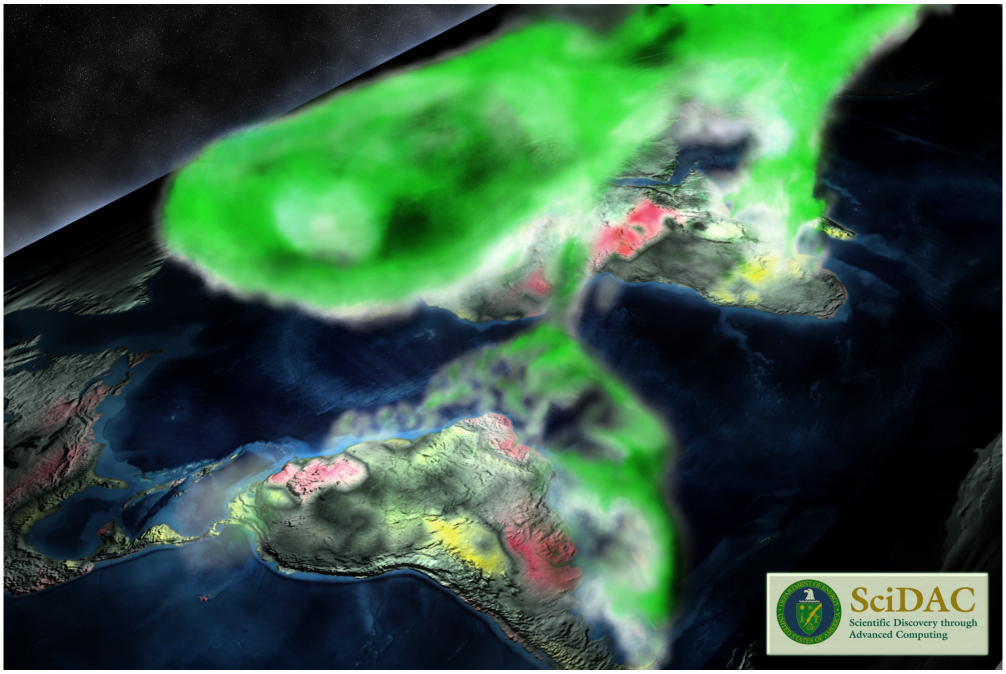


Figure 1: Visualization offers the ability to “see the unseeable.” The image shows the component of the atmospheric CO_2 concentration that results from the net ecosystem exchange (NEE), which is shown on the land surface. This “green CO2” is the flux due to the respiration of vegetation, respiration of soil microbes, and fire minus that taken up by ecosystem production. (Sample data courtesy Warren Washington, NCAR, John Drake, ORNL, and and Forrest Hoffman, ORNL; image courtesy J. Daniel, ORNL).

complexity will likely not prove adequate on contemporary and future datasets. It is generally agreed that the field of visualization is on the verge of a “phase change:” visualization is typically viewed as a way to present scientific results. However, visualization is much more than “pretty pictures:” the real potential lies in the integration of interactive visual exploration and visual analysis with exceptional computational capacity and the spectacular capabilities of the human mind and cognitive system.

While a survey of research topics is beyond the scope of this article, the research summaries that follow represent efforts targeted at a number of interrelated challenges faced by contemporary science. First, we present recent work in the area of flow field visualization (Section 3.1) that focuses on computing and displaying structure. Such techniques represent an important advance in the ability to extract and convey knowledge about complex data. This work has general applicability to a diverse range of science projects. Second, we present recent work that uses a rigorous mathematical foundation as the basis for quantitative and comparative analysis of complex data (Section 3.2). The idea here is to characterize, identify and analyze features using topology. This approach has great potential to accelerate scientific knowledge discovery as a technique that supplants a common yet imprecise comparative analysis technique known as “chi-by-eye.” A third research area addresses an orthogonal set of challenges, namely simplifying creation, modification and reuse of visualization technologies (Section 3.3). This award-winning work takes the first steps towards reducing the amount of labor and specialized knowledge required for creating visualization “pipelines.”

3.1 Efficient Computation of Coherent Structures in Fluid Flow Applications

The effective analysis of simulated flow data is required in many scientific problems ranging from fluid dynamics and magneto-hydrodynamics to climate and combustion research. As the size and complexity of the corresponding vector fields are growing, the efficient extraction of their salient structures becomes essential. The notion of Finite-Time Lyapunov Exponents (FTLE) provides a sound theoretical framework to characterize Coherent Lagrangian Structures in transient flows. Despite its conceptual simplicity, the associated computational cost is essentially prohibitive. We have developed a novel approach for adaptive computation of FTLE fields in two and three dimensions that significantly reduces the computational cost. Furthermore, for three-dimensional flows, we show that meaningful results can be obtained by restricting the analysis to a well-chosen plane. Moreover, we examine some of the visualization aspects of FTLE and introduce several new methods that benefit the analysis of specific aspects of challenging datasets.

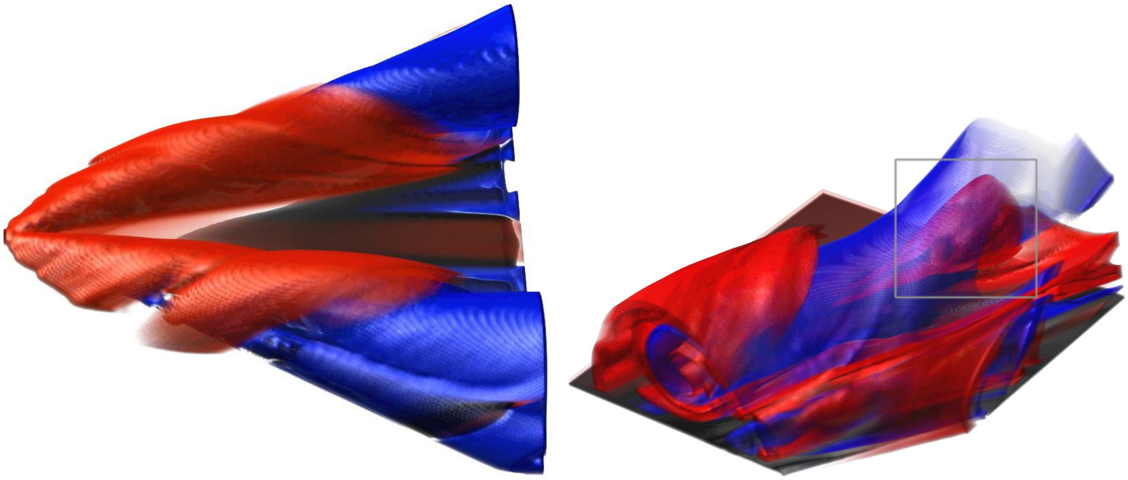


Figure 2: Volume rendering of regions of high forward (red) and backward (blue) FTLE above a delta wing. Wing edge separation and the primary attachment layer (left), with inner structures occluded. The interplay of separation and attachment structures is visible on the front face (right). The gray box highlights the separation structure that characterizes a vortex breakdown bubble. (Sample data courtesy M. Rütten, German Aerospace Center; image courtesy C. Garth, UC Davis.)

To address the challenge raised by the size and the qualitative complexity of flow vector fields resulting from modern CFD computations, scientific visualization research has explored different approaches that have in common the goal of characterizing, extracting, and visually representing salient flow structures across spatial and temporal scales (Figure 2). These methods are mainly divided into topological and feature-based approaches. While the former leverages a sound mathematical framework and allows for an objective and fully automatic post-processing, the latter explicitly integrates practically significant flow structures into the analysis at the cost of ambiguous definitions and ad-hoc methods.

In this context, the notion of Lagrangian Coherent Structures (LCS) and its quantitative assessment using the so-called Finite-Time Lyapunov Exponent (FTLE) provide a promising alternative that combines a well-articulated theoretical basis with physical intuition. Specifically, coherency in steady and transient flows can be characterized in terms of repelling and attracting manifolds. Despite the versatility and consistency of this approach, its practical application is fundamentally

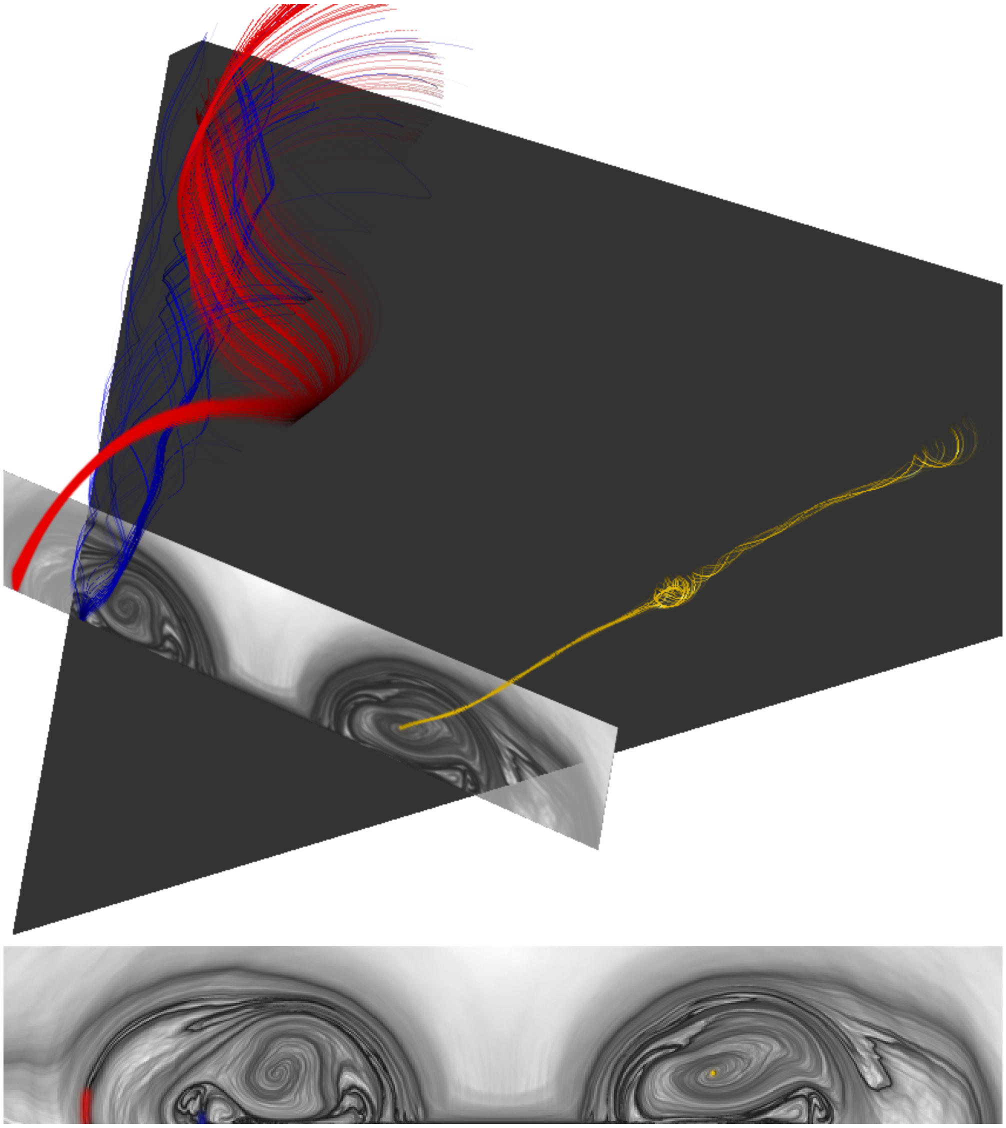


Figure 3: Pathlines are seeded according to FTLE strength, FTLE ridge lines or via a user-guided probability density function (PDF). The top image shows visualization of primary (red) and secondary (blue) separation structures; pathlines seeded according to the PDF shown in the lower image. The lower image shows the planar FTLE visualization on a section plane perpendicular to the main flow direction. Darker regions correspond to regions of high FTLE. Colored regions indicate PDF's used to see the pathlines in the upper right image. (Sample data courtesy M. Rütten, German Aerospace Center; image courtesy C. Garth, UC Davis.)

hampered by a prohibitive computational cost associated with the required advection of a dense set of particles across the spatio-temporal flow domain.

Our work has resulted in three primary contributions to the field of visual data analysis. First, we lower this computational cost by significantly reducing the number of particle advections required to perform visualization and analysis based on FTLE and LCS. We developed an incremental, data-driven refinement algorithm that exploits the coherence of neighboring particle paths to generate smooth approximations of the so-called flow map from which the FTLE is computed. This approach enables high-resolution analysis of complex 4D flows and permits to construct insightful visualization for accurate assessment of coherence. Second, we show that it is often not necessary to perform a full 3D analysis: given limited problem-specific knowledge about the flow field, it is often sufficient and in some cases even beneficial to consider FTLE on 2D subsections, further reducing compute time (Figure 3). Third, we demonstrate several new visualization methodologies based upon these new techniques on data from large-scale CFD simulations.

3.2 Topologically Based Feature Detection, Tracking and Quantitative Analysis

When a heavy fluid is placed above a light fluid, tiny vertical perturbations in the interface create a characteristic structure of rising bubbles and falling spikes known as Rayleigh-Taylor instability. Rayleigh-Taylor instabilities have received much attention over the past half-century because of their importance in understanding many natural and man-made phenomena, ranging from the rate of formation of heavy elements in supernovae to the design of capsules for inertial confinement fusion.

We have developed a new, robust method for quantitative analysis of Rayleigh-Taylor instabilities in which we extract a hierarchical segmentation of the mixing envelope surface to identify bubbles and analyze analogous segmentations of fields on the original interface plane. This approach is based on a family of robust topological techniques that enable multi-scale segmentation of scientific data for feature extraction and error-bounded quantitative analysis.

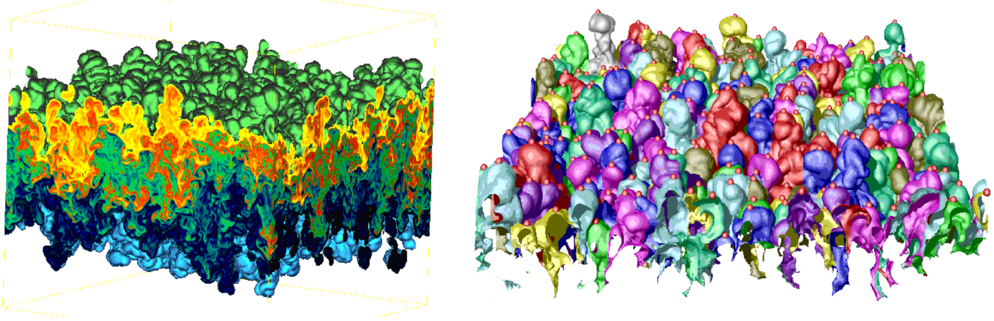


Figure 4: Turbulent mixing interface of a Rayleigh-Taylor instability simulation (left). Topological segmentation of the upper envelope highlighting in different colors the bubbles that rise during the mixing process (right). (Sample data courtesy A. Cook and W. Cabot, LLNL, SciDAC Science Application “Simulation of Turbulent Flows Hit with Strong Shockwaves; image courtesy V. Pascucci, LLNL.)

To overcome the challenge of analyzing the complex topology of the Rayleigh-Taylor mixing layer, we have developed a novel approach based on robust Morse theoretical techniques. This approach systematically segments the envelope of the mixing interface into bubble structures (Figure 4) and represents them with a new multi-resolution model allowing for the first time a multiscale quantitative analysis of the rate of mixing based on bubble count. The analysis highlighted and provided precise measures for four fundamental stages in the turbulent mixing process that previously scientists could only observe qualitatively, therefore enabling new insights and deeper understanding in this fundamental phenomenon (Figure 5).

This work has been documented in a paper, which was the winner of the Best Application Paper award at IEEE Visualization 2006 [3] and later presented at International Workshop on the Physics

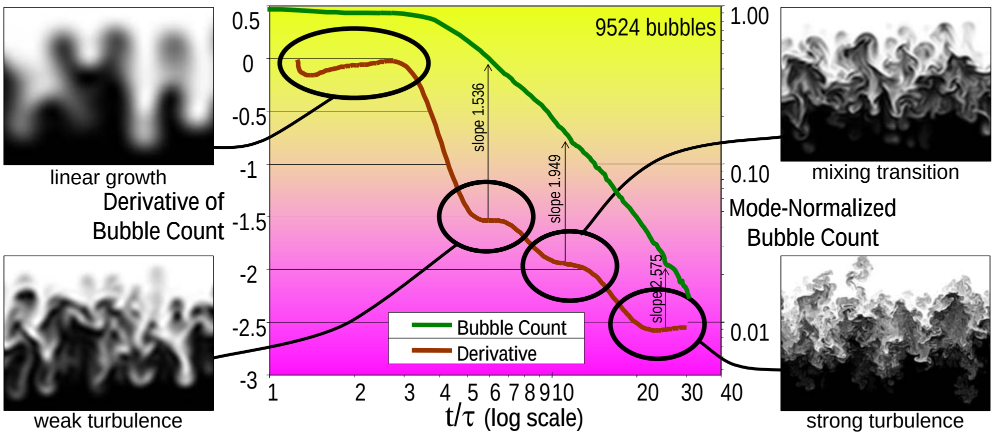


Figure 5: Time analysis of the bubble structures in the Rayleigh-Taylor mixing interface. Our approach both highlights qualitatively the four main stages of the process and quantifies the mixing rates characterizing each stage. (Sample data courtesy A. Cook and W. Cabot, LLNL, SciDAC Science Application "Simulation of Turbulent Flows Hit with Strong Shockwaves; image courtesy V. Pascucci, LLNL.)

of Compressible Turbulent Mixing [4]. Follow-up work enabled the first-ever direct comparison of two simulations based on different physics models and run with different initial conditions; the first run with one billion nodes over 758 time steps, the second run with 27 billion nodes over 220 time steps. Although comparison by superposition (e.g., "chi-by-eye") of the two simulations could not yield any meaningful result, the topological approach provided a quantitative multi-scale, feature based comparison highlighting fundamental similarities (Figure 6), which validated the lower resolution large eddy simulation (LES) with respect to the higher resolution direct numerical simulation (DNS).

3.3 Querying and Creating Visualizations by Analogy

While there have been advances in visualization systems, particularly in multi-view visualizations and visual exploration, the process of building visualizations remains a major bottleneck in data exploration. We show that provenance metadata collected during the creation of pipelines can be reused to suggest similar content in related visualizations and guide semi-automated changes. To enable the effective reuse of computational (visualization) pipelines, we introduce the idea of query-by-example in the context of an ensemble of visualizations, and the use of analogies as first-class operations in a system to guide scalable interactions. This work, which is part of VACET's forward-looking research portfolio, recently received the prestigious Best Paper award at IEEE Visualization 2007 [5].

Most visualization dataflow-based systems (e.g., AVS, SCIRun) have sophisticated user interfaces with visual programming capabilities that ease the creation of visualizations. Nonetheless, the path from "data to insight" requires a laborious, trial-and-error process, where users successively assemble, modify, and execute pipelines. In the course of exploratory studies, users often build large collections of visualizations, each of which helps in the understanding of a different aspect of their data. A scientist working on a computational fluid dynamics application might need different visualizations such as 3-D isosurface plots, 2-D plots, and direct volume rendering images.

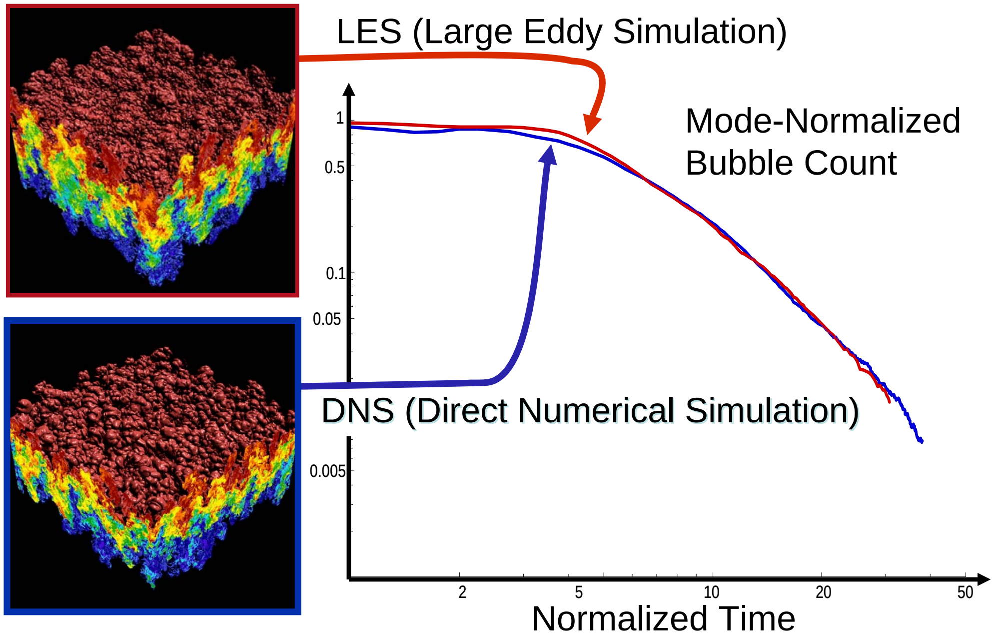


Figure 6: Feature-based comparison of two Rayleigh-Taylor instability calculations based upon different models and run with different initial conditions. The results shown here validate the Large Eddy Simulation (top) with respect to the Direct Numerical Simulation (bottom). (Sample data courtesy A. Cook and W. Cabot, LLNL, SciDAC Science Application "Simulation of Turbulent Flows Hit with Strong Shockwaves"; image courtesy D. Laney, LLNL.)

Although in general each of these visualizations is implemented in a separate dataflow, they have a certain amount of overlap (e.g., they manipulate the same input data sets). Furthermore, for a particular class of visualizations, the scientists might generate several different versions of each individual dataflow while fine tuning visualization parameters or experimenting with different data sets.

VisTrails (www.vistrails.org) is an open source system which implements a provenance model that uniformly captures changes to pipeline and parameter values during the course of data exploration. This detailed history information, combined with a multi-view visualization interface, simplifies the exploration process. It allows users to navigate through a large number of visualizations, giving them the ability to return to previous versions of any given visualization, compare different pipelines and their results, and resume explorations where they left off.

This provenance information can also be used to simplify and partially automate the construction of new visualizations. We propose a new framework that enables the effective reuse of this knowledge to aid users in performing data exploration through visualization. The framework consists of two key components: an intuitive interface for querying dataflows and a novel mechanism for semi-automatically creating and refining visualizations by analogy. The query engine is exposed to the user through a query-by- example interface whereby users query dataflows through the same familiar interface they use to create the dataflows (Figure 7). This approach lets them easily search through a large number of visualizations and identify pipelines that satisfy user-defined criteria. While the query interface allows users to identify pipelines that are relevant for a particular task, the visualization by analogy component provides a mechanism for reusing these pipelines to construct

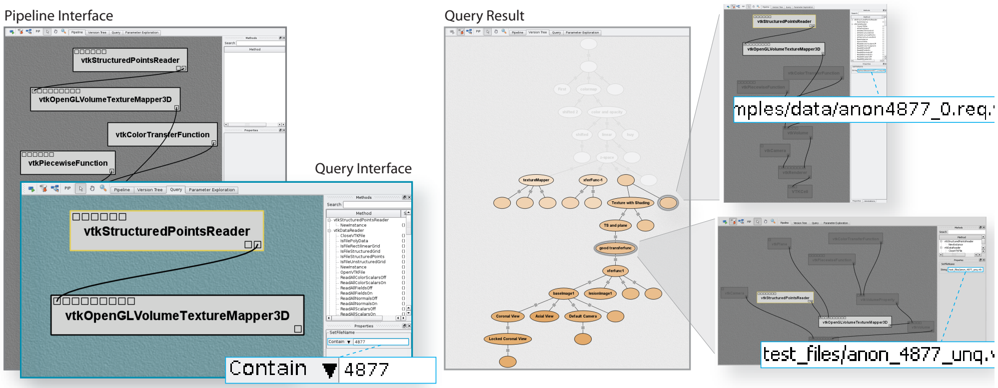


Figure 7: Querying by example. The interface for building a query over an ensemble of pipelines is the same as the one for constructing and updating pipelines. In fact, they work together: portions of a pipeline can become query templates by directly pasting them onto the Query Canvas. In this figure, the user is looking for a volume rendered image of a file whose name contains the string "4877." The system highlights the matches both at the visualization level (version tree, shown in the middle) and at the module level (shown in the right insets). (Image courtesy C. Scheidegger and C. Silva, University of Utah.)

new visualizations in a semi-automated manner – without requiring users to directly manipulate the pipeline specifications.

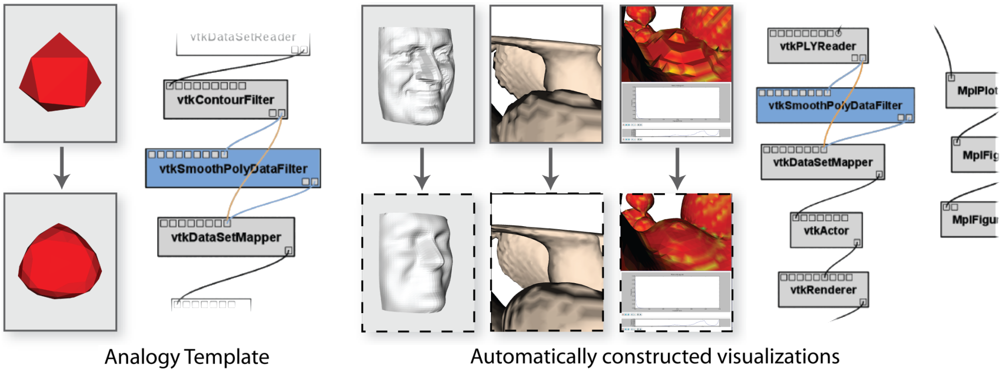


Figure 8: Visualization by analogy. The user chooses a pair of visualizations to serve as an analogy template. In this case, the pair represents a change where a file downloaded from the WWW is smoothed. Then, the user chooses a set of other visualizations that will be used to derive new visualizations. These new visualizations are derived automatically. The pipeline on the left reflects the original changes, and the one on the right reflects the changes when translated to the last visualization on the right. The pipeline pieces to be removed are portrayed in orange, and the ones to be added, in blue. Note that the surrounding modules do not match exactly: the system figures out the most likely match. (Image courtesy C. Scheidegger and C. Silva, University of Utah.)

To apply an analogy, VisTrails first determines the difference between a source pair of analogous visualizations, and then transfers this difference to a third visualization (Figure 8). The the user does not need to know the exact details of the dataflows to modify them. In addition, the analogy mechanism allows provides the basis for scalable updates: analogies can be automatically applied to a large number of pipelines simultaneously. Together, the ability to query visualization pipelines by example and to refine them by analogy are a step towards scalable pipeline development in visualization systems.

4 Production Software

A typical research project goes something like this: a researcher has an idea, or hypothesis. They set up an experimental methodology to test the hypothesis, publish results in a scientific journal, and then move on to the next project. In our world, "research" similarly is the genesis of new technology, but there is a lot more to the story. First, from whence originates the motivation for the hypothesis? Second, once a new idea is discovered, what then?

Our answers to both questions result from a look at how VACET does business. An abbreviated version of our business model is that we have forged long-term relationships with science stakeholders in which they tell us what kinds of knowledge they are looking for in their massive datasets as well as tell us about their methodology for hypothesis testing. With them, we identify the technologies needed to achieve such a capability. In some cases, we can adapt or extend existing technology, while in others, we must conceive new technology. Then, and here's where "the real action is," those ideas have to be reduced to practice – the production quality, petascale capable visual data analysis software.

As most in the software business would agree, this objective represents a formidable amount of software engineering work. We have adopted a low-risk, fast time-to-solution approach – build upon proven technology. To that end, our team uses two primary delivery vehicles that are described in the subsections that follow (Sections 4.1 and 4.2). Both are visualization applications that are the result of decades worth of research and development. By going this route, we can quickly add new capabilities to infrastructure that is production quality and petascale capable. This strategy has proven effective as we have delivered production-quality visual data analysis infrastructure for use on Adaptive Mesh Refinement data (Section 4.3).

4.1 VisIt

VisIt is an open source, turnkey application for large scale simulated and experimental data sets. Its charter goes beyond pretty pictures; the application is an infrastructure for parallelized, general post-processing of extremely massive data sets. Target use cases include data exploration, comparative analysis, visual debugging, quantitative analysis, and presentation graphics.

The VisIt product delivers the efforts of many software developers in a single package. VisIt leverages several third party libraries: the Qt widget library for its user interface, the Python programming language for a command line interpreter, and the Visualization ToolKit (VTK) library for its data model and many of its visualization algorithms. On top of that, an additional fifty man-years worth of effort have been devoted to the development of VisIt itself. The VisIt-specific effort has largely been focused on parallelization for large data sets, user interface, implementing custom data analysis routines, addressing non-standard data models (such as adaptive refinement meshes (AMR) and mixed materials zones), and creating a robust overall product. VisIt consists over one and a half million lines of code, and its third party libraries have an additional million lines of code. It has been ported to Windows, Mac, and many UNIX variants, including AIX, IRIX, Solaris, Tru64, and, of course, Linux, including ports for SGI's Altix, Cray's XT4, and many commodity clusters.

The basic design is a client-server model, where the server is parallelized. The client-server aspect allows for effective visualization in a remote setting, while the parallelization of the server allows for the largest data sets to be processed reasonably interactively. The tool has been used to visualize many large data sets, including a twenty seven billion data point structured grid (see Figure 9), a one billion point particle simulation, and curvilinear, unstructured, and AMR meshes with hundreds of millions to billions of elements. The most common form of the server is as a stand alone process that reads in data from files. However, an alternate form exists where a simulation code can link in "lib-VisIt" and become itself the server, allowing for in situ visualization and analysis.

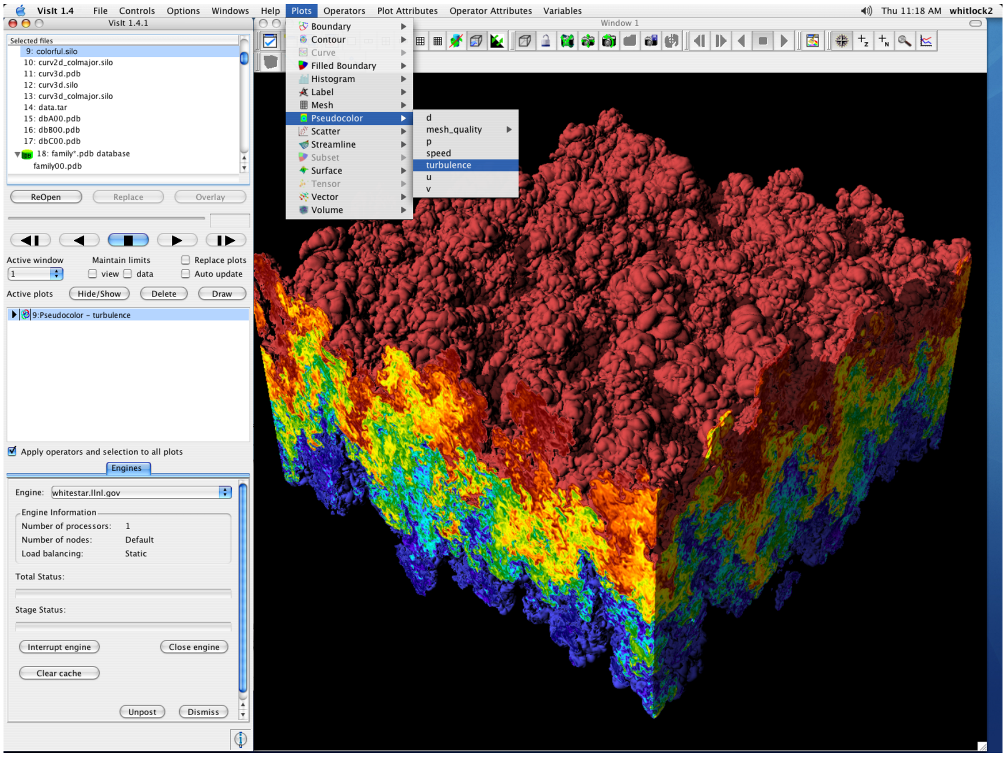
VisIt follows a data flow network paradigm where interoperable modules are connected to perform custom analysis. The modules come from VisIt's five primary user interface abstractions and there are many examples of each. There are twenty one "plots" (ways to render data), forty-two "operators" (ways to manipulate data), eighty-five file format readers, over fifty "queries" (ways to extract quantitative information), and over one hundred "expressions" (ways to create derived quantities). Further, a plugin capability allows for dynamic incorporation of new plot, operator, and database modules. These plugins can be partially code generated, even including automatic generation of Qt and Python user interfaces.

The VisIt project originated at Lawrence Livermore National Laboratory as part of the Advanced Simulation and Computing (ASC) program of the Department of Energy's (DOE) National Nuclear Security Agency, but it has gone on to become a distributed project being developed by several groups. Major hubs for the project come from VACET, ASC, and the Global Nuclear Energy Partnership (GNEP) from the DOE's Office of Nuclear Energy. The VisIt project has twenty developers from many organizations and universities, including five DOE Laboratories. VisIt received an R&D 100 award in 2005 and is downloaded approximately twenty five thousand times per year.

4.2 SCIRun

SCIRun is a scientific problem-solving environment (PSE) that allows interactive construction and steering of large-scale scientific computations [1]. A scientific application is constructed by connecting computational elements (modules) to form a program (network). The program may contain several computational elements as well as several visualization elements, all of which work together in orchestrating a solution to a scientific problem. SCIRun is designed to facilitate large-scale scientific computation and visualization on a wide range of architectures from the desktop to large supercomputers. Geometric inputs and computational parameters may be changed interactively, and the interface provides immediate feedback to the investigator.

SCIRun is being used to support the efforts of the SciDAC Center for Extended Magnetohydrodynamic Modeling in their analysis the instabilities of magnetic fields that confine the burning plasma. Within SCIRun, tools have been developed to rapidly create and analyze Poincaré plots that show the behavior of the magnetic fieldlines which have a periodic or quasi-periodic behavior as shown in Figure 10. SCIRun is also being used to produce query-driven visualization of particle-in-cell simulations that are part of the SciDAC Center for Gyrokinetic Particle Simulations of Turbulent Transport in Burning Plasmas. Here, physicists are interesting in analyzing just a few out millions of particles that contribute to turbulent transport. The query-driven aspects of SCIRun allow physicists to isolate and visualize these "trapped" particles over 100 hundreds of time steps as shown in Figure 11.



Selected files
9: colorful.silo
10: curv2d_colmajor.silo
11: curv3d.pdb
12: curv3d.silo
13: curv3d_colmajor.silo
14: data.tar
15: dbA00.pdb
16: dbB00.pdb
17: dbC00.pdb
18: family*.pdb database
family00.pdb

Active window: 1
Maintain limits ☐ view ☐ data
☐ Replace plots
☐ Auto update

Active plots
9:Pseudocolor – turbulence

☑ Apply operators and selection to all plots

Engine: whitestar.llnl.gov

Engine Information
Number of processors: 1
Number of nodes: Default
Load balancing: Static

Total Status:

Stage Status:

Figure 9: VisIt was used to visualize this 27-billion element Rayleigh Taylor instability, which simulates the mixing of heavy and light fluids. This simulation was run by the MIRANDA code on the NNSA BG/L machine. (Image courtesy H. Childs, LLNL.)

4.3 Adaptive Mesh Refinement Visual Data Analysis

Adaptive Mesh Refinement (AMR) is a highly effective simulation method for spanning a large range of spatiotemporal scales, such as astrophysical simulations that must accommodate ranges from interstellar to sub-planetary. Most mainstream visualization tools still lack support for AMR as a first class data type and AMR code teams use custom-built applications for AMR visualization. VACET has provided significant enhancements to one of its technology pillars (VisIt, which is described in more detail in Section 4.1) to provide the kind of production-quality, parallel-capable AMR visual data analysis infrastructure needed by SciDAC scientists that use AMR-based simulations. As a result, at least one team (the Applied Partial Differential Equations Center) has migrated to this new platform for most of its day-to-day work, thereby realizing a substantial cost savings: they no longer need to expend their own effort towards developing and maintaining AMR-capable visual data analysis software.

Adaptive Mesh Refinement (AMR) techniques combine the compact, implicitly specified structure of regular, rectilinear with the adaptivity to changes in scale of unstructured grids. Handling AMR data for visualization is challenging, since coarser information in regions covered by finer patches is superseded and replaced with information from these finer patches. During visualization, it becomes necessary to manage selection of which resolutions are being used for any given visualization operation. Furthermore, it is difficult to avoid discontinuities at level boundaries, which, if not properly handled, lead to visible artifacts in visualizations. Due to these difficulties, AMR support as first class data type in production visualization tools has been lacking despite the

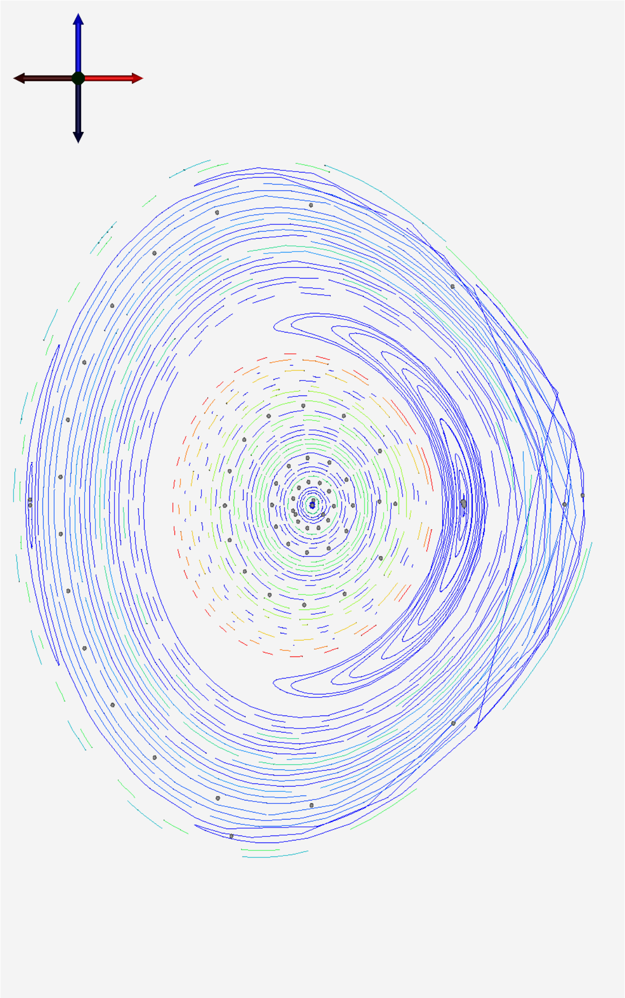


Figure 10: A Poincaré plot of magnetic field instabilities from a MHD simulation of a tokamak simulation. The instabilities appear in this image as a large “banana-shaped” region located in the middle of this cross section. (Sample data courtesy S. Kruger, Tech-X Corporation and the SciDAC Center for Extended Magnetohydrodynamic Modeling; image courtesy Scientific Computing and Imaging Institute, University of Utah.)

growing popularity and usefulness of AMR simulations.

VisIt (Section 4.1) accommodates AMR as first class data type. It handles AMR data as a special case of “ghost data,” i.e., data that is used to make computations more efficient, but which is not considered to be part of the simulation result. VisIt tags cells in coarse patches that are available at finer resolution as ”ghost” cells, allowing AMR patches to retain their highly efficient native format as rectilinear grids. It offers a rich set of production-quality functions, like pseudocolor and volume rendering plots (Figure 12), for visualization and analysis of complex data sets on parallel platforms, making it an ideal candidate to replace specialized AMR visualization tools.

The majority of our work focused on implementing a set of essential debugging features offered by ChomboVis in VisIt. These efforts improve VisIt’s handling of AMR data, both in terms of interface and performance.

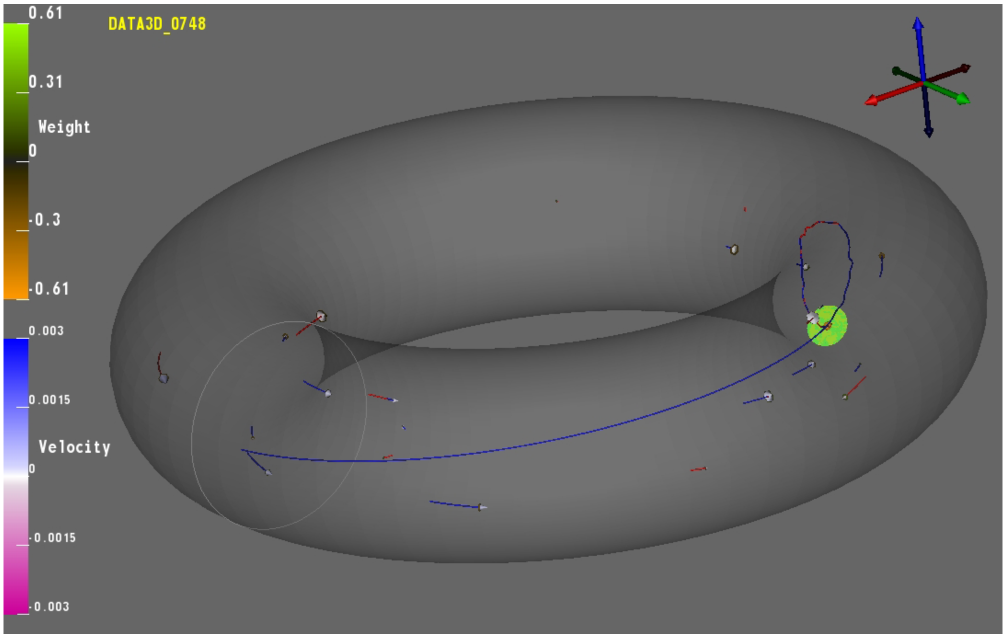


Figure 11: Only 22 of the 400 million particles in a particle in cell simulation are being displayed based on the number of times that they are magnetically trapped (red line) and de-trapped (blue line) in relation to the externally imposed magnetic field. One particle has been high lighted by electric potential that surrounds it throughout the simulation due to its extreme trajectory. (Sample data courtesy Stephane Ethier, Princeton Plasma Physics Laboratory, SciDAC Center for Gyrokinetic Particle Simulations of Turbulent Transport in Burning Plasmas; image courtesy Scientific Computing and Imaging Institute, University of Utah.)

4.4 Interface Enhancements

ChomboVis provides spreadsheet "plots" that support direct viewing of numerical values on a particular slice of a patch. This function is essential for debugging and used by AMR code development teams on a daily basis. We added these spreadsheets to VisIt as shown in Figure 13 and connected them to VisIt's "pick cell" feature allowing users to "link" them to other plots. We further added a capability to dynamically create new buttons in the VisIt interface to perform custom actions. This matches a capability that APDEC users valued in ChomboVis and allows new users to quickly navigate the tool. We also modified the VisIt selection routines to better support AMR data, allowing users to specify selections in terms of cell indices in a particular AMR level.

4.5 Performance Enhancements

We optimized handling of AMR grids in VisIt. These optimizations can save on memory by a factor of ten and also support more efficient rendering. Additional performance and memory optimizations improve efficiency for the important use case of rendering patch boundaries. VisIt previously used very general algorithm that was unnecessarily slow. Our new, specialized algorithm is an order of magnitude faster and more memory efficient.

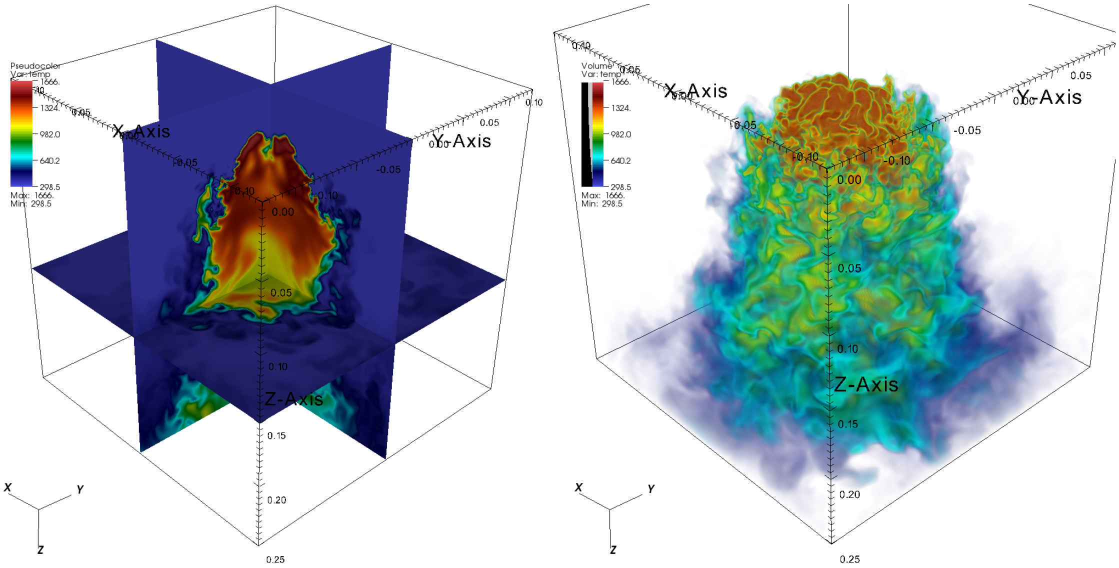
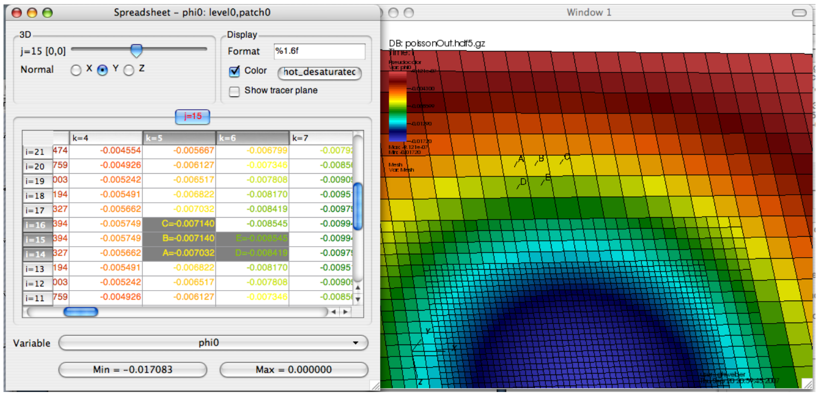


Figure 12: Production quality visualization of an AMR simulation of a hydrogen flame. The left panel shows a pseudocolor plot restricted to three axis-perpendicular slices. The right panel shows a volume rendered image of the same data. (Sample data courtesy of J. Bell and M. Day, Center for Computational Sciences and Engineering, LBNL; images courtesy G. Weber, LBNL.)



j=15 [0,0]

Normal ○ X ● Y ○ Z

Format %1.6f

☑ Color hot_desaturatec

☐ Show tracer plane

		k=4	k=5	k=6	k=7
i=21	474	-0.004554	-0.005667	-0.006799	-0.0079
i=20	759	-0.004926	-0.006127	-0.007346	-0.0085
i=19	003	-0.005242	-0.006517	-0.007808	-0.0090
i=18	194	-0.005491	-0.006822	-0.008170	-0.0095
i=17	327	-0.005662	-0.007032	-0.008419	-0.0097
i=16	394	-0.005749	C=-0.007140	-0.008545	-0.0099
i=15	394	-0.005749	B=-0.007140	E=-0.008545	-0.0099
i=14	327	-0.005662	A=-0.007032	D=-0.008419	-0.0097
i=13	194	-0.005491	-0.006822	-0.008170	-0.0095
i=12	003	-0.005242	-0.006517	-0.007808	-0.0090
i=11	759	-0.004926	-0.006127	-0.007346	-0.0085

Variable phi0

Min = -0.017083

Max = 0.000000

Figure 13: Spreadsheet plots are an important tool for debugging AMR codes. They support direct viewing of numerical data in patch cells. VisIt labels selected cells in both Spreadsheet and 3D visualizations allowing users to recognize correspondences quickly and effectively. (Sample data courtesy of P. Colella and B. van Straalen, LBNL. Image courtesy G. Weber, LBNL.)

5 Conclusion and Future Work

The SciDAC Visualization and Analytics Center for Enabling Technologies (VACET) is a highly productive effort combining the forces of leading visualization researcher from five different institutions to solve some of the most challenging data understanding problems in modern science. The VACET technology portfolio is diverse, spanning all typical visual data analysis use models and effectively balancing forward-looking research with focused software architecture and engineering that results in production quality software infrastructure. One of the key elements to VACET's success is a rich set of projects that are collaborations with science stakeholders: these efforts focus on identifying and overcoming obstacles to scientific knowledge discovery in modern, large and complex scientific datasets.

In its first year of operation, VACET has been highly productive. As recognized leaders in the field, we have generated over thirty peer-reviewed publications, approximately twenty-four of which are journal papers, including two award winning papers. Our work has had a measurable positive impact on DOE SciDAC science projects, including being adopted for day-to-day visual data analysis use thereby resulting in cost savings. Our science stakeholder projects a diverse range of SciDAC and other DOE science projects: accelerator modeling, astrophysics, climate, combustion, fusion, mathematics, and turbulence. Additionally, we have delivered approximately twenty-seven visualization presentations in the past year, some of these coordinated with the SciDAC Institute for Ultrascale Visualization, to help convey the message of visualization science to the broader scientific community.

You can read more about VACET on the web at www.vacet.org.

Acknowledgement

This work was supported by the Director, Office of Advanced Scientific Computing Research, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 through the Scientific Discovery through Advanced Computing (SciDAC) program's Visualization and Analytics Center for Enabling Technologies (VACET). This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] 2007. SCIRun: A Scientific Computing Problem Solving Environment, Scientific Computing and Imaging Institute (SCI).

[2] T. De Fanti et al. Special Issue on Visualization in Scientific Computing. *Computer Graphics*, 21(6), November 1987.

[3] D. Laney, A. Mascarenhas, P. Miller, P. T. Bremer, and V. Pascucci. Understanding the Structure of the Turbulent Mixing Layer in Hydrodynamic Instabilities. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1053–1060, 2006.

[4] P. Miller, P.-T. Bremer, W. Cabot, A. Cook, D. Laney, A. Mascarenhas, and V. Pascucci. Application of morse theory to analysis of rayleigh-taylor topology. In *10th International Workshop on the Physics of Compressible Turbulent Mixing*, 2006.

[5] Carlos Scheidegger, David Koop, Huy Vo, Juliana Freire, and Claudio Silva. Querying and Creating Visualizations by Analogy. *Transactions on Visualization and Computer Graphics, IEEE Visualization 2007*, 13(6):1560–1567, 2007.