

SciDAC Visualization and Analytics Center for Enabling Technology

E. Wes Bethel¹, Chris Johnson², Ken Joy³, Sean Ahern⁴, Valerio Pascucci⁵, Hank Childs⁵, Jonathan Cohen⁵, Mark Duchaineau⁵, Bernd Hamann³, Charles Hansen², Dan Laney⁵, Peter Lindstrom⁵, Jeremy Meredith⁴, George Ostrouchov⁴, Steven Parker², Claudio Silva², Allen Sanderson² and Xavier Tricoche²

1. Computing Sciences Division, Lawrence Berkeley National Laboratory, The University of California, Berkeley, CA 94720-8139, USA. 2. Scientific Computing and Imaging Institute, University of Utah, 72 S. Central Campus Drive, Salt Lake City, UT 84112, USA. 3. Institute for Data Analysis and Visualization, University of California, One Shields Avenue, Davis, CA 95616-8562 USA. 4. National Center for Computational Sciences, Oak Ridge National Laboratory, One Bethel Valley Road, P.O. Box 2008 MS-6016, Oak Ridge, TN 37831, USA. 5. Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA 94551 USA, USA. Website: www.vacet.org

Abstract. The Visualization and Analytics Center for Enabling Technologies (VACET) focuses on leveraging scientific visualization and analytics software technology as an enabling technology for increasing scientific productivity and insight. Advances in computational technology have resulted in an "information big bang," which in turn has created a significant data understanding challenge. This challenge is widely acknowledged to be one of the primary bottlenecks in contemporary science. The vision of VACET is to adapt, extend, create when necessary, and deploy visual data analysis solutions that are responsive to the needs of DOE's computational and experimental scientists. Our center is engineered to be directly responsive to those needs and to deliver solutions for use in DOE's large open computing facilities. The research and development directly target data understanding problems provided by our scientific application stakeholders. VACET draws from a diverse set of visualization technology ranging from production quality applications and application frameworks to state-of-the-art algorithms for visualization, analysis, analytics, data manipulation, and data management.

1. Introduction – VACET and Its Mission

The SciDAC Visualization and Analytics Center for Enabling Technologies (VACET) [1] focuses on leveraging scientific visualization and analytics software technology as an enabling technology for increasing scientific productivity and insight. Our mission is to foster scientific insight through creating and deploying effective data understanding technology that is truly responsive to the needs of our stakeholders in the scientific research community. Our stakeholders – researchers from experimental and computational sciences – are “awash in data” produced by an “information big bang” whereby their ability to collect and generate data far outpaces the ability to understand such data. This challenge is widely acknowledged to be one of the primary bottlenecks in contemporary science. The vision for our Center is to respond directly to this challenge by adapting, extending, creating when necessary and deploying visualization and data understanding technologies for our science stakeholders. As a Center for Enabling Technologies, one of VACET’s central goals is to bring cutting-edge, production quality visualization and analytics to bear on the data understanding challenges of the SciDAC scientific community in the form of well engineered, usable software. The remainder of this article highlights a few examples of how we are achieving this vision. First, we provide an overview of technical challenges that lie on the path towards petascale visual data analysis.

Second, we offer a view of how our Center's software infrastructure will be deployed. Third, we present an overview of recent visual data analysis accomplishments that span many different science applications.

2. Path to Petascale Visual Data Analysis

For some, the end result of a scientific visualization has been a colorful 3D image or animation. In many cases, a scientific researcher would “throw data over the fence” to the visualization scientist, who would then produce an image. This approach has resulted in very nice images and animations, but has not taken full value of the capabilities of visual analysis for scientific data understanding. As we move towards petascale and beyond, we are in the midst of a major paradigm shift in the field of visual data analysis. This shift reflects the realities of moving towards petascale: We can no longer think of visualization as entirely a post process, visual data analysis activities must be an integral part of the scientific knowledge discovery process. The growing data size/complexity confounds traditional visualization algorithms; new machine architectures and I/O loads require rethinking the visual data analysis pipeline.

Childs [2] describes a visual data analysis architecture that consists of modular software components that can operate in any of the four potential architectural approaches. This architecture, which forms the basis of the VisIt [3] software application, is in widespread use and runs on all modern desktop and HPC platforms. While this architecture appears to be sufficient for the present, it will likely require evolution to accommodate the demands data loads at the petascale and beyond.

VACET is adopting this architectural model for use in deploying its software infrastructure to scientific collaborators for use on HPC platforms, particularly at the DOE's open computing facilities (NERSC, NCCS). This architecture also serves as one of the primary platforms for conducting the applied research in visual data analysis that will provide solutions responsive to the needs of petascale applications. At the HPC centers, and within VACET in particular, we are attempting to provide solutions that go beyond traditional visualization. Visual analysis goes beyond simple generation of imagery and includes data exploration, visual code debugging, comparative analysis, quantitative analysis, and presentation graphics. Our goal is to deploy infrastructure for all of these, for full analysis, code comparison, and verification and validation. In this, we will need the assistance of application scientists, as physics experience will be required. Our codes will act as facilitators, providing parallel infrastructure, releases on many platforms, documentation, and support.

3. Example Collaborator Accomplishments

We now give examples of VACET visual data analysis with some of our SciDAC collaborators.

3.1. Accelerator Modeling

Plasma and laser wakefield-driven accelerators produce accelerating fields thousands of times those achievable in conventional radio-frequency accelerators, potentially extending the frontiers of high-energy physics and enabling laboratory scale ultrafast radiation sources. Because the plasma response is highly nonlinear, large-scale, self-consistent particle simulations in 3D are important to understand and optimize this new acceleration process. Codes that model particle-based physics, including VORPAL, Osiris, and GTC, are capable of extreme levels of parallelism and a correspondingly large amount of output data. The data understanding challenge in this case is to focus visual data analysis and presentation on data deemed to be interesting. Figure 1 shows particles that have been trapped in a laser wakefield simulation and that are undergoing wakefield acceleration. These particles are shown in conjunction with the plasma density field. (Data courtesy C. Geddes, LBNL; Image courtesy C. Siegrist, LBNL)

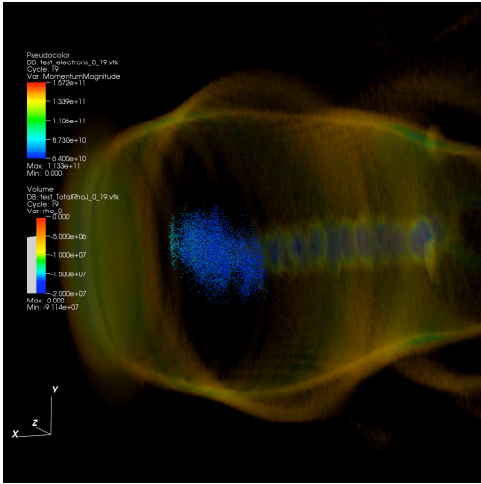


Figure 1. Laser wakefield simulation.

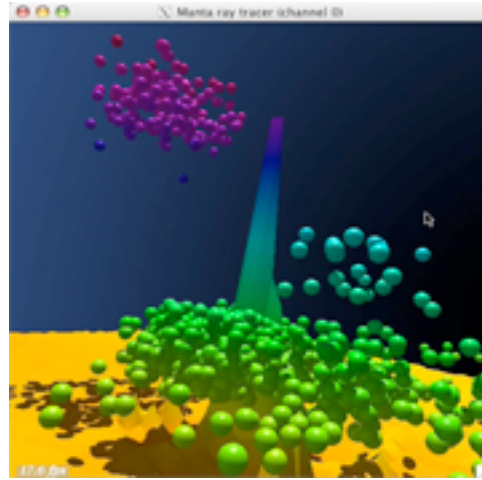
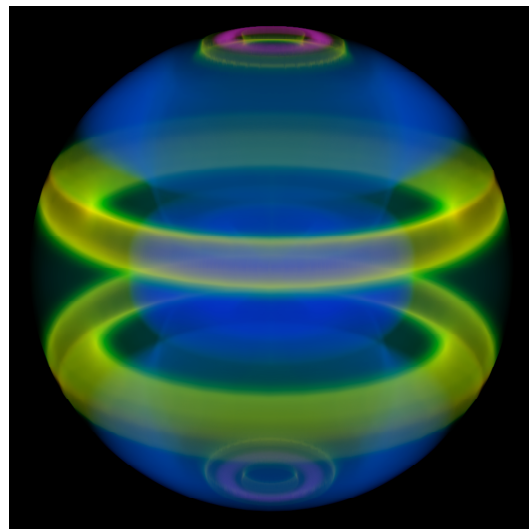
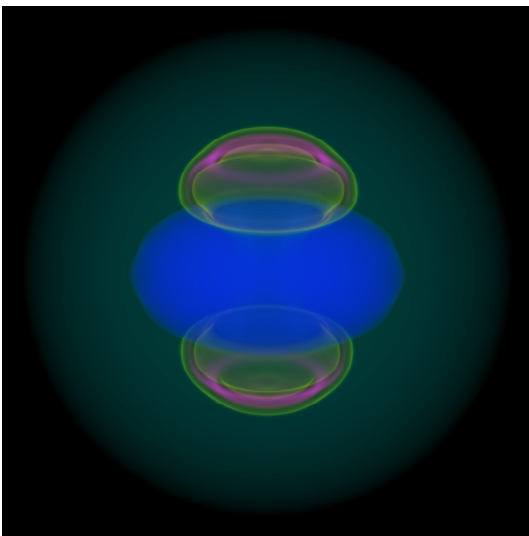


Figure 2. Acceleration simulation particle visualization.

An alternate presentation of wakefield-acceleration simulation output, Figure 2, shows density as a height field and individual particles whose velocity exceeds a user-specified threshold. In this image, we are exploring using high performance, interactive ray tracing technology to generate photorealistic images at interactive rates. Photorealism plays a key role in facilitating rapid comprehension of three-dimensional structure and depth relationships in complex scientific data (Data courtesy P. Hamill, Tech-X Corp., image courtesy S. Parker, Utah).

3.2. Astrophysics

The Computational Astrophysics Consortium aims to develop a full understanding, achieved through numerical simulation, of how supernovae of all types explode and how the elements have been created in nature. Validation of simulation results will occur through ongoing comparison with astronomical observations. These simulation codes will model physical, chemical, radiation and gravitational phenomena that occur at a vast range of scales, and will produce vast quantities of large and complex data. Early progress includes visualization results, shown in Figures 3 and 4, of a code that models radiation and hydrodynamic processes in a core-collapse supernova. This is the first step towards



Figures 3 and 4 show volume rendering of two time steps of the entropy field from a simulation modeling the radiation/hydrodynamic processes in a core collapse supernova explosion (Data courtesy A. Burrows, U. of Arizona; Image courtesy C. Siegerist, LBNL).

providing production-quality, parallel-capable visual data analysis infrastructure for one of several interrelated code teams.

3.3. Climate/Environment

In the area of climate modeling, the focus has recently shifted from studying isolated simulations of, for example, the atmosphere to trying to understand the intricate coupling between atmospheric and oceanographic chemistry and land and ocean ecological systems. The ultimate goal is to understand the dynamics of the global carbon cycle on time scales ranging from decades to centuries. To this end, VACET is collaborating with the Community Climate System Model (CCSM) Consortium and the Earth Science grid to develop and deploy advanced visualization and data analysis tool. One of VACET's technologies, VisTrails [6], was recently featured in a CDAT Newsletter [7].

3.4. Combustion

As part of DOE's INCITE program, combustion researchers have used the S3D code to perform three-dimensional Direct Numerical Simulation (DNS) of a turbulent, non-premixed $H_2/CO/N_2$ -air flame with detailed chemistry. VACET has provided a custom data reader to directly load S3D output data, thereby eliminating costly data conversion steps and accelerating time to discovery, as well as provide one-on-one consulting to help the DNS team to conduct their own visual data analysis.

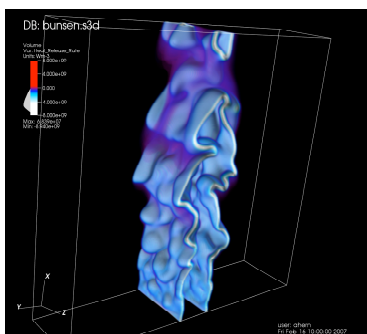


Figure 5. Heat release computed by S3D during Direct Numerical Simulation of a turbulent, non-premixed $H_2/CO/N_2$ -air flame with detailed chemistry. (Data courtesy of R Sankaran, ORNL and J Chen, SNL; Image courtesy of S Ahern, ORNL.

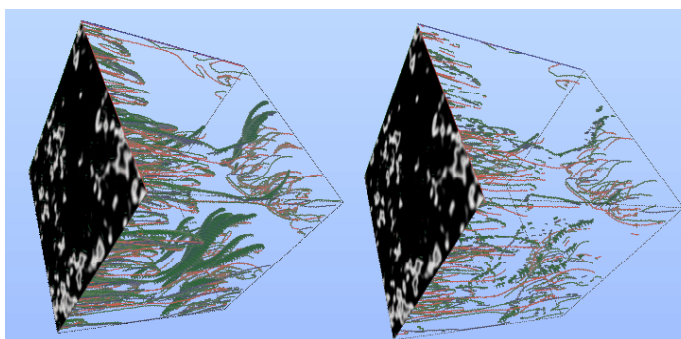


Figure 6. Feature identification and tracking over time. Here, the features consist of auto-ignition kernels that combine over time then undergo an extinction and reignition process. The count, spatial positioning and topological relationships of such features provide the basis for the type of quantitative analysis needed to support advances in scientific knowledge discovery (data courtesy J. Chen, SNL-CA, Image courtesy P.-T. Bremer, A. Mascarenhas, V. Pascucci, LLNL).

In recent work, we have developed a method for computing and displaying the correlation between variables in a multivariate dataset. This technique provides new insights into the relationships between variables not possible with standard visualization techniques.

3.5. Fusion

Computational fusion includes modeling many different regimes: wave heating and current drive, microturbulence and turbulent transport, magnetohydrodynamics and device stability, and energy transport across all regimes. VACET has interactions with fusion teams studying each of these areas, as well as teams focusing on integrated/coupled codes to model the entire device.

With respect to visual data analysis from the microturbulence studies, which use particle-in-cell codes, one significant challenge is to focus visual data analysis processing and presentation on the subset of large and complex data deemed to be interesting. One approach, which is being studied and applied by both the SciDAC VACET and the SciDAC Institute for Ultrascale Visualization, uses

multidimensional data range selection using a technique known as “parallel coordinates” to select a subset of data deemed to be interesting. Here, particles that satisfy a given weight criteria seem to follow magnetic field line structure in this particular example.

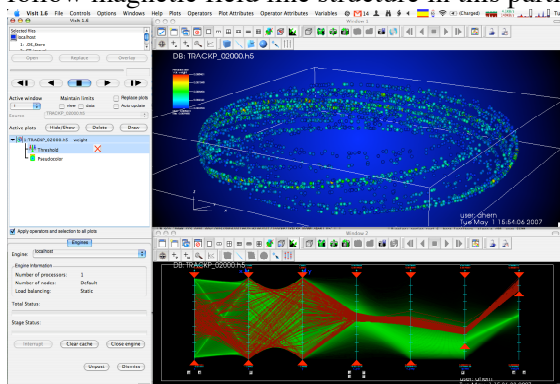


Figure 8. The extraction of particles of interest from a GTC PIC simulation by way of multidimensional data filtering, parallel coordinates display, and linked views in the VisIt visualization system. (Data courtesy of S. Ethier, PPPL; image courtesy of S. Ahern, ORNL).

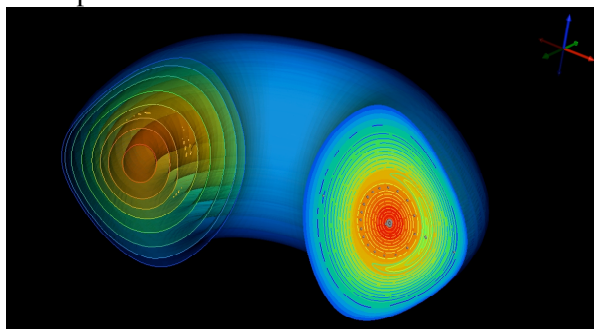


Figure 9 shows a visualization of a silo fusion MHD data set showing a series of isotemperature surfaces with a Poincaré plot of the magnetic field overlaid using SCIRun [8]. The analysis of the Poincaré plot located a set of magnetic islands with their “O” points shown with a gray sphere. (Data courtesy R. Sanchez ORNL; Image courtesy A. Sanderson, Utah).

In the development of magnetic fusion reactors, physicists need the ability to analyze the magnetic fields that confine the burning plasma. The corresponding magnetic fieldlines have a periodic or quasi-periodic behavior. We have developed a novel, highly efficient algorithm for rapid computation and display of Poincaré plots that has been deployed to our fusion collaborators.

3.6. Applied Mathematics/AMR

The APDEC center deploys software that solves multi-scale and multi-physics problems based on finite difference and finite-volume methods on logically-rectangular structured grids combined with block structured adaptive mesh refinement (AMR) to represent multi-scale behavior and then applies these techniques to application areas such as astrophysics, combustion, and fusion. This creates multiple visualization and analysis challenges. The goal is to extend VisIt's AMR handling so that it will be suitable for APDEC stakeholders, providing the techniques they need, even for very large scale simulations. Early progress includes extending VisIt to provide the techniques that AMR code developers expect.

References

- [1] VACET web site: <http://www.vacet.org>.
- [2] Childs, H. Architectural Challenges and Solutions for Petascale Postprocessing. In this Proceedings.
- [3] VisIt web site: <http://www.llnl.gov/visit>.
- [4] Bell G, Gray J and Szalay A, Petascale computational systems. IEEE Computer, 39(1):pp. 110–112).
- [5] http://www.scidac.gov/Conference2007/speaker_bios/geddes_abs.html.
- [6] VisTrails web site: <http://www.vistrails.org>.
- [7] <http://www-pcmidi.llnl.gov/software-portal/Newsletter/Vol3/news.html>.
- [8] SCIRun: A Scientific Computing Problem Solving Environment, Scientific Computing and Imaging Institute (SCI), <http://software.sci.utah.edu/scirun.html>.