NEXT-GENERATION Visualization Technologies: Enabling Discoveries at EXTREME Scale

The SciDAC Institute for Ultra-Scale Visualization aims to enable extreme-scale knowledge discovery by advancing the state of visualization technologies, fostering awareness of and communication about new visualization technologies, and placing these technologies into the hands of application scientists.

The development of supercomputers transformed the pursuit of science. With supercomputing's rapid growth in capability and capacity, scientists expect to study problems of unprecedented complexity with exceptional fidelity and to address many new problems for the first time. However, the size and complexity of the data output of such large-scale simulations present tremendous challenges to the subsequent data visualization and analysis tasks; they create a growing gap between the scientists' ability to simulate complex physics at high resolution and their ability to extract knowledge from the resulting massive datasets. The Institute for Ultra-Scale Visualization (Ultravis Institute), funded by the U.S. Department of Energy's (DOE) SciDAC program, will close this gap by leading community efforts to develop advanced visualization technologies that enable knowledge discovery at extreme scale.
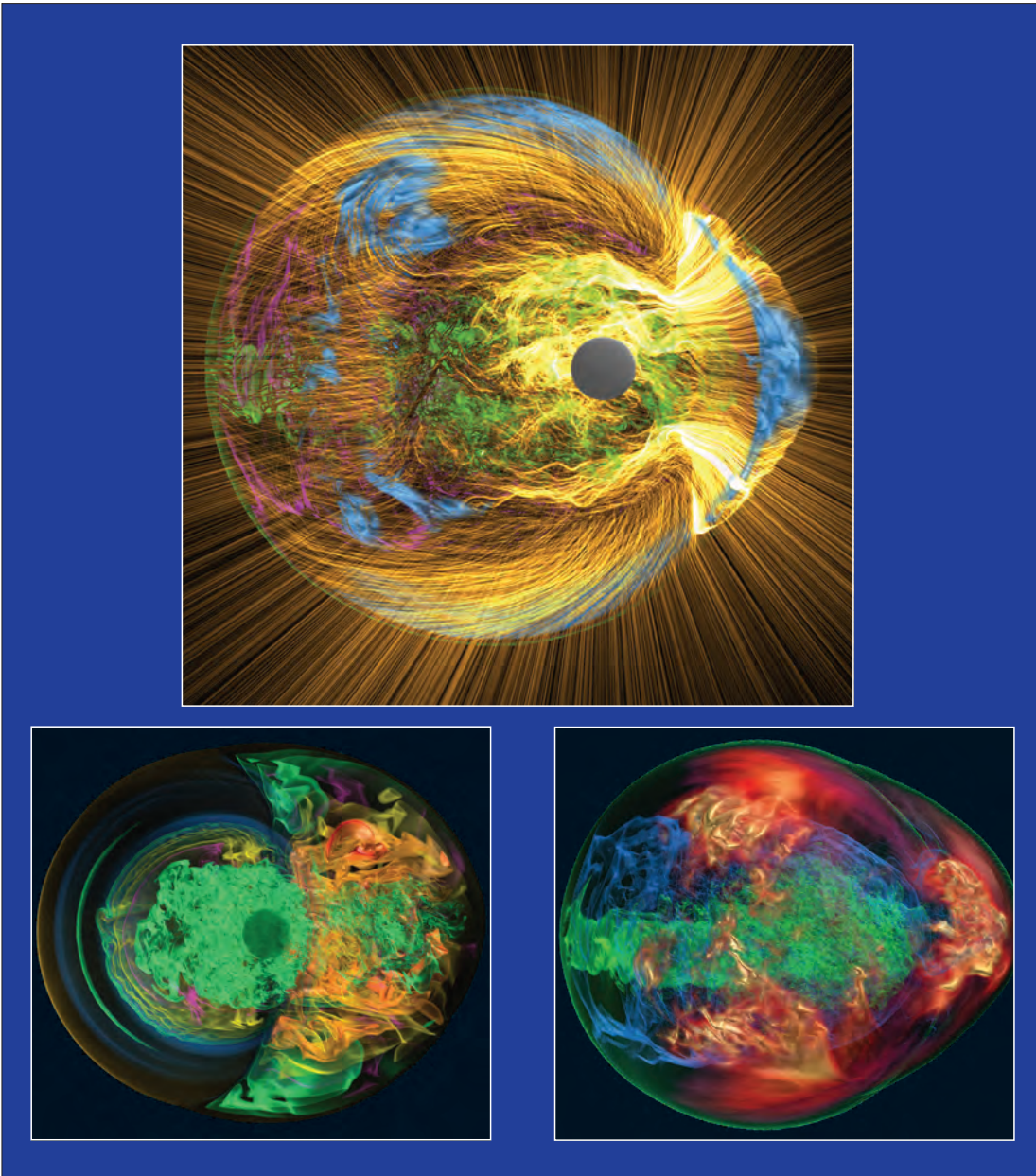
Visualization—the creation of vivid pictures from simulation output in the form of arrays of numbers—has become an indispensable tool for scientists. But visualization is not without its challenges, especially considering the time-varying, multivariate aspect of simulation output. We have identified and pursued several high-priority visualization research topics that will provide the greatest benefit to the SciDAC community as well as the broader area of scientific supercomputing and discovery. These topics include application-driven visualization, multivariate and multidimensional exploration, complex adaptive-grid data visualization, parallel visualization, *in situ* visualization, and next-generation visualization tools.

Researchers at North Carolina State University have conducted the highest-resolution supernova simulations to date. They used the DOE Leadership Computing Facility to study the development of a rotational instability in a supernova shockwave when it was only a fraction of a second old, deep in the core of the dying star. The hundred terabytes of data output by each run of the simulation contained the greatest details and dynamic features ever modeled. The advanced visualization techniques developed by the Ultravis Institute made it possible to completely and

The size and complexity of the data output of large-scale simulations present tremendous challenges to the subsequent data visualization and analysis tasks
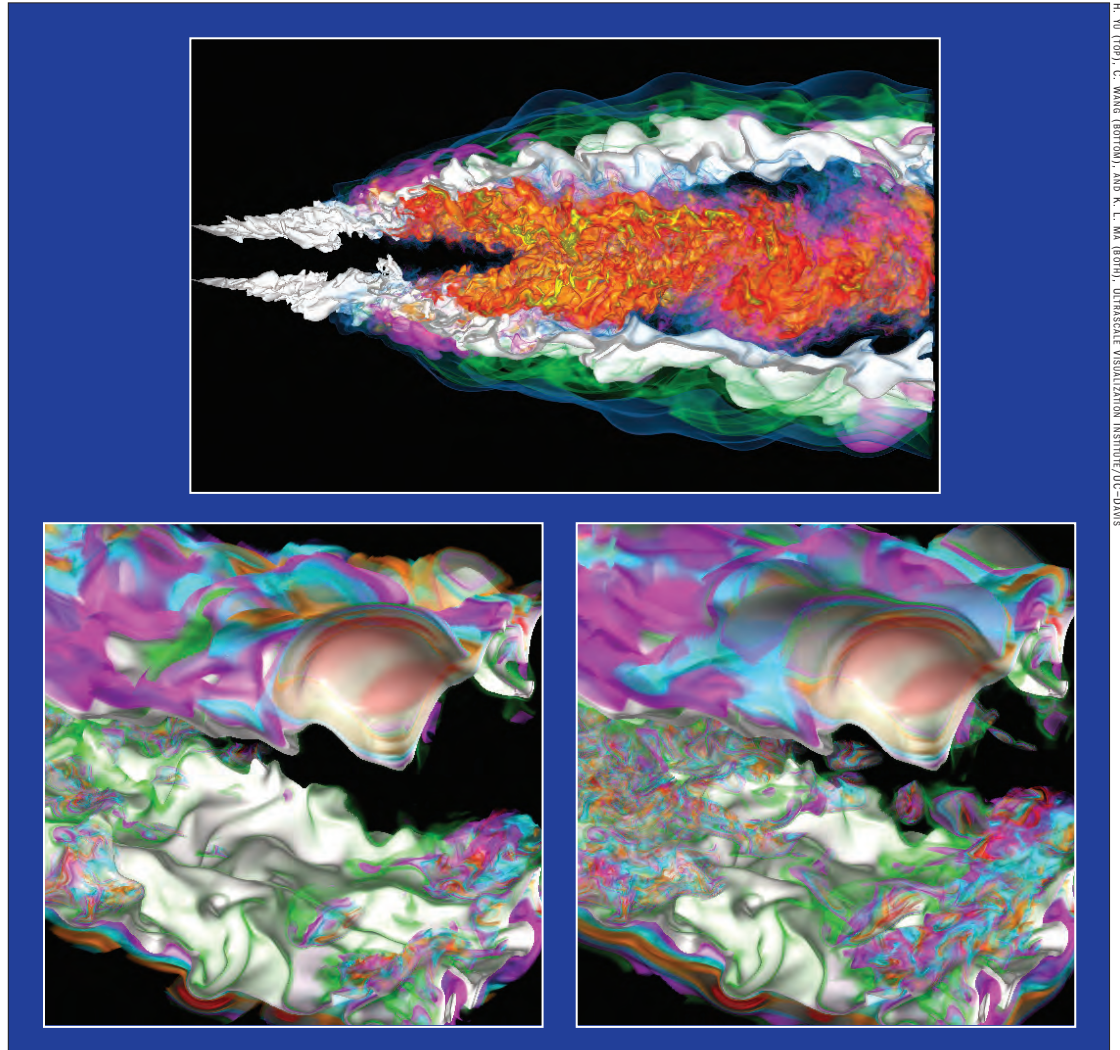
Figure 1. *Top, simultaneous visualization of pathlines and the angular momentum field obtained from a supernova simulation performed by Dr. John Blondin. Bottom, volume rendering of the entropy (left) and angular momentum (right) fields. New parallel and graphics processing unit (GPU)-accelerated rendering techniques enable interactive visualization at levels of detail and clarity previously unattainable.*

vividly reveal the details and actions hidden in the arrays of data, as shown by the images in figure 1. The initially spherical supernova shock develops a wave pattern in the counter-clockwise direction. This spiraling wave gathers more angular momentum as it grows. The flow tracers in the top image illustrate the deflection of the radially infalling gas by the distorted accretion shock, feeding a rotational flow pattern around the neutron star at the center of the supernova. This shock instability helps drive the supernova shock out of the core and leads to the spin-up of the neutron star left behind.

Application-Driven Visualization

A common approach to visualizing a large amount of data is to reduce the quantity sent to the graphics pipeline by using a multiresolution data representation coupled with a compression method such as quantization or transform-based compression. This approach, however, does not explicitly take domain knowledge or visualization tasks into account for data representation and reduction. In particular, compressing high-precision floating-point data based solely on values can only result in limited savings. Further reduction is possible by using the typicality that only a small subset of the

We have attempted to directly incorporate domain knowledge and tasks into the whole data reduction, compression, and rendering process.

Figure 2. *Top, simultaneous visualization of mixture fraction and HO_2 data obtained from a turbulent combustion simulation performed by Dr. Jackie Chen. Bottom, visualization of the same data using a distance field representation of HO_2. The left image depicts the HO_2 distribution near the feature isosurface (in white). The right image shows a wider coverage of HO_2 around the feature surface. This application-driven approach not only allows scientists to see previously hidden features but also enables more aggressive encoding of the data according to the distance field, leading to greater storage saving.*

data is of interest in analysis. We advocate an application-driven approach to compressing and rendering large-scale, time-varying scientific simulation data. Most scientists have certain domain knowledge and specific visualization tasks in their minds. In the context of time-varying, multivariate volume data visualization, such knowledge could be the salient isosurface of interest for some variable, and the visualization task could be observing spatio–temporal relationships among several other variables in the neighborhood of that isosurface. We have attempted to directly incorporate such domain knowledge and tasks into the whole data reduction, compression, and rendering process.

Scientists at Sandia National Laboratories are able to perform three-dimensional (3D) fully-resolved direct numerical simulation of turbulent combustion. With full access to the spatially and temporally resolved fields, direct numerical simulation plays a major role in establishing a fundamental understanding of the microphysics of turbulence–chemistry interactions. For example, to understand the dynamic mechanisms of extinction and re-ignition in turbulent flames, scientists need to validate known relationships and reveal hidden ones among multiple variables. In the study of a lifted turbulent jet, the scientific interest is twofold: the overall structure of the jet flame and the lifted flame base region to understand what is stabilizing the lifted flame. More specifically, the visualization task is to show a particular stoichiometric mixture fraction isosurface together with another field, such as HO_2 (hydroperoxy radical) or OH (hydroxy radical),

Unstructured grids and adaptive mesh refinement methods are increasingly used in large-scale scientific computations for the modeling of problems involving complex geometries and phenomena at varying scales.
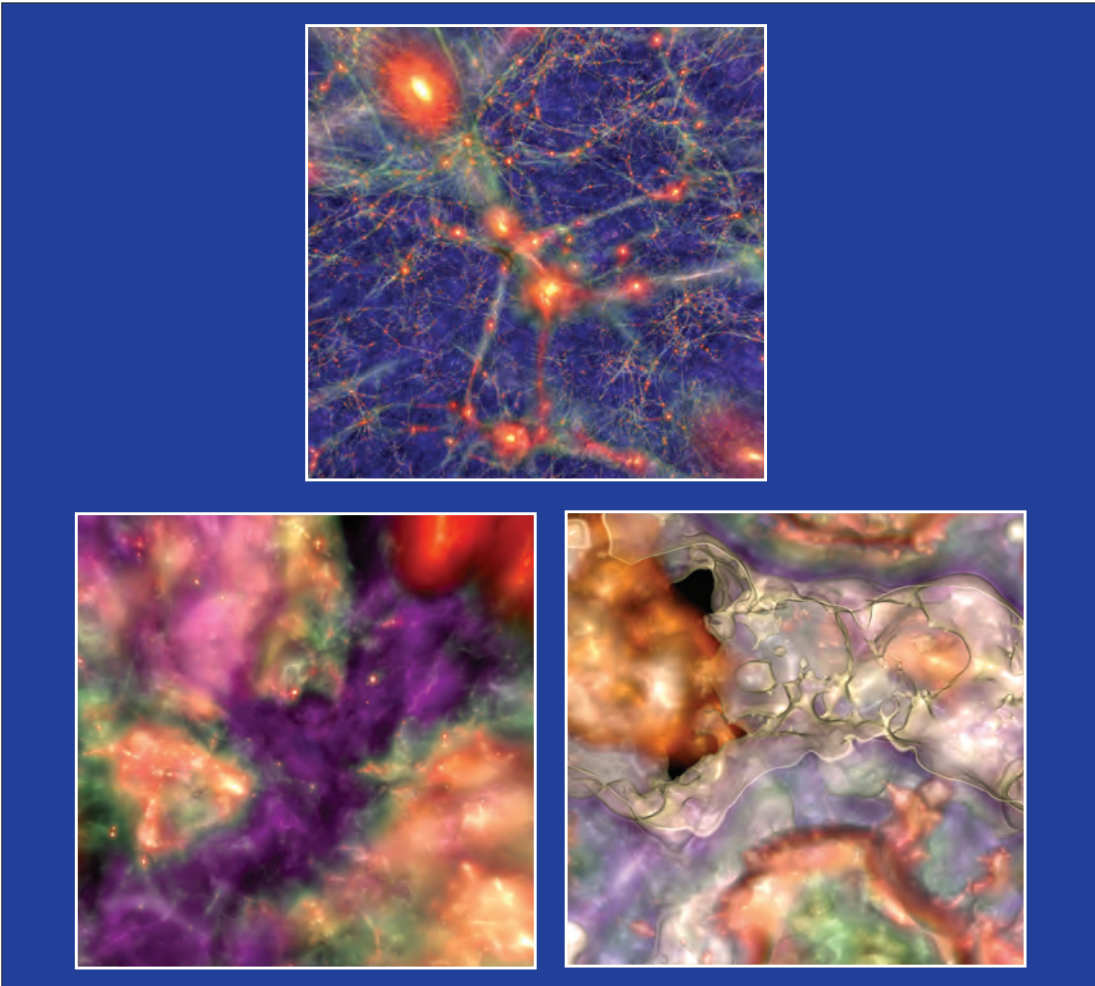
Figure 3. *Visualization of density data (top) and gravitational potential data (bottom) output from a cosmology simulation performed by Dr. Brian O'Shea. The Ultravis Institute has created a new rendering technology for visualizing data on 3D adaptive grids at previously unavailable quality and interactivity.*

because researchers want to see how each field behaves near the isosurface. Given this knowledge, we encoded selected scalar fields based on each grid point's distance from the mixture fraction isosurface. For example, we can more aggressively compress the regions farther away from the features of interest. Figure 2 shows visualizations made using two different distance values for the HO_2 field. Scientists can choose to visualize HO_2 distribution around the isosurface within a selected distance from the surface. The region farther away from the surface, which would otherwise obscure the more interesting features near the mixture fraction isosurface, is not displayed or de-enhanced. Such a fine level of control on revealing spatial features locally was not previously available to scientists. This new capability will suggest new and novel ways to visualize and explore data, allowing scientist to study their data to a much greater extent. Finally, the encoding resulted in a 20-fold saving in storage space, which was particularly critical when graphics processing unit (GPU)-accelerated rendering

was used. In this way, we can better cope with the limited video memory space of commodity graphics cards to achieve interactive visualization. The images generated from data with and without the distance-based compression are almost visually indistinguishable.

Adaptive-Grid Data Visualization

Unstructured grids and adaptive mesh refinement (AMR) methods are increasingly used in large-scale scientific computations for the modeling of problems involving complex geometries and phenomena at varying scales. By applying finer meshes only to regions requiring high accuracy, both computing time and storage space can be reduced. On the other hand, the use of adaptive, unstructured discretizations complicates the visualization task, because the resulting datasets are irregular both geometrically and topologically. The need to store and access additional information about the grid's structure can lead to visualization algorithms that incur considerable memory and computational
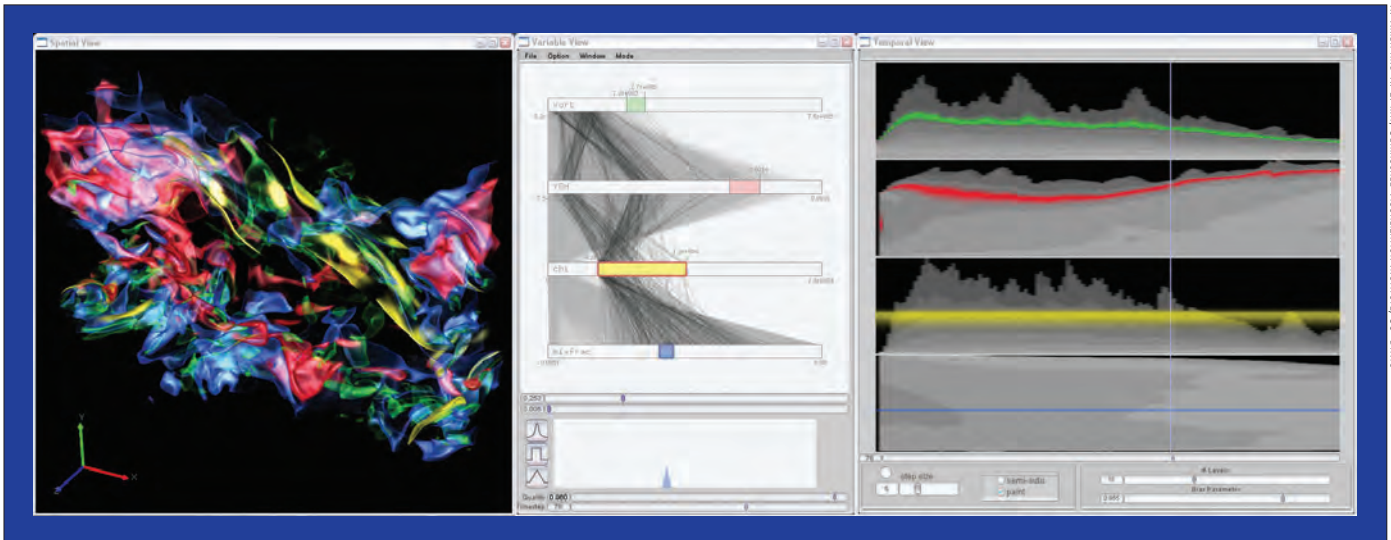
Figure 4. *Interface for tri-space visual exploration of time-varying, multivariate volume data. From left to right, the spatial view, variable view, and temporal view of the data are given.*

overheads. Hardware-accelerated volume rendering can offer real-time rates but is limited by the size of the video memory and complexity of the rendering algorithm. While algorithms and tools are introduced for visualizing simple unstructured AMR grids, several areas of study are beginning to use more sophisticated mesh structures, consisting of nested, overlapped, or higher-order elements. We are developing new approaches to render these data. Produced from a cosmology simulation, figure 3 (p15) shows visualization results using our new algorithm to render AMR data with nested elements without re-meshing. Because we do not resample the mesh into a uniform one, nor do we break the nested elements, our renderer can handle data orders of magnitude larger than conventional algorithms can. Both quality and rendering performance are superior compared to existing systems.

Interfaces for Time-Varying, Multivariate Data Visualization

A visualization tool's usability is largely determined by its user interface. Past visualization research largely focused on designing new visual representations and improving performance visualization calculations. The design and deployment of appropriate user interfaces for advanced visualization techniques began to receive more attention only a few years ago. Interface design has therefore played a major role in several Institute projects, in particular enhancing scientists' ability to visualize and analyze time-varying, multivariate data.

One visualization interface designed for exploring time-varying, multivariate volume data consists of three components that abstract the complex exploration of different spaces of data and visualization parameters. An important concept realized here is, the interface is also the visualization itself. In figure 4, the rightmost panel displays the time histograms of the data. A time histogram shows how the distribution of data values changes over the whole time sequence and can thus help the user identify time steps of interest and specify time-varying features. The middle panel attempts to display the correlation between each pair of variables in parallel coordinates for the selected time step. By examining different pairs of variables the user can often identify features of interest based on the correlations observed. The leftmost panel displays a hardware-accelerated volume rendering enhanced with the capability to render multiple variables into a single visualization in a user-controllable fashion. Such simultaneous visualizations of multiple scalar quantities allow the user to more closely explore and validate simulations from the temporal data space, parallel-coordinate space, and the 3D physical space. These three components are tightly cross-linked to facilitate what we call tri-space data exploration, offering scientists new power to study their time-varying, multivariate volume data.

A similar interface design effectively facilitates visualization of multidimensional particle data output from a gyrokinetic simulation performed by scientists at the Princeton Plasma Physics Laboratory. Depicting the complex phenomena associated with particle data presents a challenge because of the large quantity of particles, variables, and time steps. By using two modes of interaction—physical space and variable space—our system allows scientists to explore collections of densely packed particles and discover interesting features within the data. Although single variables can be easily explored through a one-dimensional transfer

By using two modes of interaction—physical space and variable space—our system allows scientists to explore collections of densely packed particles and discover interesting features within the data.
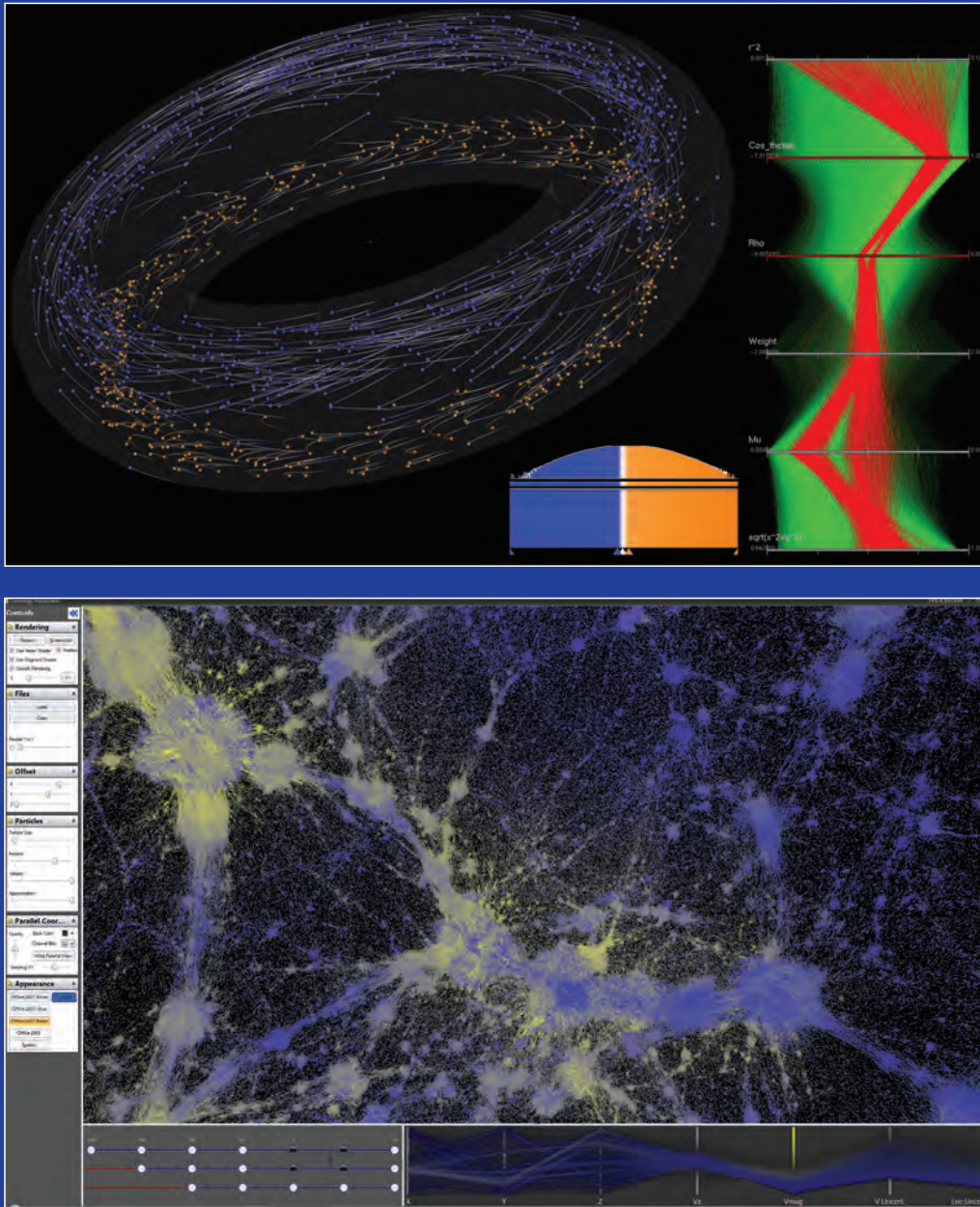
Figure 5. *Top, multidimensional visualization of a gyrokinetic particle simulation with a parallel coordinates interface using six axes, from top to bottom: toroidal coordinate, trap particle condition, parallel velocity, statistical weight, perpendicular velocity, and distance from the center. The visualization highlights trapped particles, those changing direction frequently. Bottom, a similar user interface for comparative analysis of different cosmological simulation codes and approximations.*

function, we again turn to a parallel coordinates interface for interactively selecting particles in multivariate space. In this manner, particles with deeper connections can be separated and then rendered using sphere glyphs and pathlines. With this interface, scientists are able to highlight the location and motion of particles that become trapped in turbulent plasma flow (figure 5, top). The bottom panel in figure 5 shows a similar design for comparative visualization of cosmological simulations. Using

this interface, scientists at the Los Alamos National Laboratory are able to more intuitively and conveniently compare different simulation codes or different approximations using the same code. Our approach to multivariate data visualization has also been adopted by general-purpose visualization tools such as VisIt.

Orthogonal to the above designs, we have also developed an interface allowing scientists to visualize specific multivariate temporal features using
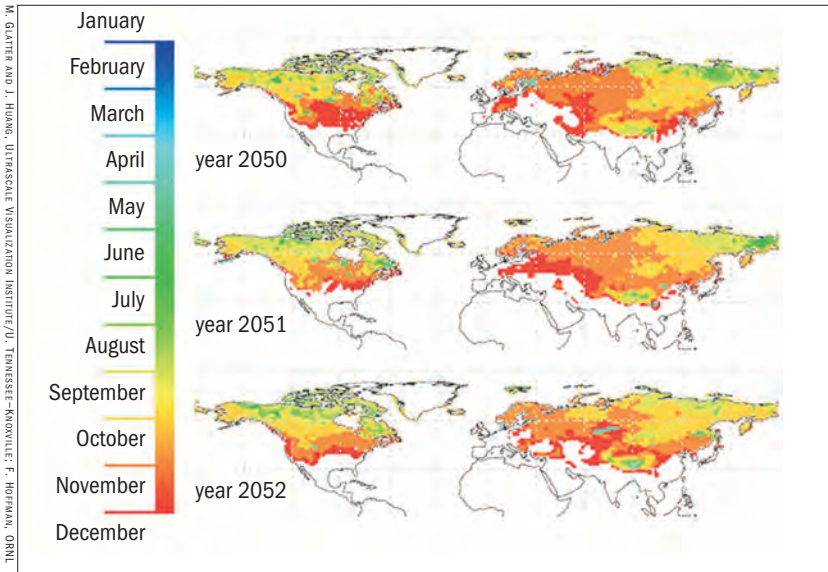
Figure 6. *An uncertainty-tolerant visualization created from using seeReg, a visualization language interface, shows variation in time of the first "major" snowfall in the northern hemisphere in a C-LAMP simulation.*

powerful constructs resembling regular expressions for textual search. Through a succinct visualization language interface, which we call seeReg, a user can specify partial hypotheses for visualization. The power of seeReg stems from its capability to automatically expand on partially specified features and reveal all possible matches in the dataset in a succinct view. Figure 6 shows a visualization of DOE C-LAMP (Carbon-Land Model intercomparison Project) climate modeling data using seeReg to reveal variation in time of the first "major" snowfall in the northern hemisphere. The years shown from top to bottom are 2050, 2051, and 2052. "First Snow" is detected by months of event with an expression specifying that the user is not interested in any variation of ground cover of snow below 0.7 units. As soon as the snow ground cover increases by more than 0.7 units, our domain expert regards the event as the first major snowfall. The value of 0.7 is determined empirically by the user in a highly interactive visualization exploration. This visualization is novel in that it allows fuzzy user knowledge to be directly used in creating visualization. This perspective of uncertainty visualization has not been presented before in the field. In particular, it nicely complements traditional multivariate visualizations created with a parallel coordinate type of user interface.

Parallel Visualization

Scientists generally do not look at all of the terabytes of data generated by their simulations, because they do not have a tool capable of displaying the full extent of the data at the original precision. The usual practice is to down-sample the data in either the temporal or the spatial domain so it is possible to look at the reduced data on a desktop computer, which defeats the purpose of running the high-resolution simulations. A feasible solution to this problem is to use parallel visualization, which capitalizes the power of the same supercomputers used to run the simulations. While many parallel visualization algorithms have been developed over the years, very few are usable due to their limited scalability and flexibility. It is the primary goal of the Ultravis Institute to make parallel visualization technology a commodity for SciDAC scientists and the broader community. We have been experimenting with the whole process of parallel visualization, including input/output (I/O) and coordination with the simulations rather than rendering algorithms in isolation, on the new generation of supercomputers such as the IBM Blue Gene/P and the Cray XT4 for assessing the feasibility of post-processing visualization as well as *in situ* visualization.

A distinct effort of ours is the development of a scalable parallel visualization method for understanding time-varying 3D vector-field data. Vector-field visualization is more challenging than scalar-field visualization, because it generally requires more computing for conveying the directional information and more space for storing the vector field. We have introduced the first scalable parallel pathline visualization algorithm. Particle tracing is fundamental to portraying the structure and direction of a vector flow field, because we can construct paths and surfaces from the traced particles to effectively characterize the flow field. However, visualizing a large time-varying vector field on a parallel computer using particle tracing presents some unique challenges. Even though the tracing of each particle is independent of others, a particle may drift to anywhere in the spatial domain over time, demanding interprocessor communication. Furthermore, as particles move around, the number of particles each processor must handle varies, leading to uneven workloads. Our new algorithm supports levels of detail, has very low communication requirements, and effectively balances load. This new capability enables scientists to see their vector-field data in unprecedented detail, at varying abstraction levels, with higher interactivity. An example is the upper image in figure 1 (p13), showing pathline visualization superimposed with volume rendering of the angular momentum field.

For massively parallel rendering, parallel efficiency is often determined by the image compositing step, which combines the partial images generated by individual processors into a complete image. Because this step requires interprocessor communication, it can easily become the primary bottleneck for scalable rendering if not carefully done. We have experimentally studied existing

Through a succinct visualization language interface, which we call seeReg, a user can specify partial hypotheses for visualization.

parallel image compositing algorithms and found binary swap the most scalable one. The binary swap algorithm, nevertheless, has an inherent limitation in processor utilization. Consistently, high efficiency is obtained only when the number of processors is a power of two. We have developed a variation of the binary swap algorithm, which we call a 2–3 swap. Our experimental study on both the IBM Blue Gene/P and Cray XT4 shows this new algorithm is scalable to any number of processors. We are making our binary swap compositing implementation available to others who develop parallel renderers.

In Situ Visualization

Due to the size of data typically produced by a large-scale simulation, visualization is almost exclusively a post-processing step. Even though it is desirable to monitor and validate some simulation stages, the cost of moving the simulation output to a visualization machine could be too high to make interactive visualization feasible. A better approach is either not to move the data or to keep to a minimum the portions of the data that must be moved. This approach can be achieved if both simulation and visualization calculations are run on the same parallel supercomputer, so the data can be shared, as shown in figure 7. Such *in situ* processing can render images directly or extract features, which are much smaller than the full raw data, to store for on-the-fly or later examination. As a result, reducing both the data transfer and storage costs early in the data analysis pipeline can optimize the overall scientific discovery process.

In practice, however, this approach has seldom been adopted. First, most scientists were reluctant to use their supercomputer time for visualization calculations. Second, it could take a significant effort to couple a legacy parallel simulation code with an *in situ* visualization code. In particular, the domain decomposition optimized for the simulation is often unsuitable for parallel visualization, resulting in the need to replicate data for accelerating visualization calculations. Hence, scientists commonly store only a small fraction of the data, or they study the stored data at a coarser resolution, which defeats the original purpose of performing the high-resolution simulations. To enable scientists to study the full extent of the data generated by their simulations and for us to possibly realize the concept of steering simulations at extreme scale, we ought to begin investigating the option of *in situ* processing and visualization. Many scientists are becoming convinced that simulation-time feature extraction, in particular, is a feasible solution to their large data problem. During the simulation time, all relevant data about the simulated field are readily available for the extraction calculations.
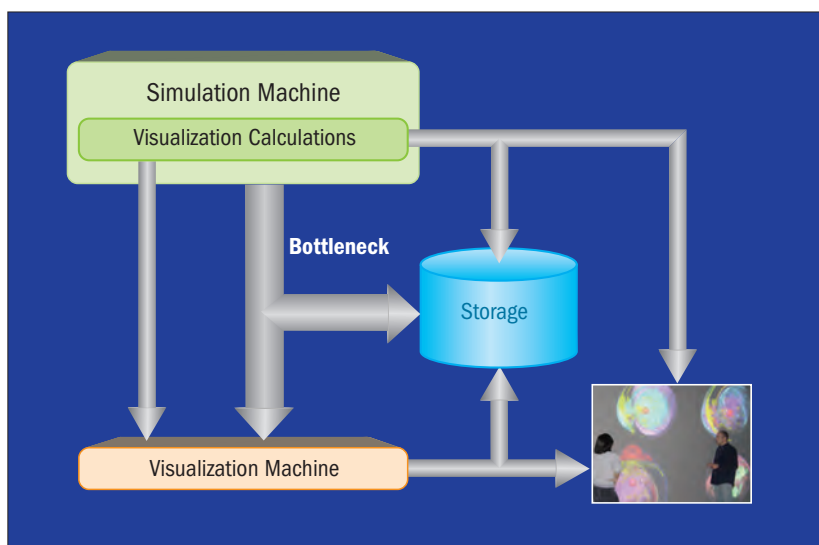
Figure 7. *Transferring the whole data generated by a petascale simulation to a storage device or a visualization machine could become a serious bottleneck, because I/O would take most of the supercomputing time. A more feasible approach is to reduce and prepare the data in situ for subsequent visualization and data analysis tasks. The reduced data, which typically are several orders of magnitude smaller than the raw data, can greatly lower the subsequent data transferring and storage costs.*

It may also be desirable and feasible to render the data *in situ* to monitor and steer a simulation. Even in the case that runtime monitoring is not practical due to the length of the simulation run or the nature of the calculations, it may still be desirable to generate an animation characterizing selected parts of the simulation. This *in situ* visualization capability is especially helpful when a significant amount of the data is to be discarded. Along with restart files, the animations could capture the integrity of the simulation with respect to a particularly important aspect of the modeled phenomenon.

We have been studying *in situ* processing and visualization for selected applications to understand the impact of this new approach on ultrascale simulations, subsequent visualization tasks, and how scientists do their work. Compared with a traditional visualization task that is performed in a post-processing fashion, *in situ* visualization brings some unique challenges. First of all, the visualization code must interact directly with the simulation code, which requires both the scientist and the visualization specialist to commit to this integration effort. To optimize memory usage, we must share the same data structures with the simulation and visualization codes to avoid replicating data. Second, visualization workload balancing is more difficult to achieve because the visualization must comply with the simulation architecture and be tightly coupled with it. Unlike parallelizing visualization algorithms for standalone processing where we can partition and distribute data best suited for the visualization calculations, for *in situ* visualization,

It is the primary goal of the Ultravis Institute to make parallel visualization technology a commodity for SciDAC scientists and the broader community.
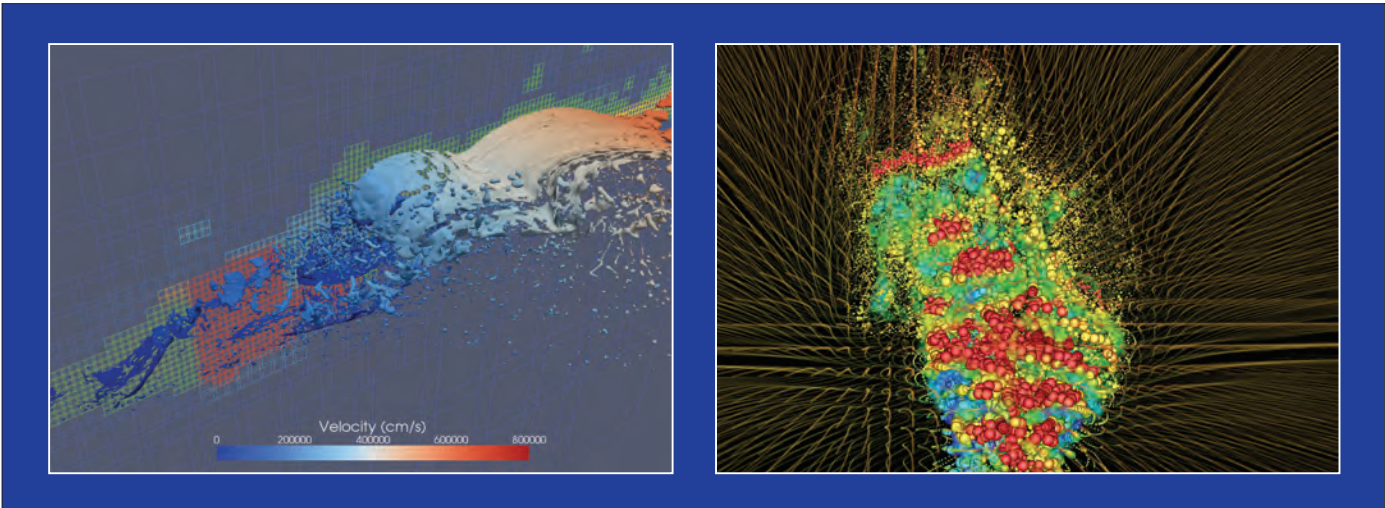
Figure 8. *Some core functionalities for next-generation visualization technologies implemented in the ParaView framework. Left, fragmentation analysis using large-scale AMR grids. Right, optimization of hydroelectric turbines.*

the simulation code dictates data partitioning and distribution. Moving data frequently among processors is not an option for visualization processing. We need to rethink to possibly balance the visualization workload so the visualization is at least as scalable as the simulation. Finally, visualization calculations must be low cost with decoupled I/O for delivering the rendering results while the simulation is running. Because visualization calculations on a supercomputer cannot be hardware accelerated, we must find other ways to simplify the calculations such that adding visualization would only detract a very small fraction of the supercomputer time allocated to the scientist.

We have realized *in situ* visualization for a terascale earthquake simulation. This work also won the HPC Analytics Challenges competition of the Supercomputing Conference because of the scalability and interactive volume visualization we were able to demonstrate over a wide-area network using 2,048 processors of a supercomputer at the Pittsburgh Supercomputing Center. We were able to achieve high parallel efficiency exactly because we made the visualization calculations, that is, direct volume rendering, to use the data structures used by simulation code, which removes the need to reorganize the simulation output and replicate data. Rendering is done *in situ* using the same data partitioning made by the simulation, so no data movement is needed among processors. We are beginning work on a new earthquake simulation that will be petascale and include the structures of buildings above ground, which also demands more work for the visualization part.

To realize *in situ* data reduction and visualization, we need to essentially create a new visualization and data-understanding infrastructure. We are also developing *in situ* data encoding algorithms, index-

ing methods, incremental 4D feature extraction and tracking algorithms, and data-quality measuring methods for better supporting post-processing visualization. We must emphasize that visualization researchers in isolation cannot achieve this important goal. Through SciDAC, we have obtained commitments from science collaborators to pursue this research together.

Visualization Tools

New visualization technologies cannot reach a wide audience without being incorporated into visualization tools that can be leveraged by the scientist who needs to understand the data. Building visualization tools, in particular general-purpose tools that can be used in a wide variety of domains, includes the daunting task of combining advanced visualization techniques, which often have conflicting requirements on the application framework, data types, and interaction patterns. For maximum applicability, such tools may also need to perform well on a wide variety of architectures: from single-user laptops, to remote interactive access of specialized parallel visualization hardware, to massively parallel supercomputers.

To deploy next-generation visualization technologies, we primarily use the ParaView application and framework. In order to use ParaView, or any other visualization framework, to deploy the latest visualization research, the basic underlying framework must have fundamental support. For ParaView, this includes the necessary adaptive data structures, temporal extent controls, and parallel scalability. Figure 8 shows two demanding applications using ParaView. ParaView's modular architecture and multiple customization paths, including plug-ins and scripting, allow for quick integration of multiple visualization technologies. Figure 9

We must emphasize that visualization researchers in isolation cannot achieve this important goal. Through SciDAC, we have obtained commitments from science collaborators to pursue this research together.
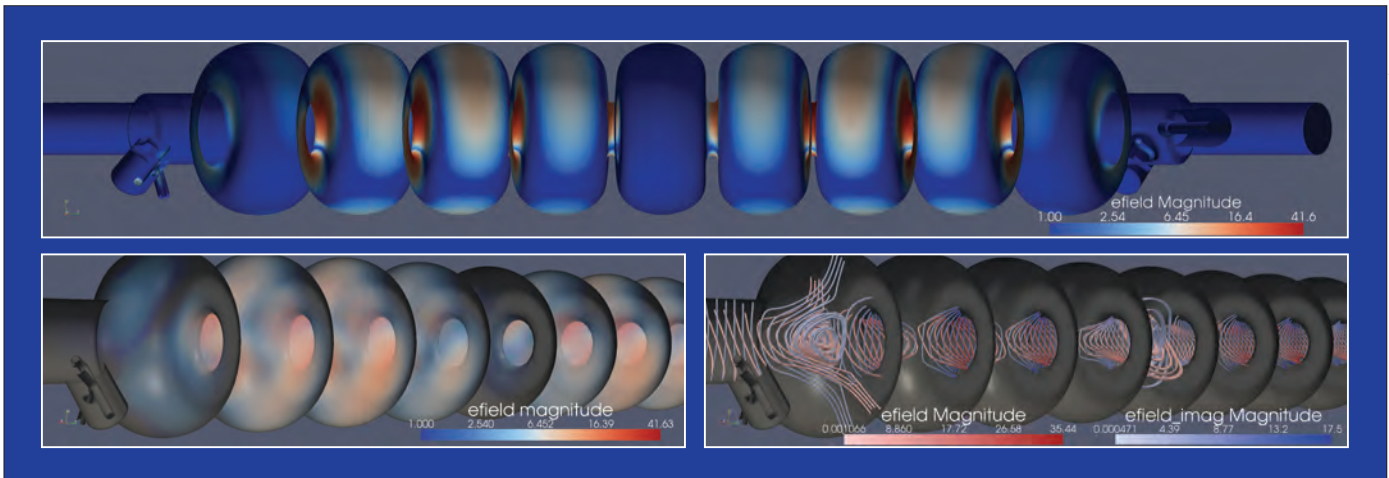
Figure 9. *Visualization of a simulation of a low-loss accelerating cavity using ParaView by researchers at SLAC. Top, electric field magnitude on the cavity surface. Bottom left, electric field magnitude within the cavity's volume. Bottom right, direction of real and imaginary electric fields.*

demonstrates how a customized reader and rendering module used with ParaView allows a variety of visualization techniques to be applied to the data that scientists at the Stanford Linear Accelerator Center need to study.

In addition to adopting the latest visualization technologies as they are developed, visualization tools must also adapt to the changing landscape of requirements and supporting hardware. For example, petascale visualization tools may soon need to exploit new parallel paradigms in hardware such as multiple cores, multiple GPUs, and cell processors. Another critical consequence of petascale computing is larger datasets generated by simulations; I/O, particularly file I/O, is a more critical concern than ever before. To maintain scalability in file reads and writes, parallel coordinated disk access must be an integrated part of our visualization tools.

Petascale computing also provides the facility for several new modes of analysis. With higher fidelity simulations comes the ability to quantify uncertainty. Comparative analysis is also important for both ensuring the validity of a simulation as well as verifying the results correspond to the physical phenomenon. Ensembles of runs can have higher fidelity or larger dimensions than ever before on petascale supercomputers. All three of these analyses are underrepresented in our current visualization tools, and this is a problem we must correct for effective analysis to continue.

Conclusion

Petascale and exascale computing are right around the corner. Will we have adequate tools for extracting meaning from the data generated by extreme-scale simulations? The timely investment made by the DOE SciDAC program in ultra-scale visualization ensures the challenges will be addressed. In this article, we point out some of the grand challenges facing extreme-scale data analysis and visualization and present several key visualization technologies developed by the Ultravis Institute to enable knowledge discovery in the petascale and exascale regimes. The Institute combines the pursuit of basic research in visualization algorithms, interaction techniques, and interface designs with a unique approach of working one-on-one with SciDAC application scientists to accelerate the adoption of new technologies by their science teams. This allows us to demonstrate the success of newly developed technologies in a particular area of study, providing other science teams with an opportunity to see how each technology is applied. Matured technologies are generalized and deployed using ParaView to benefit the larger scientific community. The Institute has also launched a series of outreach and education activities that foster interaction both within the community and between visualization experts and application scientists. The impact of these basic research efforts together with the outreach and education activities not only accelerates the development of ultra-scale visualization tools but also helps strengthen the presence of SciDAC in the academic community. ●

Further Reading

http://ultravis.org

The impact of these basic research efforts together with the outreach and education activities not only accelerates the development of ultra-scale visualization tools but also helps strengthen the presence of SciDAC in the academic community.