Remote and Distributed High Performance Scientific Visualization – Terabytes to Insights

E. Wes Bethel, John Shalf, Cristina Siegerist, LBNL
Ken Joy, Kwan-Liu Ma, Bernd Hamann, UCD

.

The ever growing size and complexity of data from scientific simulations is creating a bottleneck in analyzing and understanding simulation results. In addition, while the computations tend to be performed at centralized facilities, analysis is performed by distributed teams of scientists at remote locations. Our work focuses on combining fundamental science-driven visualization research to accelerate data understanding in both a remote and distributed context.

Vision. Application scientists face a challenge – the need to analyze and understand ever larger and complex datasets grows more acute on a daily basis, yet visualization and analysis technology and interactive resources have not evolved at a pace similar to scientific simulations and centralized computing facilities. Our vision is to respond to the needs of the science applications community by providing, through close collaborative research, domain-specific tools that answer the challenges posed by large and complex data. Our response consists of two broad thrust areas. First, fundamental research in visualization algorithms provides solutions tailored to specific domain challenges. Second, research in visualization architectures and deployment mechanisms helps to make advanced visualization technology accessible to domain scientists.

Exploring Large Accelerator Modeling Data Sets. Working closely with the Advanced Computing for 21st Century Accelerator Science and Technology SciDAC ISIC, our group has created domain-specific visual analysis tools suitable for use in exploring large and complex time varying, six-dimensional datasets produced by accelerator simulations. The PPaint application, shown in Figure 1, combines two complementary techniques to accelerate data understanding. Regions of interest – in this case a beamline halo – are interactively selected using a painting metaphor in a 2D projection. The selected regions are clearly visible in multiple time-varying, 3D views, helping to understand phenomena in complex, time-varying data. We developed a novel visualization technique well suited for use on 6D data (position, phase) that helps to visually distinguish between particles moving towards or away from the viewer.

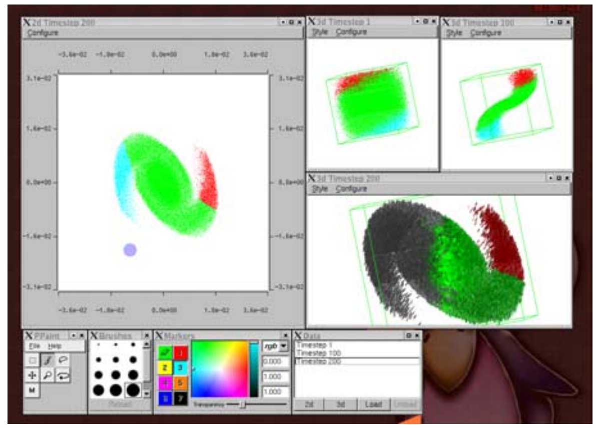


Figure 1. The PPaint Application combines interactive selection with a novel visualization technique to facilitate exploration and understanding of large and complex accelerator simulation data.

Another domain-specific application, PartView, is capable of using high performance parallel I/O libraries combined with a pipelined architecture to permit interactive browsing of very large, remotely located simulation datasets.

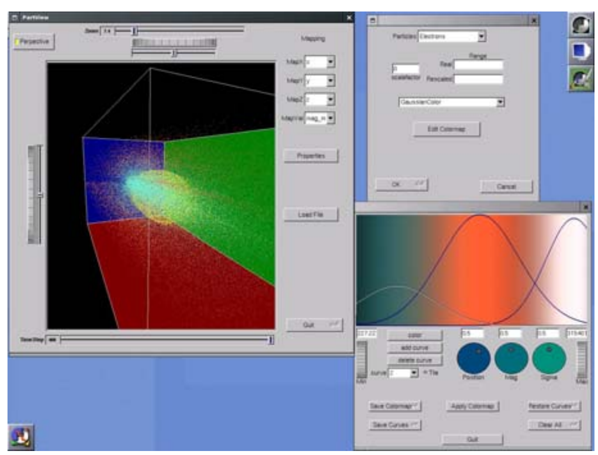
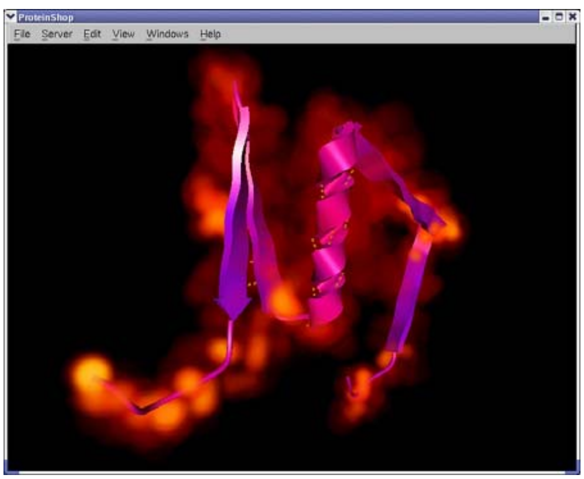


Figure 2. The PartView application supports interactive visual data analysis of very large accelerator simulation datasets using a combination of pipelined-parallel data processing combined with a domain-specific "dashboard."

Supported by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Predicting Protein Structures.
The ProteinShop application (see proteinshop.lbl.gov) uses a combination of domain-specific interaction techniques combined with parallel energy optimization codes to determine 3D protein structures given a sequence of amino acids. Working closely with Silvia Crivelli (LBNL), the LBNL Visualization Group developed a novel inverse kinematics algorithm to support physically and chemically meaningful molecular manipulation. ProteinShop can also display computed internal energy while the parallel optimization code runs on a remote supercomputer. ProteinShop is used by the LBNL team in the biannual CASP competition (predictioncenter.llnl.gov), and has submitted a proposal to NIH for long term protein structure prediction research funding using ProteinShop as a programmatic cornerstone. The LBNL Visualization group's paper entitled "Interactive Protein Manipulation" won the "Best Application Paper" award at IEEE Visualization 2003.



Performance Modeling and Pipeline Optimization. One of the greatest challenges in remote and distributed scientific visualization is a deceptively simple concept: making complex yet powerful technology accessible to non-experts. A typical deployment paradigm is to have individual work tasks – software components – placed onto distributed resources most appropriate for the task at hand. One component might reside close to the data to minimize expensive data I/O operations, while another might reside on the researcher's workstation to maximize interactive performance. In the case of visualization, the best match between component and resource for any given workflow component may change over the course of a visualization session. The change may come as a result of a specific performance target goal, or in response to a user changing an algorithmic parameter.

Starting with a basic performance model derived for visualization algorithms, our research has shown that it is possible to select an optimal placement of workflow components onto distributed resources in polynomial time. This important result indicates that dynamic pipeline reconfiguration is computationally inexpensive and may be easily applied to complex workflows consisting of dozens of distributed components. The eventual goal is to alleviate the scientist of the burden of the time consuming manual component placement onto distributed resources.

Latency Tolerant Visualization and Graphics Algorithms. The term "latency tolerant" means that a given visualization technique provides interactive performance response regardless of the capacity or latency characteristics of the underlying network transport mechanism. This past year, we have developed a suite of platform-neutral media encoders that generate QuickTime VR Object Movie format media, as well as HTML/JavaScript-based "workalikes." This form of media provides the ability to perform 3D exploration of time-varying, 3D scientific visualization results using a standard web browser, and is especially well-suited for use in offline, out-of-core, post-processing visualization settings – the environment typical of large computing centers. Future work will focus on research to support multiresolution representations of such media to provide the greater degree of visual accuracy required for scientific research.

Deployment Technologies for Remote and Distributed Visualization. Deployment is the process of transitioning research prototypes into field-usable tools. While "the Grid" offers much promise in terms of easing deployment of remote and distributed computing, much work remains to realize such a goal. To use heterogeneous computing resources, it is first necessary to stage executables for remote use – a tedious process at best. We developed a novel method of packaging multi-architecture binaries to support rapid deployment on heterogeneous, Grid-based resources.

Further information: http://vis.lbl.gov/
Contact Information:
E. Wes Bethel, Lawrence Berkeley National Laboratory, Phone: 510-486-7353, ewbethel@lbl.gov

Supported by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.