As data sets quickly approach petascale sizes and beyond, there’s a pressing need to address how to visualize them at extreme scales. To deal with the challenges, next-generation visualization software must change. Researchers must develop new visualization algorithms and systems that leverage the emerging software and hardware technology. Data visualization and analysis must also be more tightly integrated with the simulations generating the data.

Emerging architectures provide a unique opportunity to develop new visualization software. Modern supercomputers already have more than 100,000 cores, and one-million-core machines are on the horizon. The number of cores available to desktop machines has been steadily increasing. Extremely powerful graphics coprocessors and new programming models provide additional computation and graphics processing power, facilitating a tight coupling between applications and graphics.

Data storage and communication issues also profoundly affect visualization system efficacy. This is because processor performance improvements have consistently outpaced data access rates for disks. So, simulations will generate more data than we can effectively store and access with current hardware and software. Researchers must address the issues of data movement by looking into new ways to construct an end-to-end visualization pipeline.

In This Issue

Data sets’ increasing size will seriously challenge conventional visualization systems that focus on postprocessing. A major bottleneck in the traditional visualization pipeline is data movement, which includes writing simulation results to disk, reading the data back to the data analysis and visualization machines, and redistributing the data or images among the different computation nodes for load balancing and result gathering.

In “Extreme Scaling of Production Visualization Software on Diverse Architectures,” Hank Childs and his colleagues study how data-parallel visualization algorithms scale to run on today’s massive supercomputers. They report on runs on many of the world’s top supercomputers and examine performance at massive scales: using at least 16,000 processing cores analyzing trillions of cells. Important results include the success of a data-parallel visualization approach as well as the significant bottleneck that I/O becomes at this scale.

One way to achieve fast data-parallel visualization is to use ghost data: copies of adjacent data elements from neighboring processors. With these copies, you can compute appropriate boundary conditions between processors without communicating data between them. In “Parallel and Streaming Generation of Ghost Data for Structured Grids,” Martin Isenburg and his colleagues describe a memory-efficient way to generate ghost data. With their algorithm, you can create ghost data in a streaming, incremental manner, without loading the entire data set in memory.

We anticipate that future data analysis and visualization components will be much more tightly integrated with the simulation code. In situ visualization—producing visualization while the simulation runs—lets scientists analyze data at the original resolution without incurring significant I/O overhead. In “In Situ Visualization for Large-Scale Combustion Simulations,” Hongfeng Yu and his colleagues discuss in depth the challenges of in situ visualization, the integration of visualization modules into large-scale combustion simulations running on Oak Ridge National Laboratory’s Cray XT5, design decisions and optimization strategies, and performance results. They also share lessons
they learned and provide future research directions.

The explosive growth of data isn't only a phenomenon in applications involving scientific computing. It's also becoming common in biomedical imaging applications as a result of rapid advances in optical and electron microscopy. The advanced imaging techniques promise opportunities to enhance our understanding of how the human brain functions. In “Ssecret and NeuroTrace: Interactive Visualization and Analysis Tools for Large-Scale Neuroscience Data Sets,” Won-Ki Jeong and his colleagues present two tightly coupled systems that allow interactive exploration and analysis of large-scale microscope images. The research aims to reconstruct neural circuits and the mammalian nervous system. NeuroTrace focuses on interactive segmentation and 3D visualization of high-resolution electron microscope data sets by combining the 2D level set and a 3D tracking technique. Ssecret is a client-server remote-visualization system that allows real-time exploration of 2D volume slices. The article also focuses on data management across a cache hierarchy involving disk, memory, and video memory in GPUs.

Visualization essentially is a means for people to express and communicate ideas. It also plays an important role in creating an exploratory data analysis environment in which people can verify their hypotheses and seek answers to questions related to the underlying domain problems. From this viewpoint, it’s safe to assume that almost every visual-analysis task at some point will require a meeting in which people at different locations exchange ideas and discuss their findings. In “Ultrascale Collaborative Visualization Using a Display-Rich Global Cyberinfrastructure,” Byungil Jeong and his colleagues describe a unified hardware and software environment that links multiple high-resolution tiled displays to facilitate remote collaboration. To manage the parallel graphics streams between the rendering nodes and tiled-display nodes, this environment employs Scalable Adaptive Graphics Environment (SAGE) middleware. Besides discussing user interaction with SAGE and various applications of it, Jeong and his colleagues highlight its successful implementation at different institutions.

The five articles in this issue represent the state of the art in extreme-scale production visualization software, efficient domain decomposition, in situ visualization, biomedical-image segmentation and remote viewing, and display-rich collaborative cyberinfrastructure. We anticipate that interactive visualization will continue to be an integral part of the workflow to understand ultrascale scientific applications. As the number of computing cores continues to increase and the hierarchy of computation and memory becomes deeper, even more challenging research issues will arise. The work described in this issue represents only the beginning of another wave of research that will eventually lead to breakthrough ultrascale-visualization approaches.

James Ahrens is the visualization team leader in Los Alamos National Laboratory’s Computer Science for High-Performance Computing Group. His research interests include methods for visualizing extremely large scientific data sets, distance visualization, and quantitative and comparative visualization. Ahrens has a PhD in computer science from the University of Washington. Contact him at ahrens@lanl.gov.

Han-Wei Shen is an associate professor in Ohio State University’s Department of Computer Science and Engineering. His research interests are scientific visualization and computer graphics. Shen has a PhD in computer science from the University of Utah. Contact him at hwshen@cse.ohio-state.edu.