In Partnership with the Electronic Visualization Lab, the National Center for Microscopy and Imaging Research, and the Scripps Genome Center

Calit2-CEOA Exhibit Showcases e-Science Advances Based on OptIPuter Technologies

Booth 1647

This exhibit showcases applications leveraging "OptIPuter" technology, which improves our ability to collaborate. By using this technology, researchers are advancing science in the following ways:

- Creating a virtual distributed laboratory through visualcasting (#1, below)
- Studying seafloor landslides (#2)
- Creating real-time sensor networks to monitor the world's oceans (#3)
- Creating wireless networks and a data management system to deliver data in real time (#4)
- Implementing a satellite communications network to provide continuous Internet connectivity for oceanographic research ships and platforms (#5)
- Visualizing Earth science datasets in 3-D (#6)
- Mining the oceans for novel proteins and the genes that encode them (#7)
- Visualizing and interacting with protein structures (#8)
- Developing integrated collaborative cyberinfrastructure to support multi-scale, multi-modal, multi-site (biomedical) science (#9)

OptIPuter is one of the largest Information Technology Research grants, which was awarded by the National Science Foundation in 2002. Its mission is to enable collaborating scientists to interactively explore massive amounts of previously uncorrelated data by developing a new network-centric cyberinfrastructure for a number of this decade's e-science shared information technology facilities.

The OptIPuter is a tightly integrated cluster of computational, storage, and visualization resources, linked over parallel, dedicated optical networks across campus, metro, national, and international scales. The OptIPuter runs over multi-1 to 10-Gbps lambdas, with advanced middleware and network management tools and techniques to optimize transmissions so distance-dependent delays are the only major variable.

OptIPuter partners are designing and developing unique software services to support data-intensive scientific applications. These services use highly distributed resources, such as compute clusters, mass storage systems, visualization systems, and parallel optical communications channels. These resources can be discovered, assembled, utilized, and reconfigured dynamically. One component of the software system is the Distributed Virtual Computer (DVC) middleware and its dependencies. LambdaRAM is the distributed shared-memory middleware that allows multiple OptIPuter clusters at remote sites to participate in a single memory mapping. This empowers applications to access multidimensional terabyte datasets over optical networks. The Scalable Adaptive Graphics Environment (SAGE), developed by UIC's EVL, is specialized middleware that enables real-time streaming of extremely high-resolution graphics and high-definition video from remotely distributed rendering and storage clusters to scalable display walls over high-speed networks. Another enables data resource discovery and use, and another contains protocols for data transport and light-path provisioning. The distribution mechanism for this software is the SDSC Rocks automated configuration technology.

Image: A hawk is captured soaring above SDSU's Santa Margarita Ecological Reserve by a real-time ROADNet camera.

OptIPortals—SDSC Rocks/SAGE implementations running on clusters driving tiled displays that are connected to other sites over multi-gigabit links—are being deployed at partner sites. The OptIPortal is a 21st century PC, and SAGE serves as a window manager on tiled displays, enabling users to bring up multiple streamed applications and view them simultaneously.

University of California, San Diego (UCSD), and University of Illinois at Chicago (UIC) lead the research team. Academic partners include Northwestern University, San Diego State University, University of Southern California/Information Sciences Institute, University of California, Irvine, Texas A&M University, and University of Illinois at Urbana-Champaign/ National Center for Supercomputing Applications. Affiliate partners include the U.S. Geological Survey/National Center for Earth Resources Observation and Science, NASA, University of Michigan, Purdue University, University of Amsterdam and SARA Computing and Network Services in The Netherlands, CANARIE and the Communications Research Centre in Canada, the Korea Institute of Science and Technology Information in Korea, and the National Institute of Advanced Industrial Science and Technology in Japan. Industrial partners include Big Bangwidth, Calient Networks, Chiaro Networks, Glimmerglass Networks, IBM, Lucent Technologies, Rincon Research Corporation, Sun Microsystems, and Telcordia Technologies.

The OptIPuter receives major funding from NSF via cooperative agreement OCI-0225642 to UCSD. For more information, see www.optiputer.net or contact Maxine Brown, maxine@evl.uic.edu.

1. SAGE and Visualcasting Create a Virtual Distributed Laboratory

Electronic Visualization Laboratory (EVL) and the University of Illinois at Chicago (UIC); Contact: Maxine Brown, maxine@evl.uic.edu

EVL/UIC demonstrates its Scalable Adaptive Graphics Environment (SAGE) middleware and visualcasting service to create a virtual, distributed laboratory on the SC06 show floor. Multiple OptIPuter partners, using differently sized tiled displays in multiple research booths, share and discuss high-resolution content supported by multiple high-definition video streams. Visualcasting, a high-performance SAGE network service, extends SAGE to support distributed collaboration by letting users share application content. SAGE visualcasting supports the distribution of high-definition video for teleconferencing among the sites, as well as streams of scientific data (including NASA animations, Louisiana State University and Scripps Institution of Oceanography simulations, and National Center for Supercomputing Applications visualizations) over 10-Gbps networks locally via SCInet and nationally via National LambdaRail.

2. Studies of the Gaviota Slide Offshore Southern California and EM Modeling

Scripps Institution of Oceanography and British Petroleum America; Contact: Atul Nayak, anayak@ucsd.edu

As marine geoscientists image seafloor morphology with higher and higher resolution, evidence of the widespread occurrence of seafloor landslides grows. Because mass-wasting events pose a serious threat to subsea production installations, pipelines, and coastal communities, it is important to understand the mechanics of these geohazards. We are studying a seafloor landslide off the coast of Santa Barbara, California (Goleta and Gaviota slides).

At SC06, we present a 3-D flythrough of the Santa Barbara region where Scripps and BP have set up an instrumentation facility. The flythrough highlights underwater bathymetry at the landslide study regions, data collected from a CHIRP sonar survey in 2004, and visualization of the new seafloor geodetic techniques using optical fibers, an array of precise acoustic transponders, and an AUV-borne precision mapping system. The visualizations are presented using the SAGE framework and Magic Carpet software.

3. Cyberinfrastructure for Ocean Observing Systems

Laboratory for the Ocean Observatory Knowledge Integration Grid (LOOKING) and the Southern California Coastal Ocean Observing System (SCCOOS); Contacts: Matthew Arrott, marrott@ucsd.edu (LOOKING); Atul Nayak, anayak@ucsd.edu (SCCOOS)

The development of instrument arrays and grids (power/compute/storage/visualization) that can provide coherent cyberinfrastructure to support real-time sensing networks of our oceans is an important focus area for CEOA and Calit2. Two projects led by UCSD are noteworthy. LOOKING (lookingtosea.ucsd.edu) seeks to create an autonomous sensor network that responds to changes in user requirements, technologies (software or middleware), and ocean observatory life cycle needs. One of the most significant deliverables of LOOKING is the design of a proposed architecture for the NSF program ORION.

Working interactively with local, state, and federal agencies, resource managers, policy makers, educators, scientists, and the public, SCCOOS is improving the delivery of coastal observations to enable better management of the coastal ocean environment. SCCOOS brings together coastal observations in the

Southern California Bight to provide information to address issues in climate change, ecosystem preservation and management, coastal water quality, maritime operations, coastal hazards, and national security. The website (www.sccoos.ucsd.edu) serves automated and manual shore station data; near real-time surface current, wave height, and surface winds data; and bathymetric data for southern California.

At SC06, we present technologies that impact the development of scalable, high-resolution, and continuous ocean-monitoring systems. For example, visualizations of the Huntington Beach experiment (a SCCOOS-administered project) show deployment of gliders, buoys, remotely operated vehicles, and moorings, and the functions of the individual instruments. We also display visualizations of surface wave currents and wave heights around southern California and real-time data products delivered to the SCCOOS website. We also play high-definition video from deep-sea, high-temperature venting systems off the Washington-British Columbia coastline collected by the University of Washington and the Research Channel.

4. ROADNet

Scripps Institution of Oceanography and the San Diego Supercomputer Center; Contact: Carolyn Keen, cskeen@ucsd.edu

ROADNet is enhancing our capacity to monitor and respond to changes in our environment by developing wireless networks and an integrated, seamless, and transparent information management system to deliver seismic, oceanographic, hydrological, ecological, and physical data to a variety of end users in real time. (See accompanying brochure.)

5. HiSeasNet

Scripps Institution of Oceanography; Contact: Steve Foley, sfoley@ucsd.edu

Access to the Internet is an integral part of nearly every research lab and office on land; extending this access to oceanographic ships—our seagoing laboratories—broadly impacts seagoing research activities. HiSeasNet is a satellite communications network designed to provide continuous Internet connectivity for oceanographic research ships and platforms. For the ships, HiSeasNet provides:

- Transmission of hot data in real time to shore-side collaborators.
- Basic communications—e-mail, voice, and videoteleconferencing for scientists, engineers, and crew at sea.
- Tools for real-time educational interactions between shipboard scientists, teachers, and the classroom, as well as informal science and other education and outreach activities.

6. 3-D Interactive Visualization of Heterogeneous Earth Science Datasets

Visualization Center at Scripps; Contact: Debi Kilb, dkilb@ucsd.edu

At SC06, we display these datasets collected by geologists, seismologists, and oceanographers at EarthScope sites, the RIDGE 2000 study sites, and sites in southern California. (See also the brochure.)

7. Mining the Ocean for Novel Proteins and the Genes that Encode Them

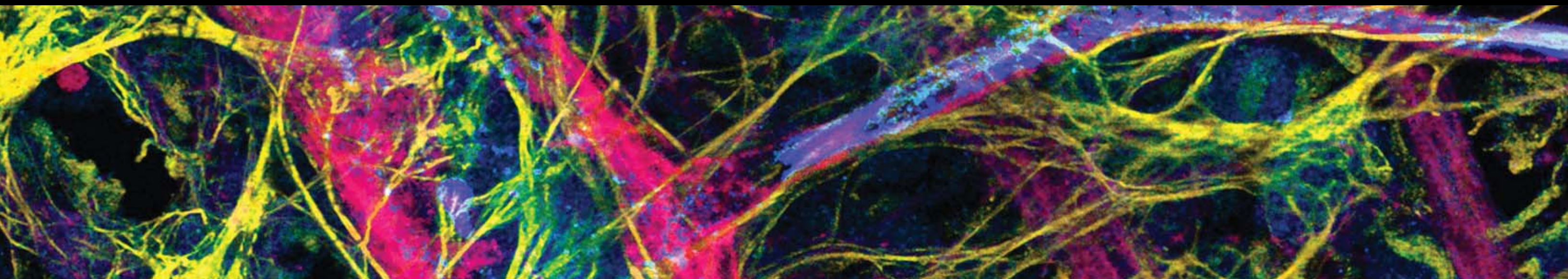
Calit2 and the Scripps Genome Center; Contacts: Lee Edsall, ledsall@genomes.ucsd.edu (ecology/metagenomics); Jurgen Schulze, jschulze@ucsd.edu (computer science)

Marine environmental (metagenomic) datasets present an opportunity for bioinformaticists to mine the ocean for novel proteins and the genes that encode for them. Metagenomic sampling, which collects the DNA sequences present in an environmental sample, allows us to ask broad questions such as

- What biological activity is taking place in this environment?
- Are there natural products here that can be useful to humans?

The Scripps Genome Center (genomes.ucsd.edu) is investigating the genome of a coral pathogen, Aurantimonas SI85-9A, in relation to 11 published marine metagenomic samples for the presence of genes encoding novel proteins of potential biomedical interest. The samples were collected at depths of up to 4,000 meters and encompass roughly 10 million potential protein-coding regions. For the purposes of annotating these large datasets, we are using the MAGPIE (Multipurpose Automated

Image: Confocal microscopy and a suite of immunohistochemical stains reveal spatial relationship of astrocytes in the rat retina. Image courtesy of Tom Deerinck, Frank Hughes, and Mark Ellisman, National Center for Microscopy and Imaging Research (NCMIR).



Genome Project Investigation Environment) software on top of a clustered visualization technology (tiled display wall). The tiled wall allows us to display a larger range of genomic information and more of the evidence data generated by an automated annotation. Our visualization software connects to the MAGPIE data bank at the Scripps Institution of Oceanography via a network connection and downloads high-resolution images with the DNA sequences. The images and evidence tables have been generated using 15G of genomic data, which was computed on a 44-node Sun Microsystems Solaris cluster.

8. *Visualizing and Interacting with Protein Data*

Calit2, Burnham Institute for Medical Research, and UCSD; Contact: Jurgen Schulze, jschulze@ucsd.edu

This demo features the PDB Browser software, developed at Calit2. It enables 3-D visualization and interaction with data downloaded in real time from the Structural Classification of Proteins (SCOP) and the Protein Data Bank (PDB), which contains more than 36,000 protein structures. Such interaction makes it possible to study the structures of biological macromolecules and their relationships to sequence, function, and disease. The PDB Browser has been implemented as a plug-in for the COVISE navigation software, developed at the High Performance Computing Center Stuttgart (Germany). The software enables the user to "fly" around and through molecules, change their sizes, and investigate potential binding sites and how the shape of a molecule changes when it changes state. It also enables collaborating researchers in different geographical locations to share and interact with each other's imagery.

Proteins are modular in architecture and are comprised of domains. The enormous diversity seen among proteins has evolved from a small number of ancestral domains, possibly in the range of a few thousand domains, and all living organisms sequenced to date are known to share a significant portion of their protein repertoire. Our software enables the user to interactively visualize biological macromolecules at an atomic level, allowing for their detailed comparative structural analysis, including investigating potential drug-binding sites.

9. *Beyond Networks: Integrated Collaborative Cyberinfrastructure*

National Center for Microscopy and Imaging Research; Contact: Ruth West, rwest@ncmir.ucsd.edu; Rajvikram Singh, raj@ncmir.ucsd.edu

Interrelated, grand challenges in biomedical research, information technology, and cyberinfrastructure provide a platform for the development of multi-scale, multi-modal, multi-site science. As biomedicine strives to understand the mechanisms of health and disease, it requires a variety of multi-scale, multi-modal data to fill the gaps in our understanding of biological systems.

Recent developments in high-bandwidth networks, display technology, and HPC have opened venues for researchers to analyze and visualize ultra-high-resolution data and remote resources collaboratively. Cyberinfrastructure thus has been tasked with the seamless integration of computation, data, instruments, visualization, and domain expertise. This virtualization of science teams and communities of all scopes and sizes will drive the development of both technology and the economy.

At SC06, we demonstrate the use of integrative, collaborative cyberinfrastructure to provide multi-site interaction with high-resolution neuroscience datasets and 3-D models of neuronal subcellular structures derived from light microscopy and electron tomography. The framework of our system is based on the popular Scalable Adaptive Graphics Environment (SAGE) and the Collaborative Visualization and Simulation Environment (COVISE)-driven tile displays linked over lambda-enabled, high-bandwidth, low-latency networks. We show that the ultra-high-resolution imaging data of the nervous system, when visualized, bridges anatomical scales spanning centimeters to nanometers. This data traverses whole brain slices, the cerebellum, and single cells, to parts of cells (dendritic branches and spines) and drills down to the levels within subcellular organelles—the interior of the mitochondrion. Apart from the analysis of 2-D and 3-D data, the demonstration also involves HD videoconferencing streams from remote collaborating sites and instruments. We emphasize the benefits of collaboratively visualizing and interacting with high-resolution data on tile displays and immersive visualization environments.

The biological structures visualized in this demonstration span the dynamic multi-scale processes at the heart of information processing in the nervous system. These structures reside in the mesoscale: a dimensional range of nm^3 to 100s of μm^3 Elucidation of how changes in structure affect differences in function within this dimensional range is fundamental to understanding phenomena such as neuroplasticity, associated with learning and memory, or the characteristic features of neuropathology such as in Alzheimer's disease or Parkinsonian disorders.