April 1, 2010
Position Paper for NSF Campus Bridging Workshop

The OptIPortal, a Scalable Visualization, Storage, and Computing Termination Device for High Bandwidth Campus Bridging

(Excerpted from Future Generation Computer Systems/The International Journal of Grid Computing: Theory, Methods and Applications, Elsevier B.V., Vol 25, issue 2, February 2009, pp. 114-123)

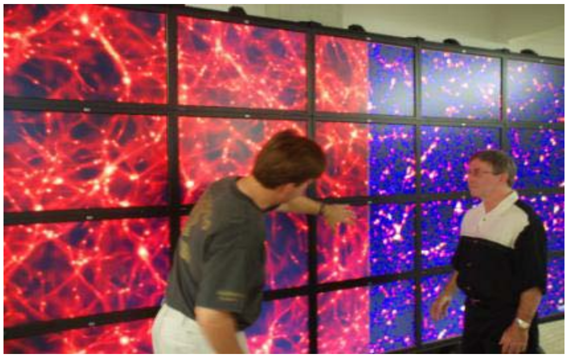
Authors: Thomas A. DeFanti[1,], Jason Leigh[3], Luc Renambot[3], Byungil Jeong[5], Alan Verlo[3], Lance Long[3], Maxine Brown[3], Dan Sandin[3], Venkatram Vishwanath[6], Qian Liu[1], Mason Katz[2], Phil Papadopoulos[2], Joseph Keefe[1], Greg Hidley[1], Greg Dawe[1], Ian Kaufman[1], Bryan Glogowski[1], Kai-Uwe Doerr[1], Javier Girado[4], Jurgen P. Schulze[1], Falko Kuester[1], and Larry Smarr[1]

Affiliations:
[1] California Institute for Telecommunications and Information Technology (Calit2), University of California San Diego (UCSD)
[2]San Diego Supercomputer Center, University of California San Diego (UCSD)
[3] Electronic Visualization Laboratory (EVL), University of Illinois at Chicago (UIC)
[4] Qualcomm, Inc.
[5] Texas Advanced Computing Center (TACC), University of Texas, Austin
[6] Argonne National Laboratory

As dedicated fiber optics are deployed on campuses to bridge between data-intensive end-users and regional or national-scale optical networks, the data flow into a user's lab will go up by two-to-three orders of magnitude. This means that the shared internet termination device, the PC or server, must be scaled up as well to maintain proper impedance matching with the data flow. This scaling must be not only in storage, but also in computing power and visualization "pixel real estate" to enable scalable analysis.

Fortunately, the National Science Foundation (NSF) funded the OptIPuter project for eight years (2002-2009), and one of its research results is the OptIPortal, just such a scalable termination device. This paper describes the software systems that have been developed for the OptIPortal and gives pointers to where one can download the software and locate recipes for the hardware requirements. The main point of the OptIPuter project was to examine a "future" in which networking was not a bottleneck to local, regional, national and international computing. This is one of the key goals for NSF's campus bridging program. OptIPortals are designed to allow collaborative sharing over 1-10 Gigabit/second networks of extremely high-resolution graphic output, as well as video streams.

The OptIPortal is constructed as a tiled display wall. OptIPortals typically consist of an array of 4 to 100 LCD display panels (1- 4-megapixels each), driven by an appropriately sized PC or cluster of PCs with optimized graphics processors and network interface cards. Rather than exist as one-of-a-kind laboratory prototypes, OptIPortals are designed to be openly and widely replicated, balancing the state of the art of PCs, graphic processing, networks, servers, software, middleware, and user interfaces, and installed in the context of a laboratory or office conference room. Some feature 3D stereo display panels (see NexCAVE and REVE below). They represent a constantly evolving technology. It is estimated that ~100 OptIPortals have been built globally and are in active use.



Some OptIPortals build on NSF's investment in the Rocks software system, which allows an end-user to easily install software across one's cluster. Since Rocks is the software environment upon which these OptIPortals are based, the hardware requirements for the OptIPortal are essentially those for Rocks, once the choice of display is made. Most of the deployments of OptIPortals have been done on commodity hardware, running Intel or AMD processors. Configurations are possible in which each computer in the cluster can drive one, two or more displays, depending on the performance and capabilities of the chosen graphics interface. OptIPortals can be optimized for specific functionality in terms of processor speed, network bandwidth, storage capacity, memory availability, and cost.

Rocks provides an easy way to configure an OptIPortal's display cluster, though OptIPortal middleware scales across different operating systems, operating system flavors and heterogeneous clusters. The middleware hides OS specific aspects and provides a cross-platform API. Locally available resources, such as the number of available graphics cards, displays and associated capabilities (resolution, swap and frame synchronization, etc.) can be probed at the device driver or the window manager level, allowing the middleware to report and adapt to hardware capabilities. Considering the number of PCs in a typical OptIPortal, mean time to failure becomes an important parameter when selecting cluster management strategies. From a system administrator's perspective, Rocks-based systems are easy to manage, largely by pruning system management overhead down to a single node.

Middleware and applications leveraging OptIPortal technology can be grouped into three major categories: stream-centric techniques, parallel distributed rendering techniques, and hybrid systems combining distributed real-time rendering and streaming within the same context. These in turn can scale from low-level visual content distribution approaches to high-performance parallel real-time rendering engines with multithread CPU support and GPU-based hardware acceleration.

OptIPortal head nodes and graphics nodes are networked together with 1 Gb/s or 10 Gb/s switches and network interface cards (NICs). Inexpensive switches allow onboard 1 Gb/s ports to be easily used, usually with a 1 Gb/s or 10 Gb/s uplink to the servers/campus networks. More expensive switches (e.g., Arista) and 10 Gb/s NICs allow much faster loading of large images and models, and, of course, facilitate streaming HD and 4K video. The latest motherboards now support up to 4 dual-ported graphics cards (GPUs), which will drive 8 two or four megapixel displays, but the load on the PC and NIC becomes quite high (just like putting a lot of disks on a PC would). One excellent feature of such systems is that GPUs can have up to 480 graphics cores and 1.792GB each, which is 1920 CUDA-programmable processors and 7GB of memory per PC. The optimal choice of OptIPortal motherboards and their NICs and GPUs is a constantly challenging design task.

1. Stream-Based Systems

SAGE (Scalable Adaptive Graphics Environment), initially funded by the OptIPuter award and now funded by NSF to harden and deploy to its growing user community, targets especially high-resolution tiled display systems, which can potentially cover all the walls and tabletop surfaces in a room, and which are interconnected to data sources and/or other OptIPortals with multi-10Gb/s optical networks. It operates on the assumption that as wall sizes increase, multiple users will naturally find a need to make full use of the available resolution to juxtapose multiple visuals and interact with them at the same time. It also assumes that it is possible for any type of application, given the appropriate middleware, to send a pixel stream to the SAGE tiled display.

SAGE middleware directs each of the incoming pixel streams from an application to the correct portion of a tiled wall allowing the system to scale to any number of streams and tiles. More importantly, it allows multiple applications on multiple distributed rendering clusters to run simultaneously and be viewed simultaneously on the tiled display, in essence, a true multi-tasking operating system for tiled displays. Anything from a parallel OpenGL application to a HD/4K video stream to a remote laptop can be displayed on the tiled display as long as the pixels from their image buffers can be extracted.

SAGE also features a capability called *Visualcasting* whereby dedicated clusters can be placed at high-speed network access points to replicate incoming pixel streams and broadcast them to multiple tiled displays at the same time, enabling users on distributed OptIPortals to look at the same visuals and therefore work collaboratively. The number of Visualcasting cluster nodes can be adjusted to suit the anticipated number of streams. This capability has been successfully demonstrated over transoceanic links. Addition of trackers or cameras for gesture input allows for richer control and interaction.

2. Parallel Distributed Rendering

Many software packages distribute visual content exclusively to multiple rendering engines in a parallel master-slave or client-server approach. A common shortcoming by most packages is scalability across multiple display tiles connected to a single machine, when the combined tile resolution exceeds the supported OpenGL display context size.

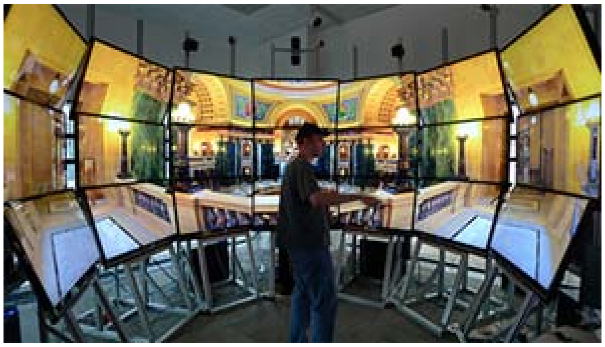
3. Hybrid Systems

CGLX explores an approach where high performance real-time parallel rendering and streaming of visual content from other applications can be combined. The middleware is based on the assumption that the rendering nodes in a cluster have sufficient CPU and GPU resources at their disposal. The framework can leverage from these resources by utilizing classical work distribution strategies in cluster systems such as culling and multi-threading for OpenGL applications and provides a freely programmable API in combination with a native container-based distributed desktop management application which accepts multiple pixel streams. To maximize the availability of network resources for data transmission related to the visualization content, CGLX implements its own lightweight network layer and message passing environment. CGLX provides users with access to parallel hardware accelerated rendering on different operating systems and aims to maximize pixel output to support high resolution tiled display systems. Natively, CGLX maps an OpenGL context to each display tile, resulting in multiple contexts when multiple displays are connected per node. This attribute makes CGLX the only fully scalable OptIPortal interface currently available.

Crucial for all distributed rendering approaches is the availability of a reliable high performance network to retrieve massive data content or to control the visualization system itself. An OptIPortal features a network solution that can provide data transfer rates up the 10Gbits/s. These maximum values can be maintained due to dedicated high performance local networks or a high-speed network grid such as OptIPuter. The access to vast amounts of distributed storage and computational resources on an OptIPortal and the additional network bandwidth enables stream-based approaches to dramatically increase their achievable performance. High performance real-time parallel visualization systems, which can also act as rendering back ends for stream-based approaches, can leverage these network resources to load and process data at remote sites and to simply stream the final results at interactive rates. This attribute of OptIPortals allows users to share, exchange and manipulate remote data sets interactively in distributed cooperative workspaces spanning the globe.

4. OptIPortal Virtual Reality/3D Systems: NexCAVE and REVE

The NexCAVE (.calit2.net/newsroom/article.php?id=1584) is a multi-panel, 3D virtual reality display that uses JVC HDTV 3D LCD screens in an array. When used with polarized stereoscopic glasses, the NexCAVE's modular, micropolarized panels and related software make it possible for a broad range of scientists — from geologists and oceanographers to archaeologists and astronomers — to visualize massive datasets in three dimensions, at a level of detail impossible to obtain on a single-screen desktop display. The NexCAVE's technology delivers a faithful and deep 3-D experience with excellent color saturation, contrast, and stereo separation. To present stereo imaging with a modified consumer HDTV, the JVC panels have a transparent surface applied that circularly polarizes alternate horizontal lines of the screen clockwise and anticlockwise. Lightweight passive polarized glasses filter out, for each eye, the corresponding clockwise or anticlockwise images. Since these HDTVs are very bright, 3-D data in motion can be viewed even with standard fluorescent room lights on. Thus, like other OptIPortals (and unlike projection-based VR systems), the NexCAVE will fit in any office or lab, and can be viewed in normal ambient light. The 10-panel, 3-column prototype at Calit2 has a ~6000x1500 pixel resolution, while a 21-panel, 7-column version built for KAUST has ~15,000x1500-pixel resolution in a semi-circular surround configuration.

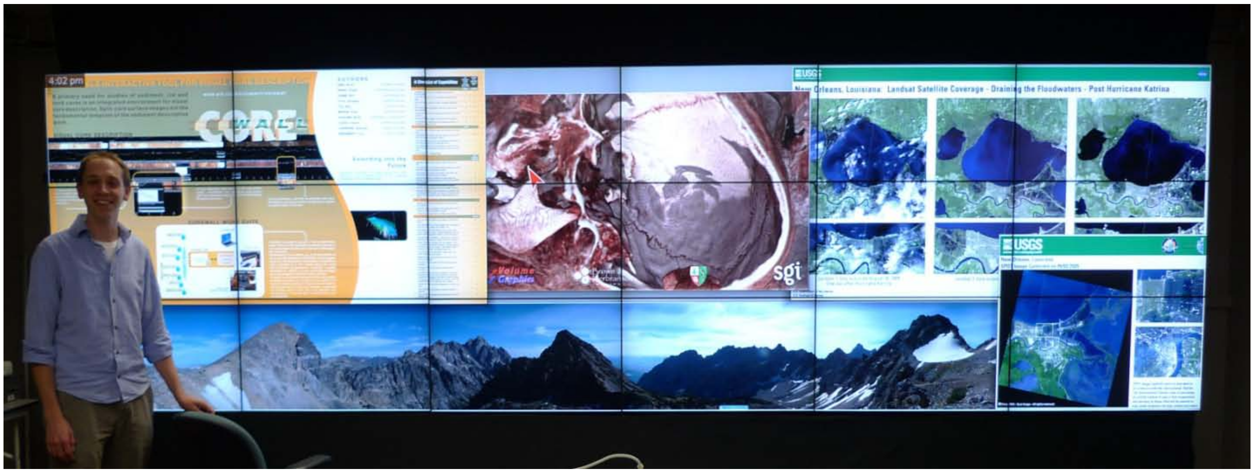


The REVE (“Rapidly Expandable Virtual Environment”) uses passive lenticular lens HDTV panel 3D technology from Alioscopy, Inc. to present very bright images to the viewer without requiring stereo glasses. The ~3:1 loss of resolution caused by autostereo spatial multiplexing is made up for by tiling the displays. Calit2 has a 6-panel REVE OptIPortal, and KAUST has an 18-panel one.

We currently support three software environments to drive the NexCAVE and the REVE: OpenCover, which is the OpenSceneGraph-based VR renderer of COVISE, CGLX, and EVL’s Electro. We use ROCKS-based OS distribution and management to quickly install and recover nodes.

5. Almost Entirely Seamless OptIPortal (the AESOP)

AESOP is a nearly borderless tiled display wall built with 46” NEC ultra-narrow bezel 720p LCD monitors. These NEC displays have inter-tile borders that are 7mm thick when tiled edge-to-edge within the framing, virtually eliminating the “window pane” effect of the “classic” OptIPortal’s 35mm tiled borders. Calit2 has one configurable 16-tile (4x4) AESOP. EVL has an 18-tile (3x6) AESOP, shown above in its Cyber-Commons room. EVL built the first AESOP in the summer of 2009 using hardware funds from an existing NSF grant, shown above in EVL’s Cyber-Commons room. (*Cyber-Commons* is EVL’s term for a technology-enhanced meeting room that supports local and distance collaboration and promotes group-oriented problem solving.) Calit2 built its AESOP shortly thereafter, also with NSF funds. These displays are scalable, support audio, and have networking and software sufficient to connect these displays to each other, to local servers, and to servers and similar displays worldwide. These OptIPortals run CGLX (UCSD) and SAGE (EVL) software, which support current and future applications.


4:02 pm
CORE
sgi
USGS
New Orleans, Louisiana: Landsat Satellite Coverage - Draining the Floodwaters - Post Hurricane Katrina
USGS