An Experimental OptIPuter Architecture for Data-Intensive Collaborative Visualization

Jason Leigh, Luc Renambot, Thomas A. DeFanti, Maxine Brown,
Eric He, Naveen Krishnaprasad, Javid M Alimohideen Meerasa,
Atul Nayak, Kyoung Park, Rajvikram Singh,
Shalini Venkataraman, Chong (Charles) Zhang
Electronic Visualization Laboratory, University of Illinois at Chicago (UIC),
Chicago, Illinois 60607
(312) 996-3002 voice, (312) 413-7585 fax, cavern@evl.uic.edu

Drake Livingston, Michael McLaughlin
Glimmerglass Networks, Inc.
26142 Eden Landing Road
Hayward, CA 94545
(510)723-1900 voice, (510)790-9851 fax, drake@glimmerglass.com

Abstract

This paper describes the OptIPuter's networking model and the visualization tools that are being developed to take advantage of the model. The model proposes the use of photonic switches to direct light through optical networks to create distributed computing pipelines for supporting large scale, interactive data exploration. This paper also describes a way to implement extremely high bandwidth multicasting over photonic networks to support high resolution graphics distribution in collaborative work involving large scale data.

1. Introduction

The OptIPuter [1] is a National Science Foundation funded project between the University of Illinois at Chicago, and the University of California, San Diego, to interconnect distributed storage, computing and visualization resources using a backplane constructed from a grid of deterministic high speed networks. In partnership with the Scripps Institute, US Geological Survey's EROS (Earth Resources Observation System) Data Center, and the Biomedical Informatics Research Network, the specific *application* goal of the project is to develop advanced computing systems to support collaborative data exploration in the Geosciences and the Neurosciences.

The OptIPuter computing model assumes that network bandwidth is no longer the scare resource, but instead the computing and storage systems that the network interconnects are the scare resource. This is based on the observation that the rate of growth of bandwidth (doubling every eight to twelve months) is far exceeding the rate of growth of computing, as predicted by Moore's law (doubling every eighteen to twenty four months). In this model the network becomes the backplane and the clusters of computing systems become the computer peripherals. For example, a cluster of computers with high end graphics cards is considered a single giant graphics card; and a cluster with terabytes of parallel file storage is considered a single giant disk drive. The challenge is to build the hardware and software architecture to realize this high level computing model, and to measure its benefits over the traditional cluster computing model.

While the bandwidth-computing *inversion* is providing the opportunity to explore this new model of computing, it also creates a significant problem. That is, while the cost of network bandwidth is dropping dramatically, the cost of the routers that will be needed to route information between these high speed links is not- in fact if the bandwidth growth curve is increasing faster than the Moore's law curve, it stands to reason that the routers needed to handle that increase in capacity will cost substantially more than the cost predicted by Moore's law. This is because at the present time routing is done electronically- that is, packets sent over an optical network

first have to be converted from photons to electrons before they can be routed.

In this paper we describe a novel OptIPuter architecture that uses photonic networks as an alternative to purely routed networks. The paper also describes ways to apply photonic networks to support collaborative data exploration. Finally the paper will describe a number of projects underway that are beginning to exploit our proposed photonic networking model.

2. The Photonic Data Exploration Pipeline

EVL has been examining an alternative approach to routers for directing data between computing systems. The solution stems from the observation that the predominant model for large scale data exploration consists of the following pipeline:

Data Sources→Data Correlation/Filtering System→Visualization System →Display System

In this pipeline, data (usually large collections of real or simulation data) are fed into a data correlation or filtering system which generates a sub-sampled or summarized result from which a visual representation is created and displayed on laptops, tiled displays, and even immersive CAVEs. In some cases several components of this pipeline may need to share a single compute cluster- for example it may be more economical in some cases to combine the data source with the data correlation component. In the context of supporting collaborative data exploration, the results of the pipeline (typically, but not necessarily limited to visualizations) will need to be multicast to multiple end points.

From a networking perspective the important pattern to recognize is that unlike web browsing on the Internet (which tends to involve users jumping from web site to web site), the connectivity between the computing components in a large scale data exploration pipeline tends to remain static once the connections are established. That is, it is unlikely that during the course of a computation a given cluster will randomly and frequently interact with a wide range of distributed clusters. This means that for the most part, fast packet-by-packet routing is unnecessary. All that is needed is an economical way to establish the connections between the major components of the pipeline.

The solution described in this paper involves the use of Photonic Switches- all-optical MEMS devices which do nothing more than direct light from one input port to an output port. MEMS, Micro Electro-Mechanical Systems, is a technology whereby tiny mechanical components are etched out of silicon wafers. Engineers use this technology to shape tiny micro-mirrors and micro-lenses in silicon and arrange these components into a small free-space optical system that focuses and redirects laser beams from inbound fibers to outbound fibers. Information contained in one or more wavelengths of light (lambdas) that comprise the light on each fiber can then be "routed" directly in the optical-domain without detection or signal regeneration.

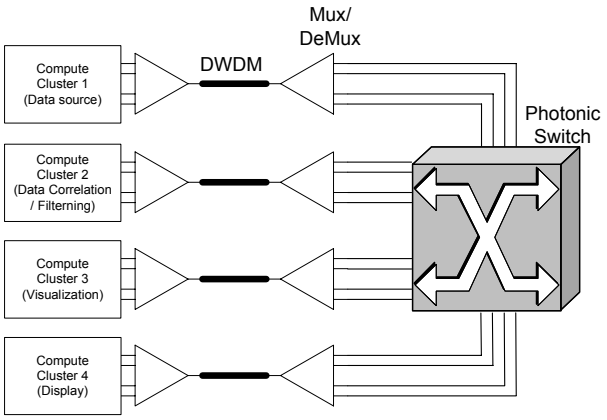

Figure 1: In the OptIPuter, a number of distributed compute clusters are interconnected via a photonic switch rather than with traditional routers. The optical transport platform performs the task of multiplexing and demultiplexing multiple light paths over a dense wave division multiplexed (DWDM), long haul network.

Vendors that produce such photonic switches include Glimmerglass Networks and Calient. Photonic switches are transparent in that the light sent along a fiber from any edge device is routed to its destination by optical reflections instead of protocol detections and signal regenerations. Photonic switches comprise a layer-1 space-switch that is configured in a star topology with many input and output fibers (e.g. 64 x 64). Fiber connections can be established in milliseconds – the switching time necessary for the micro-mirrors to move to new connection coordinates in order to properly redirect laser signals. Unlike traditional data switching products, once a fiber connection is established between any two network elements or edge devices - there is a direct, deterministic data path that is not shared. Multiple switches can be combined together to create very high-performance mesh networks that support fully dedicated high-speed data transmissions between any network elements.

By comparison with today's state-of-the-art "optical switches", photonic switches are tiny, relatively inexpensive and support massive bandwidths. Furthermore, the data rates they can switch are a function of the edge devices of the fiber network and not the switch itself. This eliminates problems of congestion often encountered in traditional routed networks. Applications can safely use aggressive transport protocols to move data between the edge devices. Also, because photonic switches are protocol transparent, new network applications can be rapidly employed as long as the edge device transmitting into the network is paired with a destination device able to receive and understand its protocol and data rate. In a local area network this can be achieved by simply connecting gigabit Ethernet cards to each other. In wide area networks however, one typically needs to multiplex several wavelengths of light onto a single Dense Wave Division Multiplexed fiber for long

haul transmission. Figure 1 illustrates how one would interconnect a collection of computing clusters using a photonic switch and an optical transport platform that performs the necessary multiplexing and demultiplexing of light paths. When combined with proper scheduling of cluster computing resources and light paths, this architecture allows applications to create multiple, simultaneous, distributed computing pipelines.

3. Photonic Multicasting

Earlier we had indicated that in order to support collaborative data exploration, one special capability required of the computational pipeline, is that at some point along the pipeline, there must be a way to multicast the stream to multiple endpoints. Again, since employing traditional electronic routers would be prohibitively expensive, a photonic solution is required. Prior research exists to *suggest* how one might build such systems[2][3][4], but there has been no attempt to *actually* build and test such systems in real applications. Glimmerglass Networks' family of REFLEXION™ photonic switches supports Photonic Multicasting that switches the content of one or more source fibers to multiple output fibers simultaneously. Standard implementations support a pair of 1:2 or 1:4 multicast configurations. Photonic multicasting can support lambda multicasting to act as a very high-performance data-stream replication device that delivers perfectly copied data-streams at any transmission speed that are always exactly in phase and remain perfectly synchronized. Because REFLEXION's multicasting technology is integrated into the photonic switch, REFLEXION can act as a network-wide data-stream resource for edge-devices participating in high-performance graphics, video, visualization, collaboration and other parallel computing applications.

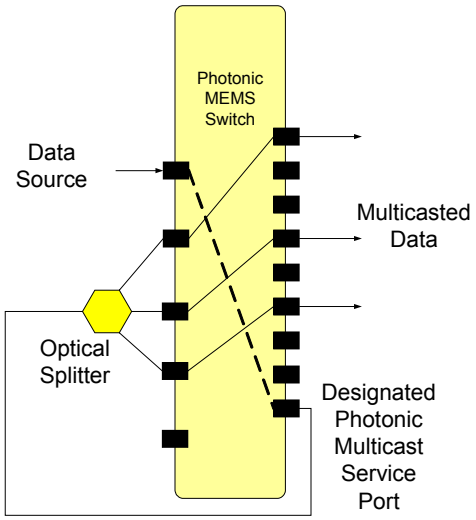

Figure 2: Schematic of how to integrate practical photonic multicasting in a photonic switch.

Figure 2 shows conceptually how photonic multicasting is achieved. The photonic switch allows an application to direct an incoming light path to any outgoing port on the switch. By designating an output port as a multicast port, one can feed the light directed to this port, into an optical splitter whose output is sent to a number of pre-assigned input ports. Using this scheme an application can choose any of the N multicast destinations on the photonic switch. Similar to the photonic switches, the optical splitters are all-photonic devices. Each optical splitter merely takes an input light signal and splits it into two or more copies of itself. Note however, that as one increases the replication-ratio from 1:2 to 1:4 and beyond, the optical power of a laser signal directed along each individual fiber is naturally reduced. Glimmerglass has tightly integrated an optical wave splitting function into its core photonic switch in order to ensure low optical loss performance. This is crucial to realize a practical implementation of this novel approach into a fiber network comprised of servers with standard GigE NICs. Depending on the particular network scenario, additional active optical amplification equipment may be required as the replication-ratio increases. REFLEXION's Photonic Multicasting architecture allows specialized implementations that support replication-ratios to 1:8 and beyond.

Note also that since photonic multicasting involves a 1-to-many split the connected hosts must use a unidirectional transport protocol for data delivery. Any acknowledgements that need to be transmitted in the opposite direction by the protocol (for example, to signal the loss or receipt of a packet) will have to travel over a separate connection. This is one of the foci of Quanta- a project to develop a suite of communication tools for extremely high speed networks[13].

In Figure 3 we illustrate how to integrate the photonic multicasting service into the photonic data exploration pipeline. In this illustration, the final visualization created by the pipeline is multicasted to several collaborating viewers.

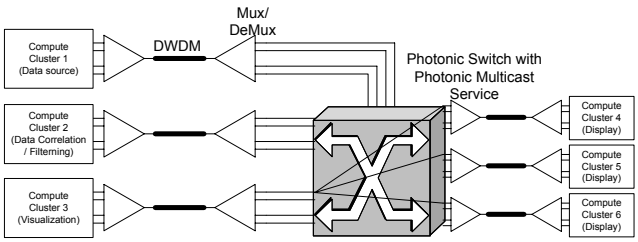

Figure 3: Integration of Photonic Multicasting into the Photonic Data Exploration Pipeline.

4. The Photonic Computing Engine

Another application of a photonic switch is as a data distribution backplane for a local compute cluster. Consider a cluster node with three optical, gigabit network interface cards. We call this a Photonic Computing Unit (PCU). Multiple PCUs can be interconnected to create a Photonic Computing Engine. Figure 4 illustrates how to create a sort-last graphics rendering engine [18] using this concept. Figure 5 illustrates how the engine is created by connecting up the input and output ports of the photonic switch. Note that since these photonic switches tend to require single-mode fiber as their input and output, pairs of

ports will have to be used to connect between any two network interface cards on the PCUs.

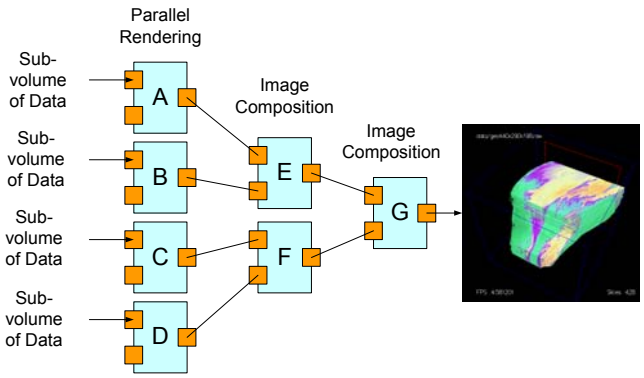

Figure 4: Photonic Computing Engine designed for Sort-Last Rendering

Extending this concept even further, Figure 6 shows how one might perform stereoscopic rendering by photonically multicasting the data to two separate rendering engines.

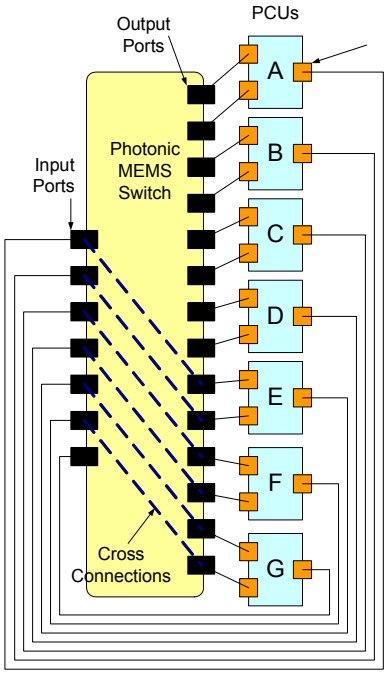

Figure 5: Creating the Sort-Last Rendering Engine using the Photonic Switch

5. Photonic Visualization Tools

A number of visualization tools are concurrently being developed to explore how to take advantage of the OptIPuter concept. These include: The Continuum- a collaborative project room that fuses together a broad range of display technologies; JuxtaVision- software to display extremely high resolution digital montages on tiled displays; TeraVision- hardware for collaborative high resolution graphics streaming; and TeraScope- a general software framework for large scale visual data exploration.

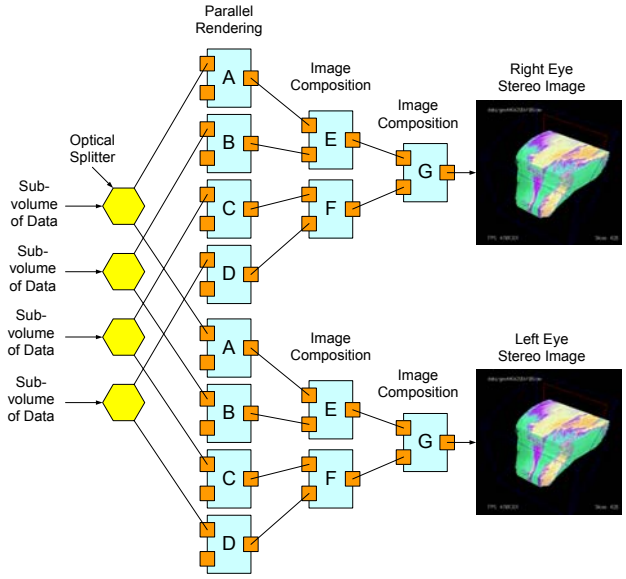

Figure 6: Taking advantage of photonic multicasting to mirror data to two Photonic Computing Engines.

5.1 The Continuum : Research in Display-Rich Collaboratories

The **Continuum** is a project between EVL, and the Technology Research, Education and Commercialization Center (TRECC) [9] in DuPage, Illinois, to research and develop the technology and human factors research needed to support distributed, collaborative project rooms. The Continuum's space (referred to as *The Continuum*) consists of a number of modular technologies that are designed to support different aspects of a collaborative campaign. These Amplified Collaboration Environments (ACEs) are connected to each other and to distributed data, computing and visualization resources. Because we see these ACEs as *synchronous* collaboration environments, the emphasis is on immediacy of access to data and visualization. This immediacy can only be achieved by massive and deterministic bandwidth connecting the collaborating sites with the distributed resources. The OptIPuter therefore plays a crucial role in supporting ACEs. At the same time, ACEs are crucial to the OptIPuter because they provide the human-interface to the OptIPuter.

The display portion of the Continuum is driven by a cluster of 8 networked PCs. The Continuum's Conferencing Module provides both H323 (via Polycom) and AccessGrid video conferencing [7]. The Immersion Module consists of a passive stereo virtual reality display for visualizing three-dimensional data sets (often called the GeoWall- by the Geoscience community) [11]. The Content Distribution Module provides a scalable LCD tiled display (called a PerspecTile) for viewing high resolution visualizations as well as mosaics of disparate visualizations. The Annotation Module employs a plasma screen enhanced with a Smartboard [8] to provide a digital whiteboard, powerpoint presentation screen and web-browser. Finally the Wireless Interaction Module allows collaborators to control the Continuum using laptops and TabletPCs as if it were one large seamless desktop.

Figure 7 shows a fully constructed Continuum-space at the Electronic Visualization Laboratory. With two Continuum-spaces already built at EVL and one at TRECC, the focus of the research now is to develop a unified software system to allow collaborators to seamlessly manipulate the information that are displayed in the Continnum. A number of researchers in the past have already addressed a variety of issues in working in ACEs [17]. Our emphasis is on supporting the collaborative manipulation of large scientific objects (data objects on the order of hundreds of gigabytes to terabytes).

Figure 7: The Continuum. Top left is a passive stereo display for showing immersive 3D content; next to it are vertically stacked plasma screens that are used for AccessGrid video conferencing; to the right of this is the plasma touchscreen. The small screens in front of the collaborators form a tiled display that can also be mounted in a 2 X 2 matrix.

5.2 JuxtaVision: Display of Large-Scale Digital Montages

Scientists are finding an increasing need to display visualizations on extremely high-resolution displays, such as EVL's 6000x3000-pixel LCD tiled display, called the PerspecTile (now dubbed the GeoWall2 by the Geoscience community). The PerspecTile is driven by a cluster of PCs. JuxtaVision is software designed specifically for scalable tiled displays to visualize extremely high-resolution images. JuxtaVision uses a networked read-only memory system called LambdaRAM [16] to provide predictive paging of image data for visualization. LambdaRAM is a system that uses the OptIPuter concept to create large memory pools to serve as data caches for overcoming access latency over long distance networks.

Figure 8 shows EVL's PerspecTile displaying Scripps Institute's satellite bathymetry for the entire Earth. The dataset is unique as it provides the first view of ocean floor structures in remote areas of the Earth. These kinds of maps have a variety of applications, including the study of plate tectonics and undersea volcanoes, to name a few.

Figure 8: JuxtaVision showing Bathymetry Data of the Entire Planet.

EVL is also applying this technology to the visualization of aerial photography images provided by USGS EROS Data Center as part of a homeland defense initiative. In this application, aerial photographs of 133 urban areas are being photographed at a 3-foot resolution. This amounts to a total of approximately 38 Terabytes of image data that JuxtaVision must be able to seamlessly navigate through.

5.3 TeraVision: High Resolution Graphics Streaming for the OptIPuter

TeraVision is a gigabit network appliance to allow scientists to share their computer screens during a collaborative work session [15]. Unique to TeraVision is that it does not require any special modification of scientists' software or hardware to share their computer screens. A scientist simply plugs the VGA or DVI output of his computer directly into the TeraVision box and it automatically captures and streams the visualization to other distantly located TeraVision boxes for remote display. Two boxes can handle stereo. Using our photonic multicast concept, multiple TeraVision boxes will be able to stream entire high-resolution tiled display images to multiple remote sites at the same time.

5.4 TeraScope: Visual Tera Mining on the OptIPuter

TeraScope is a project between EVL and the Laboratory for Advanced Computing at UIC to develop a framework for large scale, and real time data access, correlation, and visualization on the OptIPuter [16]. The goal is not to generate brute-force visualizations of terabyte data sets, but to employ data mining techniques, when possible, to create visual summaries of the data. Experience from the TeraScope project has allowed us to develop WiggleView, a tool for visualizing real-time and historically recorded seismic data from a network of seismometers distributed around the world [20]. WiggleView operates on both the GeoWall and the PerspecTile. Figure 9 shows a series of seismoglyphs that summarizes a sequence of seismic events on 15 seismometers around the world. Data is streamed in real-time from data servers administered by

the IRIS Consortium [19]- which warehouses seismic data for NSF's Earthscope cyberinfrastructure initiative.

6. Closing Remarks

This paper has introduced a novel way of connecting computing systems using photonic networks. We are currently building a photonic network consisting of several photonic switches. One is housed at the StarLight [10] facility in downtown Chicago, another is at UIC, and the third is at the University of Amsterdam (an OptIPuter partner). Concurrent with the build-out of the networking infrastructure we are beginning to prototype a variety of OptIPuter-based visualization tools. Furthermore, work is underway to build a Photonic Computing Engine for volume rendering and to compare this against traditional remote visualization techniques.

Figure 9: WiggleView showing a series of seismoglyphs that summarizes real-time seismic data from remote seismometers distributed around the world.

7. Acknowledgments

The visualization and advanced networking research, collaborations, and outreach programs at the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago are made possible by major funding from the National Science Foundation (NSF), awards EIA-9802090, EIA-0115809, ANI-9980480, ANI-0229642, ANI-9730202, ANI-0123399, ANI-0129527 and EAR-0218918, as well as the NSF Information Technology Research (ITR) cooperative agreement (ANI-0225642) to the University of California San Diego (UCSD) for "The OptIPuter" and the NSF Partnerships for Advanced Computational Infrastructure (PACI) cooperative agreement (ACI-9619019) to the National Computational Science Alliance. EVL also receives funding from the US Department of Energy (DOE) ASCI VIEWS program. In addition, EVL receives funding from the State of Illinois, Microsoft Research, General Motors Research, and Pacific Interface on behalf of NTT Optical Network Systems Laboratory in Japan.

8. References

[1] The OptIPuter : www.evl.uic.edu/cavern/optiputer

[2] G. N. Rouskas, "Optical Layer Multicast: Rationale, Building Blocks, and Challenges," IEEE Network, Jan/Feb 2003, pp. 60-65.

[3] J. Leuthold, C. H. Joyner, "Multimode Interference Couplers with Tunable Power Splitting Ratios", IEEE/OSA J. Lightwave Tech., vol. 19, no. 5, May 2001, pp.700-706.

[4] W. S. Hu, Q. J. Zeng, "Multicasting Optical Cross Connects Employing Splitter-and-Delivery Switch", IEEE Photonics Tech. Lett., vol . 10, July 1998, pp. 970-972.

[5] S. Teasley, L. Covi, M. Krishnan, and J. Olson, "How does radical collocation help a team succeed?" In proceedings of CSCW'00 (Philadelphia, Dec. 2-6), ACM Press, New York, 2000, pp. 339-346.

[6] J. S. Olson, L. Covi, E. Rocco, W. J. Miller, P. Allie (1998) "A Room of Your Own: What would it take to help remote groups work as well as collocated groups?" Short Paper the Conference on Human Factors in Computing Systems (CHI'98), 279-280.

[7] AccessGrid : www.accessgrid.org

[8] SmartTech Matisse Smartboard : www.smarttech.com

[9] Technology Research Education and Commercialization Center (TRECC) : www.trecc.org

[10] StarLight : www.startap.net/starlight

[11] The GeoWall Consortium : www.geowall.org

[12] G. N. Rouskas, "Optical Layer Multicast: Rationale, Building Blocks, and Challenges," In IEEE Network, Jan/Feb 2003, pp. 60-65.

[13] E. He, J. Alimohideen, J. Eliason, N. Krishnaprasad, J. Leigh, O. Yu, T. A. DeFanti, "QUANTA: A Toolkit for High Performance Data Delivery over Photonic Networks," to appear in Future Generation Computer Systems, Elsevier Science Press. 2003.

[14] J. Leigh, A. Johnson, K. Park, A. Nayak, R. Singh, V. Chowdhry, "Amplified Collaboration Environments," VizGrid Symposium, Tokyo, November 2002 (www.vizgrid.org).

[15] R. Singh, J. Leigh, T. A. DeFanti, F. Karayannis, "TeraVision: A High Resolution Graphics Streaming Device for Amplified Collaboration Environments," to appear in Future Generation of Computer Systems (FGCS), Elsevier Science Press. 2003.

[16] C. Zhang, J. Leigh, T. A. DeFanti, "TeraScope: Distributed Visual Data Mining of Terascale Data Sets over Photonic Networks," to appear in Future Generation Computer Systems, Elsevier Science Press. 2003.

[17] J. Borchers, M. Ringel, J. Tyler, A. Fox, "Stanford Interactive Workspaces: A Framework for Physical and Graphical User Interface Prototyping." IEEE Wireless Communications, special issue on Smart Homes. December 2002.

[18] K. Moreland, B. Wylie, C. Pavlakos, "Sort-last Parallel Rendering for Viewing Extremely Large Data Sets on Tile Displays." Proc. IEEE 2001 Symp. Parallel and Large-Data Visualization and Graphics, San Diego, California, 2001.

[19] The IRIS Consortium : www.iris.edu

[20] A. M Nayak, J Leigh, A Johnson, R Russo, P Morin, C Laughbon, T Ahern , WiggleView : Visualizing Large Seismic Datasets, Eos Trans. AGU, 83(47), Fall Meet. Suppl., Abstract U61A-0007, 2002.