The OptIPuter:

Larry Smarr inspects an iGrid visualization.

© Calit2

1047-6938/06/07-08/0040/6-$15.00 ©OSA

An Information Superhighway for Terabytes

Patricia Daukantas

Using a single computer to do scientific research is *sooooo* 20th century, according to supercomputing pioneer Larry Smarr. His OptIPuter team is laying the groundwork for linking computers all over the world via dedicated optical-fiber channels.

Once upon a time, scientific data sets came in small packages. A researcher might save some data onto a floppy disk, plot them on a two-dimensional graph, and then stare at the squiggly line for inspiration.

Today, in many fields, research generates billions, trillions or even quadrillions of bytes of data. Astronomers want to study digital photographs of the skies without traveling to remote observatories. Biomedical researchers, climate modelers and theoreticians make complex computer simulations that require them to move billions or trillions of data bytes between cities or countries.

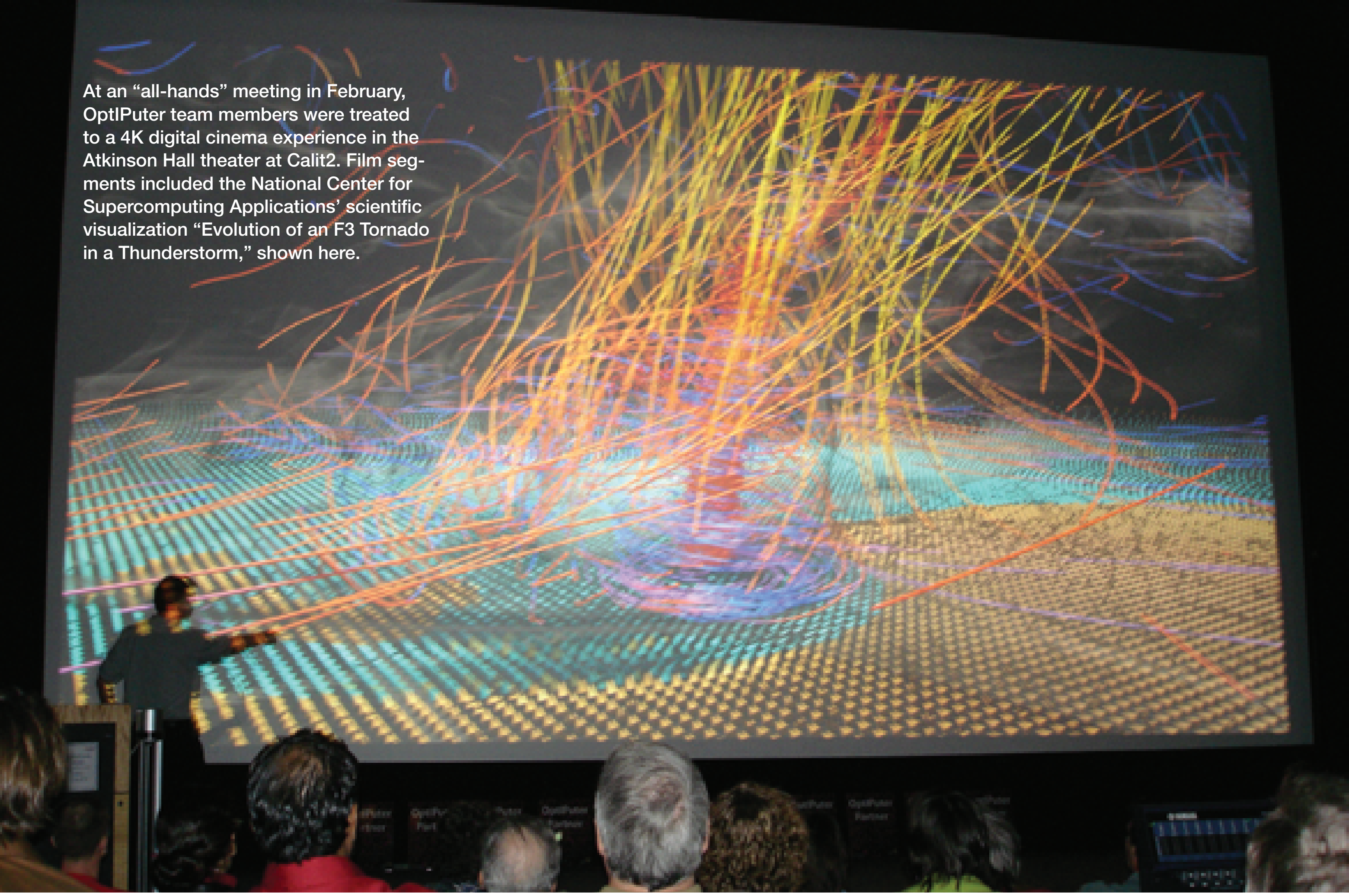
Larry Smarr envisions that such scientists will soon transfer their data through direct-connect virtual computers. Each one of these so-called OptIPuters—named for a combination of optical networking, Internet Protocol (IP) and computer storage—would link distant computer components together through easily reconfigurable optical channels that bypass standard Internet routers.

The OptIPuter project, which is now in the fourth of five years of funding from the National Science Foundation, is developing a toolbox of hardware, software and middleware that other researchers will be able to use. Team members have staged several technology demonstrations under the guidance of principal investigator Smarr, director of the California Institute for Telecommunications and Information Technology (Calit2).

An OptIPuter isn't a single, fixed entity like a desktop computer or server. Instead, it's a "cyberinfrastructure"—an evolving system for linking some of the world's most powerful computers with remote data storage libraries and "walls" of high-resolution, large-scale displays. Think of it as an optical-fiber alternative to the normal shared Internet—or, in Smarr's words, "a data freeway for science."

Smarr and his team—roughly 70 people at eight partner universities and a dozen other academic and industrial affiliates—realized that a computer's components no longer need to be all in the same room, said Steven J. Wallach, an analyst with CenterPoint Ventures of Dallas and an adviser to the OptIPuter project. The central processing units can be in one place, the data storage in a second and the dedicated visualization system in a third.

In other words, OptIPuter's connections could level the technological playing field for users of huge data sets.



At an "all-hands" meeting in February, OptIPuter team members were treated to a 4K digital cinema experience in the Atkinson Hall theater at Calit2. Film segments included the National Center for Supercomputing Applications' scientific visualization "Evolution of an F3 Tornado in a Thunderstorm," shown here.

"In essence, we saw that the world is flat before Thomas Friedman created that sound bite," said Wallach, referring to the *New York Times* columnist's recent bestselling book on globalization.

> With an OptIPuter and a dedicated connection, data sets can just flow from point A to point B without passing through any of the public Internet switches.

Transferring terabytes

Proponents of so-called grid computing have been championing the notion of widely distributed and shared computing resources for several years now.

According to Wallach, however, the OptIPuter team is taking the grid idea one step further by acting on the realization that the 30-year-old packet switching protocol that works so well for e-mail and other Internet applications was not designed for transferring terabytes of data over long distances.

That protocol, known as Transmission Control Protocol or TCP, breaks large data sets into chunks, and the routing equipment for managing the transmission of those data is relatively expensive (both financially and computationally) because every packet has to be examined along the way and reassembled at the other end. With an OptIPuter and a dedicated connection, data sets can just flow from point A to point B without passing through any of the public Internet switches. Optical fibers are playing a major role in this growing alternative.

The massive overdeployment of fiber-optic cable that occurred during the telecommunications bubble, which burst in 2000, has provided the OptIPuter team and other high-bandwidth networking proponents with lots of available infrastructure.

The burst radically changed the economics of bandwidth and the business model of networking, Smarr said. Because of the effective oversupply of bandwidth, the business models of networking changed, and it became possible for an institution, or even an individual scientist, to imagine getting a 20-year lease from a carrier for a personal 10-Gbit/s channel. Smarr calls such a wavelength-division-multiplexed channel a "lambda," after the Greek letter frequently used to denote wavelength.

Smarr has his own 10-Gbit/s channel for cross-country networking demonstrations. Dubbed the CAVEwave, the channel runs from the University of California, San Diego (UCSD), through the Pacific Northwest GigaPoP network testbed in Seattle to the University of Illinois at Chicago via the StarLight optical network exchange.

Right now, the OptIPuter uses the fiber-optic network of a university-industry consortium called the National LambdaRail (NLR), Wallach said. One fiber-optic link can carry 40 to 80 channels, each of which constitutes a channel that can be leased to an end user.

Whither the Internet?

The OptIPuter won't replace the shared Internet, Smarr said. Hundreds of millions of e-mail and Web users will continue to use it for anywhere-to-anywhere connectivity.

"If all we were doing was e-mail, we'd be completely happy with dial-up modems," said Thomas A. DeFanti, a Calit2 research scientist and one of four co-principal investigators on the OptIPuter project.

But there are many scientific applications that require, for example, 100 times the bandwidth that the shared Internet can provide. Yet there are only a few endpoints that need such enormous data pipes. For example, the largest elementary particle accelerator on the planet, the Large Hadron Collider (LHC), will switch on at CERN in Switzerland next year. Some LHC experiments will generate several terabytes of data per second.

"If you want to be a competitive particle physicist, you have to be able to get to that data," Smarr said. But trying to move such large data sets over the shared Internet would just cripple it, he added.

The Internet2 consortium, which links 207 research universities, provides 20-50 Mbit/s of connectivity among its users, Smarr said. But dedicated channels can transmit huge data sets at 10 Gbit/s (which is 10,000 Mbit/s).

In one analogy, computer users had to share a single mainframe long ago, Smarr said. They would submit their batch jobs and wait around, because they had no idea when their results would come back. The same program could take a variable time to run because the computing resource (the mainframe) had to be divided up in an unpredictable way.

When personal computers came along, individual users could have one all to themselves and no longer had to compete with other users for computing resources. A program would take the same amount of time to run on the single computer. "That was a tremendous help in terms of productivity," Smarr said.

Likewise, a single person can use a dedicated channel to move lots of data in a completely predictable way, Smarr said. He predicts that these clear optical pipes will be used to access remote scientific instruments, or perhaps to stream high-definition video between divisions of a multinational company.

In some situations, DeFanti said, researchers want people in more than one location to be able to peruse a set of detailed raw images or computationally massaged graphics, and such browsing can chew up a lot of bandwidth.

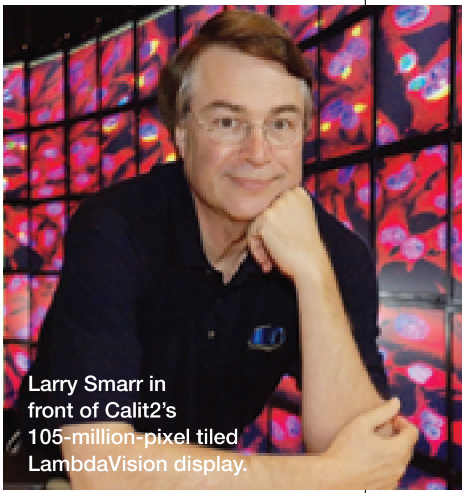
In essence, DeFanti said, the OptIPuter gives large computing clusters direct connections to each other through circuits that avoid going through routers whose data-sharing protocols, such as TCP, don't work well with real-time data deluges. The OptIPuter is working at a lower layer of networking intelligence than the expensive routers that handle the packet-switched traffic of the shared Internet.

[The Death and Rebirth of Supercomputing]

In the mid-1970s, Seymour Cray and his company, Cray Research, introduced the Cray-1 Supercomputer, the first of a series of machines designed to out-calculate almost all other existing computers. During the 1980s, Cray and his competitors built large super-mainframes around a handful of highly advanced processors.

About 15 years ago, however, computer engineers—as well as scientists who wanted better access to big, fast systems—started building computing "clusters" out of large numbers of commodity microprocessors working in parallel. Such clusters relied on newly developed software to break a complex problem into smaller pieces that could be solved by each processor in parallel and reassembled the results into a solution.

The parallel-processing model took off fast. In just a few years, institutions such as the U.S. Energy Department's weapons laboratories were purchasing these clusters for their most detailed scientific models. Most of the companies that built the older types of supercomputers got out of the business, leading analysts to proclaim the "death of supercomputing."



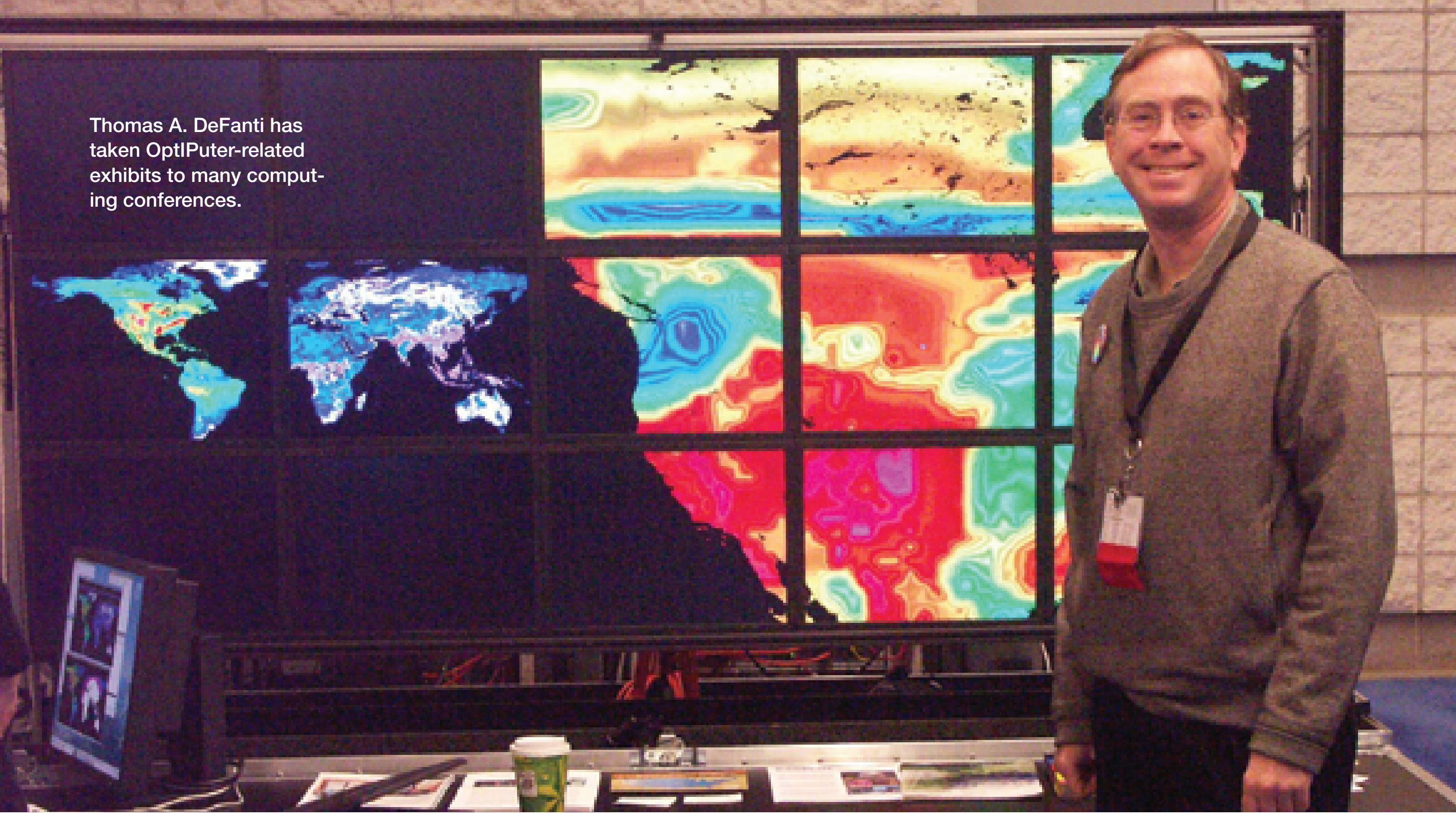
Larry Smarr in front of Calit2's 105-million-pixel tiled LambdaVision display.

But the death knell was premature. Thanks to grass roots efforts and ongoing research, supercomputing technology continued to advance—far beyond what Cray and his contemporaries had initially thought possible.

In the mid-1980s, Calit2 director Larry Smarr recommended that the National Science Foundation launch a national supercomputer center for the U.S. scientific community. The result was a set of five regional centers, including the National Center for Supercomputing Applications, which Smarr headed for 15 years before moving to the Calit2 university consortium.

Meanwhile, Steven J. Wallach, whose work was chronicled in Tracy Kidder's bestselling book *The Soul of a New Machine*, helped develop a new Data General "minisupercomputer." He later co-founded a supercomputer company called Convex Computer Corp., which Hewlett-Packard Co. acquired in 1994, and briefly helped manage an optical-switch company called Chiaro Networks Inc.

Today, massively parallel machines that can do trillions of floating-point operations per second ("teraflops") lead the revitalized supercomputing industry (often called "high-performance computing"), and the average PC has more raw processing power than the early Cray supercomputers, which weighed several tons each. A Web site, www.top500.org, tracks the world's fastest computers.



Thomas A. DeFanti has taken OptIPuter-related exhibits to many computing conferences.

© Calit2

The economics of bandwidth

The cost of low-level switching is about one-tenth the cost of routing because the end users are effectively doing more of the work themselves by specifying exactly where they want their data to flow, DeFanti said.

> The cost of low-level switching is about one-tenth the cost of routing because the end users are effectively doing more of the work themselves by specifying exactly where they want their data to flow.

"The service-level expectations when you do your own switching are different from when someone is providing a full-service network to you," DeFanti said.

"As Larry [Smarr] says, the idea of the OptIPuter is to go to the end of the rainbow where the bandwidth is the cheapest thing," DeFanti said. "Bandwidth will get cheaper and cheaper as we get more and more photonics involved and less electronics."

The OptIPuter team has installed optical switches from Glimmerglass Networks Inc. of Hayward, Calif., for testing and demonstrating the network between two of its major partner campuses, UCSD and the University of Illinois at Chicago.

For a one-day test in August 2005, researchers at NASA's Goddard Space Flight Center in Greenbelt, Md., used OptIPuter technology to connect their laboratory to the Scripps Institution of Oceanography at UCSD. The link ran from San Diego to Chicago via the CAVEwave; it then traveled from Chicago to another NLR point-of-presence in McLean, Va., and finally to Greenbelt through a separate channel.

Through the transcontinental channels, scientists at NASA Goddard were able to call up 20-GB Land Information System data sets at Scripps, render them with a UCSD visualization cluster, and show them on a large tiled display wall in Maryland.

There probably won't be a "central control" for the OptIPuter, Wallach said. Technology is in one of three phases: research, prototype or production. Smarr and his longtime collaborators tend to like making their prototypes available to the wider community. Wallach said that it wouldn't surprise him if someday there is a company called OptIPuter Systems to hook up these dedicated but temporary links between computational resources and their users.

Working out the kinks

The problems that the OptIPuter team has had to solve include finding an available channel and then making the collection of end points look like a single "distributed virtual computer" to the end user and his or her application, Smarr said. "The application people don't want to be bothered with optical fibers and lambdas."

Andrew A. Chien, the OptIPuter's software architect, was a computer science professor at UCSD until December 2005, when he joined Intel Corp. as vice president and research director. Chien and his team of UCSD students developed a middleware application, Distributed Virtual Computer (DVC), that makes the spread-out components of an OptIPuter look like a single computer to the end user's application, instead of a virtual machine built out of pieces scattered all over the globe.

In the 1980s, Smarr was a principal instigator of the NSF-funded program of regional supercomputer centers, which began with five locations and eventually got whittled down to two. However, he now calls the idea of a supercomputer center "really yesterday."

A typical research university campus might have 50 different Linux clusters, "none of which are centers of anything," Smarr said. Those local clusters can be set up as the terminators of a channel, just as a PC might be considered an endpoint of the Internet. For especially big projects, a researcher might tap into several channels, each joined with an optical connector into a different node of a parallel Linux cluster.

Traditionally, Smarr said, telecommunications and computers were separate worlds—they had different standards, different vocabularies, different everything. However, Gigabit Ethernet and 10-Gigabit Ethernet are two emerging standards that will be the same in both worlds. "This is radical," Smarr said.

Smarr said that Wallach's "Petaflops in the Year 2009" keynote speech at a Dallas supercomputing conference in 2000 was one of his inspirations for launching the OptIPuter project. (In the lingo of high-performance computing, "1 petaflops" stands for 1 quadrillion floating-point operations per second.)

The litmus test for choosing the shared Internet or an OptIPuter is to match the natural data size to the available bandwidth. The Internet is good for file sizes of up to about 20 megabytes, Smarr said. But medical imaging, satellite surveys and high-definition video have file sizes much larger than that. "It's clear you can't stuff something that's 1,000 gigabytes through the shared Internet in any kind of interactive way," he said.

Potential applications

Earth science and medical imaging are the initial applications for the OptIPuter, but new applications, such as digital cinema, are emerging. Calit2, Smarr's institute, has developed a project called CineGRID to explore ways of transferring digital motion pictures.

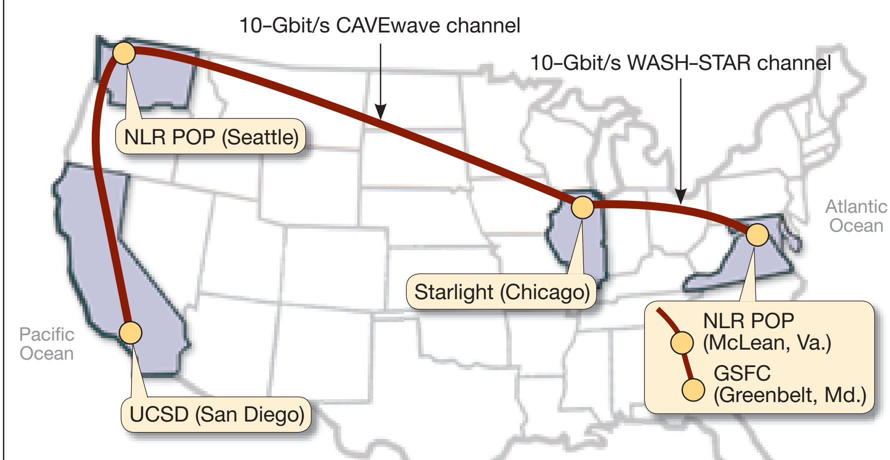
In the near future, the biggest use of the world's networks will not be e-mail or IP telephony, but video streams, Wallach said. An audio song in the popular MP3 file format is only about 3 megabytes, but a commercial DVD video disk holds about 4 gigabytes, and a high-definition DVD—which could well become the expected standard in a few years—will hold 25 gigabytes, give or take a few.

Two emerging digital-cinema standards have been dubbed 2K and 4K because they produce images that are 2,048 and 4,096 pixels wide, respectively. (Today's high-definition television has a horizontal resolution of only 1,024 pixels.) The file sizes that will be needed to send movies from studios to local theaters—at 24 frames per second times 8 megapixels per frame—could run into the hundreds of gigabytes for a typical feature film in 4K format. "You're going to need something like the OptIPuter to move that around at full resolution," Smarr said.

The motion picture industry is starting to explore secure methods of sending out digital movies to theaters, Smarr said. Digital cinema wouldn't need to go to every individual home; as in the flow of medical images to and from hospitals, there would be just a few fixed spots that would be set up to receive the high-bandwidth data transfers. Seven Hollywood studios have formed a consortium, called Digital Cinema Initiatives LLC, which has developed new technical specifications for audiovisual tracks, projectors, content storage and other aspects of digital theater.

In a way, OptIPuter file distribution would be a much simpler system than the Internet, "where the next minute you might have to get to anywhere of a hundred million endpoints in the world," Smarr said. ▲

[OptiPuter across the National LambdaRail]



The OptIPuter provided an express lane for data across the continent during an August 2005 technology demonstration.

[Patricia Daukantas (pdauka@osa.org) is senior writer/editor for *Optics & Photonics News*.]

[References and Resources]

- >> The OptIPuter project's Web site: www.optiputer.net.
- >> The National LambdaRail: www.nlr.net.
- >> The Land Information System: lis.gsfc.nasa.gov.
- >> Online publications covering high-performance computing and networking: HPCwire (www.hpcwire.com) and GRIDtoday (www.gridtoday.com).
- >> The Supercomputing conference series: www.supercomp.org.
- >> Digital Cinema Initiatives LLC, a consortium of movie studios: www.dcimovies.com/.