Building a Global Collaboration System for Data-Intensive Discovery

Distinguished Lecture

Hawaii International Conference on System Sciences (HICSS-44)

Kauai, HI

January 6, 2011

Dr. Larry Smarr

Director, California Institute for Telecommunications and Information Technology

Harry E. Gruber Professor,

Dept. of Computer Science and Engineering

Jacobs School of Engineering, UCSD

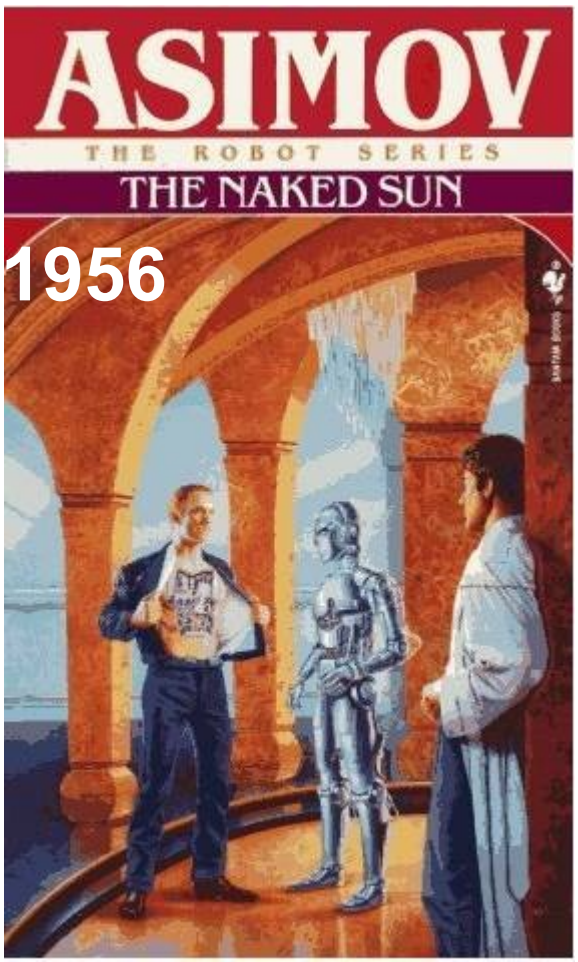
Abstract

We are living in a data-dominated world where scientific instruments, computers, and social interactions generate massive amounts of data, increasingly being stored in distributed storage clouds. Data-intensive discovery requires rapid access to multiple datasets and computational resources, coupled with a high-resolution streaming media enabled collaboration infrastructure.

The goal of this collaboration system is to allow globally distributed investigators to interact with visual representations of these massive datasets as if they were in the same room. The California Institute for Telecommunications and Information Technology has a variety of projects underway to realize this vision via the use of dedicated 10 gigabit/s optical "lightpaths," each with 1000x the typical bandwidth of the shared Internet.

I will share some examples of the use of such collaboration spaces to carry out data-intensive discovery from disciplines as diverse as bioinformatics, health care, crisis management, and computational cosmology and discuss the barriers to establishing such a global collaboration system which still remain.

Over Fifty Years Ago, Asimov Described a World of Remote Viewing

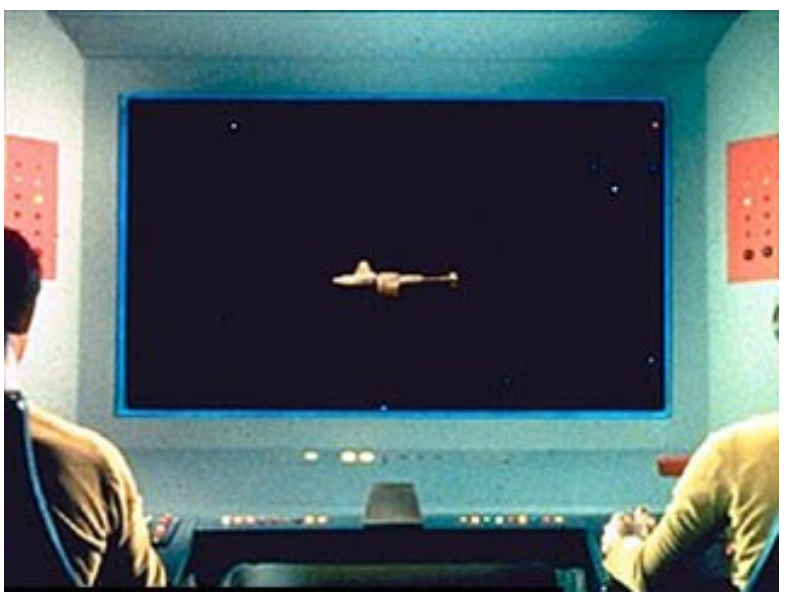


A policeman from Earth, where the population all lives underground in close quarters, is called in to investigate a murder on a distant world. This world is populated by very few humans, rarely if ever, coming into physical proximity of each other. Instead the people "View" each other with trimensional “holographic” images.

TV and Movies of 40 Years Ago Envisioned Telepresence Displays



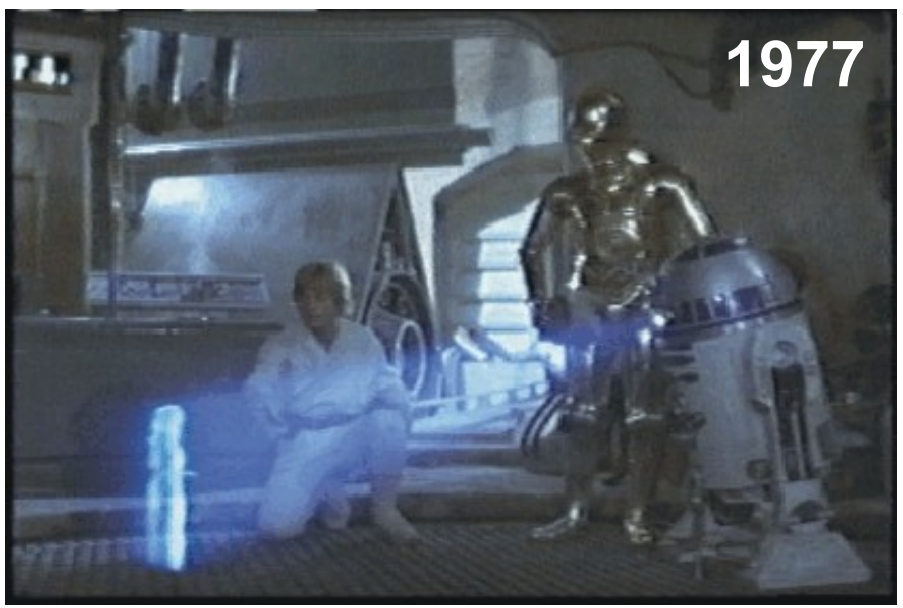






Source: Star Trek 1966-68; Barbarella 1968

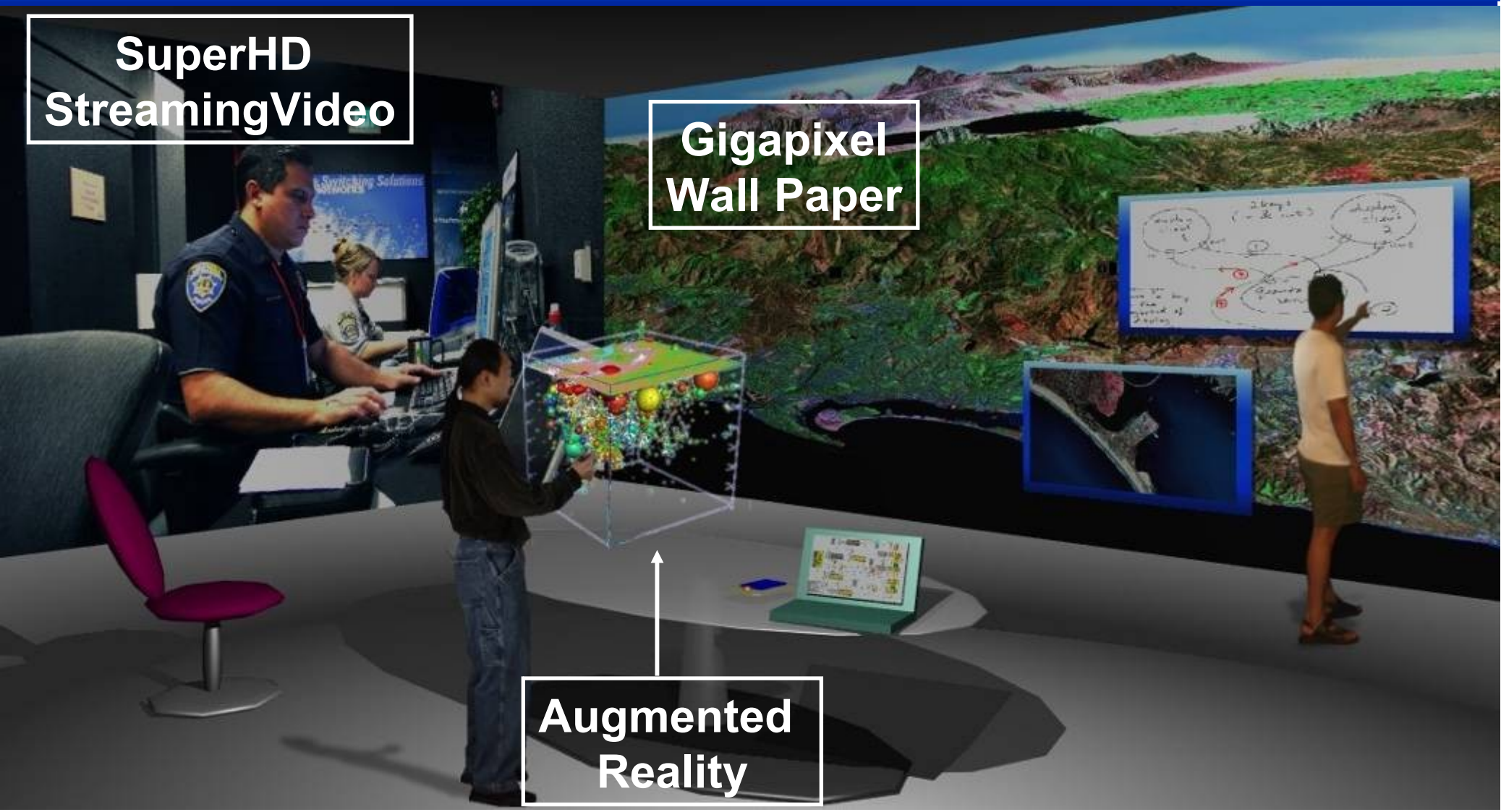
Holographic Collaboration Coming Soon?
Science Fiction to Commercialization





Over the Sixty Years from Asimov to IBM
Real Progress Has Been Being Made in Eliminating Distance
For Complex Human Interactions

A Vision for the Future: Optically Connected Collaboration Spaces



1 GigaPixel x 3 Bytes/pixel x 8 bits/byte x 30 frames/sec ~ 1 Terabit/sec!

Source: Jason Leigh, EVL, UIC

The Bellcore VideoWindow -- A Briefly Working Telepresence Experiment



(1989)

“Imagine sitting in your work place lounge having coffee with some colleagues. Now imagine that you and your colleagues are still in the same room, but are separated by a large sheet of glass that does not interfere with your ability to carry on a clear, two-way conversation. Finally, imagine that you have split the room into two parts and moved one part 50 miles down the road, without impairing the quality of your interaction with your friends.”

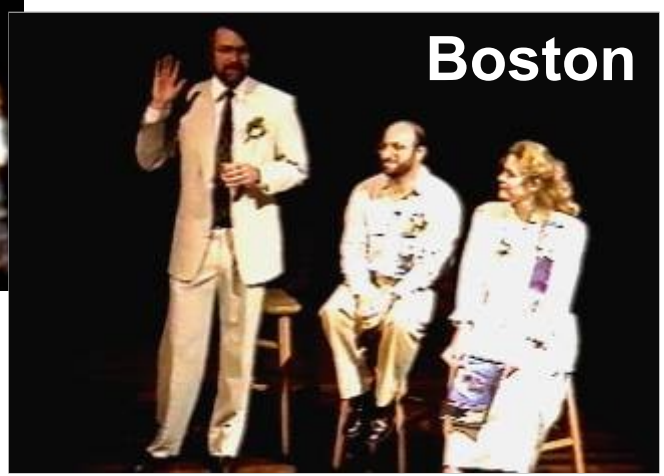
Source: Fish, Kraut, and Chalfonte-CSCW 1990 Proceedings

A Simulation of Shared Physical/Virtual Collaboration:
Using Analog Communications to Prototype the Digital Future

"What we really have to do is eliminate distance between individuals who want to interact with other people and with other computers."
— Larry Smarr, Director, NCSA

- **Televisualization:**
  - **Telepresence**
  - **Remote Interactive Visual Supercomputing**
  - **Multi-disciplinary Scientific Visualization**







"We're using satellite technology…to demo what It might be like to have high-speed fiber-optic links between advanced computers in two different geographic locations."

— **Al Gore, Senator**
Chair, US Senate Subcommittee on Science, Technology and Space

ATT & Sun

SIGGRAPH 1989

NCSA™

it²

Caterpillar / NCSA: Distributed Virtual Reality for Global-Scale Collaborative Prototyping





NCSA™

Real Time Linked Virtual Reality and Audio-Video Between NCSA, Peoria, Houston, and Germany

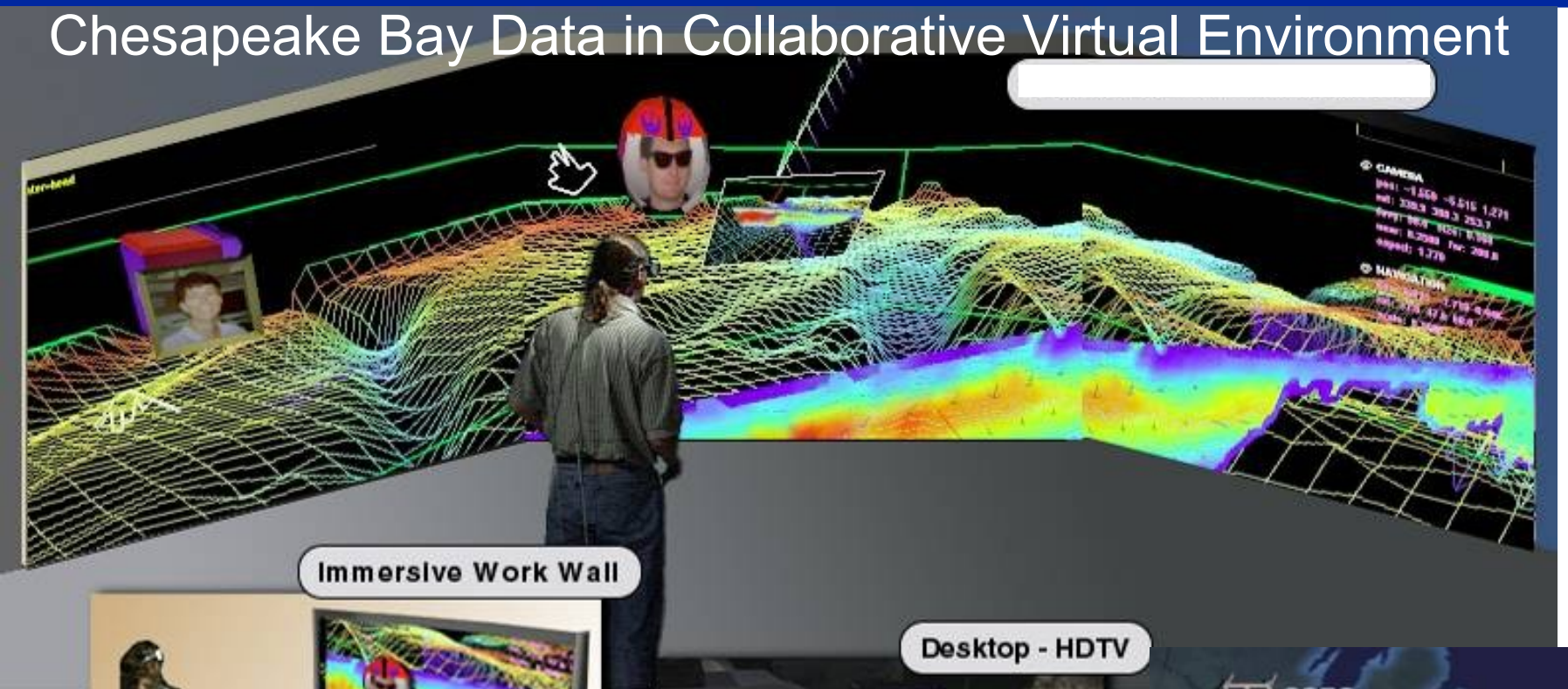
1996

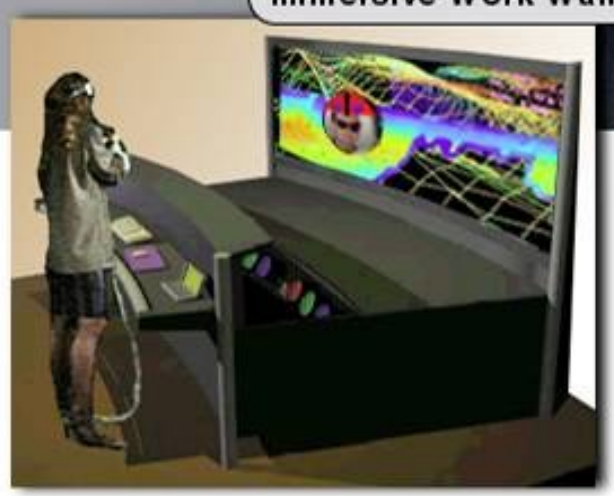
www.sv.vt.edu/future/vt-cave/apps/CatDistVR/DVR.html

Grid-Enabled Collaborative Analysis of Ecosystem Dynamics Datasets

Chesapeake Bay Data in Collaborative Virtual Environment

1997



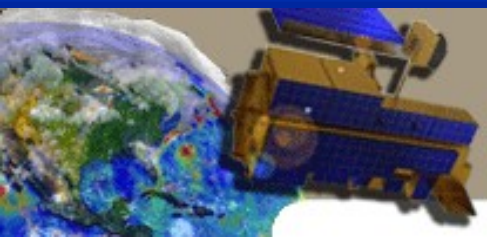






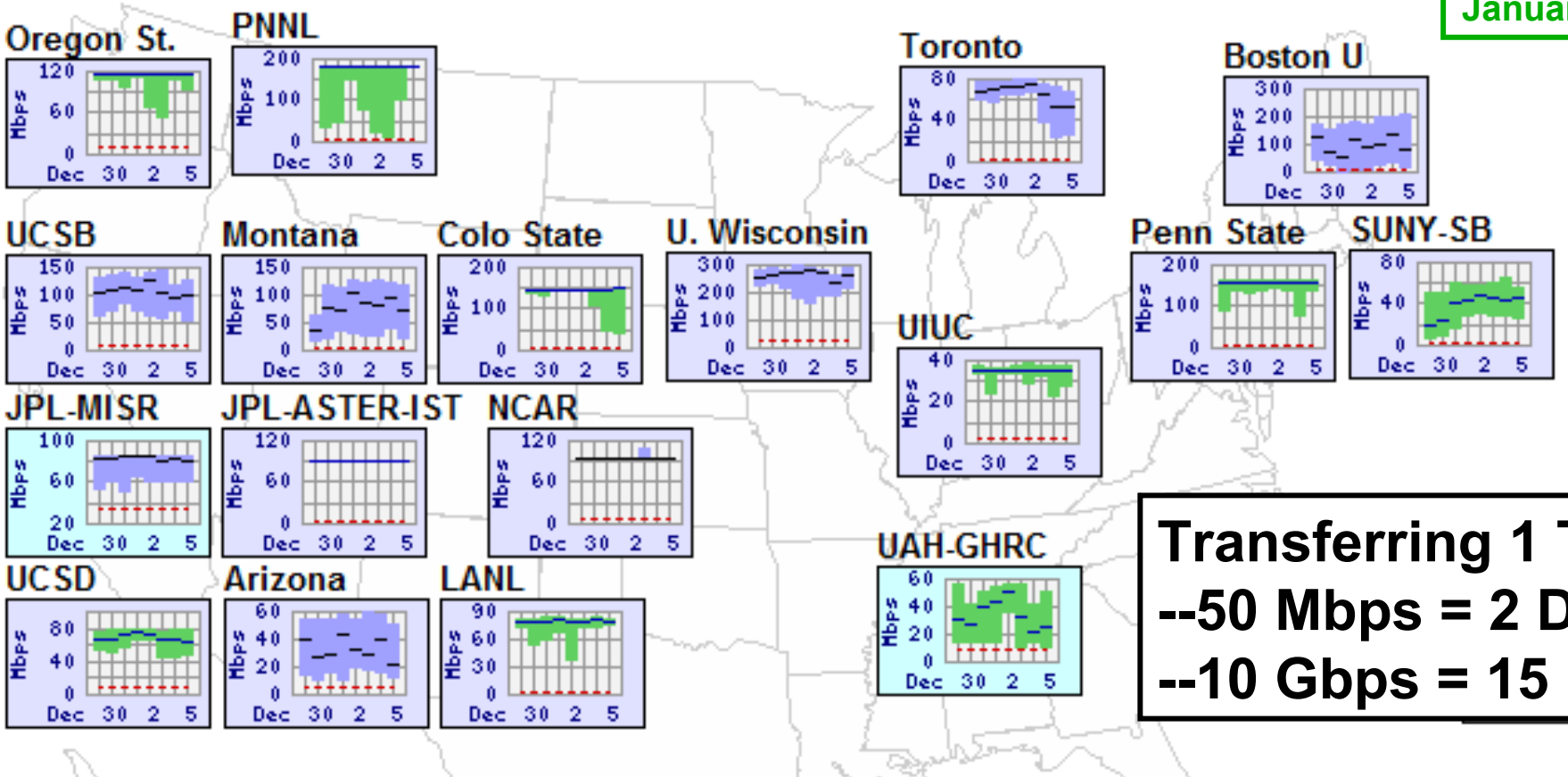
Donna Cox, Robert Patterson, Stuart Levy, NCSA Virtual Director Team
Glenn Wheless, Old Dominion Univ.
Alliance Application Technologies
Environmental Hydrology Team

Large Data Challenge: Average Throughput to End User on Shared Internet is 10-100 Mbps



ENSIGHT Active Testing TERRA Destinations

Tested January 2011



Transferring 1 TB:
--50 Mbps = 2 Days
--10 Gbps = 15 Minutes

http://ensight.eos.nasa.gov/Missions/terra/index.shtml

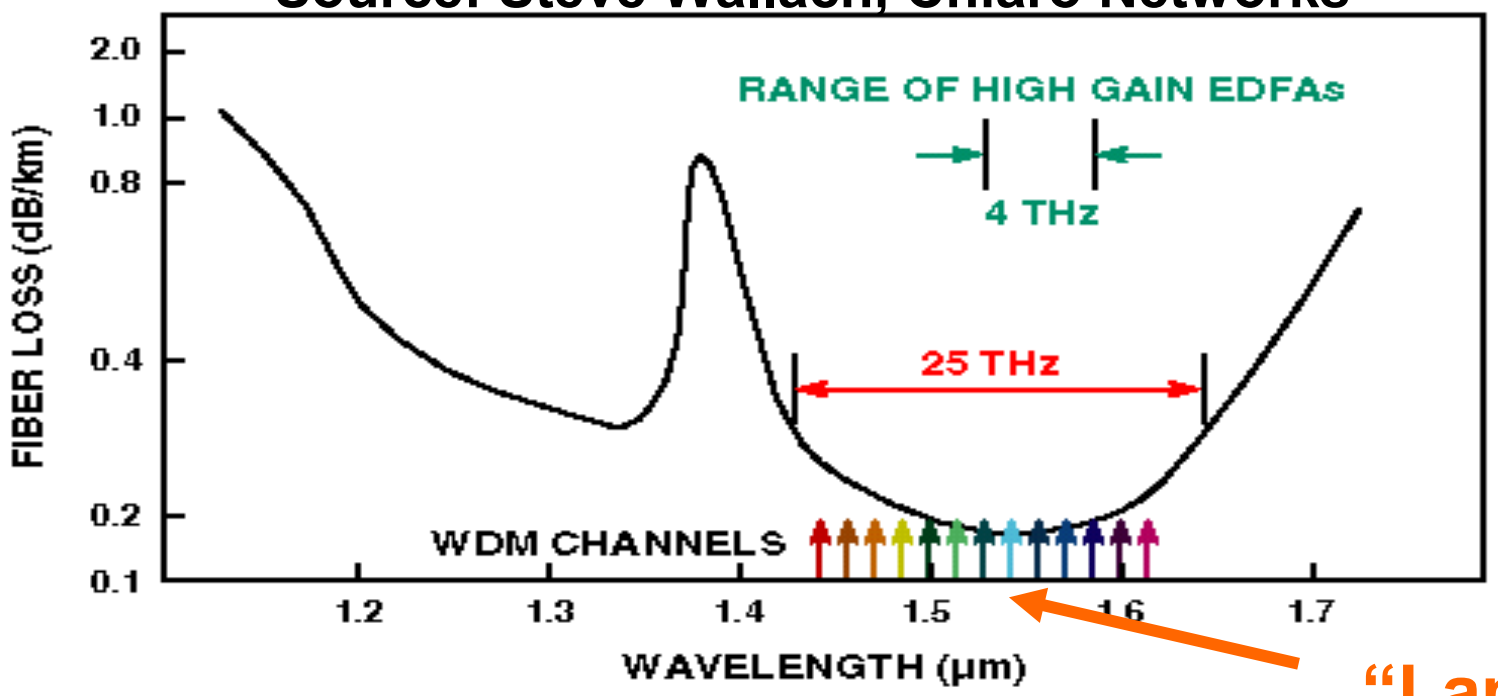
Solution: Give a Dedicated Optical Channels to Data-Intensive Users

WAVELENGTH DIVISION MULTIPLEXING (WDM)

10 Gbps per User ~ 100-1000x Shared Internet Throughput

- EXPLOITS
  - ENORMOUS BANDWIDTH OF SILICA FIBER
  - HIGH-GAIN WIDEBAND OPTICAL AMPLIFIERS

Source: Steve Wallach, Chiaro Networks

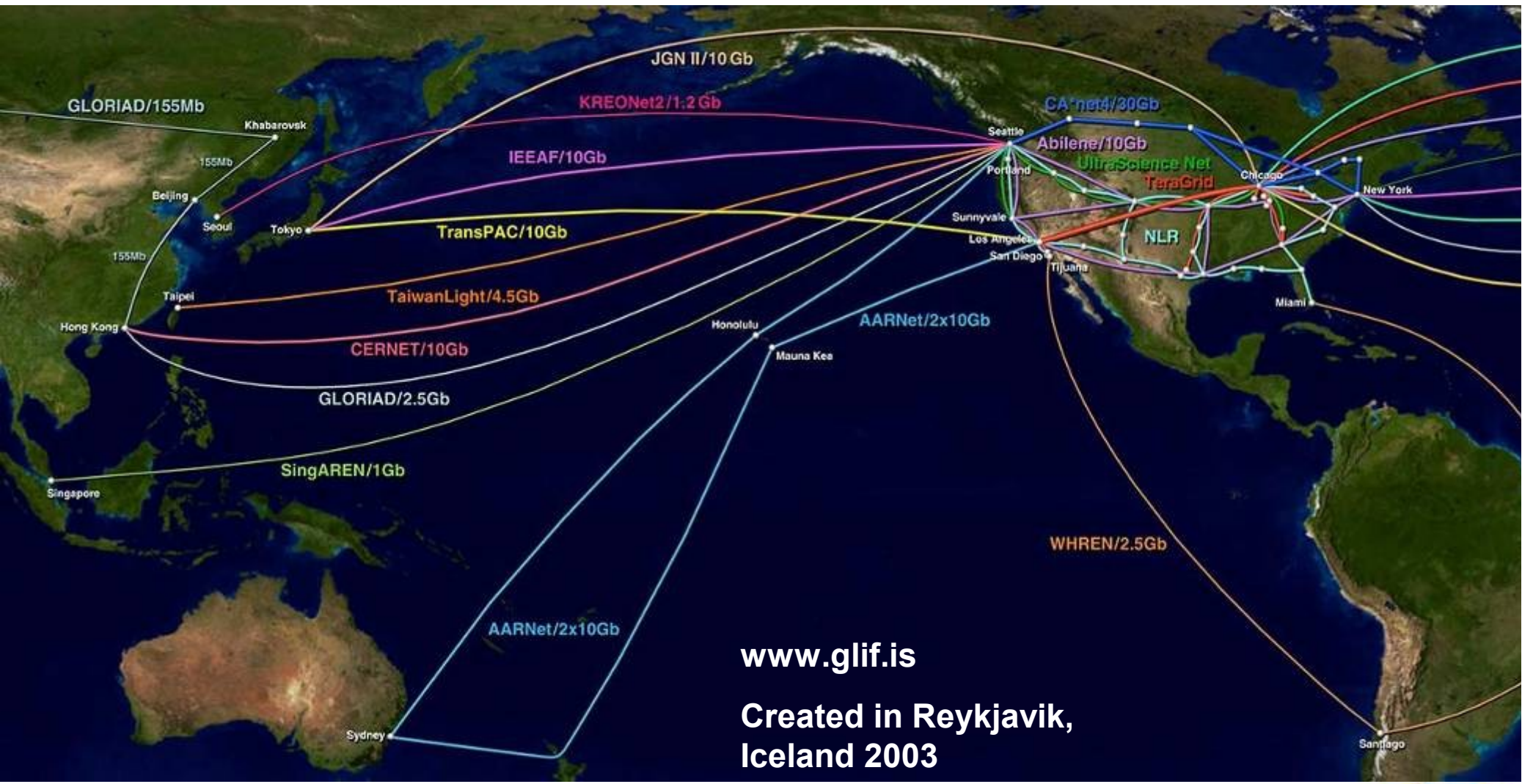


"Lambdas"

Parallel Lambdas are Driving Optical Networking The Way Parallel Processors Drove 1990s Computing

The Global Lambda Integrated Facility--
Creating a Planetary-Scale High Bandwidth Collaboratory

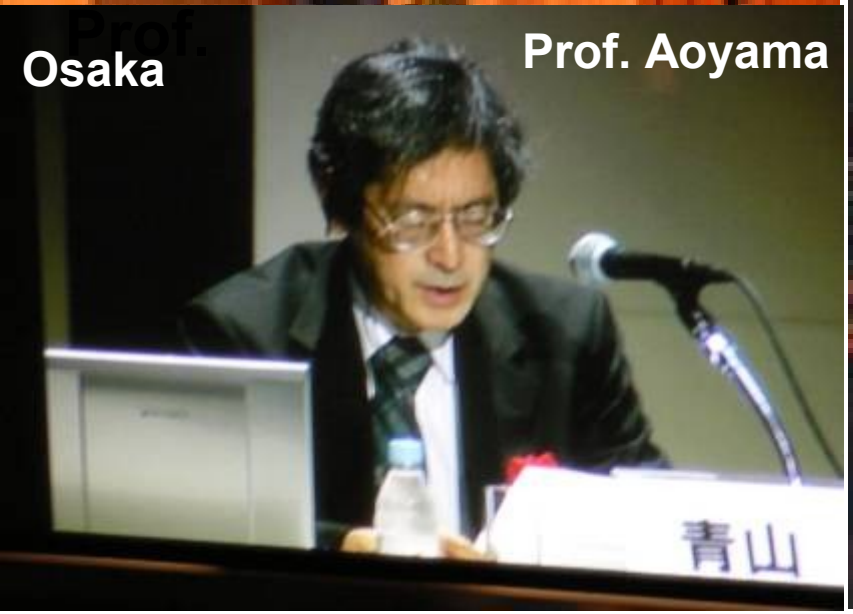
Research Innovation Labs Linked by 10G Dedicated Lambdas



Visualization courtesy of
Bob Patterson, NCSA.

High Resolution Uncompressed HD Streams Require Multi-Gigabit/s Lambdas





Telepresence Using Uncompressed 1.5 Gbps HDTV Streaming Over IP on Fiber Optics--
75x Home Cable "HDTV" Bandwidth!



"I can see every hair on your head!"—Prof. Aoyama

Source: U Washington Research Channel

Borderless Collaboration
Between Global University Research Centers at 10Gbps

Maxine Brown, Tom DeFanti, Co-Chairs

iGrid 2005

THE GLOBAL LAMBDA INTEGRATED FACILITY

www.igrid2005.org

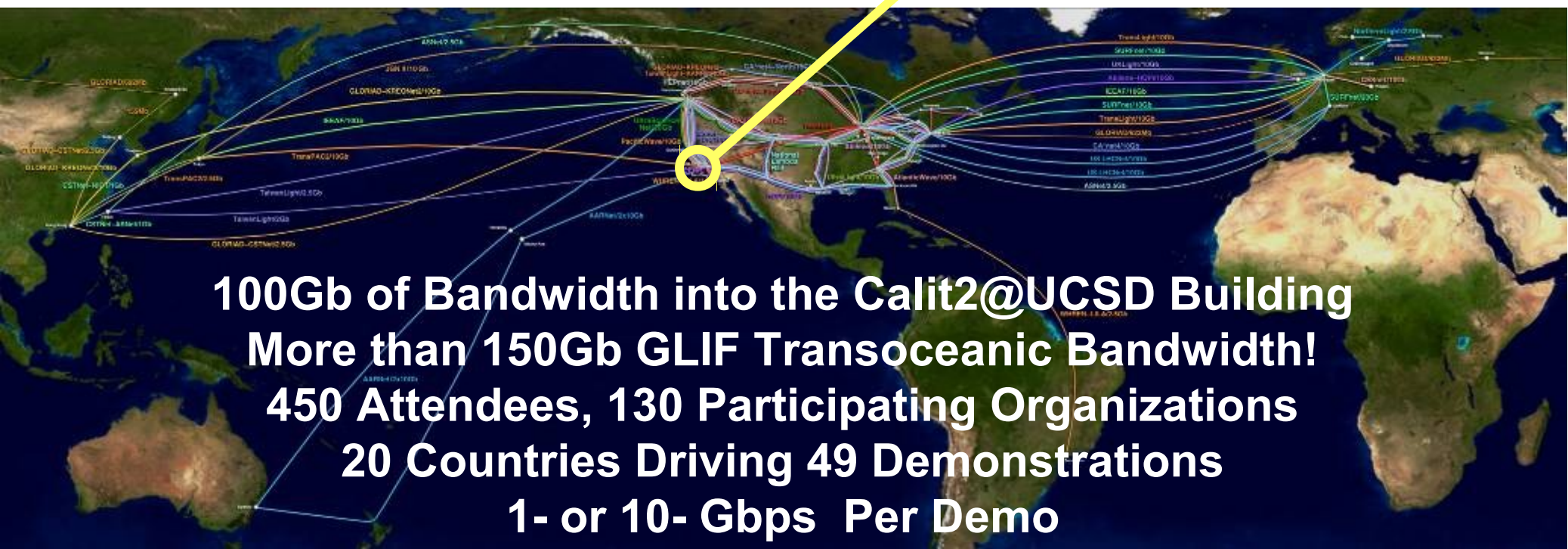


the UC Discovery Grant

September 26-30, 2005

Calit2 @ University of California, San Diego

California Institute for Telecommunications and Information Technology



100Gb of Bandwidth into the Calit2@UCSD Building

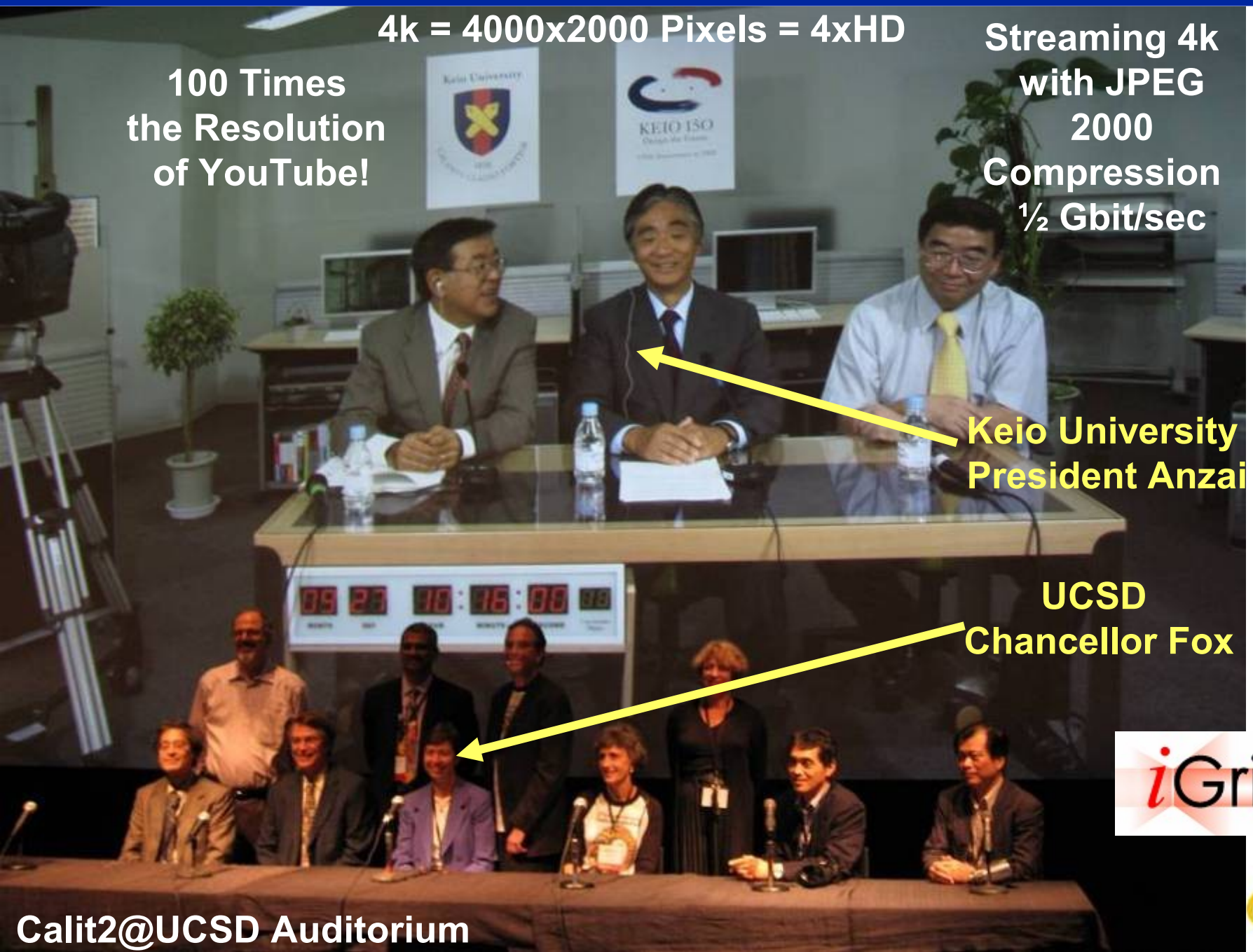
More than 150Gb GLIF Transoceanic Bandwidth!

450 Attendees, 130 Participating Organizations

20 Countries Driving 49 Demonstrations

1- or 10- Gbps Per Demo

Telepresence Meeting
Using Digital Cinema 4k Streams



CineGrid

Lays Technical Basis for Global Digital Cinema

Sony
NTT
SGI

iGrid 2005

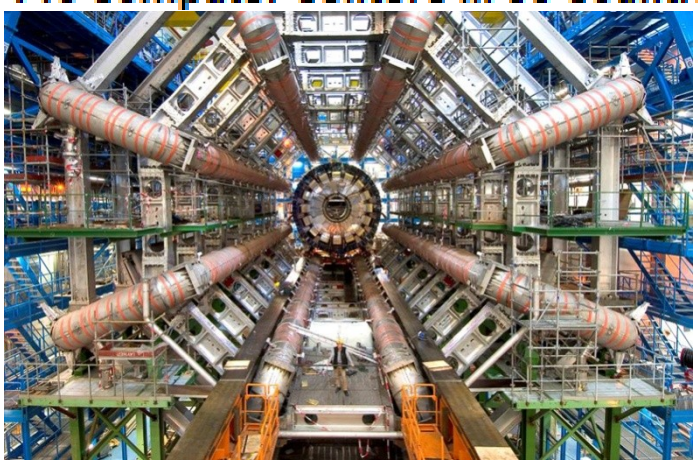
it²

The Large Hadron Collider Uses a Global Fiber Infrastructure To Connect Its Users

CERN Unveils Global Grid For Particle Physics Research

The network can pull in the IT power of more than 140 computer centers in 33 countries to crunch an expected 15 million GB of data every year.

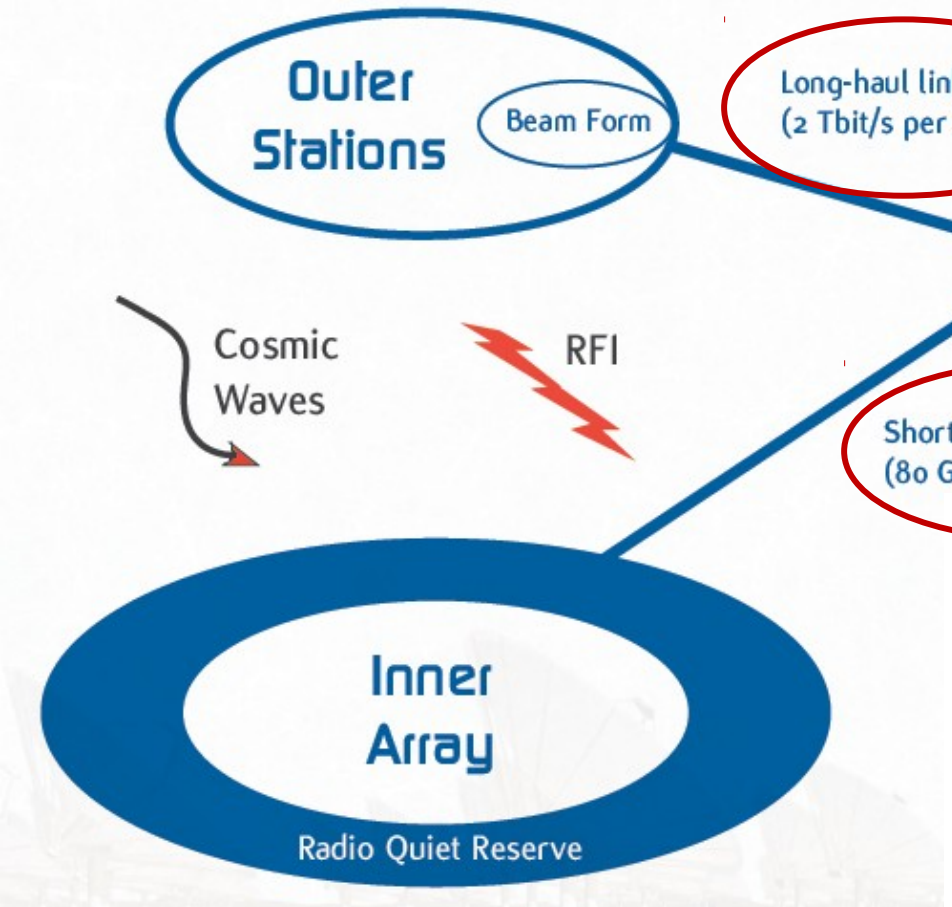
By Antone Gonsalves
InformationWeek
October 3, 2008 04:57 PM



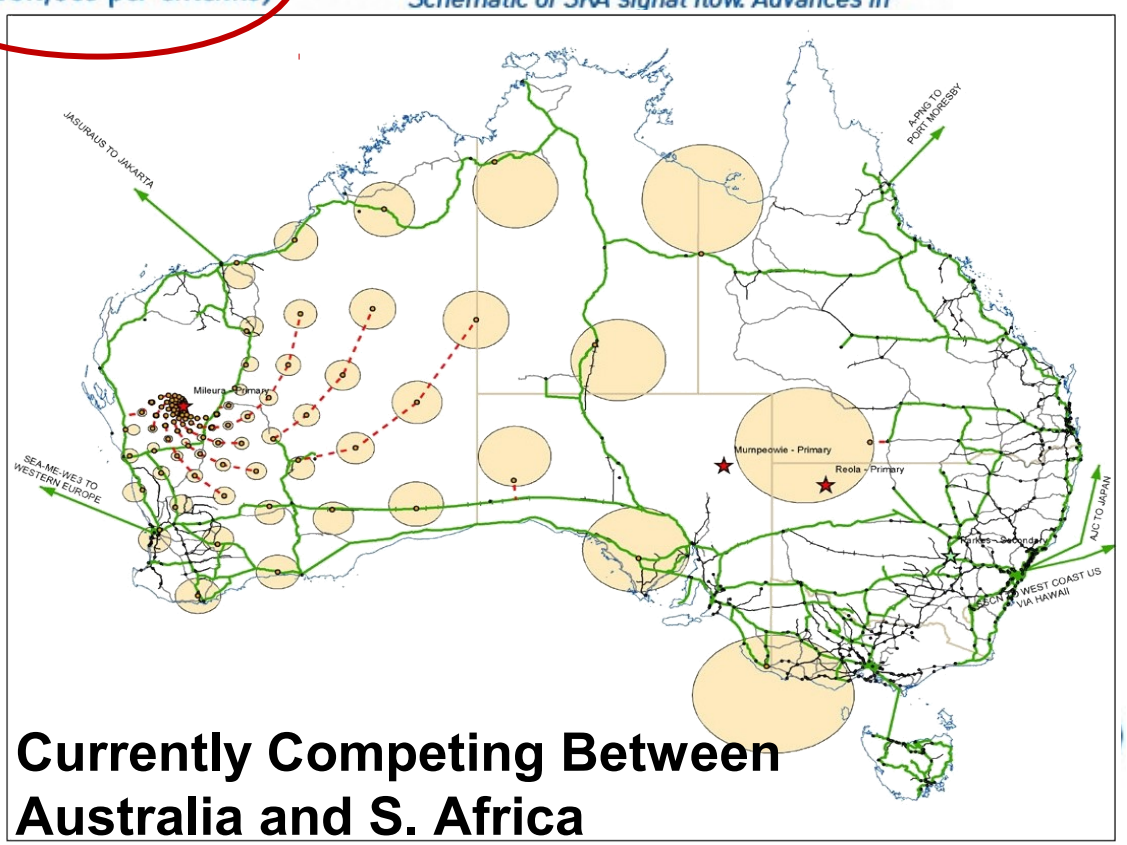
- The grid relies on optical fiber networks to distribute data from CERN to 11 major computer centers in Europe, North America, and Asia
- The grid is capable of routinely processing 250,000 jobs a day
- The data flow will be ~6 Gigabits/sec or 15 million gigabytes a year for 10 to 15 years

Next Great Planetary Instrument: The Square Kilometer Array Requires Dedicated Fiber

www.skatelescope.org



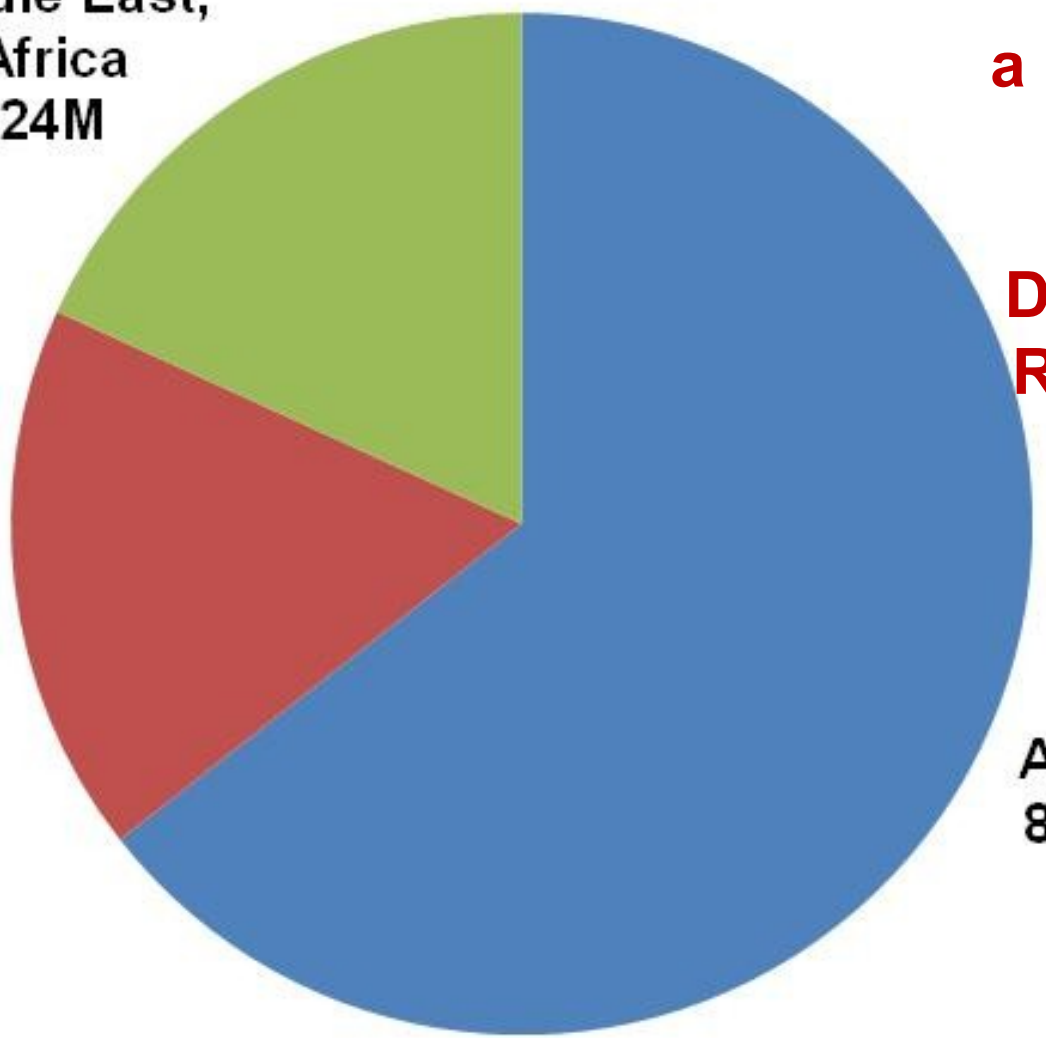
Schematic of SKA signal flow. Advances in



Transfers Of 1 TByte Images World-wide Will Be Needed Every Minute!

Currently Competing Between Australia and S. Africa

Globally Fiber to the Premise is Growing Rapidly, Mostly in Asia



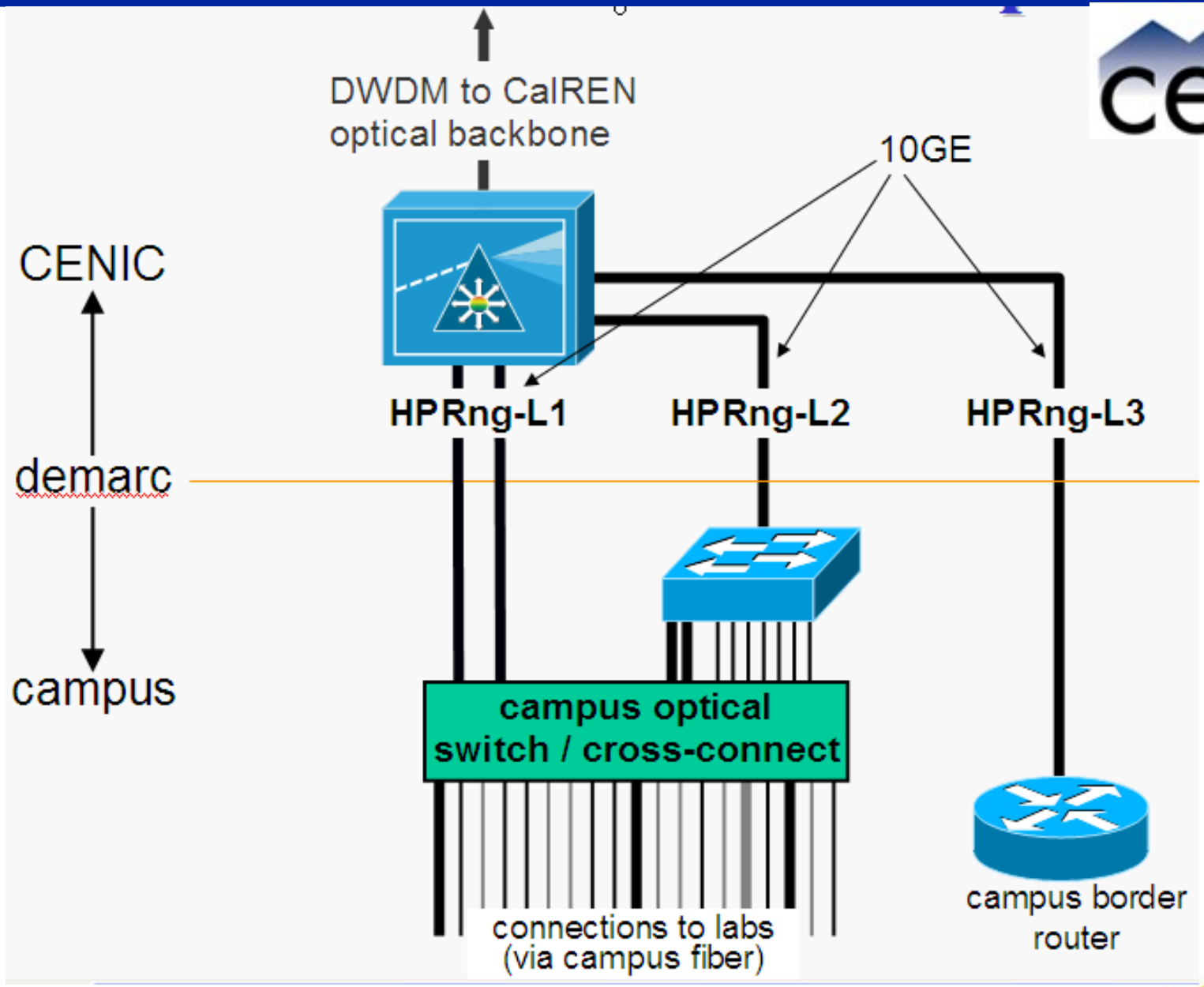
FTTP Connections Growing at ~30%/year

130 Million Households with FTTH in 2013

If Couch Potatoes Deserve a Gigabit Fiber, Why Not University Data-Intensive Researchers?

Source: Heavy Reading (www.heavyreading.com), the market research division of Light Reading (www.lightreading.com).

Campus Preparations Needed to Accept CENIC CalREN Handoff to Campus



Source: Jim Dolgonas, CENIC

Current UCSD Prototype Optical Core: Bridging End-Users to CENIC L1, L2, L3 Services

Enpoints:

>= 60 endpoints at 10 GigE

>= 32 Packet switched

>= 32 Switched wavelengths

>= 300 Connected endpoints

Approximately 0.5 TBit/s Arrive at the "Optical" Center of Campus. Switching is a Hybrid of: Packet, Lambda, Circuit -- OOO and Packet Switches

Source: Phil Papadopoulos, SDSC/Calit2 (Quartzite PI, OptIPuter co-PI)

Quartzite Network MRI #CNS-0421555; OptIPuter #ANI-0225642

SDSC

NSF

it²

Calit2 Sunlight Optical Exchange Contains Quartzite

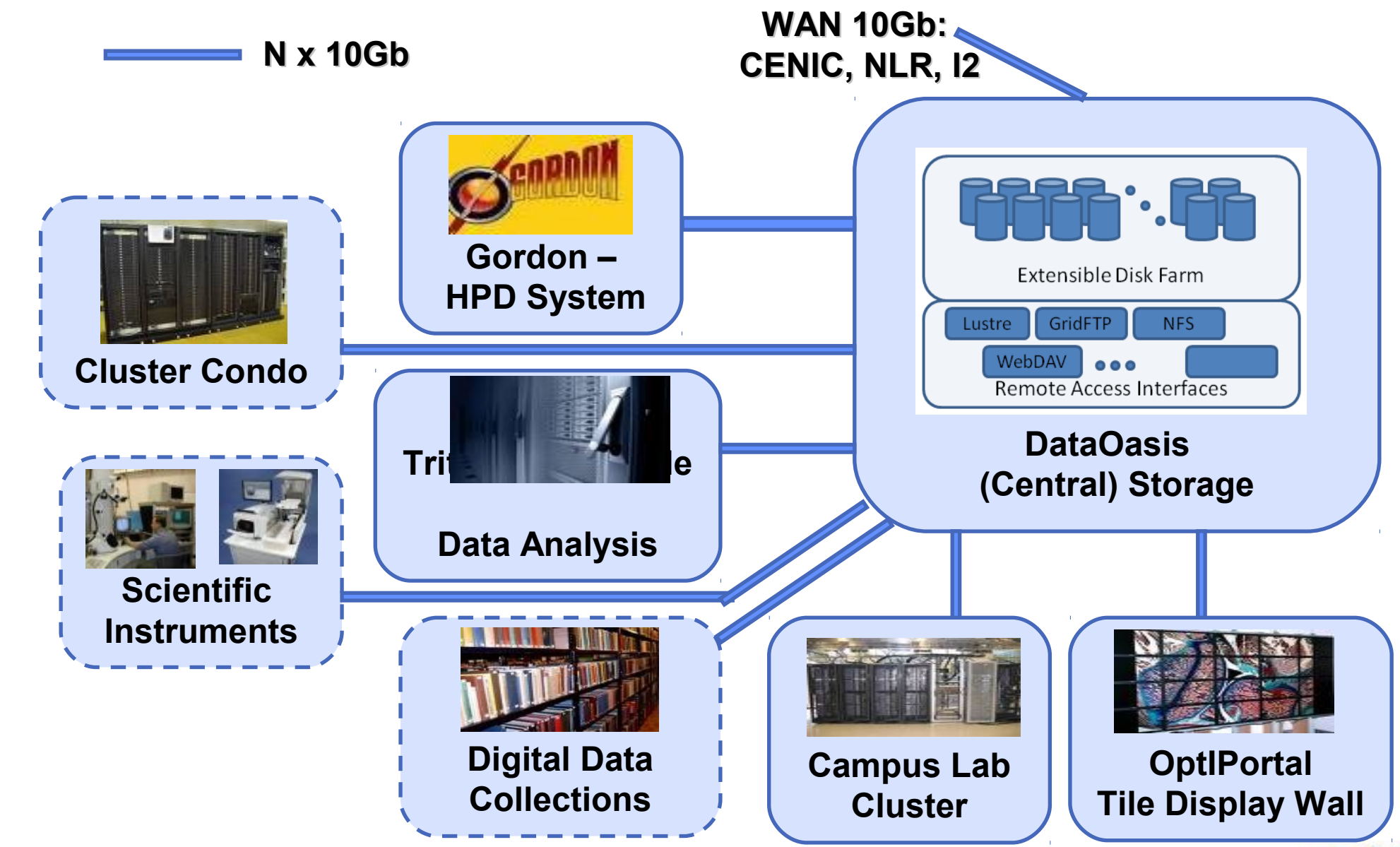


Maxine Brown, EVL, UIC OptIPuter Project Manager

SDSC

it²

UCSD Campus Investment in Fiber Enables Consolidation of Energy Efficient Computing & Storage



Source: Philip Papadopoulos, SDSC, UCSD

Data-Intensive Visualization and Analysis

The OptIPuter Project: Creating High Resolution Portals Over Dedicated Optical Channels to Global Science Data



NCMIR
NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH
at San Diego, an NIH supported resource center

Scalable Adaptive Graphics Environment (SAGE)

SAGE

evl electronic visualization laboratory

NSF

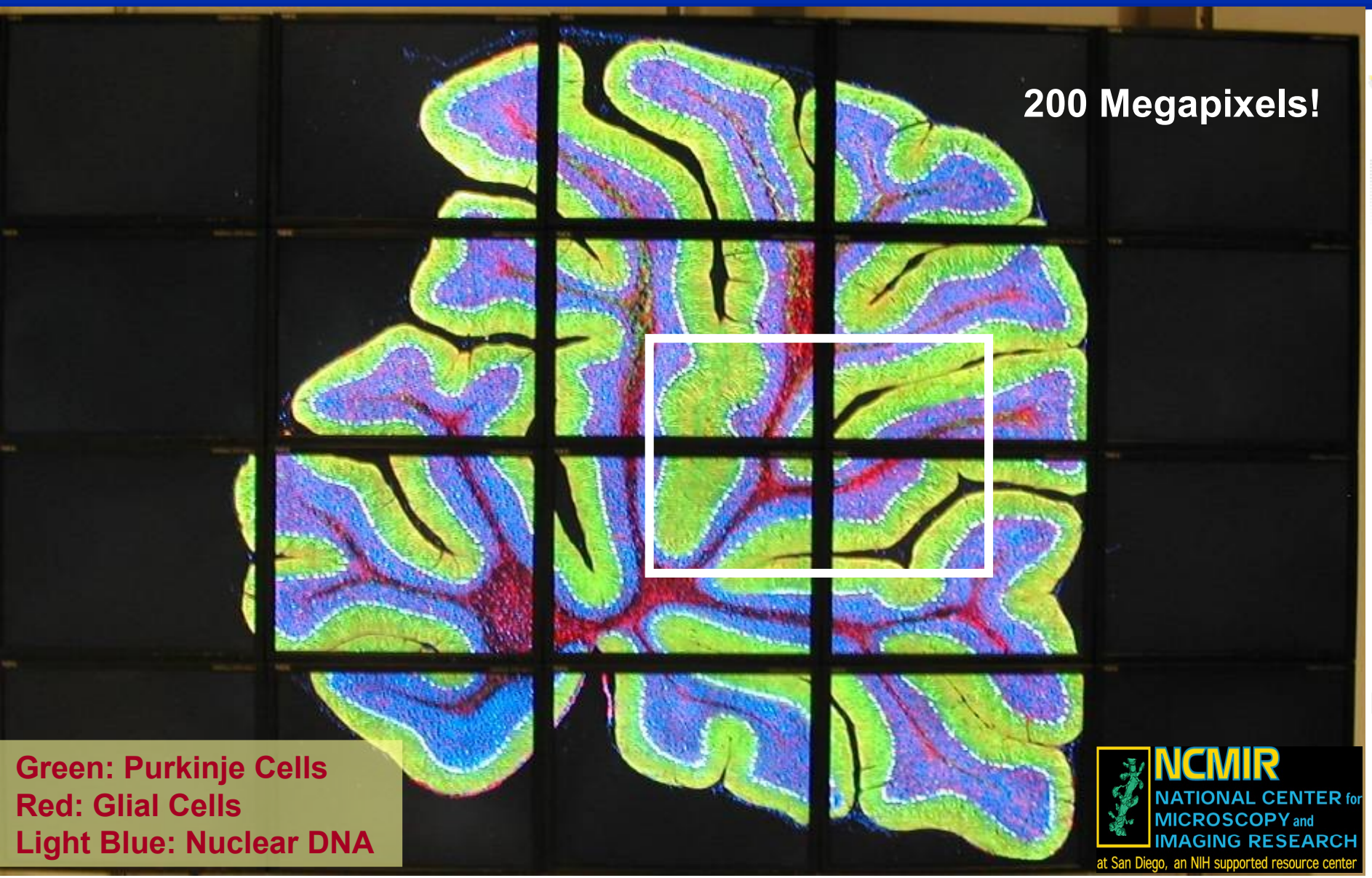
ROCKS

Picture Source: Mark Ellisman, David Lee, Jason Leigh

Calit2 (UCSD, UCI), SDSC, and UIC Leads—Larry Smarr PI
Univ. Partners: NCSA, USC, SDSU, NW, TA&M, UvA, SARA, KISTI, AIST
Industry: IBM, Sun, Telcordia, Chiaro, Calient, Glimmerglass, Lucent

it²

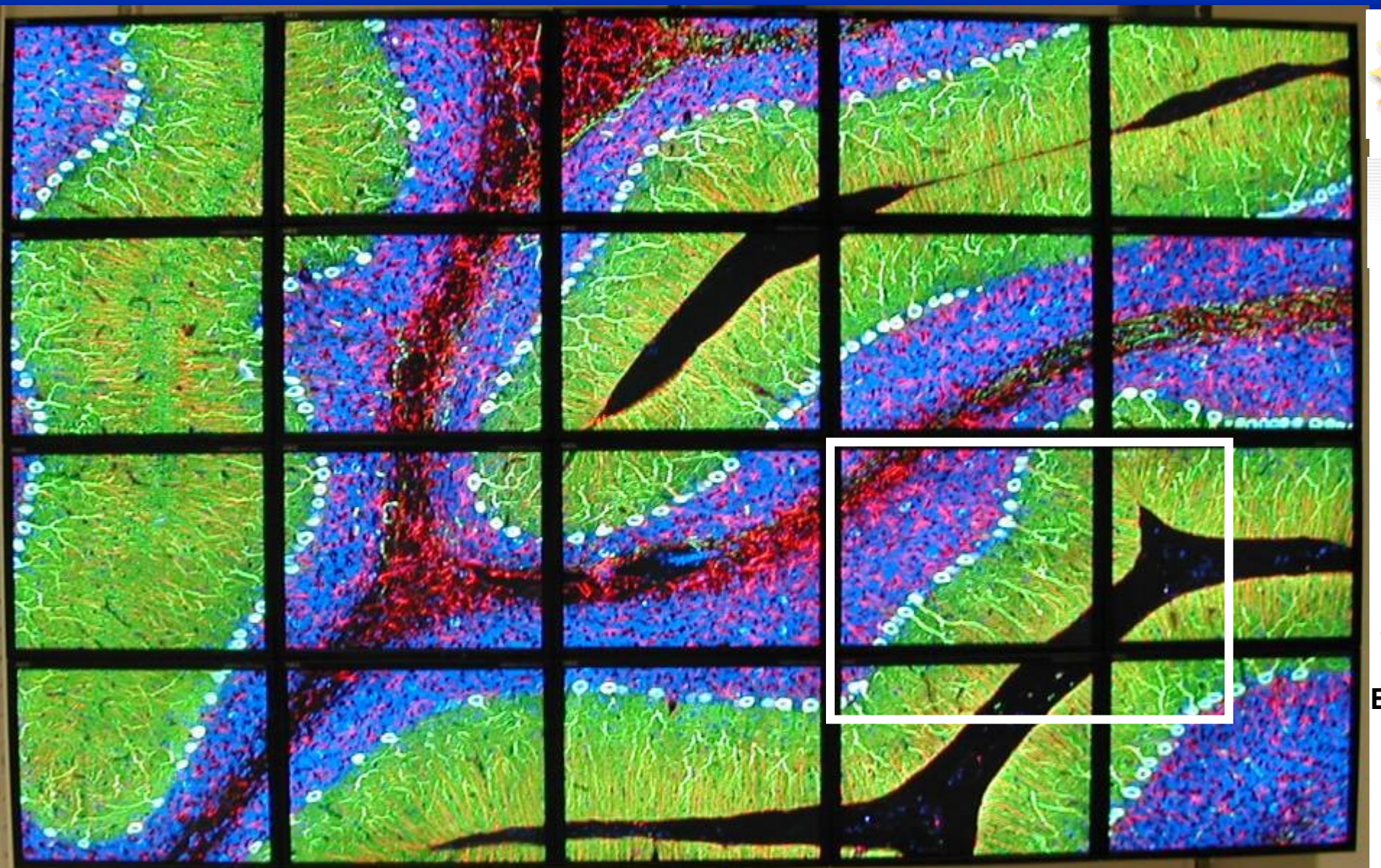
Use of OptIPortal to Interactively View Multi-Scale Biomedical Imaging



Source: Mark Ellisman, David Lee, Jason Leigh

Two-Photon Laser Confocal Microscope Montage of 40x36=1440 Images in 3 Channels of a Mid-Sagittal Section of Rat Cerebellum Acquired Over an 8-hour Period

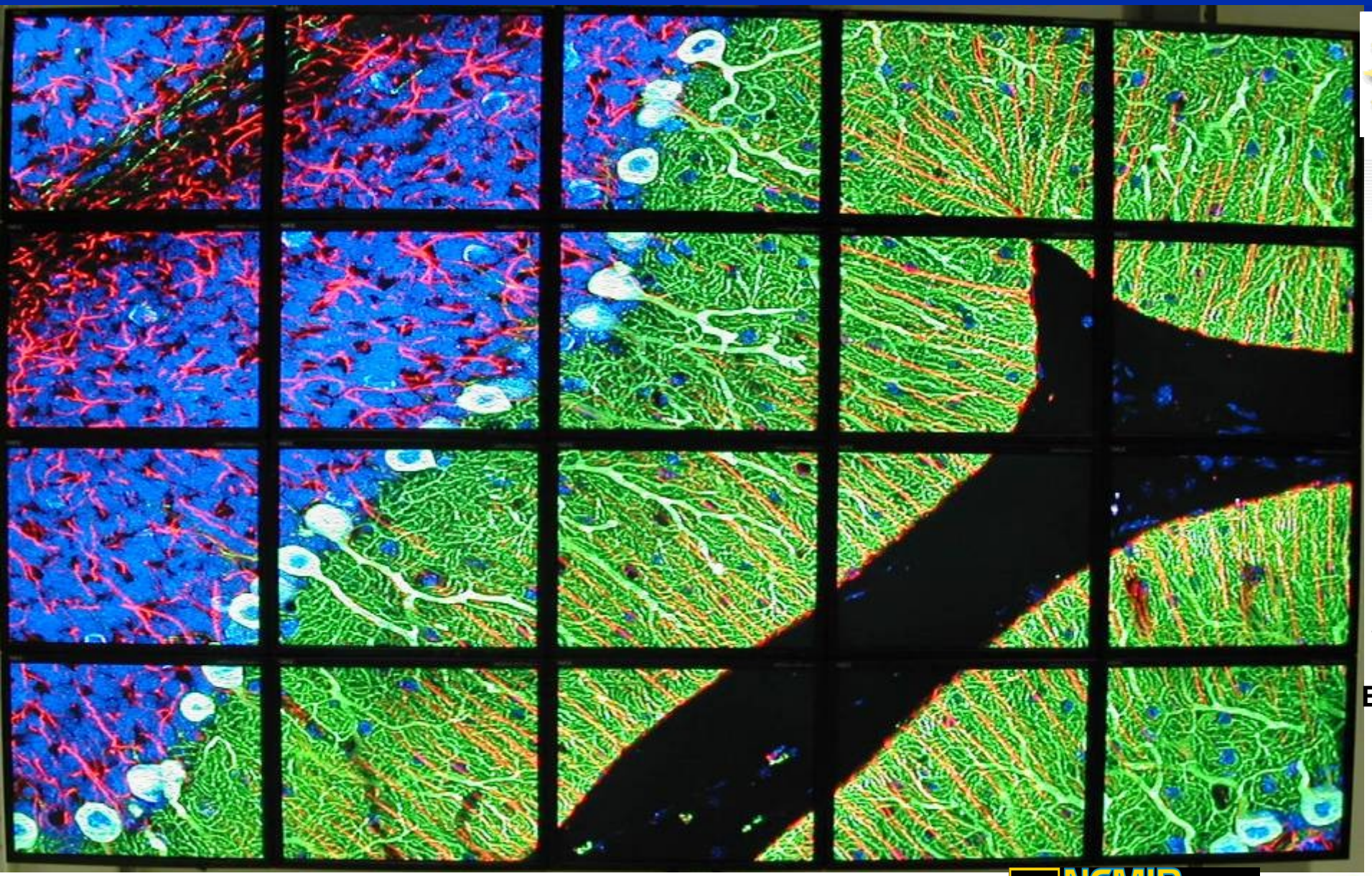
Scalable Displays Allow Both Global Content and Fine Detail



Source: Mark Ellisman, David Lee, Jason Leigh

NCMIR
NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH
at San Diego, an NIH supported resource center

Allows for Interactive Zooming from Cerebellum to Individual Neurons



Source: Mark Ellisman, David Lee, Jason Leigh

NCMIR
NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH
at San Diego, an NIH supported resource center

OptIPortals Scale to 1/3 Billion Pixels Enabling Viewing of Very Large Images or Many Simultaneous Images



Spitzer Space Telescope (Infrared)

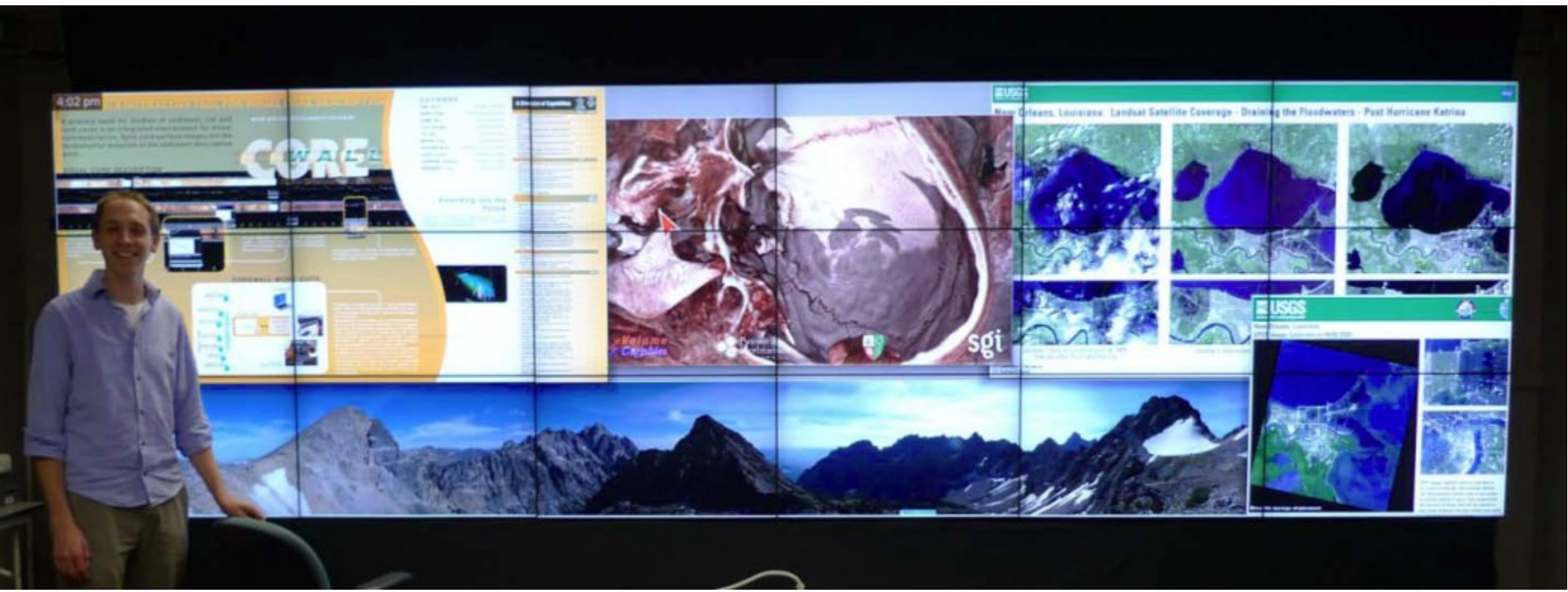


NASA Earth Satellite Images Bushfires October 2007 San Diego

Source: Falko Kuester, Calit2@UCSD

the AESOP Nearly Seamless OptIPortal

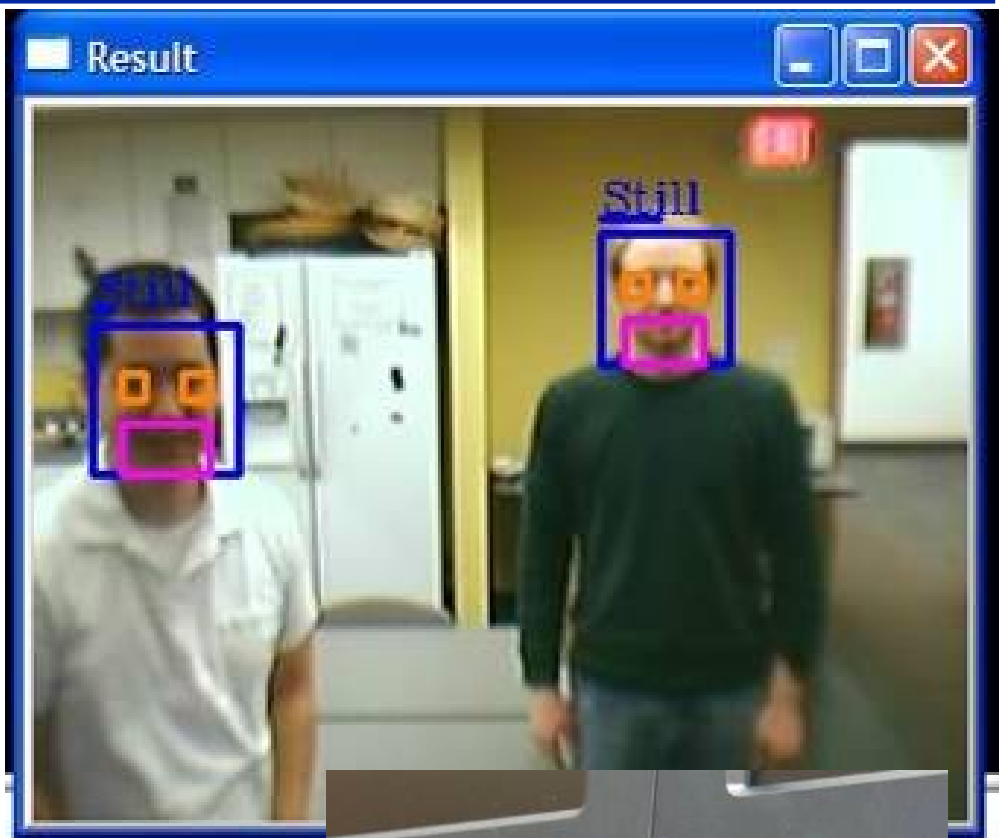
46” NEC Ultra-Narrow Bezel 720p LCD Monitors



Source: Tom DeFanti, Calit2@UCSD;

U Michigan Virtual Space Interaction Testbed (VISIT) Instrumenting OptIPortals for Social Science Research

- **Using Cameras Embedded in the Seams of Tiled Displays and Computer Vision Techniques, we can Understand how People Interact with OptIPortals**
  - **Classify Attention, Expression, Gaze**
  - **Initial Implementation Based on Attention Interaction Design Toolkit (J. Lee, MIT)**
- **Close to Producing Usable Eye/Nose Tracking Data using OpenCV**





Leading U.S. Researchers on the Social Aspects of Collaboration

Source: Erik Hofer, UMich, School of Information

High Definition Video Connected OptIPortals: Virtual Working Spaces for Data Intensive Research



NASA Supports Two Virtual Institutes

CGLX
CROSS-PLATFORM CLUSTER GRAPHICS LIBRARY

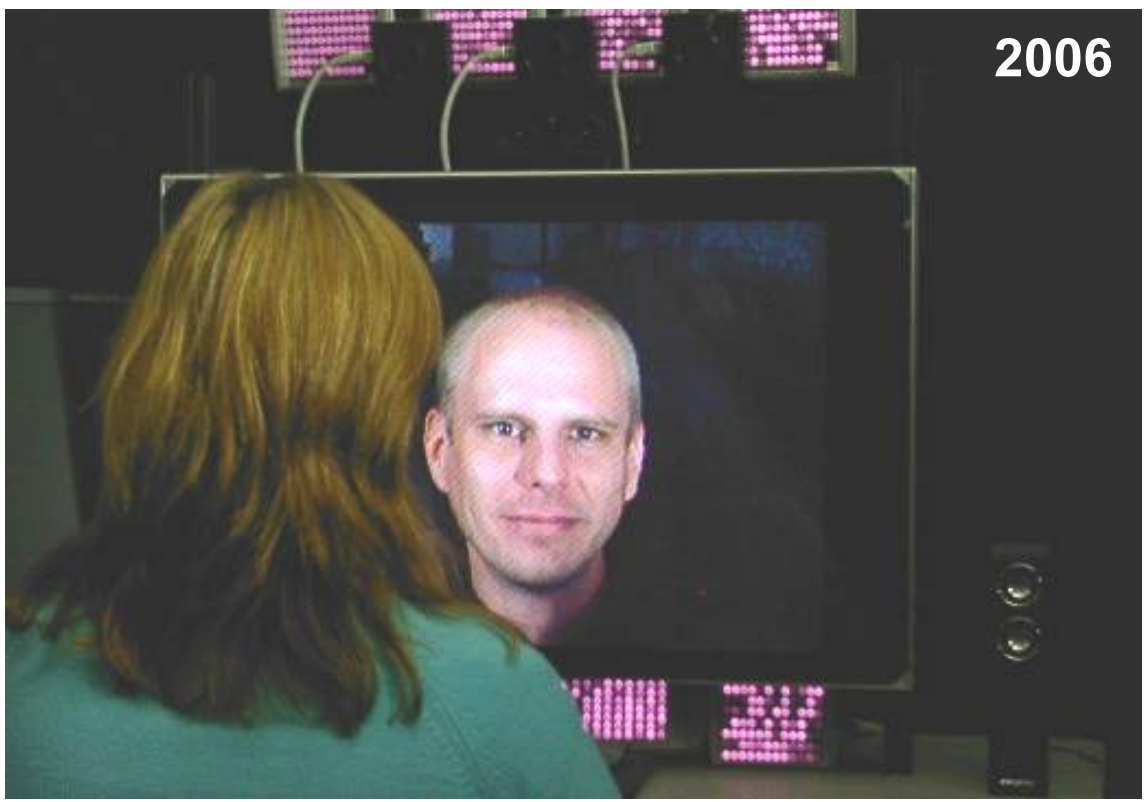
LifeSize HD

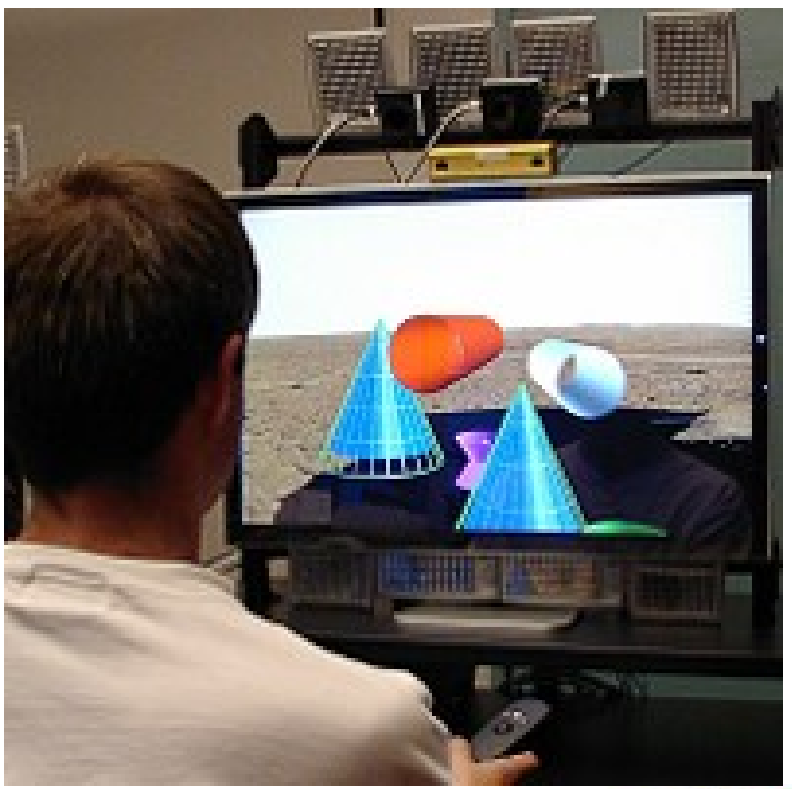
Calit2@UCSD 10Gbps Link to NASA Ames Lunar Science Institute, Mountain View, CA

Source: Falko Kuester, Kai Doerr Calit2; Michael Sims, Larry Edwards, Estelle Dodson NASA

3D Videophones Are Here!
The Personal Varrier Autostereo Display

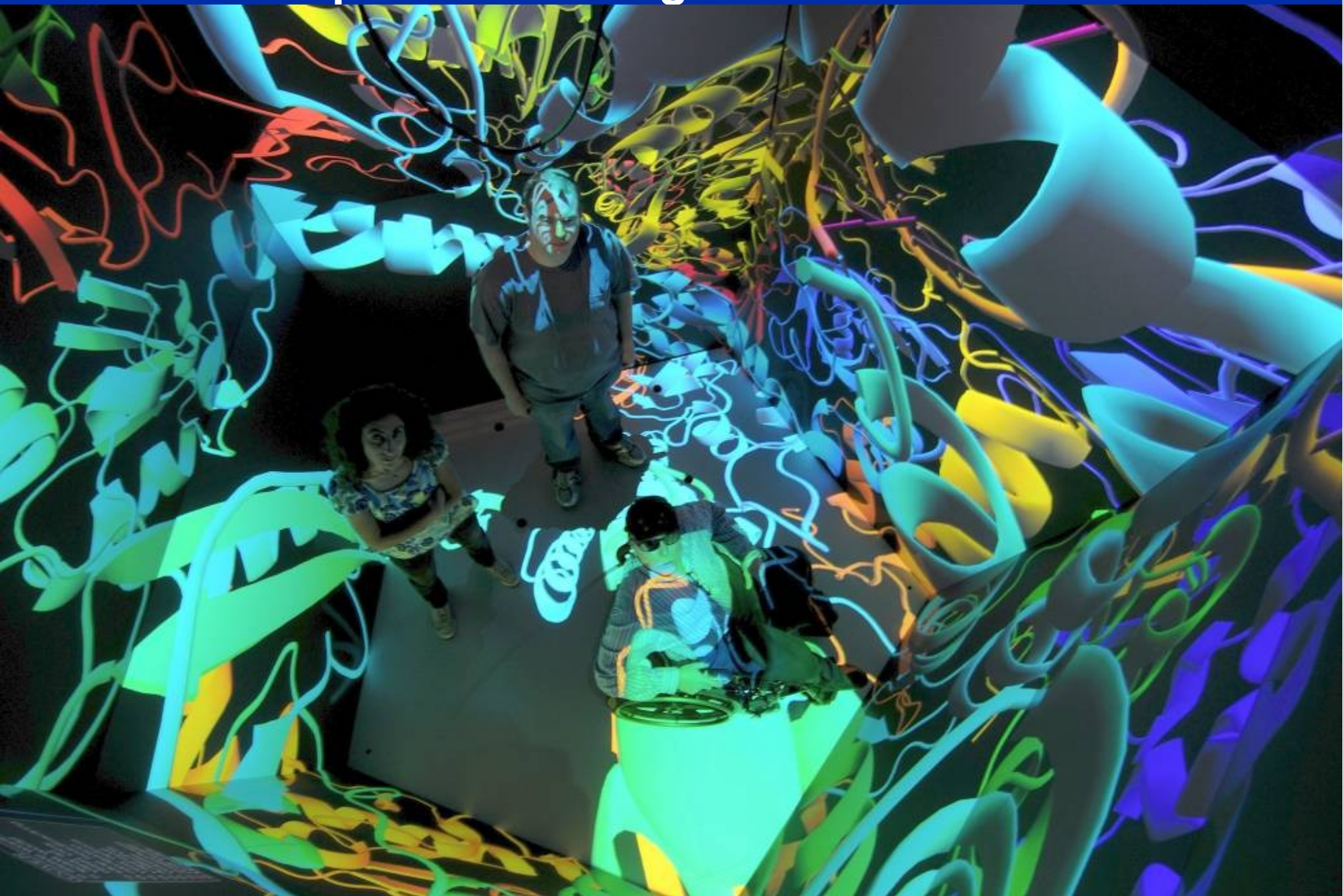
- Varrier is a Head-Tracked Autostereo Virtual Reality Display
  - 30” LCD Widescreen Display with 2560x1600 Native Resolution
  - A Photographic Film Barrier Screen Affixed to a Glass Panel
- Cameras Track Face with Neural Net to Locate Eyes
- The Display Eliminates the Need to Wear Special Glasses





Source: Daniel Sandin, Thomas DeFanti, Jinghua Ge, Javier Girado, Robert Kooima, Tom Peterka—EVL, UIC

Calit2 3D Immersive StarCAVE OptIPortal: Enables Exploration of High Resolution Simulations



3D Stereo Head Tracked OptIPortal: NexCAVE

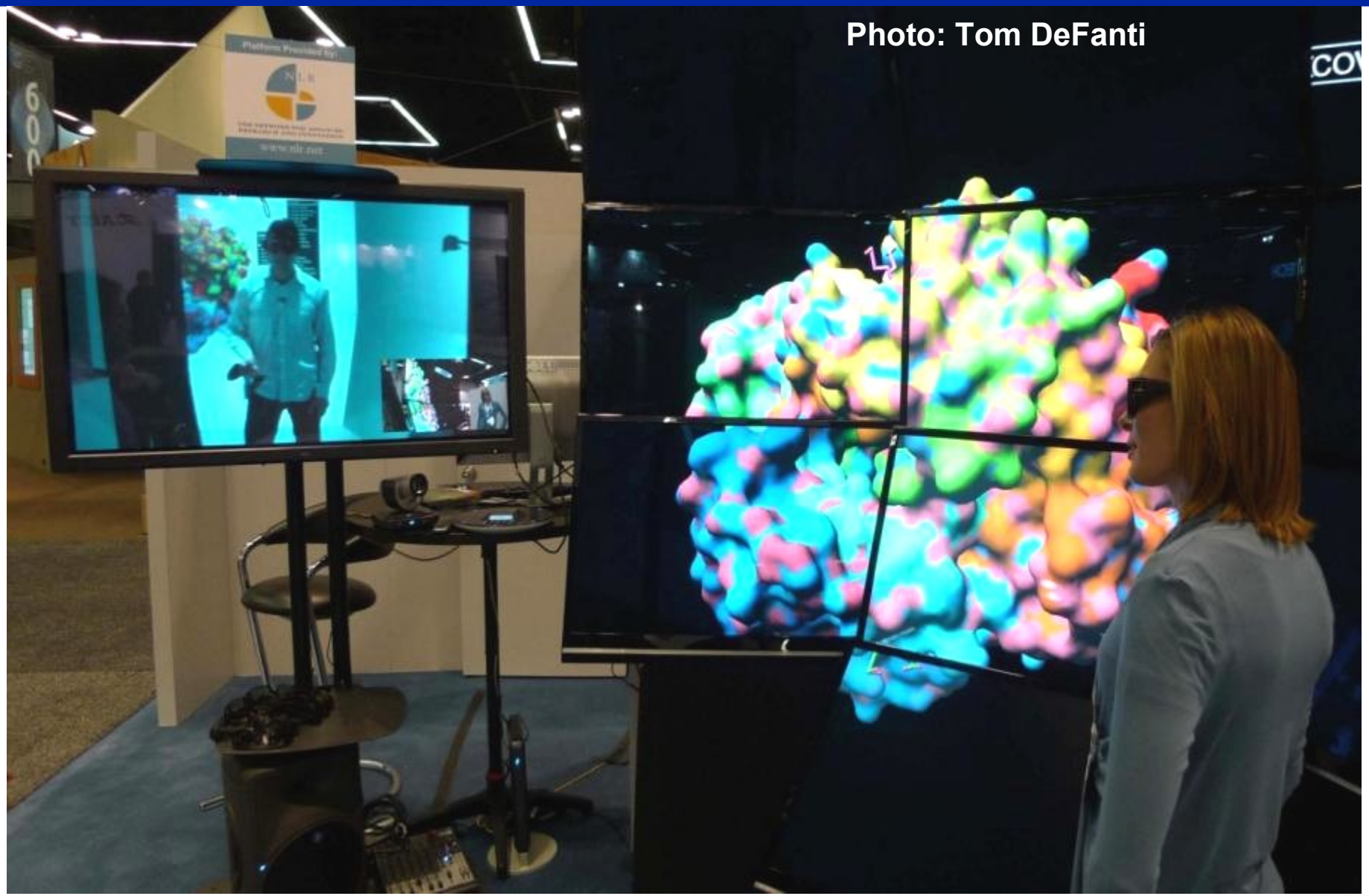


Array of JVC HDTV 3D LCD Screens
KAUST NexCAVE = 22.5MPixels

www.calit2.net/newsroom/article.php?id=1584

Source: Tom DeFanti, Calit2@UCSD

3D CAVE to CAVE Collaboration with HD Video



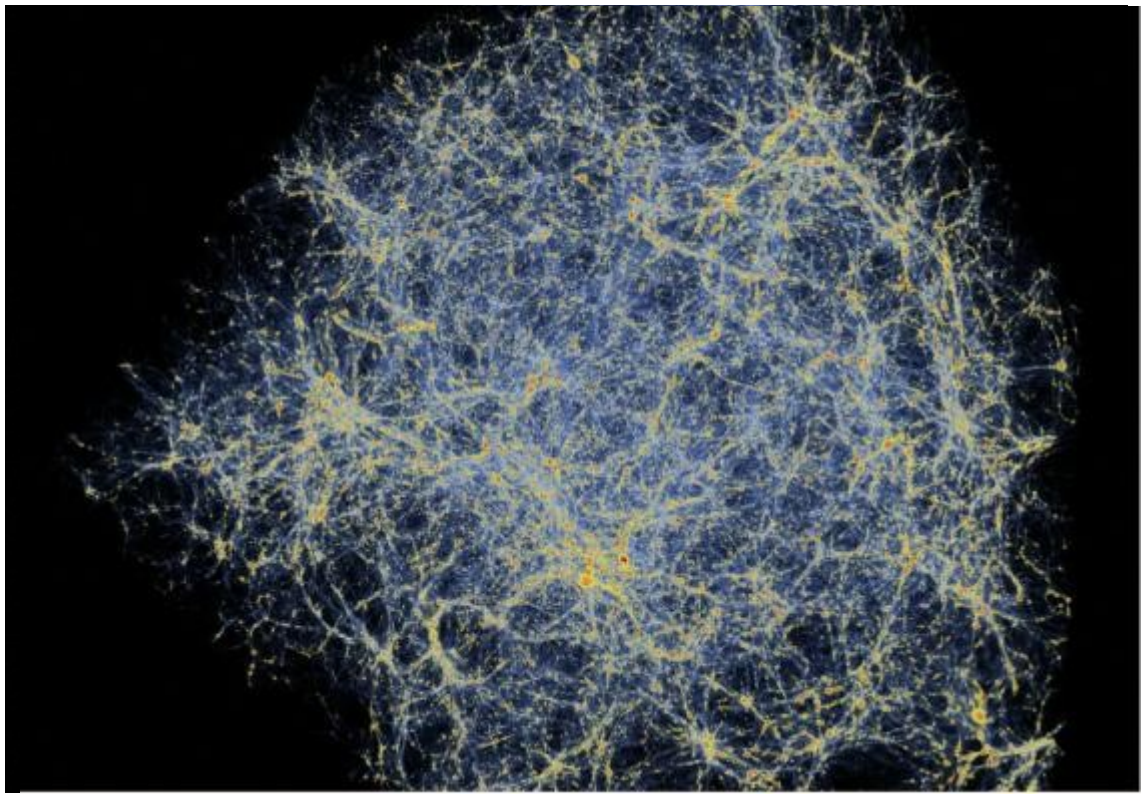
Calit2's Jurgen Schulze in San Diego in StarCAVE and Kara Gribskov at SC'09 in Portland, OR with NextCAVE

Remote Data-Intensive Discovery

Exploring Cosmology With Supercomputers, Supernetworks, and Supervisualization

Source: Mike Norman, SDSC

Intergalactic Medium on 2 GLyr Scale



Science: Norman, Harkness,Paschos SDSC
Visualization: Insley, ANL; Wagner SDSC

- 4096^3 Particle/Cell Hydrodynamic Cosmology Simulation
- NICS Kraken (XT5)
  - 16,384 cores
- Output
  - 148 TB Movie Output (0.25 TB/file)
  - 80 TB Diagnostic Dumps (8 TB/file)

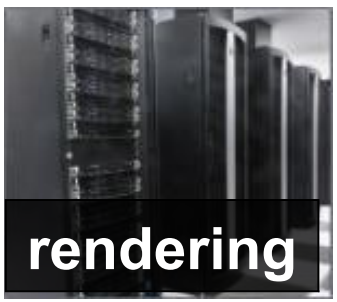
End-to-End 10Gbps Lambda Workflow: OptIPortal to Remote Supercomputers & Visualization Servers

Source: Mike Norman, Rick Wagner, SDSC

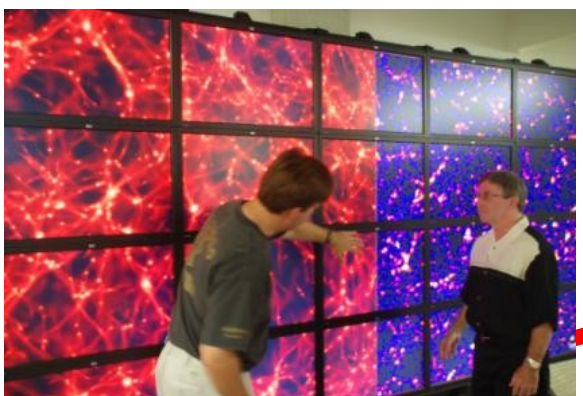
Project Stargate

Argonne NL

DOE Eureka
100 Dual Quad Core Xeon Servers
200 NVIDIA Quadro FX GPUs in 50 Quadro Plex S4 1U enclosures
3.2 TB RAM



SDSC

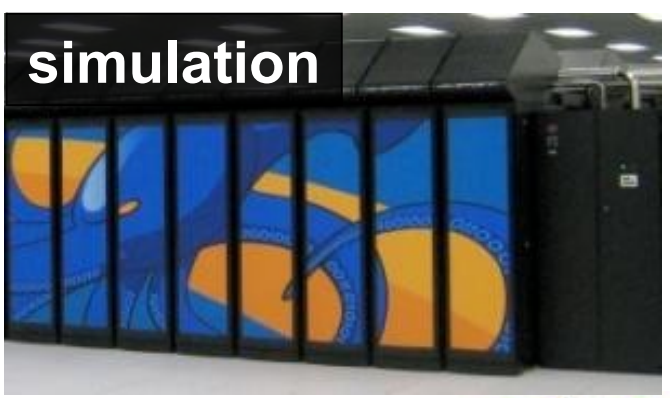


Calit2/SDSC OptIPortal1
20 30" (2560 x 1600 pixel) LCD panels
10 NVIDIA Quadro FX 4600 graphics cards > 80 megapixels
10 Gb/s network throughout

ESnet
10 Gb/s fiber optic network

NICS ORNL

NSF TeraGrid Kraken
Cray XT5
8,256 Compute Nodes
99,072 Compute Cores
129 TB RAM



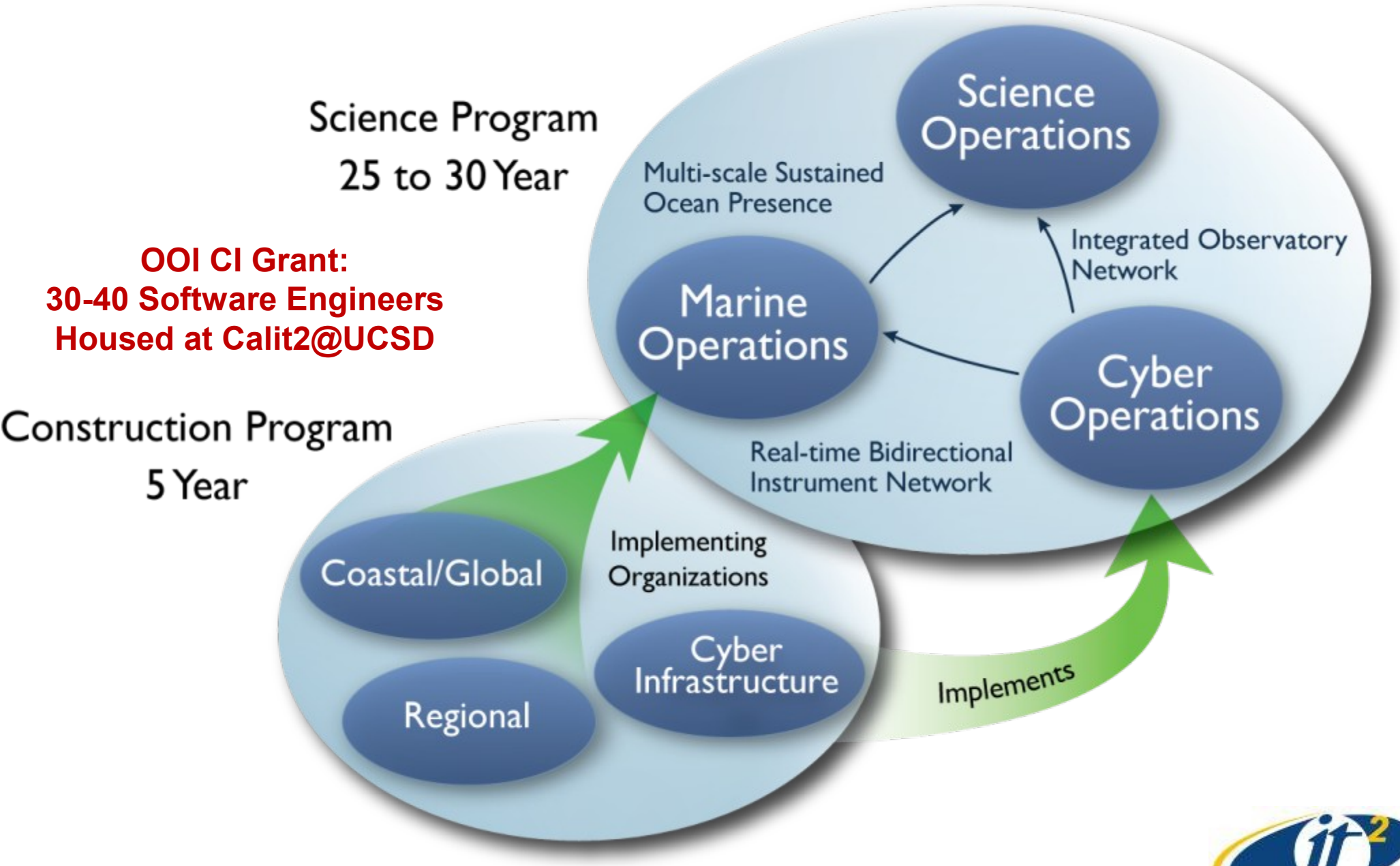
SDSC

*ANL * Calit2 * LBNL * NICS * ORNL * SDSC

it²

NSF's Ocean Observatory Initiative Has the Largest Funded NSF CI Grant

OCEAN OBSERVATORIES INITIATIVE



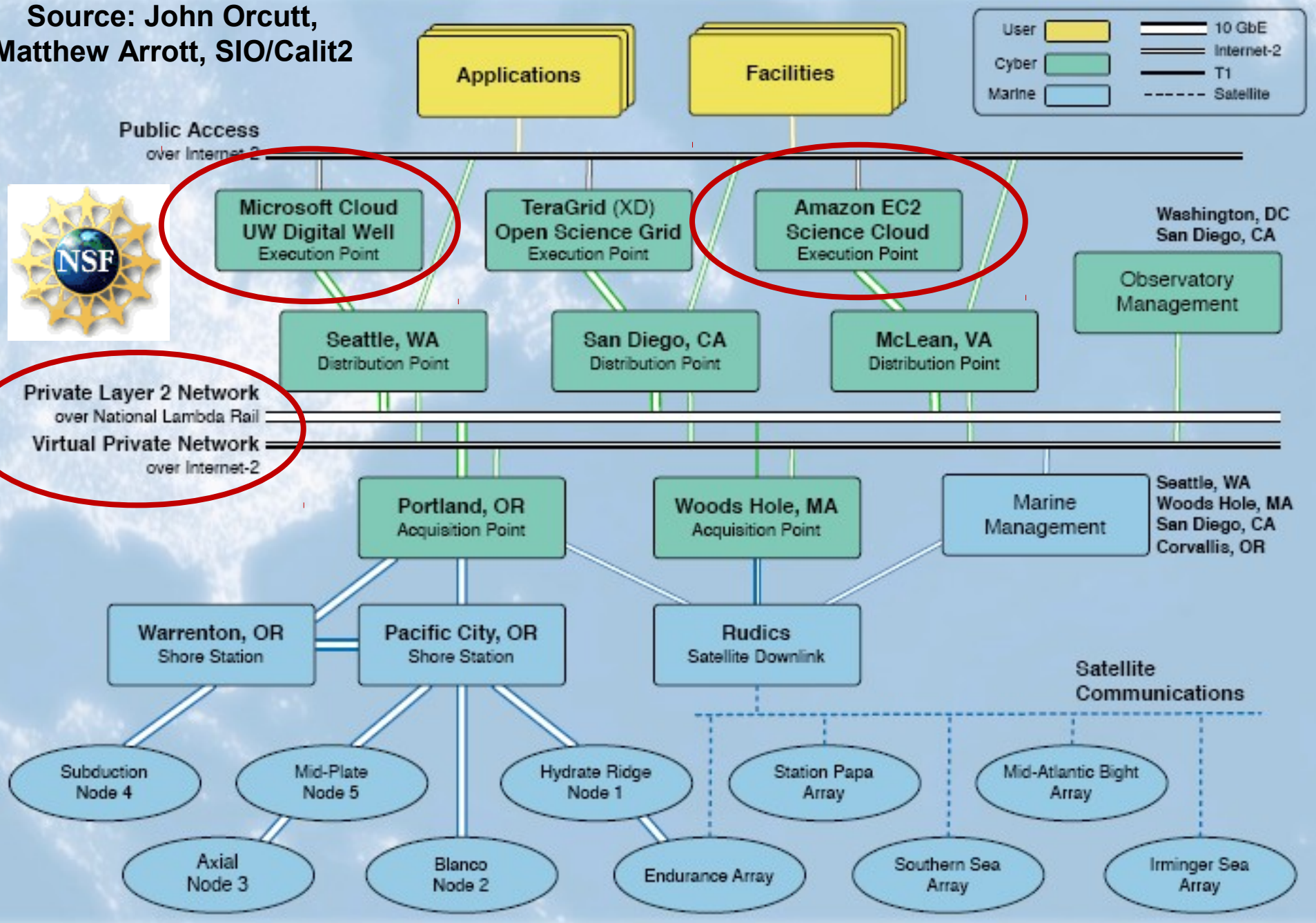
OOI CI Grant:
30-40 Software Engineers
Housed at Calit2@UCSD

Source: Matthew Arrott, Calit2 Program Manager for OOI CI

OOI CI is Built on Dedicated Optical Infrastructure Using Clouds

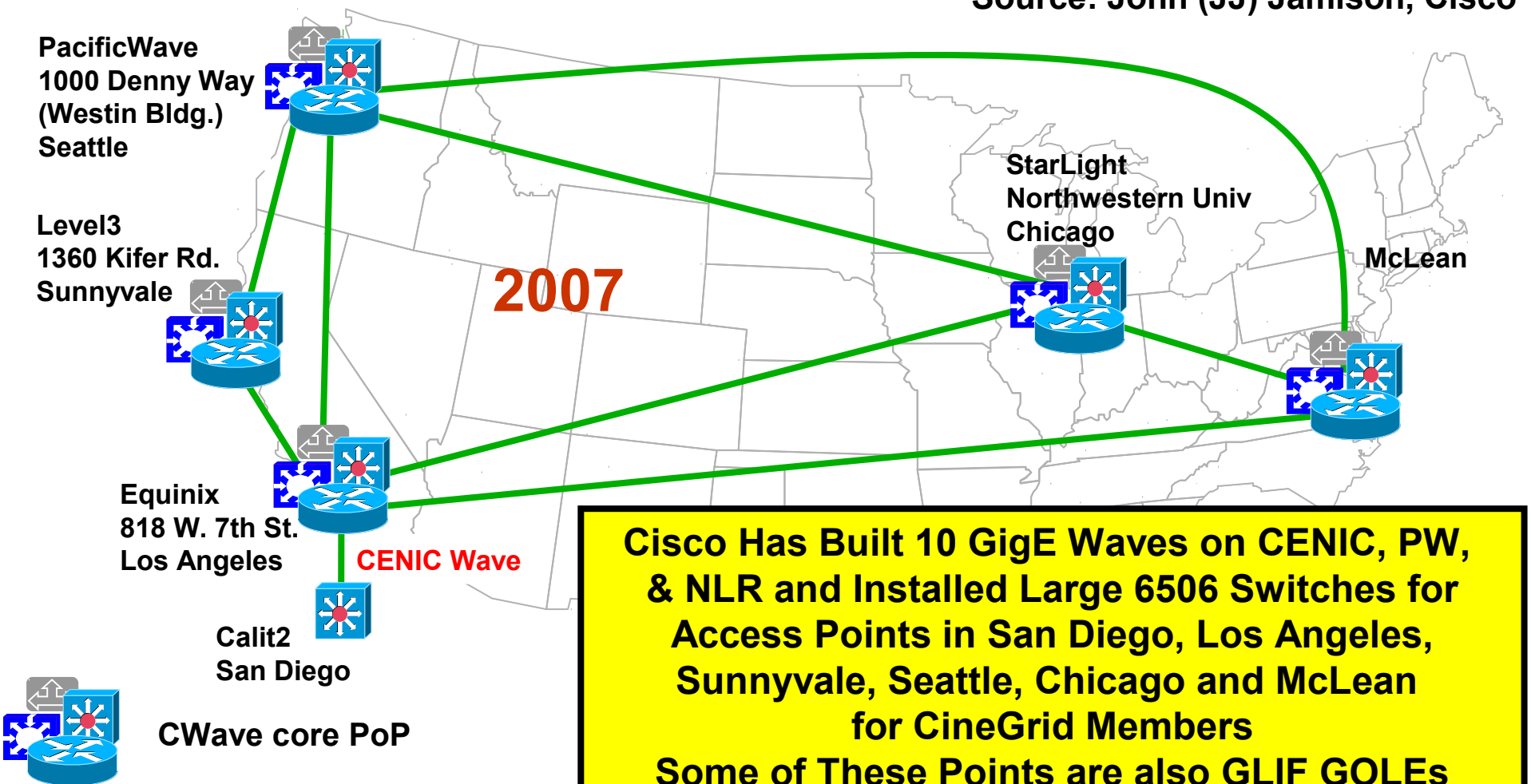
OCEAN OBSERVATORIES INITIATIVE

Source: John Orcutt, Matthew Arrott, SIO/Calit2



Cisco CWave for CineGrid: A New Cyberinfrastructure for High Resolution Media Streaming*

Source: John (JJ) Jamison, Cisco

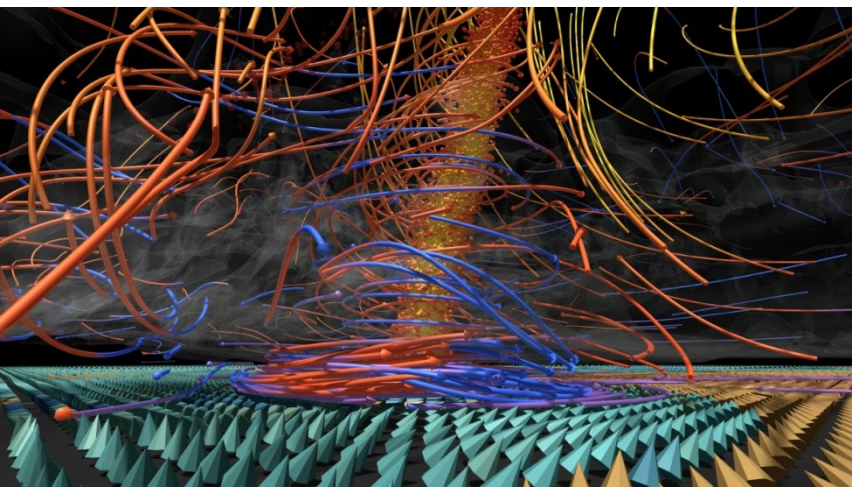


Cisco Has Built 10 GigE Waves on CENIC, PW, & NLR and Installed Large 6506 Switches for Access Points in San Diego, Los Angeles, Sunnyvale, Seattle, Chicago and McLean for CineGrid Members
Some of These Points are also GLIF GOLEs

*CTWatch QUARTERLY May 2007

CineGrid 4K Digital Cinema Projects: "Learning by Doing"

CineGrid @ iGrid 2005



CineGrid @ AES 2006





CineGrid @ Holland Festival 2007



CineGrid @ GLIF 2007

Laurin Herr, Pacific Interface; Tom DeFanti, Calit2

CineGrid 4K Remote Microscopy Collaboratory: USC to Calit2



Photo: Alan Decker

December 8, 2009

Richard Weinberg, USC

OptIPuter Persistent Infrastructure Enables Calit2 and U Washington CAMERA Collaboratory



OptIPuter

Ginger Armbrust's Diatoms: Micrographs, Chromosomes, Genetic Assembly

ROCKS



Gordon and Betty MOORE FOUNDATION

camera
Marine Microbial Ecology



NSF

it²

OptIPortals are Beginning to be Built into Distributed Centers



April 2009

Cross-Disciplinary Research at MIT, Connecting Systems Biology, Microbial Ecology, Global Biogeochemical Cycles and Climate

Linking the Calit2 Auditoriums at UCSD and UCI with LifeSize HD for Shared Seminars



Photo by Erik Jepsen, UC San Diego

Launch of the 100 Megapixel OzIPortal Kicked Off a Rapid Build Out of Australian OptIPortals



No Calit2 Person Physically Flew to Australia to Bring This Up!

Covise, Phil Weber, Jurgen Schulze, Calit2
CGLX, Kai-Uwe Doerr , Calit2
http://www.calit2.net/newsroom/release.php?id=1421

Multi-User Global Workspace:
Calit2 (San Diego), EVL (Chicago), KAUST (Saudi Arabia)



Source: Tom DeFanti, KAUST Project, Calit2

Live Remote Surgery for Teaching Has Become Routine: APAN 26th in New Zealand (2008)

August 2008

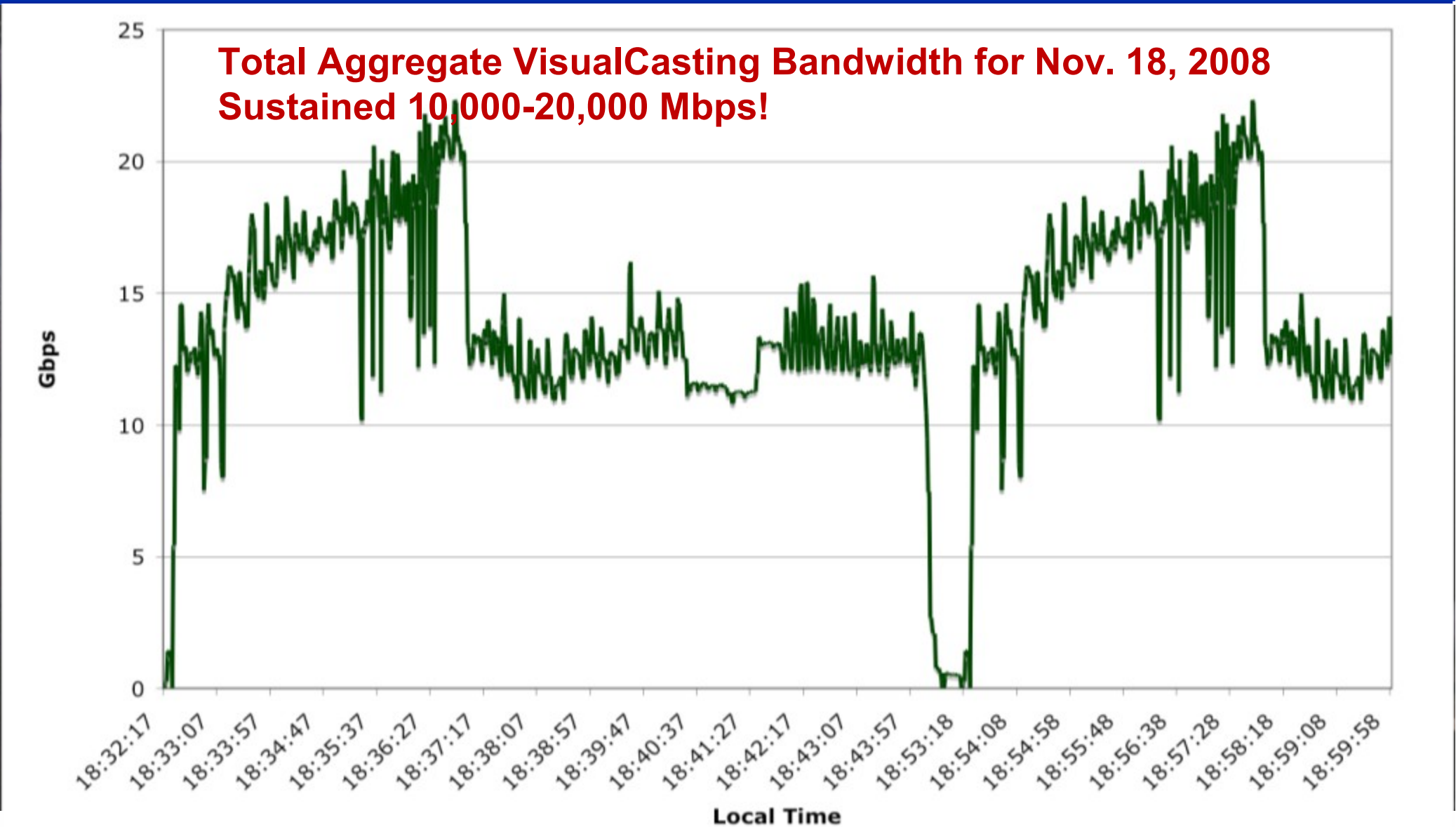


First Tri-Continental Premier of a Streamed 4K Feature Film With Global HD Discussion



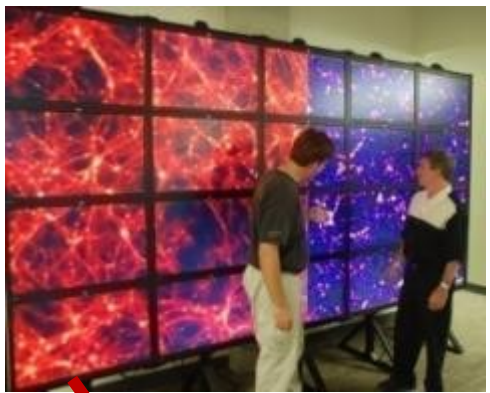
4K Transmission Over 10Gbps--
4 HD Projections from One 4K Projector

EVL's SAGE OptIPortal VisualCasting Multi-Site OptIPuter Collaboratory

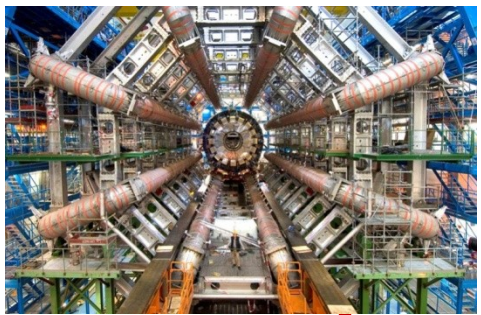


Source: Jason Leigh, Luc Renambot, EVL, UI Chicago

Academic Research OptIPlanet Collaboratory: A 10Gbps "End-to-End" Lightpath Cloud



End User OptIPortal



Instruments



HD/4k Live Video

HPC



TeraGrid



10G Lightpaths



Campus Optical Switch



Data Repositories & Clusters



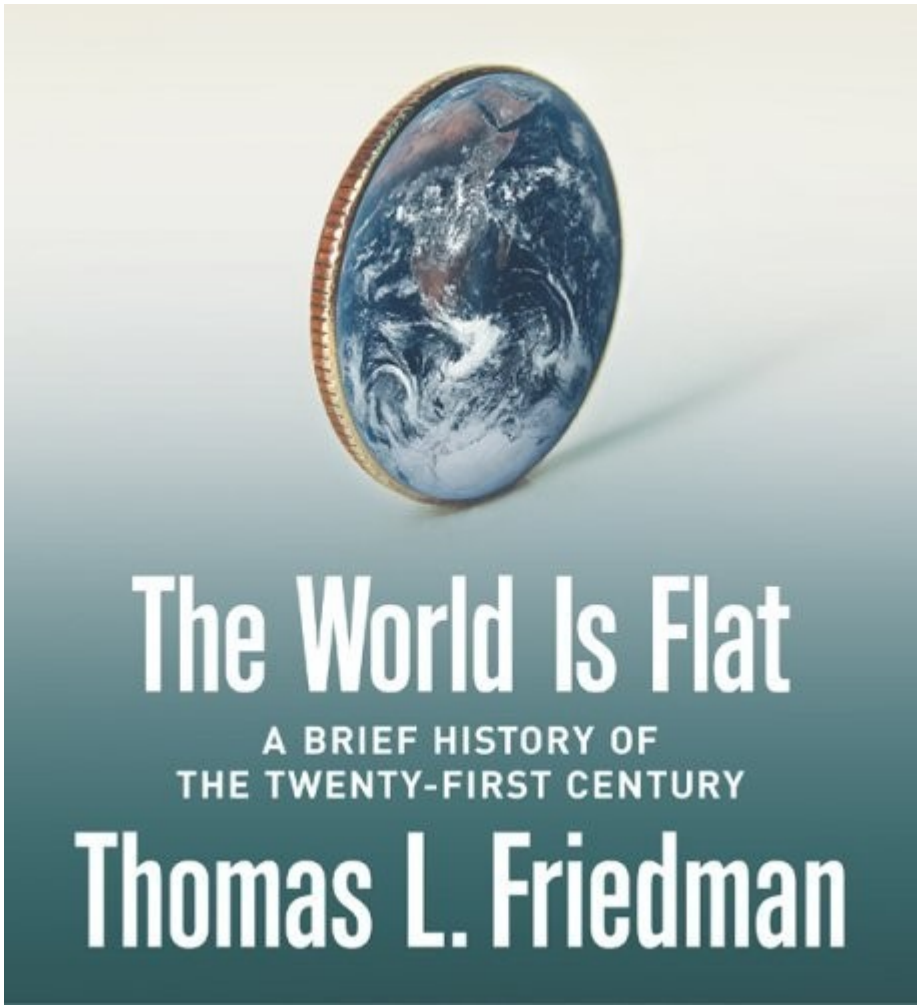
HD/4k Video Repositories



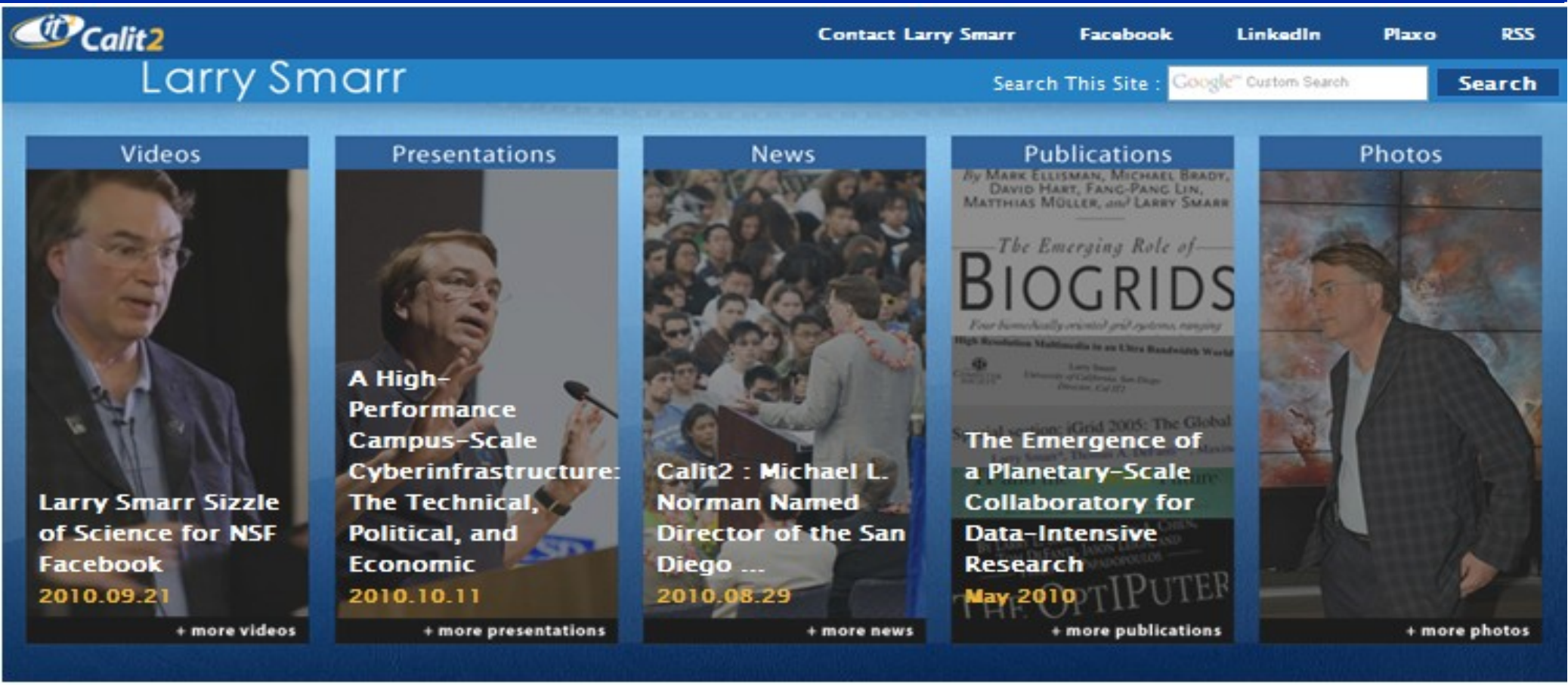


Ten Years Old Technologies--the Shared Internet & the Web--Have Made the World "Flat"

- **But Today's Innovations**
  - Dedicated Fiber Paths
  - Streaming HD TV
  - Large Display Systems
  - Massive Computing/Storage
- **Are Reducing the World to a "Single Point"**
  - How Will Our Society Reorganize Itself?



You Can Download This Presentation at lsmarr.calit2.net



ABOUT LARRY SMARR

Larry Smarr is the founding Director of the California Institute for Telecommunications and Information Technology (Calit2), a UC San Diego/UC Irvine partnership, and holds the Harry E. Gruber professorship in Computer Science and Engineering (CSE) at UCSD's Jacobs School. At Calit2, Smarr has

WHAT I'M READING

Complete List of Larry's Recommendations

Homo Britannicus: The Incredible Story of Human Life in Britain (2006)

American College and University

RECENT TWITTER POSTS

Twitter? Click To Follow Me

Ismarr: SDSC Founding Director Sid Karin and I are making a few comments on the history of the NSF Supercomputing Centers at