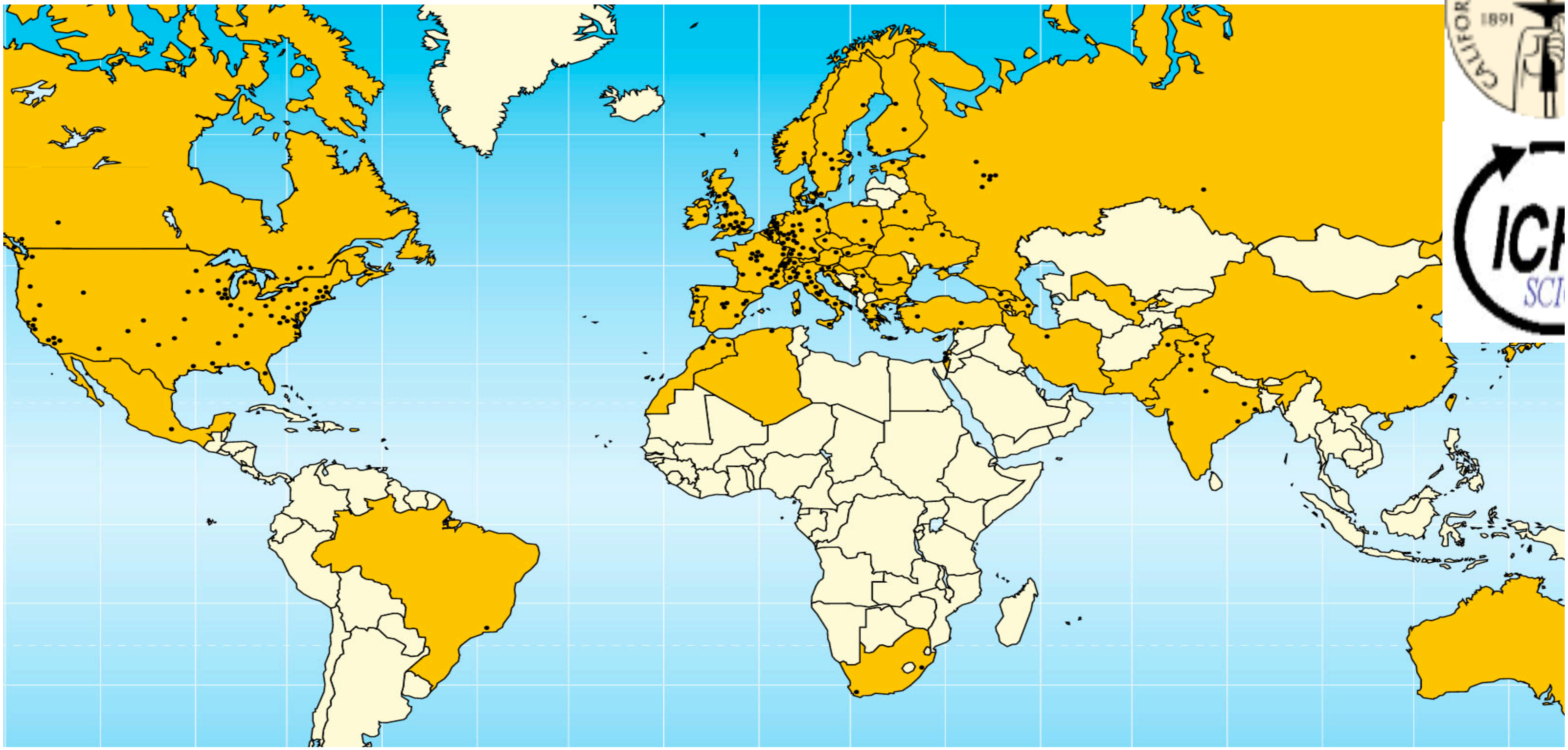
# Networks and Grids for HENP as Global e-Science



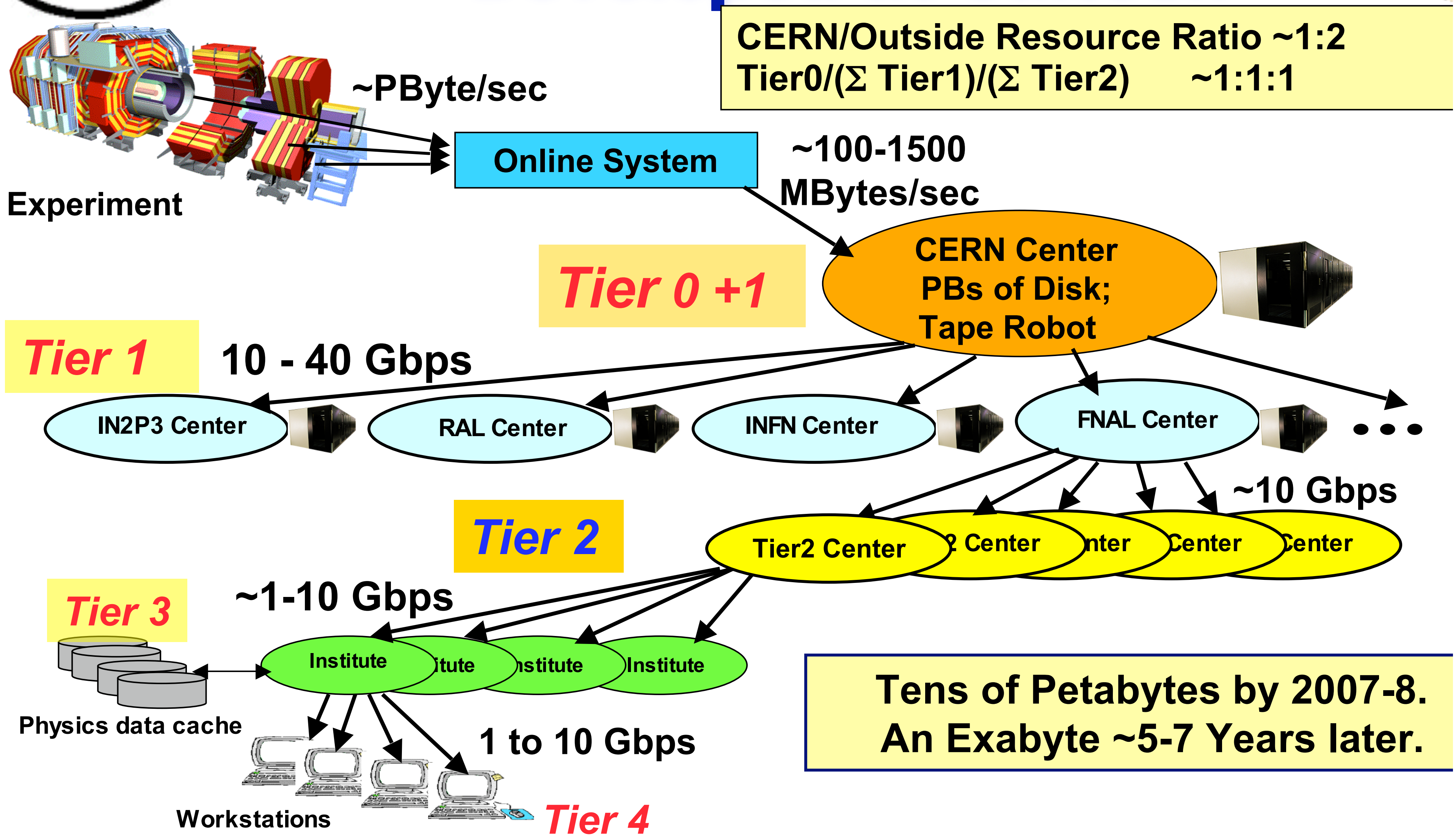

## Harvey B. Newman

**California Institute of Technology**


**SC2004, Pittsburgh**

**November 11, 2004**

# LHC Data Grid Hierarchy: Developed at Caltech




**Emerging Vision: A Richly Structured, Global Dynamic Syste**

# Challenges of Next Generation Science in the Information Age

**Petabytes of complex data explored and analyzed by 1000s of globally dispersed scientists, in hundreds of teams**

- Flagship Applications
  - High Energy & Nuclear Physics, AstroPhysics Sky Surveys:
    TByte to PByte "block" transfers at 1-10+ Gbps
  - Fusion Energy: Time Critical Burst-Data Distribution;
    Distributed Plasma Simulations, Visualization, Analysis
  - eVLBI: Many real time data streams at 1-10 Gbps
  - BioInformatics, Clinical Imaging: GByte images on dema
- Advanced integrated Grid applications rely on reliable, high performance operation of our LANs and WANs
- *Analysis Challenge:* Provide results with rapid turnaround, over networks of varying capability in different world regions

# HENP Bandwidth Roadmap for Major Links (in Gbps)

| Year | Production | Experimental | Remarks |
|---|---|---|---|
| 2001 | 0.155 | 0.622-2.5 | SONET/SDH |
| 2002 | 0.622 | 2.5 | SONET/SDH DWDM; GigE Integ. |
| 2003 | 2.5 | 10 | DWDM; 1 + 10 GigE Integration |
| 2005 | 10 | 2-4 X 10 | λ Switch; λ Provisioning |
| 2007 | 2-4 X 10 | ~10 X 10; 40 Gbps | 1st Gen. λ Grids |
| 2009 | ~10 X 10 or 1-2 X 40 | ~5 X 40 or ~20-50 X 10 | 40 Gbps λ Switching |
| 2011 | ~5 X 40 or ~20 X 10 | ~25 X 40 or ~100 X 10 | 2nd Gen λ Grids Terabit Networks |
| 2013 | ~Terabit | ~MultiTbps | ~Fill One Fiber |

**Continuing Trend: ~1000 Times Bandwidth Growth Per Decade**
**Compatible with Other Major Plans (ESNet, NLR; GN2, GLIF)**

# HENP Lambda Grids: Fibers for Physics

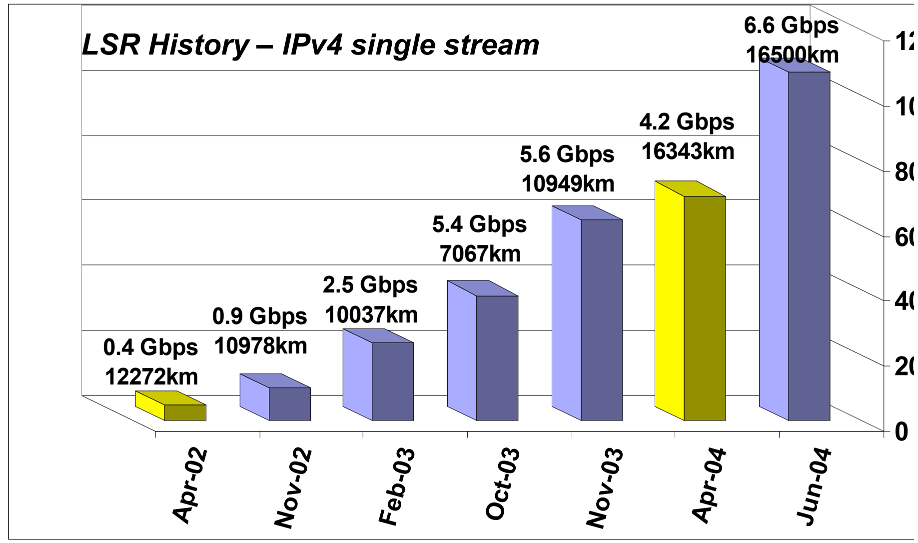
- Problem: Extract "Small" Data Subsets of 1 to 100 Terabytes from 1 to 1000 Petabyte Data Stores
- Survivability of the HENP Global Grid System, with hundreds of such transactions per day (circa 2007) requires that each transaction be completed in a relatively short time.
- Example: Take 800 secs to complete the transaction. Then

| Transaction Size (TB) | Net Throughput (Gbps) |
|---|---|
| 1 | 10 |
| 10 | 100 |
| 100 | 1000 (Capacity of Fiber Today) |

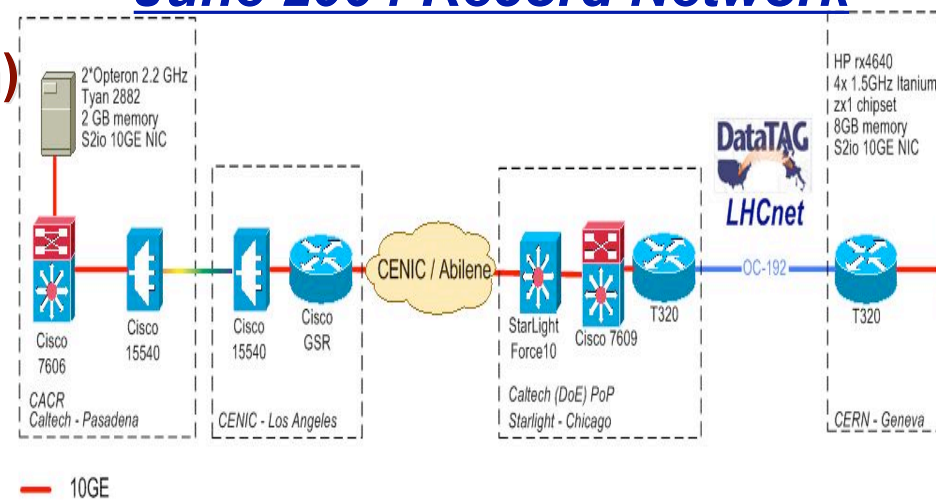
- Summary: Providing Switching of 10 Gbps wavelengths within ~2-4 years; and Terabit Switching within 5-8 years would enable "Petascale Grids with Terabyte transactions", to fully realize the discovery potential of major HENP programs, as well as other data-intensive research.

# Internet 2 Land Speed Records (LSR):
## Redefining the Role and Limits of TCP

- **Judged on product of transfer speed and distance end-to-end, using standard (TCP/IP) protocols, Across Production Net: e.g. Abilene**
- **IPv6: 4.0 Gbps Geneva-Phoenix (SC2003)**
- **IPv4 with Windows & Linux: 6.6 Gbps Caltech-CERN (15.7 kkm; "Grand Tour of Abilene") June 2004**
  - **Exceeded 100 Petabit-m/sec**
- **7.48 Gbps X 16 kkm (Linux, 1 Stream) Achieved in July**
- **11 Gbps (802.3ad) Over LAN in Sept.**
- ***Concentrate now on reliable Terabyte-scale file transfers***
  - ***Note System Issues: CPU, PCI-X Bus, NIC, I/O Controllers, Drivers***




**June 2004 Record Network**




**SC04: 6.9 Gbps X 26 kkm 11/08**

***SC04 BW Challenge: 101.1 Gbps***

# Evolving Quantitative Science Requirements for Networks (DOE High Perf. Network Workshop)

| Science Areas | Today *End2End* Throughput | 5 years End2End Throughput | 5-10 Years End2End Throughput | Remarks |
|---|---|---|---|---|
| High Energy Physics | 0.5 Gb/s | 100 Gb/s | **1000 Gb/s** | High bulk throughpu |
| Climate (Data & Computation) | 0.5 Gb/s | 160-200 Gb/s | **N x 1000 Gb/s** | High bulk throughpu |
| SNS NanoScience | Not yet started | 1 Gb/s | **1000 Gb/s + QoS for Control Channel** | Remote control anc time critica throughpu |
| Fusion Energy | 0.066 Gb/s (500 MB/s burst) | 0.198 Gb/s (500MB/ 20 sec. burst) | **N x 1000 Gb/s** | Time critica throughpu |
| Astrophysics | 0.013 Gb/s (1 TByte/week) | N*N multicast | **1000 Gb/s** | Computat' steering an collaboratio |
| Genomics Data & Computation | 0.091 Gb/s (1 TBy/day) | 100s of users | **1000 Gb/s + QoS for Control Channel** | High throughpu and steerin |

See http://www.doecollaboratory.org/meetings/hpnpw/

# *Transition* beginning now to optical, multi-wavelength Community owned or leased "dark fiber" (10 GbE) networks for R&E

## National Lambda Rail (NLR): www.nlr.ne

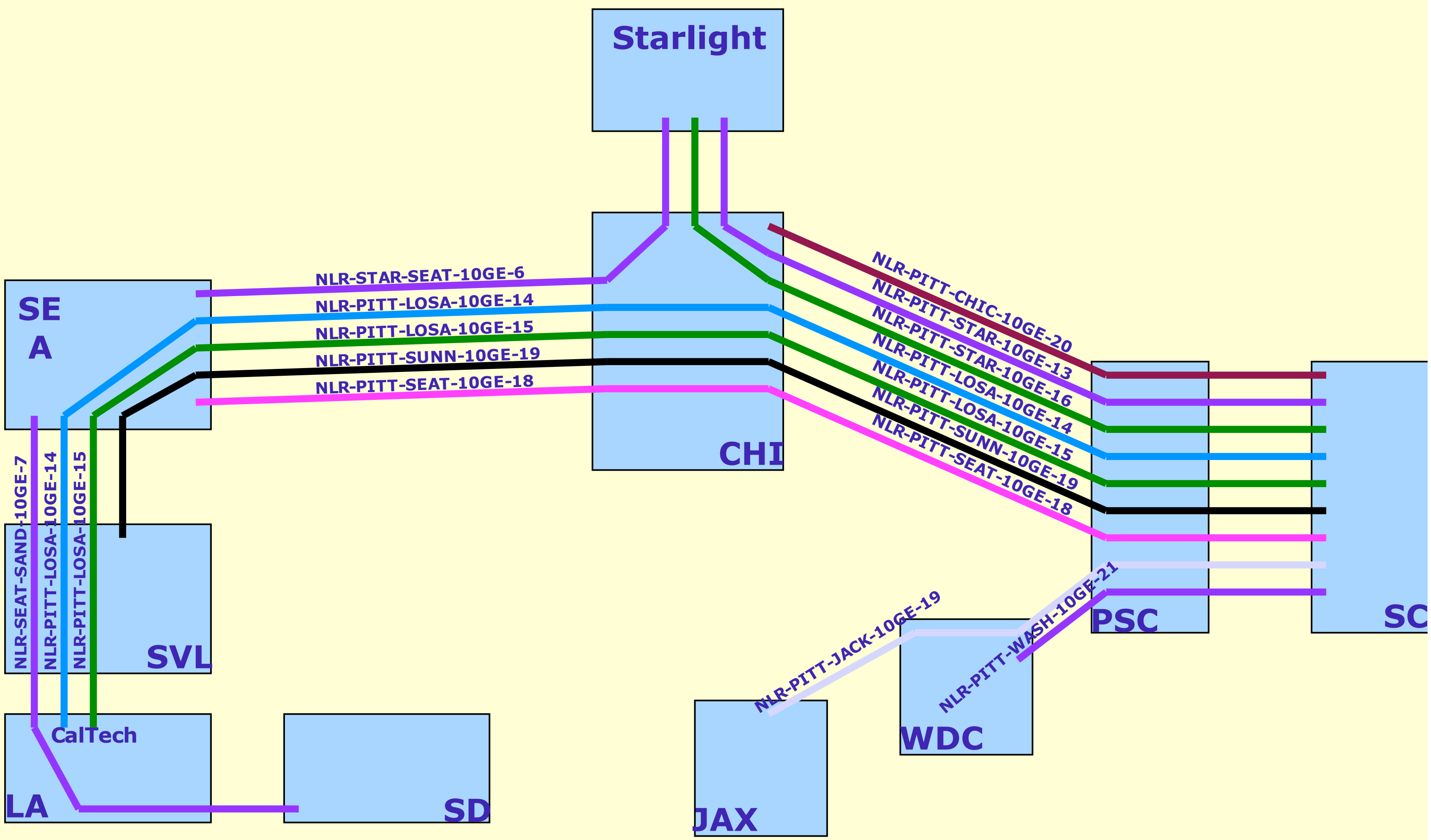

**NLR**

- Coming Up Now
- Initially 4 10G Wavelengths
- Northern Route LA-JAX by 4Q04
- Internet2 HOPI Initiative (w/HEP)
- To 40 10G Waves in Future

- Initiatives in: nl, ca, pl, cz, uk, ko, jp
- + 18 US States (CA, IL, FL, IN, …)

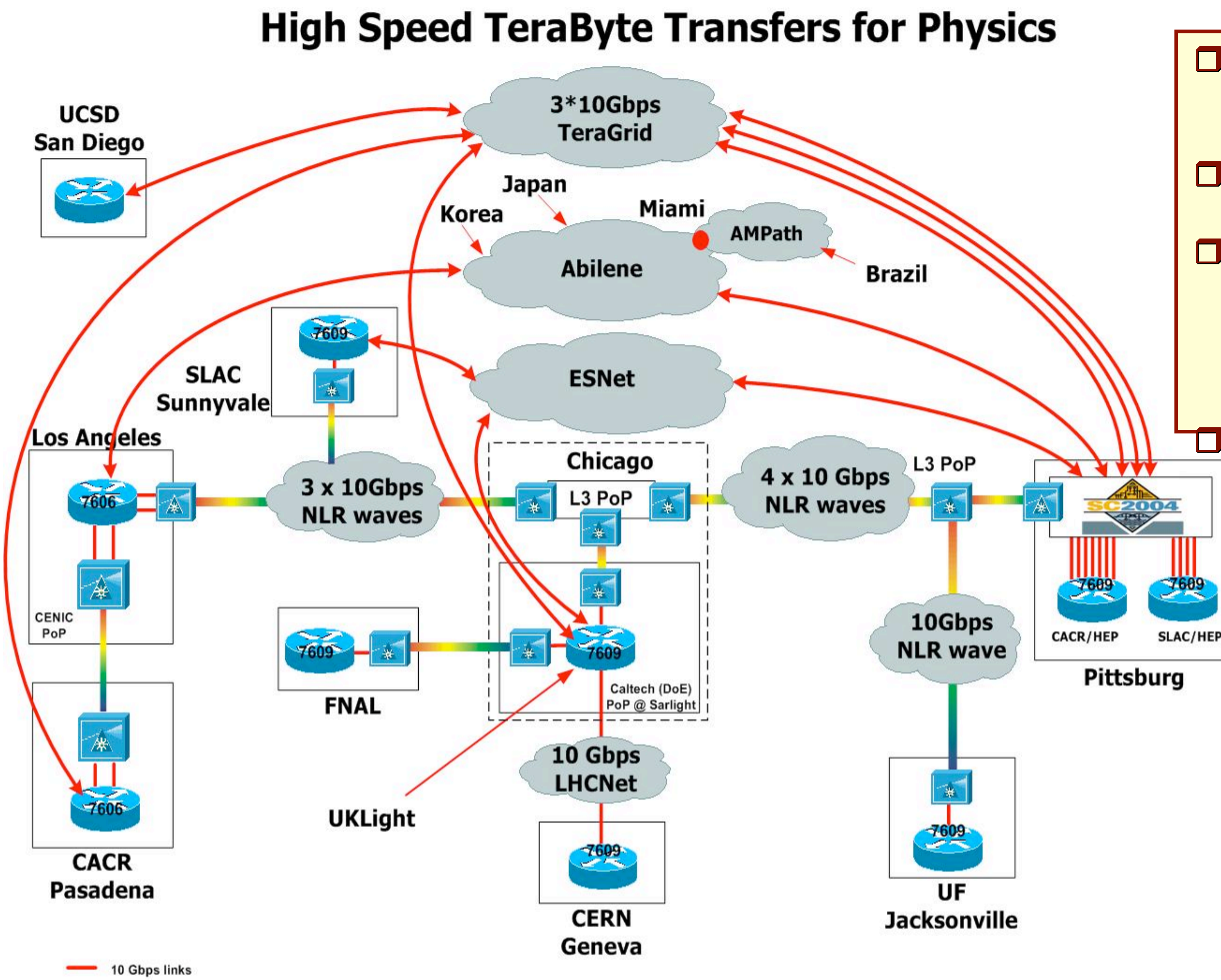
CalTech/New
FL/A
S
Opti
Ed S
UW/Rsrch

# 21 NLR Waves: 9 to SC04

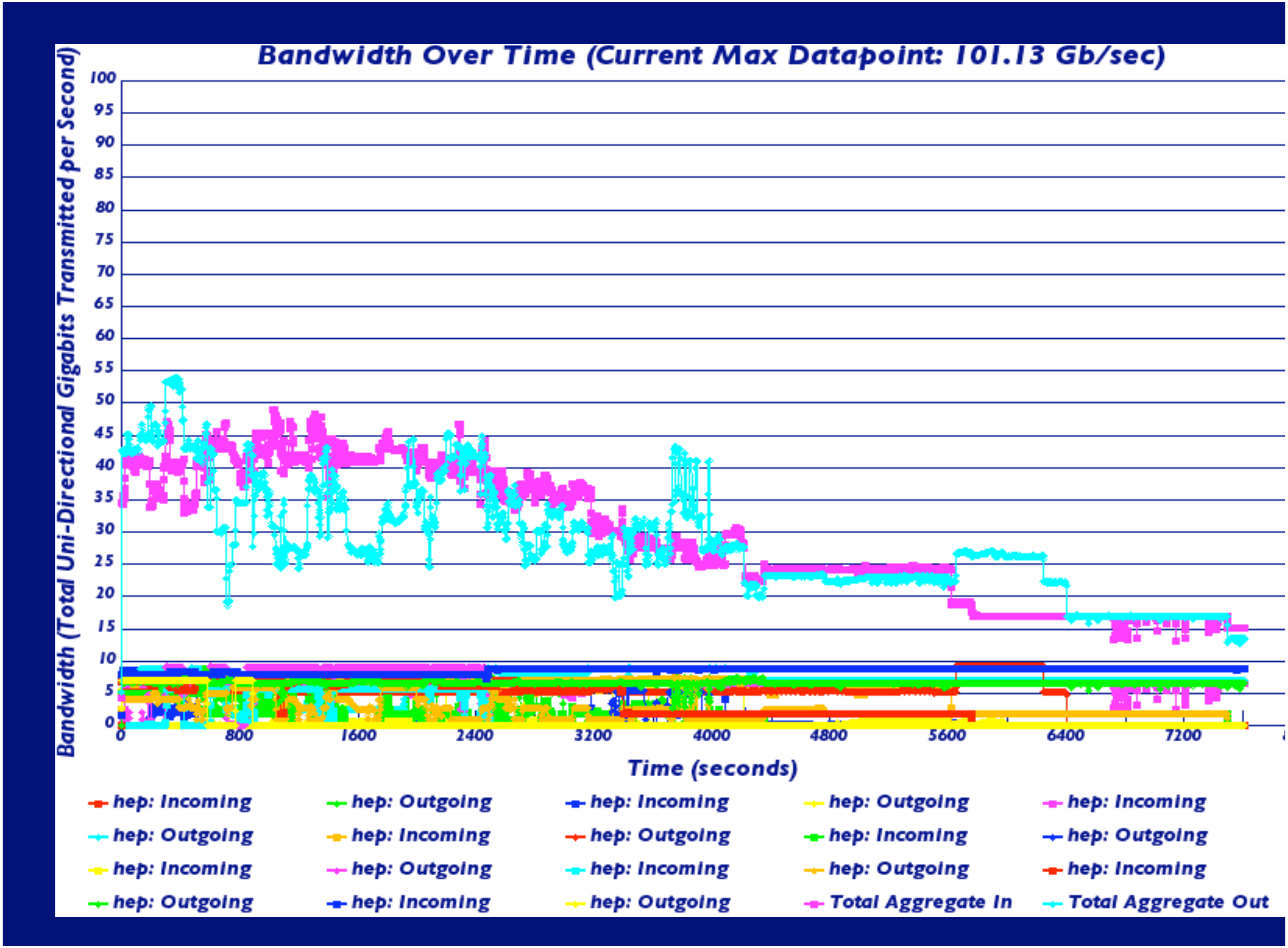



*All lines*

# SC2004: HEP network layout

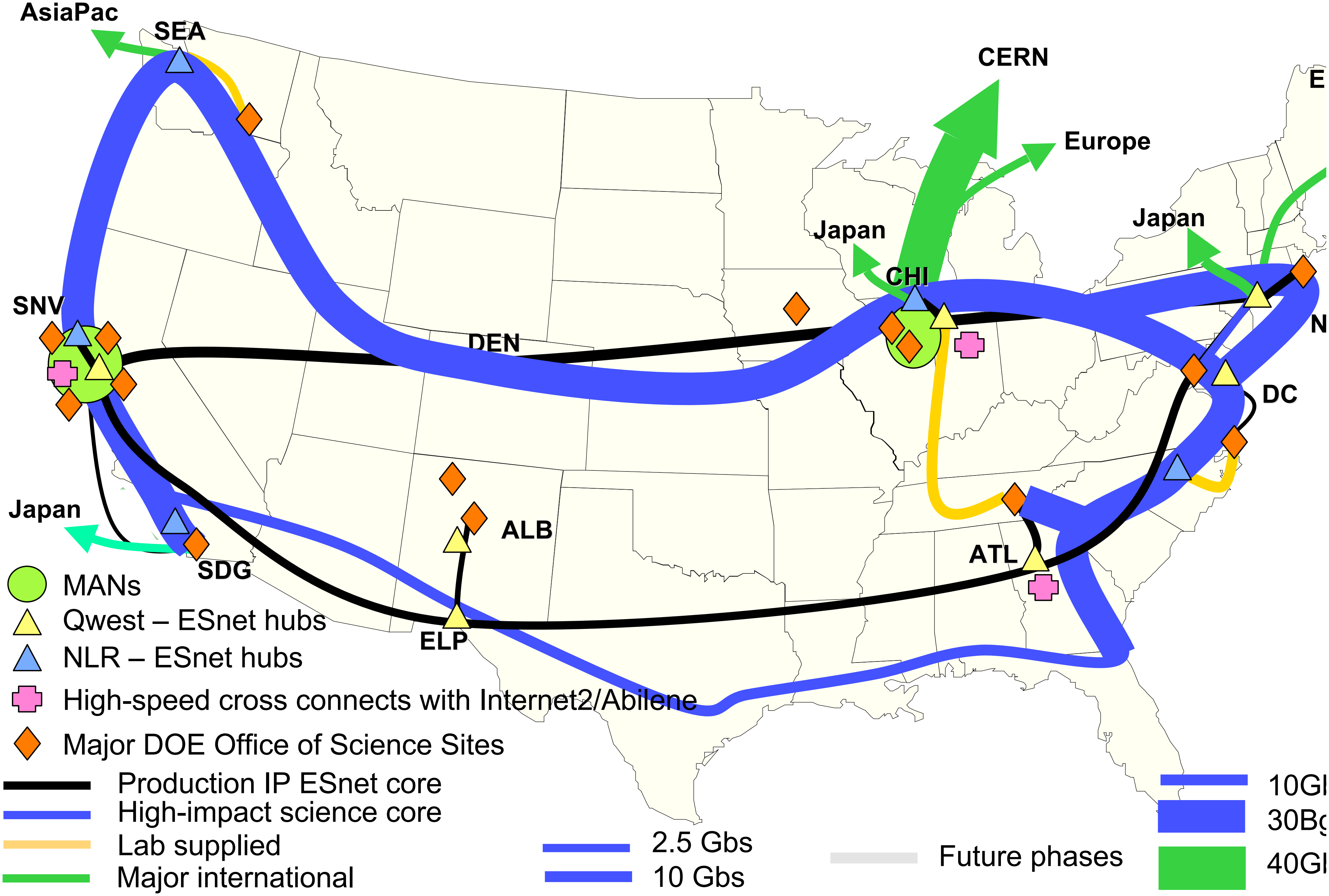



- ***Joint Caltech, FNA[L], CERN, SLAC, UF, SDSC, BR, KR, ….***
- ***10 10 Gbps waves HEP on show floor***
- ***Bandwidth challenge: aggrega[te] throughput of 101.13 Gbps achieved***
- ***FAST TCP***

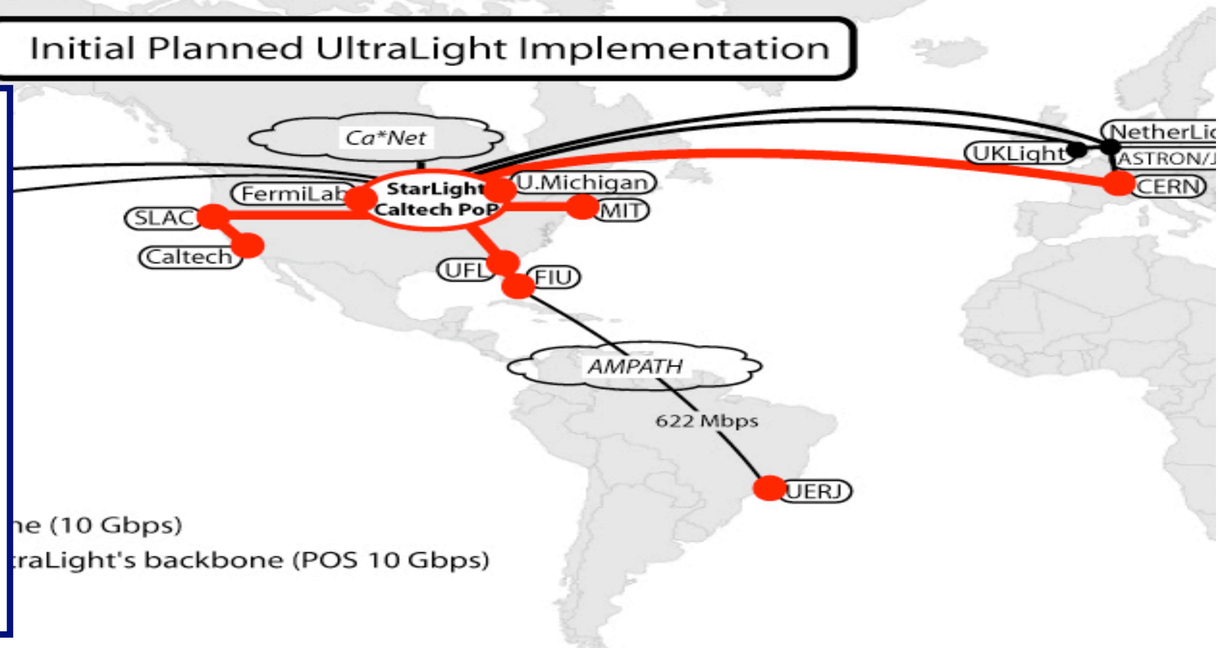


Bandwidth Over Time (Current Max Datapoint: 101.13 Gb/sec)
Bandwidth (Total Uni-Directional Gigabits Transmitted per Second)
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
800
1600
2400
3200
4000
4800
5600
6400
7200
Time (seconds)
hep: Incoming
hep: Outgoing
hep: Incoming
hep: Outgoing
hep: Incoming
hep: Outgoing
hep: Incoming
hep: Outgoing
hep: Incoming
hep: Outgoing
hep: Incoming
hep: Outgoing
hep: Incoming
hep: Outgoing
hep: Incoming
hep: Outgoing
hep: Incoming
hep: Outgoing
Total Aggregate In
Total Aggregate Out

# ESnet Beyond FY07 (W. Johnston)



# UltraLight Collaboration: http://ultralight.caltech.edu




- **Caltech, UF, UMich, SLAC,FNAL, CERN, MIT, FIU, NLR, CENIC, UCAID, Translight, UKLight, Netherlight, UvA, UCLondon, KEK, Taiwan, KNU (Korea), UERJ (Rio), Sao Paolo**
- **Cisco, Level(3)**

- **Integrated hybrid experimental network, leveraging Transatlantic R&D network partnerships; packet-switched + dynamic optical paths**
  - **10 GbE across US and the Atlantic: NLR, DataTAG, TransLight, NetherLight, UKLight, etc.; Extensions to Japan, Taiwan, Korea, Brazil**
- **End-to-end monitoring; Realtime tracking and optimization; Dynamic bandwidth provisioning**
- ***Agent-based services spanning all layers of the system,* from the optical cross-connects to the applications.**

# Research Networking in Latin America: *Just Taking Off in 2004*

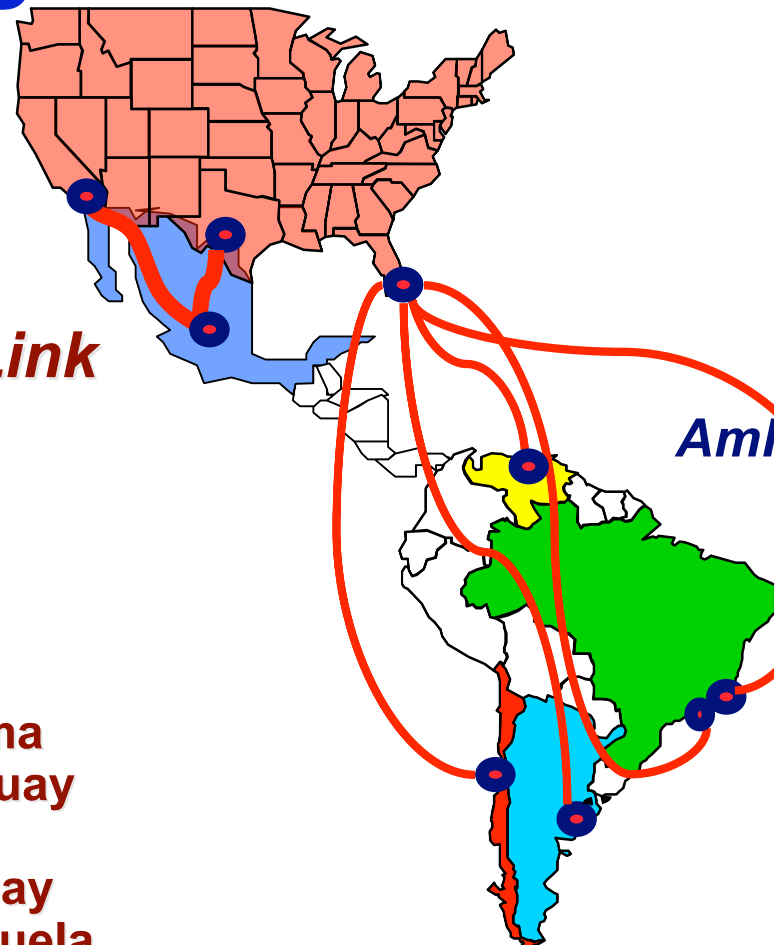
- ***AmPath*** **Provided connectivity for some Latin American countries**
  - ➡ Argentina, Brazil, Chile, Mexico, Venezuela
- ***New CHEPREO Sao Paolo-Miami Link at 622 Mbps Starting This Month***

**New: CLARA (Funded by EU)**

- **Regional Network Connecting 19 Countries:**

| | | |
|---|---|---|
| Argentina | Dominican Republic | Panama |
| Brasil | Ecuador | Paraguay |
| Bolivia | El Salvador | Peru |
| Chile | Guatemala | Uruguay |
| Colombia | Honduras | Venezuela |
| Costa Rica | Mexico | Nicaragua |
| Cuba | | |

**155 Mbps Backbone with 10-45 Mbps Spurs; 4 Mbps Satellite to Cuba; 622 Mbps to Europe**



**Also WHREN NSF Proposal: 2.5G to U**

**Brazilian HEPGrid: Rio, Sao Paolo etc.**

# GLORIAD

*Global Ring Network for Advanced Applications Development*

**www.gloriad.org: US-RUSSIA-CHINA + KOREA Global Optical Ring**

- **OC3 circuits Moscow-Chicago-Beijing since January 2004**
- **OC3 circuit Moscow-Beijing July 2004 (completes the ring)**
- **Rapid traffic growth with heaviest US use from DOE (FermiLab), NASA, NOAA, NIH and 260+ Univ. (UMD, IU, UCB, UNC, UMN… Many Others)**
- **Plans for Central Asian extension, with Kyrgyz Gov't**


**Aug. 8 2004: P.K. Young, Korean IST Advisor to President Announces**

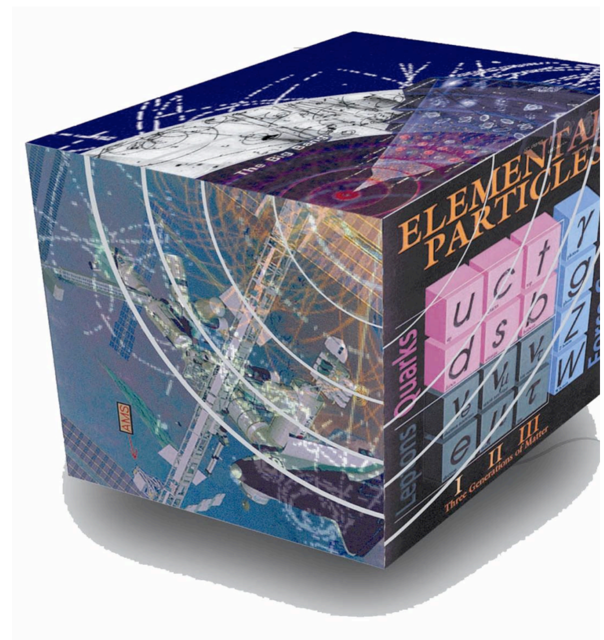
- **Korea Joining GLORIAD as a full partner**

**> 5TBytes now transferred monthly via GLORIAD to US, Russia, China**

**GLORIAD 5-year Proposal (with US NSF) for expansion to 2.5G-10G Moscow-Amsterdam-Chicago-Pacific-Hong Kong-Pusan-Beijing early 2005**
**10G ring around northern hemisphere 2007;**
**Multi-wavelength hybrid service from ~2008-9**

# International ICFA Workshop on HEP Networking, Grids and Digital Divide Issues for Global e-Science

**Dates: May 23-27, 2005**
**Venue: Daegu, Korea**




**Dongchul Son**
**Center for High Energy Physics**
**Kyungpook National University**


**ICFA, Beijing, China**
**Aug. 2004**

**Approved by ICFA**
**August 20, 2004**

# International ICFA Workshop on HEP Networking, Grids and Digital Divide Issues for Global e-Science

- **<u>Workshop Goals</u>**
  - Review the current status, progress and barriers to effective use of major national, continental and transoceanic networks used by HEP
  - Review progress, strengthen opportunities for collaboration, and explore the means to deal with key issues in Grid computing and Grid-enabled data analysis, for high energy physics and other fields of data intensive science, now and in the future
  - Exchange information and ideas, and formulate plans to develop solutions to specific problems related to the Digital Divide in various regions, with a focus on Asia Pacific, as well as Latin America, Russia and Africa
  - Continue to advance a broad program of work on reducing or eliminating the Digital Divide, and ensuring global collaboration, as related to all of the above aspects.

# Networks and Grids for HENP and Global Science

- Network backbones and major links used by HENP and other fields are advancing rapidly
  - To the 10 G range in < 3 years; much faster than Moore's Law
  - New HENP and DOE Roadmaps: a factor ~1000 BW Growth per decad
- We are learning to use long distance 10 Gbps networks effectively
  - 2004 Developments: to 7 - 7.5 Gbps flows with TCP over 16-25 kkm
- Transition to community-operated optical R&E networks (us, ca, nl, pl, c sk, kr, jp …); Emergence of a new generation of "hybrid" optical network
- ***We Must Work to Close to Digital Divide***
  - *To Allow Scientists in All World Regions to Take Part in Discoveries*
  - Removing Regional, Last Mile, Local Bottlenecks and Compromises in Network Quality are now *On the Critical Pa*
- *Important Examples* on the Road to Progress in Closing the Digital Divid
  - CLARA, CHEPREO, and the Brazil HEPGrid in Latin America
  - Optical Networking in Central and Southeast Europe
  - APAN Links in the Asia Pacific: GLORIAD and TEIN
  - *Leadership and Outreach: HEP Groups in Europe, US, Japan, & Kore*