Networks and Grids for HEP and Global e-Science

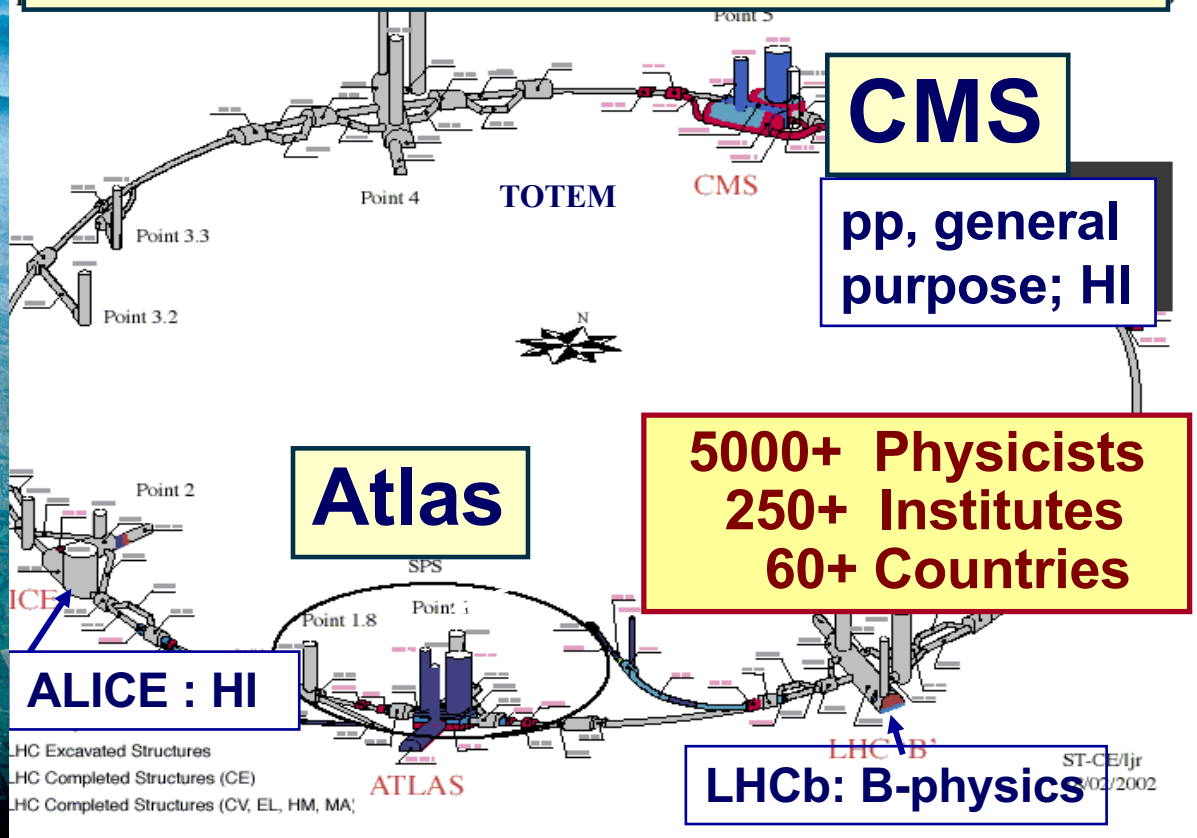
Harvey B. Newman

California Institute of Technology

CANS 2004, Miami
December 1, 2004

Large Hadron Collider (LHC)
CERN, Geneva: 2007 Start

- pp $\sqrt{s}$ =14 TeV L=10^{34} cm^{-2} s^{-1}
- 27 km Tunnel in Switzerland & France



CMS

pp, general purpose; HI

Atlas

5000+ Physicists
250+ Institutes
60+ Countries

ALICE : HI

LHCb: B-physics

Higgs, SUSY, QG Plasma, CP Violation, … *the Unexpected*

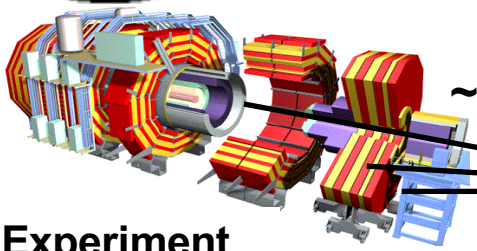
Challenges of Next Generation Science in the Information Age

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 demand**
- **Advanced integrated Grid applications rely on reliable, high performance operation of our LANs and WANs**
- ***Analysis Challenge:*** **Provide results to thousands of scientists. with rapid turnaround, over networks of varying capability in different world regions**

LHC Data Grid Hierarchy: Developed at Caltech

CERN/Outside Resource Ratio ~1:2
Tier0/(Σ Tier1)/(Σ Tier2) ~1:1:1



Experiment

~PByte/sec

Online System

~100-1500 MBytes/sec

Tier 0 +1

CERN Center
PBs of Disk;
Tape Robot

Tier 1

10 - 40 Gbps

IN2P3 Center

RAL Center

INFN Center

FNAL Center

~10 Gbps

Tier 2

Tier2 Center

Tier 3

~1-10 Gbps

Institute

Physics data cache

1 to 10 Gbps

Workstations

Tier 4

Tens of Petabytes by 2007-8.
An Exabyte ~5-7 Years later.

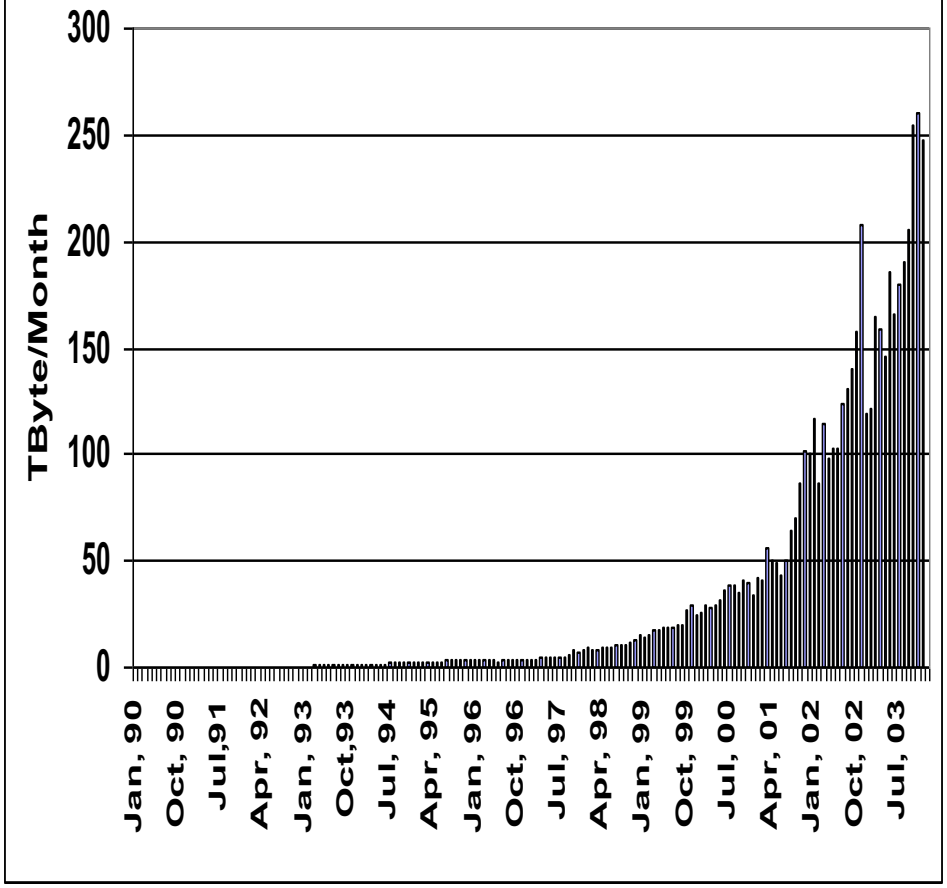
Emerging Vision: A Richly Structured, Global Dynamic System

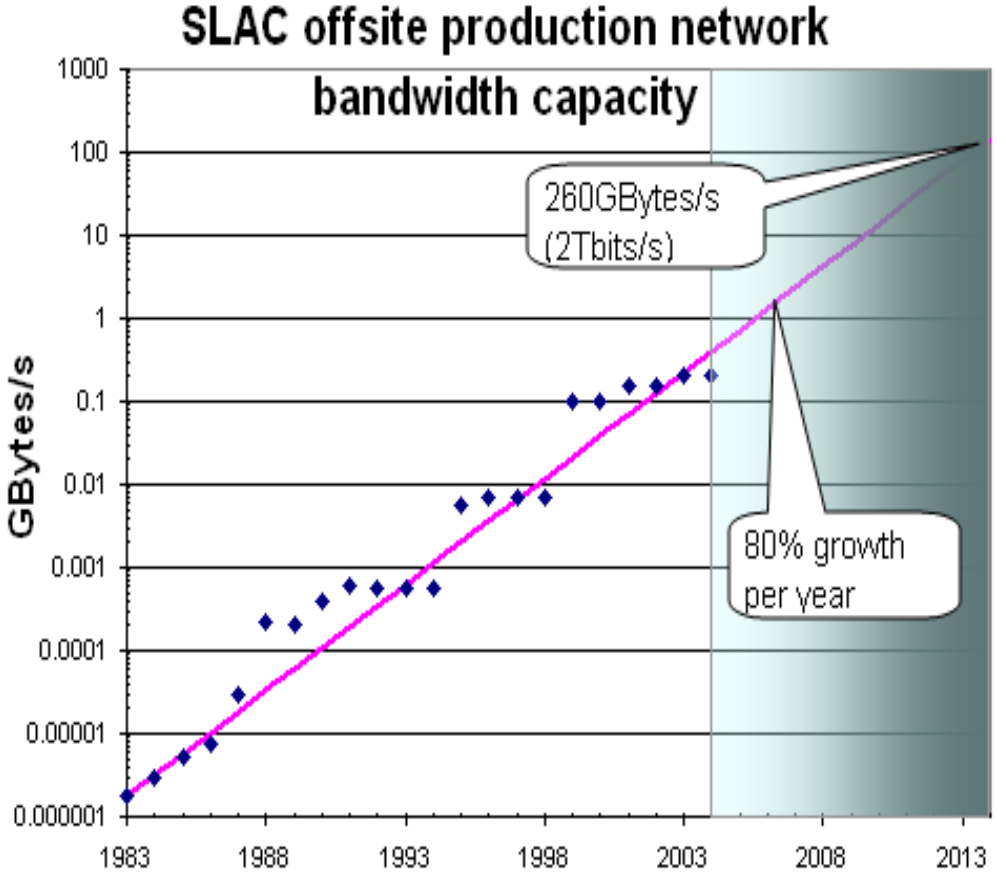
Int'l Networks BW on Major Links for HENP: US-CERN Example

- *Rate of Progress >> Moore's Law (US-CERN Example)*
  - *9.6 kbps Analog (1985)*
  - *64-256 kbps Digital (1989 - 1994) [X 7 – 27]*
  - *1.5 Mbps Shared (1990-3; IBM) [X 160]*
  - *2 -4 Mbps (1996-1998) [X 200-400]*
  - *12-20 Mbps (1999-2000) [X 1.2k-2k]*
  - *155-310 Mbps (2001-2) [X 16k – 32k]*
  - *622 Mbps (2002-3) [X 65k]*
  - *2.5 Gbps λ (2003-4) [X 250k]*
  - *10 Gbps λ (2005) [X 1M]*
  - *4x10 Gbps or 40 Gbps (2007-8) [X 4M]*
- *A factor of ~1M Bandwidth Improvement over 1985-2005 (a factor of ~5k during 1995-2005)*
- *A prime enabler of major HENP programs*
- *HENP has become a leading applications driver, and also a co-developer of global networks*

History of Bandwidth Usage – One Large Network; One Large Research Site

ESnet Accepted Traffic 1/90 – 1/04
Exponential Growth Since '92;
Annual Rate Increased from 1.7 to 2.0X Per Year In the Last 5 Years

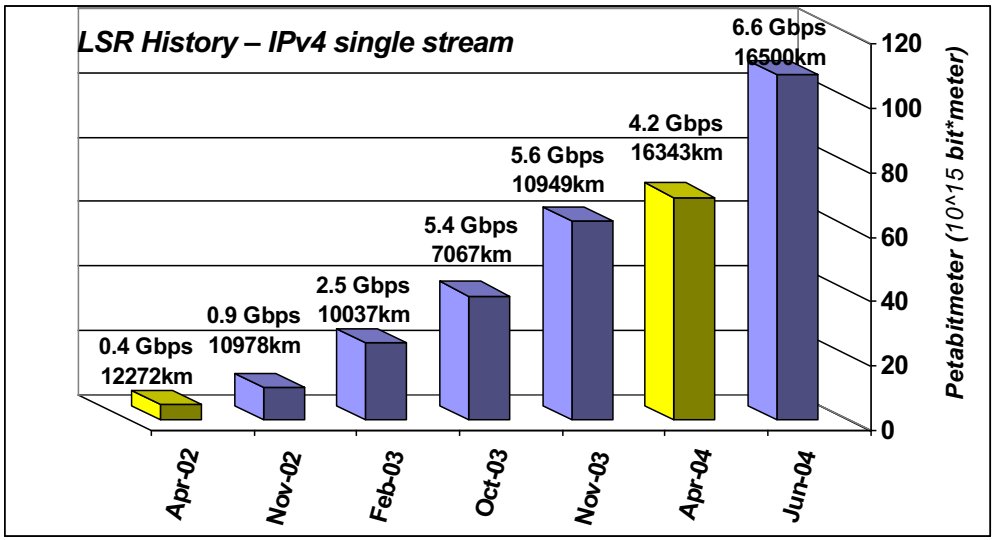




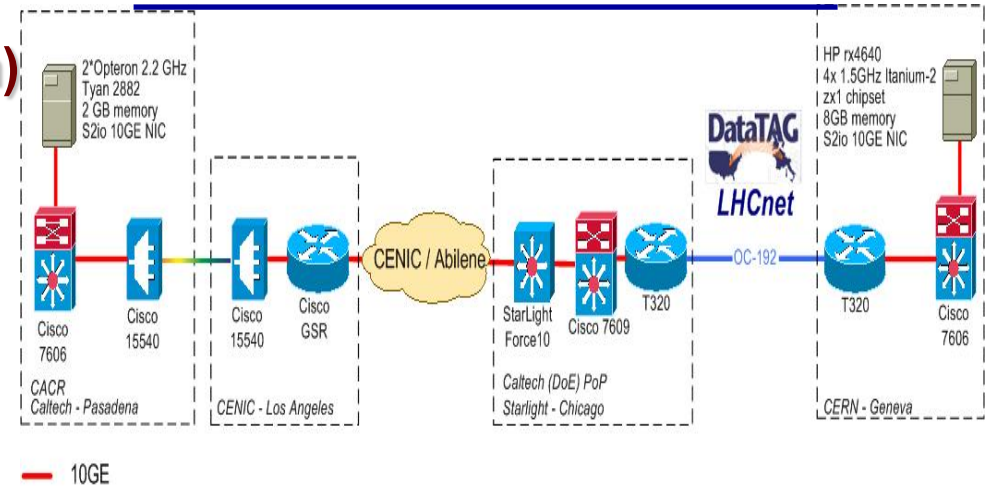
SLAC Traffic ~400 Mbps; Growth in Steps (ESNet Limit): ~ 10X/4 Years. Projected: ~2 Terabits/s by ~2014

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



LSR: 6.9 Gbps X 27 kkm 11/08/04

SC04 BW Challenge: 101.1 Gbps

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 (NLR, ESnet, USN; 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.

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 throughput
Climate (Data & Computation)	0.5 Gb/s	160-200 Gb/s	N x 1000 Gb/s	High bulk throughput
SNS NanoScience	Not yet started	1 Gb/s	1000 Gb/s + QoS for Control Channel	Remote control and time critical throughput
Fusion Energy	0.066 Gb/s (500 MB/s burst)	0.198 Gb/s (500MB/ 20 sec. burst)	N x 1000 Gb/s	Time critical throughput
Astrophysics	0.013 Gb/s (1 TByte/week)	N*N multicast	1000 Gb/s	Computat'l steering and collaborations
Genomics Data & Computation	0.091 Gb/s (1 TBy/day)	100s of users	1000 Gb/s + QoS for Control Channel	High throughput and steering

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

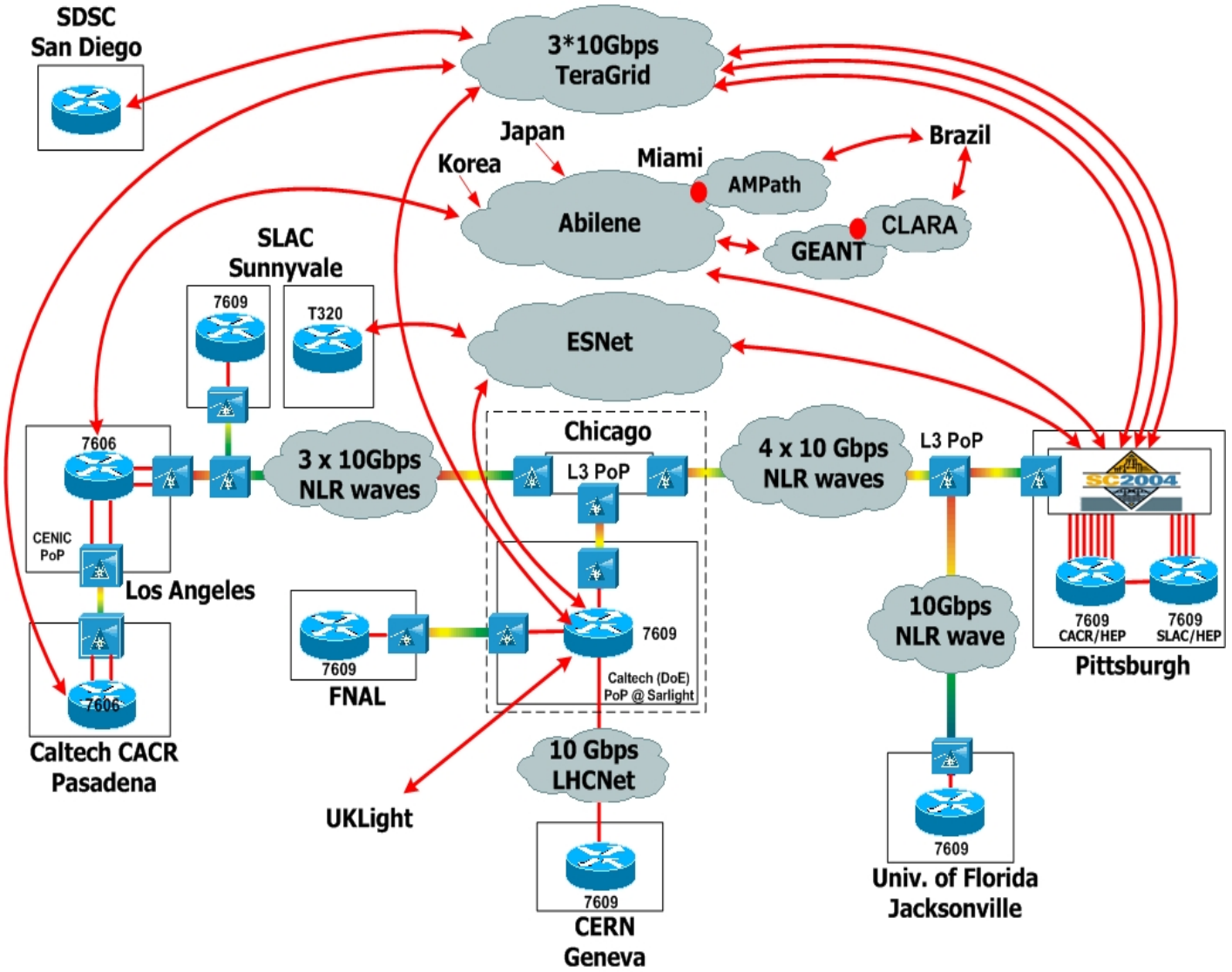
National Lambda Rail (NLR): www.nlr.net

NLR

- Coming Up Now
- Initially 4-8 10G Wavelengths
- Northern Route LA-JAX Now
- Internet2 HOPI Initiative (w/HEP)
- To 40 10G Waves in Future

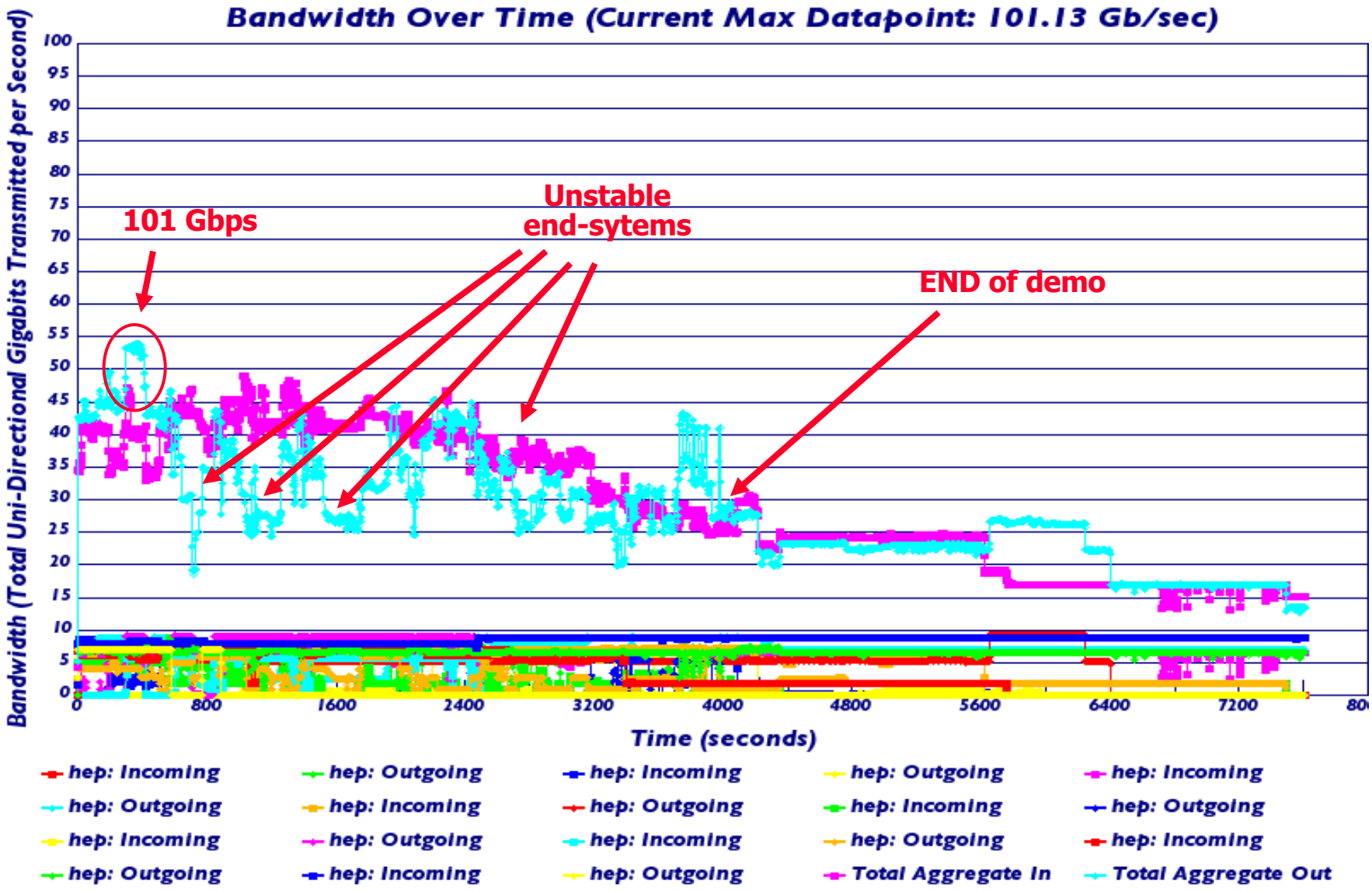
- Initiatives in: nl, ca, pl, cz, uk, kr, jp
- + 25 (up from 18) US States (CA, IL, FL, IN, …)

SC2004: HEP network layout



- *Joint Caltech, FNAL, CERN, SLAC, UF, SDSC, Brazil, Korea ….*
- *Ten 10 Gbps waves to HEP on show floor*
- *Bandwidth challenge: aggregate throughput of 101.13 Gbps achieved*
- *FAST TCP*

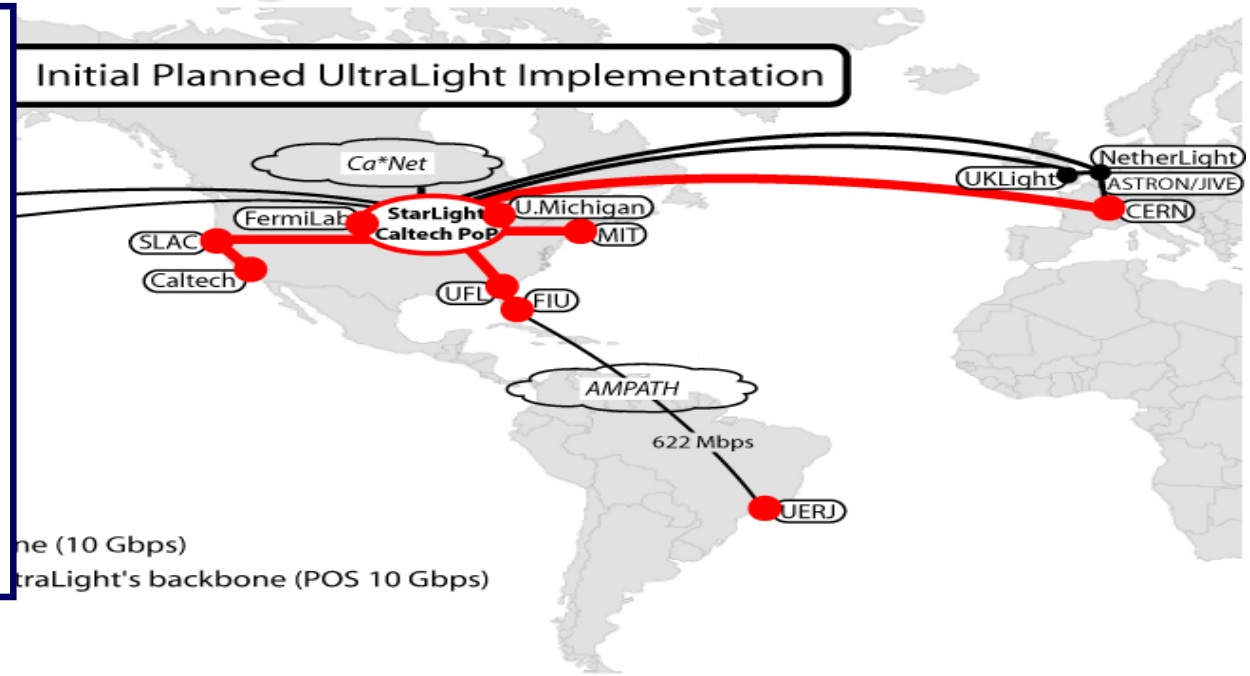
101 Gigabit Per Second Mark



Source: Bandwidth Challenge committee

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

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



- Cisco
- Next generation Information System, with the network as an integrated, actively managed subsystem in a global Grid
- Hybrid network infrastructure: packet-switched + dynamic optical paths
  - 10 GbE across US and the Atlantic: NLR, LHCNet, NetherLight, UKLight, etc.; Extensions to Korea, Brazil, Taiwan
- End-to-end monitoring; Realtime tracking and optimization; Dynamic bandwidth provisioning
- *Agent-based services spanning all layers of the system*

SCIC in 2003-2004
http://cern.ch/icfa-scic

Three 2004 Reports; Presented to ICFA in February

- *Main Report: "Networking for HENP"* [H. Newman et al.]
  - Includes Brief Updates on Monitoring, the Digital Divide and Advanced Technologies [*]
  - A World Network Overview (with 27 Appendices): *Status and Plans for the Next Few Years of National & Regional Networks, and Optical Network Initiatives*
- *Monitoring Working Group Report* [L. Cottrell]
- *Digital Divide in Russia* [V. Ilyin]

August 2004 Update Reports at the SCIC Web Site:

See http://icfa-scic.web.cern.ch/ICFA-SCIC/documents.htm

- Asia Pacific, Latin America, GLORIAD (US-Ru-Ko-China); Brazil, Korea, ESNet, etc.

ICFA Report: Networks for HENP General Conclusions

- Reliable high End-to-end Performance of networked applications such as Data Grids is required. Achieving this requires:
  - A coherent approach to End-to-end monitoring extending to all regions that allows physicists throughout the world to extract clear information
  - Upgrading campus infrastructures.
    To support Gbps flows to HEP centers. One reason for under-utilization of national and Int'l backbones, is the lack of bandwidth to end-user groups in the campus
  - Removing local, last mile, and nat'l and int'l bottlenecks end-to-end, whether technical or political in origin.
    The bandwidths across borders,
    the countryside or the city may be much less.

Problem is very widespread in our community, with examples stretching from the Asia Pacific to Latin America to the Northeastern U.S. Root causes for this vary, from lack of local infrastructure to unfavorable pricing policies.

SCIC Main Conclusion for 2003 Setting the Tone for 2004

- *The disparity among regions in HENP could increase even more sharply, as we learn to use advanced networks effectively, and we develop dynamic Grid systems in the "most favored" regions*
- *We must take action, and work to Close the Digital Divide*
  - *To make Physicists from All World Regions Full Partners in Their Experiments; and in the Process of Discovery*
  - *This is essential for the health of our global experimental collaborations, our plans for future projects, and our field.*

PingER: World View from SLAC

STANFORD LINEAR ACCELERATOR CENTER

IEPM Internet End-to-end Performance Monitoring

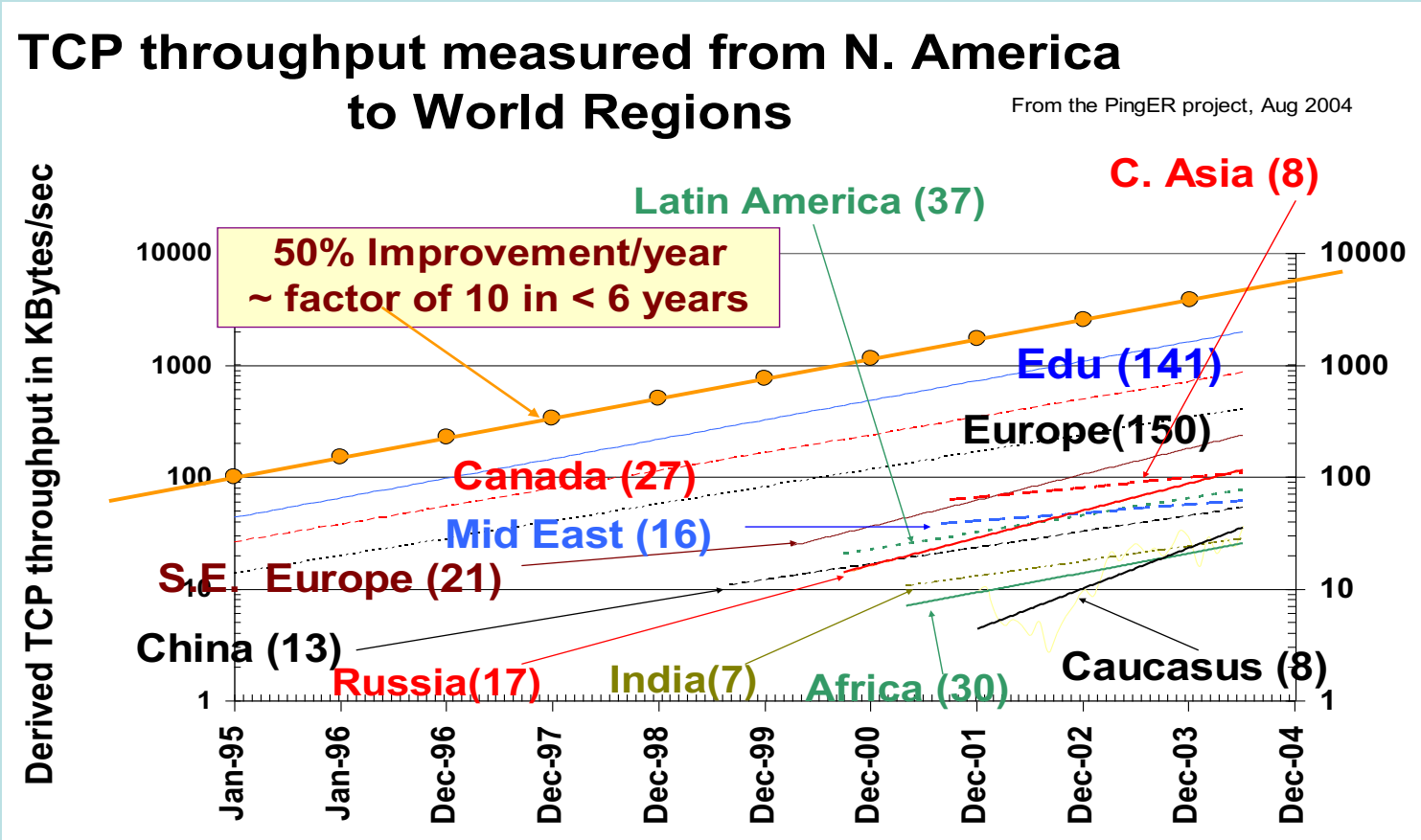
S.E. Europe, Russia: Catching up
Latin Am., Mid East, China: Keeping up
India, Africa: Falling Behind

C. Asia, Russia, SE Europe, L. America, M. East, China: 4-5 yrs behind

India, Africa: 7 yrs behind

Important for policy makers

View from CERN Confirms This View



PROGRESS in SE Europe (Sk, Pl, Cz, Hu, …)

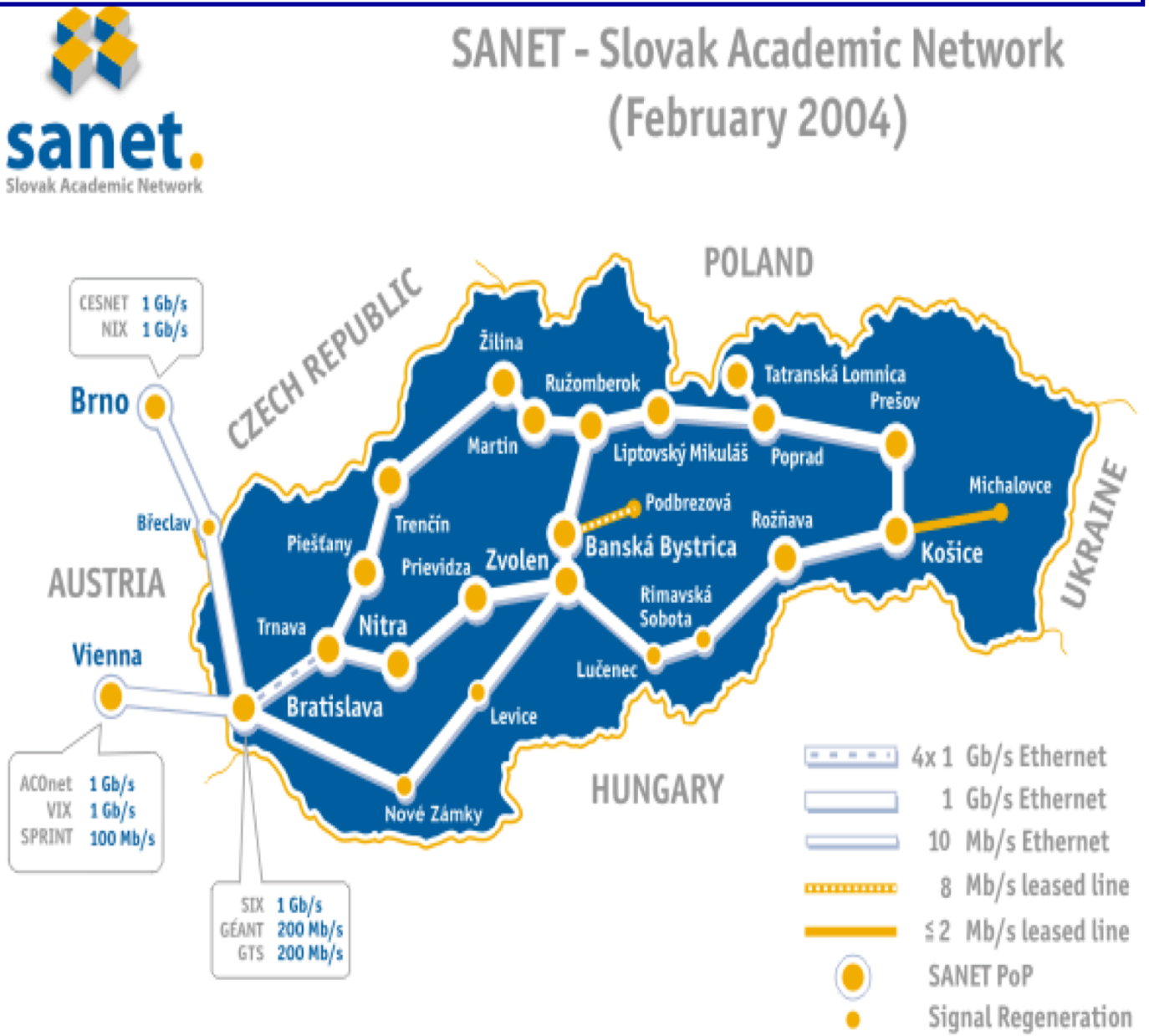
1660 km of Dark Fiber CWDM Links, up to 112 km.

1 to 4 Gbps (GbE)

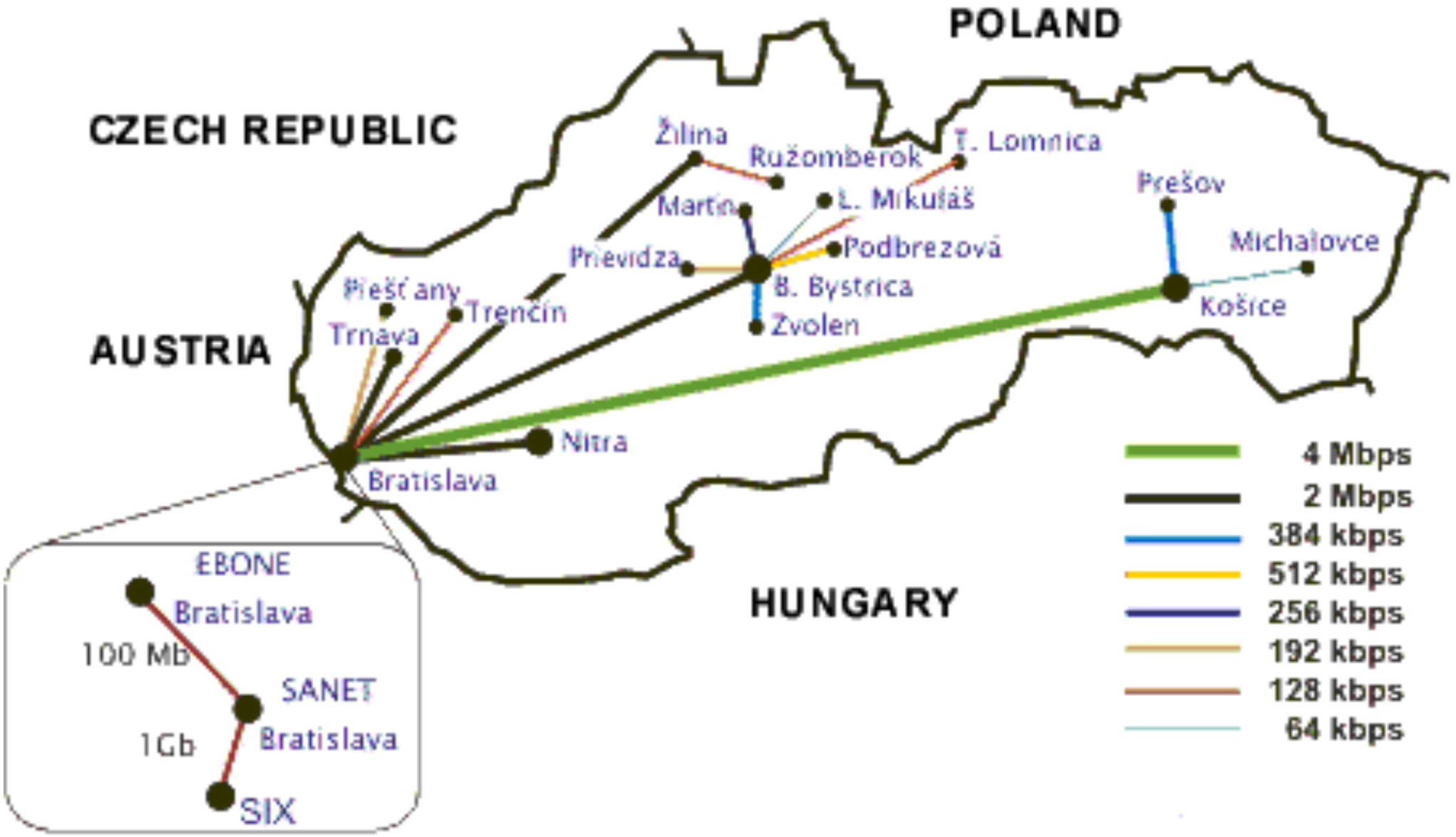
August 2002: First NREN in Europe to establish Int'l GbE Dark Fiber Link, to Austria

April 2003 to Czech Republic.

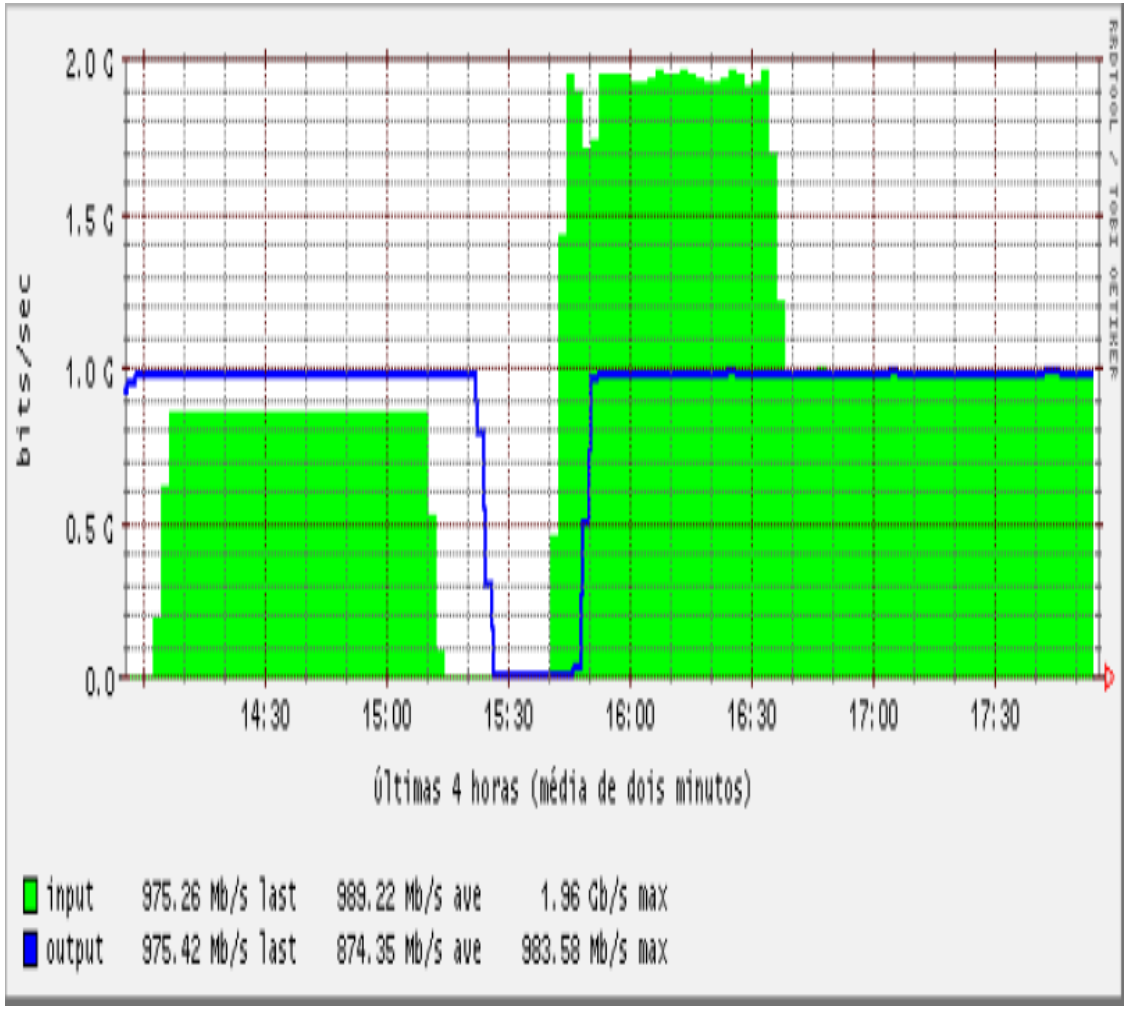
Planning 10 Gbps Backbone; dark fiber link to Poland

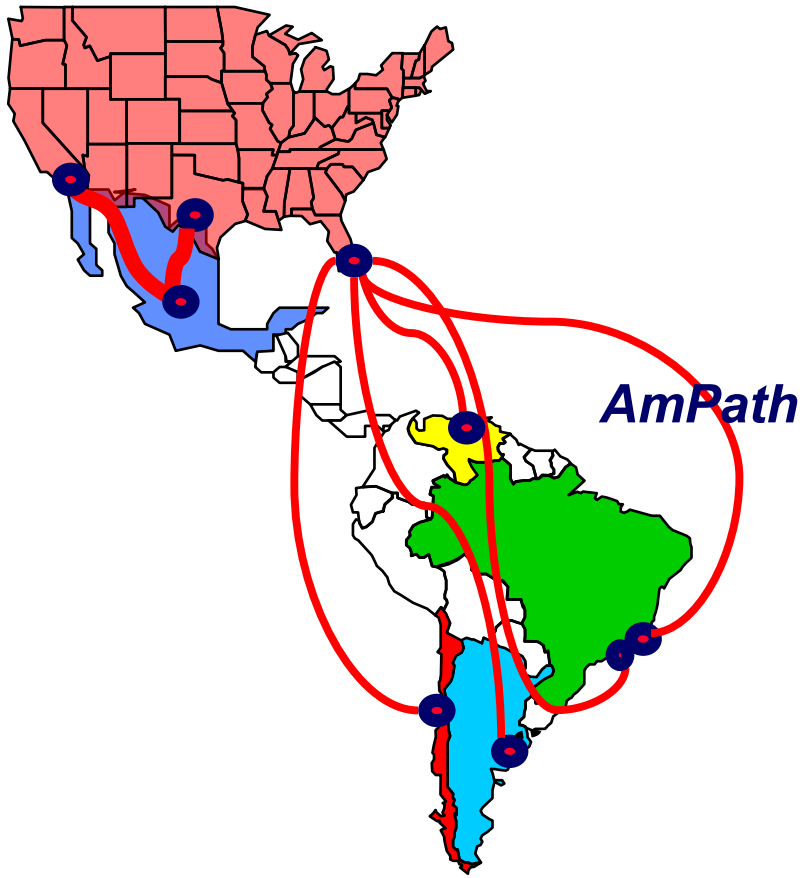


SANET - Slovak Akademic Data Network
(January 2002)



CHEPREO Link at SC04: 2.9 (1.95 + 0.98) Gbps
Sao Paulo – Miami – Pittsburgh (Via Abilene)





Brazilian HEPGrid: Rio de Janeiro, Sao Paolo; Extend Across Latin Am.

HEPGRID and Digital Divide Workshop
UERJ, Rio de Janeiro, Feb. 16-20 2004

HEPGRID AND DIGITAL DIVIDE WORKSHOP

NEWS:

Bulletin: ONE TWO
WELCOME BULLETIN
General Information
Registration
Travel Information
Hotel Registration

Tutorials

- C++
- Grid Technologies
- Grid-Enabled Analysis
- Networks
- Collaborative Systems

Theme: Global Collaborations, Grids and Their Relationship to the Digital Divide

For the past three years the SCIC has focused on understanding and seeking the means of reducing or eliminating the Digital Divide, and proposed to ICFA that these issues, as they affect our field of High Energy Physics, be brought to our community for discussion. This led to ICFA's approval, in July 2003, of the 1st Digital Divide and HEP Grid Workshop.

More Information:

http://www.lishep.uerj.br

SPONSORS

CLAF CNPQ FAPERJ UERJ

Sessions & Tutorials Available (w/Video) on the Web

CERNET2 and Key Technologies (J. Wu)

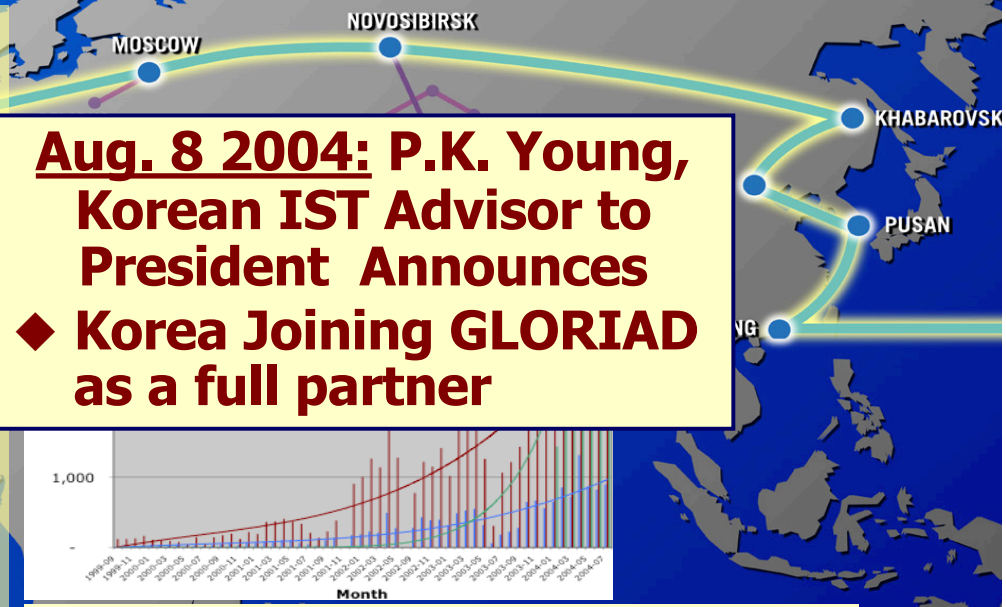
- CERNET 2: Next Generation Education and Research Network in China
- CERNET 2 Backbone connecting 20 GigaPOPs at 2.5G-10Gbps
- Connecting 200 Universities and 100+ Research Institutes at 1Gbps-10Gbps
- Native IPv6 and Lambda Networking
- Support/Deployment of:
  - E2E performance monitoring
  - Middleware and Advanced Applications
  - Multicast

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**
- **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-Busan-Beijing early 2005; 10G ring around northern hemisphere 2007 (or earlier); Multi-wavelength hybrid service from ~2008

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

- **Workshop Goals**
  - ➡ 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.

Role of Science in the Information Society; WSIS 2003-2005

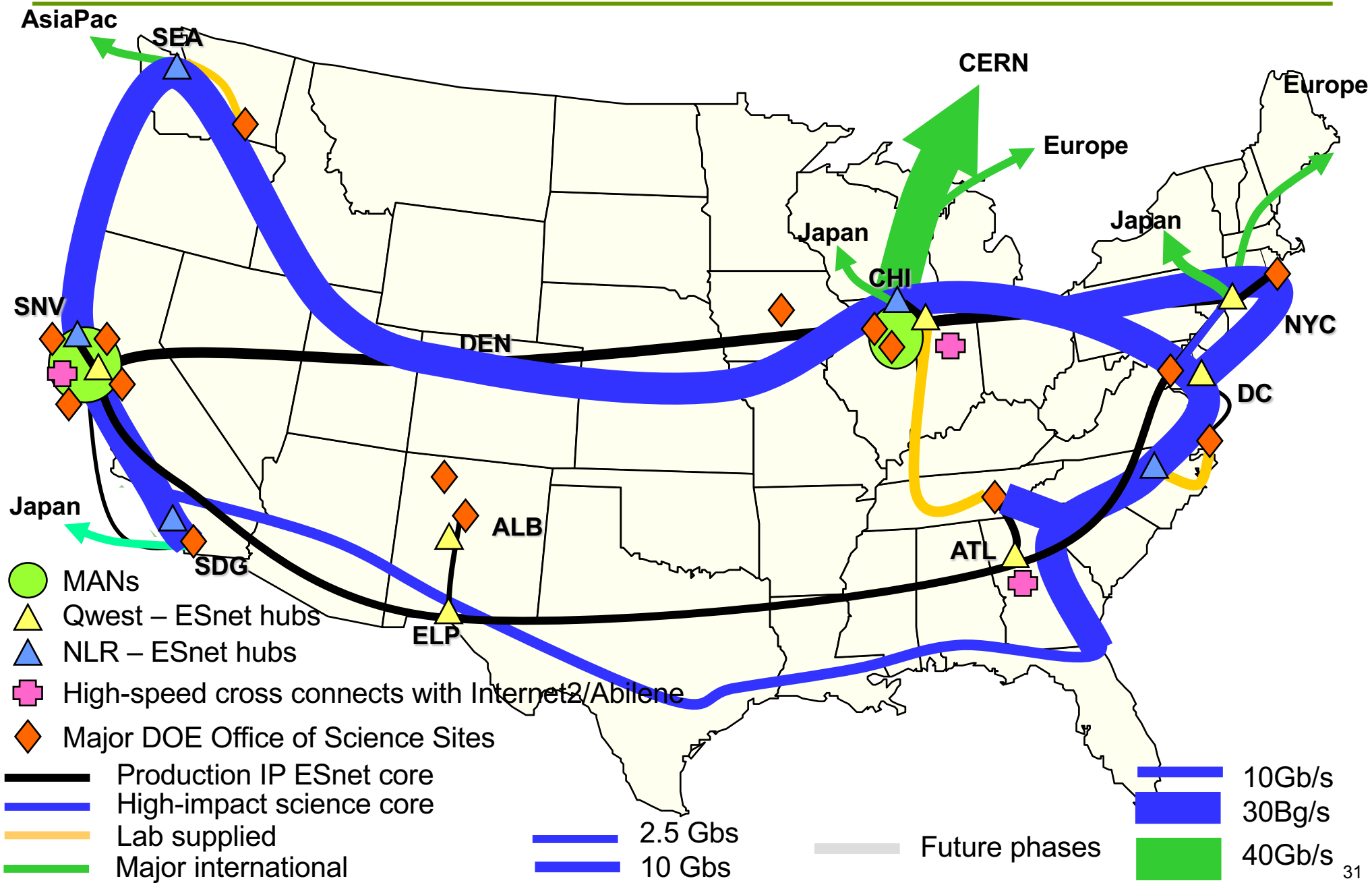
- **HENP Active in WSIS**
  - CERN RSIS Event
  - SIS Forum & CERN/Caltech Online Stand at WSIS I (> 50 Demos; Geneva 12/03)
- **Visitors at WSIS I**
  - Kofi Annan, UN Sec'y General
  - John H. Marburger, Science Adviser to US President
  - Ion Iliescu, President of Romania; and Dan Nica, Minister of ICT
  - Jean-Paul Hubert, Ambassador of Canada in Switzerland
  - …
- **Planning Underway for WSIS II: Tunis 2005**

Networks and Grids for HENP and Global Science

- Networks used by HENP and other fields are advancing rapidly
  - To the 10 G range and now N X 10G; much faster than Moore's Law
  - New HENP and DOE Roadmaps: a factor ~1000 BW Growth/Decade
- We are learning to use long distance 10 Gbps networks effectively
  - 2004 Developments: 7+ Gbps TCP flows over 27 kkm; 101 Gbps Record
- Transition to community-operated optical R&E networks (us, ca, nl, pl, cz, sk, kr, jp …); Emergence of a new generation of "hybrid" optical networks
- *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 Path*
- *Important Examples* on the Road to Progress in Closing the Digital Divide
  - CHINA CNGI Program: CERNET2, CSTNET
  - AMPATH, CHEPREO, CLARA and the Brazil HEPGrid in Latin America
  - Optical Networking in Central and Southeast Europe
  - GLORIAD (US-Russia-China-Korea)
  - *Leadership and Outreach: HEP Groups in Europe, US, China, Japan, & Korea*

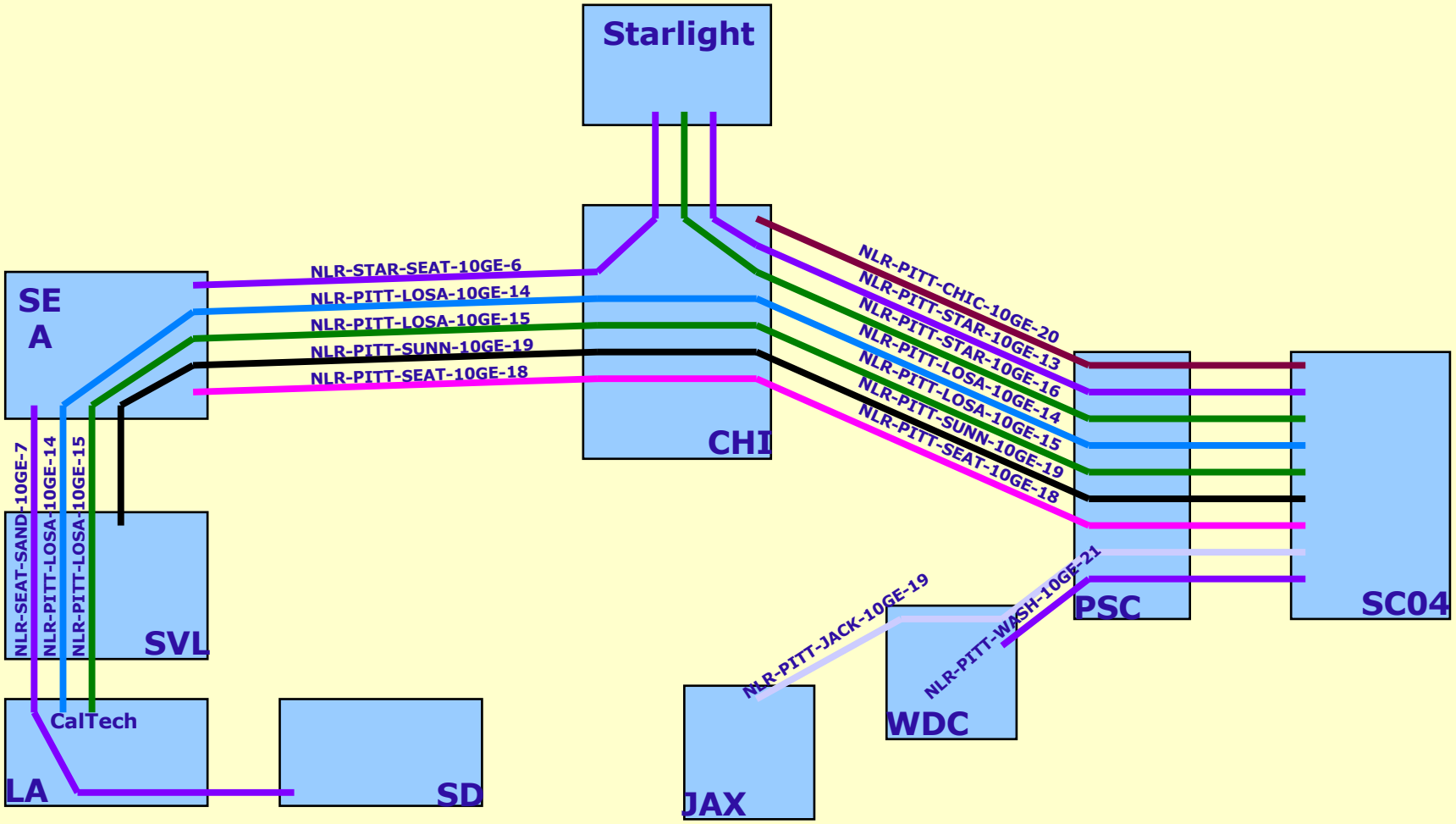
Extra Slides Follow

ESnet Beyond FY07 (W. Johnston)



CalTech/Newman
FL/Avery
SLAC
Optiputer
Ed Seidel
HOPI
UW/Rsrch Chnl

21 NLR Waves: 9 to SC04



All lines 10GE