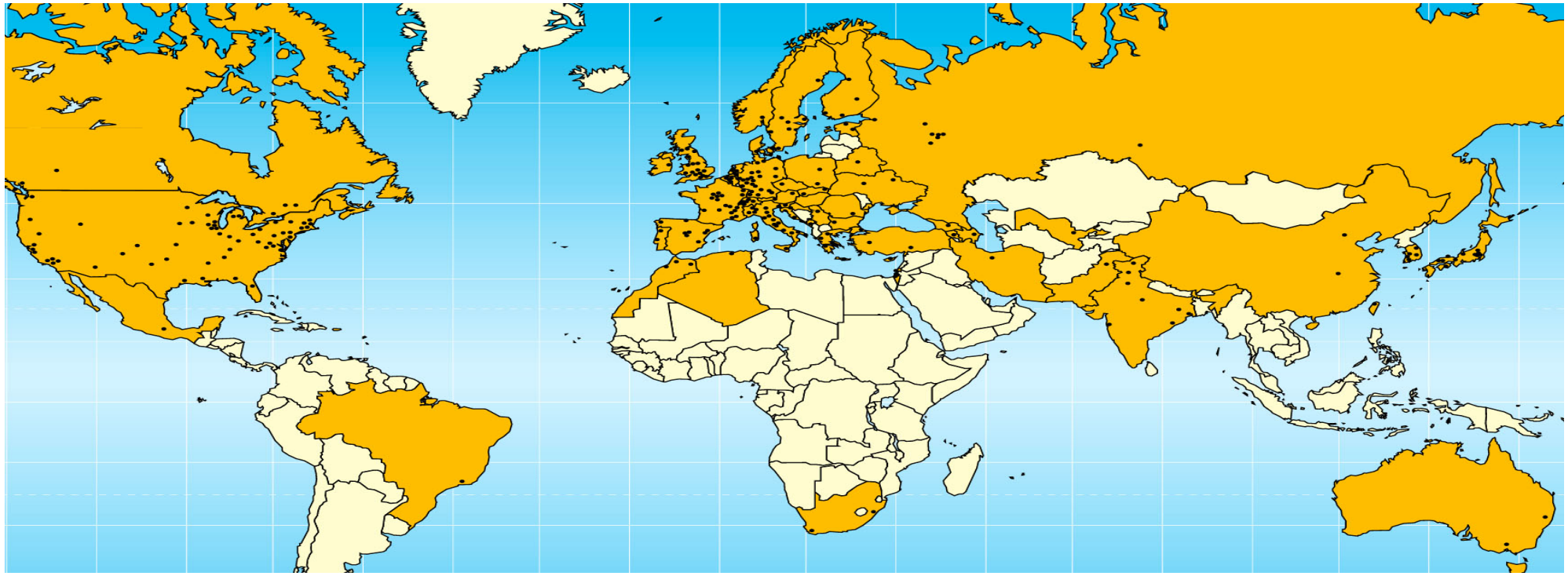
HENP Networks, ICFA SCIC and the Digital Divide



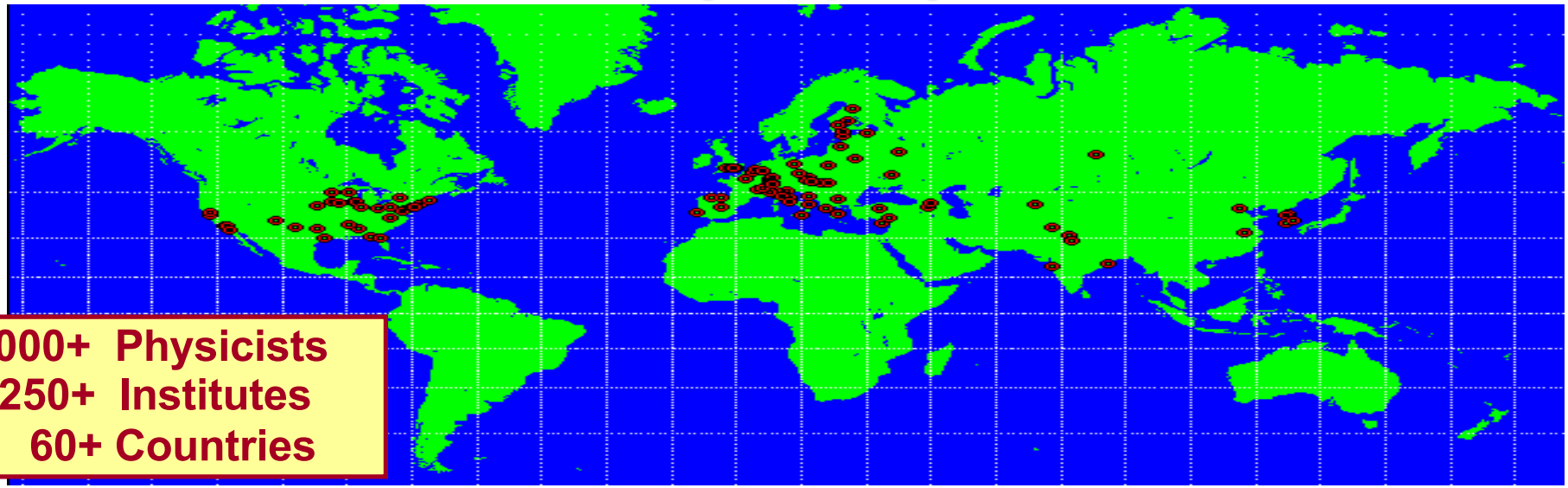
Harvey B. Newman

California Institute of Technology

AMPATH Workshop, FIU
January 31, 2003

Computing Challenges: Petabyes, Petaflops, Global VOs

- **Geographical dispersion:** of people and resources
- **Complexity:** the detector and the LHC environment
- **Scale:** Tens of Petabytes per year of data



5000+ Physicists
250+ Institutes
60+ Countries

Major challenges associated with:

Communication and collaboration at a distance

Managing globally distributed computing & data resources

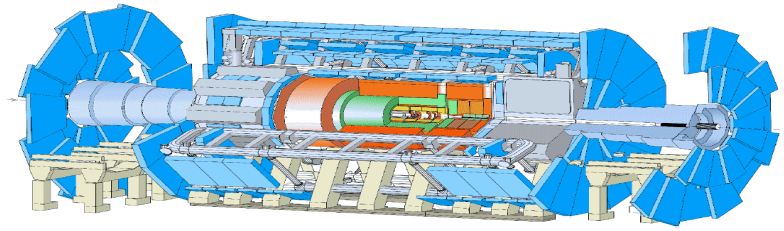
Cooperative software development and physics analysis

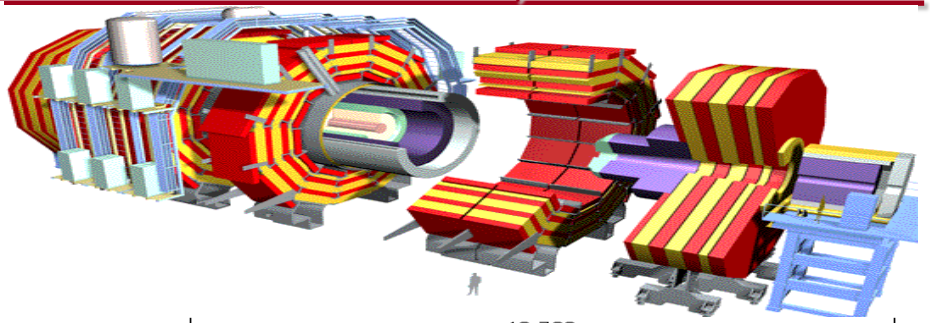
New Forms of Distributed Systems: Data Grids

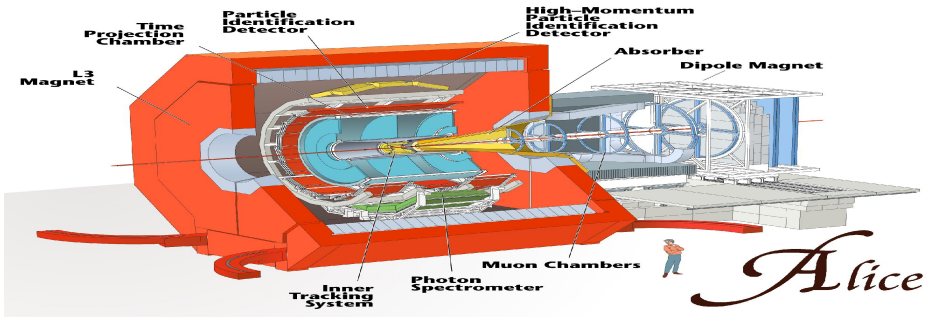
Four LHC Experiments: The Petabyte to Exabyte Challenge

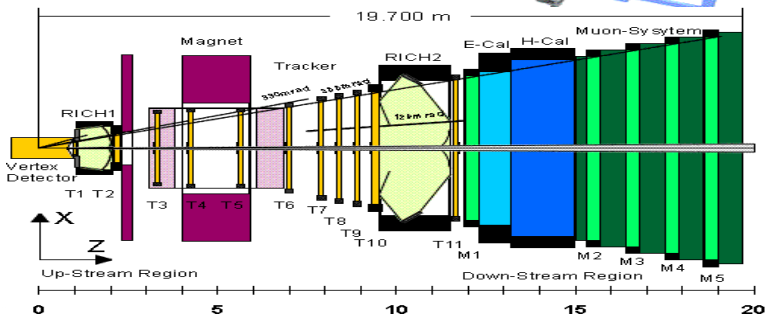
ATLAS, CMS, ALICE, LHCB

Higgs + New particles; Quark-Gluon Plasma; CP Violation









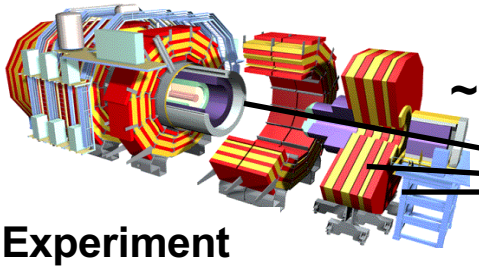
Data stored ~40 Petabytes/Year and UP;

CPU 0.30 Petaflops and UP

0.1 to 1 Exabyte (1 EB = 10^{18} Bytes)

(2007) (~2012 ?) for the LHC Experiments

LHC Data Grid Hierarchy



Experiment

~PBytes/sec

Online System

~100-400 MBytes/sec

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

Tier 0 +1

CERN 700k SI95
~1 PB Disk;
Tape Robot

HPSS

Tier 1

~2.5 Gbps

IN2P3 Center

RAL Center

INFN Center

FNAL

2.5 Gbps

Tier 2

~2.5 Gbps

Tier2 Center

Tier 3

Institute ~0.25TIPS

Physics data cache

0.1 to 10 Gbps

Workstations

Tier 4

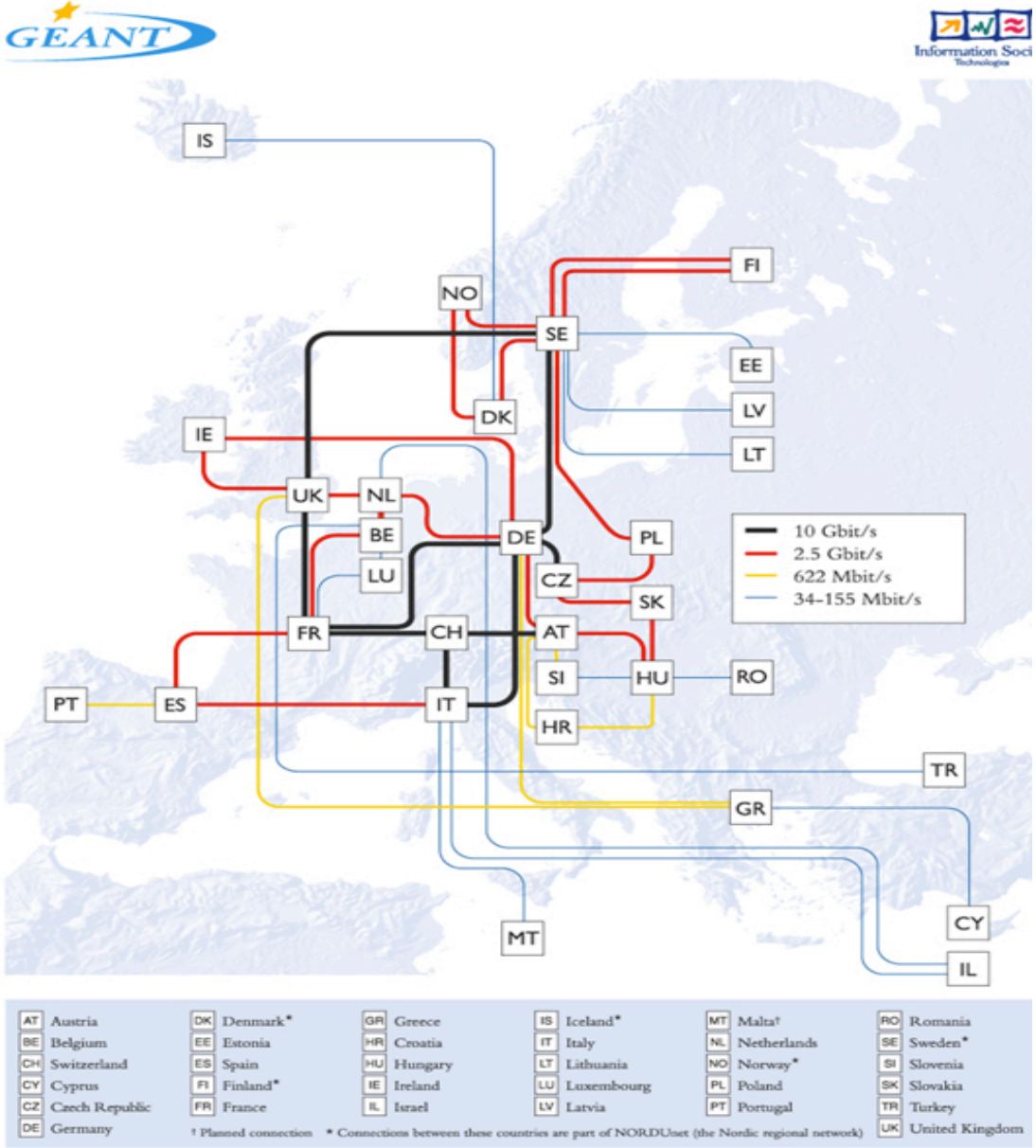
Tens of Petabytes by 2007-8.
An Exabyte within ~5 Years later.

Transatlantic Net WG (HN, L. Price) Bandwidth Requirements [*]

	2001	2002	2003	2004	2005	2006
CMS	100	200	300	600	800	2500
ATLAS	50	100	300	600	800	2500
BaBar	300	600	1100	1600	2300	3000
CDF	100	300	400	2000	3000	6000
D0	400	1600	2400	3200	6400	8000
BTeV	20	40	100	200	300	500
DESY	100	180	210	240	270	300
CERN BW	155-310	622	2500	5000	10000	20000

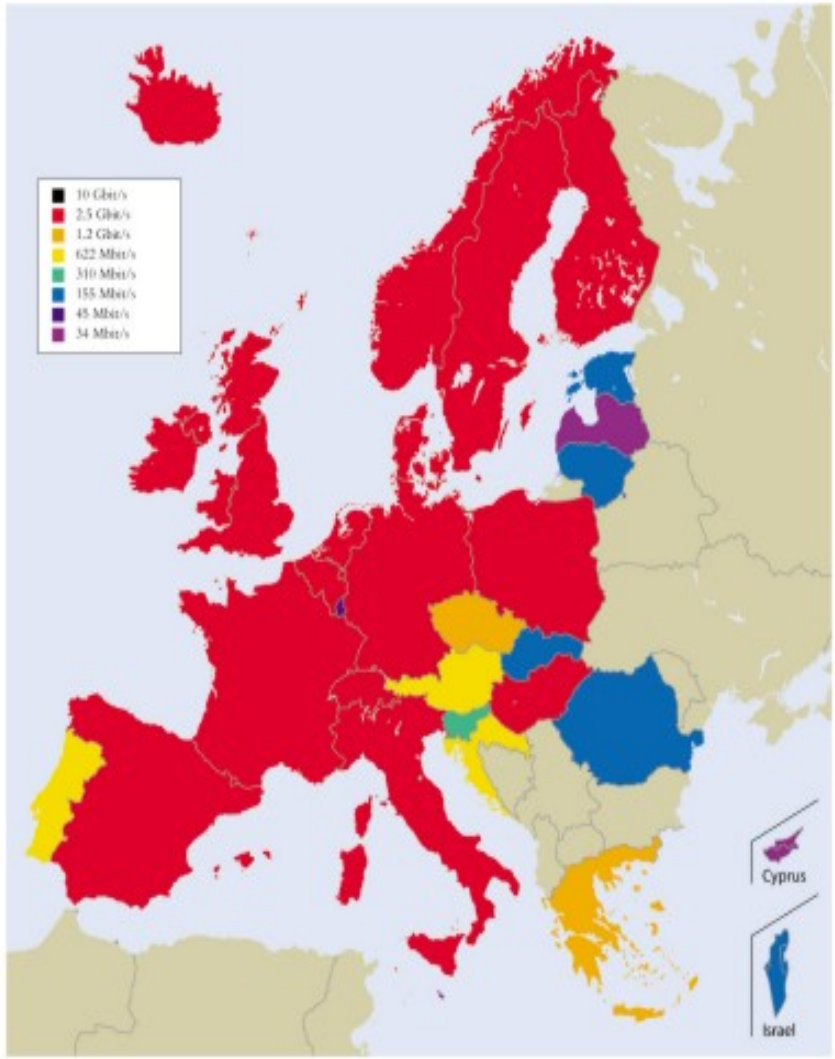
[*] BW Requirements Increasing Faster Than Moore's Law

See http://gate.hep.anl.gov/lprice/TAN



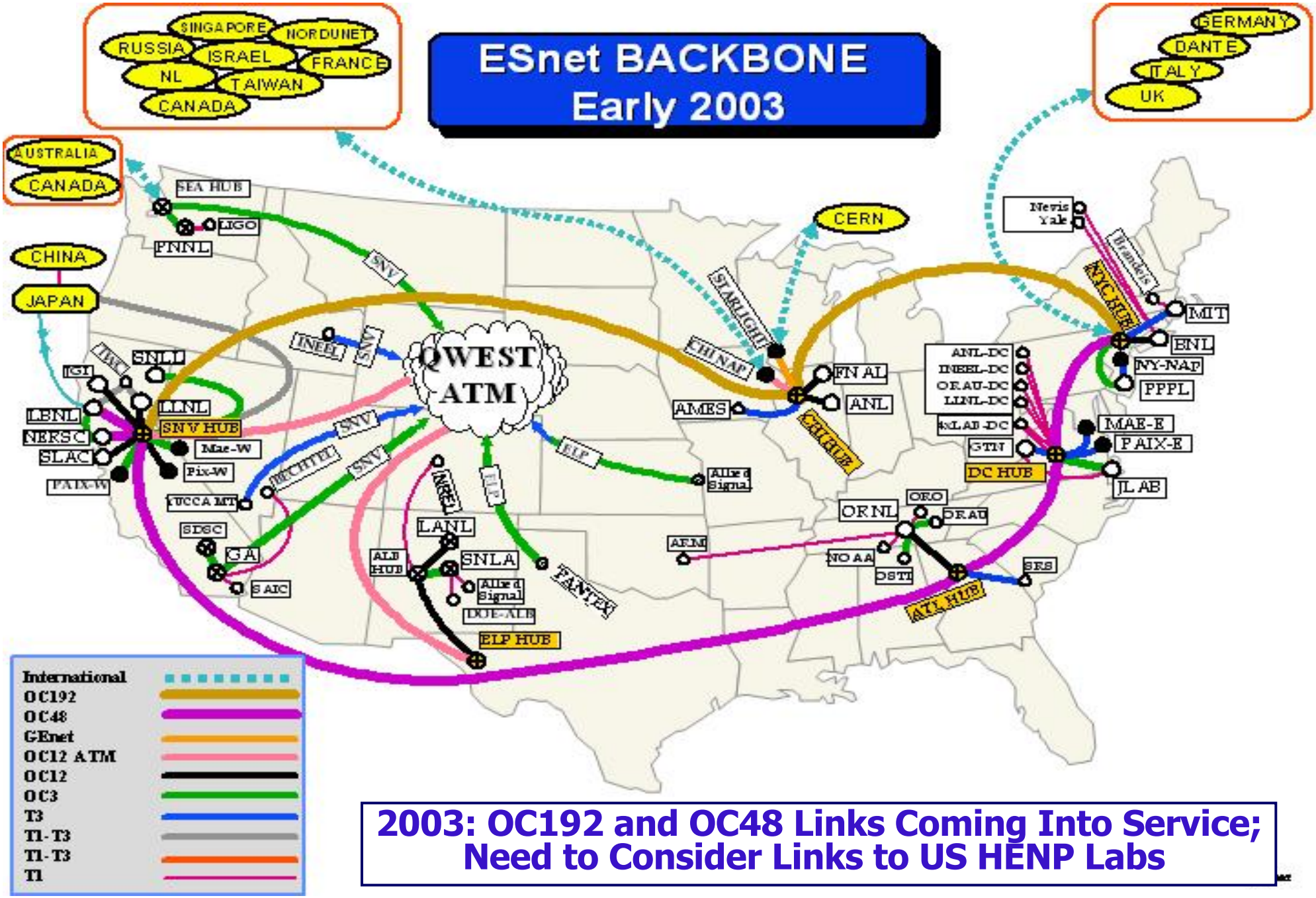
Multi-Gigabit pan-European Research Network
Backbone Topology December 2002

DANTE
www.dante.n



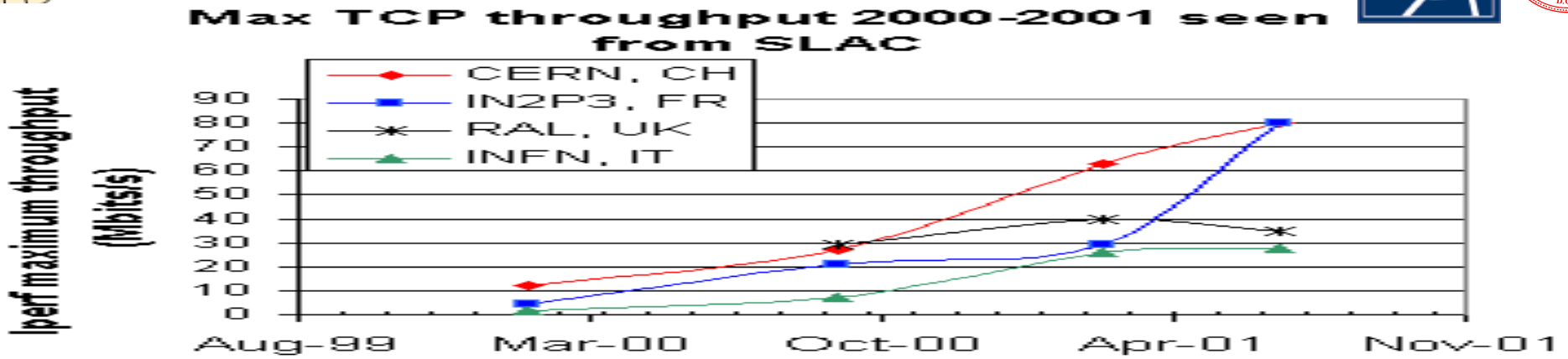
Multi-Gigabit pan-European Research Network
Backbone Access Speeds August 2002

DANTE
www.dante.net



2003: OC192 and OC48 Links Coming Into Service; Need to Consider Links to US HENP Labs

Progress: Max. Sustained TCP Thruput on Transatlantic and US Links



- 8-9/01 105 Mbps 30 Streams: SLAC-IN2P3; 102 Mbps 1 Stream CIT-CERN
- 11/5/01 125 Mbps in One Stream (modified kernel): CIT-CERN
- 1/09/02 190 Mbps for One stream shared on 2 155 Mbps links
- 3/11/02 120 Mbps Disk-to-Disk with One Stream on 155 Mbps link (Chicago-CERN)
- 5/20/02 450-600 Mbps SLAC-Manchester on OC12 with ~100 Streams
- 6/1/02 290 Mbps Chicago-CERN One Stream on OC12 (mod. Kernel)
- 9/02 850, 1350, 1900 Mbps Chicago-CERN 1,2,3 GbE Streams, OC48 Link
- 11-12/02 FAST: 940 Mbps in 1 Stream SNV-CERN; 9.4 Gbps in 10 Flows SNV-Chicago

Also see http://www-iepm.slac.stanford.edu/monitoring/bulk/; and the Internet2 E2E Initiative: http://www.internet2.edu/e2e

HENP Major Links: Bandwidth Roadmap (Scenario) 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 the Trend: ~1000 Times Bandwidth Growth Per Decade; We are Rapidly Learning to Use and Share Multi-Gbps Networks

ICFA Standing Committee on Interregional Connectivity (SCIC)

- Created by ICFA in July 1998 in Vancouver ; Following ICFA-NTF
- CHARGE:

  Make recommendations to ICFA concerning the connectivity between *the Americas*, Asia and Europe (and network requirements of HENP)

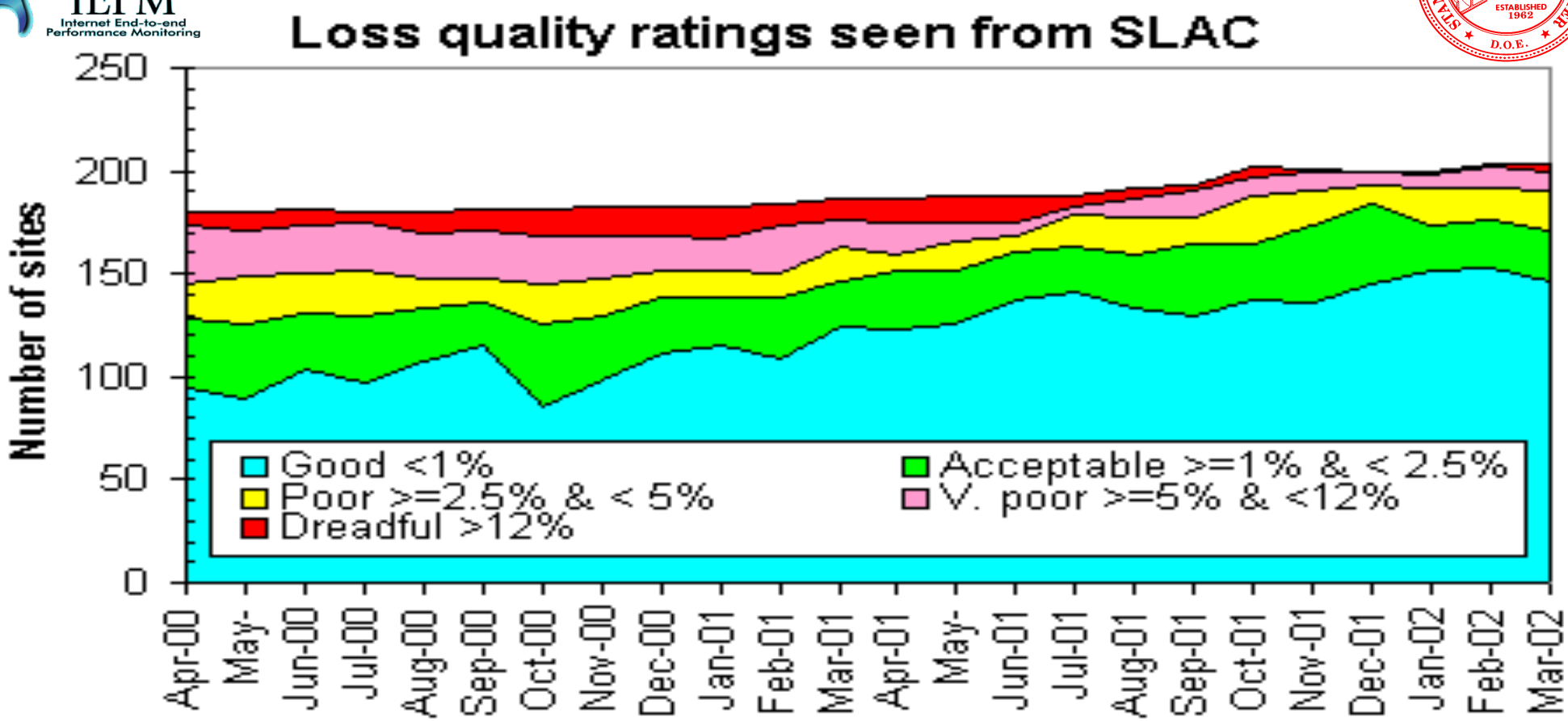
  - As part of the process of developing these recommendations, the committee should
    - Monitor traffic
    - Keep track of technology developments
    - Periodically review forecasts of future bandwidth needs, and
    - Provide early warning of potential problems
- Create subcommittees when necessary to meet the charge
- The chair of the committee should report to ICFA once per year, at its joint meeting with laboratory directors (Feb. 2003)
- Representatives: Major labs, ECFA, ACFA, NA Users, S. America

SCIC Sub-Committees

Web Page http://cern.ch/ICFA-SCIC/

- The Digital Divide: Alberto Santoro (Rio, Brazil)
  - With V. Ilyin (MSU), Y. Karita(KEK), D.O. Williams (CERN)
  - Dongchul Son (Korea), Hafeez Hoorani (Pakistan), Sunanda Banerjee (India), Vicky White (FNAL)
- Monitoring: Les Cottrell (http://www.slac.stanford.edu/xorg/icfa/scic-netmon) With Richard Hughes-Jones (Manchester), Sergio Novaes (Sao Paolo); Sergei Berezhnev (RUHEP), Fukuko Yuasa (KEK), Daniel Davids (CERN), Sylvain Ravot (Caltech), Shawn McKee (Michigan)
- Advanced Technologies: Richard Hughes-Jones, With Vladimir Korenkov (JINR, Dubna), Olivier Martin (CERN), Harvey Newman
- Key Requirements: Harvey Newman
  - Also Charlie Young (SLAC)

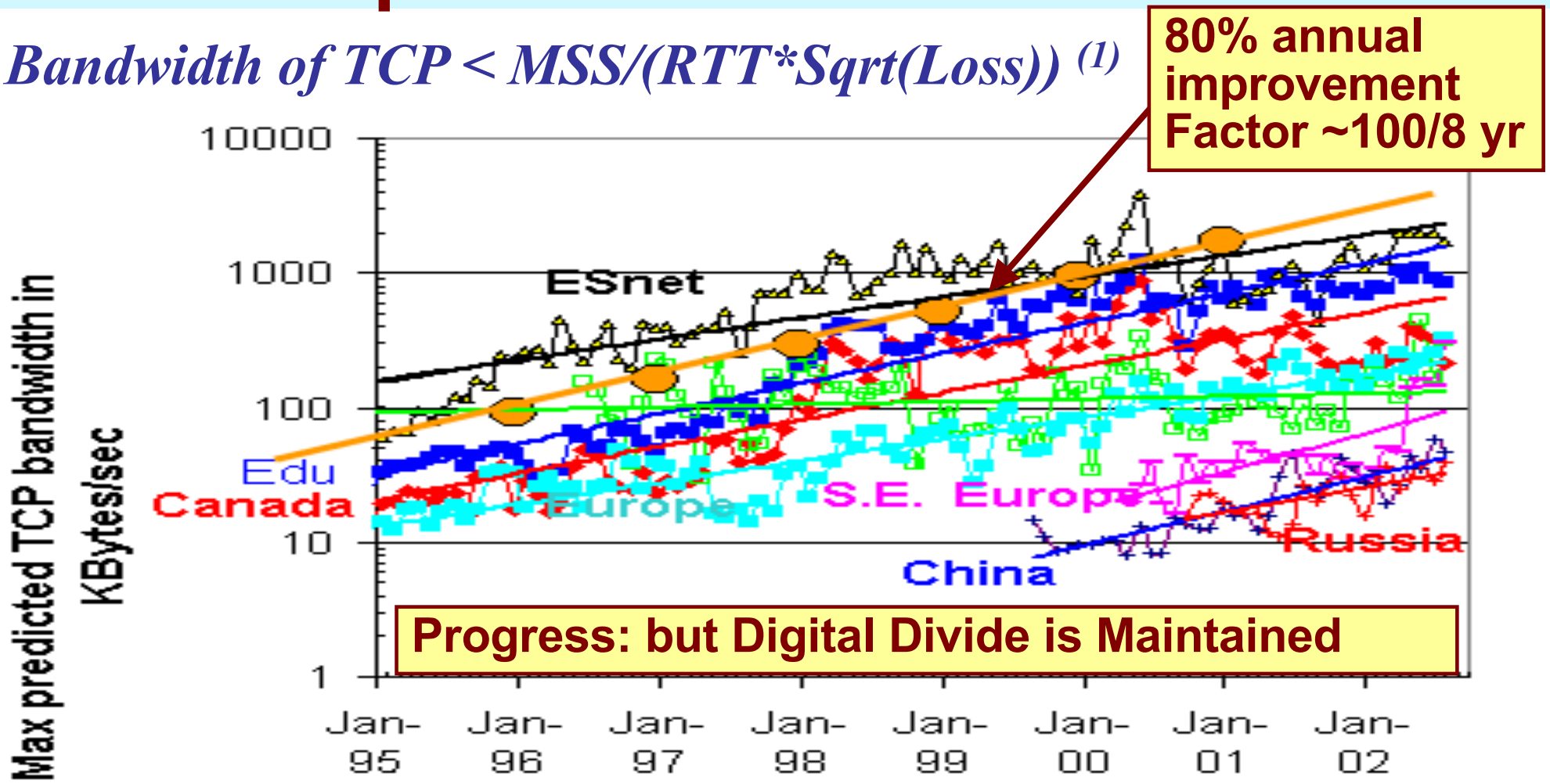
History – Loss Quality (Cottrell)



- Fewer sites have very poor to dreadful performance
- More have good performance (< 1% Loss)

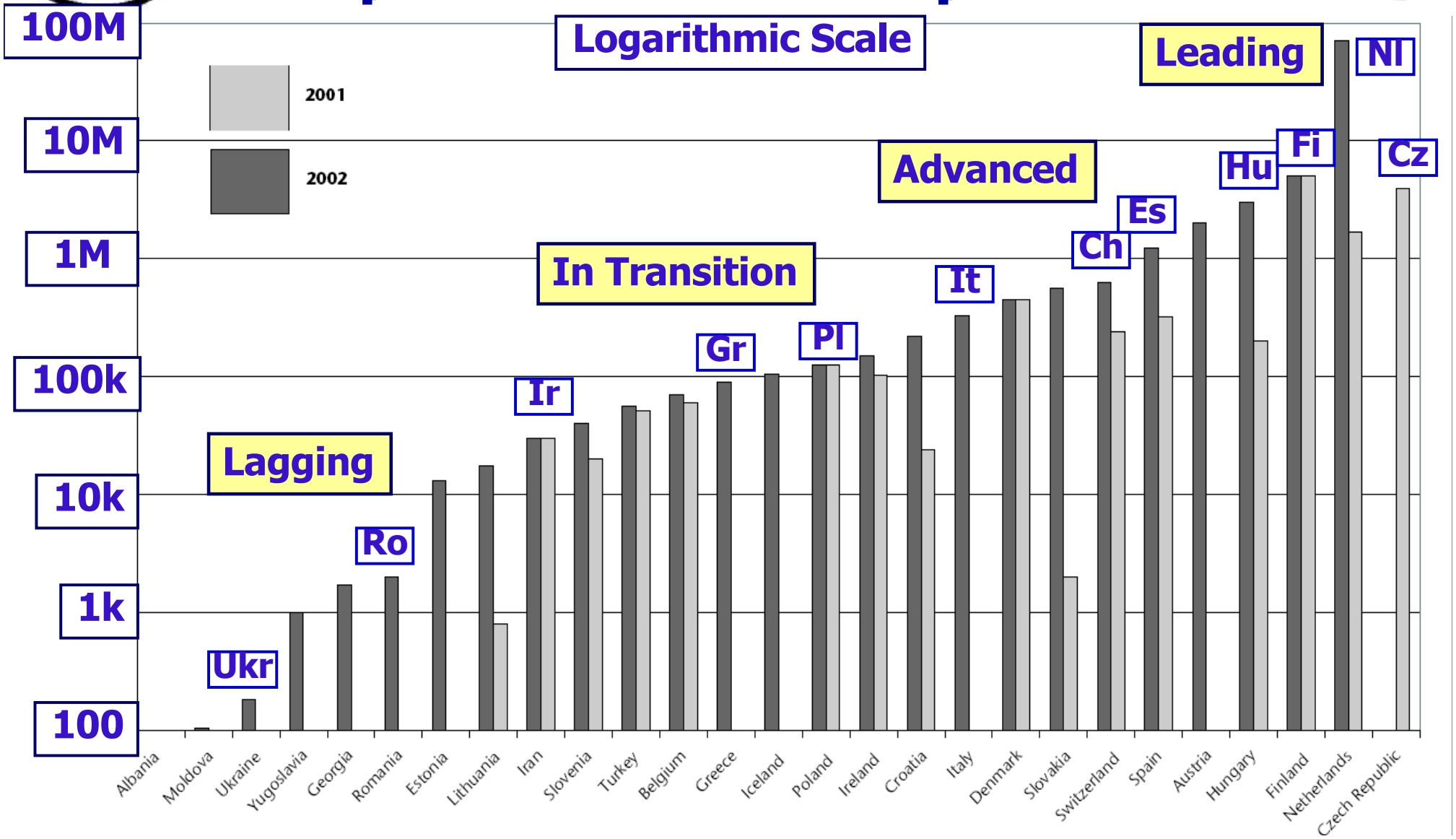
History - Throughput Quality Improvements from US

*Bandwidth of TCP < MSS/(RTT*Sqrt(Loss))* [1]

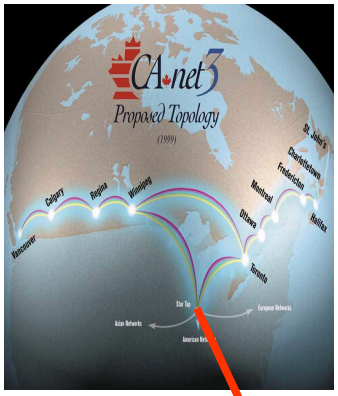


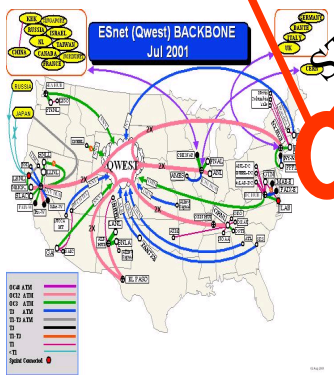
(1) *Macroscopic Behavior of the TCP Congestion Avoidance Algorithm*, Matthis, Semke, Mahdavi, Ott, Computer Communication Review 27(3), July 1997

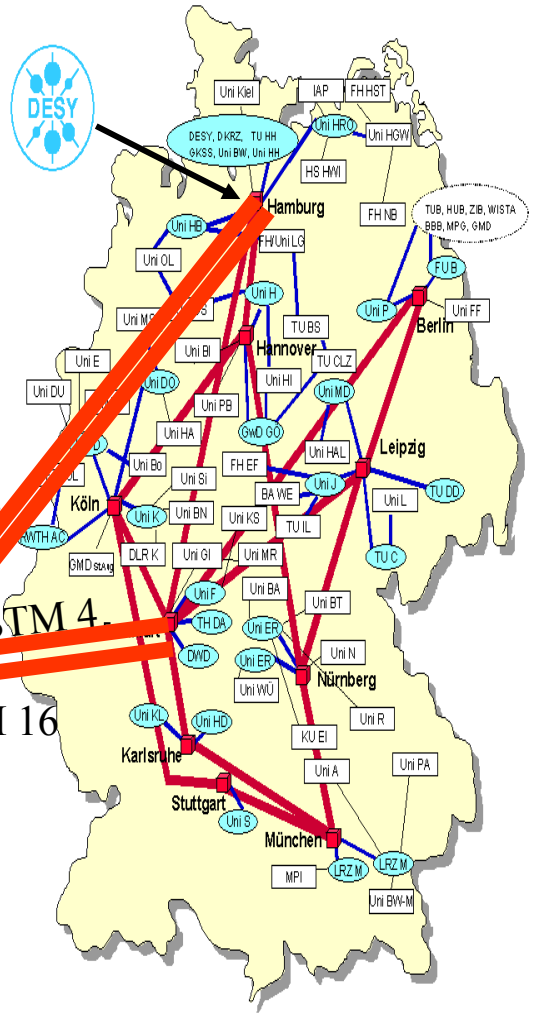
NREN Core Network Size (Mbps-km):
http://www.terena.nl/compendium/2002



National R&E Network Example
Germany: G-Win







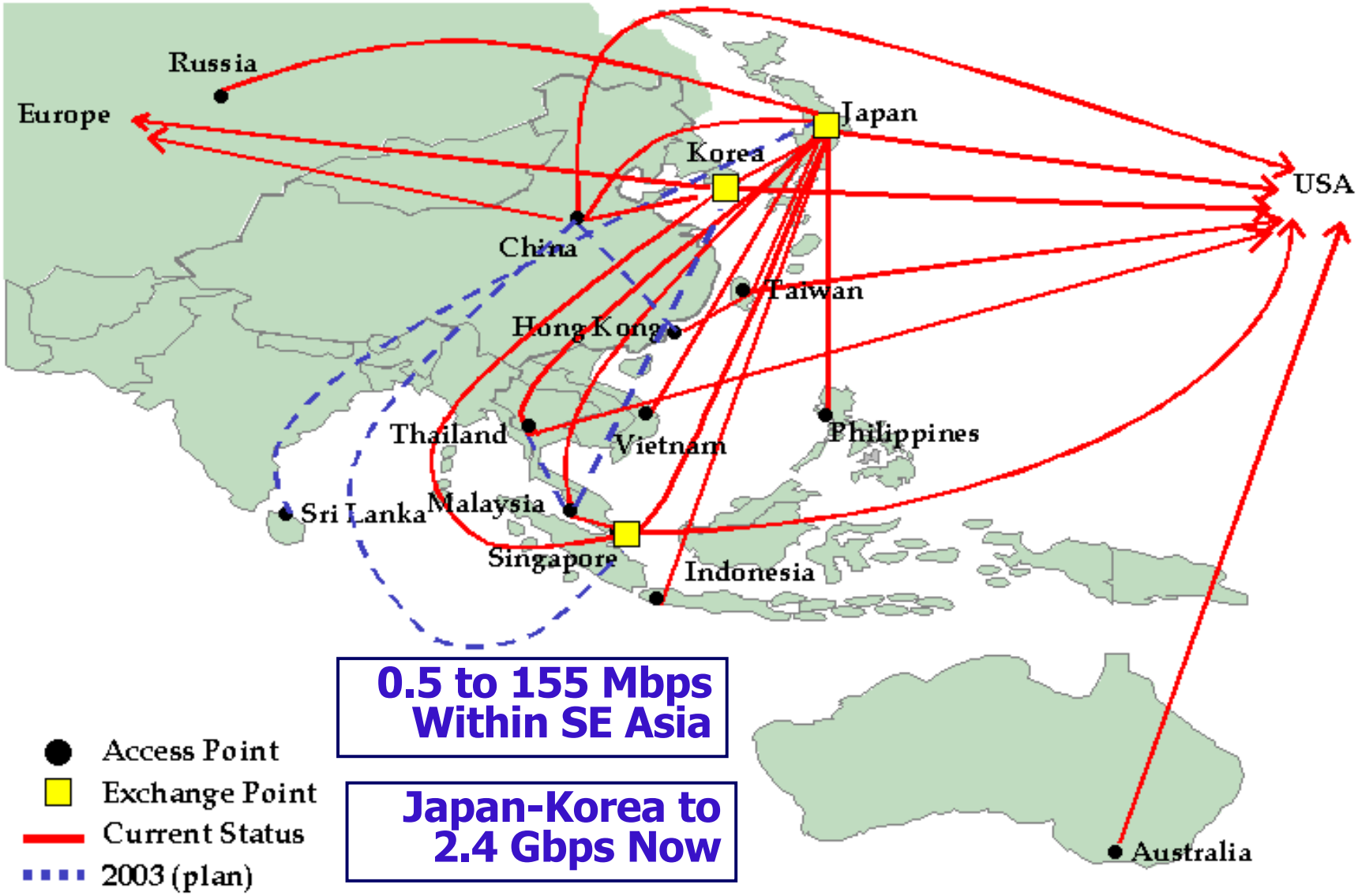
- OC48 Connections to Abilene, ESnet

Virtual SILK Highway Project :
NATO ($ 2.5 M) and Partners ($ 1.1M)

- ➡ Satellite Links to South Caucasus and Central Asia (8 Countries)
- ➡ In 2001-2 (pre-SILK) BW 64-512 kbps
- ➡ Proposed VSAT to get 10-50 X BW for same cost
- ➡ See www.silkproject.org

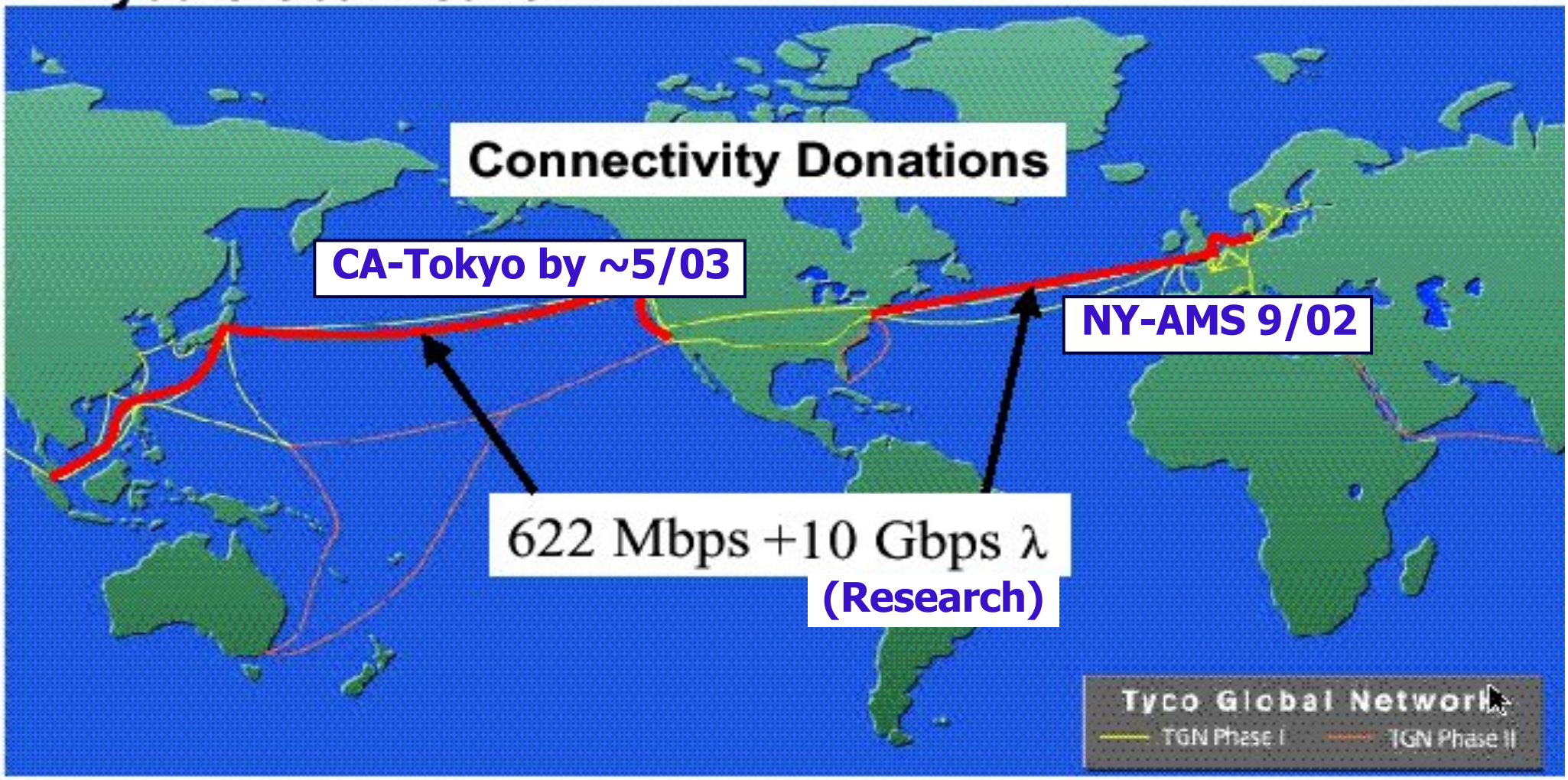
[*] Partners: CISCO, DESY. GEANT, UNDP, US State Dep., Worldbank, UC London, Univ. Groenigen

APAN Links in Southeast Asia
January 2003



Internet Educational Equal Access Foundation

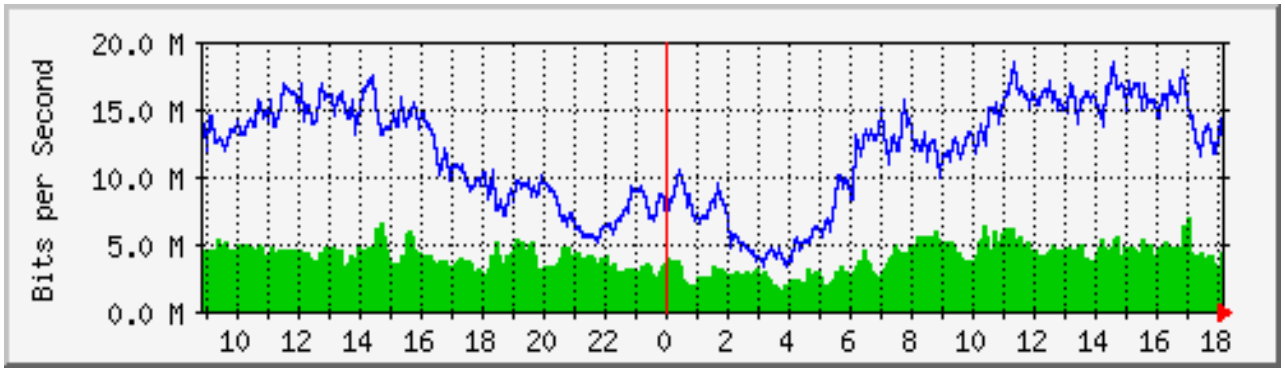
Tyco Global Network

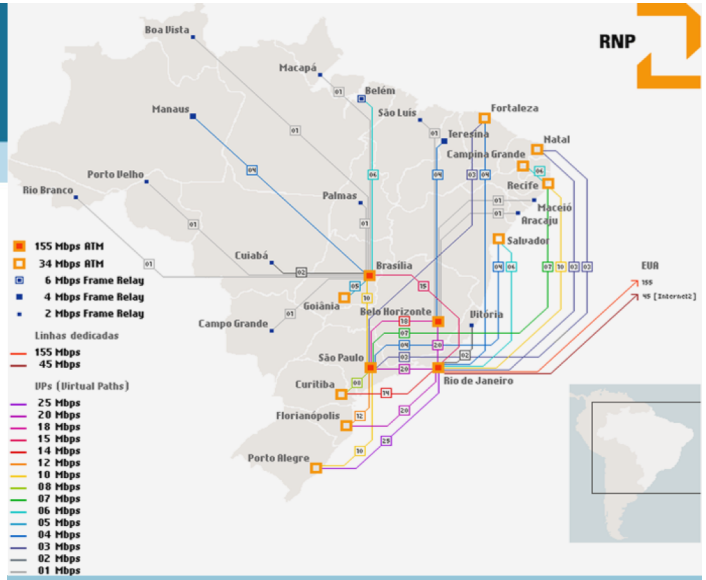


AMPATH | noc

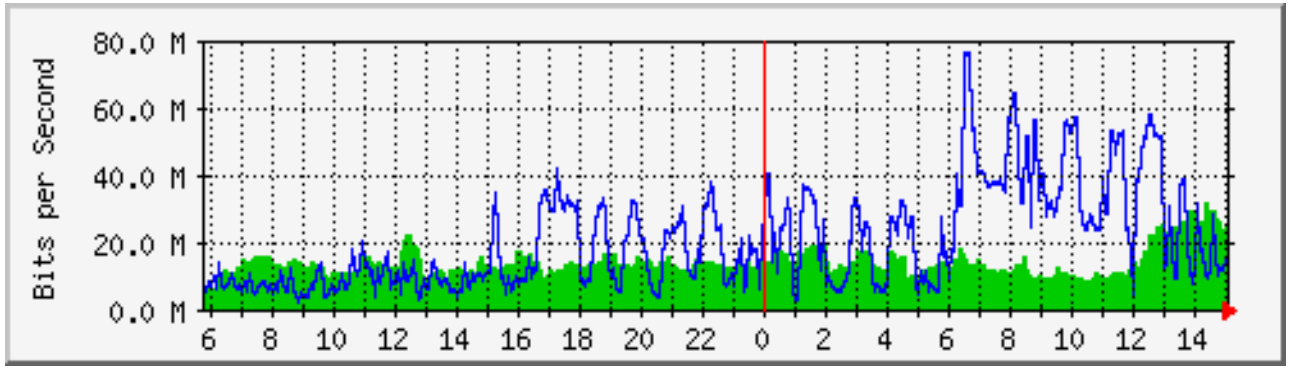
Abilene NOC | Euro-Link NOC | MIRnet NOC | STAR TAP NOC | TransPAC NOC | AMPATH Home

RNP Brazil





FIU Miami from So. America



Note: Auger (AG), ALMA (Chile), CMS-Tier1 (Brazil)

We Must Close the Digital Divide

Goal: To Make Scientists from All World Regions Full Partners in the Process of Search and Discovery

What ICFA and the HENP Community Can Do

- Help identify and highlight specific needs (to Work On)
  - ➡ Policy problems; Last Mile problems; etc.
- Spread the message: ICFA SCIC is there to help; Coordinate with AMPATH, IEEAF, APAN, Terena, Internet2, etc.
- Encourage Joint programs [such as in DESY's Silk project; Japanese links to SE Asia and China; AMPATH to So. America]
  - ➡ NSF/FIU & @LIS Proposals: US and EU to So. America
- Make direct contacts, arrange discussions with gov't officials
  - ➡ ICFA SCIC is prepared to participate
- Help Start, or Get Support for Workshops on Networks (& Grids)
  - ➡ Discuss & Create opportunities
  - ➡ Encourage, help form funded programs
- Help form Regional support & training groups (requires funding)

Networks, Grids and HENP

- Current generation of 2.5-10 Gbps network backbones arrived in the last 15 Months in the US, Europe and Japan
  - ➡ Major transoceanic links also at 2.5 - 10 Gbps in 2003
  - ➡ Capability Increased ~4 Times, i.e. 2-3 Times Moore's
- Reliable high End-to-end Performance of network applications (large file transfers; Grids) is required. Achieving this requires:
  - ➡ End-to-end monitoring; a coherent approach
  - ➡ Getting high performance (TCP) toolkits in users' hands
- *Digital Divide: Network improvements are especially needed in South America; Southeast Asia, SE Europe and Africa:*
  - ➡ *Key Examples: India, Pakistan, China; Brazil; Romania*
- Removing Regional, Last Mile Bottlenecks and Compromises in Network Quality are now
  On the critical path, in all world regions
- Work in Concert with AMPATH, Internet2, Terena, APAN; DataTAG, the Grid projects and the Global Grid Forum

New Technologies, Stewardship and e-Inclusion

- Access to and development of leading infrastructures and new classes of information-rich systems carries obligations
  - Stewardship
  - Playing a leading role in making these assets usable by a broad sector of the World Community
- Examples
  - Develop standardized toolkits, portals and Grid-enabled learning environments for wide access
    - Including from schools
  - Encourage joint programs and support from industry
  - Mandate strong education and outreach components in all medium and large research proposals (e.g. NSF)


Geoff Hall
Decoder...
Size...
Modes...
Dismiss
Guenter Fluegge
SUN_Virtual_Room
Jean-Marie BROM (IReS)
jmb@193.48.90.136/h261
5.6 f/s 213 kb/s (0%)
Geoff Hall
ghall@155.198.211.150/h261
8.0 f/s 130 kb/s (0%)
Guenter Fluegge
fluegge@137.226.33.77/h261
7.9 f/s 104 kb/s (0.1%)
Rino Castaldi ()
castaldi@193.205.77.230/h261
7.5 f/s 206 kb/s (2.5%)
Joe Incandela (University of Cal
incandel@131.225.235.137/h26
8.1 f/s 101 kb/s (0%)
Room 40-R-D10 (CERN)
CERN@137.138.77.179/h261
8.0 f/s 33 kb/s (0%)
mute
color
info...
VIC v2.9 by VRVS
Menu
Help
Quit
Virtual Rooms Videoconferencing System (VRVS)
File Edit View Favorites Tools Help
Back Search Favorites History
Address http://www.vrvs.org/
Go Links
VRVS
SUN
"CMS TRACKING S.C."
YOUR TIME (GMT +1)
DOC
Download
Schedule
JOIN
Call Someone
My Profile
Cern 40-R-D10 Room
Cern Guenter Fluegge
Infn Rino Castaldi
Rutherford Geoff Hall
Cern JEAN-MARIE BROM
Caltech Joe Incandela
Caltech Philippe Galvez
Cern danek kotlinski
MBONE
CHAT
QTIME
SHARING
H.323
CLOSE
JOIN
SET
Click to get the application
SUN_Virtual_Room
VRVS Team (@CERN)
mute
mute
50
0
Get Audio
UCL
RAT v3.2 by VRVS
Quit
Participants
40-R-D10 Room
Geoff Hall
Gregory Denis
Guenter Fluegge
JEAN-MARIE BROM
Joe Incandela
Rino Castaldi
danek kotlinski(2)
Controling the My office Camera - Net...
Camera Control
My office ROOM
Controlling Camera:
1
<--- PAN --->
TILT
Up
Slow
TILT
Down
Left
Slow PAN
Right
Zoom:
3
HELP
EXIT
Send
?
CLOSE
VRVS
Virtual Room Videoconferencing System
14600 Host Devices;
7800 Registered Users in
64 Countries
45 Network Servers
Annual Growth 2 to 3X

Next Generation Networks for Experiments: Goals and Needs

Large data samples explored and analyzed by thousands of globally dispersed scientists, in hundreds of teams

- Providing rapid access to event samples, subsets and analyzed physics results from massive data stores
  - From Petabytes by 2002, ~100 Petabytes by 2007, to ~1 Exabyte by ~2012.
- Providing analyzed results with rapid turnaround, by coordinating and managing the large but *LIMITED* computing, data handling and *NETWORK* resources effectively
- Enabling rapid access to the data and the collaboration
  - Across an ensemble of networks of varying capability
- Advanced integrated applications, such as Data Grids, rely on seamless operation of our LANs and WANs
  - With reliable, monitored, quantifiable high performance