Internet 2 International Meeting
September 19, 2005

ESnet Status Update

Joseph Burrescia, joeb@es.net
Senior Network Engineer

William E. Johnston, wej@es.net
ESnet Manager and Senior Scientist

Lawrence Berkeley National Laboratory

www.es.net

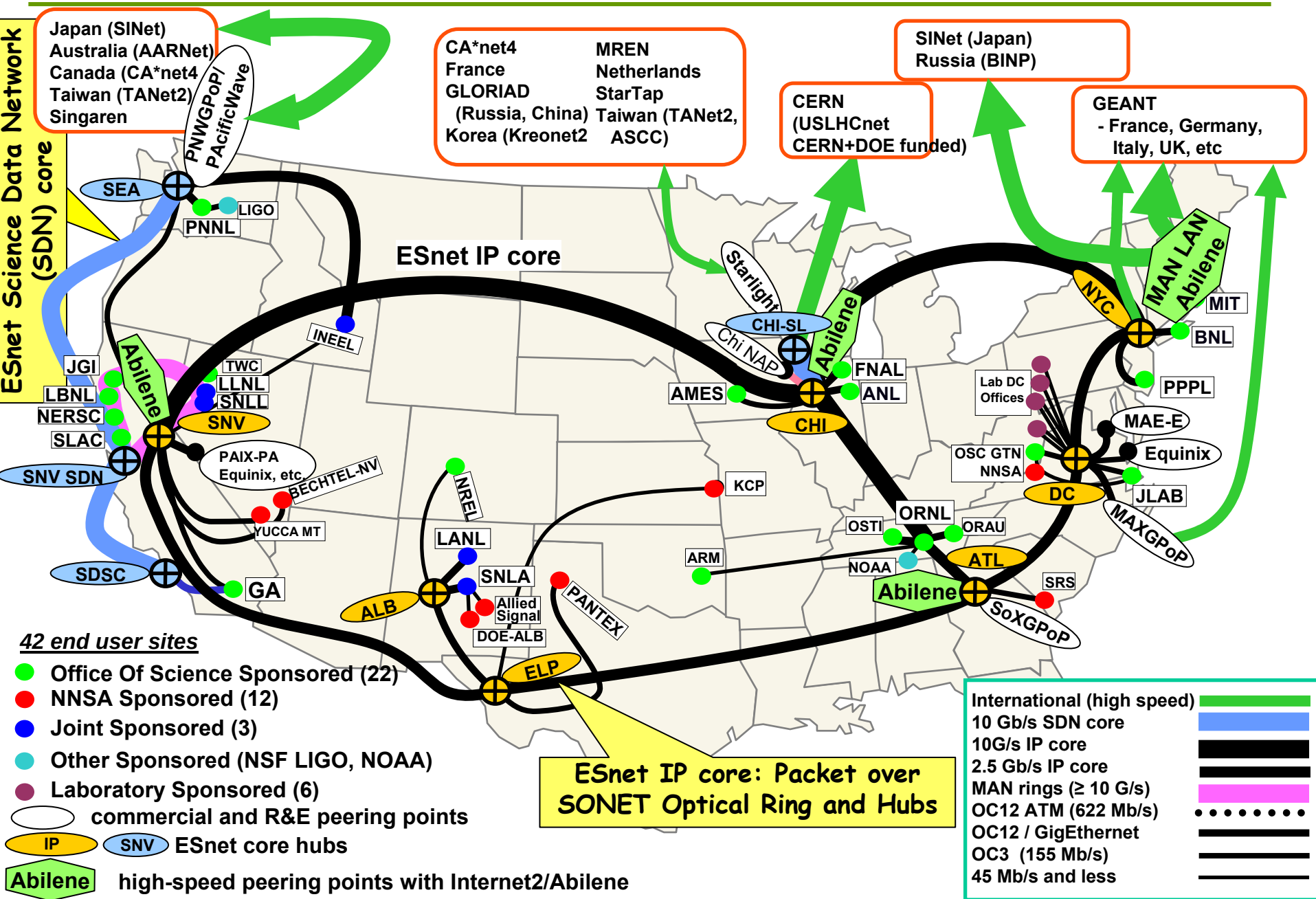
➢ DOE Office of Science Drivers for Networking

- The role of ESnet is to provide networking for the Office of Science Labs and their collaborators
- The DOE Office of Science supports more than 40% of all US R&D in high-energy physics, nuclear physics, and fusion energy sciences (http://www.science.doe.gov)
- The large-scale science that is the mission of the Office of Science is dependent on networks for
  - Sharing of massive amounts of data
  - Supporting thousands of collaborators world-wide
  - Distributed data processing
  - Distributed simulation, visualization, and computational steering
  - Distributed data management
- These issues were explored in two Office of Science workshops that formulated networking requirements to meet the needs of the science programs (see refs.)

Projected Science Requirements for Networking

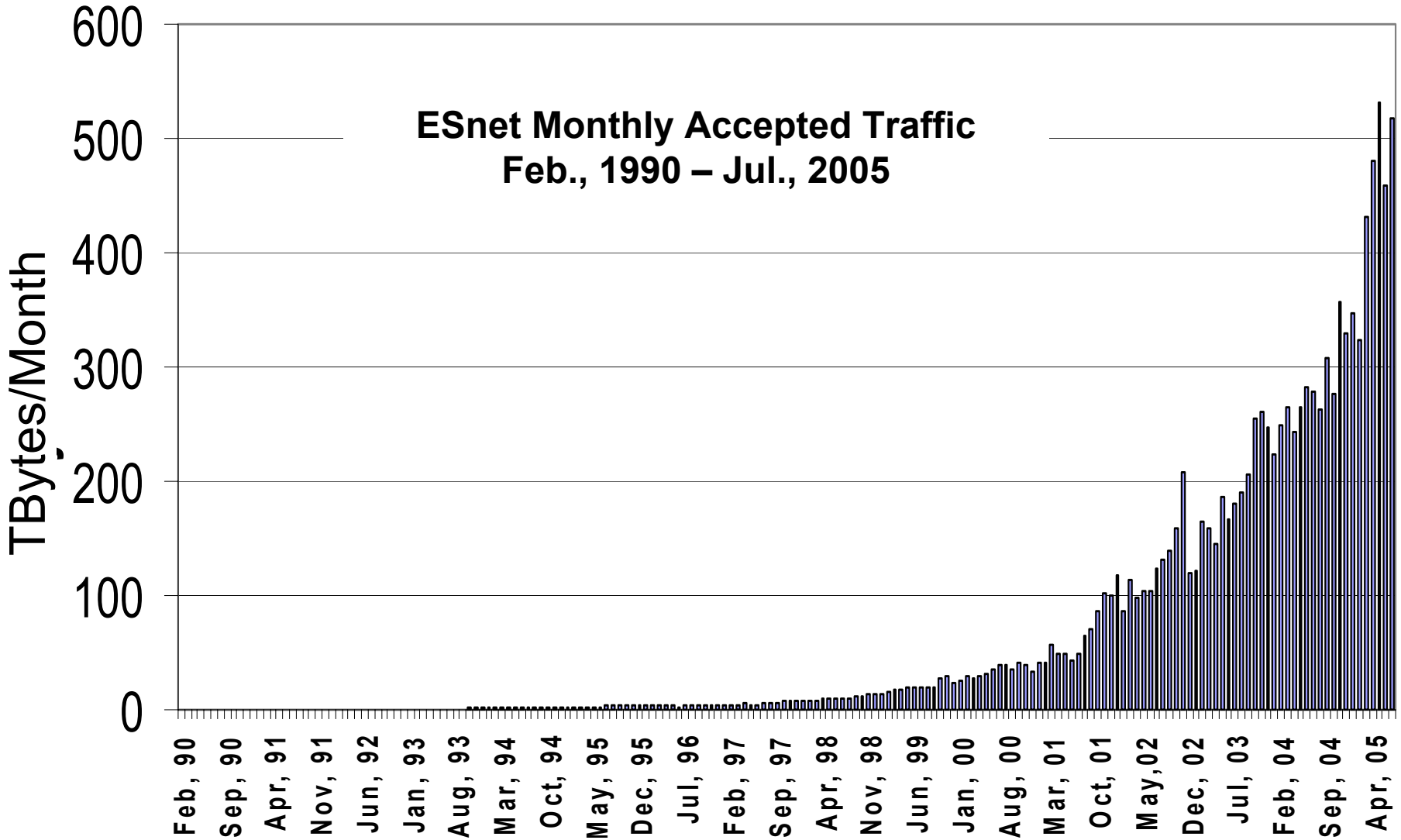
Science Areas considered in the Workshop [1] **(not including Nuclear Physics and Supercomputing)**	**Today *End2End* Throughput**	**5 years End2End Documented Throughput Requirements**	**5-10 Years End2End *Estimated* Throughput Requirements**	**Remarks**
High Energy Physics	0.5 Gb/s	**100 Gb/s**	**1000 Gb/s**	high bulk throughput with deadlines *(Grid based analysis systems require QoS)*
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**	remote control and time critical throughput (QoS)
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 (QoS)
Astrophysics	0.013 Gb/s (1 TBy/week)	**N*N multicast**	**1000 Gb/s**	computational steering and collaborations
Genomics Data & Computation	0.091 Gb/s (1 TBy/day)	**100s of users**	**1000 Gb/s**	high throughput and steering

ESnet Today Provides Global High-Speed Internet Connectivity for DOE Facilities and Collaborators



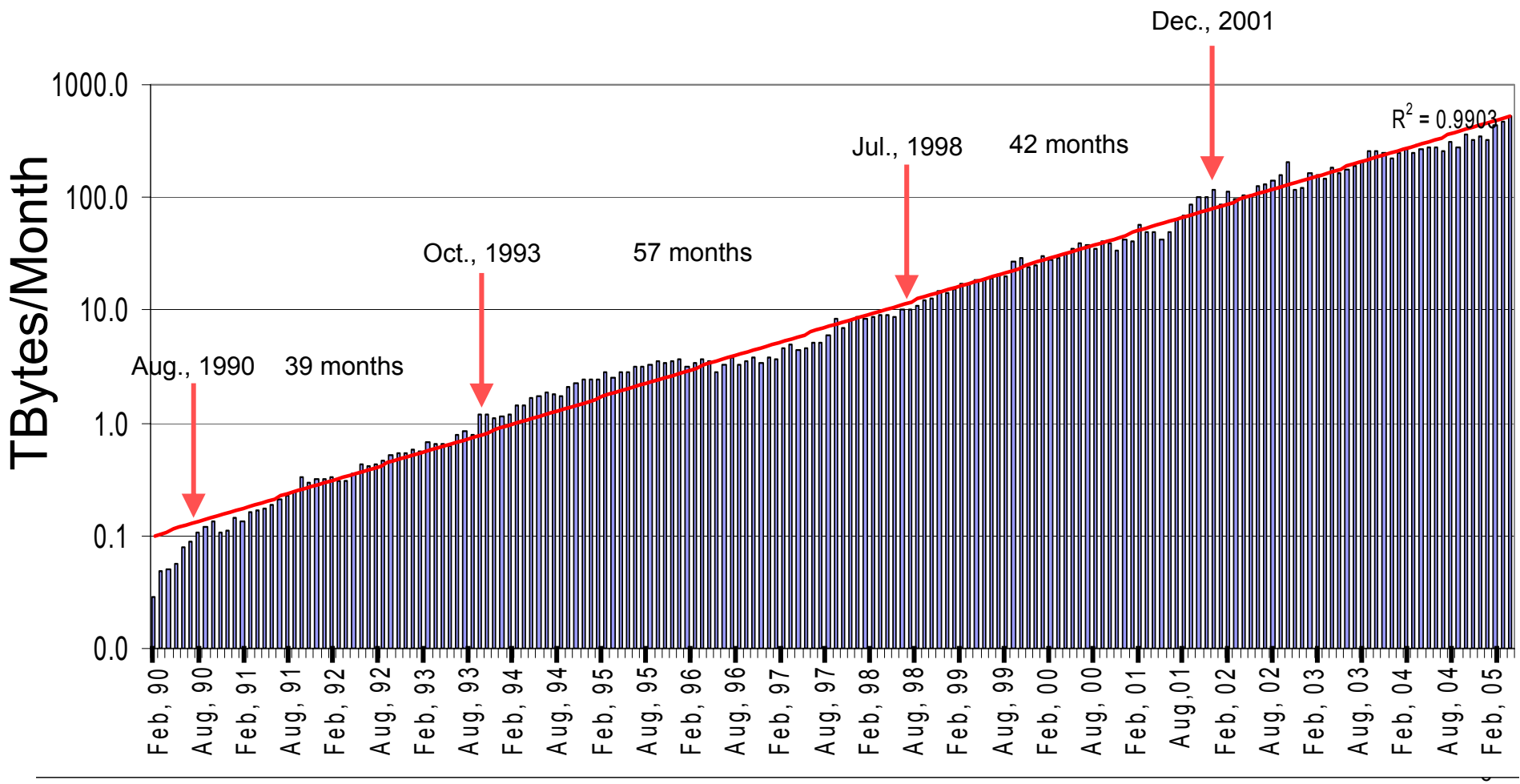
Observed Drivers for the Evolution of ESnet

ESnet is currently transporting About 530 Terabytes/mo. and this volume is increasing exponentially – ESnet traffic has increased by 10X every 46 months, on average, since 1990



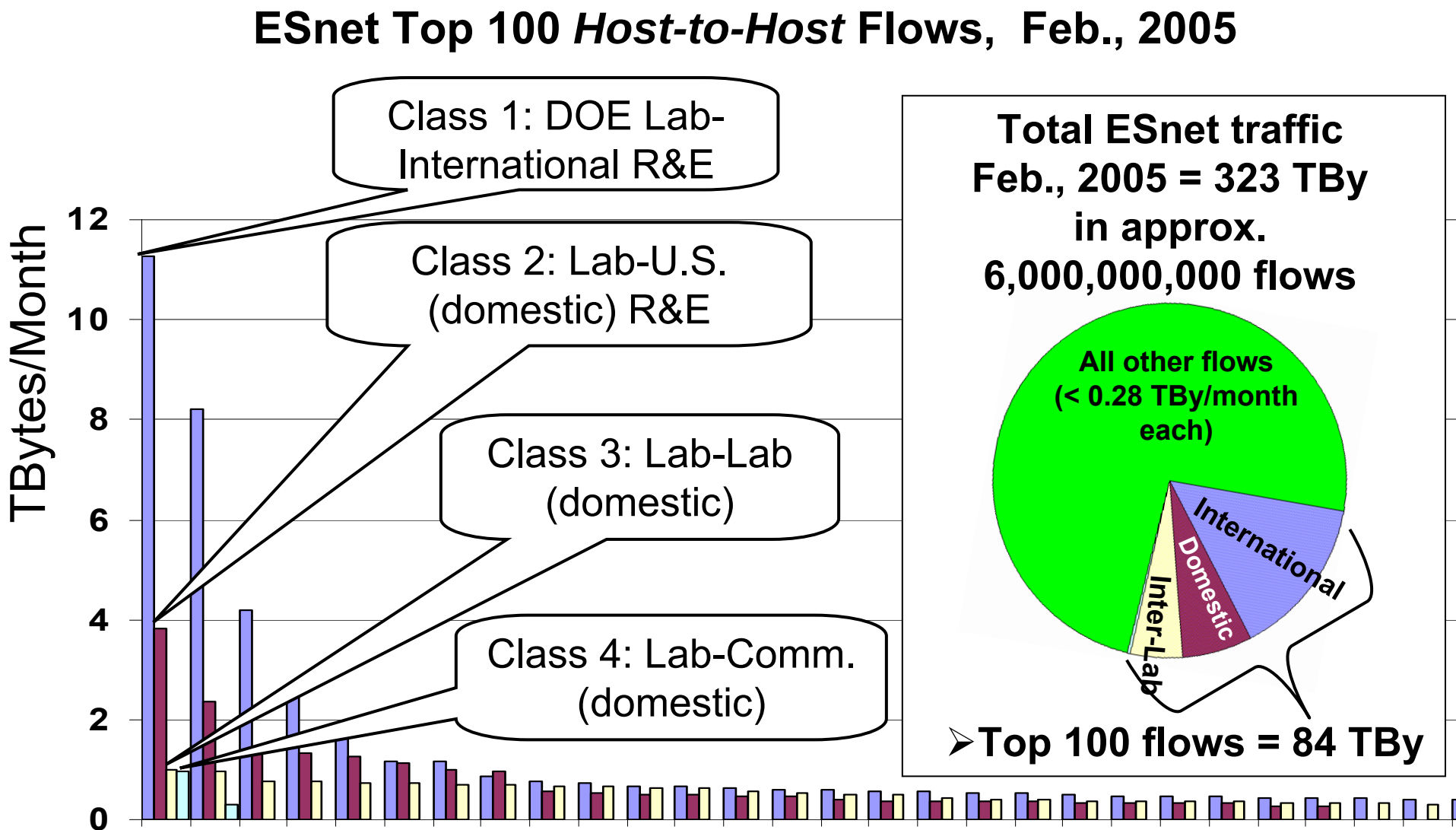
Observed Drivers for the Evolution of ESnet

ESnet traffic has increased by 10X every 46 months, on average, since 1990

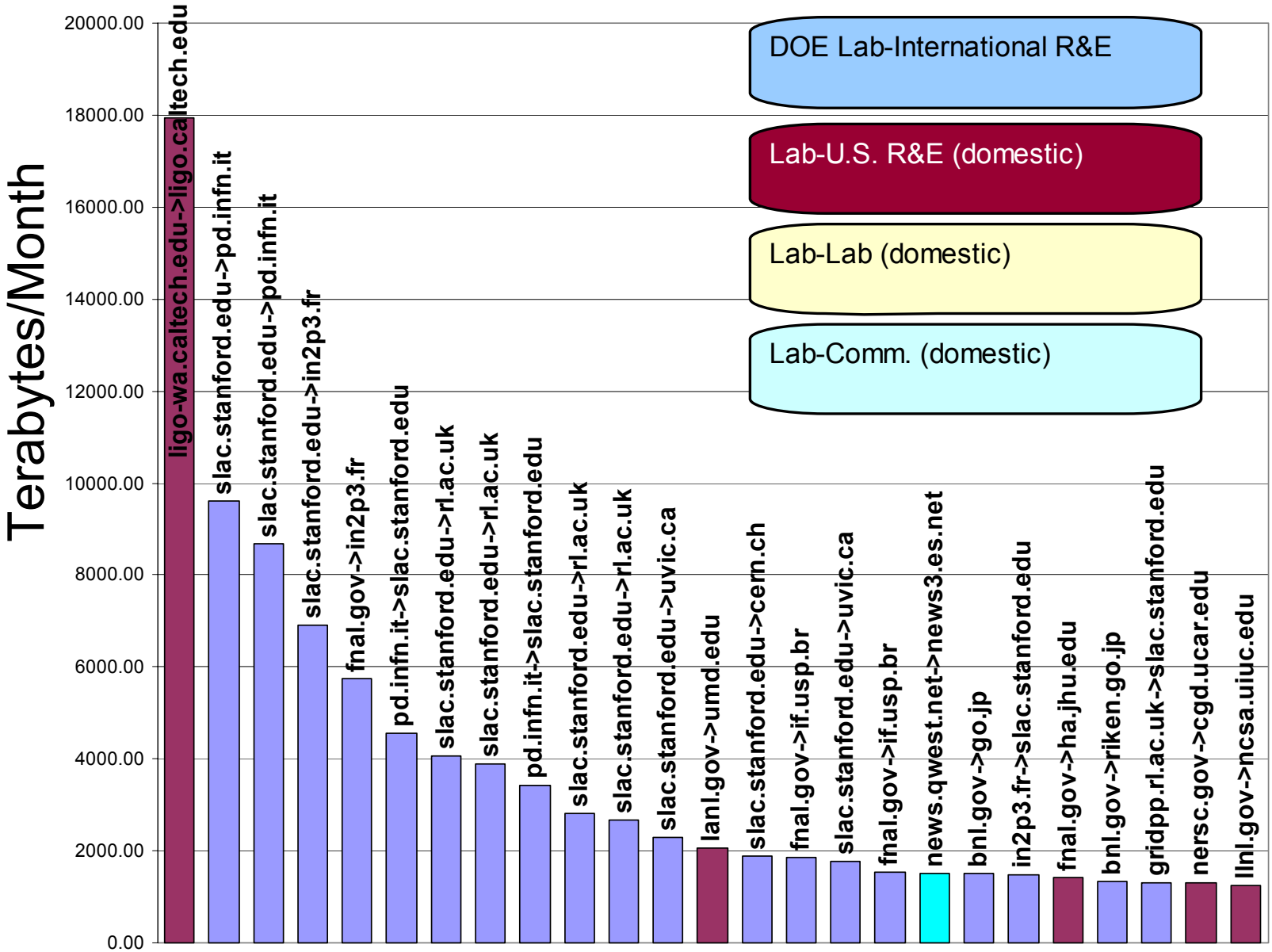


The Rise of Large-Scale Science

- **A small number of large-scale science users now account for a significant fraction of all ESnet traffic**



Source and Destination of the Top 25 Flows, Sept 2005



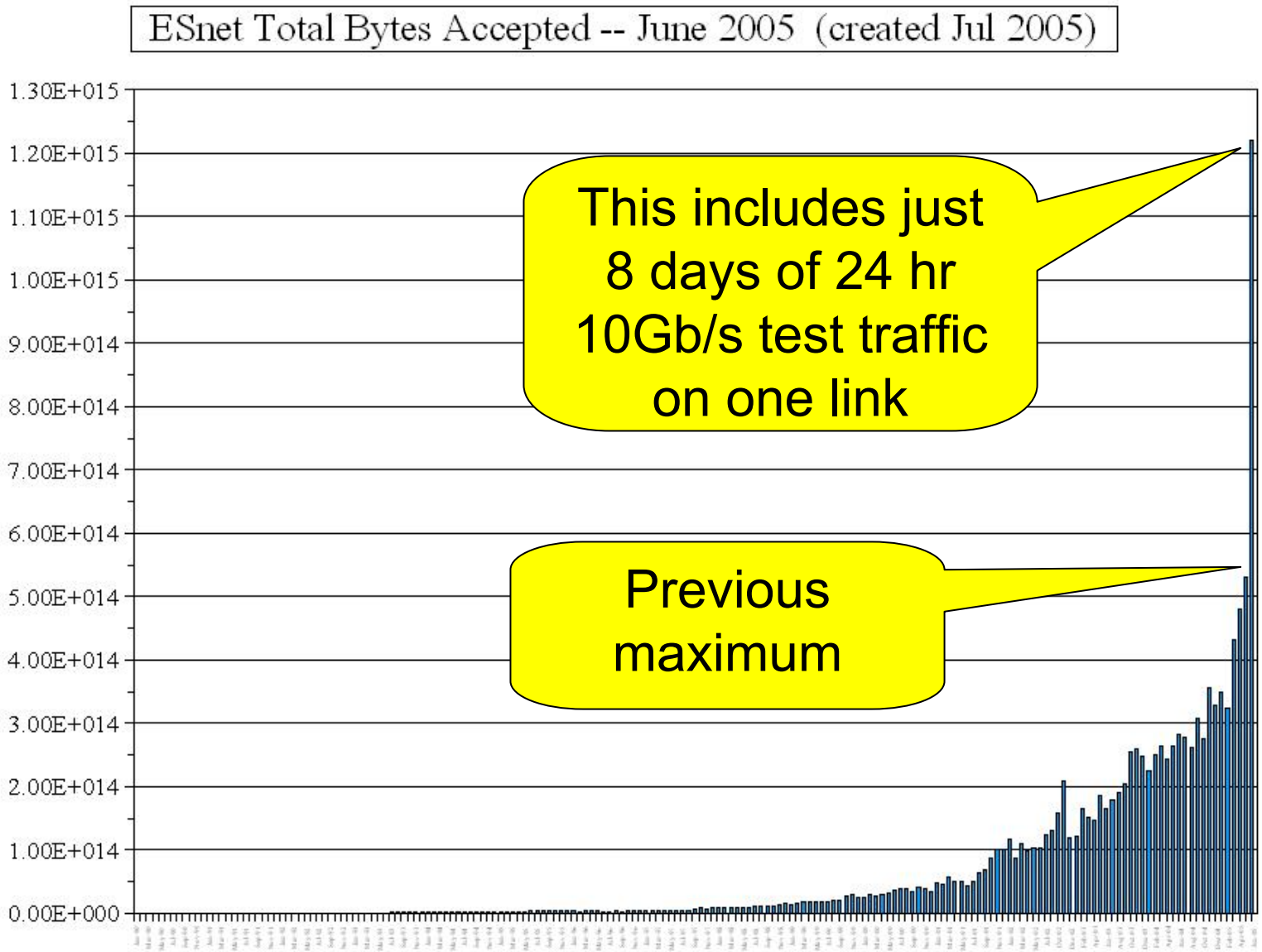
ESnet's Evolution over the Next 5-10 Years

- The current trend in traffic patterns – the large-scale science projects giving rise to the top 100 data flows that represent about 1/3 of all network traffic – will continue to evolve
  - As the LHC ramps up in 2006-07 the data to the tier 1 centers (FNAL and BNL) will increase 200-2000 times
    - A comparable amount of data will flow out of the tier 1 centers to the tier 2 centers (US universities) for data analysis
  - The National Leadership Class Facility supercomputer at ORNL anticipates a new model of computing in which simulation tasks are distributed across the central facility and a collection of remote "end stations" that will generate a lot of network traffic
  - As climate models achieve the sophistication and accuracy anticipated in the next few years, the amount of climate data that will move into and out of the NERSC center will increase dramatically (they are already in the top 100 flows)
  - Similarly for the experiment facilities at the new Spallation Neutron Source and Magnetic Fusion Energy facilities
  - Etc.

ESnet's Evolution over the Next 5-10 Years

- This evolution in traffic patterns and volume will result in (WEJ predicts)
  - The top 100 flows will become the top 1000 or 5000 flows
  - These large flows will account for 75% of a much larger total ESnet traffic volume as large-scale science data flows overwhelm everything else on the network
    - the remaining 6 billion flows will continue to account for the remainder of the traffic, which will also grow even as its fraction of the total becomes smaller
  - The current, few gigabits/sec of average traffic on the backbone will increase to 40 Gb/s (LHC traffic) and then increase to probably double that amount as the other science disciplines move into a collaborative production simulation and data analysis mode
    - This will get the backbone traffic to 100 Gb/s as predicted by the science requirements analysis three years ago

A Glimpse at the Future



ESnet's Evolution – The Requirements

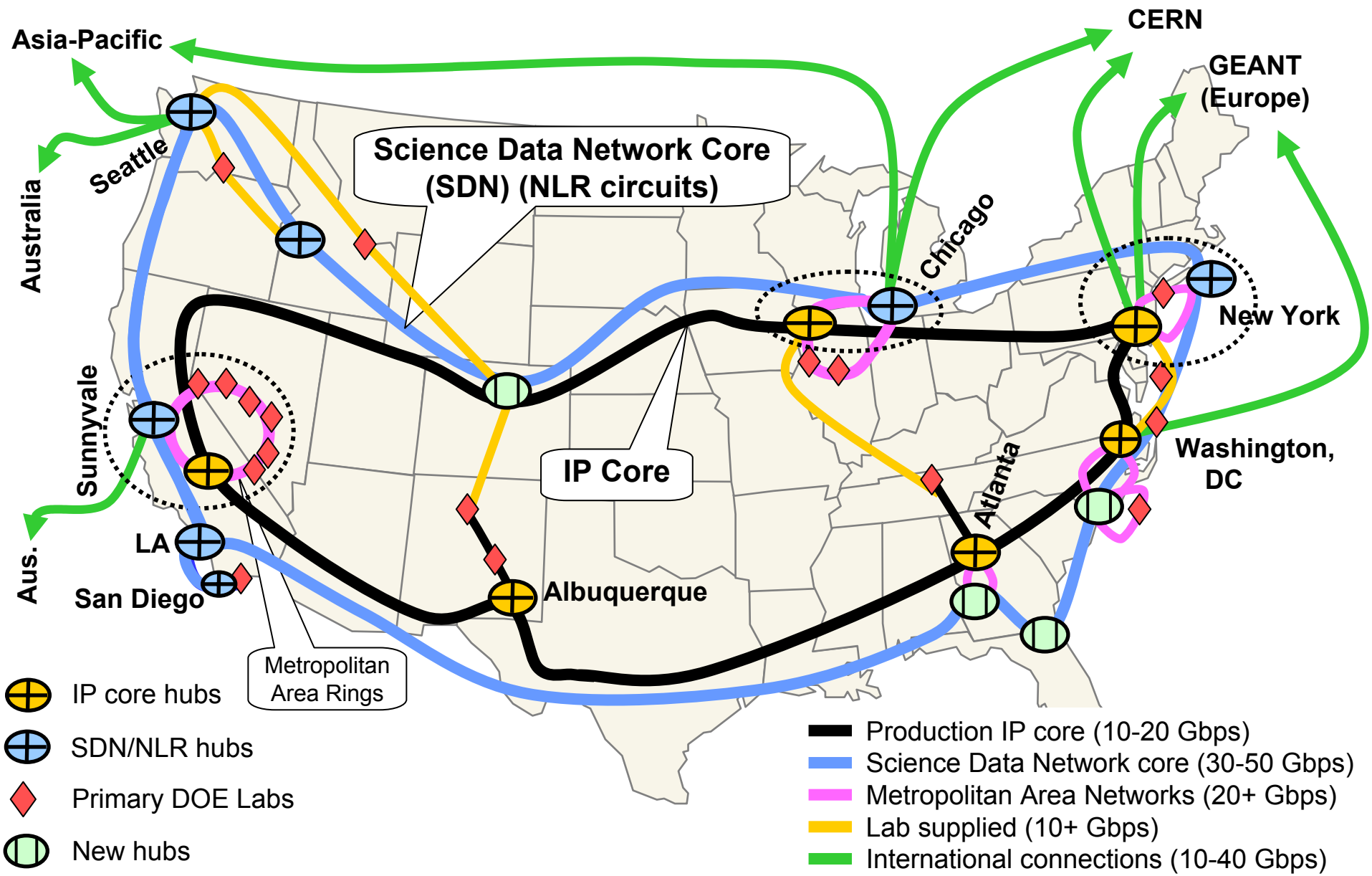
- In order to meet the identified requirements, the capacity and connectivity of the network must increase to provide
  - Fully redundant connectivity for every site
  - High-speed access to the core for every site
    - at least 20 Gb/s, generally, and 40-100 Gb/s for some sites
  - 100 Gbps national core/backbone bandwidth by 2008 in two independent backbones
- Every requirements workshop involving the science community has put bandwidth-on-demand as the highest priority – e.g.
  - Real-time data analysis for remote instruments
  - Control channels for remote instruments
  - Deadline scheduling for data transfers
  - "smooth" interconnection for complex Grid workflows

Strategy For The Evolution of ESnet

A three part strategy for the evolution of ESnet

1) ***Metropolitan Area Network*** (MAN) rings to provide
   - dual site connectivity for reliability
   - much higher site-to-core bandwidth
   - support for both production IP and circuit-based traffic
2) A ***Science Data Network*** (SDN) core for
   - provisioned, guaranteed bandwidth circuits to support large, high-speed science data flows
   - very high total bandwidth
   - multiply connecting MAN rings for protection against hub failure
   - alternate path for production IP traffic
3) A ***High-reliability IP core*** (e.g. the current ESnet core) to address
   - general science requirements
   - Lab operational requirements
   - Backup for the SDN core
   - vehicle for science services

Strategy For The Evolution of ESnet: Two Core Networks and Metro. Area Rings

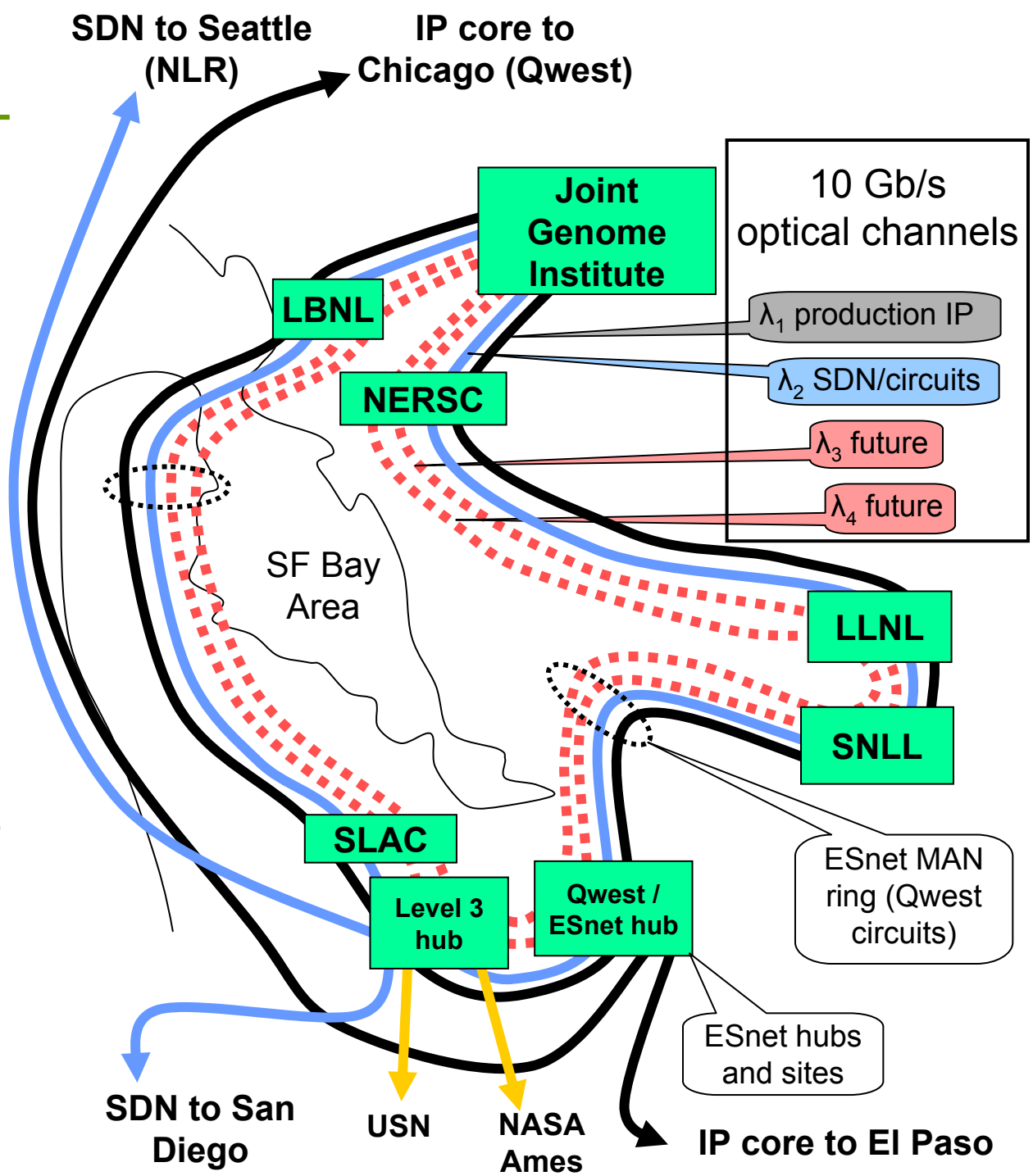


First Two Steps in the Evolution of ESnet

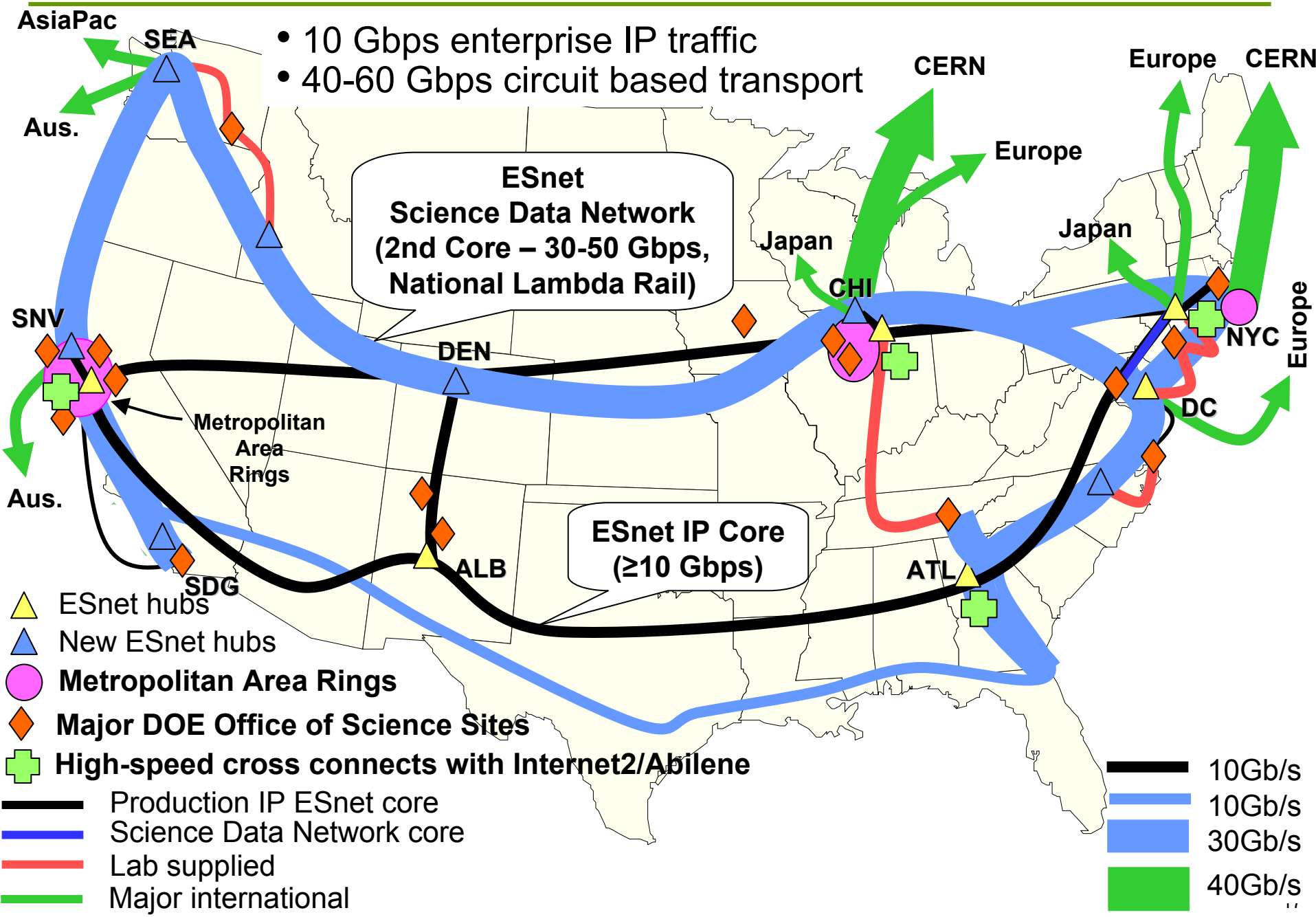
1) The SF Bay Area MAN will provide to the five OSC Bay Area sites
   - Very high speed site access – 20 Gb/s
   - Fully redundant site access
2) The first two segments of the second national 10 Gb/s core – the Science Data Network – are San Diego to Sunnyvale to Seattle

Status: ESnet BAMAN
ESnet SDN

- **BAMAN**
  - •Scheduled completion 9/31/05
  - •2 λs (2 X 10 Gb/s channels) in a ring configuration, and delivered as 10 GigEther circuits
  - •Dual site connection (independent “east” and “west” connections) to each site
  - •Will be used as a 10 Gb/s production IP ring and 2 X 10 Gb/s paths (for circuit services) to each site
- **SDN**
  - •10G SNV-SEA in production
  - •10G SNV-SD under test



➢ ESnet Goal – 2009/2010



LHC Networking and ESnet, Abilene, and GEANT

- USLHCnet (CERN+DOE funded) supports US participation in the LHC experiments and is the primary Tier 0 to Tier 1 path for the US Tier 1 data centers (FNAL and BNL)
- ESnet is responsible for getting the data from the trans-Atlantic connection points for the European circuits (Chicago and NYC) to the Tier 1 sites
  - ESnet is also responsible for providing backup paths from the trans-Atlantic connection points to the Tier 1 sites
- Abilene is responsible for getting data from ESnet to the Tier 2 sites
- The new ESnet architecture (Science Data Network) is intended to accommodate the anticipated 20-40 Gb/s from LHC to US (both US tier 1 centers are on ESnet)

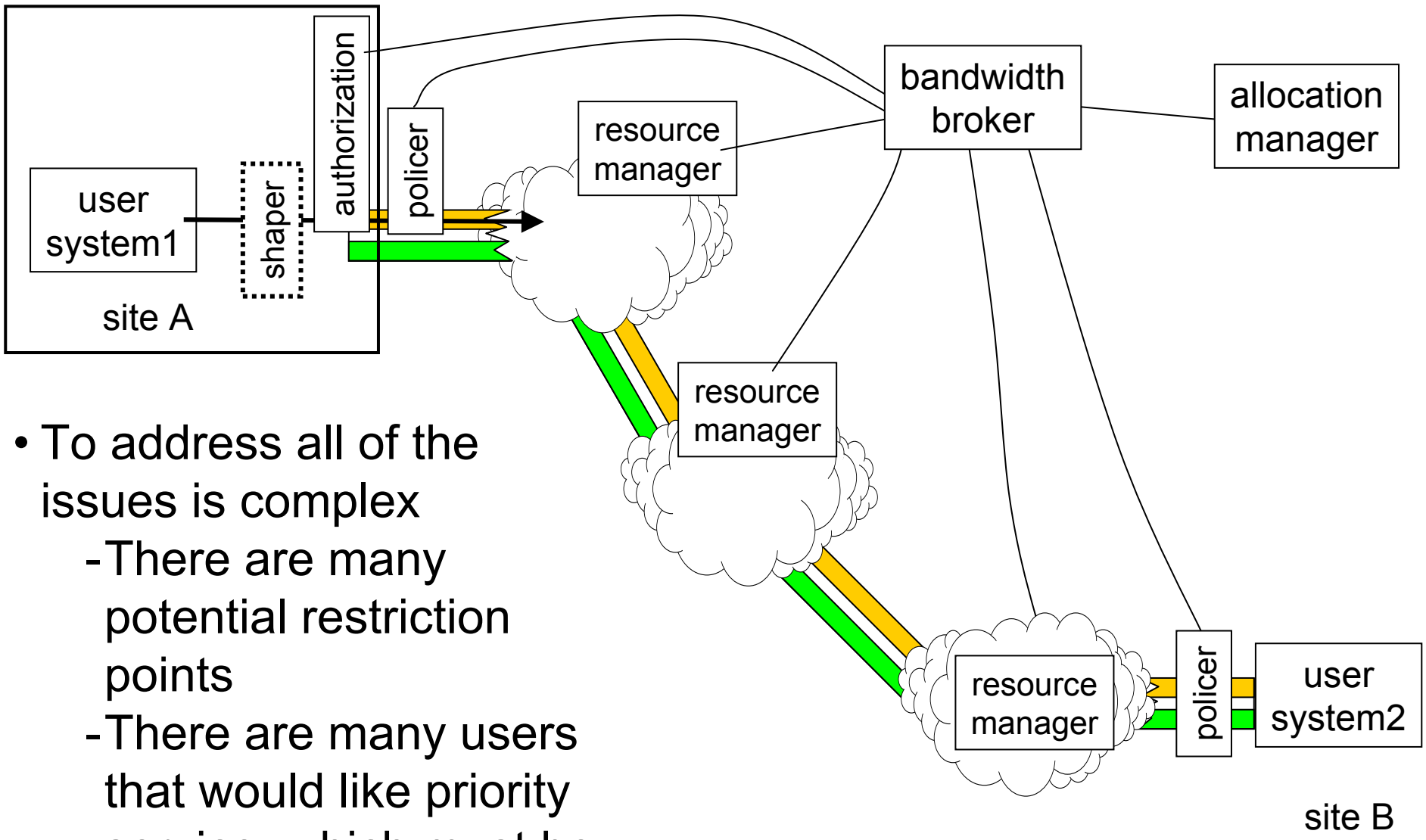
➢ Virtual Circuit Services

- New network services are critical for ESnet to meet the needs of large-scale science like the LHC
- Most important new network service is ***dynamically provisioned virtual circuits*** that provide
  - Traffic isolation
    - will enable the use of high-performance, non-standard transport mechanisms that cannot co-exist with commodity TCP based transport
      (see, e.g., Tom Dunigan's compendium
      http://www.csm.ornl.gov/~dunigan/netperf/netlinks.html)
  - Guaranteed bandwidth
    - E.g., the only way that we have currently to address deadline scheduling – e.g. where fixed amounts of data have to reach sites on a fixed schedule in order that the processing does not fall behind far enough so that it could never catch up – very important for experiment data analysis

On-demand Secure Circuits and Advance Reservation System (OSCARS)

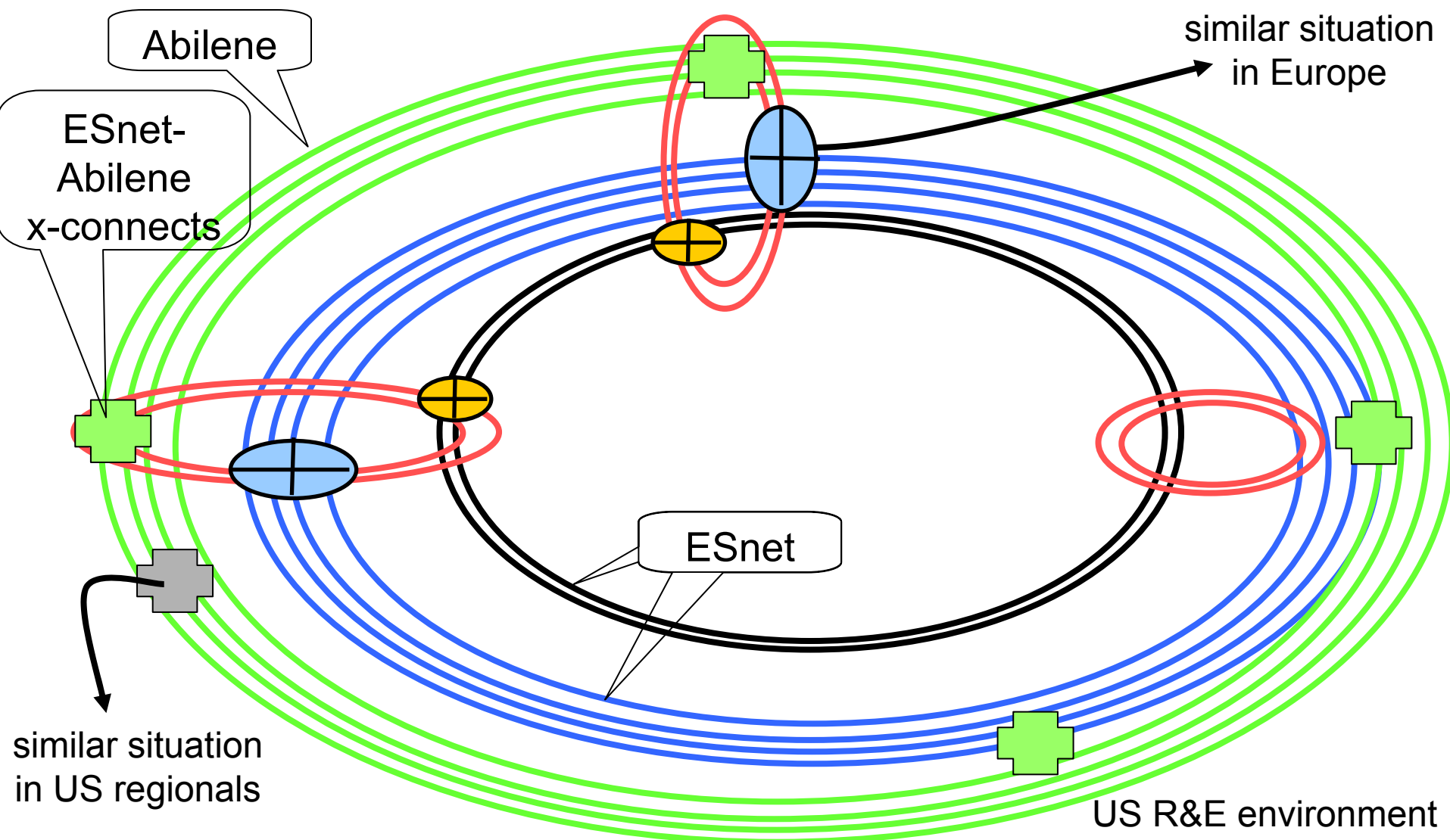
- Virtual circuits must operate across domains
  - End points will be on campuses or research institutes that are served by ESnet, Abilene's regional networks, and GEANT's regional networks – typically five domains to cross to do end-to-end system connection
  - There are many issues here that are poorly understood
- To ensure compatibility the work is a collaboration with the other major science R&E networks
  - Code is being jointly developed with Internet2's Bandwidth Reservation for User Work (BRUW) project – part of the Abilene HOPI (Hybrid Optical-Packet Infrastructure) project
  - Close cooperation with the GEANT virtual circuit project ("lightpaths – Joint Research Activity 3 project)

OSCARS: Guaranteed Bandwidth Service



- To address all of the issues is complex
  - There are many potential restriction points
  - There are many users that would like priority service, which must be rationed

Between ESnet, Abilene, GEANT, and the connected regional R&E networks, there will be dozens of lambdas in production networks that are shared between thousands of users who want to use virtual circuits – Very complex inter-domain issues



Tying Domains Together (1/2)

- **Motivation**:
  - For a virtual circuit service to be successful, it must
    - Be end-to-end, potentially crossing several administrative domains
    - Have consistent network service guarantees throughout the circuit
- **Observation:**
  - Setting up an intra-domain circuit is easy compared with coordinating an inter-domain circuit
- **Issues:**
  - Cross domain authentication *and* authorization
    - A mechanism to authenticate and authorize a bandwidth on-demand (BoD) circuit request must be agreed upon in order to automate the process
  - Multi-domain Acceptable Use Policies (AUPs)
    - Domains may have very specific AUPs dictating what the BoD circuits can be used for and where they can transit/terminate
  - Domain specific service offerings
    - Domains must have way to guarantee a certain level of service for BoD circuits
  - Security concerns
    - Are there mechanisms for a domain to protect itself (e.g. RSVP filtering)

Tying Domains Together (2/2)

- **Approach**:
  - Utilize existing standards and protocols (e.g. GMPLS, RSVP)
  - Adopt widely accepted schemas/services (e.g X.509 certificates)
  - Collaborate with like-minded projects (e.g. JRA3 (GEANT), BRUW (Internet2/HOPI) to:
    1. Create a common service definition for BoD circuits
    2. Develop an appropriate User-Network-Interface (UNI) and Network-Network-Interface (NNI)

References

1) High Performance Network Planning Workshop, August 2002

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

2) DOE Science Networking Roadmap Meeting, June 2003

http://www.es.net/hypertext/welcome/pr/Roadmap/index.html

3) OSCARS (On-demand Secure Circuits and Advance Reservation System)

http://www.es.net/oscars/index.html