Future Capabilities and Challenges for ESnet

Gregory Bell, Ph.D.

Director, Energy Sciences Network

Director, Scientific Networking Division

Lawrence Berkeley National Laboratory

Next-Generation Networks for Science PI Meeting

March 18, 2013 – Joint Bio-Energy Institute

Outline

ESnet Updates

Growth and Drivers

Programmability

Network Testbed and Research Questions

Outline

ESnet Updates

Growth and Drivers

Programmability

Network Testbed and Research Questions

Energy Sciences Network Overview

A national network, optimized for science:

- connecting 40 labs, facilities with >100 networks
- optimized for massive science data flows
- offering capabilities not available commercially
- $34.5M in FY12 (40 staff)

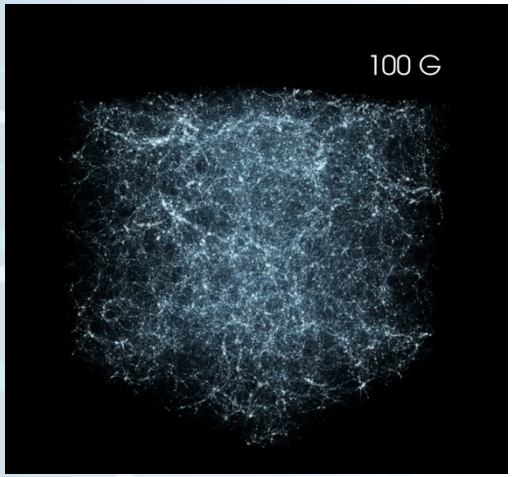
$62M ARRA grant funding:

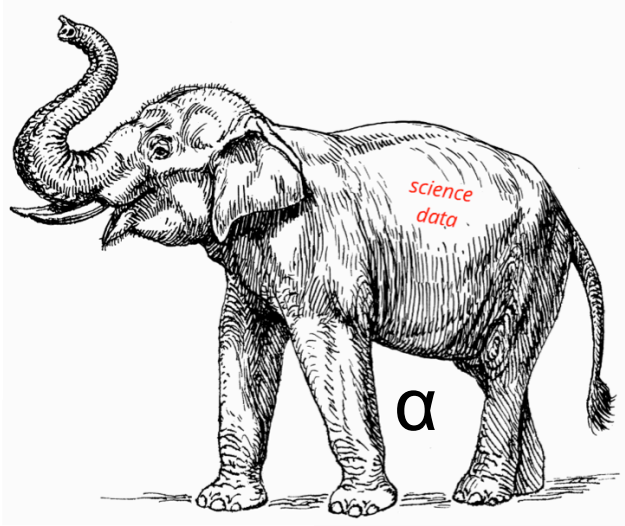
- optical fiber assets
- access to spectrum
- world's first 100G network at scale
- easier (and cheaper) scaling

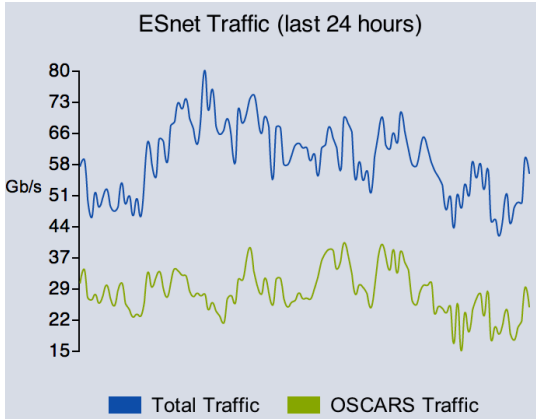
On the web:

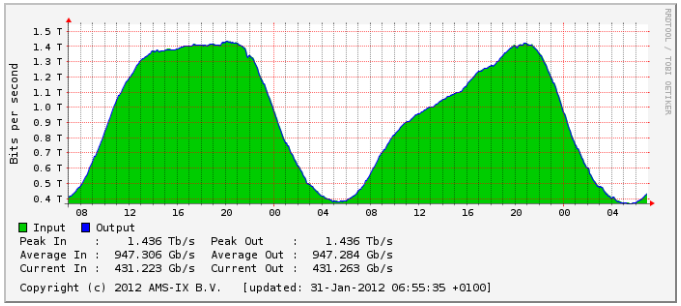
- www.es.net
- fasterdata.es.net
- my.es.net

ESnet is Optimized for the Demanding Requirements of Science Data Flows









Not the Commercial Internet

Engineered for massive traffic flows:

- emphasis on *lossless transport* for maximum performance
- bandwidth guarantees spanning multiple network domains
- distributed performance monitoring platform
- very high-speed (n x 100G) data transport, scaling cost-effectively

Optimized for DOE science missions:

- extensive scientific outreach, requirements-gathering activity
- consistent record of architectural innovation
- visualization tools for monitoring health of science data flows

Expert user support and advocacy:

- rapid diagnosis of data transport issues, around the world
- global leadership in developing network standards, middleware
- global advocacy for science-optimized network /security architectures

ESnet's Year of Unprecedented Change

The facility built its first optical network, deployed a new routing platform, transitioned to ESnet5, and said goodbye to ESnet4.

- concurrently: 6 critical staff retired, and a new facility director was chosen

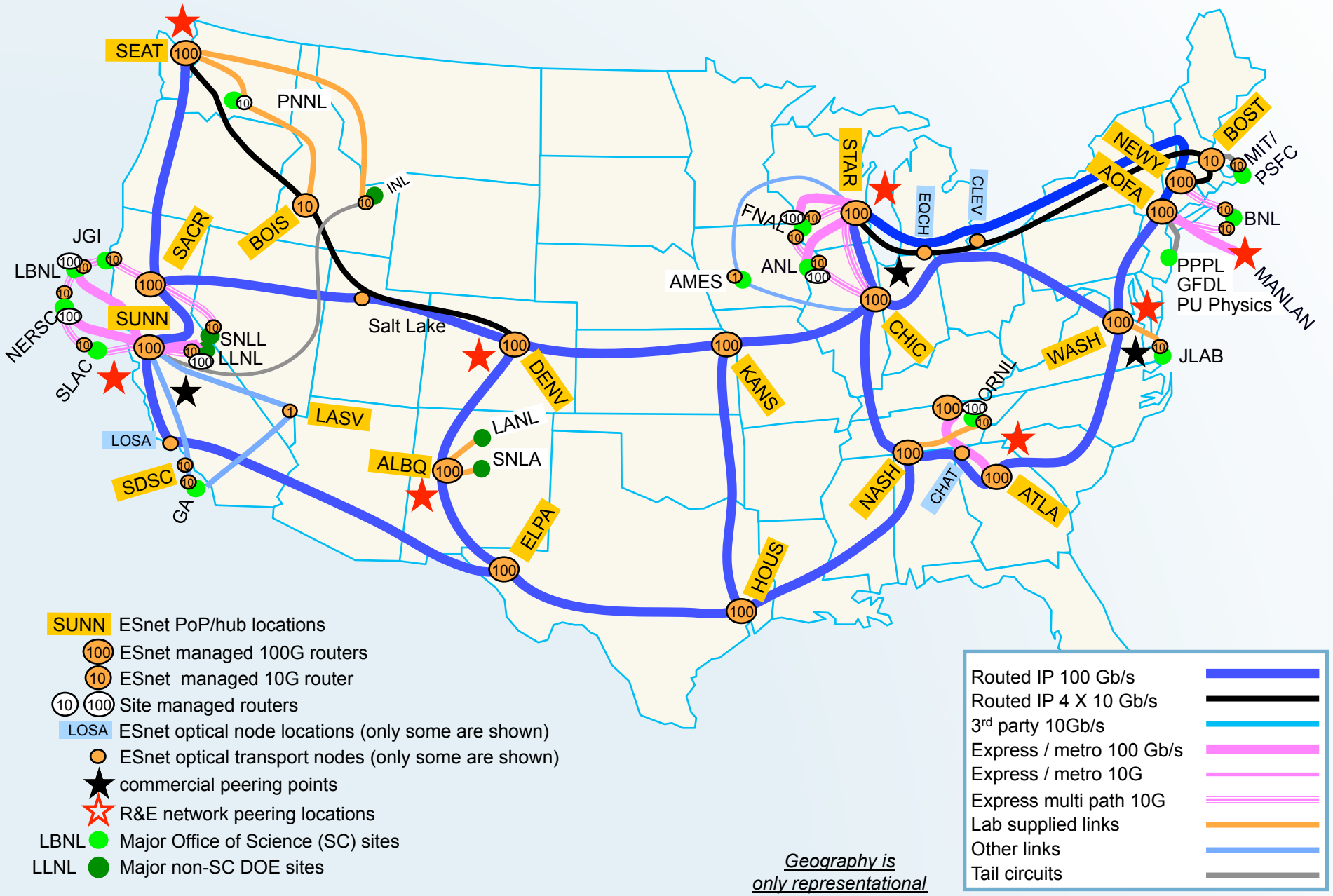
Facility completed the ANI testbed (closeout report under DOE review).

- future testbed strategy under development, due to ASCR this summer

Facility demonstrated major OpenFlow / SDN innovations:

- OpenFlow + OSCARS for stitching end-to-end path for post-TCP flows
- WAN virtual switch
- first OpenFlow optical transport, in partnership with Infinera
- NSI in OSCARS for SC12, plus ongoing leadership of standardization effort

ESnet5 January 2013



Two Awards for Deployment of ESnet5

FierceGovernment chose the ESnet5 Deployment Team as a recipient of the annual Fierce 15 award, in recognition of "federal employees and teams who have done particularly innovative things."

> "The expert execution of this science-enabling project is also worth noting."

Information Week named ESnet as one of the "top 15 innovators for 2012", among government entities at every layer: federal, state and local. This is the second time in four years ESnet has receive this award.

Outline

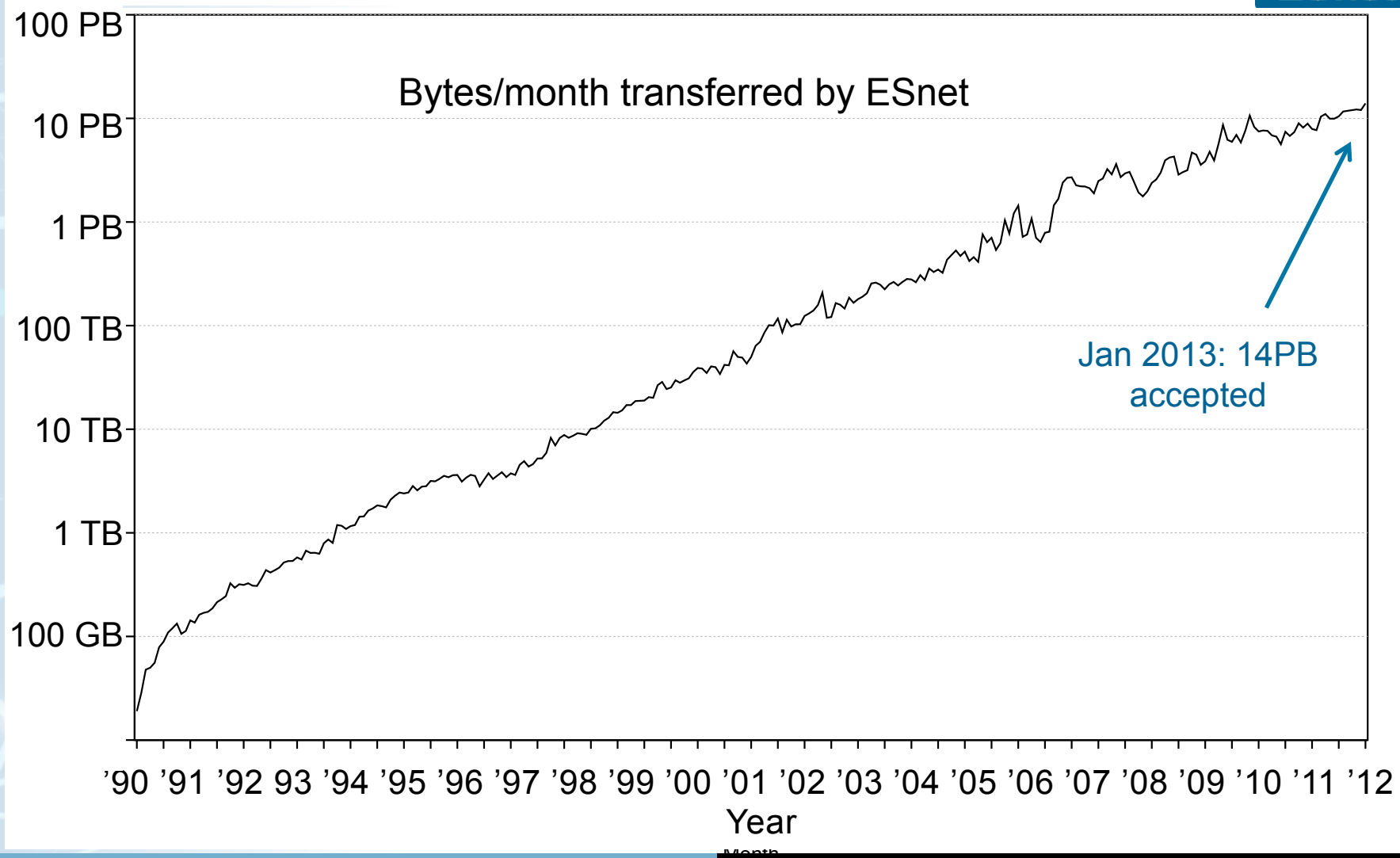
ESnet Updates

Growth and Drivers

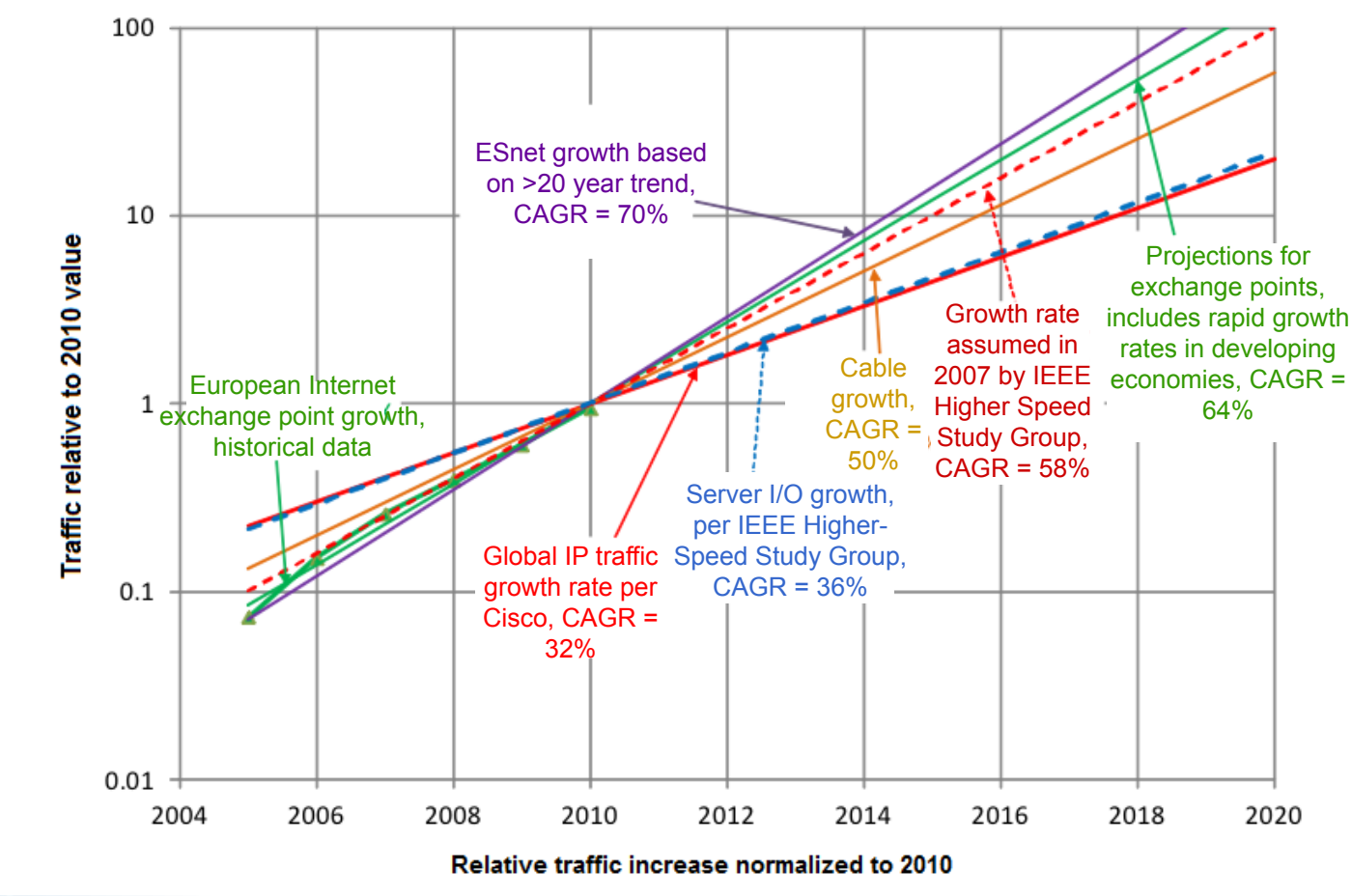
Programmability

Network Testbed and Research Questions

Unrelenting Growth in Traffic for >20 Years



Growth in Science >> Global IP Traffic



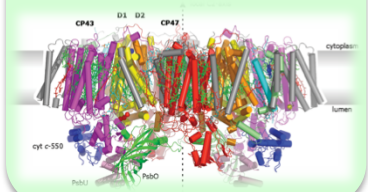
CAGR = compound annual growth rate.

Relative traffic increase for various sectors normalized to 2010, **adapted from** IEEE 802.3™ Industry Connections Ethernet Bandwidth Assessment Ad Hoc Report.

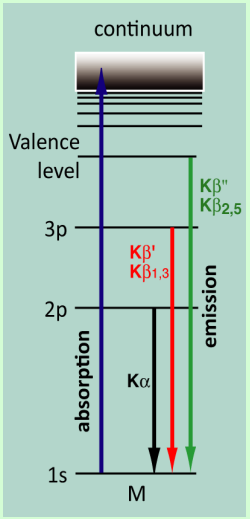
Using Linac Coherent Light Source at SLAC, Take Snapshots of Catalytic Reaction in Photosystem II

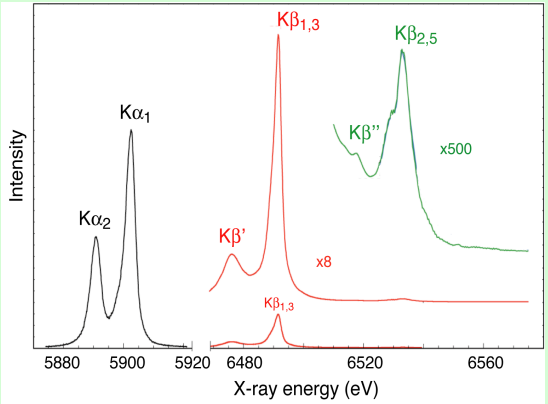
- Goal is to understand detailed mechanism of water splitting
- Linac Coherent Light Source Free-electron laser
- 50 fs X-ray pulses above the Mn absorption edge
- Diffract before destroy approach
- Simultaneously detect X-ray emission

X-ray diffraction
(structure)

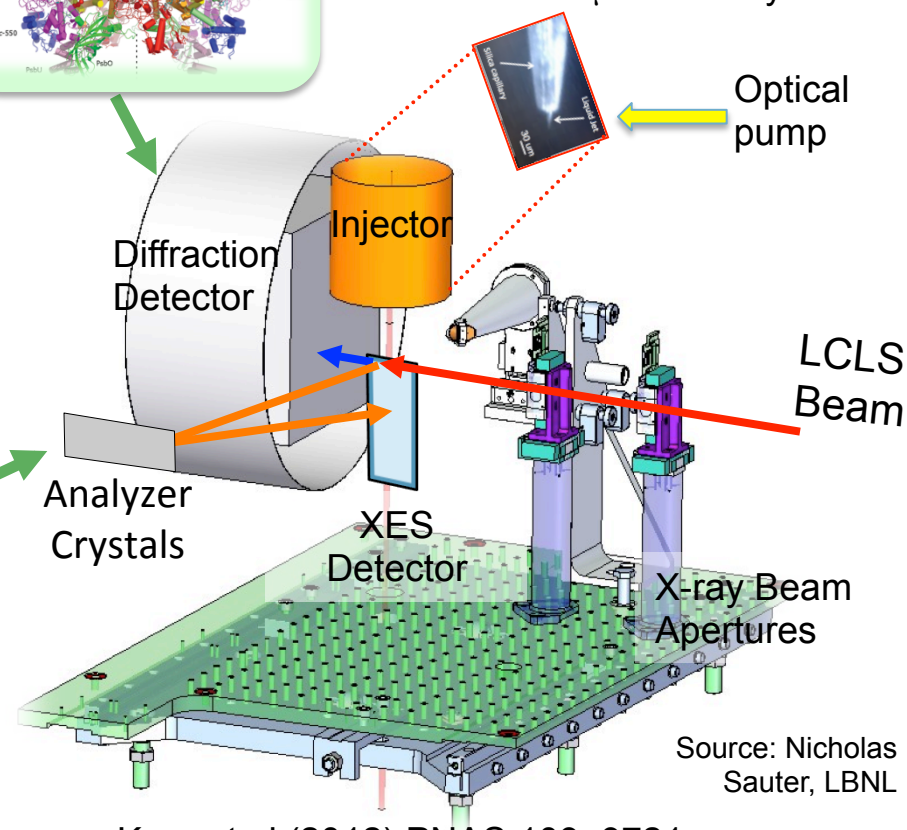


X-ray emission spectroscopy
(Chemistry at the catalytic site)





- charge density/spin state
- ligand environment

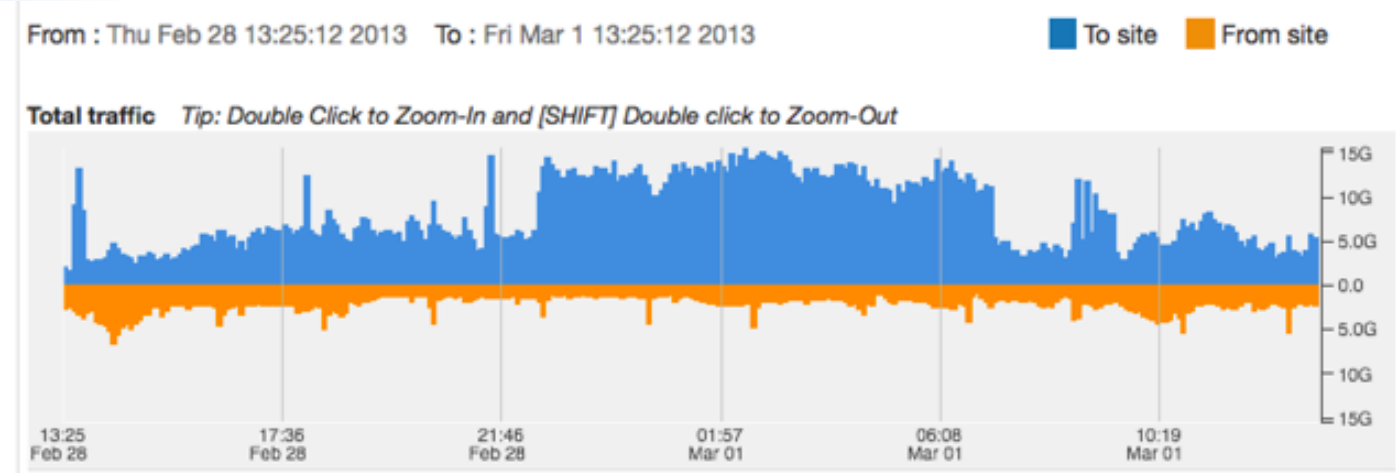


Source: Nicholas Sauter, LBNL

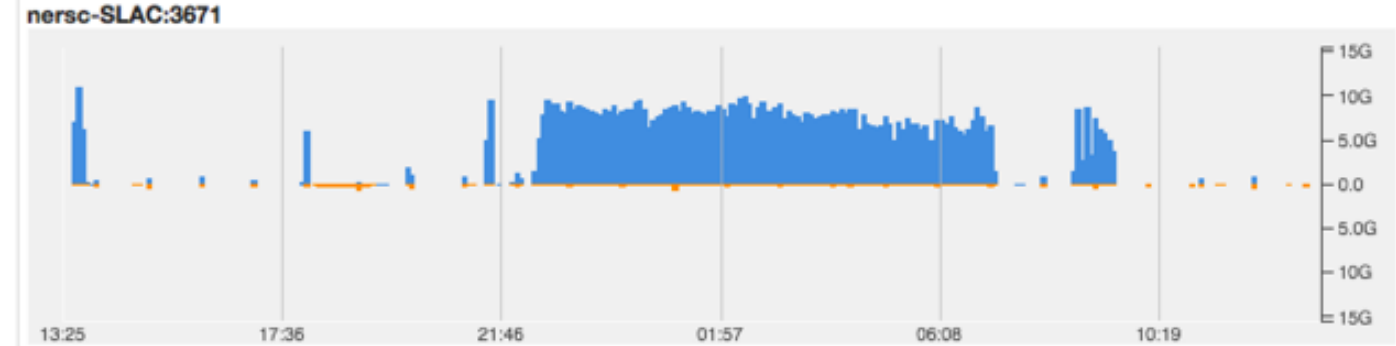
- Kern et al (2012) PNAS 109: 9721
- Sierra et al (2012) Acta Cryst D68: 1584
- Mori et al (2012) PNAS 109: 19103

SLAC / LCLS ➔NERSC (http://my.es.net)

All NERSC Traffic

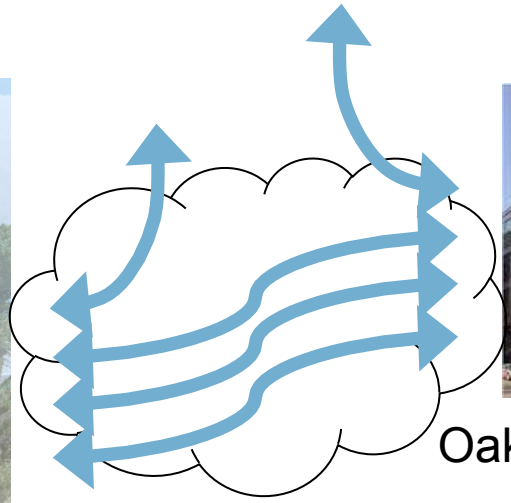


Photosystem II X-Ray Study



5 x 100G for NERSC in 2013/14

Computational Research and Theory Building, Berkeley Lab



ESnet

Oakland Scientific Facility

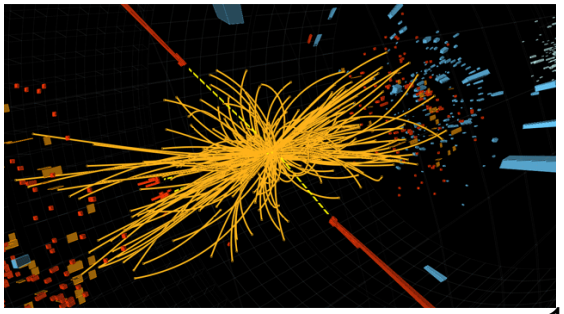
- 2 x 100G for file system links
- 1 x 100G for inter-site traffic
- 2 x 100G for WAN traffic

Evolution of LHC Data Model towards ~Remote IO

In chronological order:

1. Copy as much data as feasible to analysis centers worldwide, with hierarchical distribution.
2. Relax the hierarchy and rely on caching.
3. Use “federated data stores” to fetch *portions* of relevant data sets from remote storage (anywhere), just before they’re needed.

Increasing faith in global science networks.



Outline

ESnet Updates

Growth and Drivers

Programmability

Network Testbed and Research Questions

Three Historical Inflection Points for Global Research Networks

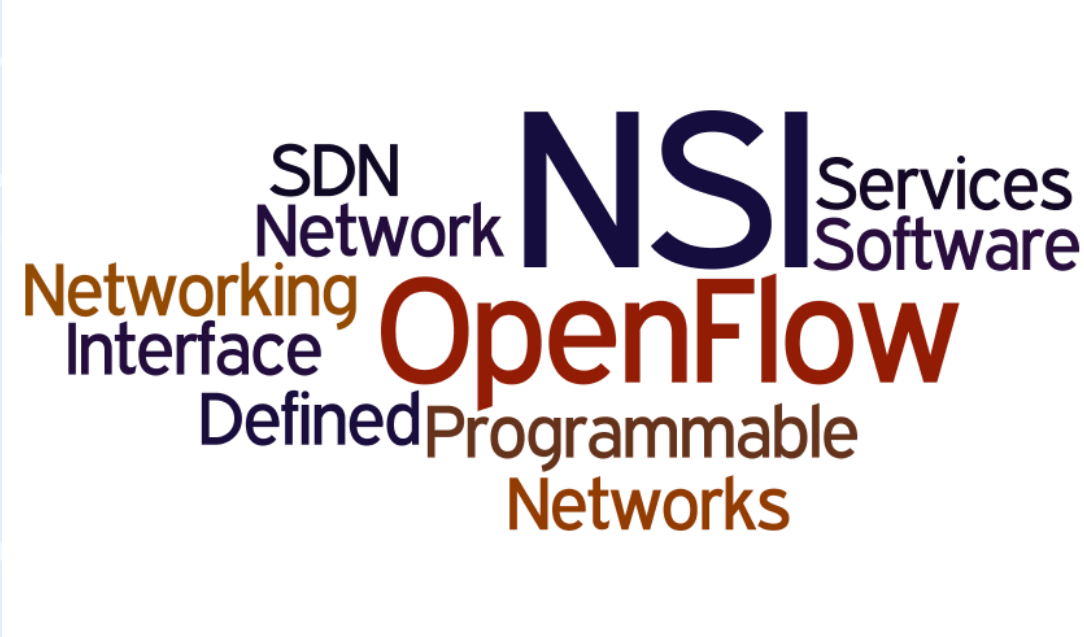
1. Abundant capacity (88 λ x 100Gbps)

2. Programmability

3. Campus architectures newly optimized for data mobility

ESnet architecture (Science DMZ) + NSF grants.

Programmability



Journey with Programmability

OSCARS, 2006-2013

<u>Insights</u>

- understand the customer
- involve the community
- open source
- experiment
- production quality

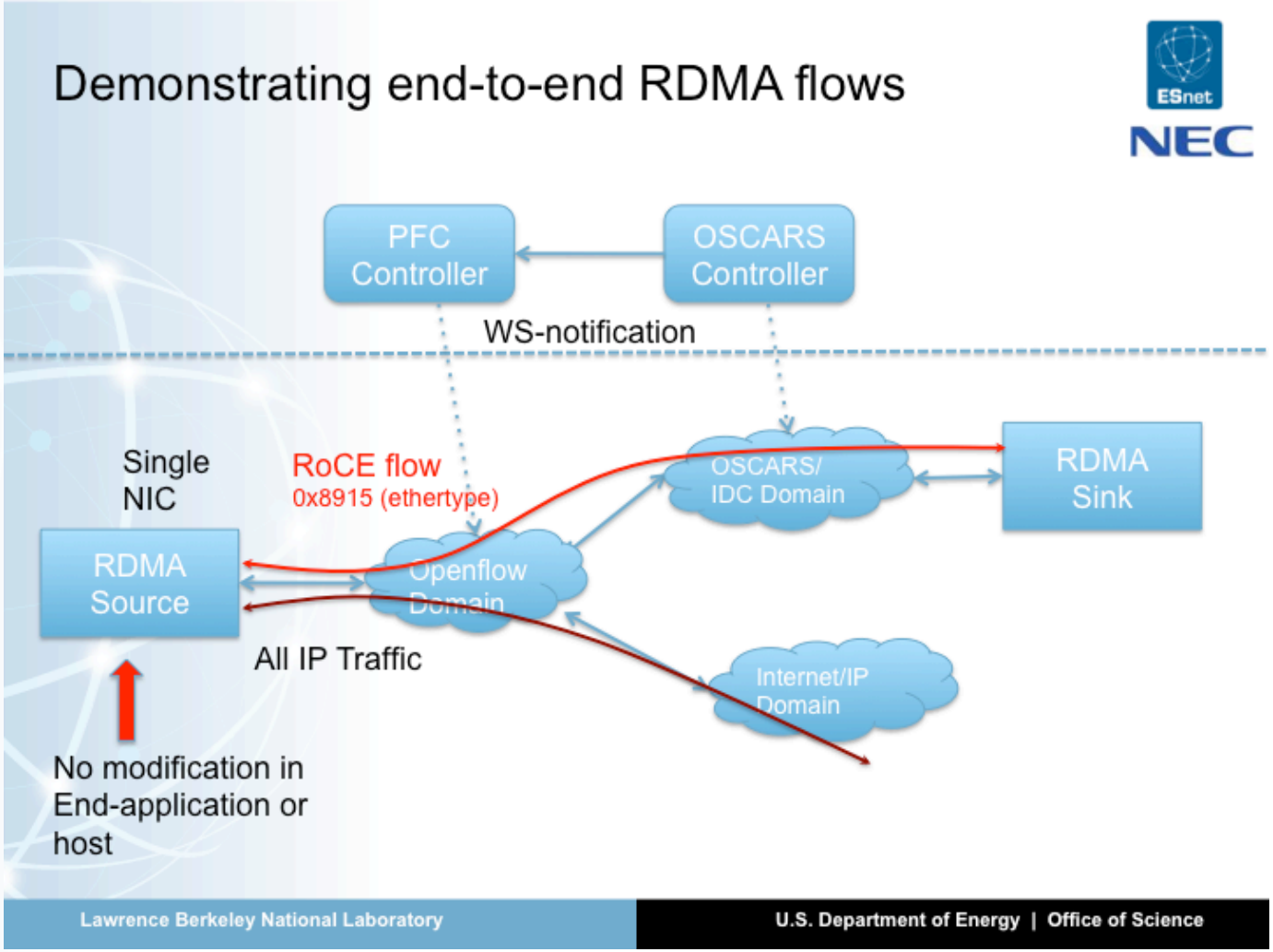
Who's using OSCARS

- Currently deployed in more than 40+ networks including wide-area backbones, regional networks, exchange points, local-area networks, and testbeds.
- In progress for an additional 11 green field deployments in 2012

Deployments using OSCARS v0.6 above dotted line

Journey with Programmability

Joint Techs Summer 2011, Fairbanks, Alaska



Insights

- SDN not immune from end-to-end problem
- 'unmodified end host' an attractive architecture

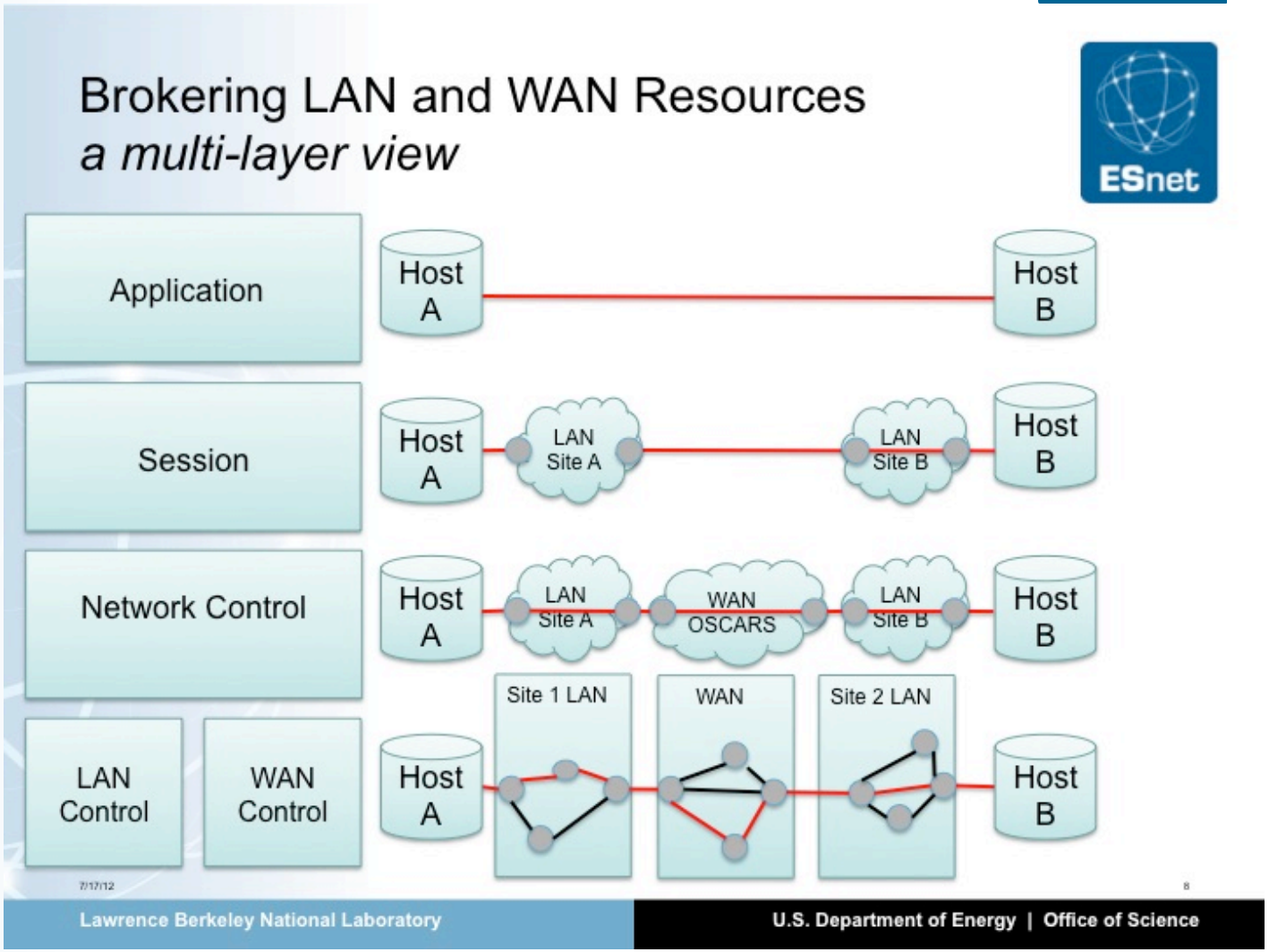
Journey with Programmability

Inaugural Open Network Summit, 2011 (Stanford) and SC 2011 (Seattle)

<u>Insights</u>

- end-to-end communication at each layer important

(even at layer 8)

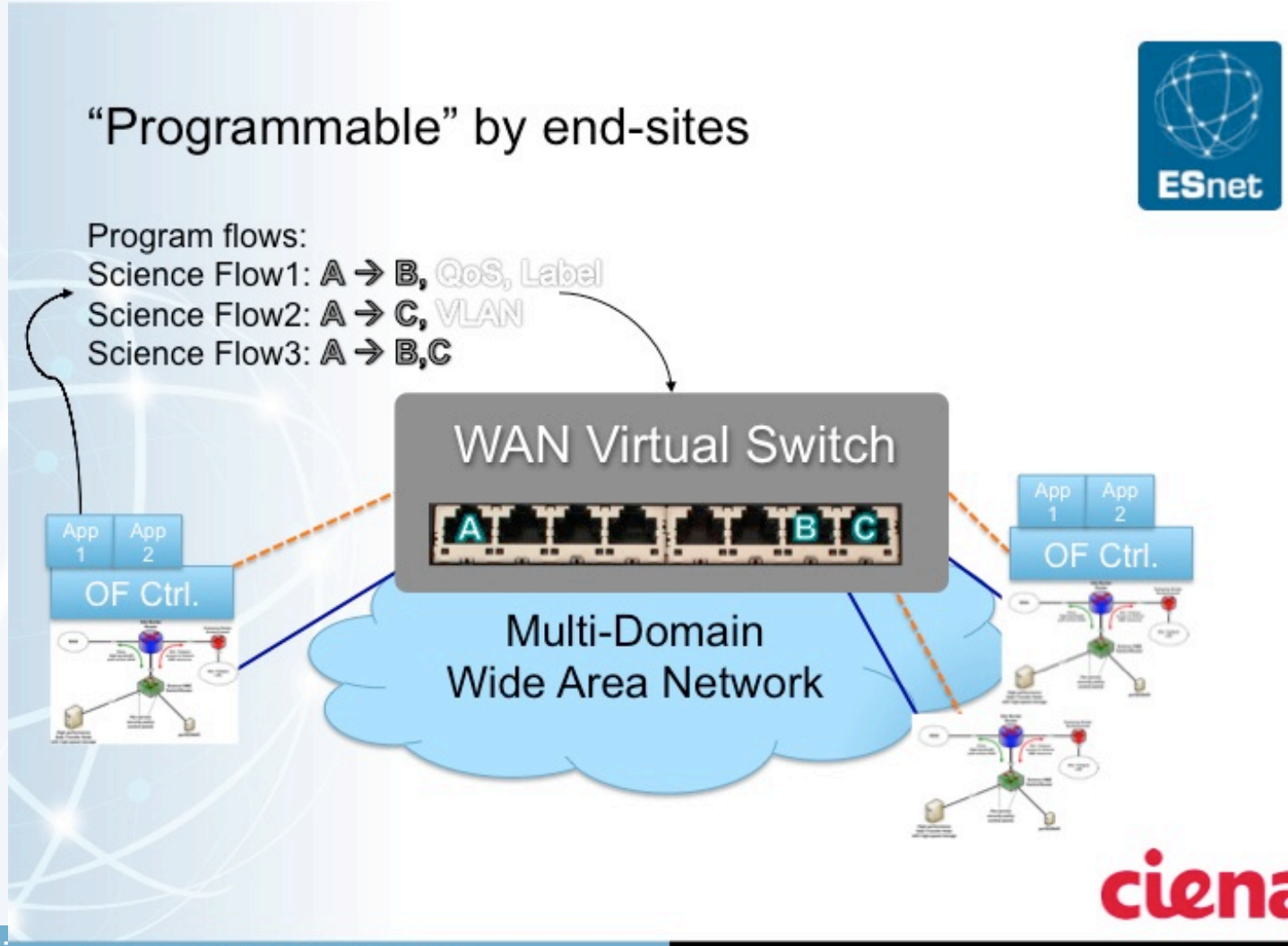


Journey with Programmability

SRS, Ciena, SuperComputing 2012, Salt Lake City

<u>Insights</u>

- 'virtual switch' abstraction in the WAN holds promise

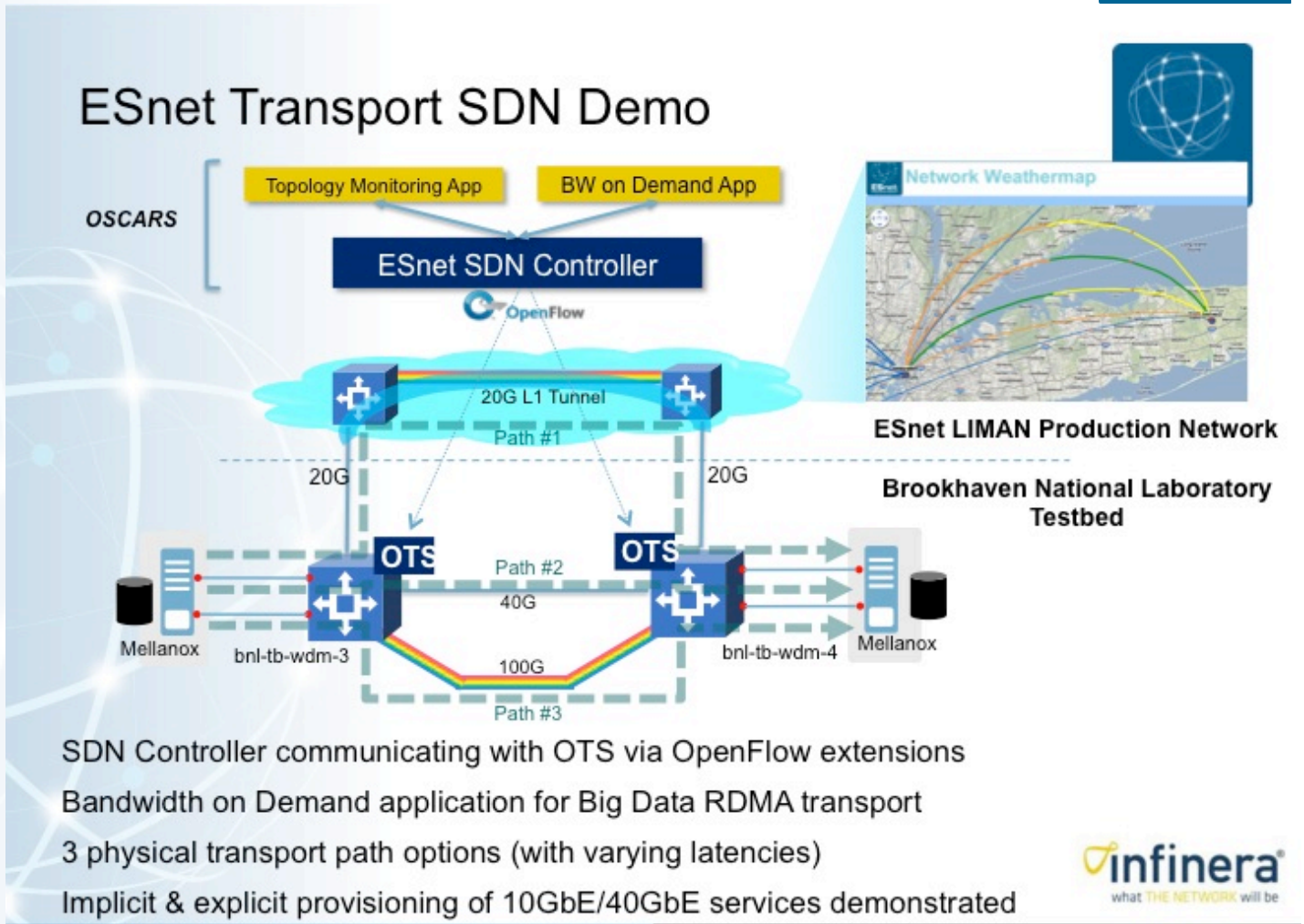


Journey with Programmability

World's first Transport SDN Demo, Infinera/ESnet/Brookhaven

<u>Insights</u>

- optical-layer automation essential for future topologies, architectures
  - 'SDN for Transport' now official ONF working group

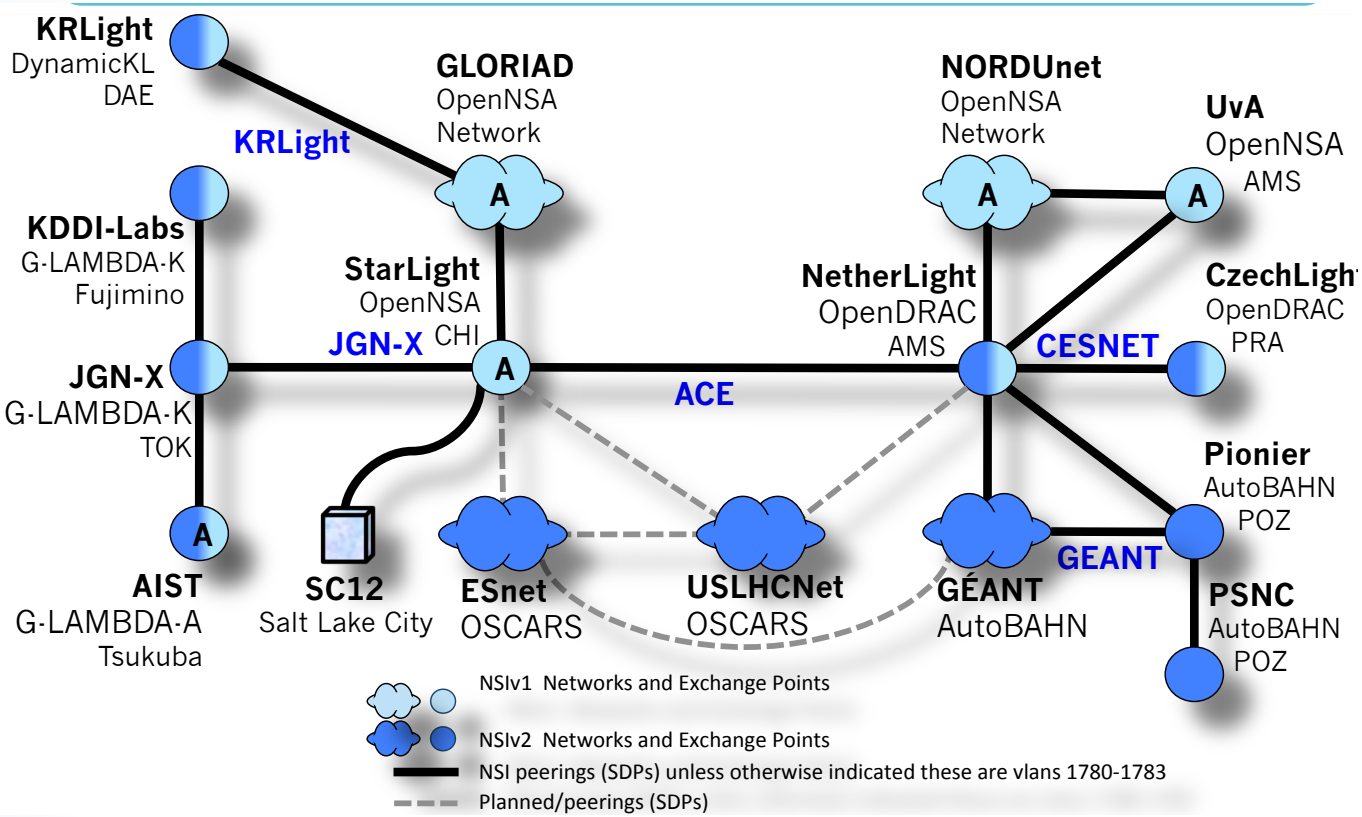


Journey with Programmability

Multi-Domain NSI 2.0 Demonstration and Tutorial, TIP 2013

Insights

- multi-domain interoperability and standards are key to wide adoption



Outline

ESnet Updates

Growth and Drivers

Programmability

Network Testbed and Research Questions

ESnet Research Testbeds

100G Testbed

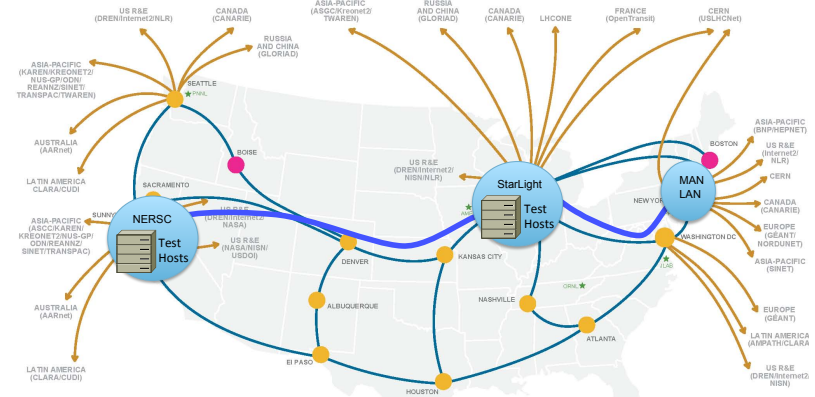
- High-speed protocol research
- Available since Jan 2012
- Dedicated 100G wave from Oakland to Chicago to NYC, plus programmable capacity on entire footprint.

OpenFlow Testbed

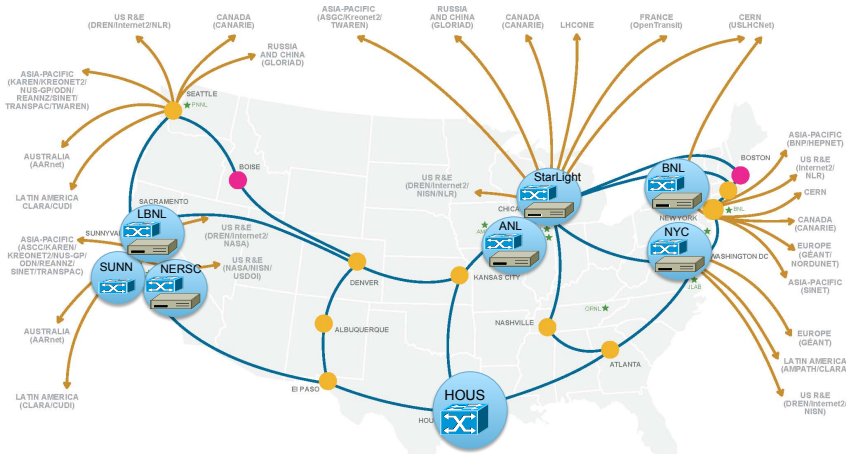
- 10G nationwide footprint

Dark Fiber Testbed

- Continental-scale fiber footprint for disruptive research



ESnet 100G Testbed



ESnet 10G OpenFlow Testbed

Publications Accepted

Yufei Ren, Tan Li, Dantong Yu, Shudong Jin, et. al, *Protocols for Wide-Area Data-intensive Applications: Design and Performance Issues*, Proceedings of IEEE Supercomputing 2012, November 12, 2012.

Zhengyang Liu, Malathi Veeraraghavan, Zhenzhen Yan, Chris Tracy, et.al, *On Using Virtual Circuits for GridFTP Transfers*, Proceedings of IEEE Supercomputing 2012, November 12, 2012.

Ezra Kissel, Martin Swany, Brian Tierney, and Eric Pouyoul, *Efficient Data Transfer Protocols for Big Data*, Proceedings of the 8th International Conference on eScience, IEEE, October 9, 2012.

Zhenzhen Yan, Chris Tracy, and Malathi Veeraraghavan, *A Hybrid Network Traffic Engineering System*, Proceedings of the IEEE 13th Conference on High Performance Switching and Routing (HPSR). June 2012, Belgrade, Serbia.

Mehmet Balman, Eric Pouyoul, Yushu Yao, E. Wes Bethel, Burlen Loring, Prabhat, John Shalf, Alex Sim, and Brian L. Tierney, *Experiences with 100Gbps Network Applications*, The Fifth International Workshop on Data Intensive Distributed Computing (DIDC 2012), June 2012.

Yufei Ren, Tan Li, Dantong Yu, Shudong Jin, and Thomas Robertazzi, *Middleware Support for RDMA-based Data Transfer,* Cloud Computing High-Performance Grid and Cloud Computing Workshop, May 2012.

G. Garzolglio, D. Dykstra, P. Mhashilkar, and H. Kim, *Identifying Gaps in Grid Middleware on Fast Networks with the Advanced Network Initiative*, International Conference on Computing in High Energy and Nuclear Physics (CHEP 2012), May 2012.

H. Pi, I. Sfiligoi, F. Wüerthwein, and D. Bradley, *Data Transfer Test with 100 Gpbs Network for Open Science Grid (LHC) Application*, International Conference on Computing in High Energy and Nuclear Physics (CHEP 2012), May 2012.

See: http://www.es.net/RandD/100g-testbed/results/

Industry Use of the Testbed

- Alcatel-Lucent used the testbed in May 2012 to verify the performance of its new 7950 XRS core router.
- Bay Microsystems used the testbed to verify that its 40 Gbps IBEx InfiniBand extension platform worked well over very long distances.
- Infinera used the testbed to demonstrate the telecommunication industry's first successful use of a prototype software-defined networking (SDN) open transport switch (OTS).
- Acadia Optronics used the testbed to test ITS 40 Gbps and 100 Gbps host NICs, and to debug the Linux device driver for its hardware.
- Orange Silicon Valley is using the testbed to test a 100G SSD-based video server
- Reservoir Labs is using the testbed to test their 100G IDS product under development

	ANI Testbed	Future Testbed
Who can use it?	DOE, .edu, industry.	DOE, .edu, industry.
Possible topologies?	Fixed and constrained by availability of physical circuits.	Programmable for each collaboration, through a combination of dedicated 100Gbps waves and OSCARS circuits on the spare capacity of the new 100Gbps ESnet footprint.
Attached resources?	Only those provided by ESnet.	Any resources provided on a temporary or permanent basis by ESnet, DOE supercomputing centers, labs, user facilities, universities, exchange points, or other network testbeds.
Bare-metal access to network components and systems?	Yes.	Yes (on ESnet-managed resources).
OpenFlow capability	Yes, with 1Gbps NEC switches.	Yes, with multi-vendor deployment on national footprint; optical OpenFlow possible.
Connectivity to other national-scale testbeds such as GENI, and those in Europe?	No.	Yes, with peering at the Starlight exchange point in Chicago.
Access to 13,000-mile dark fiber footprint for disruptive optical research?	Yes, but researchers must supply optical components and cover associated costs.	Yes, but researchers must supply optical components and cover associated costs.

Site Resources Currently Committed

NERSC

- 8 DTN nodes are connected to testbed (2x10G each) and to NERSC production GPFS
- All NERSC resources could be connected

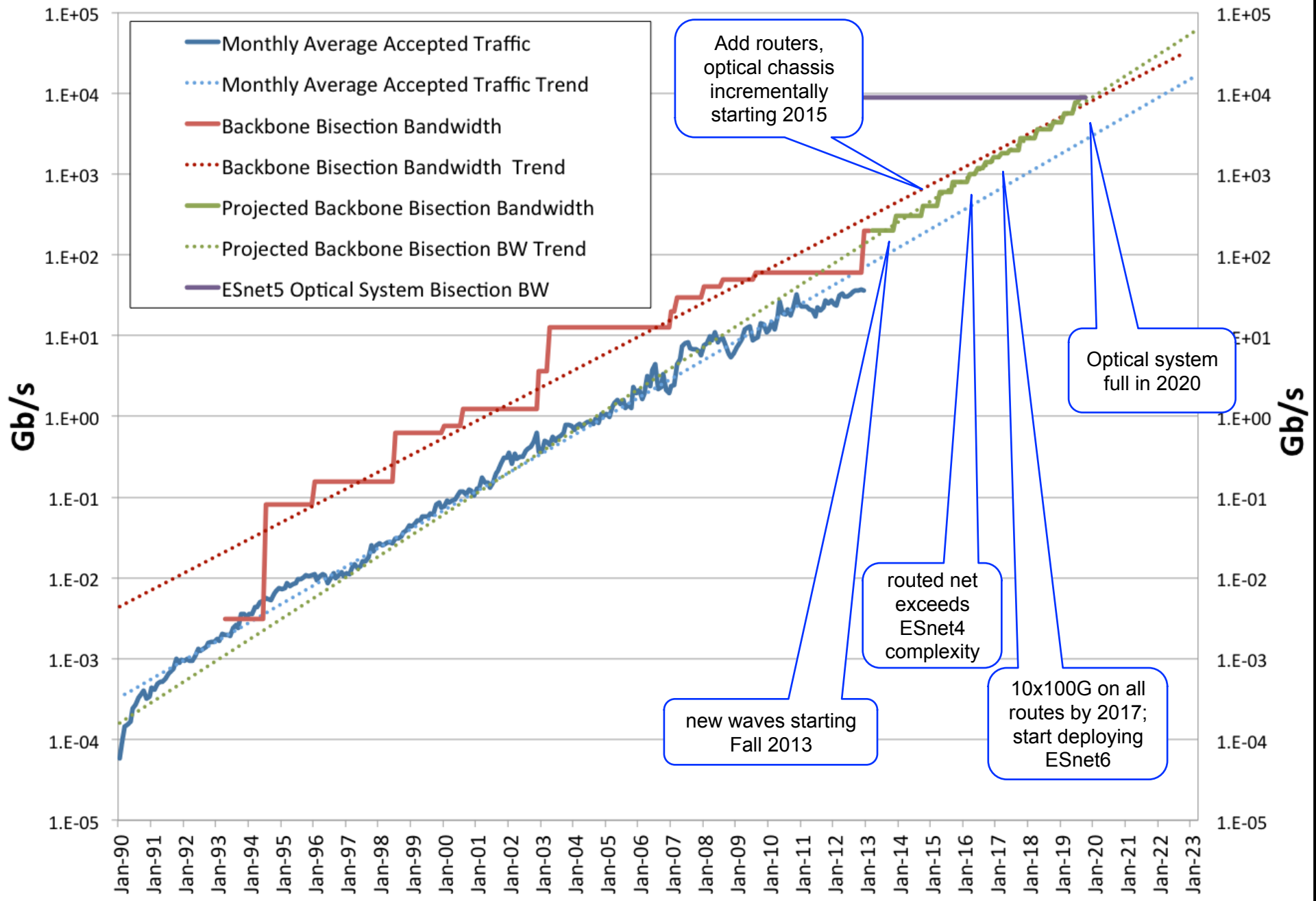
BNL

- 2 40G PCI gen3 hosts
- Juniper MX router with OpenFlow

FNAL

- 50Gbps of the 100G production network
- Nexus 7000 w/ 2-port 100GE module / 6-port 40GE module / 10GE copper module
- 6 nodes w/ 10GE Intel X540-AT2 (PCIe) / 8 cores / 16 GB RAM
- 2 nodes w/ 40GE Mellanox ConnectX®-3 / 8 cores w/ Nvidia M2070 GPU

ESnet Traffic vs Backbone Capacity



A Few Research Questions

- As data transport needs continue to grow, what is the optimal role for parallelism?
- What management challenges will arise with >100 waves?
- How can networks best support emerging remote-IO paradigms?
- How can software-defined networking minimize our capex and opex?
- As networks become programmable, what core services and abstractions will be most useful to applications and middleware?
- Can real-time inspection of n x 100G waves lead to useful discovery, correlation, visualization, or security outcomes?
- How can we develop a global, community-curated knowledgebase for recording science and CI-related metadata for IP ranges?
- How can Named Data Networking (NDN) be used in big data context?
- How can ESnet resources such as the testbed be used to validate models to simulate complex distributed systems?
- How can can ESnet make the vast quantities of monitoring data it collects (flow, perfSONAR, etc) more useful to network researchers?

Recruiting for Network Research Postdoc

ESnet Network Research Postdoctoral Fellow-75692

Organization:SN-Scientific Networking

Description

Are you an exceptional network researcher who likes working on truly challenging projects? Are you a strong software developer? Are you passionate about learning and open minded about the way that networks are built? Consider joining the research and development team for Berkeley Lab's Scientific Networking Division. At the core of the Scientific Networking Division is ESnet, the Energy Sciences Network. ESnet interconnects the US national laboratory system, is widely-regarded as a technical pioneer, and is currently the fastest science network in the world. ESnet's Advanced Network Technologies Group (ANTG) has an immediate opening for a postdoctoral researchers to work on research in projects in Software Defined Networking and other network research topics related to "Big Data".

In the upcoming year, the Scientific Networking Division will embark on an important challenge: expanding its program in applied research, development and integration. With the vision to be a pioneer in developing innovative network technologies, the Division is seeking an exceptionally-competent, flexible and innovative network researcher that is willing to think beyond the conventional. We are working at the leading edge of software-defined networking, OpenFlow, dynamic network infrastructure, network visualization, network knowledge plane, multi-domain and multi-layer architectures. The successful candidate will be the one that brings strong and diverse coding skills, focus, and ability to work with a fast-paced team.

SPECIFIC RESPONSIBILITIES:

- Research on a range of topics in high-speed networking
- Help explore the role of Software Defined Networking for Big Data
- Help with conference papers and funding proposals
- Help design and write software to demonstrate proof-of-concept tools
- Help with cutting-edge demonstrations of advanced networking concepts
- Work closely with collaborators from both Universities and other DOE labs.
- Track next generation network technologies and provide expert advice on when a technology is ready for use by a project in ESnet

https://lbl.taleo.net/careersection/2/jobdetail.ftl?lang=en&job=75692