1986–2016
ESnet
30 YEARS OF NETWORKING
AT THE SPEED OF SCIENCE

Making way for the Elephants

Inder Monga
Executive Director, ESnet
Division Director, Scientific Networking
Lawrence Berkeley National Laboratory

CHEP
San Francisco
October 14th, 2016

U.S. DEPARTMENT OF ENERGY
Office of Science

BERKELEY LAB

1986–2016

ESnet

30 YEARS OF NETWORKING

AT THE SPEED OF SCIENCE

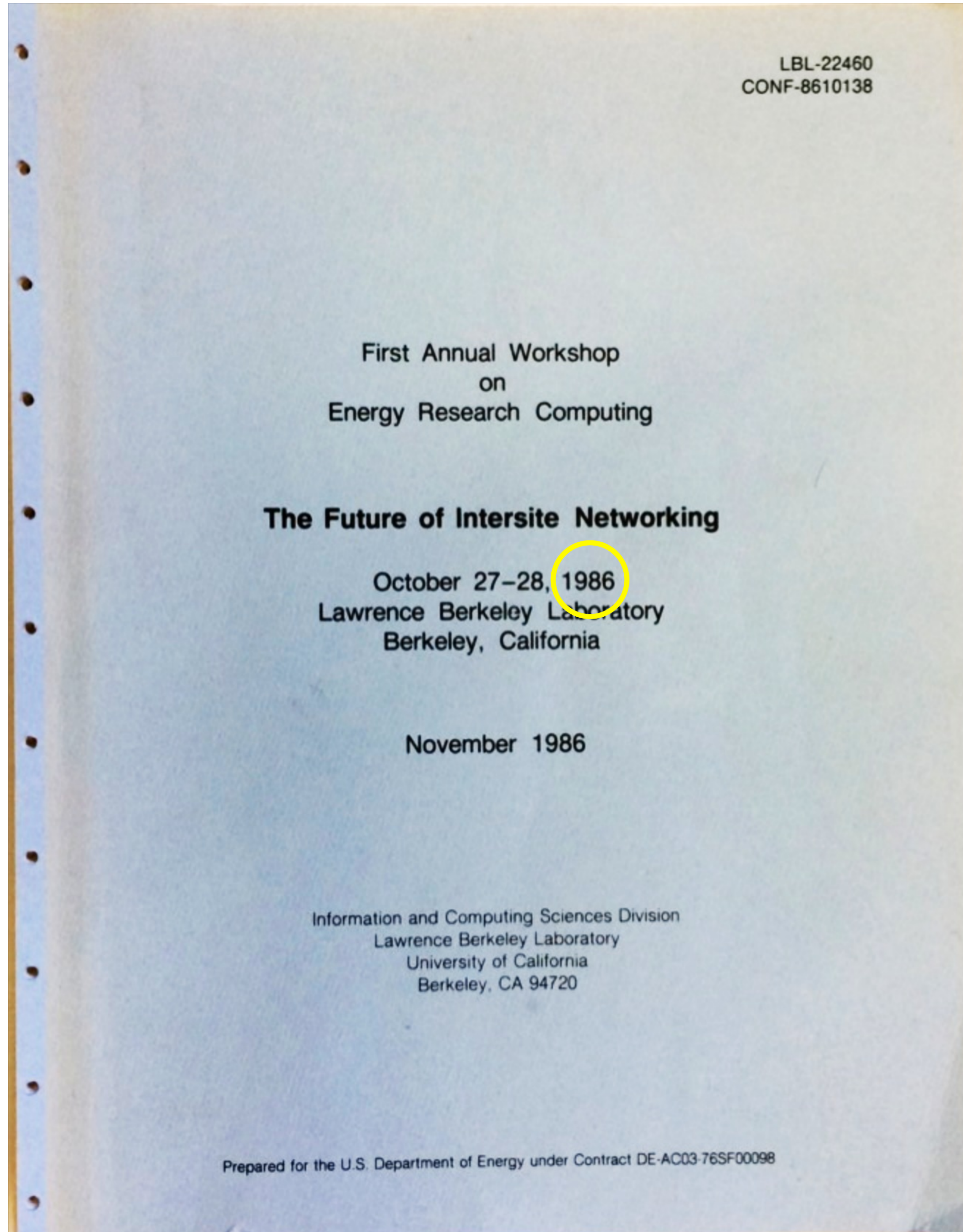
1986–2016
ESnet
30 YEARS OF NETWORKING
AT THE SPEED OF SCIENCE

Retrospective view

IBM PC Convertible
released April 3, 1986

1986–2016
ESnet
30 YEARS OF NETWORKING
AT THE SPEED OF SCIENCE

HEP and ESnet are inextricably linked



LBL-22460
CONF-8610138

First Annual Workshop
on
Energy Research Computing

The Future of Intersite Networking

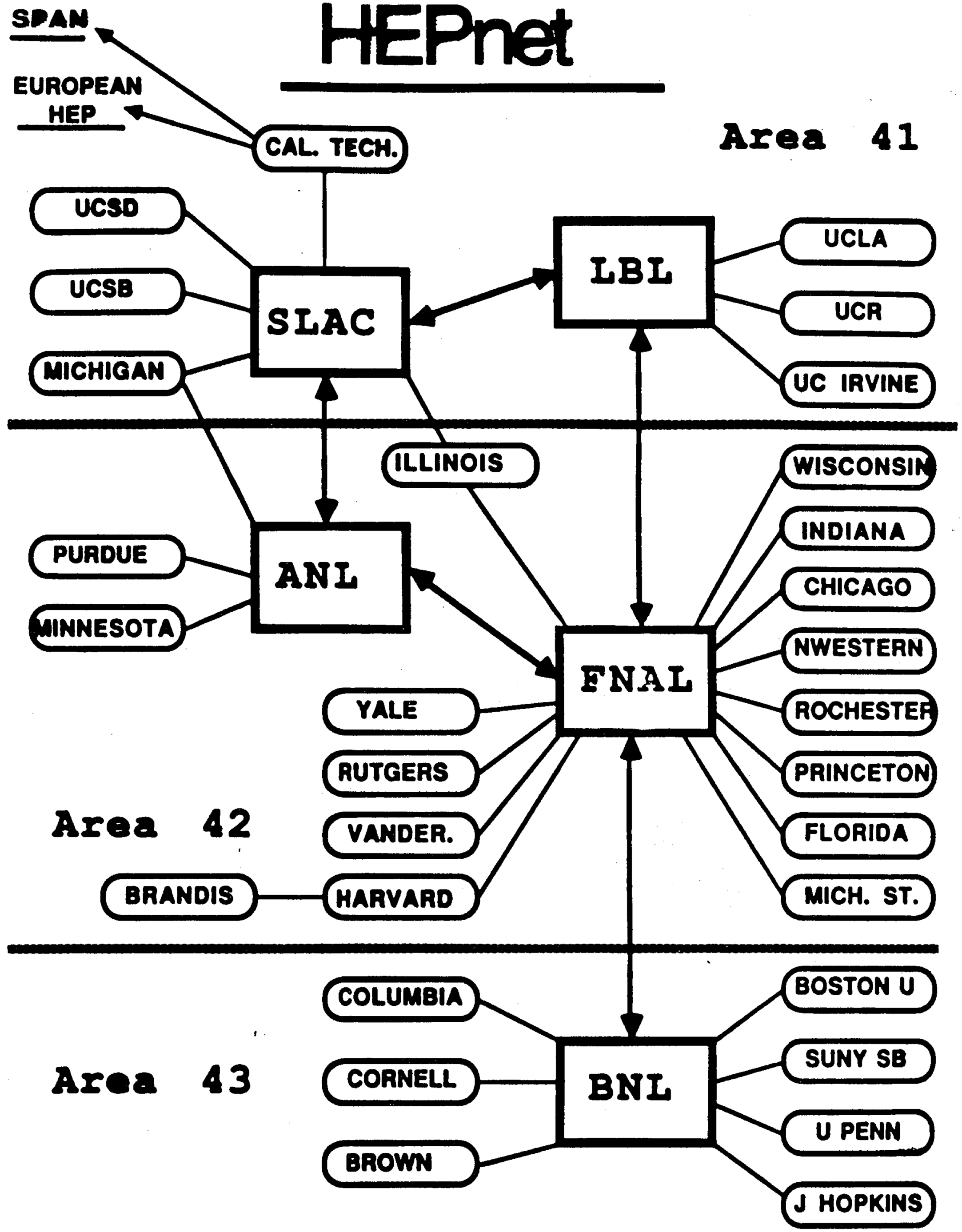
October 27–28, 1986
Lawrence Berkeley Laboratory
Berkeley, California

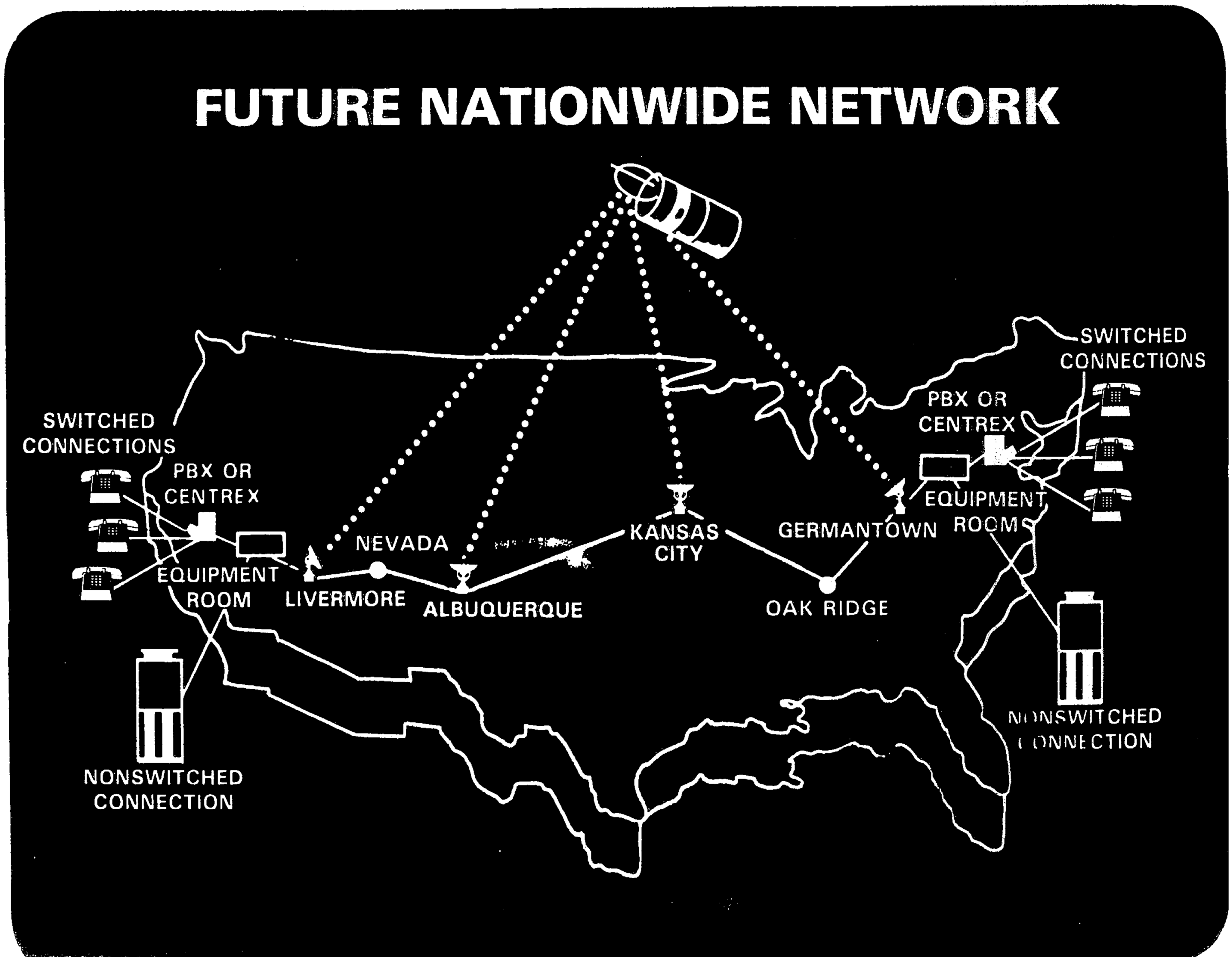
November 1986

Information and Computing Sciences Division
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

HEPnet

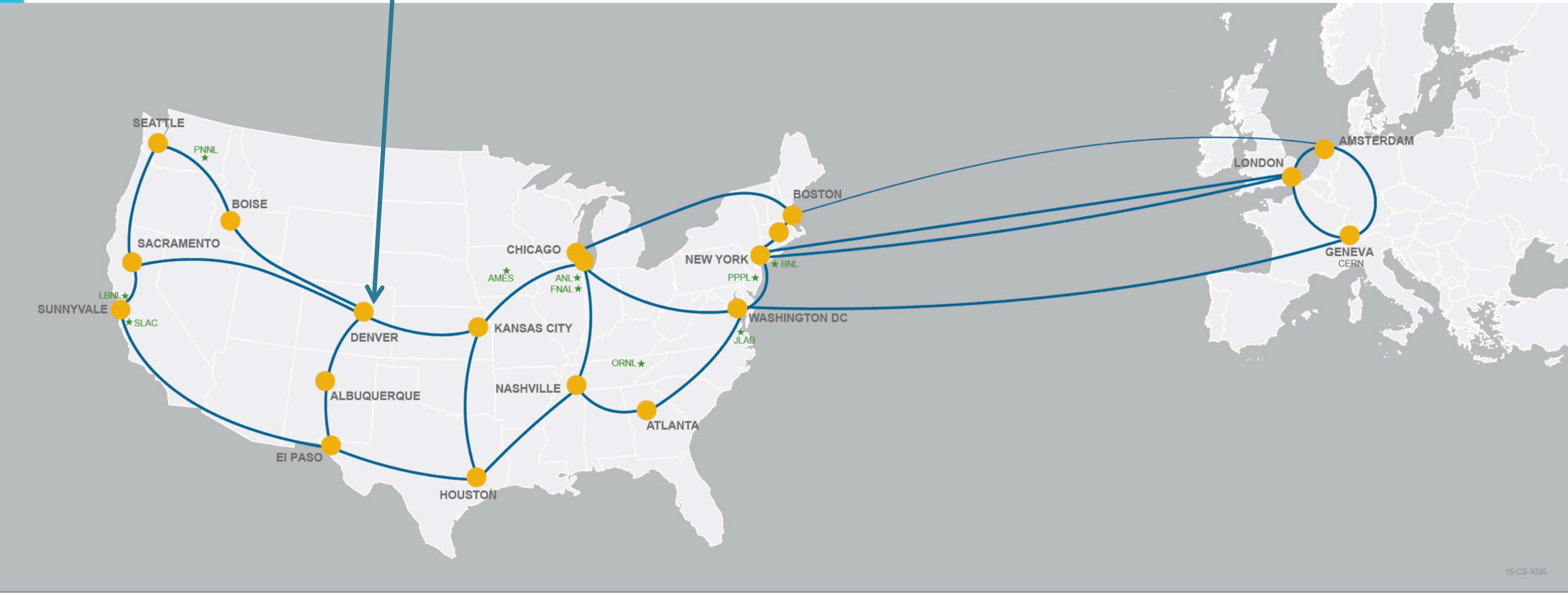

SPAN
EUROPEAN HEP
CAL. TECH.
Area 41
UCSD
UCSB
MICHIGAN
SLAC
LBL
UCLA
UCR
UC IRVINE
ILLINOIS
WISCONSIN
INDIANA
CHICAGO
NWESTERN
ROCHESTER
PRINCETON
FLORIDA
MICH. ST.
PURDUE
MINNESOTA
ANL
FNAL
YALE
RUTGERS
VANDER.
HARVARD
BRANDIS
Area 42
COLUMBIA
CORNELL
BROWN
BNL
BOSTON U
SUNY SB
U PENN
J HOPKINS
Area 43


FUTURE NATIONWIDE NETWORK
SWITCHED CONNECTIONS
PBX OR CENTREX
EQUIPMENT ROOM
NONSWITCHED CONNECTION
LIVERMORE
NEVADA
ALBUQUERQUE
KANSAS CITY
GERMANTOWN
OAK RIDGE
SWITCHED CONNECTIONS
PBX OR CENTREX
EQUIPMENT ROOMS
NONSWITCHED CONNECTION

The Present

DOE's Energy Sciences Network (ESnet):

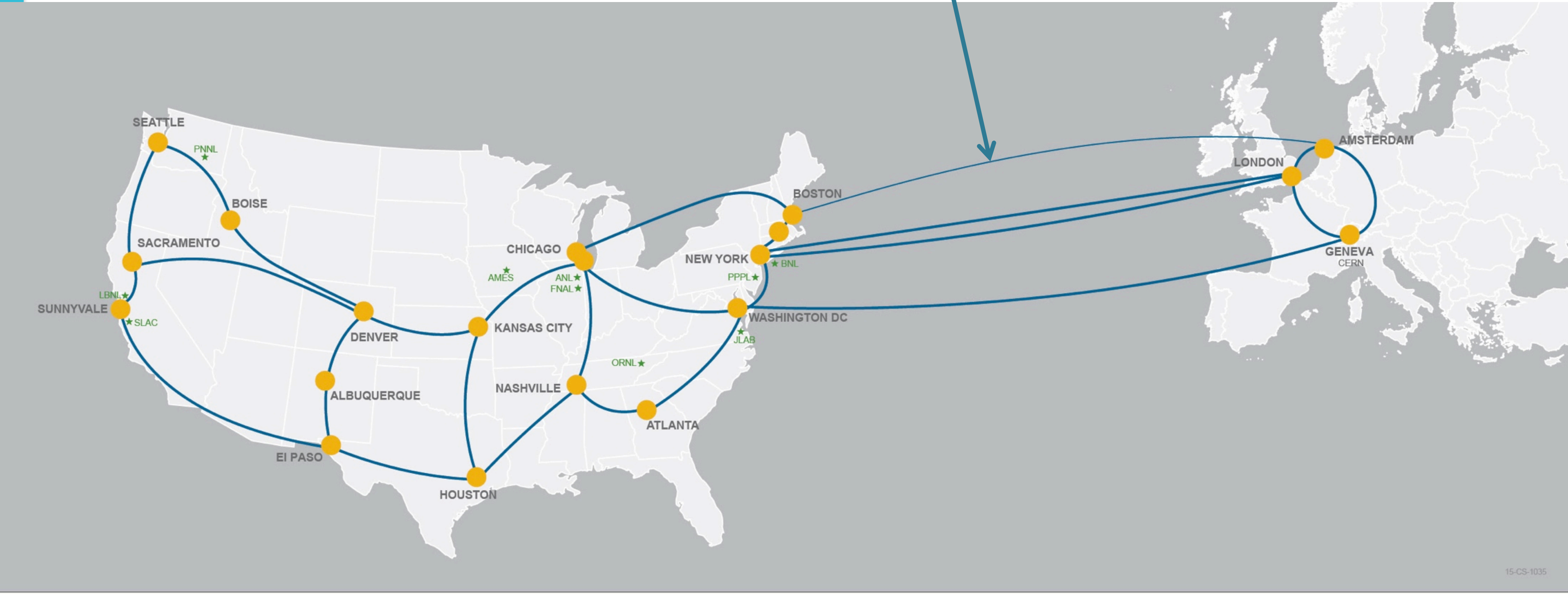
Fiber spectrum ~ 4.4 Tb of spectral capacity, 100Gbps wavelengths



ESnet 1986–2016 30 YEARS OF NETWORKING AT THE SPEED OF SCIENCE

DOE's Energy Sciences Network (ESnet):

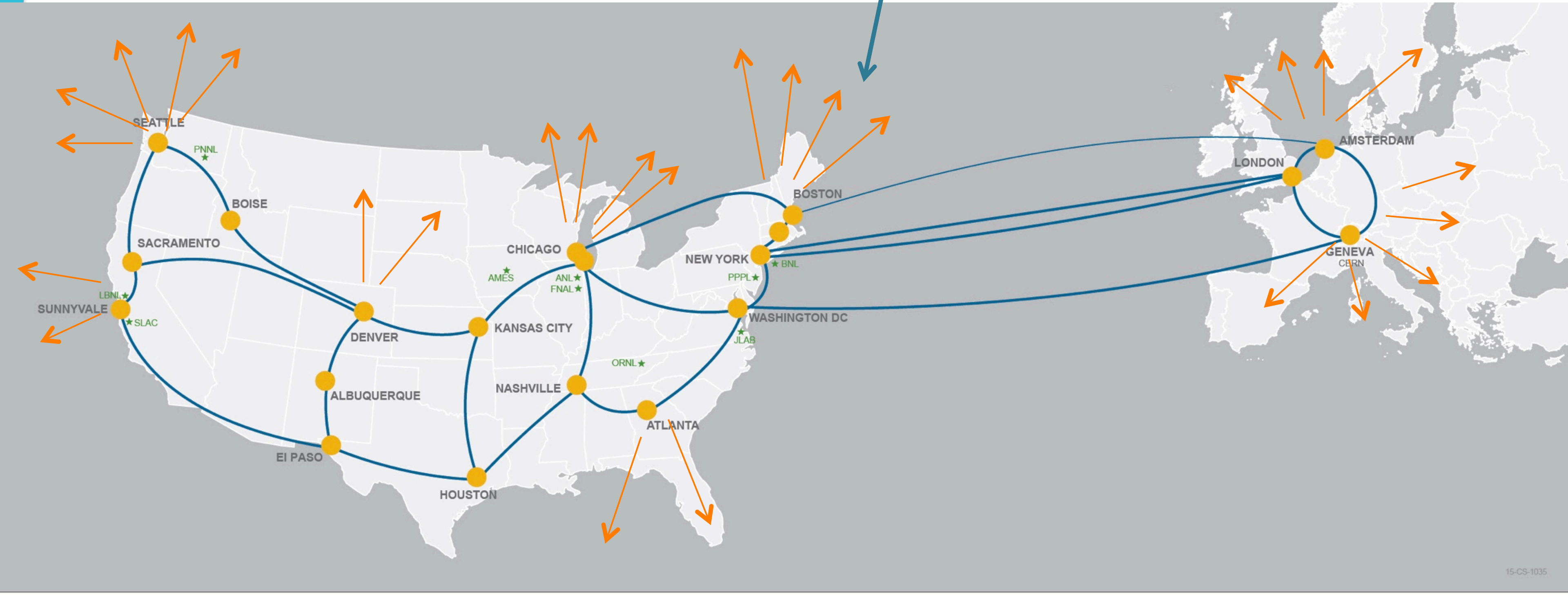
340Gbps transatlantic.



1986–2016
ESnet
30 YEARS OF NETWORKING
AT THE SPEED OF SCIENCE

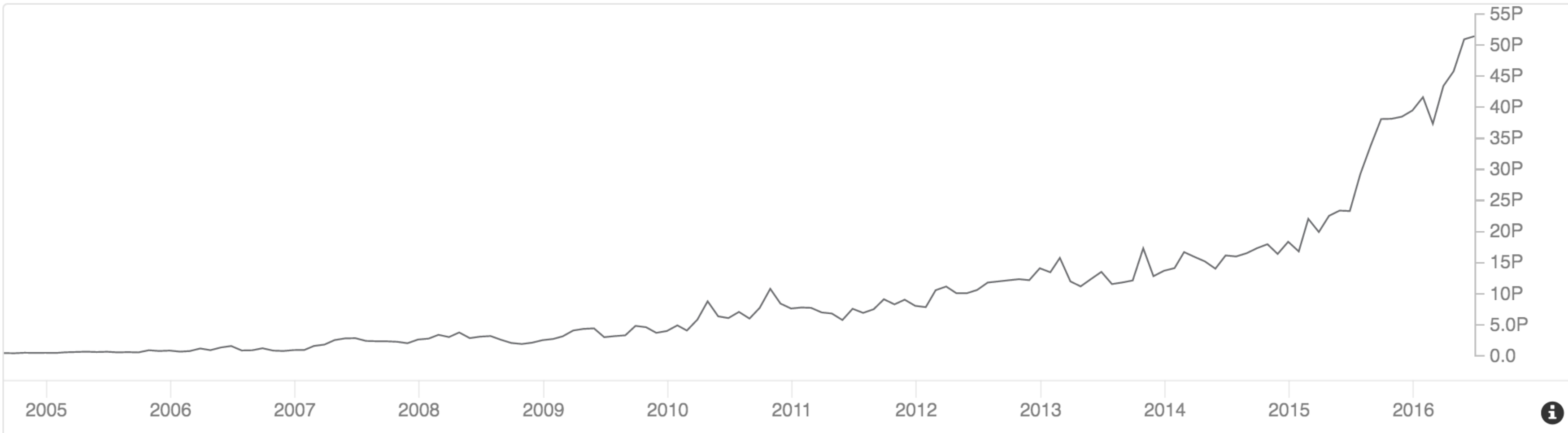
DOE's Energy Sciences Network (ESnet):

150 peers, 1.5 Tbps peering capacity.



1986–2016 ESnet 30 YEARS OF NETWORKING AT THE SPEED OF SCIENCE

Blistering rate of traffic growth (~70% yoy)

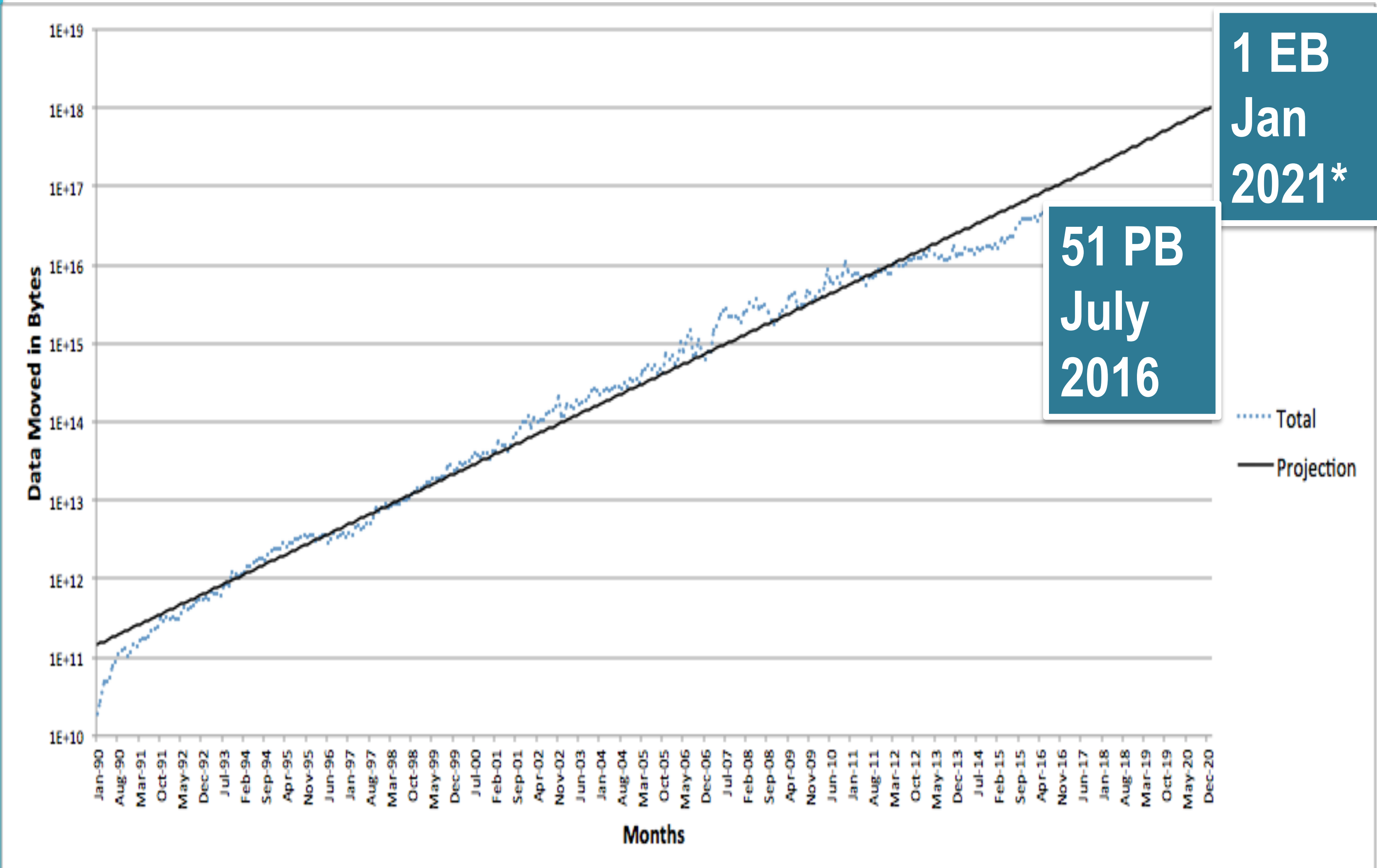


July 2016

	Bytes	Percent of Total	One Month Change	One Year Change
OSCARS	12.32PB	24.0%	-12.1%	+104%
LHCONE	18.41PB	35.9%	+4.75%	+277%
Normal traffic	20.56PB	40.1%	+7.08%	+68.1%
Total	51.29PB		+0.973%	+121%

1986–2016 ESnet 30 YEARS OF NETWORKING AT THE SPEED OF SCIENCE

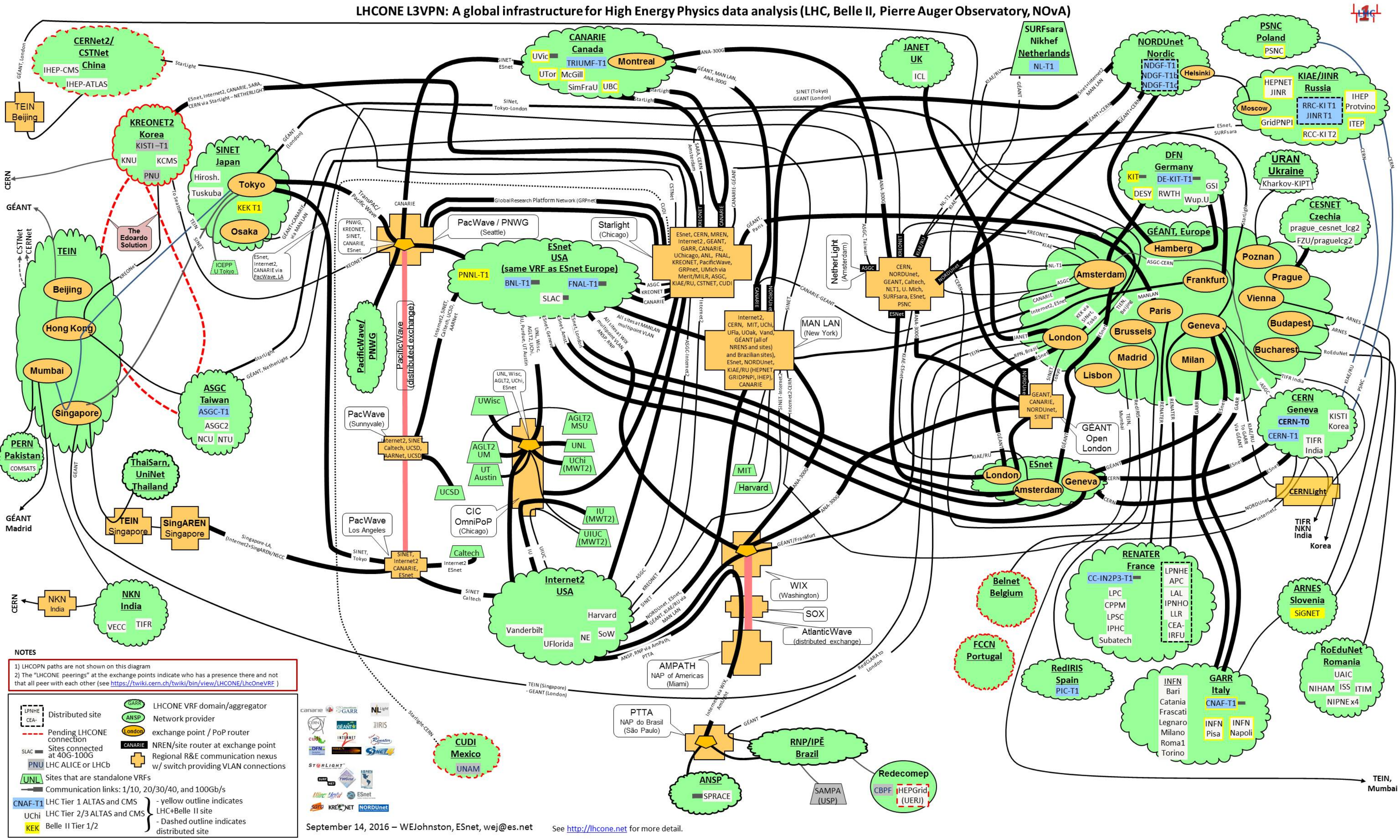
ESnet: An exascale facility in 2021



LHC Open Networking Environment (LHCONE):
How do the *networkers* support a global science effort

- An overlay network serving the LHC community
- Closed private network with governance and a security models
  - LHC compute and storage resources accessible
- Network operators can identify and segregate LHC traffic from general R&E traffic.
- Enables monitoring and management benefit the LHC community.
  - For example adding capacity, or routing it across different sets of links
- Architecture
  - R&E network providers offer LHCONE services
  - The networks on each continent exchange traffic with each other following their normal peering and customer/provider relationships.
  - A core mesh of intercontinental networks exchange LHCONE traffic with each other following standard business relationships as well.

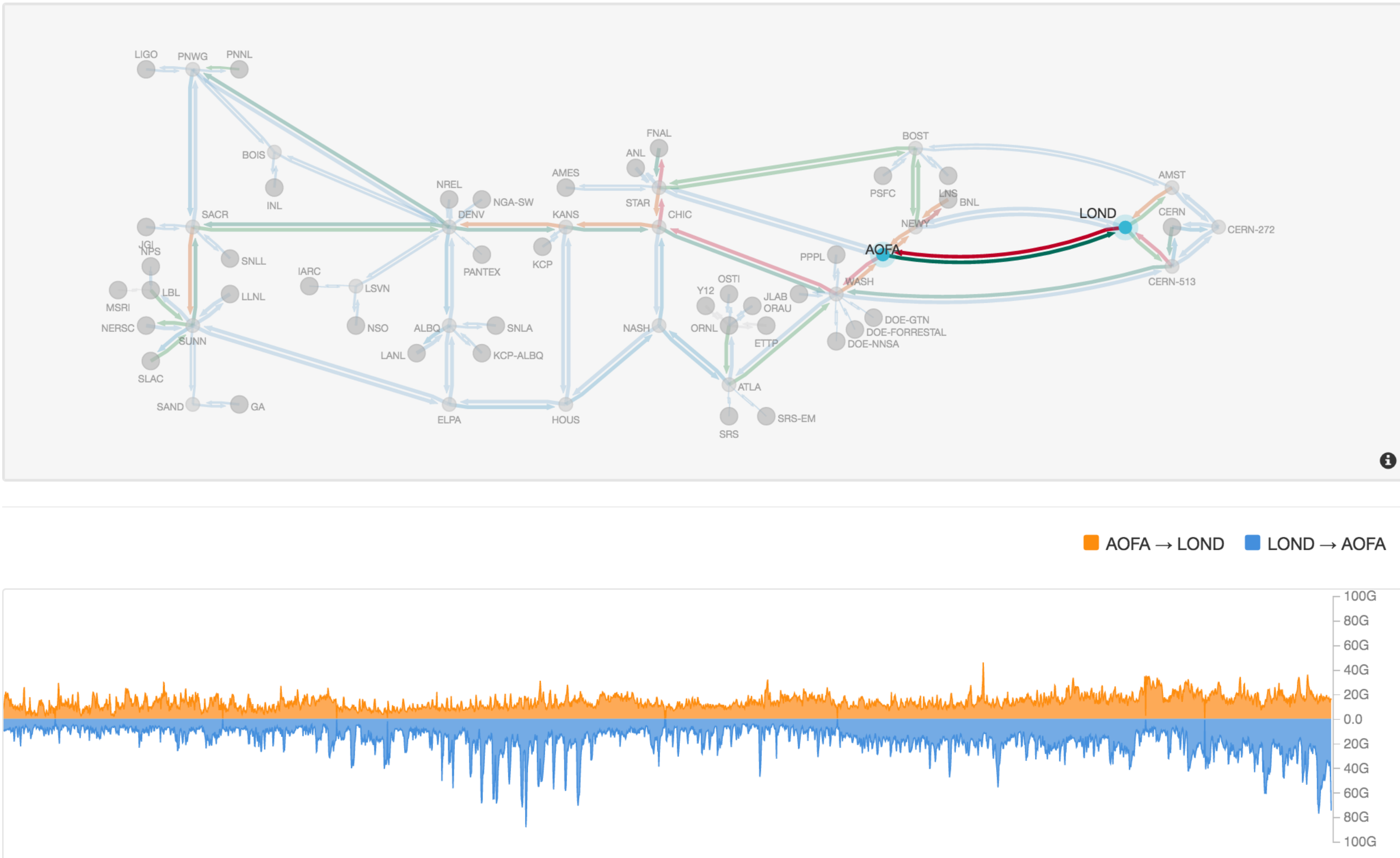
LHCONE: Dense international interconnectivity



ESnet peers with Universities for LHCONE traffic

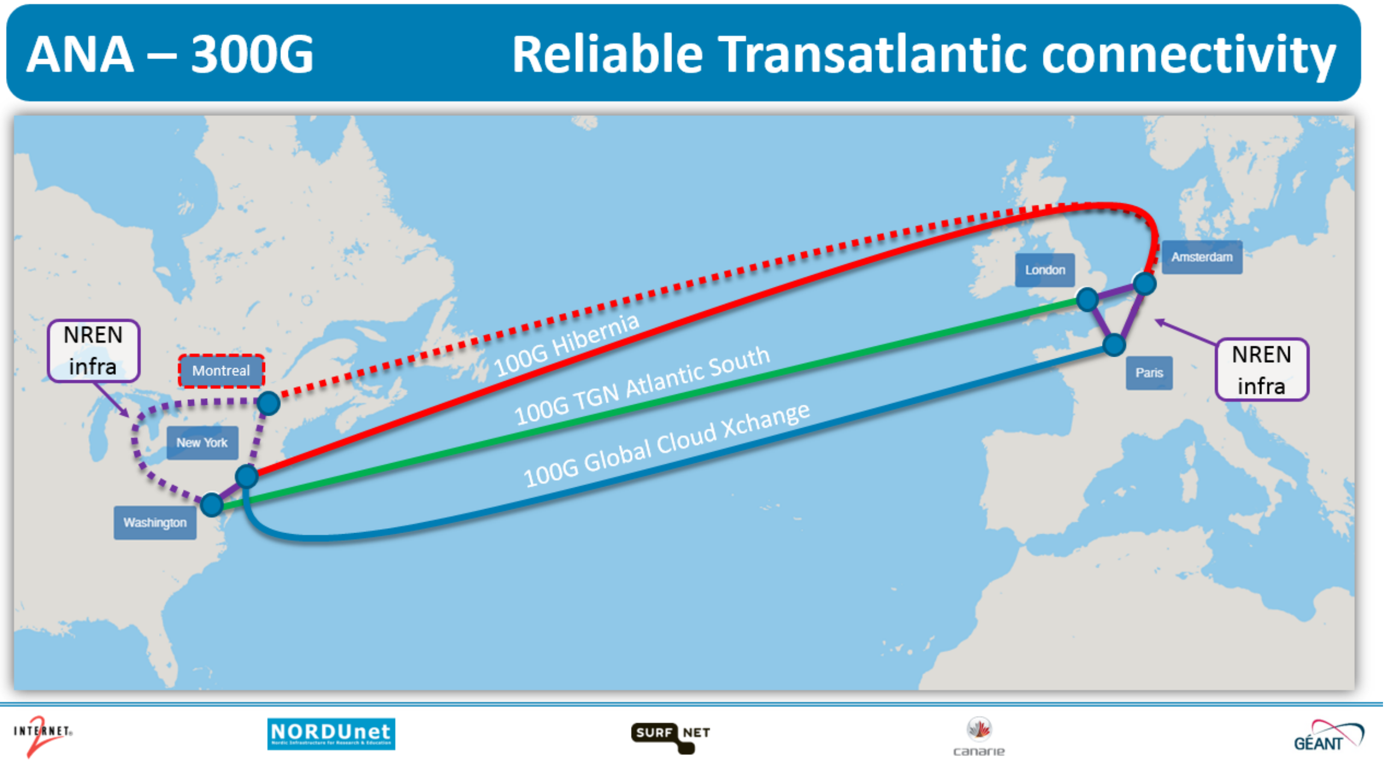
https://my.es.net/collaborations/lhcone/

Transatlantic traffic is healthy (70 G peaks)



NETWORKING
AT THE SPEED OF SCIENCE

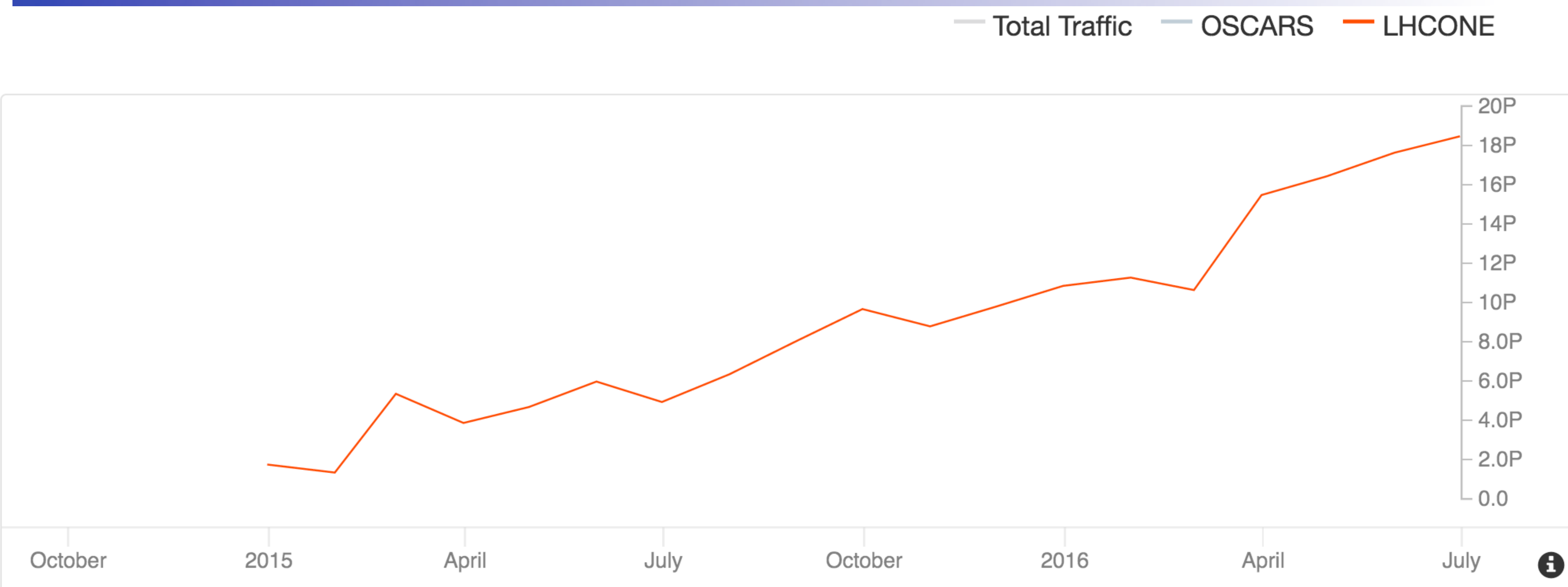
Continued Coordination with other Trans-Atlantic efforts



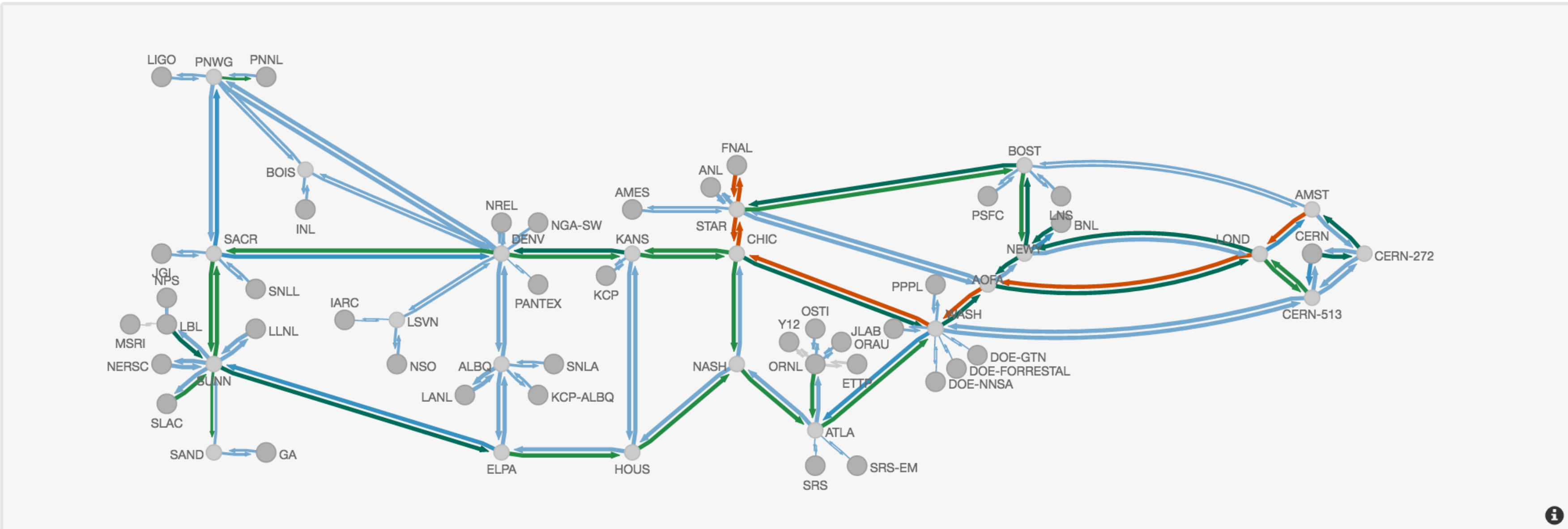
IU, partners lead $3.25M grant to improve data-sharing in Africa, North America and Europe

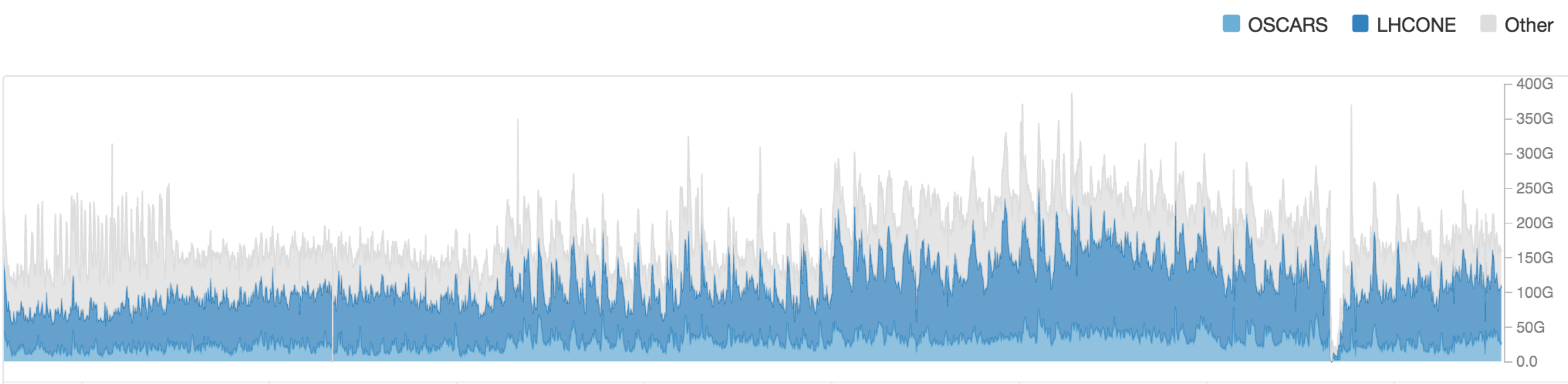
LHCONE ramps up with Run 2

From 1.7 PB in December 2014
To 18.4 PB in July 2015



Significant portion of ESnet's traffic (35%)





Belle-II joining LHCONE infrastructure, Data Challenges promise success

Data Challenge 2015-2016 : Summary

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Overall Summary

Source →	KEK	PNNL	DESY	KIT	CNAF	NAPOLI	SIGNET
KEK		4.6 Gbps	-	-	-	-	-
PNNL	4 Gbps		6.6 Gbps	6 Gbps	9 Gbps	8 Gbps	-
DESY	3 Gbps	6.6 Gbps		8 Gbps	8 Gbps	8* Gbps	3 Gbps
KIT	3.5 Gbps	> 4* Gbps	6-8 Gbps		6-8 Gbps	6 Gbps	3 Gbps
CNAF	-	8 Gbps	10 Gbps	6 Gbps		8* Gbps	3 Gbps
NAPOLI	3* Gbps	3* Gbps	3* Gbps	3 Gbps	3* Gbps		3* Gbps
SIGNET	0.8 Gbps	0.6* Gbps	5 Gbps	2-5 Gbps	5 Gbps	2* Gbps	

Columns show **upload** bandwidth

Rows show **download** bandwidth

LHCONE Network

* Site had storage/SRM related problems that were tackled later.

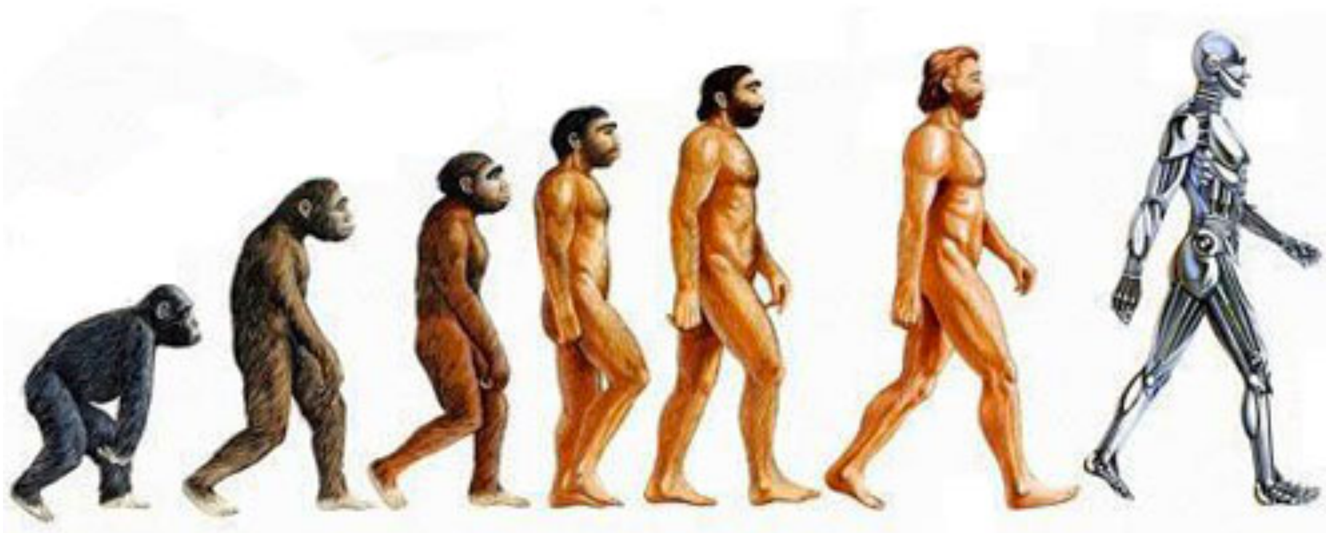
Feb. 2, 2016 - Belle II Network Data Challenge Results   Vikas Bansal   31

Two significant upgrades since test:

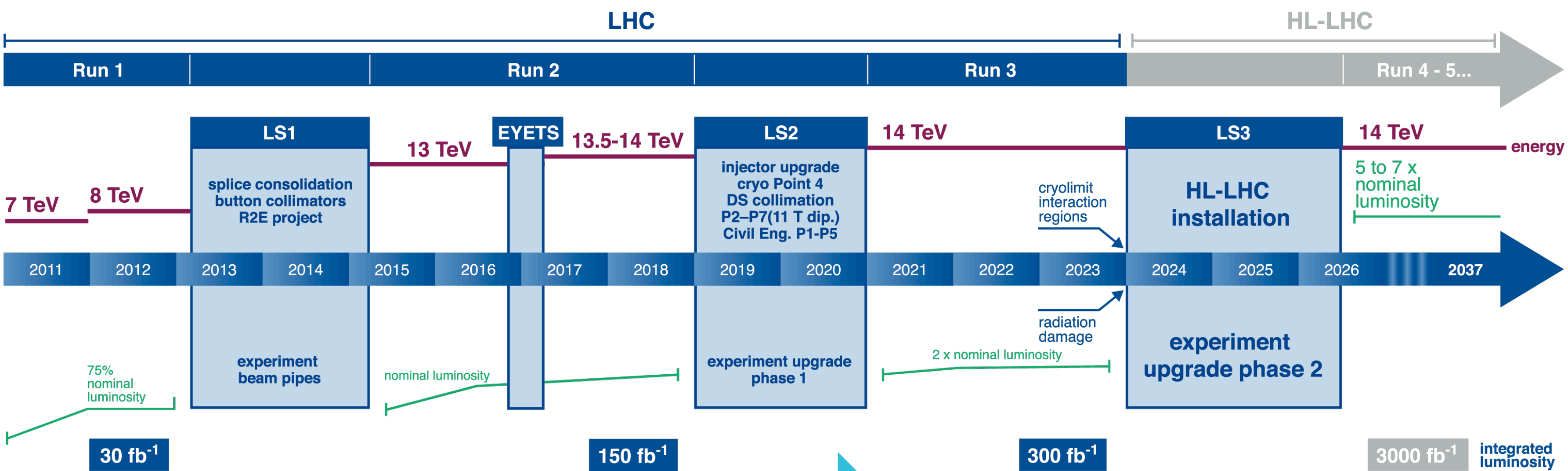
- ✓ 100G SINET (Japan-US)
- ✓ 2x10G PNNL

1986–2016 ESnet 30 YEARS OF NETWORKING AT THE SPEED OF SCIENCE

The Future



LHC and ESnet present and future



2011 - 2021

ESnet5

2014 -

European Extension

2016-2021

ESnet6 Design and Implementation

2021-

ESnet6 (*)

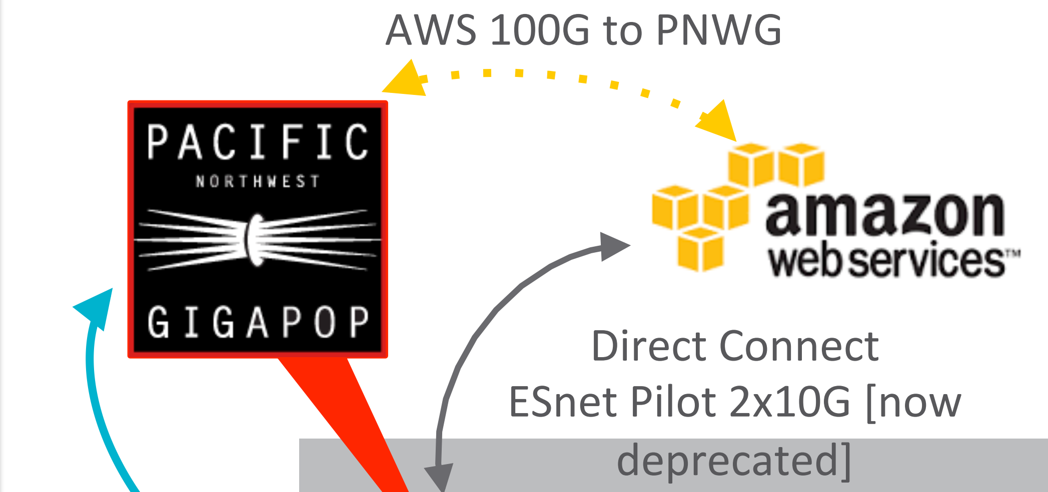
1986–2016
ESnet
30 YEARS OF NETWORKING
AT THE SPEED OF SCIENCE


CHALLENGE

Enabling the use of Cloud resources

How much investment needed for capable networking to all Cloud providers, <u>worldwide</u>

What is the right network architecture? For one NREN or across Global NRENs?



ESnet ENERGY SCIENCES NETWORK

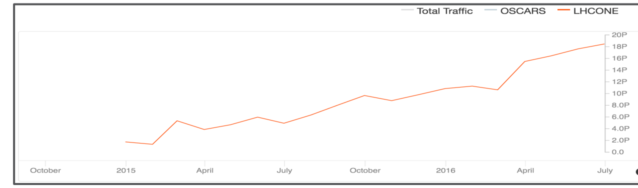
NETWORKING AT THE SPEED OF SCIENCE

Human beings have a hard time with exponentials

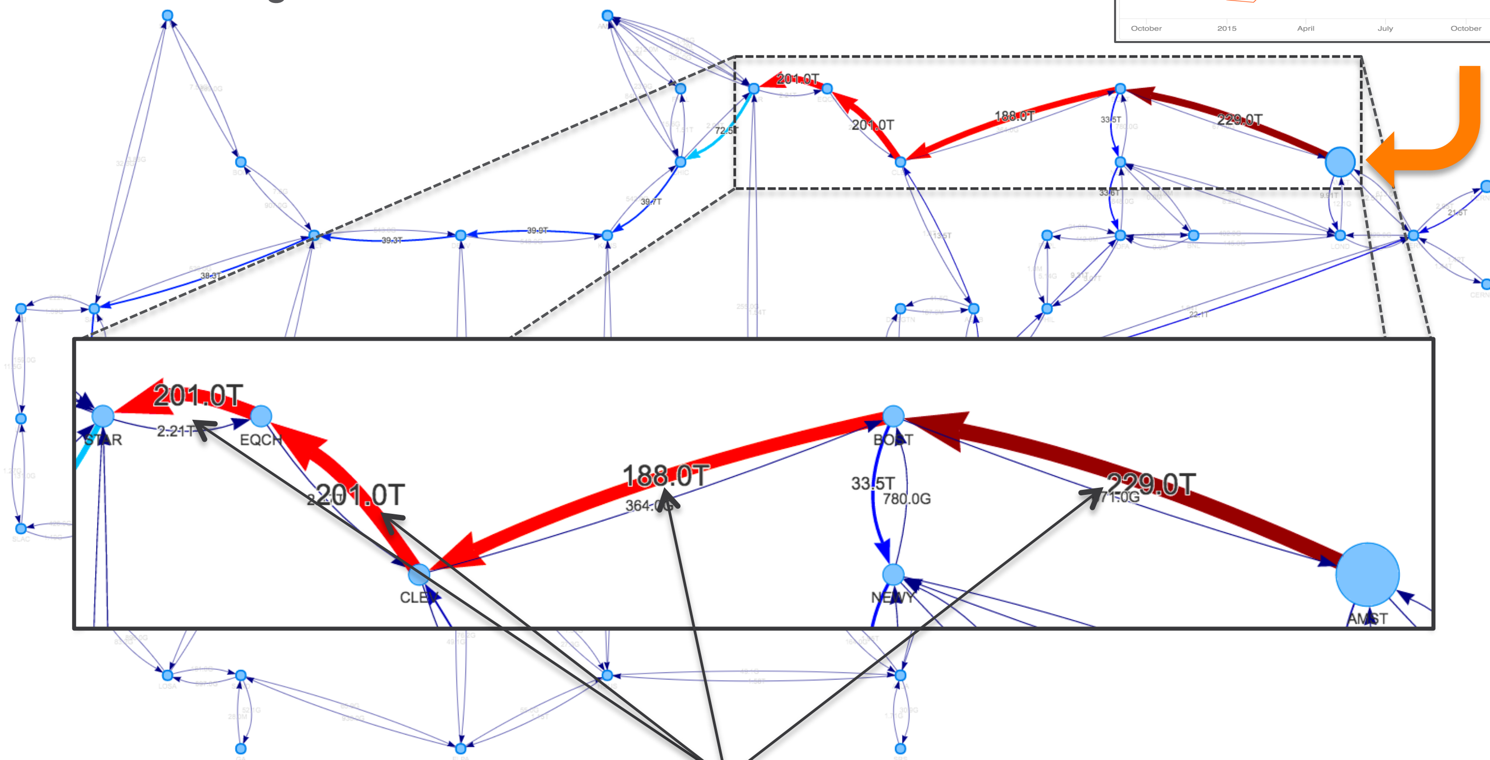
30 YEARS OF NETWORKING
AT THE SPEED OF SCIENCE

Long term modeling and capacity prediction continues to be a challenge

LHCONE ramps up ~10x in 8 months | **From 1.7 PB in December 2014 To 18.4 PB in July 2015**

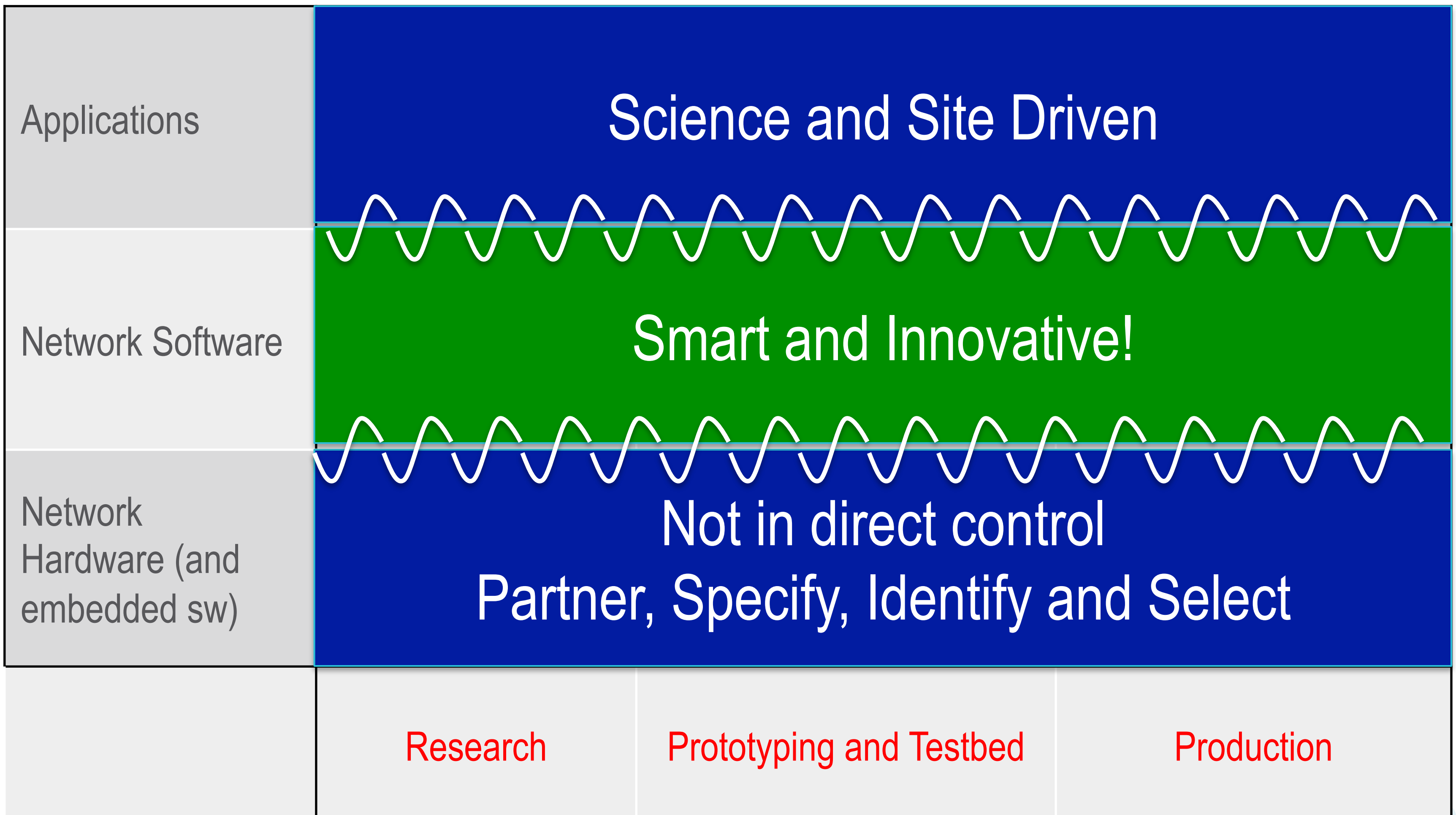


Predicted Usage Map in Jan 2030 with normalized growth rates



100+Tbps speeds at long-haul distances on a single fiber pair is outside the existing optical technology envelope

Software needed create ‘smart’ networks: Requires co-design with Science Applications/Teams

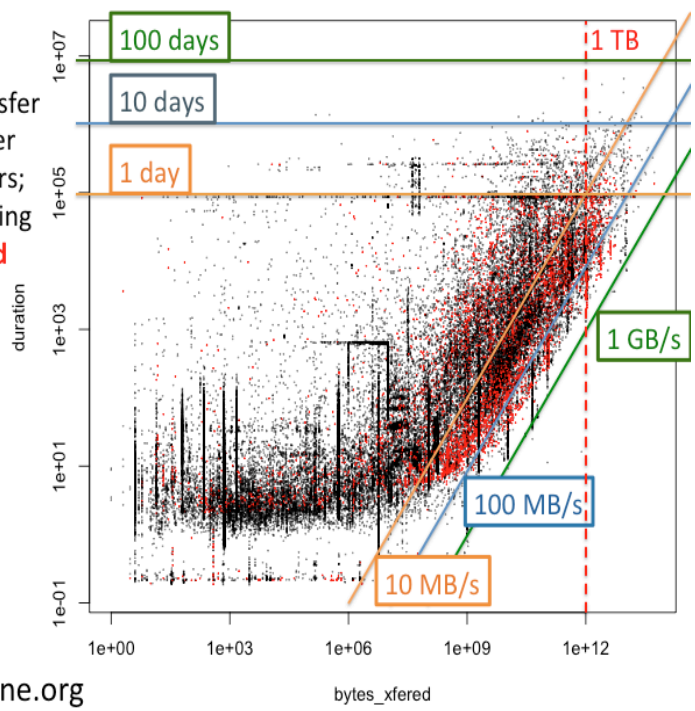


Grand Challenge: Predictable Network Transfers at scale

- Transfers over a shared network are not predictable
- Best-effort delivery can also mean worst-effort delivery

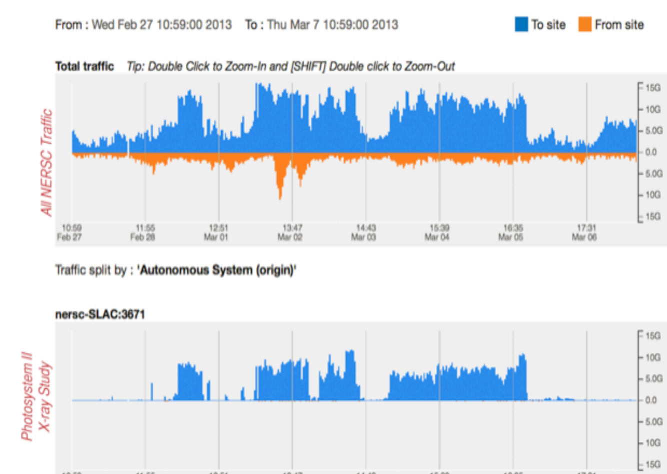
- Science workflows depend heavily on **end-to-end** data movement infrastructure, e.g. deadline scheduling, remote control, etc

Globus Transfer requests over last two years; those involving NERSC in **red**



globusonline.org

LCLS ↔ ESnet ↔ N



ESnet 30 YEARS OF NETWORKING AT THE SPEED OF SCIENCE

How do we (try to) keep up with the Physicists?

Research
Prototyping
Testbed

Production

Funded Research

Next-Generation Architectures

Network Software

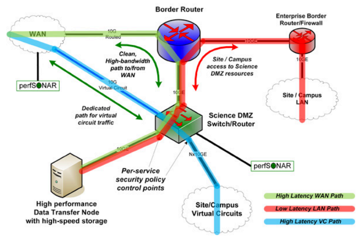
Tools Development

1986–2016 ESnet 30 YEARS OF NETWORKING AT THE SPEED OF SCIENCE

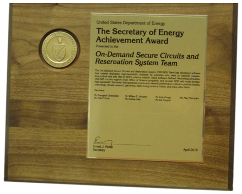
Contributions to the Community (examples)

perfSONAR

OSCARS



2013 R&D 100



Science DMZ

iPerf3

SENSE: SDN for End-to-end Networked S

Inder Monga [Lead-PI] (ESnet), Phil Demar (FNAL), Harvey Ne
Linda Winkler (ANL), Tom Lehman (UMD/MAX), Damain Haze
Mar 2016 – Feb 2019

- **Collaborators**: FERMI, ANL, Caltech, UMD/Max, NERSC, ESnet
- **Vision**: Enable National Labs and Universities to request and provision end-to-end intelligent network services for their application workflows

Goal

- Leverage the emerging Software Defined Network (SDN) ca
to-end, science networking architecture friendly to data-inte
applications

Impacts

- Present geographically distributed resources (datacenters, instruments, etc.) as components of a local facility
- Simplifies complex massive datasets distribution with coordinated, multi-domain, smart and secure services
- Enable seamless application-network interaction for new near real-time distributed computing and data analytics

SENSE SDN Control Plane Architecture for End-to-End Orchestration

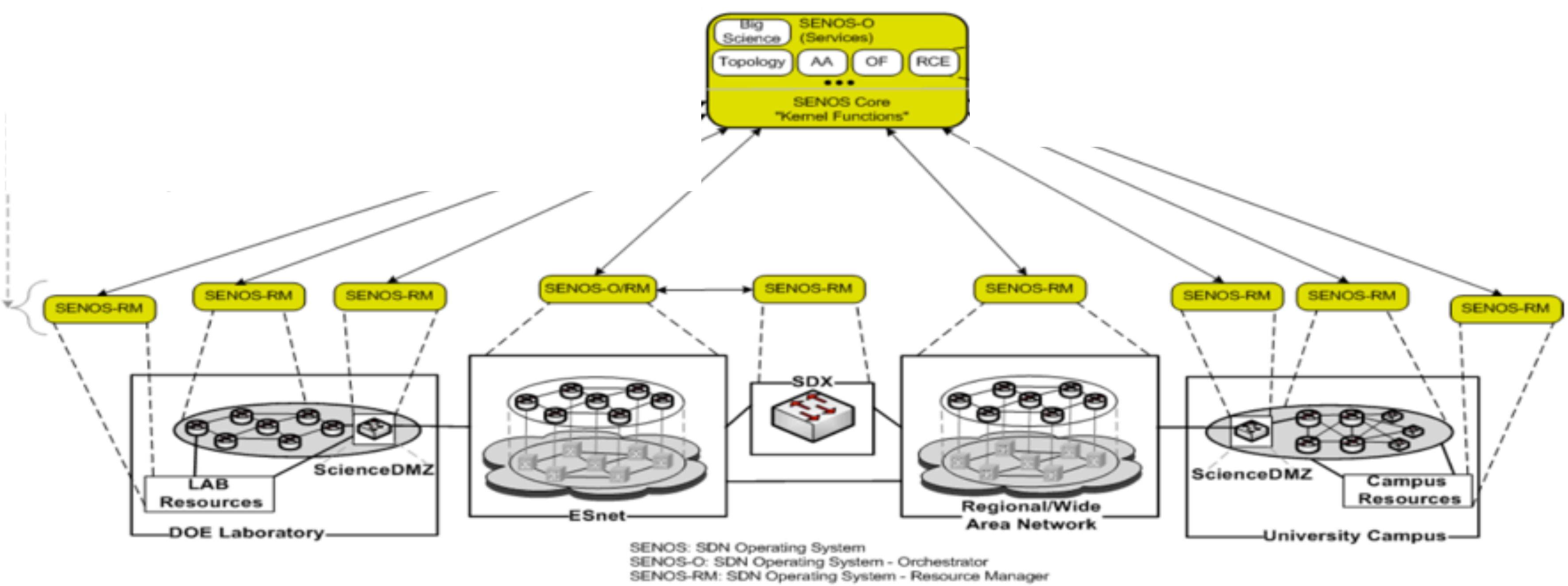
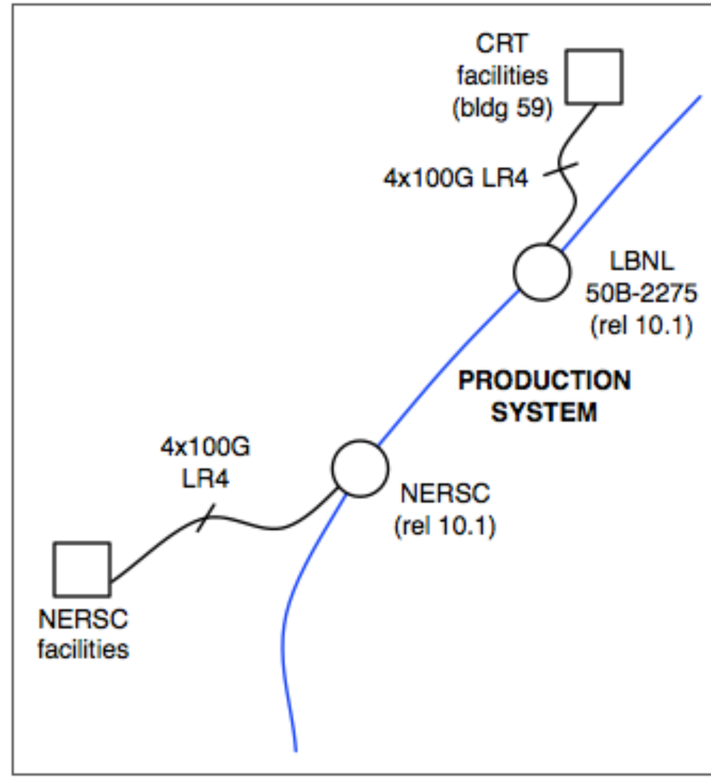


Figure 1. SENOS End-to-End Orchestration

First Production 400G Service on ESnet5

Goal: Deploy and harden a 400G production service (4x100 GigE), perform applications testing, production run.

- Two new wavelengths were provisioned, 200G per wave (2x100 GigE payload)
- Wavelength Selectable Switches (WSSs) are in the path, but are limited to 50 GHz granularity.
- On BayExpress, the production 400G circuit consumed 100 GHz of spectral bandwidth
  - 2 adjacent 50 GHz channels
  - Comes as close to a "super channel" as possible in production



1986–2016 ESnet 30 YEARS OF NETWORKING AT THE SPEED OF SCIENCE

ESnet Upgrade
Desired ~ 2021/2022, CD-0 in progress

Reference: #WOCinTech Chat

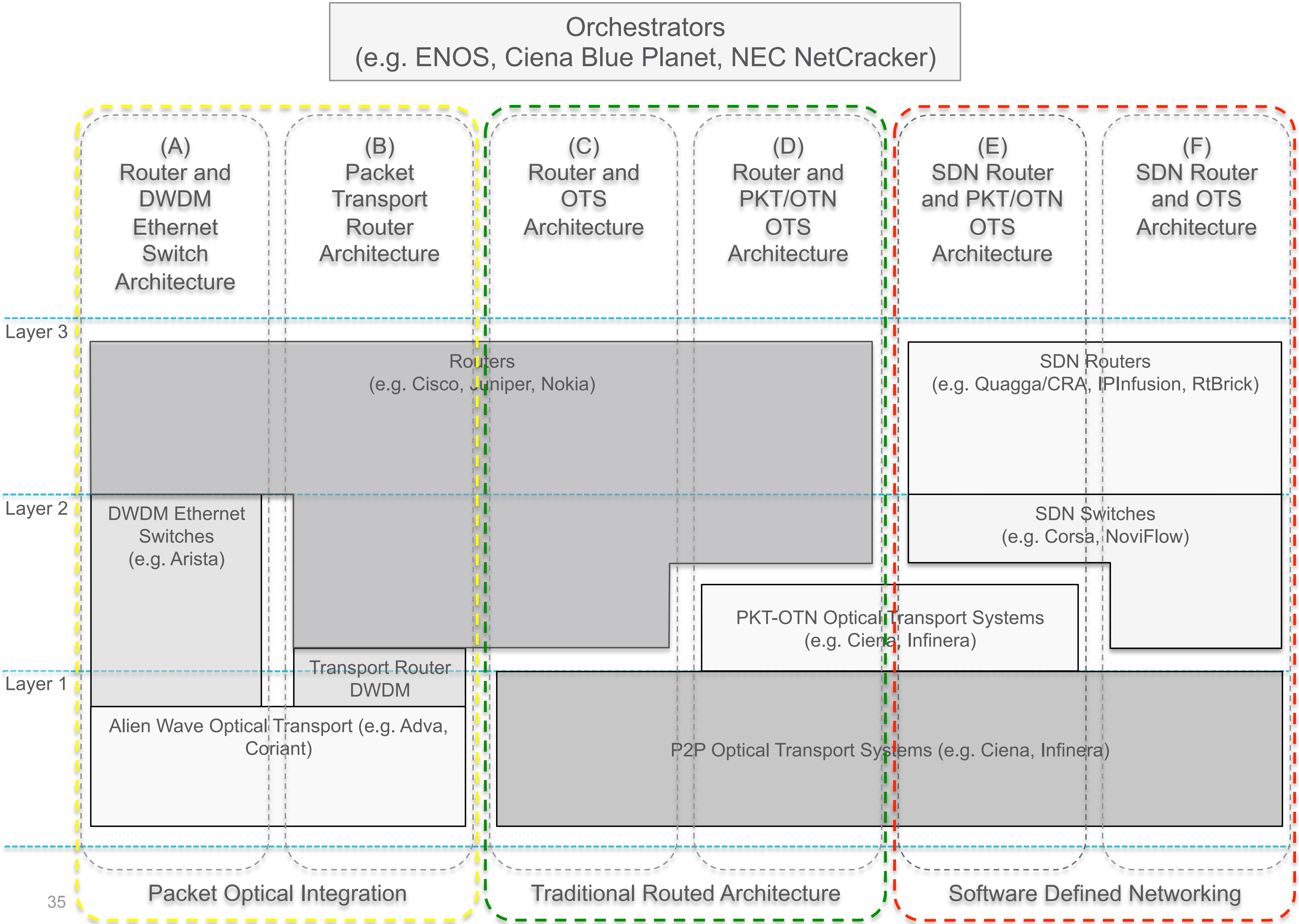
1986–2016
ESnet
30 YEARS OF NETWORKING
AT THE SPEED OF SCIENCE

ESnet6

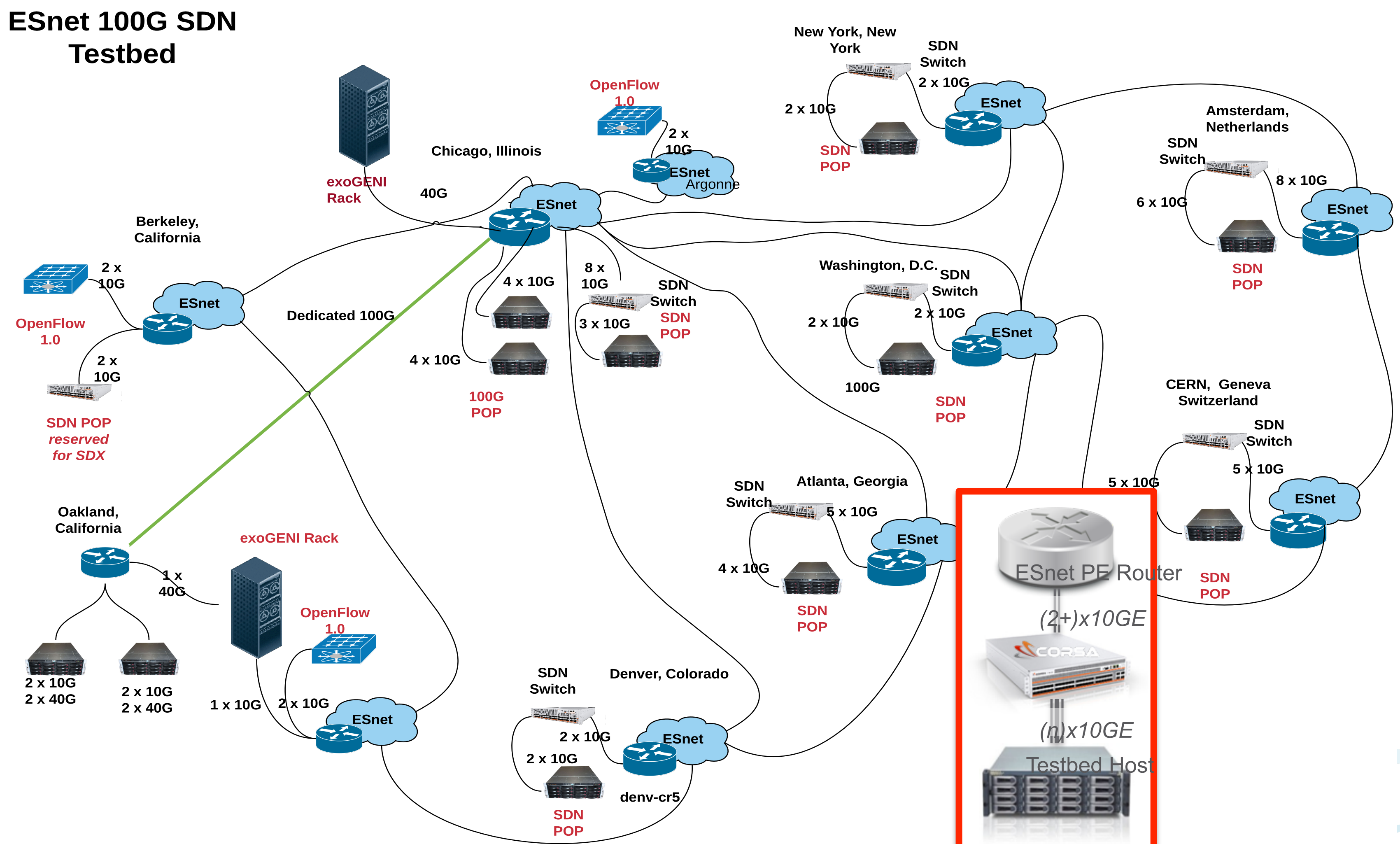
- Address the next ten years of science requirements
- Three key design goals
  - **Capacity**
    - Handle exponential traffic growth at reasonable cost
      - Support for LHC Run3 and Run4?
      - Other experiments, ex. 1 Tbps from SLAC - NERSC
  - **Reliability and Resiliency**
    - Distributed science facilities, computing, data – scientists depend on the network for their science research to work
    - Cyber-resiliency – protection against increasing level of cyber-security attacks
  - **Flexibility**
    - Compute models changing, near real-time analysis, ‘superfacility’, etc.
- Current Status: Deep within R&D cycle to explore all technology alternatives, build prototypes and engage scientists for future requirements

1986–2016 ESnet 30 YEARS OF NETWORKING AT THE SPEED OF SCIENCE

Architecture and Technologies Matrix



ESnet's 100G SDN Testbed –
significant footprint growth and dedicated bandwidth.
Focused on ESnet6 for FY17



Summary

- The network continues to successfully meets HEP's data and scientific computing needs

- Changes in computing models, data and storage could have a significant impact on network traffic, planning and future architectures
  - Example: Rapid adoption of multiple public cloud providers

- Capacity planning with ongoing exponential traffic growth will produce many technological and funding challenges in the next 10 years.

- Capacity, Reliability and Flexibility are three important criteria for designing the next generation network
  - Working collaboratively with HEP to include design constraints and criteria for next-generation is critical