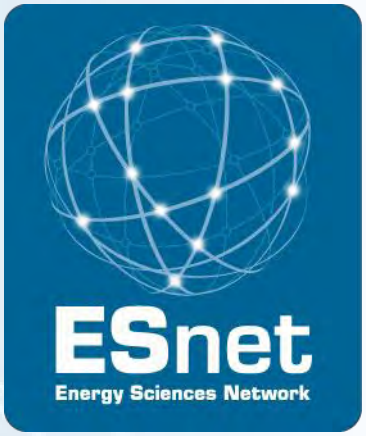



# Advanced Networking Initiative

Steve Cotter

Dept Head, Energy Sciences Network

Advanced Scientific Computing Advisory Committee (ASCAC)

Nov 2, 2011

Washington, DC


# Agenda

ESnet Overview

The Landscape in 2008/2009

Overview of the Advanced Networking Initiative

Early ANI Accomplishments

Possible ANI Demonstrations at 2011 Supercomputing Conference

# Energy Sciences Network (ESnet)

DOE National User Facility, funded by SC's Advanced Scientific Computing Research program with 2012 budget of $35M and staff of 40

- Grown from 2008 budget of $23.8M
- Received $62.4M in ARRA funding 2009 for the *Advanced Networking Initiative*

One of the largest research and education (R&E) networks in the world

- Traffic growth exceeds that in the Internet core, and carries more science data than any other network
- Connects 40+ U.S. DOE sites
- Peer with over 140 other networks (commercial and R&E)

Driven by DOE's Science Mission

- One-stop-shop resource for data transfer expertise, including network architecture, performance measurement, tools and troubleshooting
- Provide collaboration services:  voice/data/video conferencing & trust federation services to DOE scientists and their collaborators

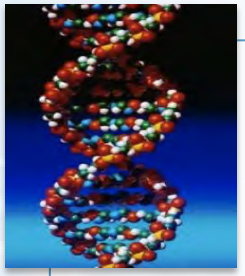
# How We're Different from the Commercial Internet

Our network is purpose-built for scientific applications:

- Handles small number of very large flows vs. billions of small flows
- Seamlessly manages highly variable traffic (peaks and troughs)
- Provides end-to-end service guarantees across partner networks
- Able to transfer/analyze/share data at a time scale dictated by the process of science
- Provides unique tools, services, overlay networks

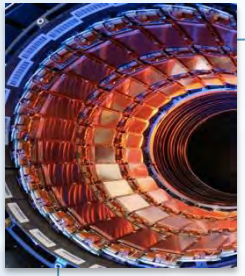
# The Landscape in 2008/2009

# Data Explosion is Occurring Everywhere



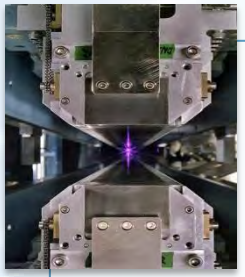
**Genomics**

- Sequencer data volume increasing 12x over the next 3 years
- Sequencer cost decreasing by 10x over same time period



**High Energy Physics**

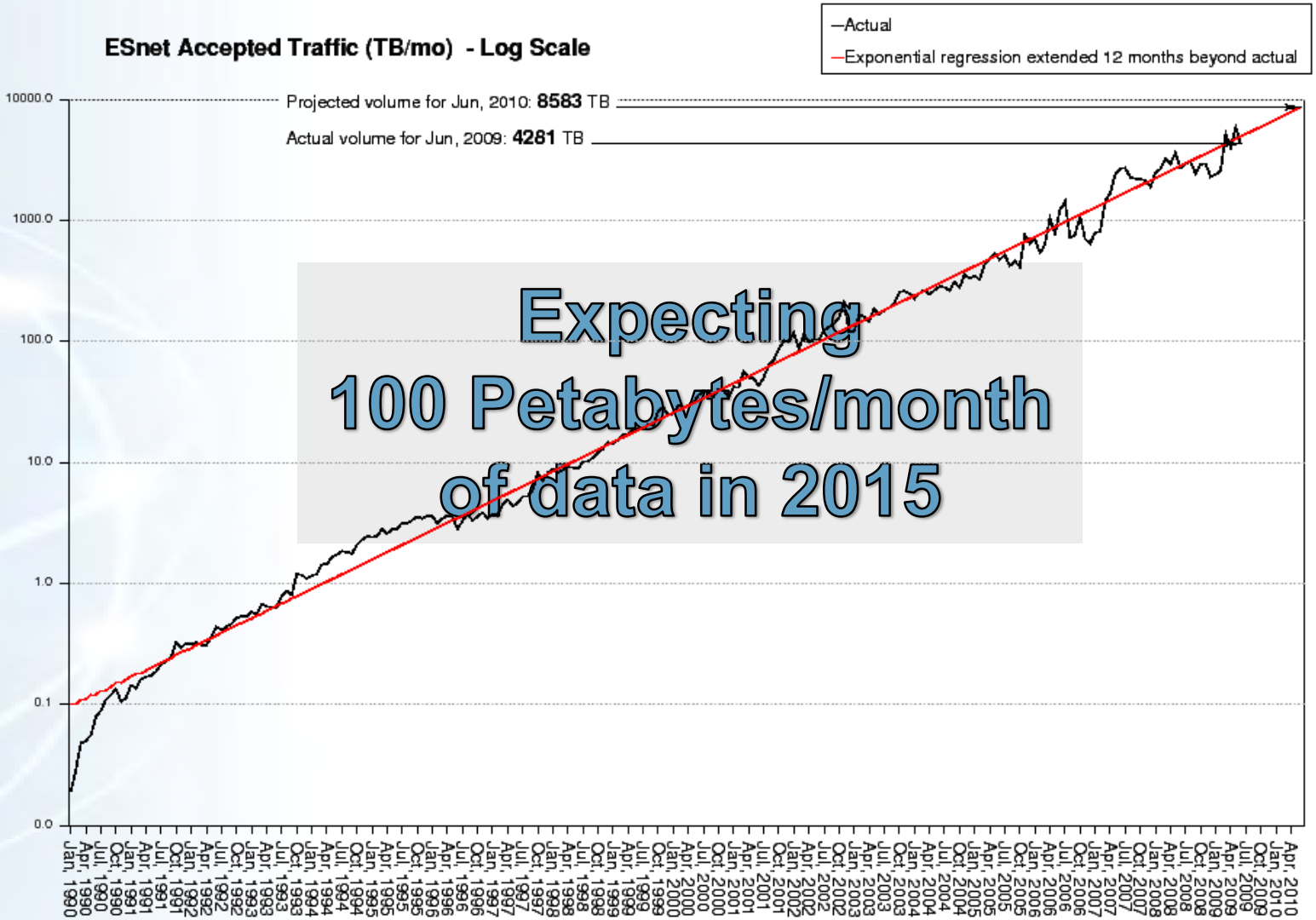
- LHC experiments produce & distribute petabytes of data/year
- Peak data rates increase 3-5x over 5 years



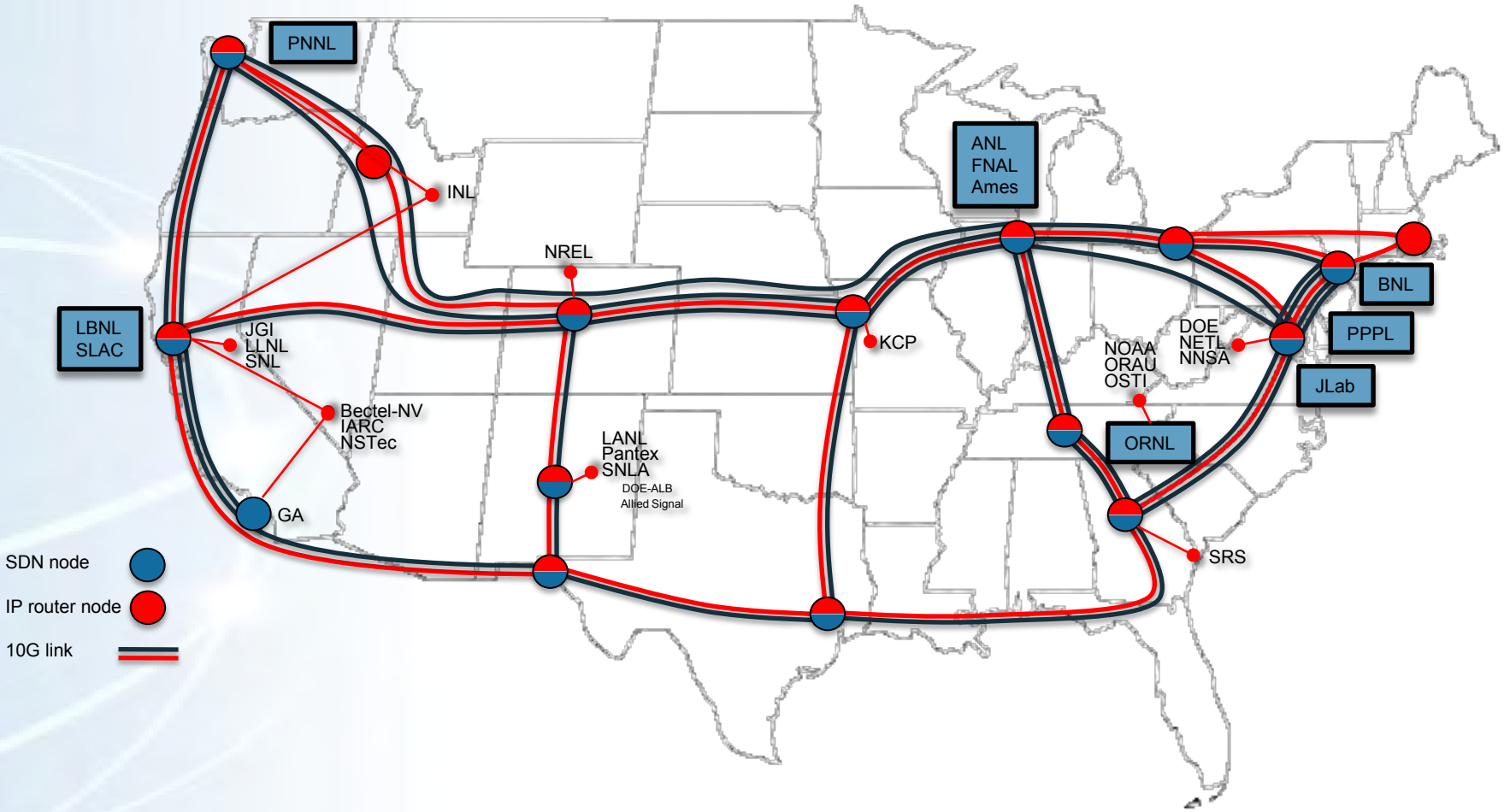
**Light Sources**

- Many detectors on a Moore's Law curve
- Data volumes rendering previous operational models obsolete

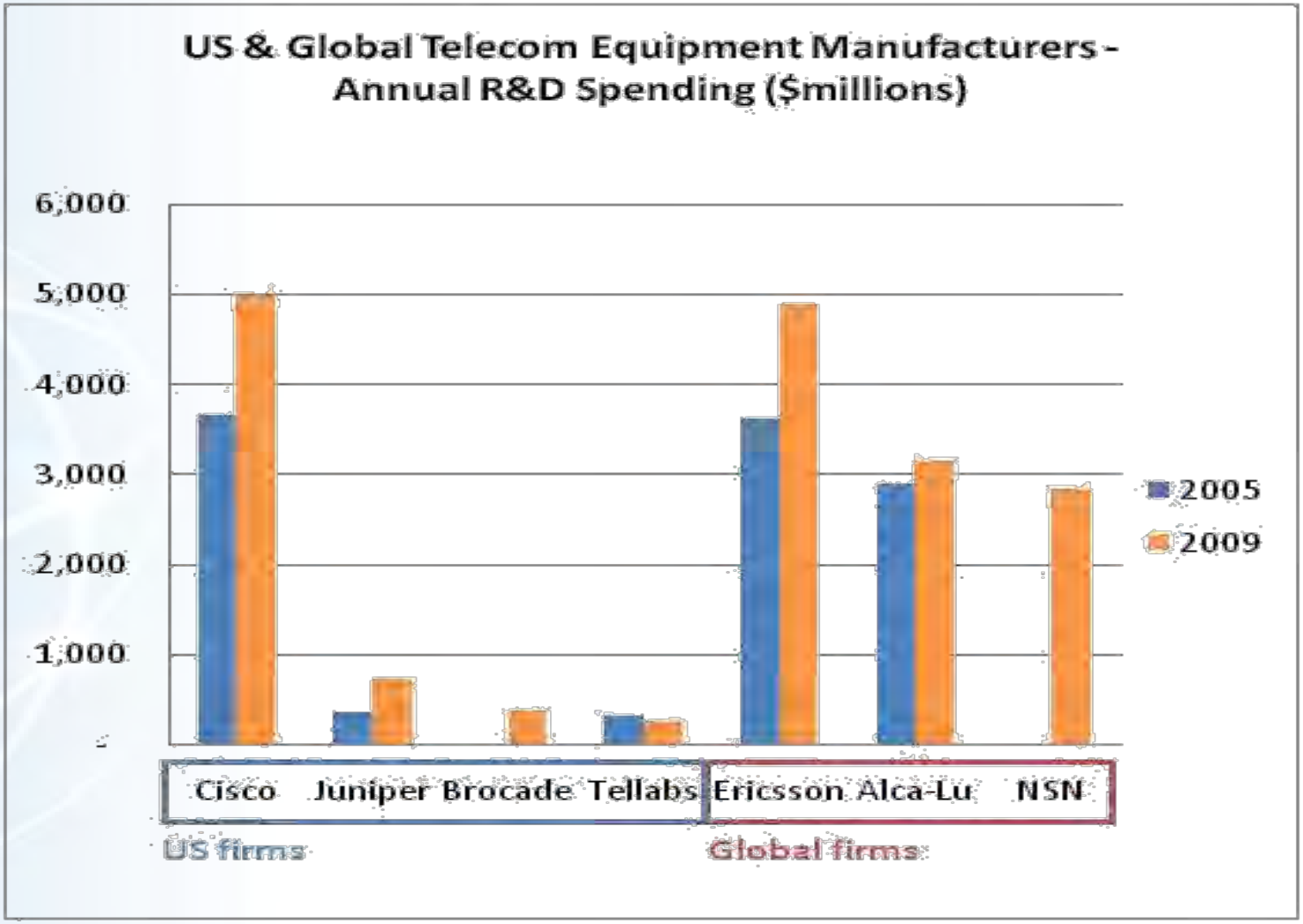
# Historical Traffic Growth



# ESnet4 Topology in 2009

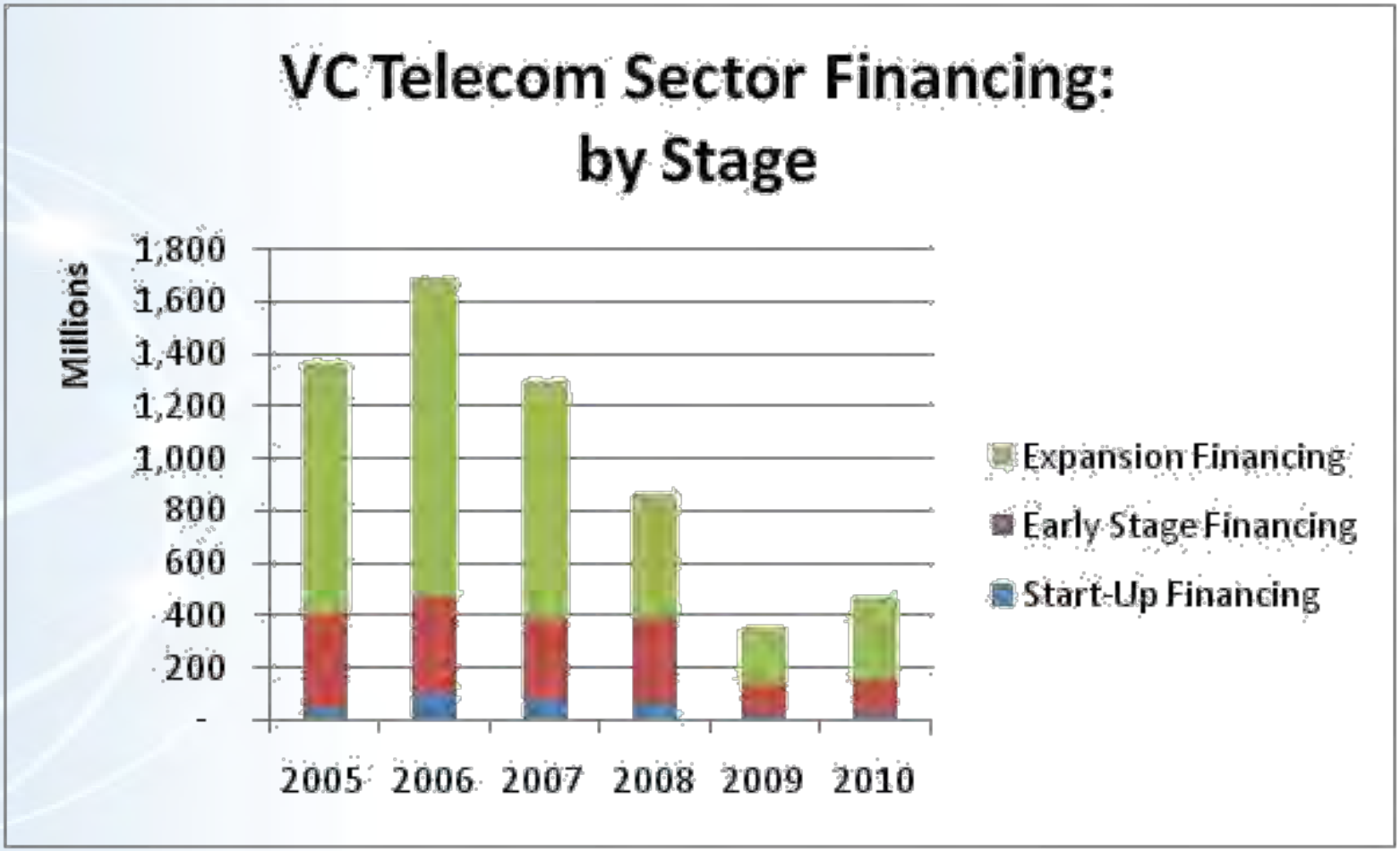


# U.S. Investment Was Falling Behind




Source: Company financial reports, 2010.

# VC Investment in the Telecom Sector:
## Declined >75% from 2006 to 2009




Source: PWCMoneyTree, 2010.

# Telecom Equipment Market Leaders

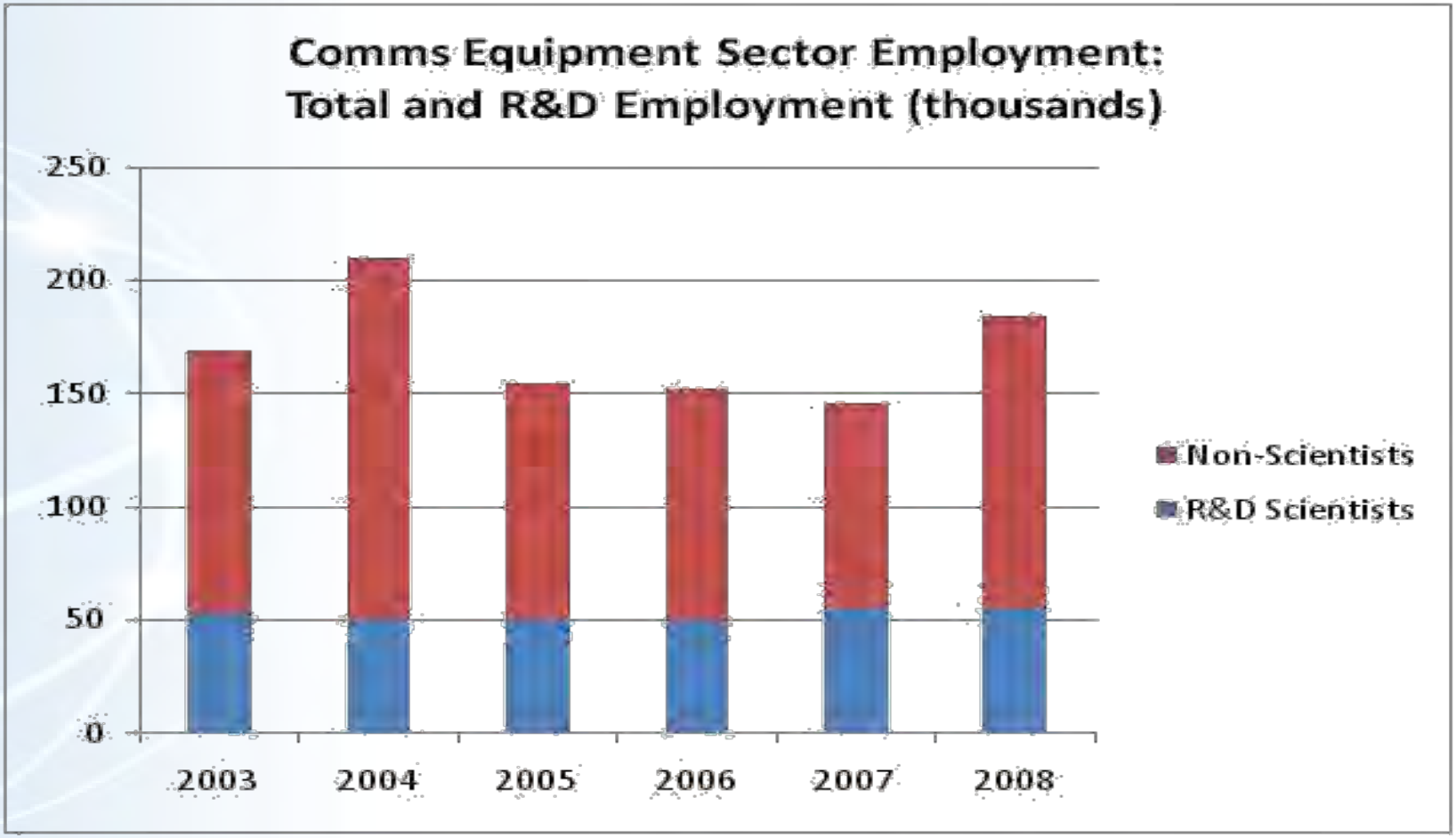
## U.S. companies in red

| | Wireless | Data | Optical | Copper Access | Fiber Access | IPTV Infrastructure | Professional Services |
|---|---|---|---|---|---|---|---|
| **Leader** | Ericsson | Cisco | Alcatel-Lucent<br>Huawei | Alcatel-Lucent<br>Huawei | Huawei | Cisco<br>Motorola (NSN) | Ericsson |
| **Challenger** | Alcatel-Lucent<br>Huawei<br>NSN | Juniper<br>Huawei<br>NSN<br>Alcatel-Lucent | Fujitsu<br>Ciena (Nortel) | ZTE | ZTE<br>Alcatel-Lucent<br>Motorola (NSN) | Huawei | Alcatel-Lucent<br>NSN |
| **Ones to Watch** | ZTE | ARRIS<br>Ericsson<br>Tellabs | ADTRAN<br>Adva<br>Infinera<br>Tellabs | ADTRAN<br>Calix<br>Ericsson | ADTRAN<br>Calix<br>Ericsson | Alcatel-Lucent<br>Ericsson | IBM |
| **Laggard** | Motorola (NSN) | Brocade | Ericsson<br>NSN | Tellabs | Tellabs | | |

Source: Morgan Keegan, 2010.

# R&D Scientists =

## 1/3 of Jobs in Telecom Equipment Sector




Source: NSF, 2010.

# Overview of the Advanced Networking Initiative (ANI)

# Advanced Networking Initiative

Goal: To accelerate by several years the commercialization of 100 Gigabit per second (Gbps) networking technologies by deploying a demonstration prototype national network and conducting research and development on an advanced network testbed facility.

100Gbps Prototype National Network

- 4 sites (DOE supercomputing centers, international exchange point)
- Stretch goal of including other SC sites

Network Research Testbed Facility

- Dark fiber IRU
- Advisory committee
- Infrastructure for researchers

# Early Challenges

Prototype Network:

- No 100Gbps standard at the time (ratified 1 year after award)
- Router and transport equipment costs unknown

Telecom industry had cut back on infrastructure investments

- No carrier was planning 100Gbps network upgrades in the ANI timeframe
- Consolidation in the industry meant less dark fiber was on the market

# 2009 Progress

Set up tabletop testbed right away (Phase 1)

- Purchased $2M in optical equipment, OpenFlow switches, routers, measurement devices
- Formed Advisory Committee (Google, Amazon, academia, labs)
  - Issued first call for research projects
- Hired staff to run testbed, manage network deployment

Bought dark fiber on Long Island to get to BNL (non-ARRA)

- Leveraged for testbed Phase 2

Began series of meetings with equipment vendors, carriers, researchers, other agencies

# 2010 Progress

Issued Transport RFP in May (standard ratified in June)

- After multiple requests for extension, ended Aug
- Finalist selected late Oct, began negotiations

Router RFP delayed

- Vendors still developing products
  - Alpha version equipment from 3 vendors evaluated in testbed

Testbed

- Received 2 rounds of research proposals
  - 13 research projects supported
- End of year, prepared equipment for redeployment in Phase 2

# 100Gbps Prototype Network

Consists of two elements:

- Backbone network
  - Long haul links between 4 locations
  - Access to 50% of capacity (up to 44 10/40/100Gbps wavelengths)
- Metropolitan networks
  - Last-mile infrastructure to the 4 locations
  - Access to 100% of capacity (up to 88 10/40/100Gbps wavelengths)

# 100Gbps Prototype Network

Combines ANI funding with Internet2 stimulus funds to build full national footprint

Internet2/Level3 Communications/Indiana Univ. manage the optical equipment and supporting infrastructure

Uses Ciena Activeflex 6500 optical equipment

- Backbone network:  chassis and fiber owned by Internet2, but ESnet purchases and owns transponder cards
- Metropolitan networks:  All equipment and fiber owned by ESnet
- Ability to provision wavelengths between any two add/drop or regeneration locations on network

Uses Alcatel-Lucent 7750 routers

- 14 chassis deployed with 33 100Gbps interfaces

# Bay Area Metro Ring

SLAC: Complete 12/14/11

- Build fiber laterals

NERSC: Complete 11/19/11

- Build fiber laterals, diverse entrance to building

LBNL: Complete 2/20/12

- Diverse fiber path for first time

JGI: Complete 2/25/12

- Build fiber laterals, diverse entrance to building

LLNL/Sandia: Complete 12/10/11

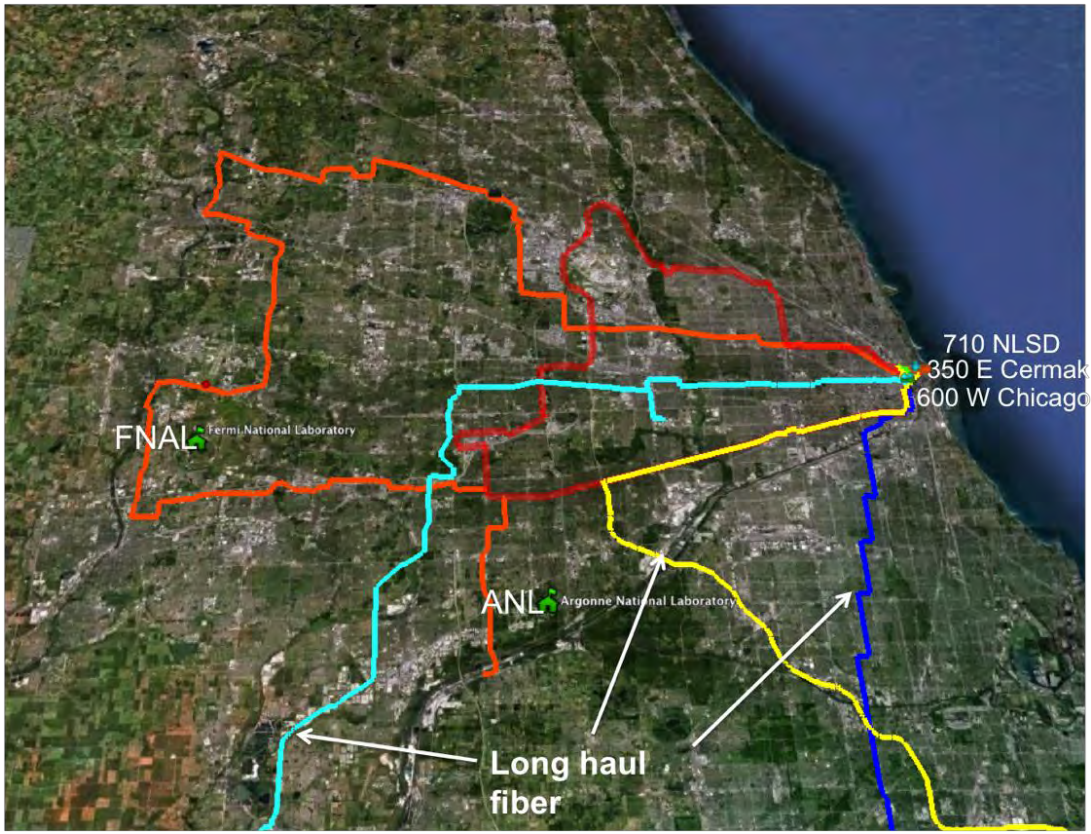
# Chicago Area Metro Ring

ANL: Complete 11/08/11

- Build fiber laterals

FNL: Complete 12/21/11

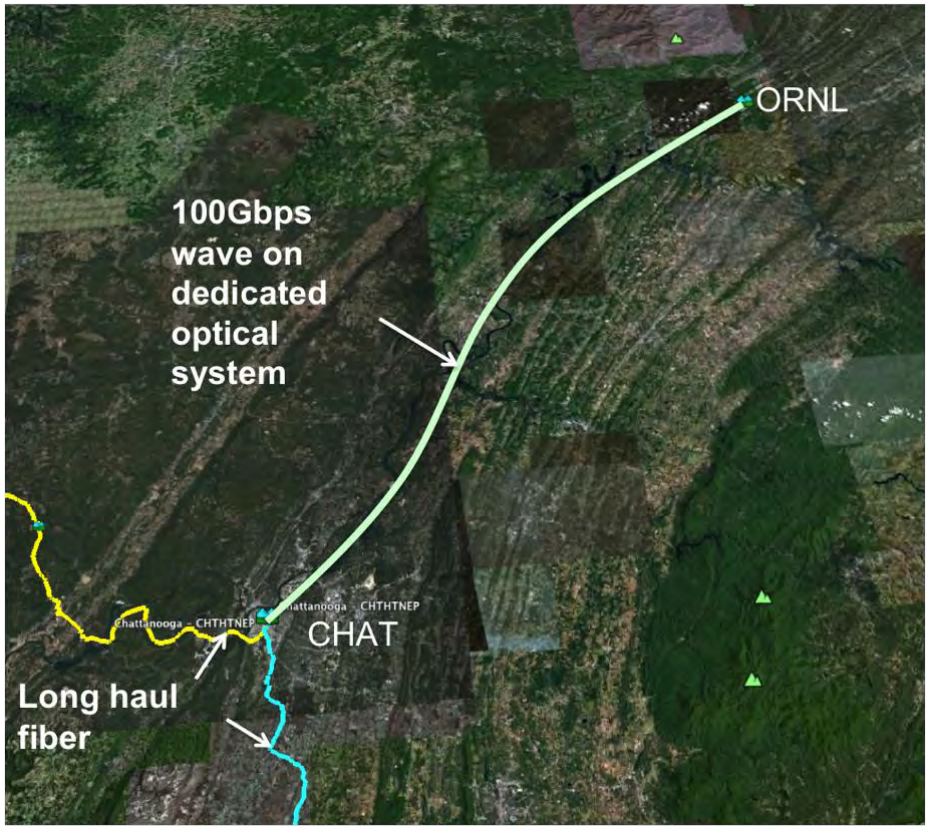
- Exploring alternate path solution using city of Batavia fiber for diversity with FNL fiber



# Oak Ridge Metro Connection

ORNL: Complete 11/09/11

- Only 1 carrier with fiber into campus
- Carrier would not sell fiber - instead getting wavelength services on dedicated optical system
- ESnet can get up to 67 100Gbps wavelengths (6.7 Tbps capacity)



# Long Island Metro Ring

Northern portion of ring:
Complete 11/21/11

- Used for testbed and production traffic to BNL
- Diverse entrances onto BNL campus
- Connect to 100Gbps network at two locations

# Remaining Timeline

Four primary locations operational before end of December 2011 (hopefully by SC11)

Rest of national network build completed by July 2012

Transition to production status complete before November 2012

# 100Gbps Network Dedicated to Science

Only science networking facility in the world with:

- Production-class owned infrastructure for cost-effectiveness and scalability
- Infrastructure capable of supporting disruptive national-scale network research

Optical Transport Network

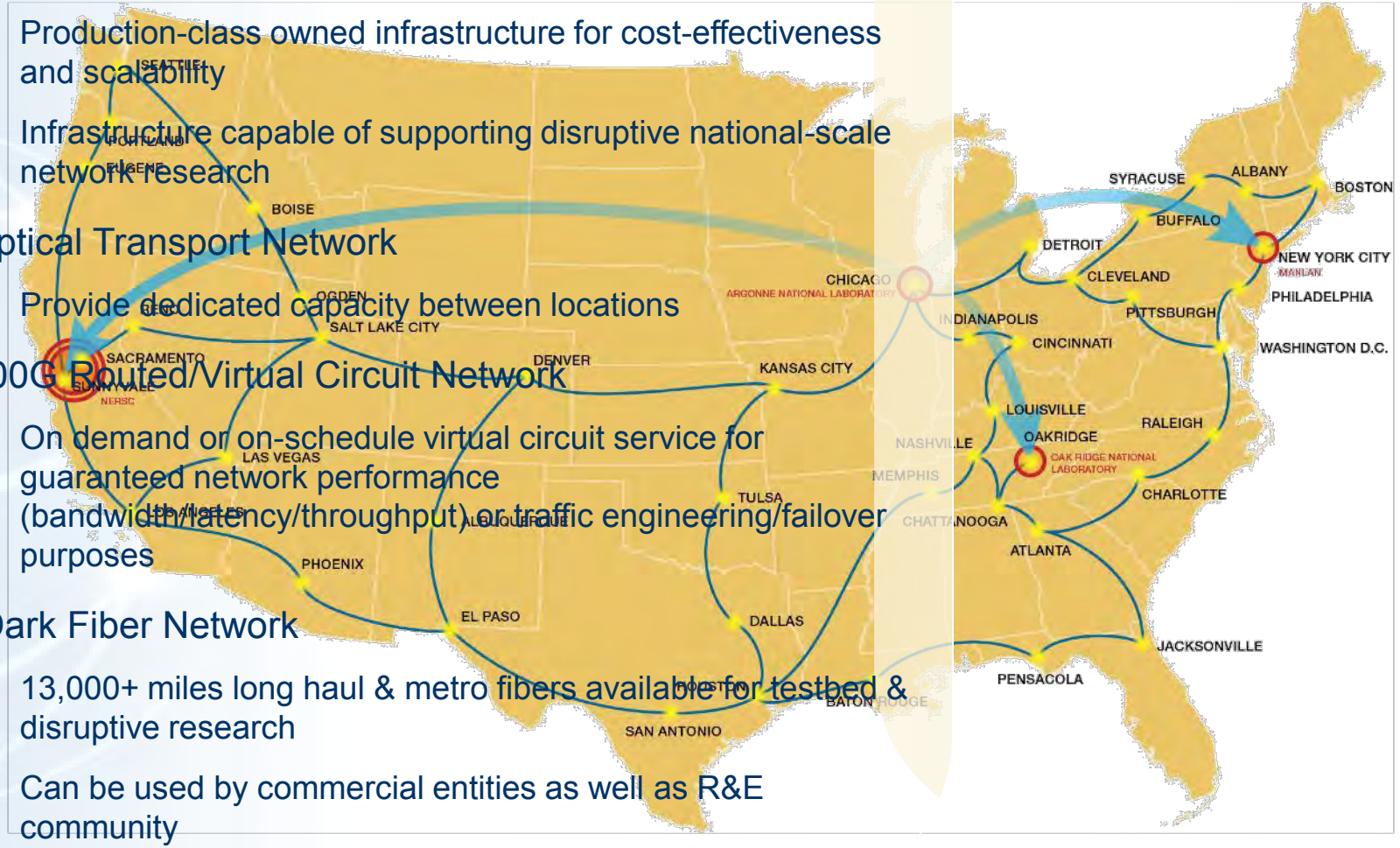
- Provide dedicated capacity between locations

100G Routed/Virtual Circuit Network

- On demand or on-schedule virtual circuit service for guaranteed network performance (bandwidth/latency/throughput) or traffic engineering/failover purposes

Dark Fiber Network

- 13,000+ miles long haul & metro fibers available for testbed & disruptive research
- Can be used by commercial entities as well as R&E community



# Early ANI Accomplishments

# Early ANI Accomplishments

Monitoring and Visualization of Energy consumed by Networks (MAVEN)

* Real-time view of live network energy consumption and environmental conditions


World's First Transcontinental 100Gbps Link

* Between New York, Chicago, Denver, and Sunnyvale creating a coast to coast path of nearly 4,000 miles

World's First 40Gbps RDMA over long haul network over Infiniband

* Public-Private Collaborative Demo Leveraged New ESnet Advanced Networking Initiative (ANI) Testbed

End-to-End Circuit Service at Layer 2 (ECSEL)

* Using OpenFlow on the campus and OSCARS in the wide area, demonstrated a 'zero configuration' architecture for dynamic, end-to-end virtual circuits

# Carbon Cycle 2.0


New network is first 100G network in the world instrumented for power and environmental monitoring that aims to:

- Create an energy-aware network ecosystem through research and practice
  - Tools, equipment, PDUs, meters, business models to incorporate traffic proportionality
- Build tools to collect and visualize live network energy consumption
  - Flexible meta-data to create customized views (power consumed per path, per location, per layer)
- Create open datasets for network energy-efficiency research
  - Working with industry and standards bodies
- Catalyze adoption of theoretical research/experiments by industry
  - Energy proportionality will require redesign of network equipment
  - Establish metrics based on quantified improvements against baseline

# Supported Testbed Research Projects

By early 2012, ~30 research projects using the testbed

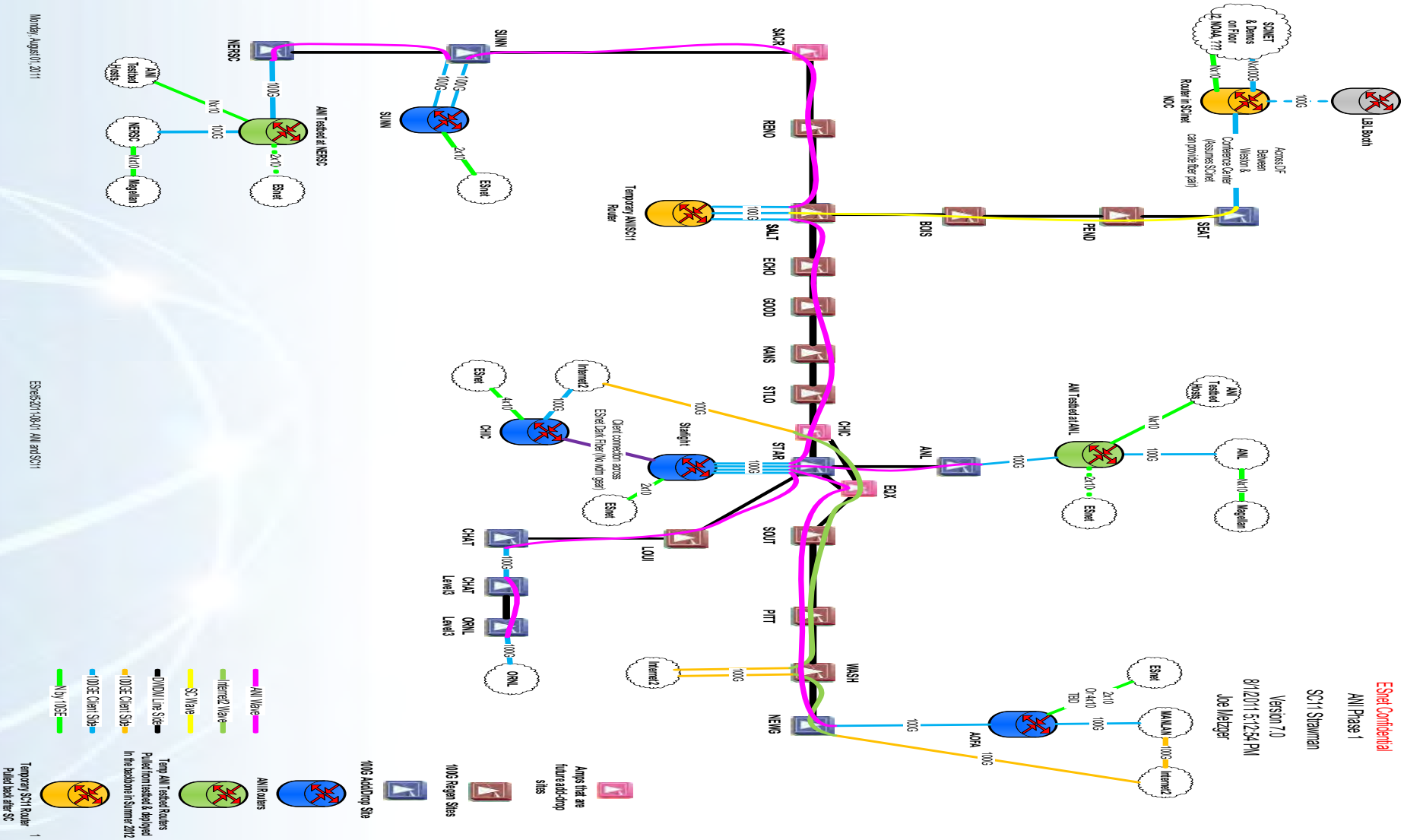
Potential breakthroughs from this research include:

- Ability to scale TCP well beyond 10Gbps: 2 projects
- Alternative transport protocols that scale better than TCP (e.g.: RDMA): 2 projects
- Ability to easily create end-to-end circuits: 2 projects
- 100Gbps host interface: 1 project
- Ability to scale application middleware to 100Gbps and beyond: 6 projects
- “Unconditionally secure” Continuous Variable Quantum Key Distribution (CV-QKD) technology: 1 project

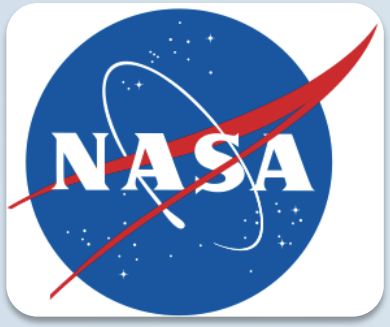
For more information see: http://www.es.net/RandD/advanced-networking-initiative/current-testbed-research/

# Possible ANI Demonstrations at 2011 Supercomputing Conference

# ANI Network for SC11



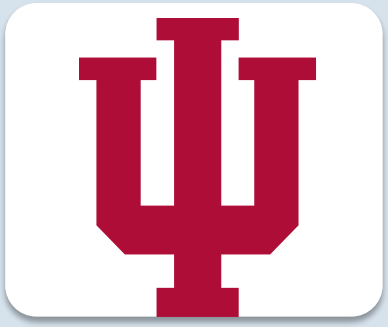
# SC11 Demos Utilizing ANI




* Data transfer from NASA Goddard to Seattle
* Demonstrating 100Gbps data transfers on low-cost commodity hardware
* Multiple collaborators
* Part of SCinet Research Sandbox


* Data transfer from NERSC to ORNL and NERSC to ANL
* 36 Terabytes of Climate data
* Transfer to use GridFTP and RDMA-FTP
* Disk to memory and disk to disk


* Data transfer from NERSC to ANL
* Data transfer will use new FTP100 tool that is RDMA based



* Data transfer from Bloomington to Seattle
* 100Gbps Lustre FS demo
* Part of SCinet Research Sandbox

# SC11 Demos Utilizing ANI


* Cosmology data visualization at 100Gbps
* Data transfer from NERSC to Seattle
* Comparison visualization at 10Gbps


* Prototype 100Gbps HEP data transfer, NERSC to ANL
* Data Analysis performed at ANL using Condor


* Data transfer looped from Seattle to Chicago (>4400 mile path)
* 40Gbps RDMA over long haul network over Infiniband


* 40Gbps data transfer from University of Michigan to Seattle
* LHC data
* Caltech, CERN/LHC collaboration



U.S. DEPARTMENT OF
ENERGY


UNIVERSITY OF
CALIFORNIA