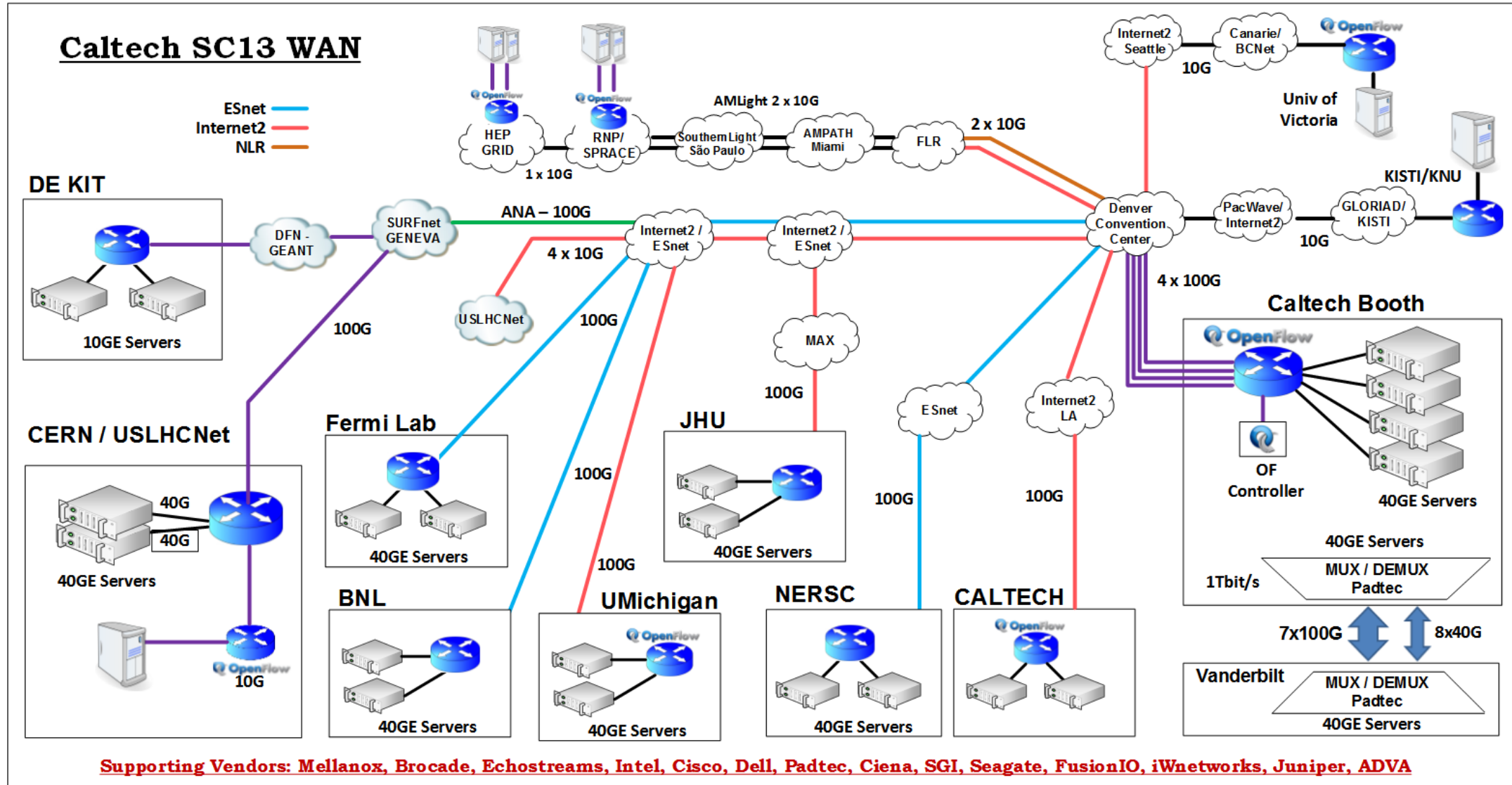


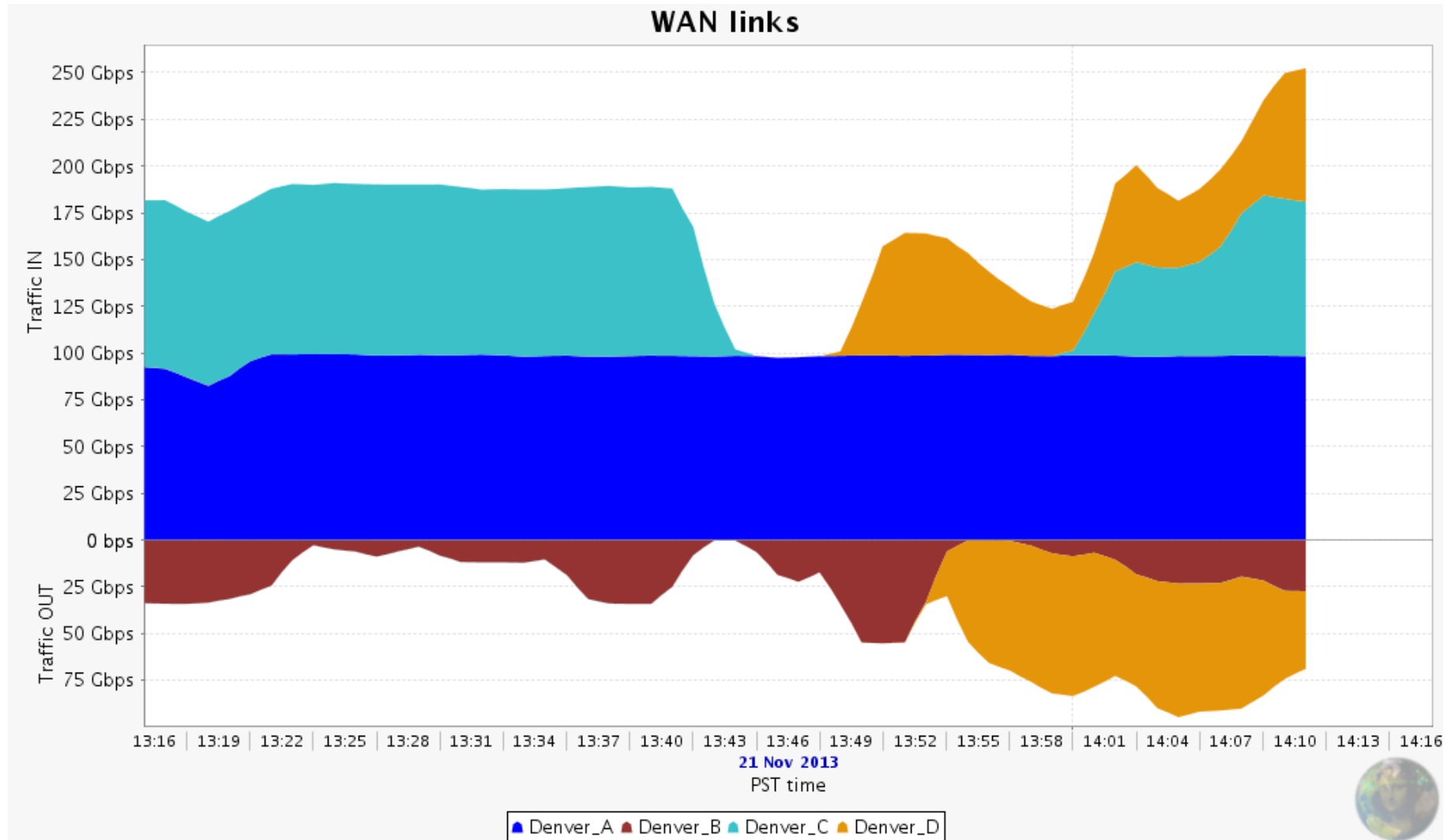
# SC13 Data Movement WAN and ShowFloor

Azher Mughal  
Caltech

# WAN Network Layout



# WAN Transfers



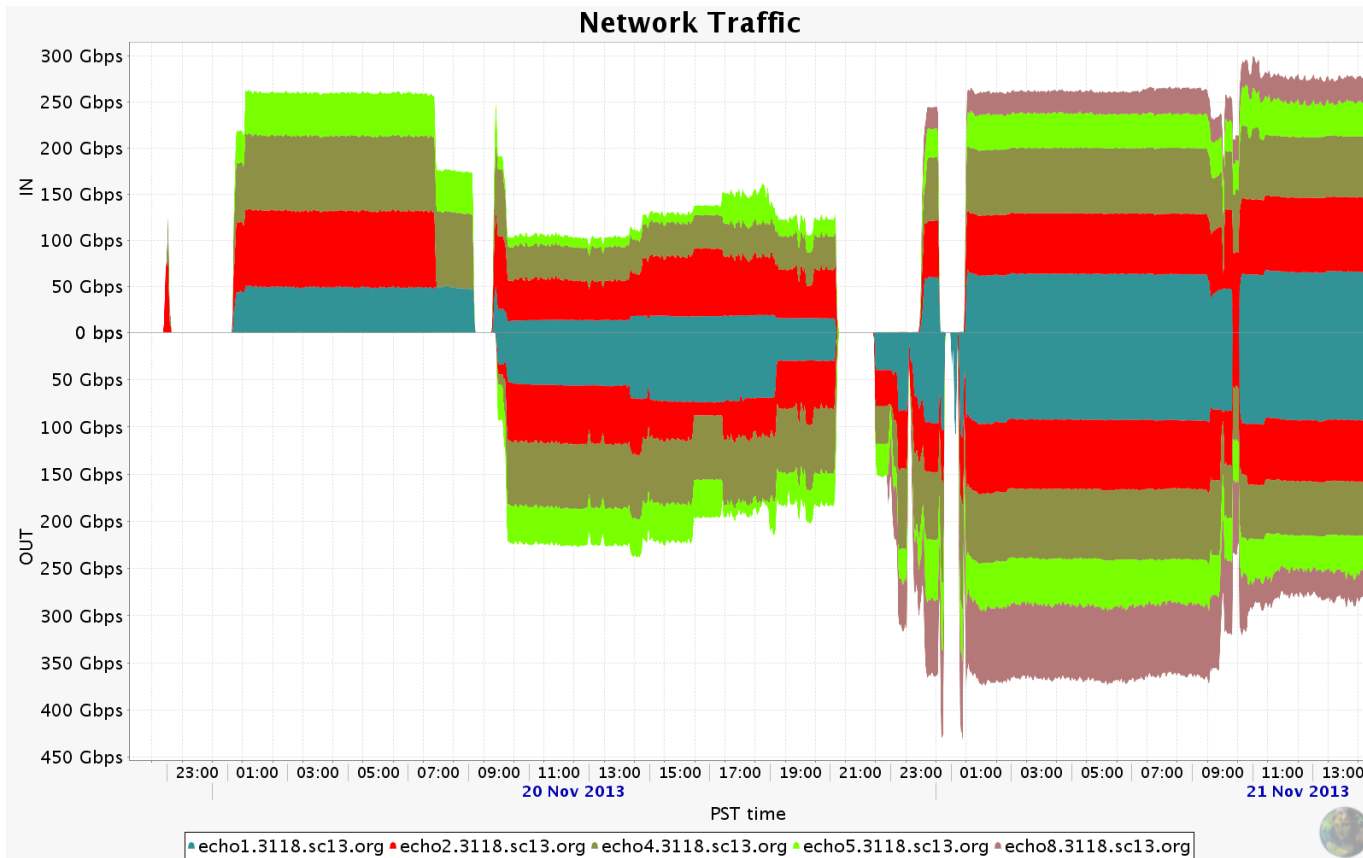
# From Denver Show Floor



# TeraBit Demo

7x 100G links

8 x 40G links



# Results

## SC13 – DE-KIT

- 75Gb from Disk to Disk (couple of servers at KIT – Two servers at SC13)

## SC13 BNL over Esnet:

- 80G over two pair of hosts, memory to memory

## NERSC to SC13 over ESnet:

- Lots of packet loss at first, then removed the Mellanox switch from the path, and then the path was clean
- Consistent 90Gbps, reading from 2 SSD host sending to single host in the booth.

## SC13 to FNAL over ESnet:

- Lots of packet loss; TCP max around 5Gbps, but UDP could do 15G per flow.
- Used 'tc' to pace TCP, and then at least single stream TCP behaved well up to 15G.

But using multiple streams was still a problem. This seems to indicate something in the path with too small buffers, but we never figured out what.

## SC13 – Pasadena Internet2:

- 80G read from the disks and write on the servers (disk to memory transfer). Link was lossy the other way.

## SC13 – CERN over Esnet:

- About 75Gb memory to memory. Disks about 40Gb

# Post SC13 – Caltech - Geneva

- About 68Gb
- 2 pair of Servers used
- 4 Streams per server
- Each Server around 32Gbps
  
- Single stream stuck at around 8Gbps ??



# Challenges

- Servers with 48 SSD Disks - Adaptec Controllers
  - 1GB/s per controller (driver limitation, single IRQ)
- Servers with 48 SSD Disks - LSI Controllers
  - 1.9GB/s per controller
  - Aggregate = 6 GB /s out of 6 controllers (still working)
- Sheer number of resources (servers+switches+NICs+man-power) needed to achieve Tbit/sec