Results From
NASA High End Computing (HEC)
WAN File Transfer Experiments/Demonstrations
Super Computing 2012 (SC12)

Bill Fink
Bill.Fink@nasa.gov
Computational And Information Science and Technology Office (CISTO)
High End Computer Networking Team (HECN), Code 606.1
NASA Goddard Space Flight Center
November 11-15, 2012

Sample Results From NASA HEC WAN File Transfer Experiments/Demonstrations At SC12

- **<u>NASA Partners in "Using 100G Network Technology in Support of Petascale Science" Special SC12 Demonstration/Evaluation Experiments</u>**
- DOE's Earth Sciences Network (ESnet), Sidera Networks, Super Computing Convention's SCinet, Northwestern University International Center for Advanced Internet Research (iCAIR), University of Illinois at Chicago's Laboratory for Advanced Computing (LAC), Starlight, University of Maryland College Park's Mid-Atlantic Crossroads (MAX)
- Vendors who loaned equipment: Brocade, Ciena, Cpacket, Fujitsu, Hot Lava, Intel, LSI, Supermicro
- SC12 suport from NASA/GSFC (Jarrett Cohen) and NASA/ARC (Gina Morello and Harjot Sidhu)

Sample Results From NASA HEC WAN File Transfer Experiments/Demonstrations At SC12

Special SC12 Demonstration/Evaluation Experiments

- Use of three custom NASA/HECN Team built network-testing-servers deployed into the LAC/iCAIR and NASA Exhibit Booths and at GSFC, capable of:
- >90Gbps bi-directional tcp memory-to-memory data flows over a 100G path, 96Gbps between the booths uni-directional disk-to-disk file copies (using SSDs)
- Demonstrate/Evaluate interoperability between multiple vendor 100G products from Alcatel, Brocade, Ciena, Fujitsu, over SCinet, ESnet, Starlight, Sidera Networks, and MAX/DRAGON
- Just prior to SC12, obtained 115Gb/s back-to-back (four 40GEconections between our servers).
- Due to a 3+ day delay in getting equipment, resolving a raid controller problem, and unexpected flow control issues, had insufficient time to run final tests. Achived 84Gbps TCP WAN memory-to-memory (believe we could have achieved ~80Gbps disk-to-disk if hadn't run out of time).

High End Computer Networking (HECN) Team

Bill Fink
Acting Project Lead

Paul Lang
Network Engineer

Aruna Muppalla
Network Engineer

Jeff Martz
Network Engineer

Mike Stefanelli
Network Engineer

Pat Gary
(In Memoriam)

Pre-SC12 Successes

* Faster Raid Controllers
* Faster Processors
* 3-port 40GE NICs (2nd pair received 10/26, power outage 10/27-10/28, GSFC closed 10/29-10/30 because of hurricane Sandy, shipped servers to SC12 on 11/5)
* split load between 2 processors (16 cores total) each processor has 1 NIC (use two 40GE) and two raid controllers, total of 16 2-disk raid arrays (four arrays per raid controller, each array ~8Gb/s
* 115Gb/s disk-to-disk back-to-back transfer (over four 40GE interfaces).
* GSFC to Scinet NOC loop – 98 Gb/s bi-directional UDP with no drops

SC12 Issues

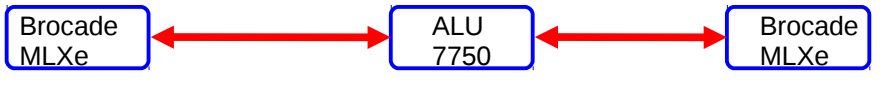
* Acadia Optronics 100G NIC delayed due to fabrication problems
* Unable to acquire switch/router with both 40 and 100G for demo.
* Routers delayed in shipping, arrived Monday evening
* Second 8-port 10G line card for routers delivered Tues. 11AM
* Disks dropping off-line (raid controllers overheating, had to rearrange equipment and remove lid to server, issue not experienced on the server in the computer room at GSFC)
* Optimal configuration for 40G NICs not ideal for 10x10G muxponder.
* WAN flow control issues 40G to 10G - investigated possible reconfiguration solutions including swap out of 3-port 40GE NICs with 6-port 10GE NICs, but WAN connection was lost just as the test was ready ~3PM Thursday.

Using 100G Network Technology in Support of Petascale Science

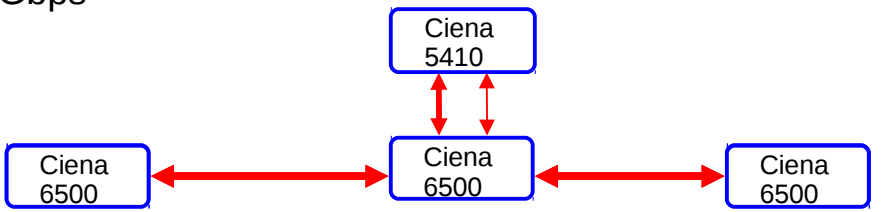
A Collaborative Initiative Among NASA, ESnet, Sidera Networks, SCinet, Northwestern/iCAIR, UIC/LAC, Starlight, UMD/MAX/DRAGON

SC12 Demo Summary

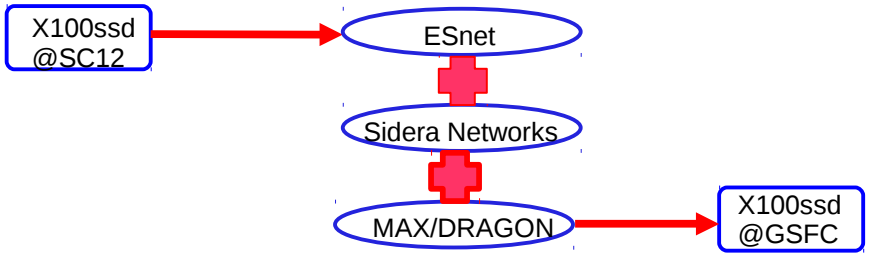
1x100Gbps full-duplex across:
- uni-directional tcp: 99Gbps
- uni-directional disk-to-disk: 96.1 Gbps



10x10Gbps (100G) Muxponder
- uni-directional tcp: 99.04Gbps
- uni-directional disk-to-disk: ~80Gbps



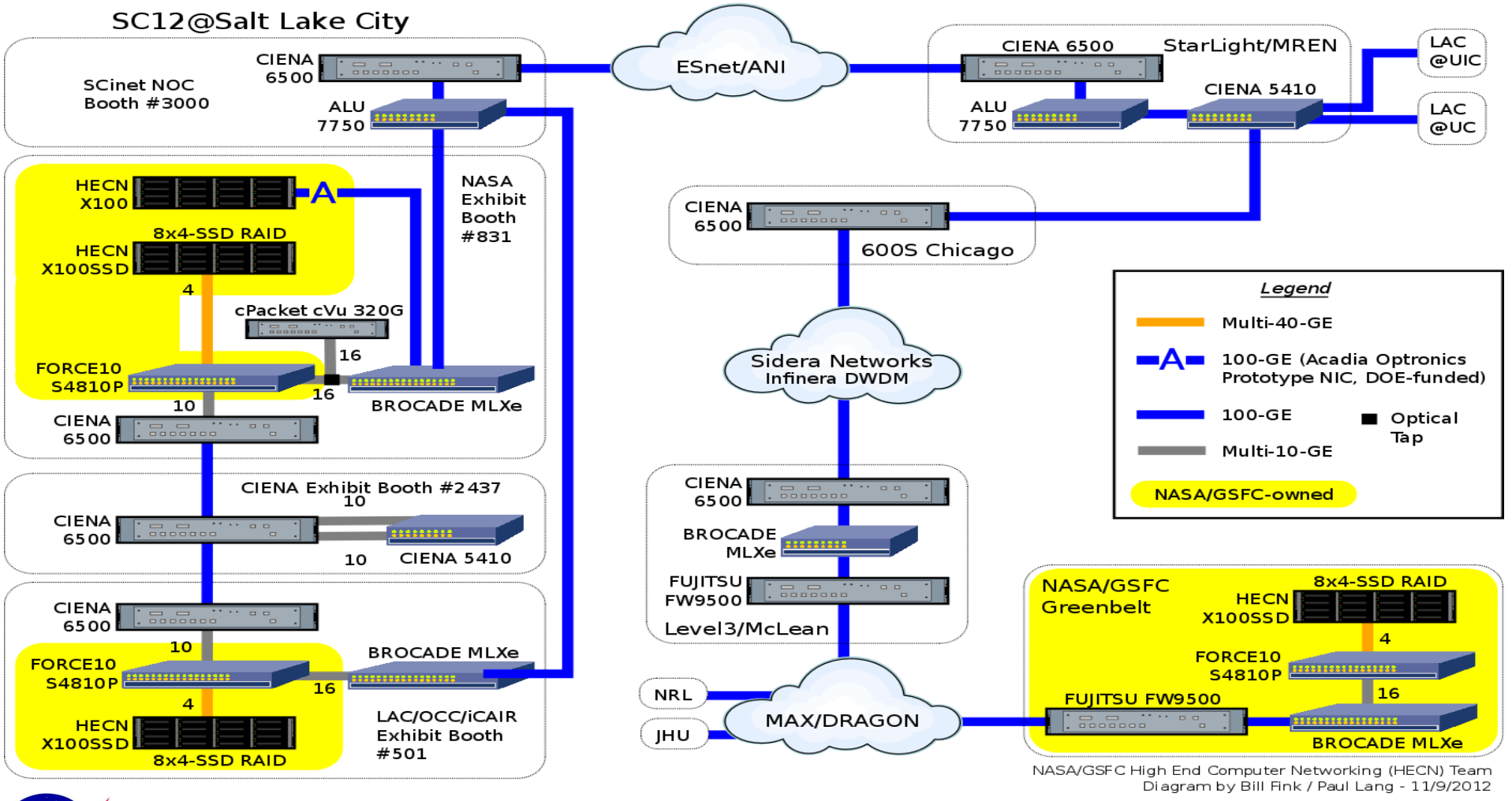
1x100Gbps full-duplex across:
- uni-directional tcp:83.9Gbps



Testbed Environment

Evaluations/Demonstrations of 100 Gbps Disk-to-Disk File Transfer Performance Across LANs & WANs

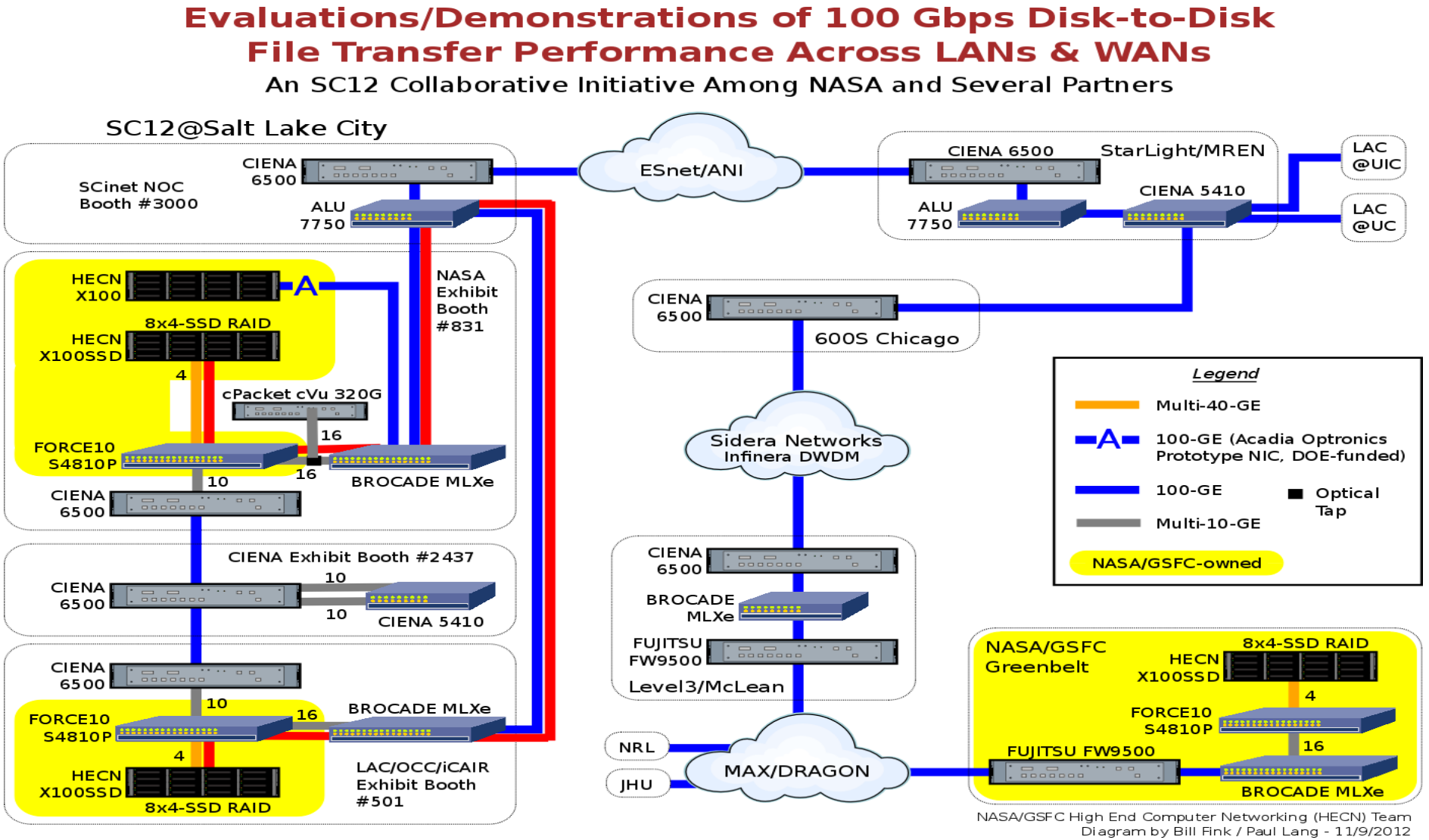
An SC12 Collaborative Initiative Among NASA and Several Partners



NASA/GSFC High End Computer Networking (HECN) Team
Diagram by Bill Fink / Paul Lang - 11/9/2012

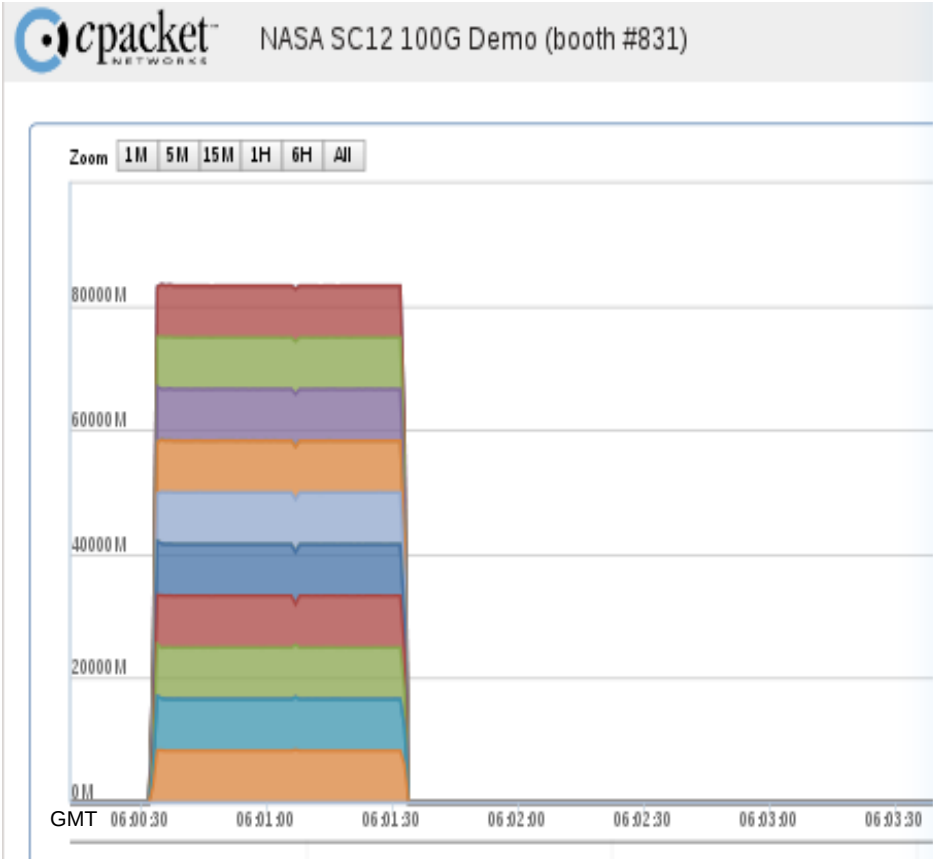
NASA to LAC Booth via ALU 7750 & Brocade MLXe

(red path) 96.1Gbps disk-to-disk



NASA to LAC Booth via Alcatel & Brocade

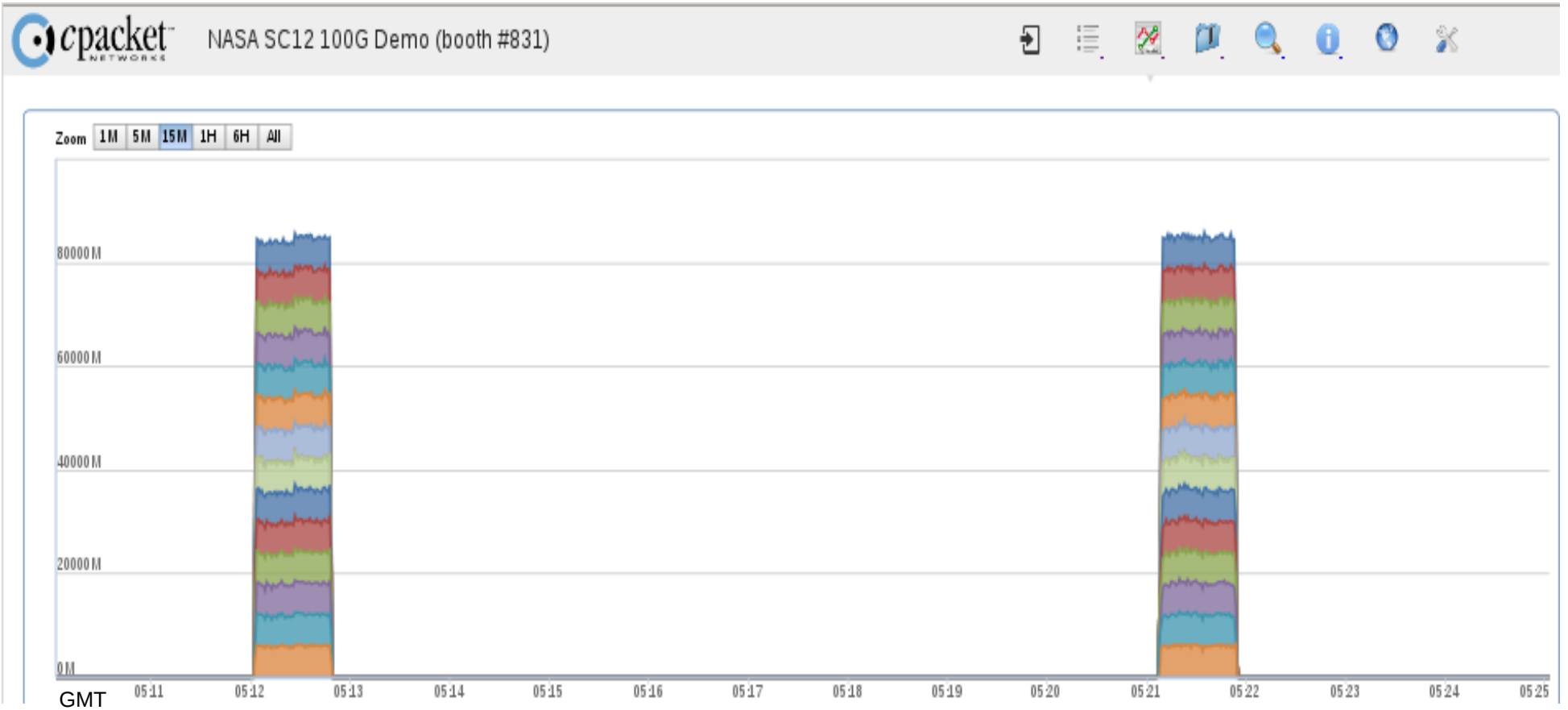
nuttcp tcp memory-to-memory test results (99Mbps)



Note: Two streams are not included in the above graph.

NASA to LAC Booth via Alcatel & Brocade
nuttscp disk-to-disk test results (96.1Gbps)

(512GB in ~46 seconds, 16 streams of 32GB)



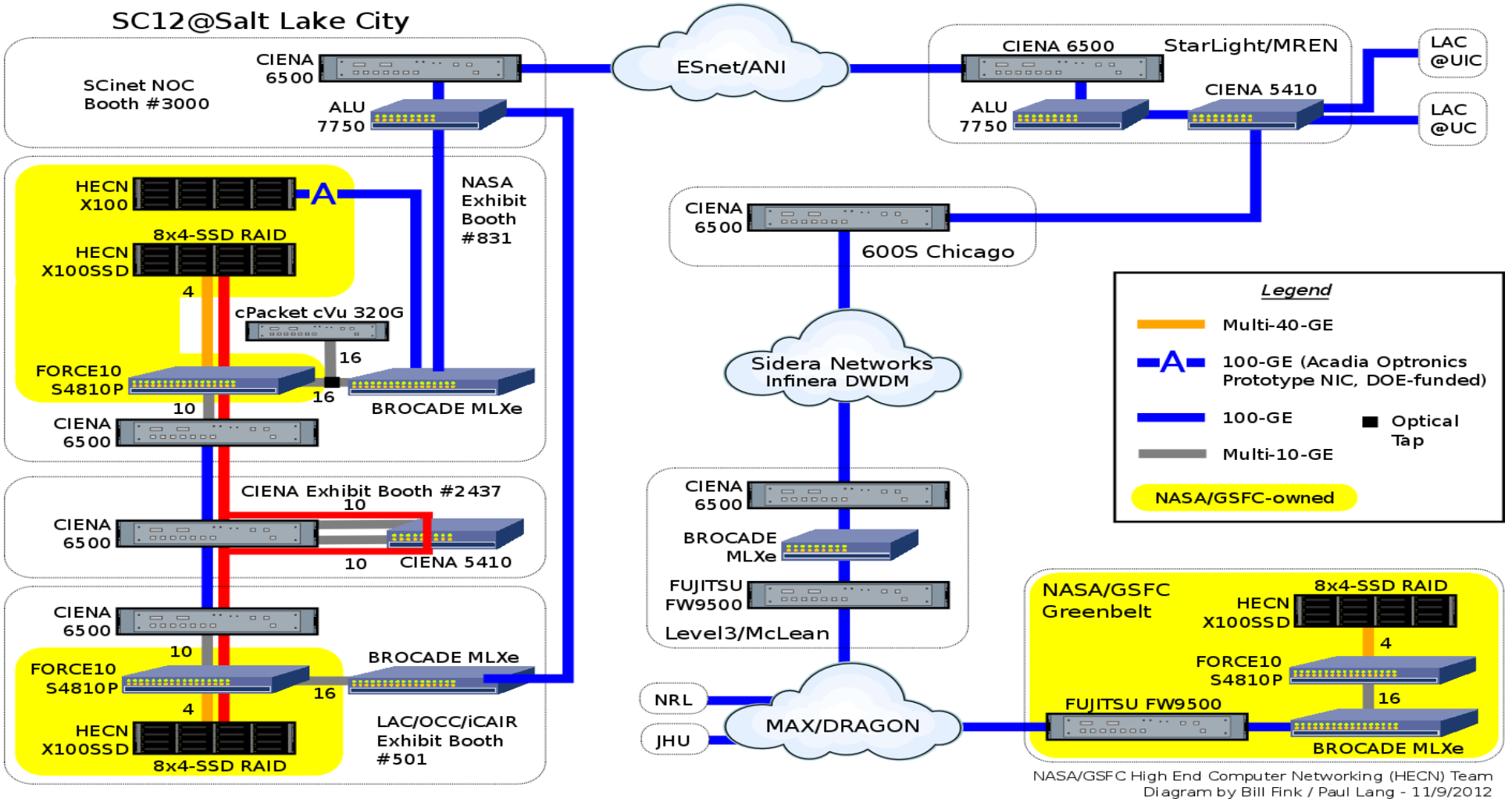
Note: Two streams are not included in the above graph.
Above: GB = 1024x1024x1024 and Gbps = 1,000,000,000 bits/second

NASA to LAC Booth via Ciena 5410 & 10x10G Muxponder

(red path)

Evaluations/Demonstrations of 100 Gbps Disk-to-Disk File Transfer Performance Across LANs & WANs

An SC12 Collaborative Initiative Among NASA and Several Partners



NASA/GSFC High End Computer Networking (HECN) Team
Diagram by Bill Fink / Paul Lang - 11/9/2012

NASA to LAC Booth via Ciena 5410 and x10G Muxponder

(no graph – due to Ciena not being in the path through the cPacket)

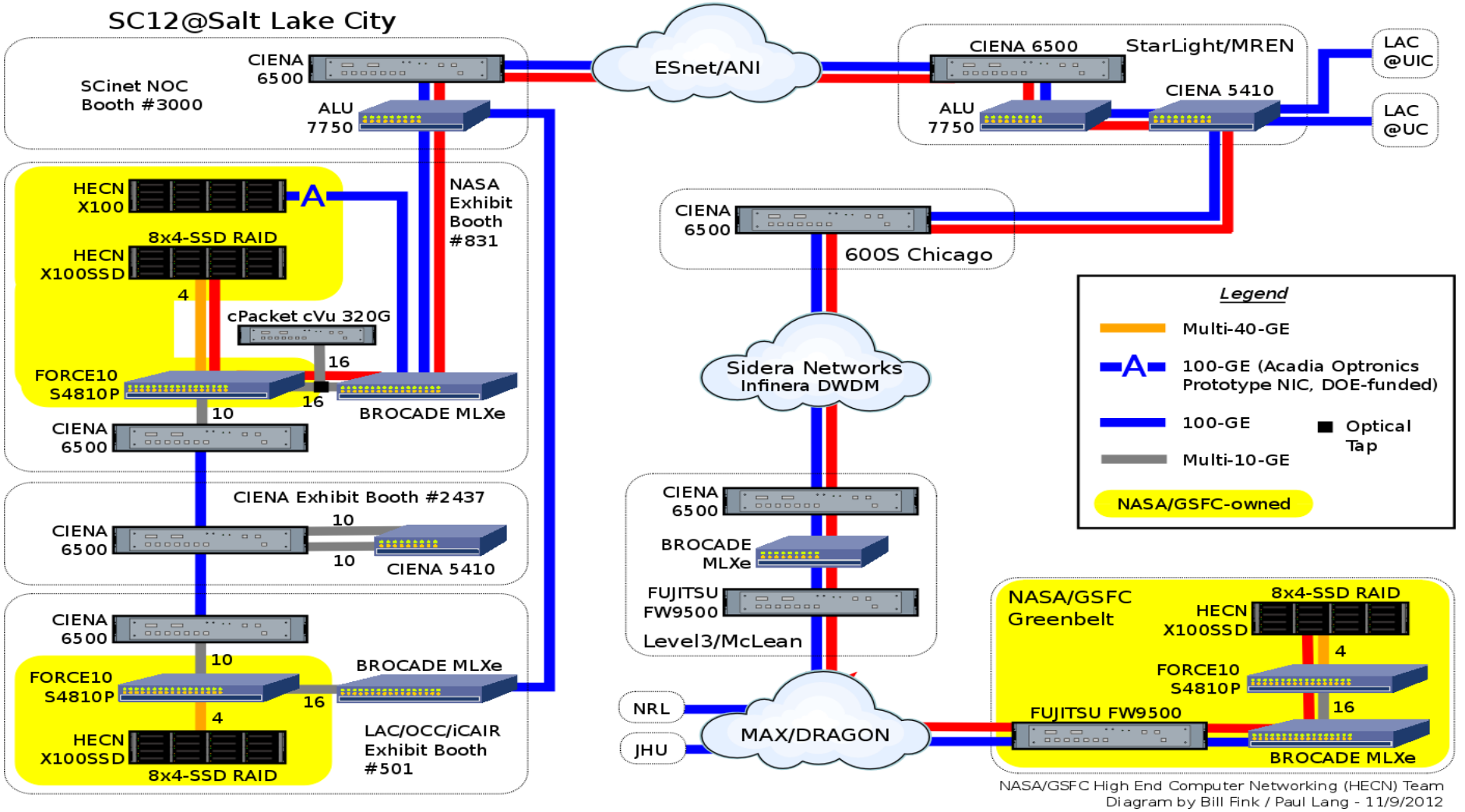
* nuttcp tcp memory-to-memory uni-directional 98.93Gb/s

* nuttscp disk-to-disk uni-directional transfer 79.86Gb/s (only used 10 streams each stream capable of 8Gb/s, two disks at 4Gb/s each, would have needed to modify the number of disks in each stream to get 96Gb/s with 10 streams)

NASA Booth to GSFC (red path)

Evaluations/Demonstrations of 100 Gbps Disk-to-Disk File Transfer Performance Across LANs & WANs

An SC12 Collaborative Initiative Among NASA and Several Partners

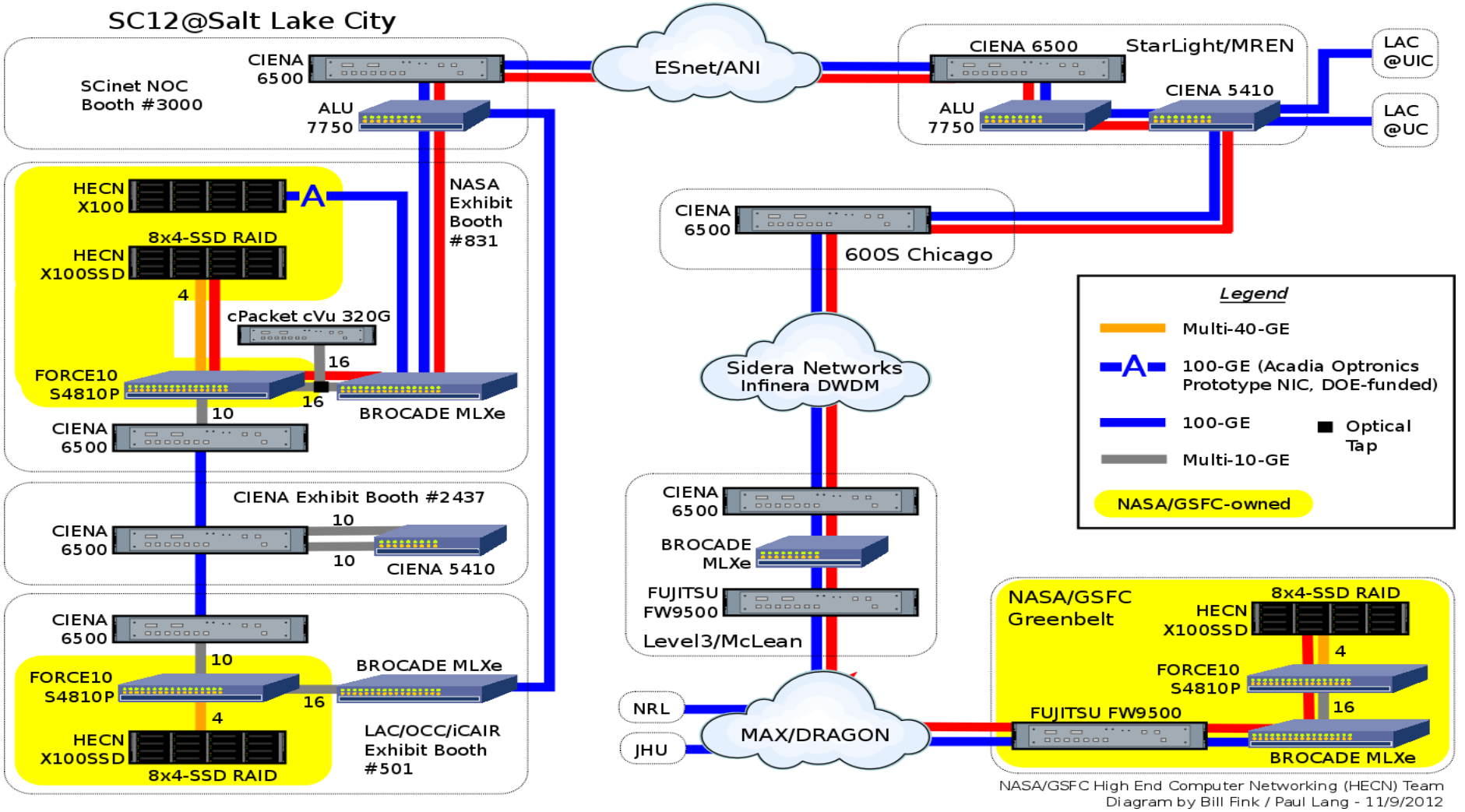


NASA/GSFC High End Computer Networking (HECN) Team
Diagram by Bill Fink / Paul Lang - 11/9/2012

NASA Booth to GSFC (red path)

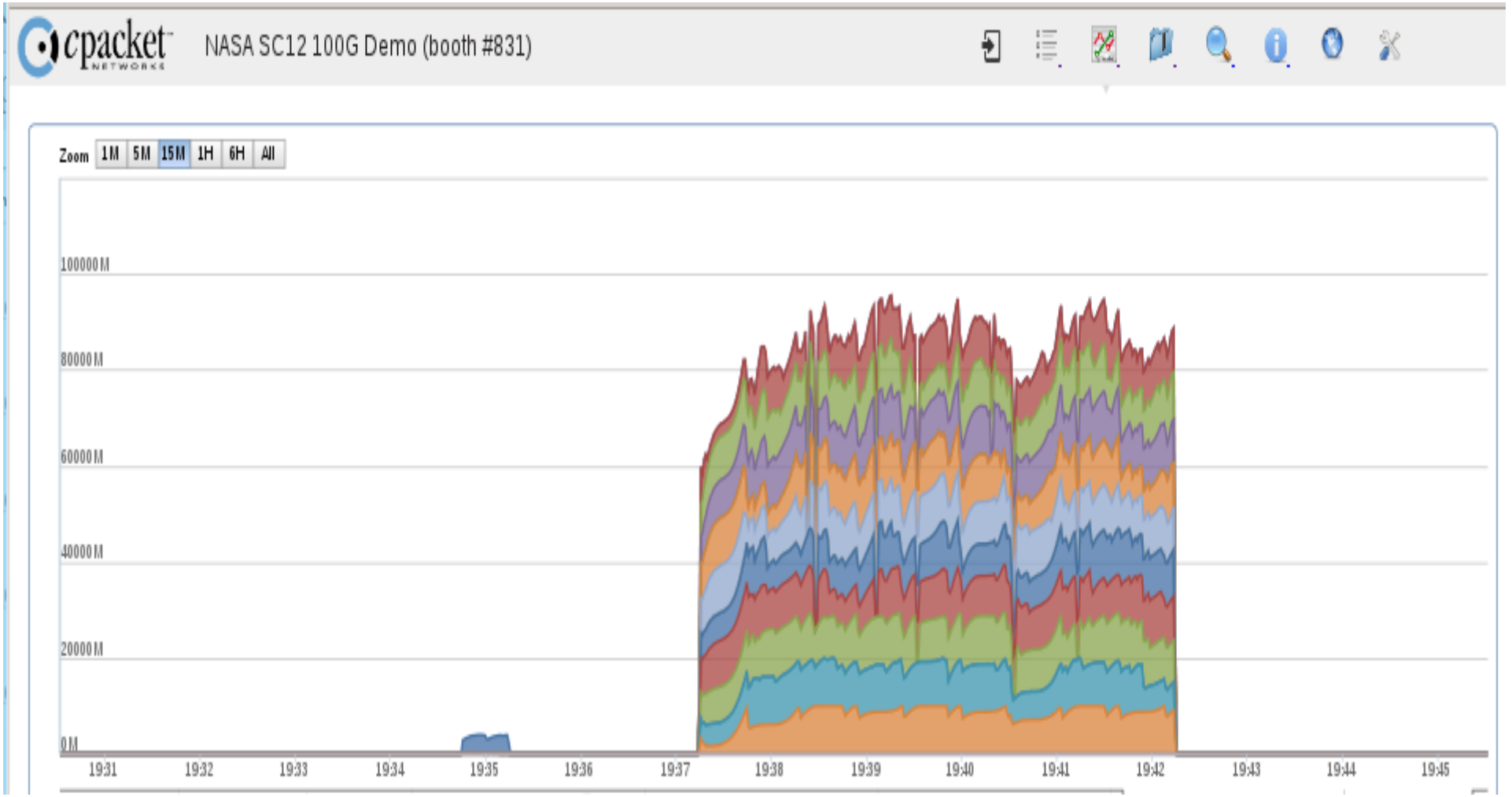
Evaluations/Demonstrations of 100 Gbps Disk-to-Disk File Transfer Performance Across LANs & WANs

An SC12 Collaborative Initiative Among NASA and Several Partners



NASA Booth to GSFC

nuttcp uni-directional tcp memory-to-memory results (83.9Gbps)



Possible SC13 demos

* Acadia Optronics 100G NIC
* Single File Transfer rather than multiple simultaneous file transfers
* 40 and 100G demo (no 10G) or maybe all 100G.
* Burst rate adjustment to assist in speed mismatches (100/40G -> 10G, or 100G -> 40G)
* Open-Flow path
* UDP file transfer methods.

X100SSD Parts/Price List

Description	Model#	Price	Quantity	Subtotals
SM 3u 16bay chassis	836TQ-R800B	$869	1	$869
IcyDock 2.5" adaptor	MB882SP-1S-2B	$17	16	$272
SM front pannel cable	CBL-0084	$3	1	$3
Fan extension cable	12" 3-pin fan extension	$1	1	$1
Red Greatland	18" Slimline SATA adapter	$6	1	$6
SM FDD tray for 2.5"	CP-220-83601-0B	$8	1	$8
Power extension cable	8" 8pin power extension	$8	1	$8
SM 2u 24bay chassis	216A-R900LPB	$869	1	$869
SM motherboard	X9DR3-F	$465	1	$465
Intel Xeon processors	E5-2687W	$1,875	2	$3,750
CPU cooler	Dynatron R14 2U	$46	2	$92
WD 500GB system disk	WD5000BPKT	$87	1	$87
DDR3 ECC Reg (4x8GB)	KVR1600D3D4R11SK4/32G	$279	2	$558
HotLava 3port 40GE NIC	3QF3A60a1	$2,200	2	$4,400
LSI 9271-8i raid	LSI00330	$658	4	$2,632
OCZ Vertex3 SSD	VTX3-25SAT3-120G	$105	32	$3,360
SAS-SATA cable	CBL-SFF8087OCF-10M	$16	4	$64
intSAS-intSAS	1 meter SFF-8087 cable	$20	6	$120
1M ext SAS cable	DMX-8088.8088-01M	$40	4	$160
SAS int/ext adapter	SAS-AD8788-2	$55	2	$110
SAS int/ext adapter	SAS-AD8788-4	$115	1	$115

SM = Supermicro. WD = Western Digital				$17,949
10M 40GE active cable	FCBG414QB1C10 (Finisar)	$400	4	$1,600

				$19,549