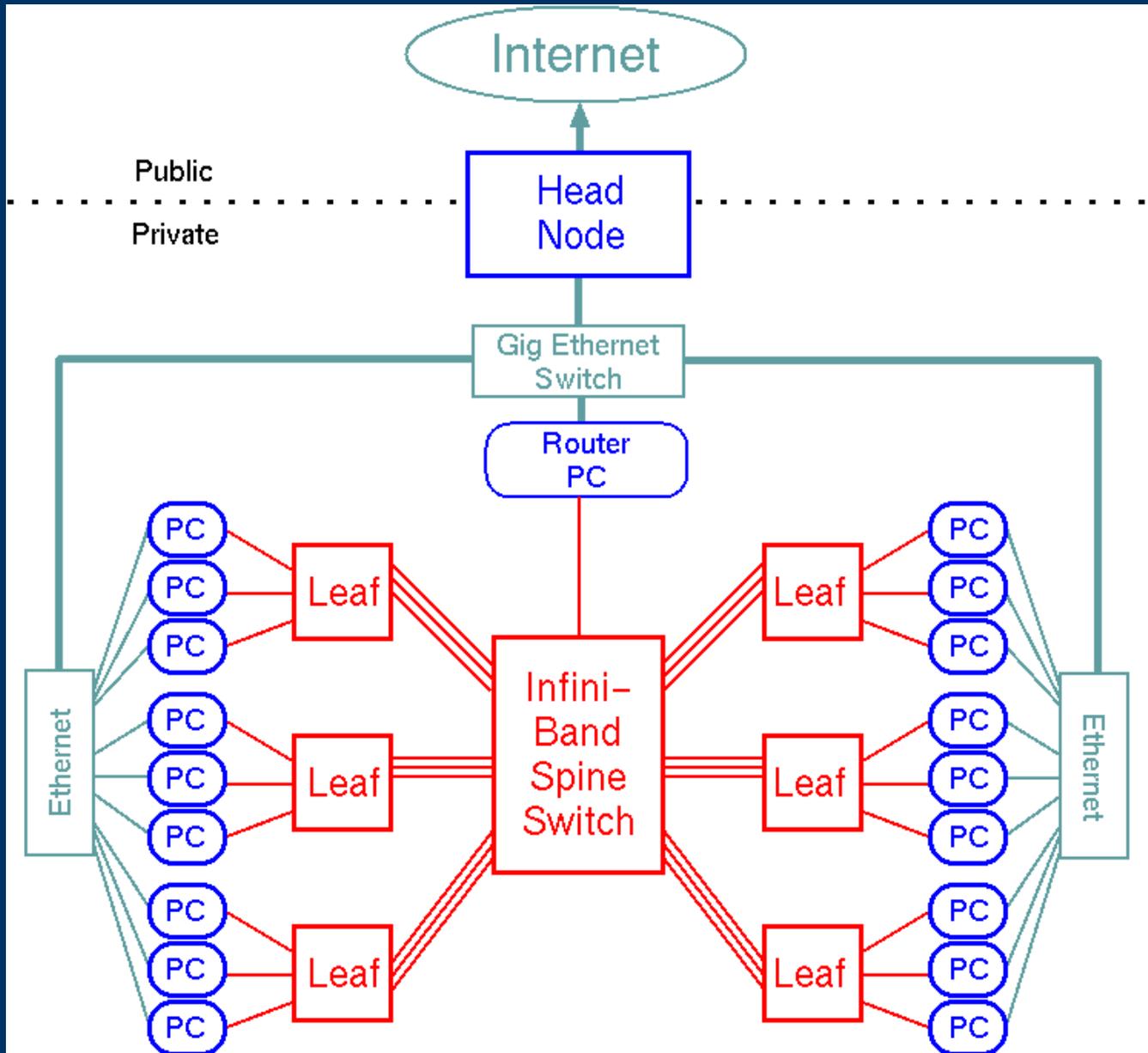


FY06 Procurement

***Don Holmgren
Lattice QCD Computing Project
Review
Cambridge, MA
May 24-25, 2005***

Standard Cluster Layout



Cost Basis – FY2005 Fermi Cluster

		Per Node Cost	Count	Total Cost
Nodes	“Buckner” Pentium 640 3.2 Ghz, 1GB memory	\$1015	260	\$263.9K
Infiniband	144-Port Spine	\$43.6K	1	\$43.6K
	24-Port Leaf	\$3318	16	\$53.1K
	HCA	\$442	260	\$114.9K
	Cables	\$70	400	\$28.0K
	Per Node Cost	\$931/\$909	260/272	\$242K
Ethernet	Switches, Cables	\$21	260	\$5.5K
Infrastructure	Serial ports, shelves, PDUs	\$51	260	\$13.1K

JLAB Cluster

- Assumed configuration:
 - Approximately 180 processors
 - Either Xeon (Pentium 6xx) or dual Opteron
 - Infiniband (or gigE mesh)
 - Pricing:
 - Estimate \$700 per node maximum for Infiniband
 - Based on FY05 cost + Mellanox estimates
 - Estimate \$1100 total for each Pentium 4 node
 - Based on FY05 cost
 - Higher Infiniband cost or lower Opteron cost may shift best price/performance to AMD
 - Release to production: March 2006
 - 450 Gflop/sec sustained (1:1 DWF:asqtad)

FNAL Cluster

- Assumed configuration:
 - 800 processors
 - Either 1066 MHz FSB Pentium 4 or dual Opteron
 - Infiniband
 - Pricing:
 - Estimate \$700 per node maximum for Infiniband
 - Based on FY05 cost + Mellanox estimates
 - Estimate \$1100 total for each Pentium 4 node
 - Based on FY05 cost
 - Higher Infiniband cost or lower Opteron cost may shift best price/performance to AMD
 - Release to production: September 2006
 - 1.8 Tflops/sec sustained (1:1 DWF:asqtad)

Schedule Details

- Details of the FNAL procurement
- JLab procurement is essentially a 1/4 scale version of FNAL, but starts earlier

WBS ID	Task Name	Start	Finish	3rd Quarter			4th Quarter			1st Quarter			2nd Quarter			
				Jul	Sep	Nov	Jan	Mar	May	Jul	Sep	Nov	Jan	Mar	May	
11	1.01.01.01	FY06 FNAL- Procure and deploy system	Mon 10/3/05	Thu 9/21/06												
12	1.01.01.01.01	Develop prototype	Mon 10/3/05	Thu 3/16/06												
13	1.01.01.01.02	Procure production hardware	Wed 2/1/06	Fri 8/4/06												
14	1.01.01.01.02.1	Procure network hardware	Wed 2/1/06	Thu 7/20/06												
27	1.01.01.01.02.2	Procure cluster computer hardware	Wed 2/1/06	Fri 8/4/06												
28	1.01.01.01.02.3	Procure miscellaneous hardware	Wed 3/15/06	Thu 6/15/06												
42	1.01.01.01.03	Upgrade site HVAC & power	Mon 10/3/05	Wed 7/5/06												
52	1.01.01.01.04	Integrate cluster	Mon 7/3/06	Thu 9/21/06												
54	1.01.01.01.04.01	Create system test procedure	Mon 7/31/06	Mon 8/28/06												
55	1.01.01.01.04.02	Prepare cluster space	Mon 7/3/06	Wed 8/2/06												
56	1.01.01.01.04.03	Install cluster computers	Tue 8/1/06	Thu 9/21/06												
62	1.01.01.01.05	Commissioning	Thu 8/17/06	Thu 9/21/06												
63	1.01.01.01.05.1	Applications-level testing	Thu 8/17/06	Thu 9/14/06												
64	1.01.01.01.05.2	Update documents	Thu 8/17/06	Thu 9/14/06												
64	1.01.01.01.05.3	FY 06 Release 1.8 Tflops to operation	Thu 9/21/06	Thu 9/21/06												

FNAL Details

- Computer room reconstruction
 - Oct. 1, 2005 - July 1, 2006
 - Existing building has 1.5 MW power, 110 tons AC
 - Sufficient for FY05 + FY06
 - Not enough capacity for FY07
 - Proposed expansions: additional 2.0 MW power, 180 tons AC
 - Sufficient for 3000+ processors
 - Budgeted off-project (FNAL base)

FNAL Details

- Prototyping
 - Oct. 1, 2005 – Feb. 1, 2006
 - Evaluate:
 - Single data rate vs double data rate Infiniband
 - 1066 FSB Intel ia32/x86_64 CPUs
 - Opteron motherboards with PCI-Express
 - PathScale Infinipath (Infiniband physical layer)
 - Dual core (AMD, Intel)

FNAL Details

- Network procurement
 - Feb. 1, 2006 – May 1, 2006
 - Infiniband based on FY05 cluster, prototyping
 - Understand oversubscription
 - Alternatives if reject Infiniband:
 - Myrinet
 - PathScale Infinipath
 - Quadrics
 - Leaf and Spine design based on FY05 results
 - Standard FNAL RFP process

FNAL Details

- Computer procurement
 - Feb. 1, 2006 – July 1, 2006
 - Node choice based on prototyping
 - Standard RFP process

FNAL Details

- Infrastructure design and procurement
 - March 1, 2006 – June 15, 2006
 - Components:
 - Ethernet (control and service network)
 - Serial lines (consoles, possibly out-of-band IPMI)
 - Layout of racks/shelves
 - Cable design
 - PDU's

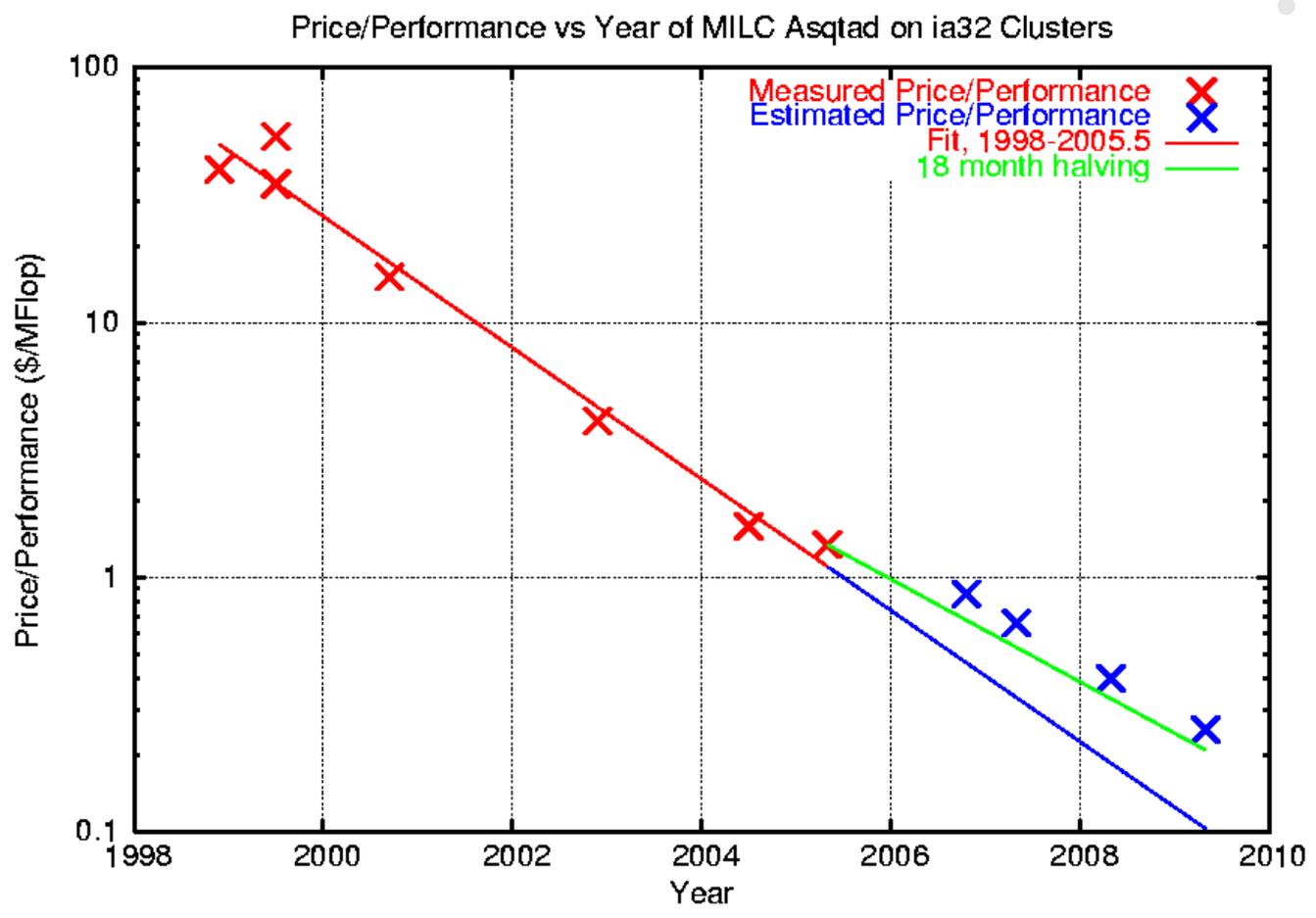
FNAL Details

- Integration and Testing
 - July 1, 2006 – Sept 15, 2006
 - Consists of:
 - Computer room preparation
 - Rack/shelf assembly and cabling
 - Node installation
 - OS installation via network
 - Network configuration
 - Unit testing
 - Application testing
 - Release to production at end

Build-to-Cost

- Fixed budget for equipment
 - # of nodes determined by per node cost
 - Performance determined by node count
 - Also, determined by components
 - Performance risk management:
 - Use conservative performance estimates
 - Use 18 month doubling times, even though we've seen faster for these applications
 - Add float to vendor roadmaps
 - Delay purchases to catch significant new components
 - Performance improvement must offset delay
 - Spending on user support versus equipment
 - Try to run lean on user support
 - But, must maximize science output
 - Evaluate annually and shift funds between effort and equipment

Performance Milestones - FY06-FY09



Measured and estimated asqtad price/performance

- Blue crosses derive from our "deploy" milestones
- Green line uses 18 month halving time

Questions?