The LQCD Project

Paul Mackenzie, FNAL
Don Holmgren, FNAL
Chip Watson, JLab

3rd International Lattice Field Theory Network Workshop
JLab
Oct. 2-6, 2005

DoE lattice QCD funding

There have been several DoE sources.

SciDAC: R&D for software, prototype clusters.

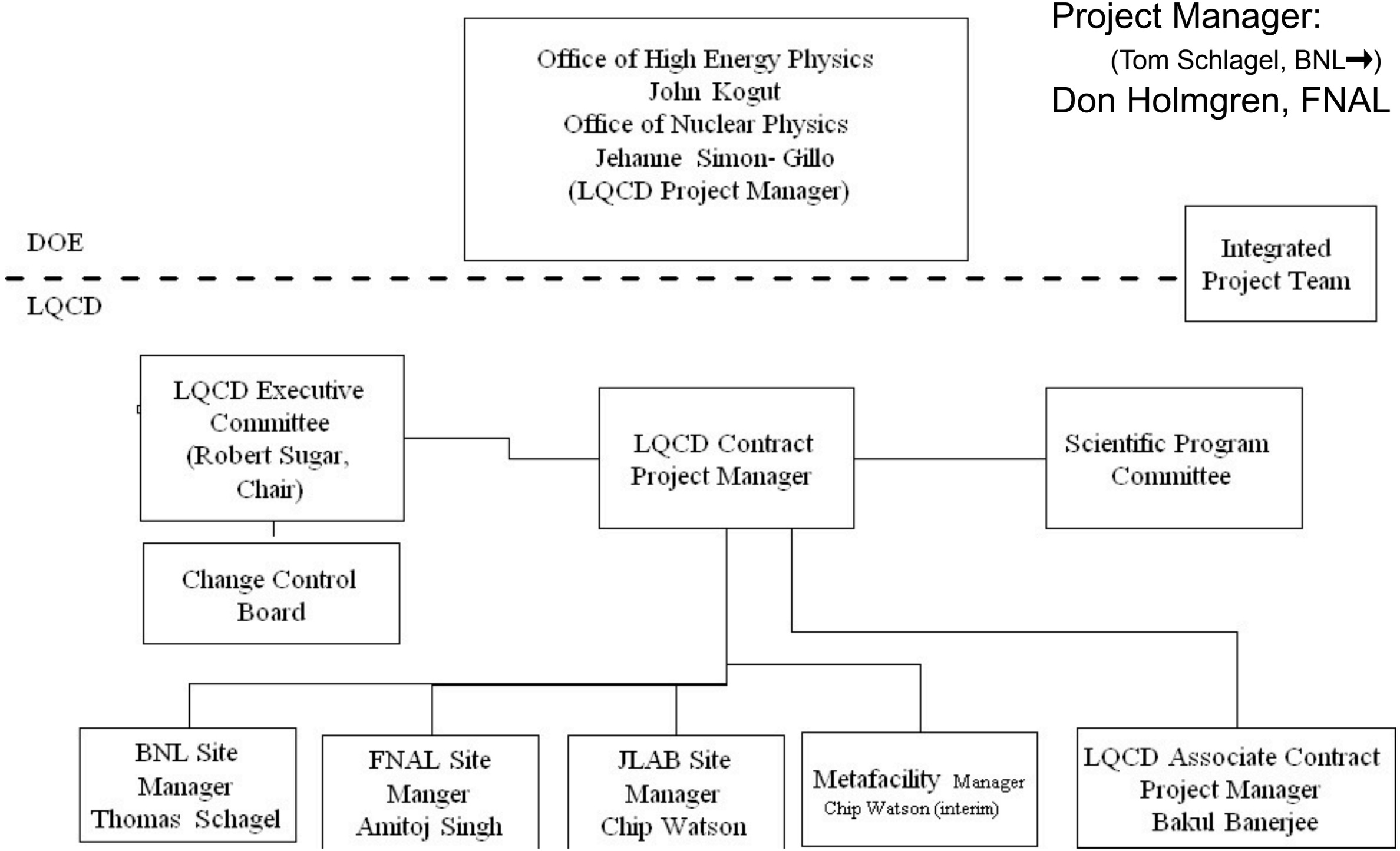
	2001	02	03	04	05	06	07	08	09	10
SciDAC	~$2 M/year					??				
				~$2.5M/y						
LQCD Facilities Project						~$2.5M/year				

QCDOC

Annual hardware upgrades.
Clusters for at least first year or two.

LQCD is a DoE "project"

Project Manager:
(Tom Schlagel, BNL→)
Don Holmgren, FNAL



Project plans

End of 2005 status

Installation	New 2005 teraflops	Total 2005 teraflops
BNL	4.2	4.2
FNAL	.86	1.0
JLab	.46	.65
Total	5.5	5.8

"The project"

Focus of this talk.

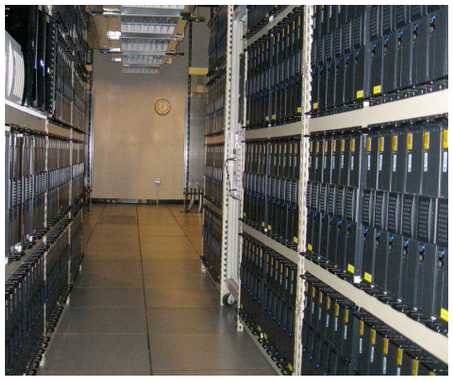
Year	$M	New teraflops	Total teraflops
2005		5.5	6
2006	2.5	2.0	8
2007	2.5	3.1	11
2008	2.5	4.2	15
2009	1.7	3.0	18

2005 Fermilab cluster

See Chip Watson's talk for current JLab clusters.

June, 2005.
256 nodes.
(=> 512 nodes).

2005 components



512 node P4 cluster.

Improvements relative to ’04 128 node cluster:

o 2.8 GHz P4 processor => 3.2 GHz

o DDR memory => DDR2

o PCI => PCI express 4x

1.06 GB/sec => 1 GB/sec in each direction.

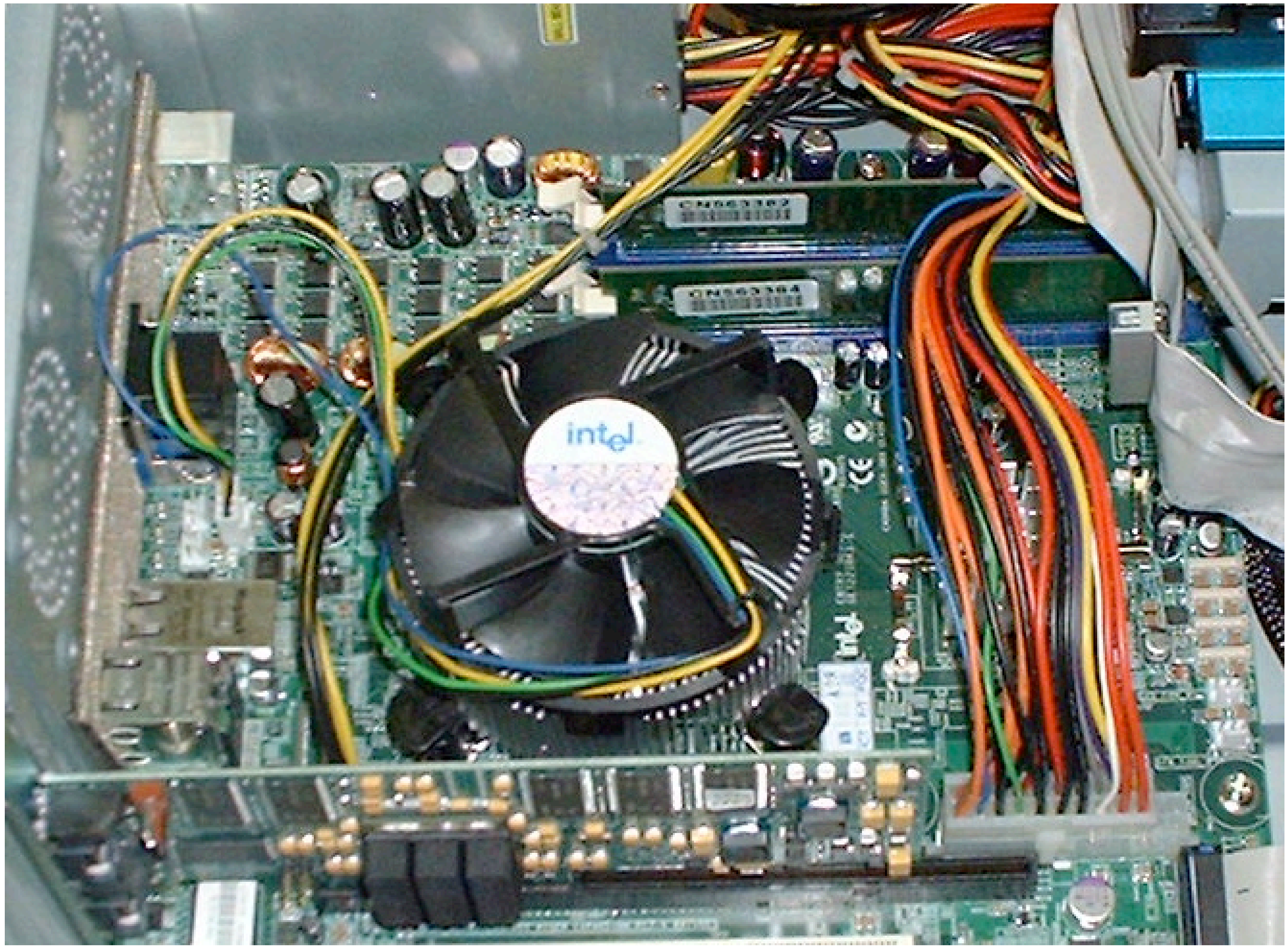
o Myrinet => Infiniband 4x

4 µsec latency, 450 MB/sec bandwidth=> 3.5 µsec latency, 2 GB/sec.

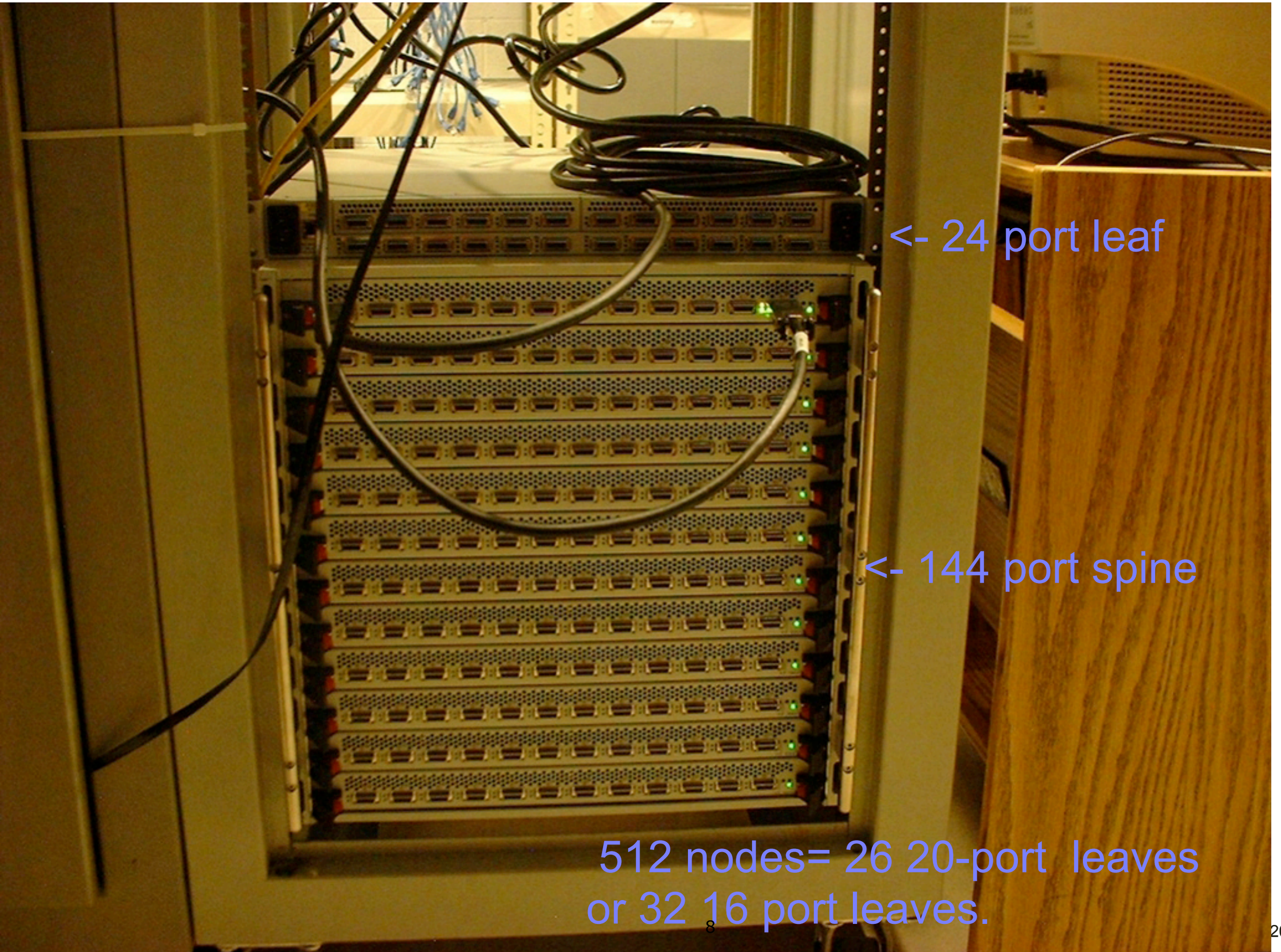
2005 node

3.2 GHz P4

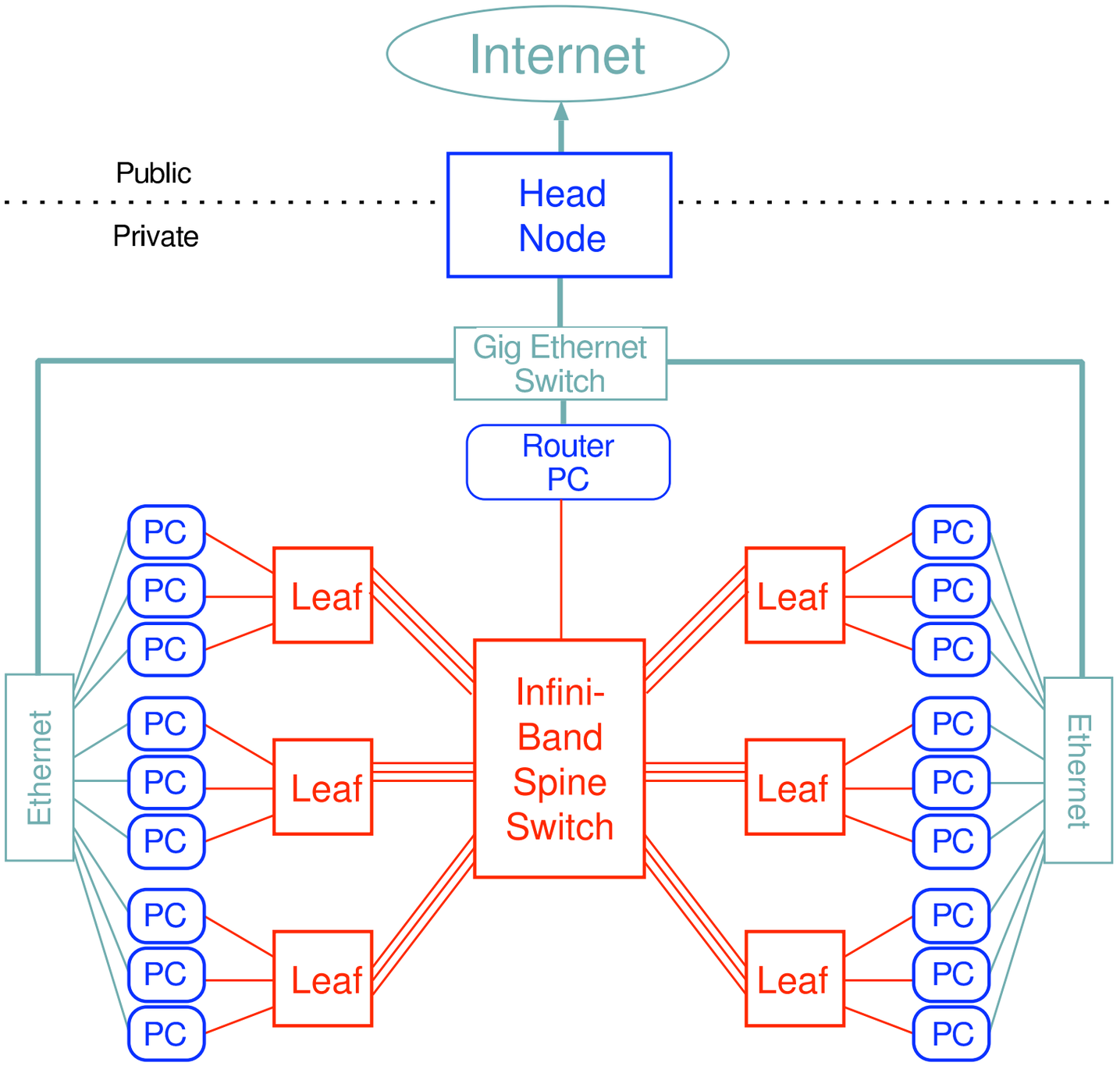
Infiniband card



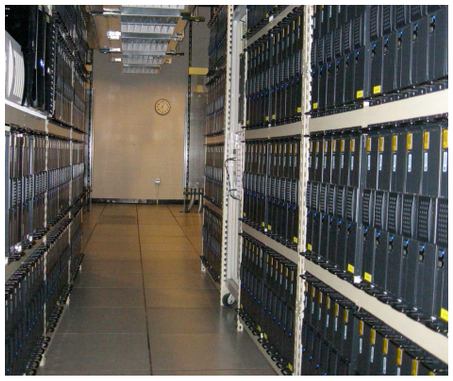
Infiniband


<- 24 port leaf
<- 144 port spine
512 nodes= 26 20-port leaves
or 32 16 port leaves.

Infiniband switch design



2005 history



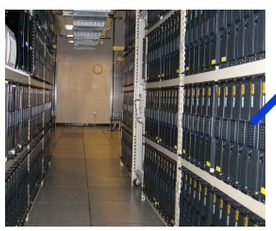
- June, machine up.
  First 256 nodes brought into friendly user production.
  - June-Sept., configuration generation:
    - a=0.18 fm
    - a= 0.15 fm
- Sept. 30, user queue started on half of new system.
- Oct. 17, next 256 nodes to be delivered.
- November, full 512 node system in operation.

http://lqcd.fnal.gov

02 cluster

04 cluster

05 cluster



Fermilab Lattice Gauge Theory Computational Facility

http://lqcd.fnal.gov/

Getting Started Latest Headlines

Lattice Gauge Theory Computational Facility

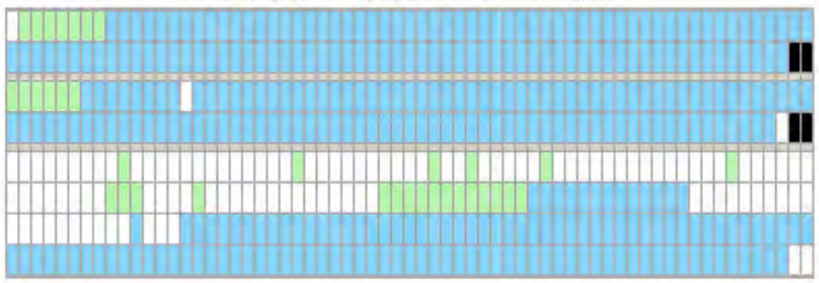
Fermilab Fermilab at Work Theoretical Physics Dept. Distributed Systesms Projects Group Integrated Systems Development Dept. Fermilab Computing Division

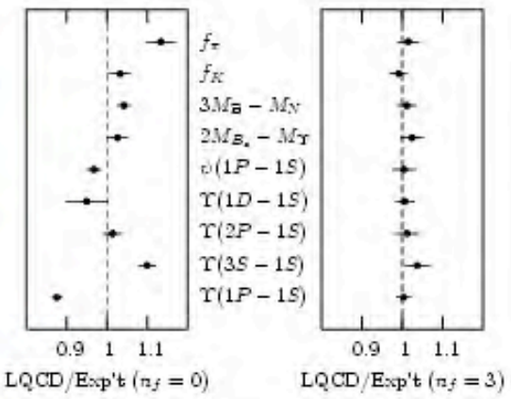
Fermilab operates large clusters of computers for lattice quantum chromodynamics, as part of the national computational infrastructure for lattice QCD established by the Department of Energy. Their goal is the understanding of the strong dynamics of quarks and gluons, which is beyond the reach of the traditional perturbative methods of quantum field theory. A central goal of the groups using the computers is the accomplishment of the calculations required to extract from experiment the fundamental parameters of the Standard Model of particle physics.

Physics Program

LQCD Cluster Status

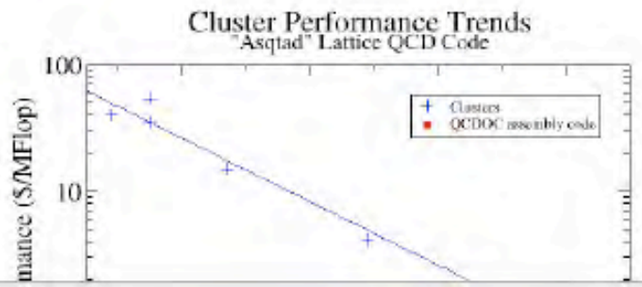
Tue Oct 4 10:02:24 CDT 2005





The QCD clusters

The QCD cluster future



- Fermilab Lattice QCD
- Lattice 2004
- Commodity Hardware for Lattice QCD
- DOE, SciDAC Support for Lattice QCD Computing

LQCD User Information:
- New accounts and renewed accounts
- Basic Computer Security Training Requirement
- Notes for Users
- Kerberos Notes for Users
- Additional Kerberos Information
- Transferring Files Between USQCD Sites
- LQCD-Users Mail Archive
- Draft Run Time Environment Specification
- Presentation on User Environment
- Run Time Environment Primer

System status
- Cluster Status
- New Muon Temperature
- Help: lqcd-admin@fnal.gov
- LQCD-Admin Mail Archive

Done

05 runs

Fill out asqtad data set on coarser lattices.

Operations counts for unquenched improved staggered configuration generation.

	a (fm)	m light	m heavy	Ns	Nt	Volume	CG l	CG h	Ops/site	steps	Ops/traj	traj	TF years
Fermilab	0.18	0.0492	0.082	16	48	196608	170	142	1164472	50	1.14E+13	3000	0.0011
		0.0328		16	48	196608	170	142	1164472	50	1.14E+13	3000	0.0011
		0.0164		16	48	196608	170	142	1164472	100	2.29E+13	3000	0.0022
		0.0082		16	48	196608	500	142	1556182	200	6.12E+13	3000	0.0058
	0.15	0.0484	0.0484	16	48	196608	138	138	1121740	100	2.21E+13	3000	0.0021
		0.029		16	48	196608	206	138	1202456	100	2.36E+13	3000	0.0023
		0.0194		16	48	196608	281	138	1291481	100	2.54E+13	3000	0.0024
		0.0097		16	48	196608	430	138	1468344	150	4.33E+13	3000	0.0041
		0.0048		16	48	196608	890	138	2014364	333	1.32E+14	3000	0.0126
Supercomputers	0.125	0.04	0.05	20	64	512000	170	142	1164472	50	2.98E+13	3000	0.003
	"coarse"	0.03	0.05	20	64	512000	212	142	1214326	50	3.11E+13	3000	0.003
		0.02	0.05	20	64	512000	253	127	1245188	75	4.78E+13	3000	0.005
		0.01	0.05	20	64	512000	426	127	1450539	150	1.11E+14	3000	0.011
		0.007	0.05	20	64	512000	583	127	1636898	200	1.68E+14	3000	0.016
		0.005	0.05	24	64	884736	893	143	2023860	333	5.96E+14	3000	0.057
QCDOC	0.09	0.0124	0.031	28	96	2107392	352	189	1436295	125	3.78E+14	3000	0.036
	"fine"	0.0062	0.031	28	96	2107392	687	189	1833940	250	9.66E+14	3000	0.092
		0.0031	0.031	40	96	6144000	1400	189	2680271	500	8.23E+15	3000	0.786
	0.06	0.008	0.02	42	144	10668672	355	300	1571613	188	3.15E+15	3000	0.301
		0.004	0.02	42	144	10668672	1030	300	2372838	375	9.49E+15	3000	0.907
		0.002	0.02	60	144	31104000	1050	300	2396578	750	5.59E+16	3000	5.339

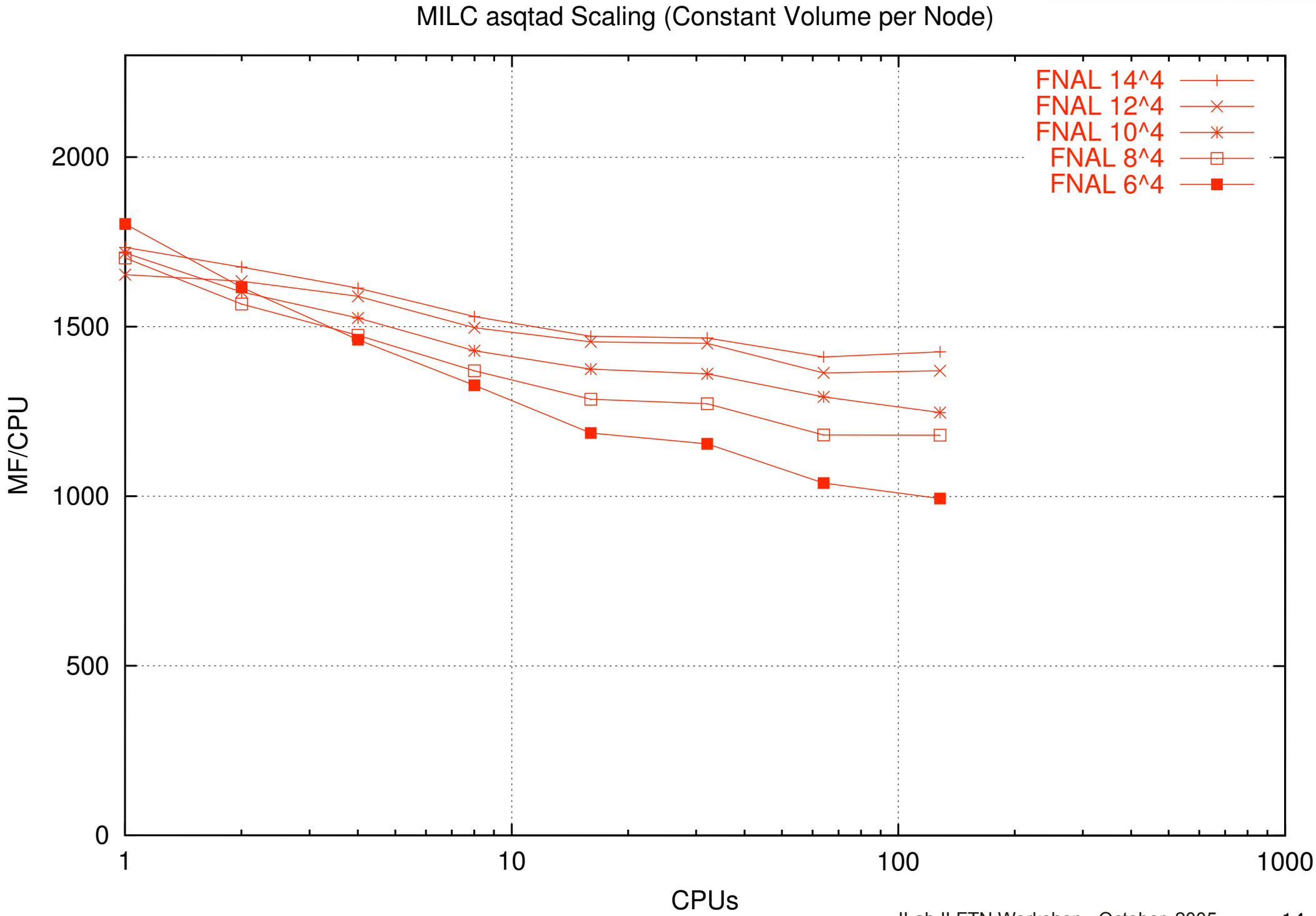
Ops/site=1187*(CG l + CG h)+794128.

05 runs

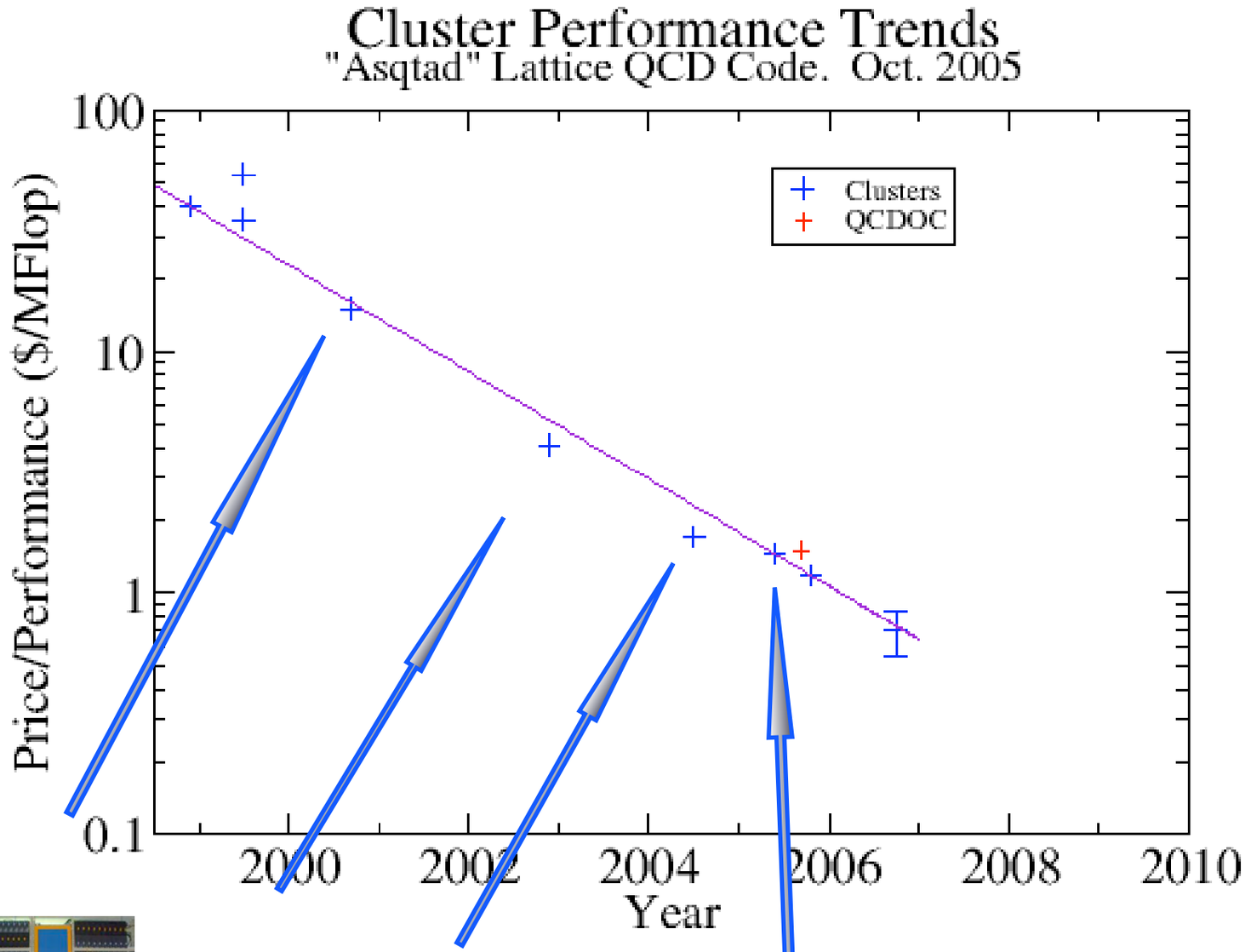
a (fm)	m_l/m_s	V	Nodes	MF/node	trajectories
0.12	0.1	24^3x64	108		
0.15	1.0	16^3x48	48	1215	1600-3200
	0.5		48	1215	1985-3200
	0.4		64	1185	2400-3200
	0.2		48	1200	2090-3200
	0.1	20^3x48	128	1200	0-3200
0.18					

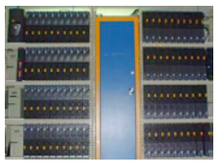
Performance benchmarks

asqtad CG.



05 price/performance





40^3x96 asqtad CG on 256 nodes.

1300 MF/node.

June 256 nodes:

Node + Infiniband = $1015+$890 = $1900.

$1.45/MF.

Oct. 256 nodes:

Node + Infiniband = $838+$660 = $1500.

$1.15/MF.

Expected performance of 512 node 2005 system:

512 x (1.3-1.4 GF) = 650-700 GF.

Most will be used for valence analysis of configurations generated on the QCDOC.

2005 allocations for staggered configuration generation on the QCDOC:

a (fm)	ml/ms	volume	TF years
0.09	0.1	40^3 x 96	0.43
0.06	0.4	48^3 x 144	0.55
0.06	0.2	48^3 x 144	1.02

These could have been generated on 512 clusters
(by combining cold and hot starts), had that been necessary.

2006 hardware plans

End of SciDAC,
First year of “the project”.

Clusters.

Alternatives: Blue Gene, more QCDOC.

Expected component improvements:

Infiniband integrated on motherboard. NOPE.

800 MHz bus => 1066MHz.

Dual core chips.

Fully buffered Dimms, *independent memory busses*.

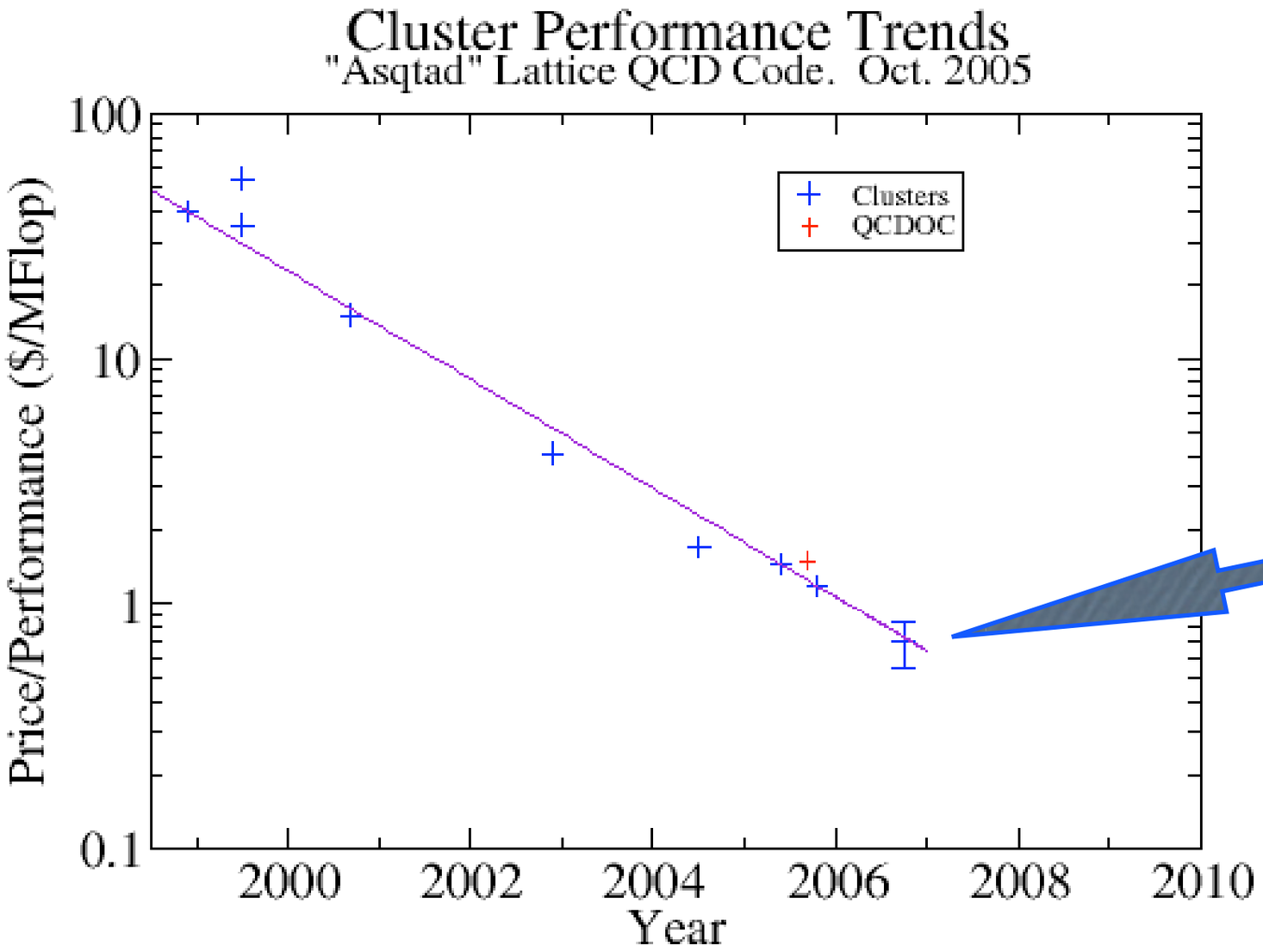
2006 system plans

$1.77M for hardware.
  1280 processors (?)=640 dual Intel nodes,
  4 GF/node,
  2+ TF system.

Schedule for specifying and purchasing 2006 clusters.

	2005				2006			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Benchmarking								
RFP								
Operations								

06 price/performance goals



Targeted price/performance of the 2006 system: $0.75+/-0.15 $/MF.

Assumes 1066MHz bus, dual core chips, fully buffered Dimms

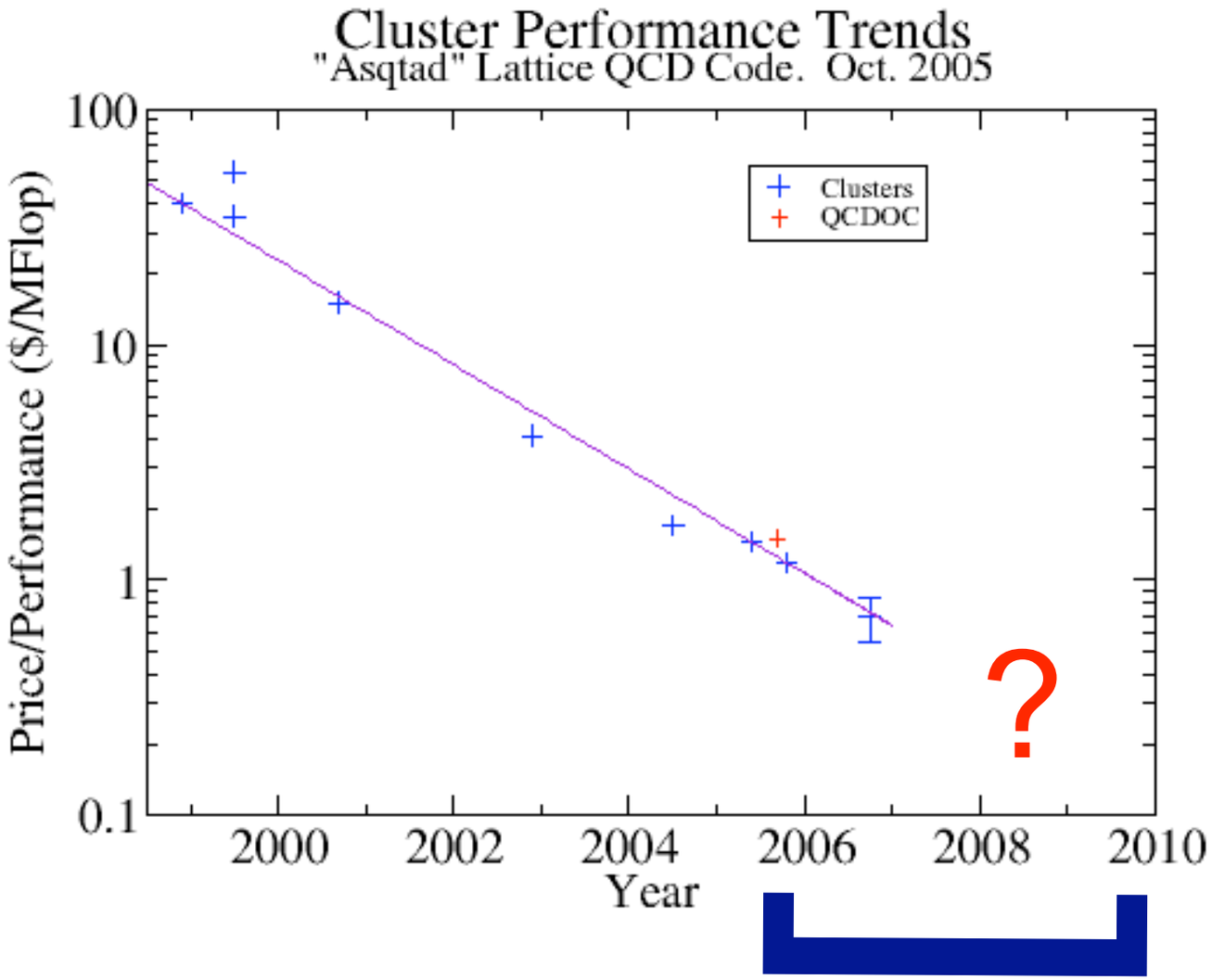
Beyond 2006 cluster components

Coming improvements:

Pretty soon, ‘07?

- 1066=>1333 FSB
- Infiniband on board
- Infinband 4x=>24x.
  - Already 12x between switches.
  - 06: 12x switch to node.
  - 07: 24x.
- Dual issue FPUs (2=>4 ops/cycle)
- Dual core => quad core

2007 and beyond?



Year	$M	New teraflops	Total teraflops
2005		5.5	6
2006	2.5	2.0	8
2007	2.5	3.1	11
2008	2.5	4.2	15
2009	1.7	3.0	18

LQCD project