Future Computing Resources for Lattice Gauge Theory

Don Holmgren

Fermilab

djholm@fnal.gov

PXPS12

June 19, 2012

Short Term

Hardware Resources

- At present, US lattice theorists who are members of USQCD have access to these computing resources:
  - ANL's Intrepid (BlueGene/P) via Incite allocations to USQCD
  - ORNL's Jaguar (XT5 → XK6) via Incite allocations to USQCD
  - FNAL's conventional (JPsi, Ds) and GPU-accelerated (Dsg) clusters via USQCD allocations
  - JLab's conventional (9q, 10q, 12s) and GPU-accelerated (9g, 10g) clusters via USQCD allocations
  - Other DOE and NSF resources via grants to individual PIs
- New hardware will come online later this year at JLab (12g GPU cluster) and BNL (BG/Q)
- The FNAL and JLab hardware are purchased and operated under the DOE SC LQCD-ext project
  - FY2012 – FY2014 plans were approved at our last annual review (May 16-17, 2012)

SC LQCD-Ext Project

- A five year (FY10-FY14) extension of the four year (FY06-FY09) SC LQCD project, which built and operated hardware dedicated to LQCD calculations
- SC LQCD had a total cost of $9.2M (hardware: $5.87M = 64%, operations: $2.95M = 32%, project mgmt: $0.38M)
- USQCD successfully argued for the LQCD-ext project with an increase in budget, with a five year total of $18.15M (hardware: $10.4M = 57%, operations: $6.9M = 38%)
- In 2009, through the nuclear physics program office, an additional $5M (hardware: $3.7M = 74%, operations: $0.93M =19%) for a dedicated LQCD facility at JLab was granted as part of the American Recovery and Reinvestment Act (ARRA)
- We will need to start the process for a follow-on project soon

Near Term Resources – BG/Q

- USQCD will have access to 10% of a BlueGene/Q “DD2” hardware prototype rack via LQCD-ext operations funds
  - Full rack is 210 TF (peak),1024 nodes (16K cores)
  - Optimized double (single) precision DWF inverter performance is 42.5 GF (62.5 GF) per node, or 20.7% (30.5%) of peak
  - James Osborn reports MILC staggered performance on ANL BG/Q (“Mira”) hardware as 12.5% of peak
  - Considerable USQCD effort is on-going to optimize software in anticipation of significant Incite time at ANL
- ANL “Mira” will have 48 BG/Q racks. If USQCD receives the same 3.5% fraction of the BG/Q that they were allocated on the BG/P in 2012, this is the equivalent of 1.7 “Q” racks.
- LQCD-ext will decide by August whether to use FY13 funds for a half- or full-rack of BG/Q at BNL, or to use the funds for clusters

Near Term Resources – LQCD-ext

Cluster capacity by year for SC LQCD-ext (if no BG/Q at BNL):

	FY2012	FY2013	FY2014
Computing hardware budget (excluding storage)	$1.875M	$2.46M	$2.26M
Capacity of new cluster deployments, TFlop/s	10-15	15-22	22-33
Total cluster delivery, TF-yrs (JPsi Core-hrs)	35 (203M)	50-55 (290M-319M)	61-78 (354M-452M)
Capacity of new GPU deployments MGPU-Hrs/Yr	2.9-4.3	4.6-6.9	7.5-11.2
Total GPU delivery, MGPU-hrs/Yr	4.7	7.6-9.0	12.2-15.9

- Baseline computing hardware budgets are shown. Future increases in storage would lower numbers.
- Ranges reflect 40%-60% budget allocations to conventional vs. GPU clusters, TBD annually
- JPsi = ~ 50 M core-hrs, Ds = ~ 130 M core-hrs, FNAL + JLab = ~ 650 GPUs
- MGPU-Hrs/Yr figures are based on FY11-model GPUs (NVIDIA “Fermi”)
  - New GPU models will deliver more than 1 “Fermi” hour per wall-clock hour
  - 2012 GPU allocations total 4.7MGPU-Hrs (JLab 9g/10g = 3.6M, FNAL DsG = 1.1M)

Storage

Storage Resources (LQCD-ext Only)

- Disk:
  - FNAL: 540 TB now, + 200 to 300 TB per year through 2014 (rate increasing)
  - JLab: 380 TB now, + 100 to 200 TB per year through 2014 (rate increasing)
  - BNL: 10% of 440 TB now, could grow to some fraction of a total of 2.5 PB
- FNAL and JLab primary storage areas use Lustre (a love/hate relationship for the admins)
- FNAL read rates over several months have averaged about 3 TB/hour sustained, with hour or longer peaks to 12 TB sustained (aggregate across all clusters)
  - With optimal file layouts and striping, sustained rates could go 2 to 3 times higher
  - We don't have a good handle on either rates delivered to individual jobs, or to rates required by different job types – user feedback is requested
- Inter-site file transfers? Storage of data from (at?) ANL Mira and ORNL Titan?

Storage and Workflow

- It is a common practice in the LQCD workflows to use small files, and file and directory names, to record parameters and job states, and somewhat larger files to hold (usually in ascii) summary numeric data that is later used in fitting
  - At both Fermilab and JLab, the scale of this practice is on the order of 10-100 million individual files
  - As I reported at the All Hands Meeting in April, this is not a good match to Lustre because of latencies in accessing metadata (i.e., doing "ls –l" or the equivalent). Half of our files (about 50 M), but only 1% by volume (less than 5 TB), are used for this purpose
  - Since these 50 M files hold the "state" of many simultaneous campaigns owned by a large number of allocations, they are clearly precious and we do our best to maintain good incremental backups
  - This is not sustainable in the long term. The community should think about expanding the practice of moving to databases to hold campaign states

Storage and Workflows

- During SciDAC2, there was a subproject that concentrated on workflow systems for LQCD – initially the idea was to help move to database-oriented workflows
  - This is a very difficult problem
  - Existing workflow systems (Kepler, Pegasus, Swift, Askalon) all had (different) fundamental problems that made them a difficult match to existing LQCD practices, and they were all GRID-centered
  - LQCD workflows that we examined were more complex than those systems generally addressed, with more parameterization and greater need for persistence and error detection/recovery across many months
  - As is typical with LQCD, the field has already created many (scripting) solutions to complex job campaigns. Porting or converting these sets of scripts to workflow languages requires a lot of debugging with only a distant payoff

Longer Term

- In LQCD-ext project plans, we rely on extrapolating from long term trends to predict future price/performance
- In 12 years of building clusters for LQCD, I have never known Intel/AMD/Motorola/IBM to have more than 2 (very occasionally 3) years in their roadmaps, and they often take fairly sharp turns in reaction to the market
- Networking follows processors, as it depends on I/O chipsets. Vendor roadmaps are rarely longer 2 years.
- So, everything further out than 3 years is pure speculation

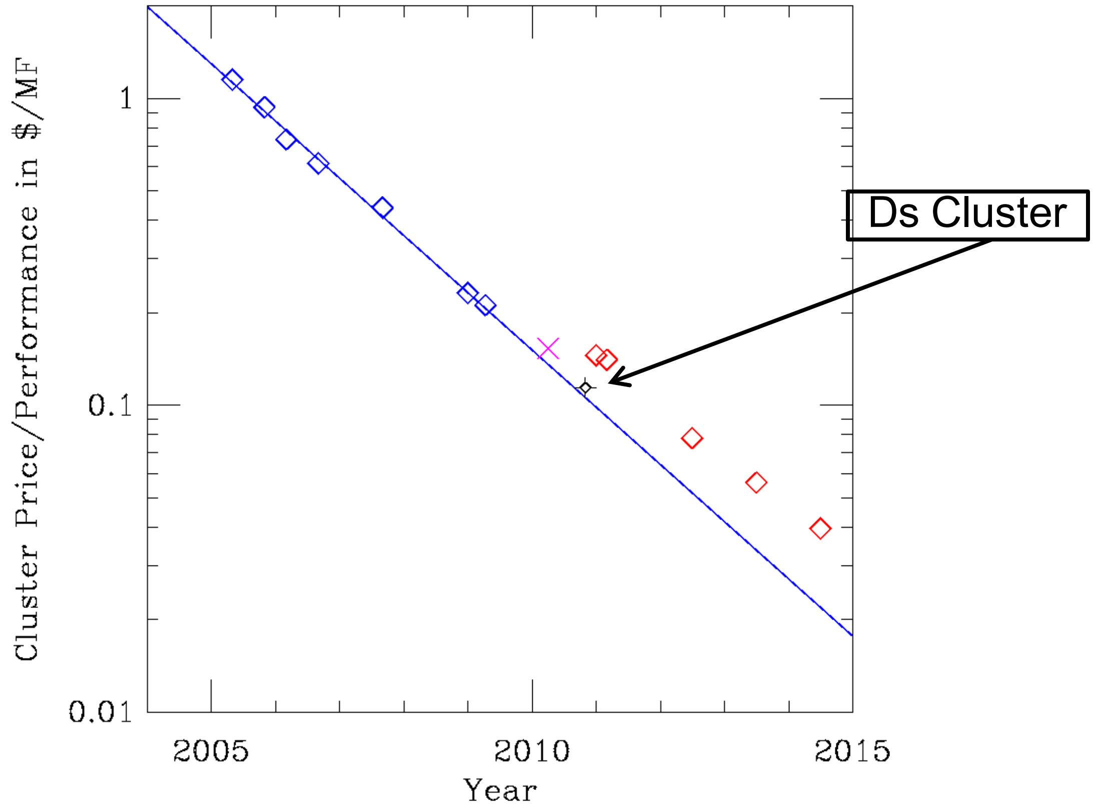
Computational Requirements

- Either memory bandwidth, ~~floating point performance~~, or network performance (bandwidth at message sizes used) will be the limit on performance on a given parallel machine
- On single nodes memory bandwidth in the constraint that limits performance
- On parallel computer clusters, the constraint is either memory bandwidth or network performance, depending upon how many nodes are used on a given job
  - Network performance limits strong scaling:
    Surface area to volume ratio increases as more nodes are used, causing relatively more communications and smaller messages

Commodity Hardware

- LQCD has long benefited from riding down the commodity price/performance curve (Moore's Law)
  - Prior to 2006, new processor models consistently had higher clock frequencies, improving execution units, and steadily increasing memory bandwidth (through faster interfaces)
  - From 2006 forward, processing capability increases per socket have resulted from multiple cores, rather than higher frequency
  - With AMD leading the way, multiple processor sockets (SMP motherboards) and later multi-core processors used NUMA (non-uniform memory access) architectures to effectively widen memory buses. Intel finally followed suit in 2009 (Nehalem).
  - Inherently parallel numerical algorithms (e.g. those used in LQCD) implemented with message passing easily realized the benefits of these developments. The latest USQCD AMD cluster has 32 cores/node.
  - *It is not clear whether the commodity market will support continuing increases in core counts on x86 processors. Is the Beowulf honeymoon ending??*

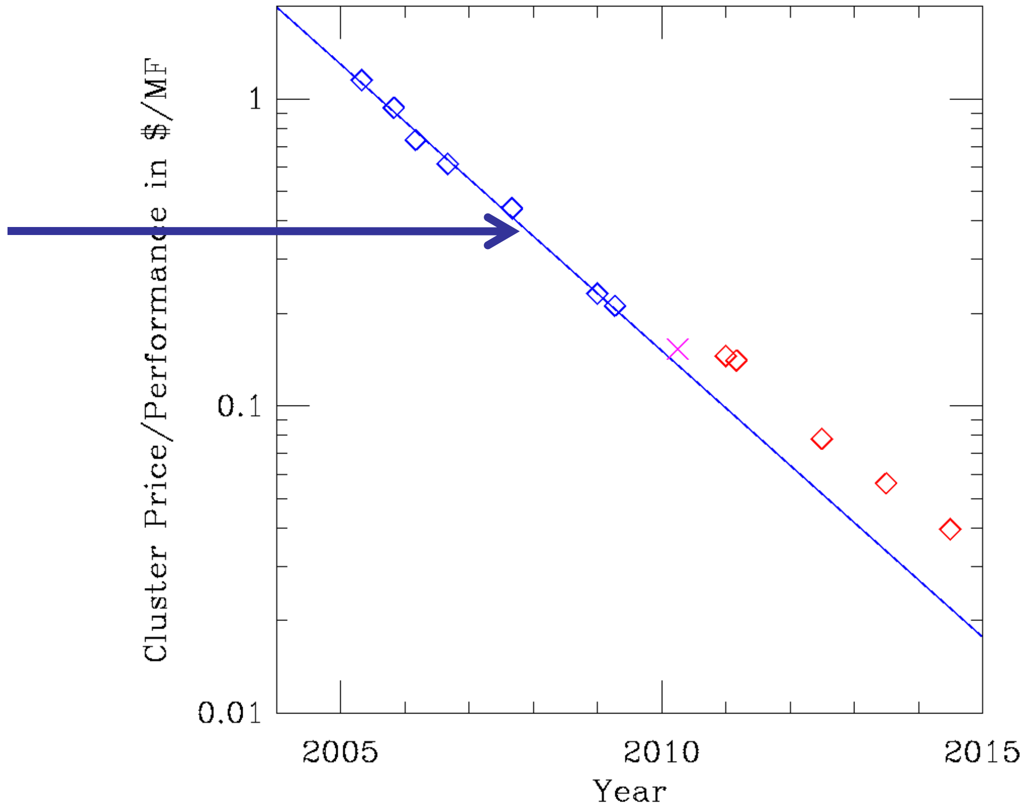
LQCD Commodity Price/Performance Trend



Cluster	Price per Node	Performance/Node, MF	Price/Performance
Pion #1	$1910	1660	$1.15/MF
Pion #2	$1554	1660	$0.94/MF
6n	$1785	2430	$0.74/MF
Kaon	$2617	4260	$0.61/MF
7n	$3320	7550	$0.44/MF
J/Psi #1	$2274	9810	$0.23/MF
J/Psi #2	$2082	9810	$0.21/MF
10q	$3461	22667	$0.15/MF
Ds	$5810	50810	$0.114/MF

Cost and Performance Basis

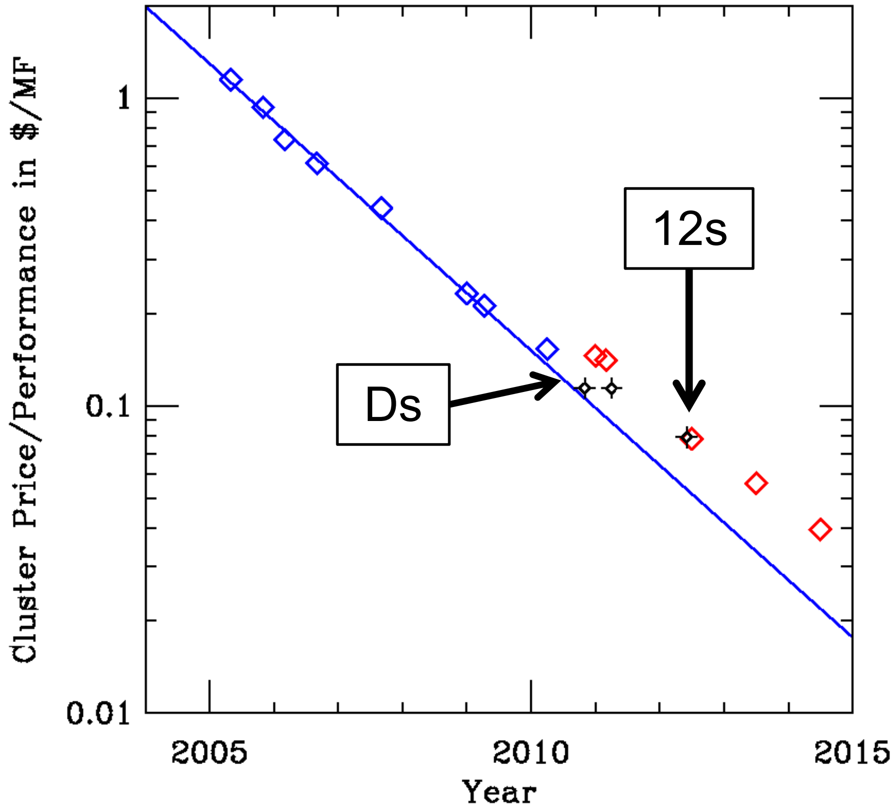
Fit is to the blue diamonds, slope gives halving time of 1.613 years



From 1998 to 2005, the slope was steeper (1.25 years), however, the clusters studied were significantly smaller and network hardware was less expensive.

Intel and AMD

- AMD Opteron has captured most of the “wins” for LQCD/LQCD-ext
  - Kaon, JPsi, 7n, Ds
  - They consistently had better memory bandwidth and lower prices
  - “Magny-Cours” (up to 48 cores in a 4 socket system) has been a workhorse platform for LQCD since 2010
  - The successor, “Bulldozer”, has good ideas, but AMD has not revised sockets to add memory channels, so “Interlagos” (socket compatible with Magny-Cours) is no better than its predecessor for LQCD
- We will likely use Intel for clusters for the next several years
  - “Sandy Bridge” is (finally) memory-bandwidth competitive with Magny-Cours, and dual-socket SB is as fast for LQCD as a four-socket Opteron
  - Next year’s “Ivy Bridge” will have only modest improvements
  - PCIe implementation (pins moved to CPUs) is better on Intel, which improves bandwidth for Infiniband and for GPUs
  - Pricing without real HPC competition from AMD is a worry



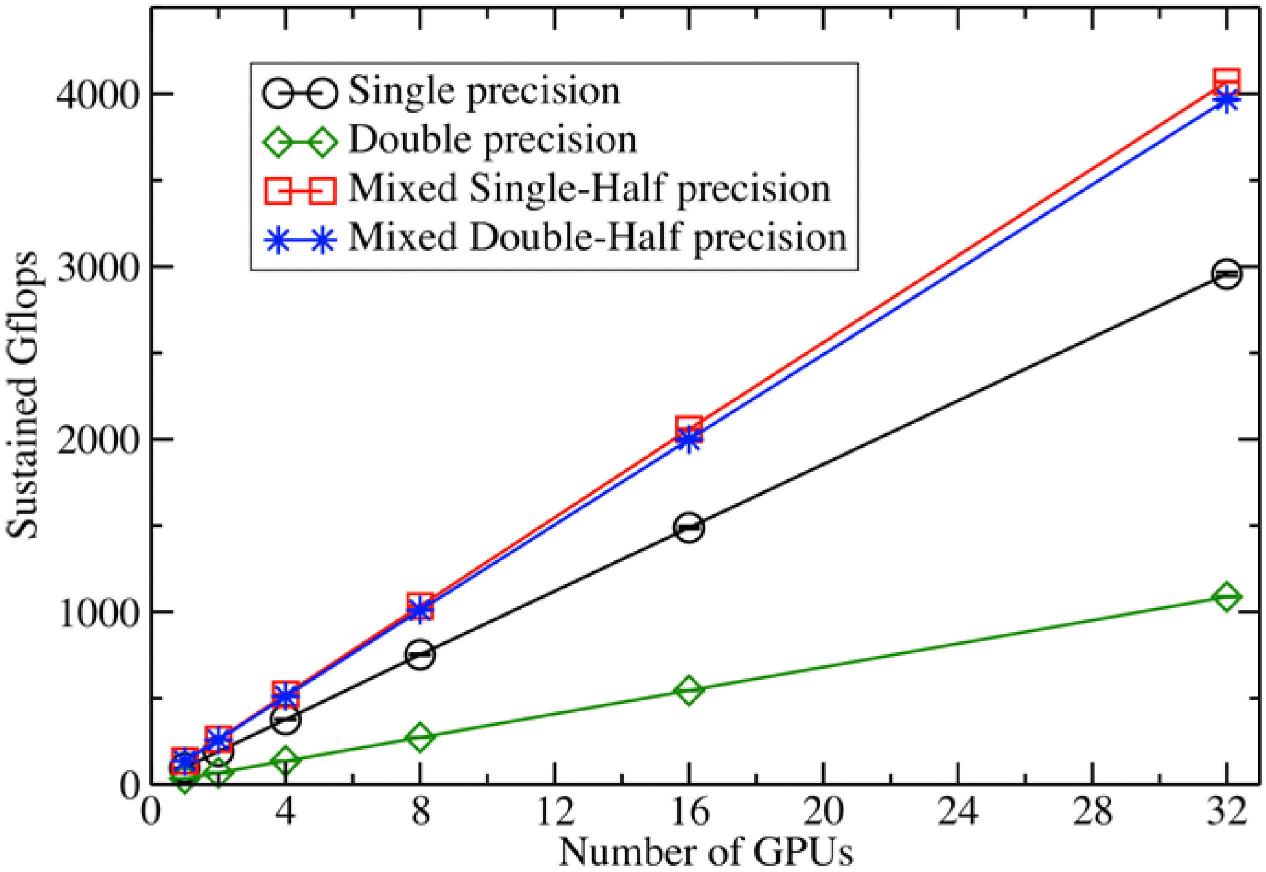
Reasons why LQCD conventional clusters are moving above the trend line:

- Intel was at least one year late in bringing out server-class Sandy Bridge CPUs
- Quad-socket Sandy Bridge systems are still not available
- Roadmap changes at AMD have (indefinitely?) delayed the previously planned “Socket G2012” version of Bulldozer which would have greatly increased memory bandwidth
- AMD changes have resulted in less price pressure on Intel for HPC hardware, and likely enabled Intel to delay Sandy Bridge and Ivy Bridge since Nehalem/Westmere were still profitable

GPU-Accelerated LQCD

- Lattice theorists in Europe started using GPUs in 2005, coding in OpenGL with Cg
- Barros *et al*. (http://arxiv.org/abs/0810.5365) at Boston U. implemented a Wilson-Dirac operator in CUDA in 2007 with about a 10x speedup over conventional x86 hardware
- The Boston U. work evolved under the SciDAC-2 program into *QUDA* (http://lattice.github.com/quda/)
  - Initially a single GPU code (http://arxiv.org/abs/0911.3191) for Wilson-Dirac
  - Mixed-precision algorithms were adapted and exploited to improve performance
  - Reduced representations were used to minimize memory operations and increase performance
- *QUDA* has since further evolved:
  - Support for other actions: Clover, twisted-mass, asqtad/HISQ (with Indiana U. and NCSA), domain wall
  - Multi GPU support (Boston U. + JLab, http://arxiv.org/abs/1011.0024)
- LQCD GPU talk at SC11:
  - "Scaling Lattice QCD beyond 100 GPUs" http://arxiv.org/abs/1109.2935

Weak Scaling Results with Mixed Precision Optimizations



Gflops shown are *effective* (*i.e.* they are the equivalent Gflops that would be observed on conventional CPUs for the same rate of work on the same problem.

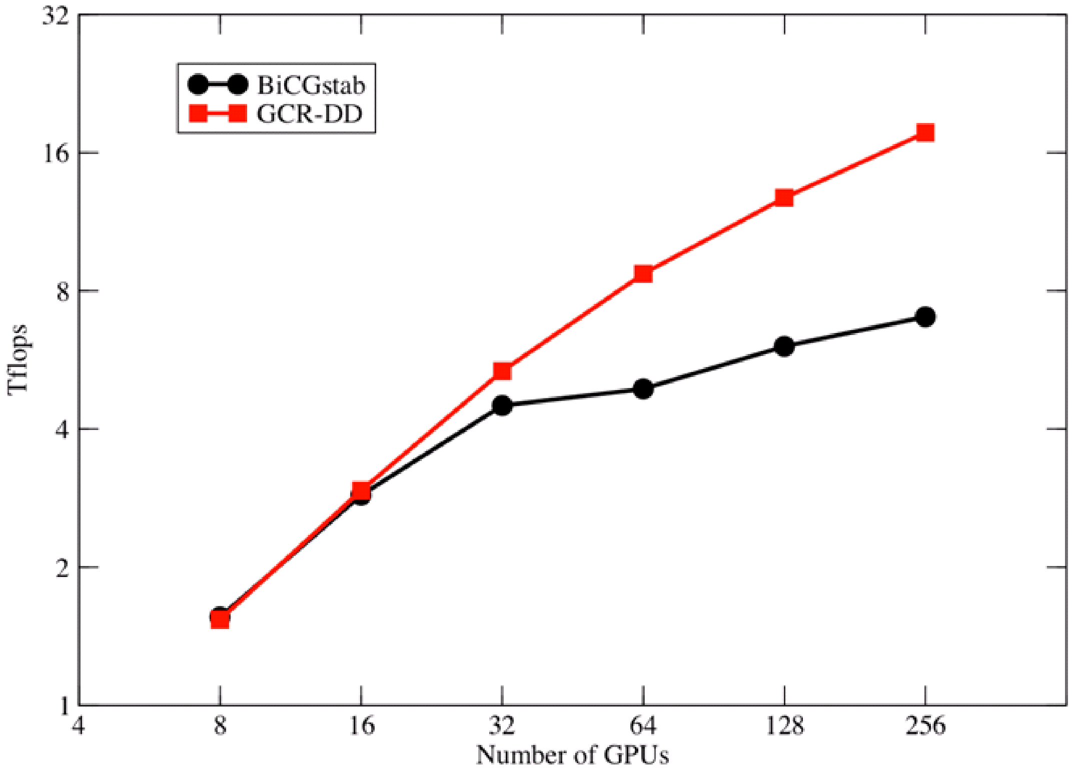
A Ds (32-core) node sustains ~ 50 GF (single)

Data are from the Jefferson Lab “9g” cluster, using GTX-285 GPUs.

These optimizations (mixed precision, reliable updates) are also relevant to conventional processors.

Algorithms can be more important than hardware for performance.

Strong Scaling Performance

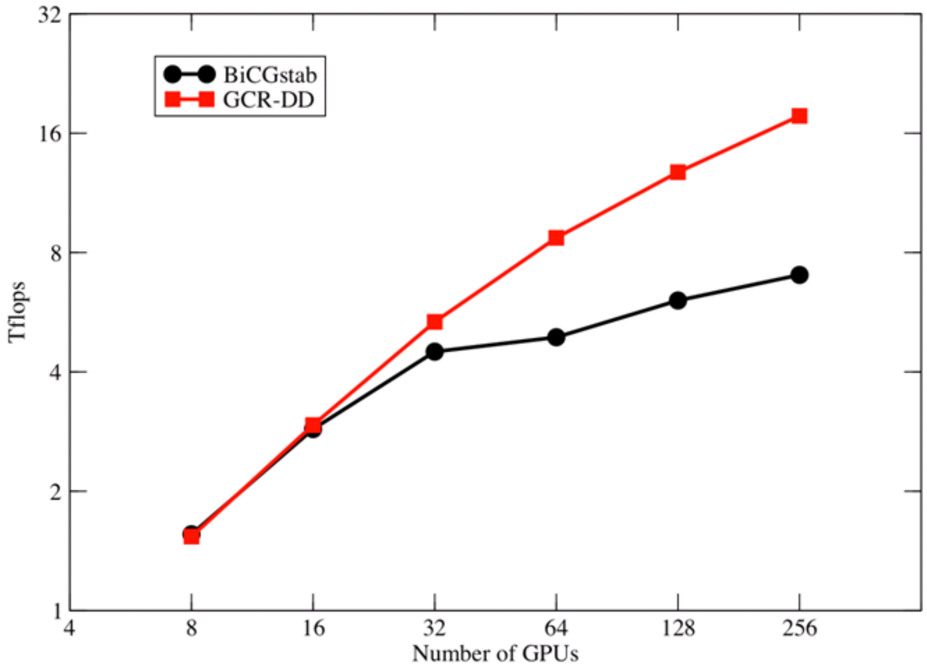


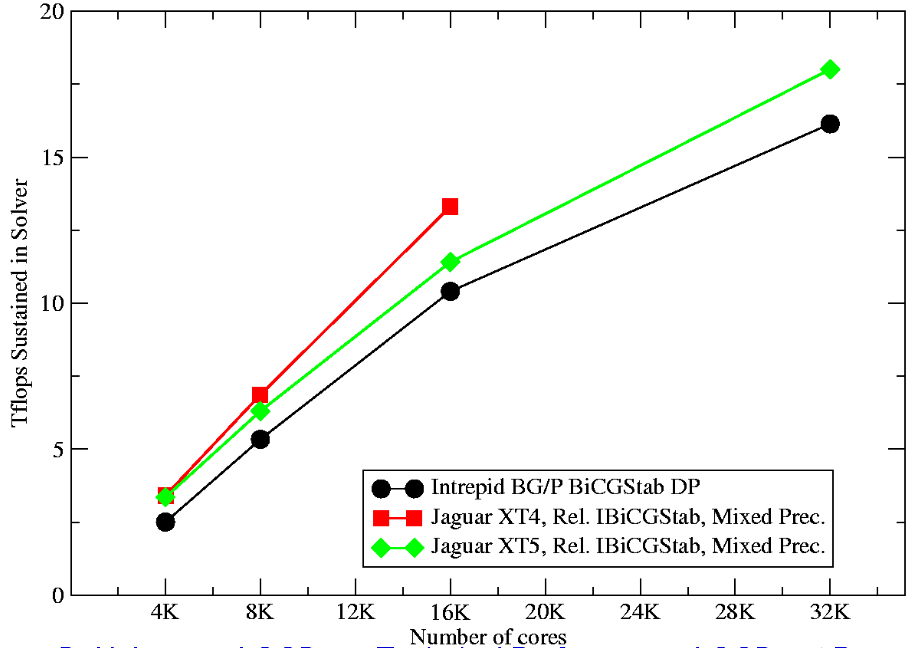
Data taken on the Edge GPU cluster at LLNL (nVidia Tesla M2050 GPUs).

TFlops are effective.

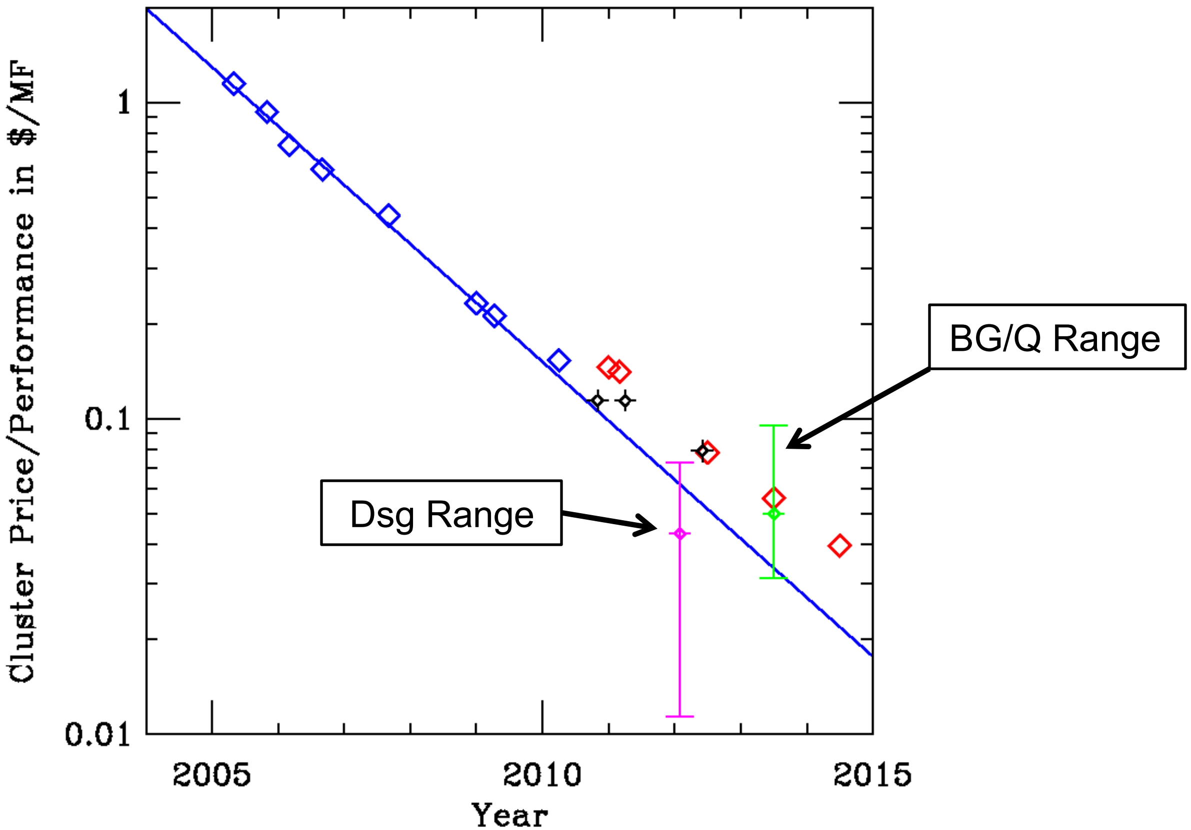
Data are for a Wilson-Clover problem of size 32^3 x 256.

The “GCR-DD” data use a communication-minimizing solver that relies on additive Schwarz pre-conditioners to perform domain decomposition.





Strong scaling results for the same problem (Wilson-Clover 32^3 x 256) on the Edge GPU cluster at LLNL (top), and on the ORNL XT4 and XT5 and ANL Intrepid BG/P (bottom).



With the very important caveat that neither machine has production software for all actions of interest (but clusters do), we can add the Dsg GPU-accelerated cluster equivalent performance and estimated BG/Q performance (values from high to low price/performance):

- Dsg: MILC HISQ / Isotropic Clover / Wilson (all three are in production)
- BG/Q: 15% Peak @ $3M/Rack (MILC HISQ) / $ 0.05/MF / DWF @ $2M/rack

Why Are GPUs Cost Effective?

- Memory bandwidth
  - A Ds node (32-core AMD) has ~ 108 GB/sec aggregate memory bandwidth, measured using cores and all 8 NUMA nodes. Cost per node, including Infiniband, case, rack, etc.: $5800
  - A single Dsg GPU card (M2050) has about the same (107 GB/sec) performance for about $1300, not including host, cast, rack, networking, etc.. But, a single host can hold 2 to 4 GPUs.
  - Both x86 processors and GPUs have excess floating point capacity relative to their ability to feed operands to their FPUs for LQCD algorithms.
- Downsides?
  - Difficult programming. To date almost all LQCD programming has been done by a handful (order 10) experts. Only parts of LQCD code (primarily inverter, now some force terms) have been ported to GPU and other parts run on host CPU.
  - Smallish memory (3, 6, now 9 GB per card) → large problems must use slow host memory and/or multiple GPUs, which introduces network and PCI bottlenecks
  - Excess flops, not enough memory or I/O bandwidth

Potential GPU Competitor – Intel “MIC”

- “Many Integrated Core” Architecture, to be called Xeon Phi
- Latest version, Knights Corner, is a GPU-like card and has:
  - More than 50 Pentium-type cores
  - At least 8 GB of GDDR5 memory (same type as GPUs use)
- K.C. will be used on the TACC’s “Stampede” supercomputer
- Said to be considerably easier to program than GPUs
  - but with fairly wide vector units (compared with SSE’s 4-wide and AXP’s 8-wide), so simply recompiling existing codes will not deliver optimal performance
- Performance not announced, but K.C.-generation likely (significantly) less than NVIDIA Kepler GPUs
- Available in late 2012 or early 2013
- We can predict, like GPUs, excess flops, insufficient memory bandwidth, communications bottlenecks

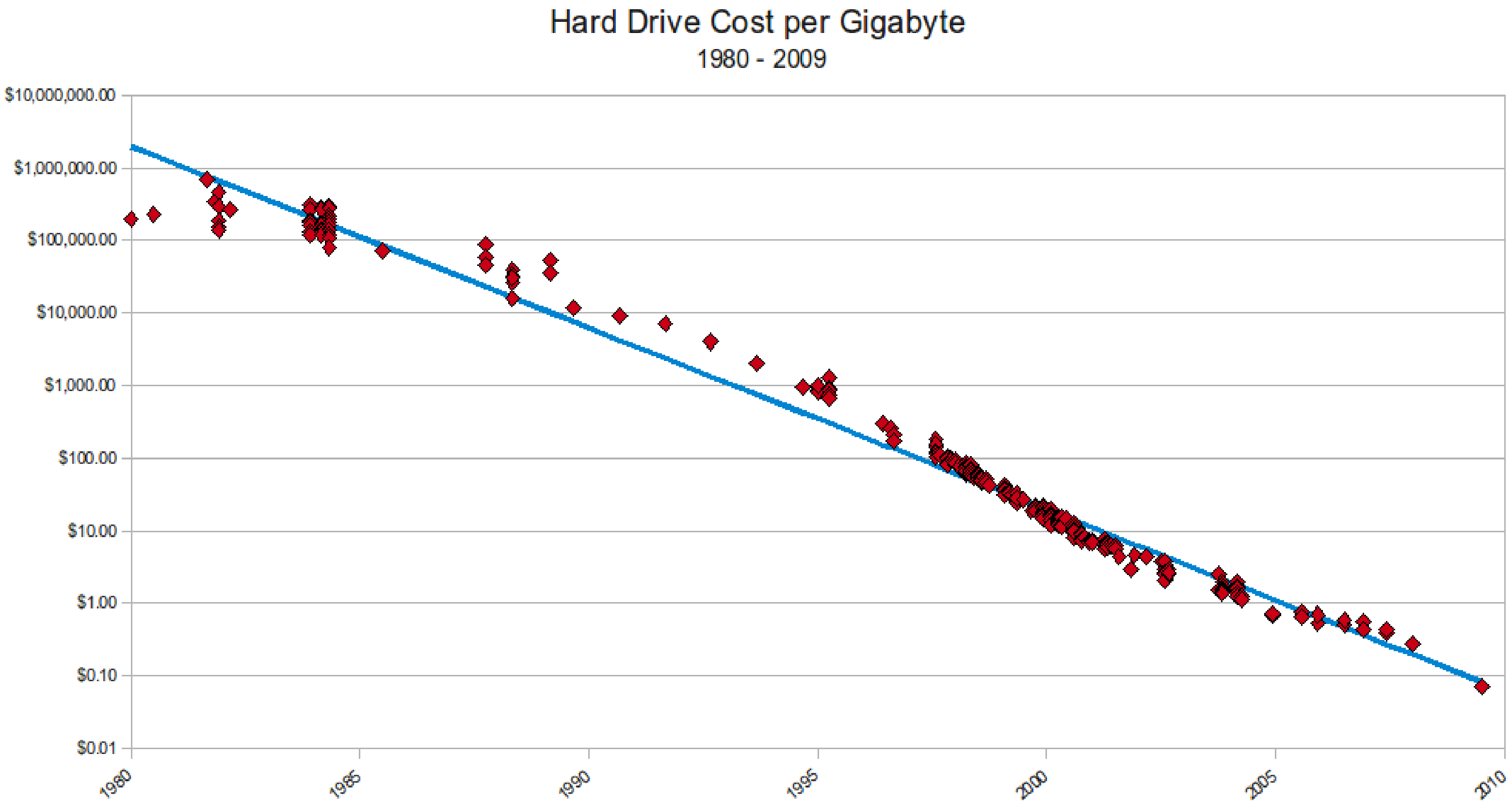
Processor Speculations

- It does not seem likely that Intel or AMD will make significant profits extending x86 architectures with higher core counts
  - Outside of LQCD and other parallel codes, utilization of high core counts for common software is difficult
  - To meet LQCD's memory bandwidth needs, "wider" memory buses would be needed to match core counts (i.e., more channels, more NUMA domains). This leads to more complex sockets – 1944 pins already in the AMD G34 socket, and 2011 pins in the Intel LGA 2011 socket for Sandy Bridge – and higher DIMM counts. AMD has postponed or canceled their new socket (G2012).
  - Perhaps "3D" (stacked) memory will come to our rescue. 3D stacking of flash memory chips is already commonplace.
- Cluster-based LQCD computing will likely fall off the commodity price/performance curve, except for GPU-like (NVIDIA, MIC) architectures
- Expect very high (GPU) or high (MIC) core counts, excess floating point capacity, small fast local memory and slow (latency, bw) global memory
  - Software approaches? Frank Winter's JIT looks promising as a way of separating off a per-hardware-type customization layer

Speculations…

- Hybrid ("fusion") processors
  - GPU, and MIC, computing is heterogeneous – programs run on both conventional host processors and on specialized GPU or MIC cores
  - GPUs and MIC processors use add-in PCIe cards
  - AMD and Intel have both experimented with "fusion" processors, which have both conventional and GPU-like cores on the same processor die
    - Currently (for LQCD), a terrible way to provide GPU FLOPs because conventional (slow) host memories are used
    - If significantly faster and larger memory were available, such as 3D-memory, such processors could be well suited to LQCD

Storage


Hard Drive Cost per Gigabyte
1980 - 2009
$10,000,000.00
$1,000,000.00
$100,000.00
$10,000.00
$1,000.00
$100.00
$10.00
$1.00
$0.10
$0.01
1980
1985
1990
1995
2000
2005
2010

Storage

- Disks, like processors, have followed an exponential trend (“Kryder’s Law”)
- Also (like processors), although data density (compute capacities) continues to increase, I/O bandwidth (memory bandwidth) has not kept pace
  - We’ve played very recently with SSD-based storage and have seen much better throughput and metadata performance
  - Even if rotating media areal density limits are not reached for a while, I/O bandwidth may force a transition to other technologies (expect cost increases). Kryder’s paper (IEEE Trans. Magnetics, **45**, 3406) is a bit discouraging
- With software like Lustre, we can provision single name spaces with petabytes of space, but:
  - File system checks (fsck) take days or weeks
  - Directory listings (“find”) of the entire namespace take days
  - Movement to/from tape and between sites is bandwidth constrained
  - Software evolution is needed

Power

- The various exponential curves have been holding for a long time – can we expect them to continue? It has been a mistake to bet against them in the past.
- In 20 minutes I haven't touched on a fundamental problem that argues against indefinite Moore's Law trends:
  - Power (see citations in http://tinyurl.com/7kklbwv)
    - ICT (information and communications technology) alone produces CO_2 emissions equivalent to the entire aviation industry. Direct emissions of internet + ICT = 2-3% of worlds emissions. ICT growth rate is the highest of any sector (4 to 6 years doubling time). A small server generates the CO_2 of an SUV getting 15 mpg.
    - The costs to power and cool a server over a 3 year lifetime now exceed the initial cost of the server
    - Energy efficiency (Flops/KWHr) is doubling more slowly (19 mos) than Top500 performance (#1: 13.6 mos, #500: 12.9 mos)

Questions?