

**FY17/FY18 Alternatives Analysis
for the
Lattice QCD Computing Project Extension II
(LQCD-ext II)**

Operated at
Brookhaven National Laboratory
Fermi National Accelerator Laboratory
Thomas Jefferson National Accelerator Facility

for the
U.S. Department of Energy
Office of Science
Offices of High Energy and Nuclear Physics

Version 1.0

Revision Date
November 14, 2017

PREPARED BY:

Bob Mawhinney, Columbia University
Co-Site Architect for LQCD at Brookhaven National Laboratory
Member of the RIKEN-BNL Research Center

CONCURRENCE:



William N. Boroski
LQCD-ext Contract Project Manager

11/14/2017
Date

Lattice QCD Computing Project Extension II (LQCD-ext II)
Change Log: Alternatives Analysis for FY17/FY18 Procurement

Revision No.	Description	Effective Date
0.1	Document created from FY16 document.	April 6, 2017
0.2	Reworked sections 1 to 3	May 5, 2017
0.3	Included comments from May 5 Acquisition Committee meeting. Updated sections 4 and 5	May 12, 2017
0.4	Minor updates throughout document to reflect combined FY17/FY18 procurement and purchase of time on Institutional Clusters. Major updates to sections 4 and 5 and the inclusion of KNL and Skylake benchmarks.	November 6, 2017
0.5	Revised sections 1 to 3 to be consistent with purchase of node-hours model. Redid sections 4 and higher to explain benchmarks, BNL cluster options and alternatives. Major reworking of document.	November 8, 2017
0.6	Small corrections from Project Manager and Acquisition Advisory Committee	November 13, 2017
0.7	Final edits to technical details of BNL clusters and phrasing in a few places	November 14, 2017
1.0	Approved version.	11/14/2017

Table of Contents

1	Introduction.....	4
2	FY17/FY18 Goals	5
3	Hardware Options.....	6
4	Benchmarks	12
5	Clusters.....	14
6	Alternatives	15
6.1	Alternative 1: A 45-30-25 partition of funds between Skylake, BNL IK and BNL IC nodes.	15
6.2	Alternative 2: A 0-30-70 partition of funds between Skylake, BNL IK and BNL IC nodes.	16
7	Discussion.....	17
8	Conclusion	17

1 Introduction

This document presents the analysis of FY17/FY18 alternatives for obtaining the computational capacity needed for the US Lattice QCD effort within High Energy Physics (HEP) and Nuclear Physics (NP) by the Lattice QCD Computing Project Extension II, abbreviated LQCD-ext II and called the Project in this document. This analysis is updated at least annually to capture decisions taken during the life of the Project, and to examine options for the next year. The technical managers of the Project are also continuously tracking market developments through interactions with computer and chip vendors, through trade journals and online resources, and through computing conferences. This tracking allows unexpected changes to be incorporated into the project execution in a timely fashion.

This analysis differs from those of previous years in that, beginning this year, the Project will be purchasing computer time in units of node-hours on various clusters at BNL and, likely, FNAL. This document is concerned solely with the purchase of node-hours on clusters at BNL. (In previous years, the Project purchased the appropriate hardware for LQCD computing needs and also paid for the operation of that hardware out of the total Project budget.) The alternatives discussed herein are constrained to approximately fit within the current budget guidance of the project for node-hour purchases at BNL in FY17/FY18:

- \$0.75M from FY17 for the purchase of computing time in FY18 at BNL, and
- ~ \$0.60M in FY18 (subject to final budgetary approval from DOE) for the purchase of computing time in FY18 at BNL.

This constraint provides funding to meet some of the requirements of the field for enhanced computational capacity, under the assumption of expanding resources at ANL and ORNL already planned by the Office of Science (SC), and under the assumption that a reasonable fraction of those resources is ultimately allocated to Lattice QCD.

All alternatives assume the continued operation of the existing resources from the FY09-FY17 LQCD Computing Project until those resources reach end-of-life, i.e., until each resource is no longer cost effective to operate, typically about 5 years.

The hardware options discussed in this document for FY17/FY18 are: a conventional CPU cluster, a GPU-accelerated cluster, a Xeon Phi Knights Landing (KNL) cluster, or some combination of these. The interconnect options are either Infiniband or Intel's Omnipath network. Conventional clusters can run codes for all actions of interest to USQCD. Optimized multi-GPU codes for solving the Dirac equation are available for HISQ, Wilson, clover, twisted mass, and domain wall fermions, using conventional Krylov space solvers. Recently, GPU-based implementations of multigrid Dirac solvers for clover fermions have been completed. For KNL with Wilson, clover and HISQ fermions, optimized inverter software is available and incorporates JLab's QPhiX code generator. Also for KNL, the Grid software package (Boyle and collaborators from the UK) has highly tuned solvers for domain wall fermions, as well as various types of Wilson and staggered fermions. Unlike GPU clusters, however, KNL clusters can run all codes for all actions of interest to USQCD, though un-optimized codes will not run nearly as efficiently as optimized codes.

As mentioned, in FY18 the project is moving to an operating model where computing time for LQCD is to be provided at BNL (and FNAL, although computing at FNAL is not discussed in this

document) from the purchase by the Project of node-hours on institutional clusters located at the labs. These lab-operated facilities provide computing hardware, selected in discussion with the Project, to support the needs of the LQCD community.

2 FY17/FY18 Goals

The project baseline called for deployment in FY17 of 45 Teraflops per second (TF) of sustained performance, based upon extrapolations of price performance of Intel x86 cores and NVIDIA Tesla GPUs and using the projected baseline budgeting for the project. In this baseline model the project purchases and operates dedicated hardware for LQCD. The project baseline assumes the use of 50% of the compute budget for conventional x86 nodes, and 50% for GPU-accelerated nodes. With the introduction of the KNL, as detailed below, the performance difference between x86 nodes and GPU-accelerated nodes has changed and, in terms of the price for a given performance, the optimal type of nodes depends on the requirements of the LQCD calculations being done. Further blurring the distinction between conventional x86 nodes and GPU-accelerated nodes, new, 32-core conventional x86 CPUs from Intel (Skylake) and AMD (Zen) have recently appeared. In the discussion in this document, the goals are to provide the target of 45 TF of computing power, scaled by changes in our overall budget, using our standard benchmarks and also to optimally meet the overall needs of the user community for the target LQCD jobs for the near future.

Regarding the 45 TF deployment goal for FY17, the project deployed 16 TF in FY17 at JLab by expanding the KNL-based 16p cluster from 192 nodes to 264 nodes. No other resources were deployed in FY17, due to a budgetary hold put on allocated FY17 hardware funds by DOE. The hold on FY17 funds has been released and the Project has been given guidance to use these funds to augment the existing Project hardware portfolio by purchasing node-hours of computing time from BNL in FY18.

The choice of a 50:50 split between x86 nodes and GPU-accelerated nodes in our baseline forecast was driven by the recognition that not all user jobs can run on GPUs, either due to not (yet) available software or the need for more memory and/or internode bandwidth than is available on GPU-accelerated nodes. Similar restrictions appear for the current analysis, making it important to understand the performance of possible hardware solutions for a variety of likely LQCD jobs of different sizes. The ability of x86 solutions to run all parts of USQCD codes gives this hardware target an advantage in user's ease-of-use.

In FY17, the project decommissioned systems purchased in 2010 and 2011. There was also some attrition of systems purchased in 2012. This reduction in capacity was partially offset by a 40-node allocation arranged by the project on the BNL Institutional Cluster (each node includes dual NVIDIA K80 GPUs) that went into production in early January 2017. The project stopped paying for a maintenance contract for the IBM BlueGene/Q half-rack at BNL in April 2017 and it ceased paying operations support at the end of FY17. At the time of this document, the BGQ half-rack is no longer available to the Project.

In our baseline model, sustained performance on conventional clusters is defined as the average of single precision DWF and improved staggered ("HISQ") actions on jobs utilizing 128 MPI ranks. In our last cluster procurement at FNAL, the 128 MPI ranks were spread out over 8 nodes, to include the effects of internode communication in the performance. "Linpack" or "peak"

performance metrics are not considered, as lattice QCD codes uniquely stress computer systems, and their performance does not uniformly track either Linpack or peak performance metrics across different architectures. GPU clusters or other accelerated architectures are evaluated in such a way as to take into account the Amdahl's Law effect of not accelerating the full application, or of accelerating the non-inverter portion of the code by a smaller factor than the inverter, to yield an "effective" sustained teraflops, or an equivalent cluster sustained performance. Effective GPU TF are based on benchmarks developed in FY 2013 to assess the performance of the NVIDIA GPUs used on the various project clusters on HISQ, clover, and DWF applications, and reflect the clock time acceleration of entire reference applications. As new codes and hardware have become available, we have adjusted our ratings to reflect a balance of LQCD calculations. For project KPI's, effective TF are equivalent to TF when combining CPU and GPU values.

3 Hardware Options

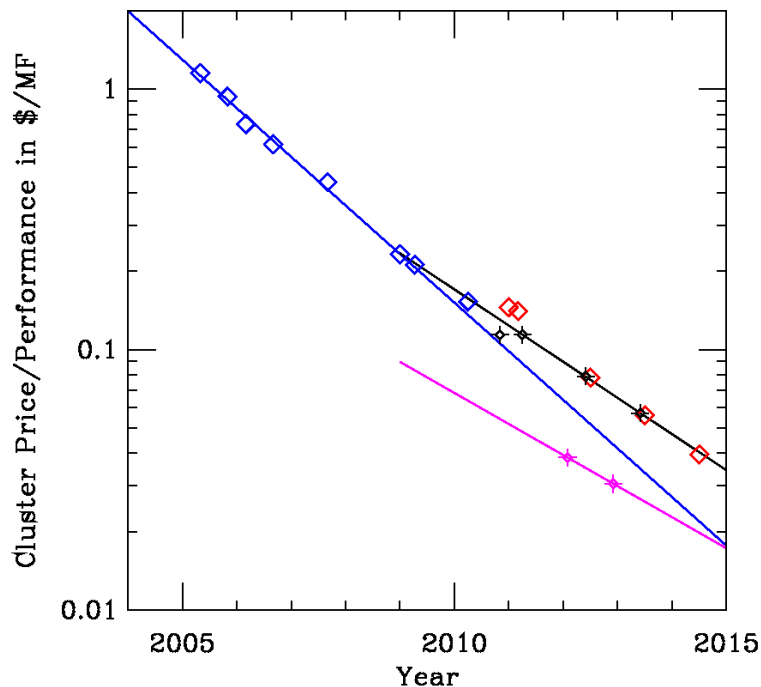
Here we discuss the strengths and weaknesses of the various types of computer hardware available at this time to guide our recommendations of the preferred types of hardware that the Project should purchase node-hours on.

The following types of hardware are considered in this analysis:

1. A conventional cluster, based on x86 (Intel or AMD) processors with an Infiniband or Intel's Omni-Path network.
2. A GPU accelerated cluster, based on Intel host processors, an Infiniband network, and NVIDIA GPU accelerators.
3. An Intel Xeon Phi Knights Landing (KNL) cluster with either an Infiniband network or Intel's Omni-Path network

Overview of Hardware Trends

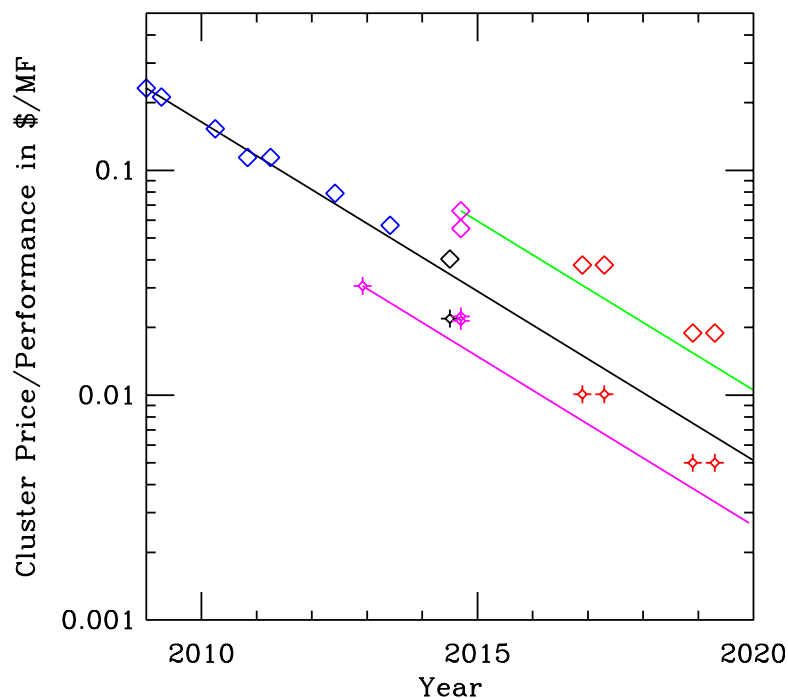
For the LQCD-ext II initial reviews, our baseline performance was developed from our experience with running both conventional clusters (since 2005) and GPU clusters (since 2009). USQCD has tracked price/performance on LQCD Infiniband-based conventional clusters deployed at Fermilab and JLab since 2005. The plot below shows these cost trends, along with exponential fits to two subsets of the data, through 2013. Also included are data and an extrapolation line for GPU-accelerated clusters.



Here, the blue line is the least-squares fit to the clusters purchased between 2005 and 2011, shown as blue diamond symbols. The red diamond symbols are baseline goals used in the LQCD-ext project plan. The black line is the fit to the points from 2009 through the FY13 cluster, Bc. The magenta line connects the points corresponding to the two GPU clusters which were not memory rich, Dsg and 12k.

What is clear from this graph is that the price performance curve has a bend in 2010 such that the performance doubling time per dollar slowed from around 18 months to around 24 months. Tesla class GPUs with ECC provide roughly 4 times as much performance per dollar, but demonstrate roughly the same 24 month doubling time.

For the LQCD-ext II project, we developed our baseline goals using a 24-month doubling time. Dropping data from before 2009, the figure below shows our experience (2014 and earlier) and our forecast (2016 and beyond). (This figure was produced in 2014.) Here the blue diamonds are the LQCD-ext and ARRA clusters. The black diamond is the original estimate for our FY14 purchase. The magenta diamonds, which are noticeably above the trend line, represent the pi0 cluster purchased at FNAL. The lower magenta diamond reflects higher than anticipated costs from manufacturers, due in part to the effective departure of AMD from the HPC cluster market. The upper of the two diamonds represents the price-performance of pi0 with larger than usual system memory (128 GBytes/node) and a 5-year warranty. The red diamonds are forecasts for “future” clusters (from a 2014 perspective) with purchases split over fiscal year boundaries. The graph also shows points with magenta stars, representing the two GPU clusters, ARRA 12k and pi0-g, along with a GPU trend line.



It is important to note that the larger memory for the pi0 cluster that was deployed in 2014 was needed for the calculations being started at that time. The trend to larger memory footprint for LQCD jobs has become the norm in much of the USQCD community. The larger memory is used to store eigenvectors, or other reusable intermediate solutions, for the operators of interest and these then markedly speed up the calculation of quark propagators and other observables. This change has resulted in the need for larger memory on the computer partition being used, as well as for increased I/O bandwidth to disk and an accompanying increase in disk storage size. In the last year, a number of groups have made substantial progress in reducing the size of these reusable intermediate solutions, somewhat decreasing the rate of growth of memory and storage requirements. The LQCD community is also generating larger lattices on the Leadership Class Facilities (LCF), and these lattices require larger memory when observables are measured on Project provided computing resources.

Overview of Allocation Requests and the LQCD Hardware portfolio

For the allocation year, 7/1/2017 to 6/30/2018, USQCD users submitted proposals to use essentially all of the available time on LQCD GPU clusters. For the conventional clusters, proposals exceed the available time by a factor of 2.49 and for the KNL cluster, the oversubscription is a factor of 2.19. Since the KNL is an x86 based machine, codes which run on conventional clusters will run without modification on KNL nodes, although achieving high performance on a KNL node generally requires more carefully crafted code than on a conventional cluster node.

At the end of FY17, the LQCD-ext II project retired the 512 node BGQ machine at BNL. This machine has run only 512 node jobs for the last year, which means a physical memory size of 8 TBytes was available to the user. A number of users were running jobs for which this memory size was barely adequate. In addition to user demand for cluster nodes, there is also substantial demand for reasonable size memory in the partitions available to users. About 30% of the cluster time at FNAL from 7/1/16 to 5/1/17 has been used for jobs with memory sizes exceeding 4 TBytes. This, coupled with the large memory size in all the jobs run on the BGQ, indicates that this acquisition should target machines capable of running QCD codes well on partitions with 5 TBytes of memory, or more. This leads us to target a machine that will run user jobs efficiently in the 16 to 32 node range, with between 128 and 256 GBytes of memory per node.

Conventional Clusters

The continued demand by USQCD users for conventional cluster time shows the usefulness of this hardware platform for LQCD calculations. Pi0 is entering its fourth year of operation and a successor platform is needed for this workload. (As we will discuss further in the KNL section below, KNL nodes are able to take on this workload without requiring code rewrites.) The presence of both GPU accelerators and the KNL is having some impact on conventional cluster nodes. Both Intel and AMD have introduced new, conventional CPU chips with up to 32 cores. Intel's new Skylake chip is available now and we provide benchmarks below. In addition to the large core count, Skylake also implements the AVX-512 instruction set that was previously only available on the KNL. Skylake does not have the 16 GBytes of on-chip MCDRAM that is available on the KNL. On the KNL, the MCDRAM provides a substantial amount of memory with very high bandwidth. The AMD Ryzen chip is a re-entry of AMD into the x86 desktop and server market and offers LQCD users the advantages of a second manufacturer in this market.

In comparing Skylake with KNL, Skylake has: 1) full-featured Xeon cores which give better performance in our limited testing for compiled codes not employing carefully tuned code, 2) less memory bandwidth which may impact performance and 3) a larger price per node, possibly by up to a factor of 2. This basic cost of purchase is a "raw" cost and does not include operational costs. The imminent retirement of the Fermilab Bc cluster and the approaching retirement of the Fermilab Pi0 cluster will soon leave no conventional clusters available to USQCD users. To meet the demand of USQCD users, whose codes do not or cannot run efficiently on a GPU or KNL cluster, a Skylake cluster with a single-rail Infiniband network is an attractive choice, despite its higher cost per node-hour. Our expectations for Ryzen are similar at this point.

An alternative cluster scenario would be a Broadwell based machine. Broadwell only has AVX-2, and not AVX-512, so it would not perform at the level expected by Skylake (or KNL). However, since it is nearing the end of its lifetime, it would become an interesting option if the price were very low. (The nodes of the existing BNL Institutional Cluster contain a Broadwell processor and 2 K80 GPUs.)

For a conventional x86 cluster, single-rail EDR Infiniband (100 GBytes/s) is a well-understood option, with Omni-path as an interesting option for the Skylake chip. For Broadwell nodes, 128 GBytes/node would be the minimum memory and systems with larger memory (256 GBytes)

possible. For SkyLake, with 6 buses per chip (like KNL), memory sizes would be 192 GB or 384 GB.

GPU Accelerated Clusters

For those calculations for which optimized software is available, GPU-accelerated clusters offer a substantial improvement in price/performance compared with conventional clusters. The LQCD hardware portfolio includes the 12k cluster at JLAB, the pi0-g cluster at FNAL and 40 dedicated nodes of the BNL Institutional Cluster (BNL IC), which has dual K80 GPUs on each node (4 logical GPUs). For USQCD users, and software developers, GPU nodes will continue to be a major focus, since not only are there substantial GPU resources in LQCD hardware, but the LCFs are deploying large GPU based machines.

The newest GPU from NVIDIA, the P-100, offers not only an improvement on the traditional large core count performance associated with GPUs, but also the NVLINK technology for connecting GPUs through peer-to-peer technology, which allows separate GPUs to access each other's memory, without going through the host (for motherboards supporting this). (At the time of this writing, BNL is adding more nodes to its existing Institutional Cluster and these nodes have P-100 GPUs.)

We have benchmarked optimized GPU codes on K80 nodes (at the BNL IC) and on P-100 nodes, available as test platforms. We have run extensively optimized codes, written by Kate Clark of NVIDIA (Kate did her PhD research in lattice QCD), on these platforms. We see good performance for problem sizes that are small enough to fit on a single node. For example, on a single node of the BNL IC a single precision domain wall fermion solver on a $24^3 \times 96$ local volume gives about 1.3 TFlops/s of sustained performance (a single node has 2 K80s).

The difficulty with the powerful GPU nodes come when one wants to run on a large enough number of GPUs that off-node communications is required. Here our tests show that with 100 GBit/s EDR IB, the performance on a 16-node system for the same DWF running as in the previous paragraph drops dramatically to about 0.7 TF per node. Given our target of running large jobs on 16 to 32 nodes, the scaling of the GPU clusters is not adequate for our purposes.

It is also important to note that the previous discussion focused on highly tuned code for the conjugate gradient solvers. For realistic workloads, where part of the code is not highly tuned and the GPUs play little role, the performance is a comparison between the speed of the CPU (a 36 core Broadwell processor on a BNL IC node) and the x86 processor of a non-GPU cluster node. These speeds are likely not markedly different. However, for each node of the BNL IC (raw cost about \$20k/node), we would have 4 nodes of a KNL system (raw \$5k/node), for example, so the parts of the code that do not use the GPUs will run up to 4 times faster on x86 cluster than on the BNL IC.

The broader LQCD hardware portfolio includes GPUs and these will continue to be an important part of our hardware strategy. NVIDIA just announced the new Volta GPU, which is a major step in their product line (they claim 1.5x in performance over the P-100) and which will be part of the DOE's Summit computer.

Xeon Phi / Knights Landing Cluster

We have had substantial experience with KNL clusters in the last 12 months, both through the 264 node KNL cluster at JLAB and also through a 144-node cluster at BNL, used for a number of BNL computing projects, including lattice QCD. (We will refer to the BNL cluster used for its institutional projects as the BNL IK computer.) The JLAB KNL cluster is connected by single-rail Omni-path (100 GBits/s) with an oversubscribed network (32 nodes per switch, 16 uplinks, so jobs of 32 nodes or less could be done with no oversubscription). The BNL-IK cluster has dual-rail Omni-path, a full fat-tree network and 1 TBytes of SSD per node. Both clusters have 192 GBytes of memory per node. As mentioned earlier, the KNL has support for AVX-512 and also has 16 GBytes of high-bandwidth on-chip MCDRAM.

We have benchmarked a number of USQCD application codes on the KNLs at JLAB, BNL and also at TACC and Theta at ANL. (Tables of benchmarks are in Section 4.) Heavily optimized single node codes give sustained performance in the 700-900 GFlops range for staggered, clover and domain wall fermions. The Grid code of Peter Boyle, running for domain wall fermions with a local volume of 24^4 in single precision, gives 300 GFlops/s for a conjugate gradient solver on single and dual rail systems of 16 nodes. (The first table in Section 4 shows that the Dirac operator application runs at 400 GFlops/s for similar conditions.) On the BNL IK, the Grid code gives 250 Gflops/s when running on 128 node systems. For the MILC code, with QPhiX optimizations and a 32^4 local volume running on 8 nodes in double precision, performance is around 75 Gflops for a double precision multi-shift solver. For MILC C code, i.e. without any optimizations beyond what the compiler will do, and using OMP, the multi-node performance is about 56 GFlops/s.

One conclusion that our benchmarks have led us to is that for current lattice sizes and calculations that run on up to 32 nodes, there is no need for dual-rail networks. If one went further in the strong scaling limit and tried to run on 128 nodes, dual-rail would be necessary. Jobs of this size would likely not be allocated enough time by the USQCD Scientific Programming Committee to make any real progress on such a large calculation. A single rail system is 10-15% cheaper than a dual rail system.

An important issue with the KNL cluster option is system reliability and usability for 16 to 32 node partitions. During the last 6 months, both the JLAB KNL and the BNL IK have been undergoing upgrades to their BIOS/OS to improve stability and to weed out weak hardware nodes. During its first 6 months of operation, the JLAB system was used almost exclusively for single node jobs, with high reliability, and, beginning in May 2017, a single user had begun running steadily on a 64-node partition. On the BNL-IK cluster, users are currently running well on 32 nodes and larger partitions. Both KNL clusters show decreasing node performance over time as more jobs are run, likely due to fragmentation of memory. This has been improved by software and firmware upgrades and, while operational reliability of the KNL clusters for multi-node jobs has clearly improved, these systems are requiring more human resources for operation than a similar sized conventional cluster.

As seen in the purchase of the JLAB KNL in FY16/FY17, such a system will meet our target performance goals. Since it is an x86 architecture, all of USQCD code will run on this platform, with optimized code getting very good performance. 192 GBytes of memory per node has proven

adequate and the BNL IK has an additional 0.8 TBytes of SSD storage on each node for temporary local storage.

4 Benchmarks

BNL currently has two clusters available for the Project to purchase time on: the BNL IC and the BNL IK. A third possible cluster, composed of Skylake nodes, would represent a conventional cluster option for the Project. The BNL IC nodes have 2 K80 GPUs and are an efficient resource for single, or few node, jobs. Larger node count jobs that we are targeting here would be more efficiently run on a KNL or Skylake cluster. In this section, we give benchmarks for single node jobs on the BNL IC and large node-count jobs on a KNL or Skylake cluster.

The current BNL IC has 2 K80 GPUs on each node and 40 nodes of this cluster are already available to the Project. USQCD has established a benchmark for their allocations of time on this cluster, using a performance of 1 K80 hour = 2.2 K40 hour = $2.2 * 224$ Jpsi equivalent core-hour = $2.2 * 2.24 * 1.22$ Gflop/s-hour = 601 GFlop/s – hour. Thus with 2 K80s per node, a BNL IC node is rated at 1.2 TFlops/s for LQCD. This good performance makes this a desirable platform for single node jobs.

Results for KNL and Skylake clusters are shown in the two tables in this section. The results for domain wall fermion (DWF) codes are given for up to 8 nodes of Skylake, as these benchmarks were run before larger partitions were available. For MILC code, results are given for up to 64 Skylake nodes. Note that single node KNL and Skylake codes will give performance of about 700 Gflops (pure solver, rating is lower), but here we are primarily interested in the performance for 8 to 32 node jobs. (Boyle’s Grid code gets 692 as a best case, from the first table, and Robert Edwards Chroma code runs at about 700 GFlops for contractions.)

The first table presents results for highly optimized domain wall fermion (DWF) codes, part of Peter Boyle’s Grid package, running on both KNL and Skylake test machines. The Skylake test nodes are connected by Mellanox 4X EDR with 100 Gigabits per second peak speed, or 25 GBytes/s bidirectional bandwidth. The KNL results are from the BNL KNL system, with dual rail Intel OPA interconnect with a peak bidirectional interconnect bandwidth of 50 GBytes/s. The tests reported here were run primarily by Meifeng Lin of BNL.


The results of primary interest here are for 8 node jobs, comparing Skylake and KNL. The local volume most relevant for a calculation of this size is $L=24$. Here the 8 node Skylake performance, with overlapping computation and communication, achieves a performance of 384 GFlops/node on the 6142 Skylake and 400 GFlops/node on the KNL. The 6142-based node has 16 cores per socket and 2 sockets, for a total of 32 cores, compared to the 64 cores on the KNL, but runs at a higher clock speed. One clearly sees that the total performance per node is essentially identical for these two choices. The lack of the high-bandwidth MCDRAM on the Skylake is not impacting the performance. The 32 cores of a Skylake node can compute as rapidly as the 64 cores of a KNL node due to their higher clock speed.

Skylake/KNL DWF Benchmark

Model	# nodes	ppn	Threads/rank	L=16 (GF/node)	L=24 (GF/node)	L=32 (GF/node)
6134	1	2	8	403	365	386
6136	1	2	12	592	542	503
6142	1	2	16	653	516	555
6142	2	2	16	486	441	493
6142	4	2	16	376	388	427
6142	4	2	16	474	422	448
6142	8	2	16	317	340	387
6142	8	2	16	421	384	417
6148	1	2	20	818	565	524
6150	1	2	18	757	570	485
6150	2	2	18	554	435	452
6150	4	2	18	408	374	399
6150	8	2	18	332	337	
8168	1	2	24	935	655	617
8168	2	2	24	637	553	463
KNL	1	1	60	692	687	661
	2	4	15	366	554	545
	2	8	7	375	508	565
	4	4	15	284	435	445
	8	8	7	279	400	450
	16	8	7	258	382	428

Comms-compute overlap

ARS OF COVORY


U.S. DEPARTMENT OF ENERGY

BROOKHAVEN
 NATIONAL LABORATORY

A CENTURY OF SERVICE

A similar comparison between Skylake and KNL has also been done for MILC code. Here, two versions of MILC code are involved: one is highly optimized code using the QPhiX library and the second is compiled MILC code where the only optimization is the use of OMP. This second case is included to reflect performance on parts of the LQCD code base that are not well optimized.

The following table shows the results of these benchmarks. Focusing on the 16 nodes, L=32 column (this local lattice volume is the target size for computations running on a hardware partition of this size), one sees that the 6148 with or without QPhiX gives essentially the same performance (55.0 compared to 53.0 GFlops/s). On a KNL with QPhiX, the performance is somewhat better, at 68.0 GFlops/s and without QPhiX, but with OMP, the KNL performance is 56.0 GFlops/s

Skylake/KNL MILC Benchmark 2

CPU	Socket/ Cores/ Ranks/HT pernode	code type	1node L=24	1node L=32	8nodes L=32	16 nodes L=32	32 nodes L=32	64 nodes L=32
6142	2/32//32	No Qphix	57.0	56.5	52.2			
6148	2/40//40	No Qphix	59.1	58.1				
6148	2/32/8/32	with QPhiX		59.0	58.0	55.0	52.0	52.0
6148	2/32/16/32	No QPhiX with OMP		60.0	57.0	53.0	55.0	50.0
6150	2/36//36	No Qphix	55.5	51.8	49.1			
KNL	1/64/1/64	with QPhiX		115.0	75.0	68.0	61.0	64.0
KNL	1/64//256	with Qphix	25.8	38.5		24.4		
KNL	1/64/16/64	No QPhiX with OMP		68.0	56.0	56.0	53.0	50.0
KNL	1/64/64/64	No Qphix/No OMP	65.0	56.0	50.0	45.0	44.0	42.0

Based on MILC multishift CG code from Steve Gottlieb
 White: benchmark data from Zihua Dong, Meifeng Lin
 Gold: benchmark data from Karthik Raman w/ Intel OPA
 Green: benchmark data from Gottlieb w/ Cray Aries



5 Clusters

BNL currently has two clusters available for the Project to purchase time on: the BNL IC and the BNL IK. A third cluster, made of Skylake nodes, could be made available to the Project, if the Project's analysis shows that cluster to be useful for LQCD. Across these cluster combinations, the LQCD project has access to 200 TBytes of a 1 PByte GPFS storage with a peak bandwidth of 24 GBytes/s. In this section, we describe the cluster options available at BNL.

BNL IC: The BNL IC entered production in January, 2017 with 108 nodes. Each node contains two sockets of Broadwell CPUs, with a total of 36 cores, 2 K80 GPUs, 256 GB of DRAM, a 1.8TByte SAS local disk drive, a non-blocking EDR interconnect fabric and access to the high-performance, GPFS-based storage with up to 24 GBytes/s of bandwidth via EDR. BNL is expanding the BNL IC, with 18 additional nodes in production from September 2017, with these nodes having P-100 GPUs replacing the K80s. Conditional on funding, BNL is planning to add

90 more nodes with P-100s, for a total machine size of 216 nodes. A node-hour on the BNL IC is charged at \$0.99.

BNL IK: The BNL KNL cluster has 144 KNL nodes with 192 GBytes of memory per node, a dual rail Omnipath interconnect, 0.8 TBytes of temporary SSD storage per node 1 GByte/s of I/O throughput and access to the high-performance GPFS-based storage with up to 8 GBytes/s of bandwidth via LNET routers from Omnipath to Infiniband. 72 KNL nodes are allocated to other stakeholders, so the Project could purchase time on up to 72 nodes of the BNL IK. There is an option to expand the BNL IK by about 32 nodes; however, the procurement time for adding nodes to the KNL cluster is estimated to be 4-6 months. A node-hour on the BNL IK is charged at \$0.51.

BNL SC: BNL has recently installed a Skylake cluster for High Throughput Computing (HTC) applications to support experiments such as RHIC, ATLAS, DUNE, and LSST. This HTC system does not have a fast interconnect. To support Project calculations, BNL could provide a HPC Skylake cluster to the Project. This would consist of nodes containing two Xeon Gold 6150 (18 cores each) CPUs and 192 GBytes of DRAM. The interconnect would be single rail EDR Infiniband, with up to 24 GBytes/s of bandwidth to the GPFS storage system. Unless the Project requests to use node-hours on this HPC cluster, there is no set time-frame for its deployment at BNL. However, if the Project requests node-hours from this machine, BNL estimates that it will deploy 64 nodes of a HPC BNL SC to production by February 1, 2018 since this machine could be acquired as an option on the HTC BNL SC. A node-hour on the BNL SC will be charged at \$0.91.

6 Alternatives

The Project has \$750k of FY17 funds available for immediate use to purchase node-hours of computing at BNL. There may be additional funds for the project in FY18, but these are uncertain at this time. There are three possible clusters on which to purchase node hours and some constraints on the balance of node-hours between the options. In particular, if the Project does not request a substantial fraction of the time on the HPC BNL SC, such a machine may not be available in FY18. Similarly, adding nodes to the BNL IK could come at the expense of the nodes in the HPC BNL SC. If the node count for the HPC BNL SC is pushed too low, it is not worthwhile to support this additional cluster.

Given the benchmarks and machine options, we can now present the alternatives in order of decreasing preference.

6.1 Alternative 1: A 45-30-25 partition of funds between Skylake, BNL IK and BNL IC nodes.

- a) Purchase all of the time on 64 nodes of a Skylake cluster for 8 months in FY18, starting February 1, 2018*

b) Purchase all of the time on 64 nodes of the BNL IK cluster for 9 months in FY18, starting January 1, 2018

c) Purchase all of the time on 30 nodes of the BNL IC cluster for 9 months in FY18, starting January 1, 2018

Analysis: This is our first choice, as it gives the Project access to the maximum diversity of platforms, while allowing calculations to run on hardware that is best suited to the calculation and the performance of the software. The BNL IC and BNL IK give very similar cost performance for single node jobs. For single node jobs, the BNL IC has a performance per node of 1.2 TFlops/s, compared to 0.7 GFlops for a KNL node. This ratio, 1.7, is very close to the ratio of costs, $0.99/0.51 = 1.94$. This gives the Project two very cost competitive options for single node jobs.

For multi-node jobs, the tables show that the performance per node for the KNL and Skylake clusters are very similar at the 15% level. The Skylake nodes cost more per node-hour, by a factor of $0.91/0.51 = 1.8$. The Skylake nodes are a valuable addition to the Project's hardware portfolio, since they have full-capability Xeon cores which offer better out-of-order execution than on KNL, have the most bandwidth to the GPFS disk system and should give the best performance for compute intensive parts of calculations (such as contractions) which are outside of the inverters and rely on not easily optimized code. In addition, we note again that the conventional cluster at FNAL, pi0, is oversubscribed by a factor of 2.49.

There are some codes that get very good performance on a KNL cluster and these can take advantage of the 64 nodes on the BNL IK (and also the Project's KNL cluster at JLAB). Experience to date has shown that achieving good performance on the KNL is challenging and running the same codes on different lattice sizes can lead to very poor performance, due to network bandwidth limitations. These have not been seen in Skylake tests, which gives a high-performance multi-node capability for codes that do not run well on the KNL.

The project would initially request this balance of funding for the various clusters and there remains the possible of moderate levels of adjustment throughout the fiscal year, depending on the needs of other groups using these resources.

6.2 Alternative 2: A 0-30-70 partition of funds between Skylake, BNL IK and BNL IC nodes.

a) Purchase no time on a Skylake cluster in FY18

b) Purchase all of the time on 64 nodes of the BNL IK cluster for 9 months in FY18, starting January 1, 2018

c) Purchase all of the time on 82 nodes of the BNL IC cluster for 9 months in FY18, starting January 1, 2018

Analysis: This second alternative would entail the Project not purchasing any time on a Skylake cluster and BNL would be unlikely to deploy such a cluster this fiscal year. Time would come solely from the BNL IC and BNL IK. The BNL IK is a cost-effective option for single node jobs, but this would leave USQCD without a viable conventional cluster to take over the load from pi0, as it ages. The BNL IC can be run as a conventional cluster, using just the Broadwell CPUs, but this is wasting resources with the GPUs being idle. There is a possibility that BNL would enlarge the BNL IK by around 30 nodes, which the Project could request, and this would give more multi-

node job capability. This capability would only be on KNL nodes, which suffer from not being as general purpose as the Skylake nodes in Alternative 1. In addition, the BNL IK has lower bandwidth to disk than a Skylake cluster, due to the bridging required to go between Omnipath and Infiniband.

The project also has access to the 264 node KNL cluster at JLAB, so there is substantial KNL capacity for the jobs which run well on that platform.

7 Discussion

The goal of this alternatives analysis is to determine the scenario which gives the Project the maximum available computing power for the entire range of calculations in the USQCD portfolio. The analysis this year has changed noticeably from previous years, in that we will be buying time on institutional clusters at BNL, rather than buying a single machine which we will run for 5 years.

The first alternative is very compelling and far superior to the second option. By committing to running on 64 nodes of Skylake at BNL, the lab will provide this cluster and also allow us to spend the rest of our \$750k to purchase time on the BNL IK and BNL IC. These two platforms have demonstrated their cost-effectiveness for USQCD calculations. The addition of a Skylake cluster gives us a natural successor to the conventional clusters at FNAL and a computer which can provide good performance for codes that do not perform well on KNL or GPU based nodes.

It is also important that the HPC BNL Skylake Cluster can be brought on line rapidly.

8 Conclusion

The preferred path forward for the LQCD-ext II Project is Alternative 1, in which there is the rapid deployment of a Skylake cluster for HPC at BNL. The Project will purchase all of the node-hours on the 64 nodes of this machine from February 1, 2018 through the end of FY18. The project will also purchase all of the time on 64 nodes of the BNL KNL from January 1, 2018 to the end of FY18 and all of the time on 30 nodes of the BNL IC from January 1, 2018 to the end of FY18.