Lattice QCD Extension II Computing Project
(LQCD-ext II)

Response to Recommendations from the 2016 DOE Annual Progress Review of the LQCD-ext II Computing Project

Compiled by

Paul Mackenzie
Chair, USQCD Executive Committee

Robert D. Kennedy
Associate Contractor Project Manager, LQCD-ext II Computing Project

October 24, 2016

This page intentionally left blank

LQCD-ext II 2016 Annual Progress Review
Response to Review Recommendations

INTRODUCTION

On June 28-29, 2016, the U.S. Department of Energy (DOE) Office of High Energy Physics and the Office of Nuclear Physics conducted an Annual Progress Review of the LQCD-ext II (LQCD Extension II) project. The review was held at the Thomas Jefferson National Accelerator Facility and resulted in a written report that contained no formal recommendations. However, the report did contain seven suggestions to help improve project effectiveness and impact. This document summarizes the project response to these suggestions, along with subsequent actions taken.

RESPONSE TO SUGGESTIONS

Suggestion #1: Only 50% of the active users responded to the survey. The reviewers would like to see an effort by the project team to increase this percentage.

Report Section: LQCD-ext II Review – Progress towards Scientific and Technical Milestones

Response: The project has improved the survey response rate for PI's by communicating through the SPC and the Allocation Process the expectation that an award of LQCD resources entails a civic duty to complete a user survey. We have not been as successful communicating this to active users. The project will work with the EC and SPC to communicate to PI's that this civic duty applies to all LQCD users, not just the PI's. We will reinforce this in the user survey email announcement and the presentation of survey results at All Hands meetings. We will consider means to hold active users accountable as we have been able to with PI's through the Allocation Process.

Suggestion #2: **Physics deliverables are the ultimate objective of the project. The definition and documentation of science milestones seem to be quite good in some LQCD areas, and lacking in others. The project should develop procedures to document scientific milestones uniformly over all the LQCD areas so that the project can track their annual progress more quantitatively.**

Report Section: **LQCD-ext II Review – Progress towards Scientific and Technical Milestones**

Suggestion #2: The idea of more specific physics milestones has been raised in past reviews. There has definitely been progress in this aspect of documentation, but the improvement has not been uniform across the full spectrum of LQCD physics topics. USQCD management should continue their efforts to persuade their users across all LQCD areas to pay more attention to physics milestones. (See also comments under charge bullet #2.)

Report Section: **LQCD-ext II Review – Effectiveness of USQCD, Scientific Impact, Procedures and Related Activities**

Response: We agree with this suggestion and will continue our efforts to better document our physics goals and milestones. We are glad that the review panel sees progress and agree that more is desirable. As the report notes, progress has been easier in some parts of our program than others. Prediction of expected future uncertainties is easiest in those areas in which complete, well understood uncertainty budgets have existed the longest, such as in hadronic weak matrix elements. Predictions of future uncertainties are less reliable in areas in which uncertainties are just coming under control, such as in much of nuclear physics. We plan to try to make progress quantifying our goals throughout our program.

Suggestion #3: **The reviewers would like the project team to seek broader engagement with the users on future architecture needs. One possibility would be to add specific questions in the user survey to seek input on, for example, currently available clusters and hardware preferences for future procurements.**

Report Section: **LQCD-ext II Review – Technical Design and Scope for FY2016**

Suggestion #3: **The project should consider adding an open-ended question concerning desired architecture or resources (memory, queue management, ticketing, etc) to the annual user survey. The idea is to allow for more bottom-up information flow within the collaboration on the hardware acquisition process.**

Report Section: **LQCD-ext II Review – Effectiveness of USQCD, Scientific Impact, Procedures and Related Activities**

Response: We appreciate the purpose of this suggestion is to better engage users on future architecture needs. We believe the user survey is not the ideal venue to do so, since these questions would defeat the goal of maintaining a simple user survey that requires little time to complete. We propose instead that the project work with the SPC to broaden the existing questions on future architectures and resource needs that are in the annual Call For Proposals, where PI's are asked to document their computing and storage resource requirements. Also, the new Architectural Review committee allows some users to participate more directly in the computing acquisition process for broader user input as well as communicating more broadly to users how the acquisition process works at the host laboratories.

Suggestion #4: **The architectures currently considered by the project team are very similar to those available at the ASCR NERSC Center and the two Leadership Computing Facilities (LCFs). Since those three centers have strong track records in hardware procurement, allocation processes, and user support, the project team should develop closer communications and interactions with them.**

Report Section: LQCD-ext II Review – Technical Design and Scope for FY2016

Response: Indeed, the architectures are very similar. The project will address this topic more directly in reviews in the future.

There is communication now between the USQCD community and the NERSC + LCF community on new architectures as both communities are engaged with each other and with computing vendors directly in venues such as:

- HEPiX
- SuperComputing
- Intel HPC Events (invitation only, includes discussions with architects)
- NVIDIA GTC (GPU conference, open)

USQCD members from all major collaborations/projects have worked alongside NERSC + LCF representatives in vendor-sponsored events, such as the Code Readiness Exercise (a.k.a. Dungeon Session) held at NERSC, to adapt software codes to exploit new architectures before they are deployed. USQCD has adopted elements of the NERSC allocation policy, for example to impose a penalty on projects which do not use their quarterly portion of their allocation.

The vendor relationships of the LQCD Project's Site Architects are already exploited to obtain the best technical information, optimal pricing, and vendor support. This is effective in part because the vendors know that the Site Architects are engaged with both the USQCD community and NERSC + LCF community. The project can reach out to the NERSC + LCF community to assess potential user support commonalities.

Suggestion #5: **There appears to be a considerable burden on the users to port their codes to new architectures delivered by the project. While the LQCD-ext II project is entirely a hardware project and doesn't provide support for software development, it should consider, as part of user support, a role in connecting the users with ASCR researchers in order to advance algorithmic development and code tuning/optimization. The SciDAC program provides a mechanism to achieve this. In fact, the project's documents and presentations referenced the SciDAC Program several times. The SciDAC Program funds computational mathematicians and computer scientists through the SciDAC Institutes. Perhaps the LQCD-ext II team should invite members from the SciDAC Institutes to participate in the All-Hands Meetings in order to facilitate interactions and collaborations between LQCD and math/computer science communities.**

Report Section: **LQCD-ext II Review – Technical Design and Scope for FY2016**

Response: We acknowledge that there is a considerable burden on users. To address this, USQCD has a long history of working with vendors to allow aggressive adoption of more performant architectures. The LQCD project will address this topic more directly in reviews in the future.

USQCD members from all major collaborations/projects have worked alongside NERSC + LCF representatives in vendor-sponsored events, such as the Code Readiness Exercise (a.k.a. Dungeon Session) held at NERSC, to adapt software codes to exploit new architectures before they are deployed. In fact, some past and current USQCD members work directly with or for vendors in the design of new hardware architectures to improve features critical to Lattice QCD performance. Examples include:

- Balint Joo, TJNAF, consults with Intel on Intel Phi architecture
- Kate Clark, former USQCD researcher, works at NVidia on NVidia architecture
- Columbia University involvement with IBM on the BlueGene architecture

USQCD has a standing Software Committee whose web page describes the SciDAC-2 QCD API, software documentation, and committee membership: http://usqcd.fnal.gov/software.html. Note the presence of LQCD-ext II Site Architects on the USQCD Software Committee to ensure communication and consistency between the software (USQCD) and hardware (LQCD-ext II) planning.

A number of USQCD members are heavily involved in SciDAC-funded projects, such as Balint Joo at TJNAF, Chulwoo Jung at BNL, James Osborne at ANL, and Alexei Strelchenko at FNAL. While members of the SciDAC Institutes can certainly be invited to USQCD All Hands meetings, there already exist avenues for communication and sharing of experience between the USQCD community and SciDAC Institutes.

Suggestion #6: The budget allocation for hardware in the last year of the project (FY19) is significantly larger than other years. There doesn't seem to be a good rationale for this decision. The reviewers questioned if it is a wise decision to have a large procurement during the final year of the project.

Report Section: LQCD-ext II Review – Feasibility and Completeness of Budget and Schedule

Response: (Bill Boroski)

The decision to allocate a significant fraction of the project budget to hardware in the last year of the project is driven solely by the funding profile of the project. The funding profile determines the level and timing at which funds are made available to the project.

The annual funding profile for the project was established by the DOE when the project was awarded CD-2/3 approval on October 1, 2014. The funding profile, shown in Table 1, includes the combined annual HEP and NP planning budgets that allow for the acquisition, deployment, and operation of computing facilities necessary to achieve approved project goals. Ideally, annual budgets would have been larger in the earlier years of the project, which would have allowed the project to front-load the hardware budget so that the large hardware purchases could be made earlier in the project lifecycle. Unfortunately, due to budget constraints, a front-loaded budget profile was not possible, so the project is allocating as much budget to hardware as possible, as early as possible, within the limits of the annual budget plan allocations.

Table 1. Approved Annual Funding Profile for the LQCD-ext II Project (in millions)

FY15	FY16	FY17	FY18	FY19	Total
2.00	3.00	3.00	3.00	3.00	14.00

For any given year, the allocation of funds available for new hardware is determined by first taking into account the level of funding necessary to support personnel costs and then allocating remaining funds to that year's hardware budget. The level of funding necessary for personnel support in any given year is derived from a detailed staffing model that takes into account the volume of hardware in production operation during that year. The model has been refined over 10+ years of operating experience and has proven to be quite accurate in determining the fraction of annual budget that must be allocated to adequately fund personnel for operations and new hardware deployments.

Figure 1 shows the annual planning budgets for the LQCD-ext and LQCD-ext II projects over the lifecycles of these projects, broken down into three categories:

1. Personnel
2. Travel, M&S, and Management Reserve.
3. Compute/Storage Hardware.

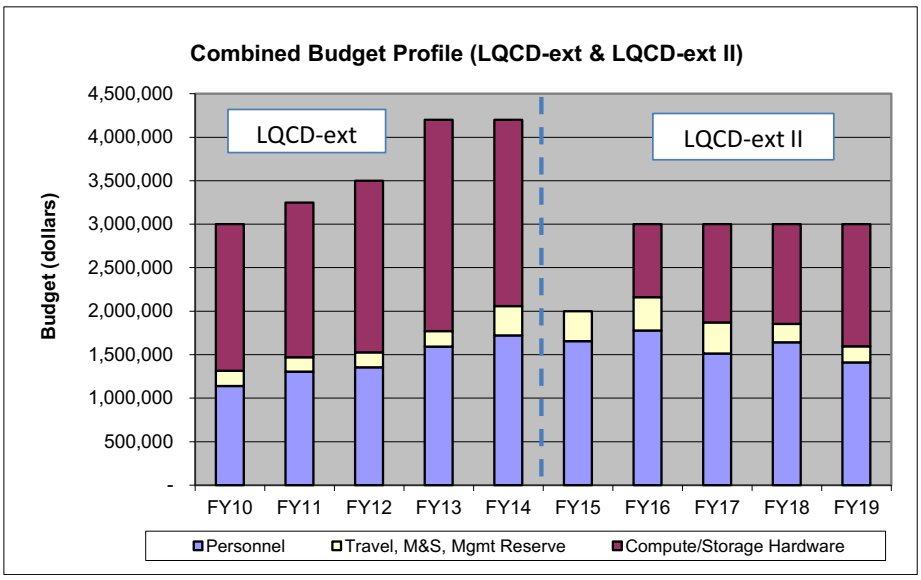

Figure 1. Planned budget profile for LQCD-ext and LQCD-ext II projects, showing distribution by budget categories.

The reason that the personnel budget stays relatively flat over FY15-19 in terms of dollars is that in the first few years of the LQCD-ext II project, the LQCD-ext II project is operating and supporting hardware deployed during the LQCD-ext project. As the older hardware reaches end of life and is retired, corresponding personnel support costs also decrease, which allows for a higher fraction of the annual budget to be allocated to hardware. This is precisely what is occurring in the latter years of the LQCD-ext II project and why hardware allocation fraction is larger in the later years. Table 2 shows the fraction of the total budget allocated to new hardware purchases over 10-year period.

Table 2. Fraction of total annual budget allocated to hardware, by year.

FY10	FY11	FY12	FY13	FY14	FY15	FY16	FY17	FY18	FY19
56%	55%	56%	58%	51%	0%	28%	38%	38%	47%

The USQCD community intends to solicit additional funding to continue acquiring, deploying and operating dedicated computing hardware beyond FY2019. If funding proposals are approved, then LQCD project operations will continue beyond FY2019 and the large hardware acquisition in FY2019 will be executed as planned to continue supporting the scientific program. However, contingency plans will also be developed in the event that a proposal for continued funding is not approved, in which case the planned FY19 hardware budget may be repurposed to support the operation of existing hardware through planned end of life. Development of these contingency plans will occur in FY2017.

Suggestion #7: **The project team stated that the BNL institutional cluster would be acquired using funds from non-DOE sources. This fact should be fully documented to avoid any potential misunderstandings or interpretations that the LQCD-ext II project acquired ~40 nodes on the BNL institutional cluster using additional DOE funds from outside the LQCD-ext II project.**

Report Section: **LQCD-ext II Review – Change Request: Add cluster-hosting at BNL to the baseline project plan**

Response: The Institutional Cluster (IC) is funded by a wide variety of sources, including programmatic funding from Basic Energy Sciences (BES), the Center for Nanofunctional Materials (CFN), Computational Materials Center, BNL institutional funds, NY state funds and buy-in from smaller projects funded by agencies such as ASCR, BER and others. As currently configured, DOE funds pay for approximately 1/3, and non-DOE funds pay for the remaining 2/3 of the IC. NY state funds will nearly double the size of the IC in FY17, and the DOE-funded share of the IC will then drop to approximately 1/6.