Department of Energy

Review Committee Report

on the

LATTICE QUANTUM CHROMODYNAMICS (LQCD) PROJECT

May 2006

Intentionally Blank

EXECUTIVE SUMMARY

The LQCD project supports the development and operation of a large scale dedicated computing facility capable of sustaining over seventeen (17) teraflop/s for the study of Quantum Chromodynamics (QCD) that will play an important role in expanding our understanding of the fundamental forces of nature and the basic building blocks of matter. The hardware is currently or will be housed at Brookhaven National Laboratory (BNL), Fermi National Accelerator Laboratory (FNAL), and Thomas Jefferson National Accelerator Facility (TJNAF or JLab), and operated as a single distributed computing facility, which will be available to lattice gauge theorists at national laboratories and universities throughout the United States. The project started in FY 2006 and is to be completed in FY 2009. The funding for this project supports the acquisition and operation of ~13 Tflops that, when combined with existing hardware, will yield a system capable of over 17 Tflops. The total cost is $9.2M. The President's FY 2006 budget allocated $2.5M to begin the project.

In a February 14, 2006, memorandum, Robin Staffin, Associate Director of Science for High Energy Physics and Dennis Kovar, Associate Director of Science for Nuclear Physics requested that the Office of Advanced Scientific Computing Research organize and conduct an Annual Technical Progress Review of the ongoing Lattice Quantum Chromodynamics (QCD) Computing Initiative. The purpose of the review was to evaluate the continued significance and relevance of the LQCD project; the progress toward scientific and technical milestones as presented in the project's IT Exhibit 300; the status of the technical design and proposed technical scope for FY 2007; the feasibility and completeness of the proposed budget and schedule, including workforce availability; and the effectiveness of the management structure.

On May 25-26, 2006 a progress review of this project was conducted at FNAL to evaluate the progress of the project against the FY 2006 goals as well as to review the FY 2007 hardware installation and operations plan. The review committee found that:

- The LQCD management and staff are doing an excellent job of meeting the project goals and facilitating the QCD science.
- They are bringing in the systems on time and on budget with focus on the end science goals.

- The LQCD project has effectively implemented most of the recommendations from the last year.
- The plan for procuring compute systems for the future seems very solid.

While the LQCD team presented a clear plan for compute system procurements in FY07 and beyond they did not present a complete facility plan for FY07 or future years. The committee urged the project team to work with FNAL staff to ensure that the specifications of the FNAL contributed computing facility will be adequate for the needs of the project. In addition, the committee suggests that the project develop integrated plans for data management, programming and user environments and facilities management to complement their compute system procurement plans.

The LQCD team is considering a plan to merge the FY08 and FY09 procurements into one which was encouraged by the committee. Such an advancement of funds would allow the project to deliver significantly greater resources in the last two years of the project and reduce the administrative costs of conducting the procurements.

Overall, the committee believed that the LQCD team has adopted project management principles and techniques effectively and appropriately, and implemented the project management recommendations from last year's reviews. The primary challenges remain in the execution of the project: determining the right allocation of funding towards staff support vs. hardware capacity, and improving the effectiveness of current and future systems. This balance must trade off the total hardware capability that can be purchased against the increased effectiveness of the users due to a more robust support infrastructure. The goal should be to deliver the most science for the funds invested.

CONTENTS

1. Introduction

Over the past six years, members of the United States lattice gauge theory community have worked together to plan the computational infrastructure needed for the study of Quantum Chromodynamics (QCD). The majority of the members of the community are involved in this effort. Research and development performed during this period has provided the groundwork for the construction of production hardware beginning in FY 2006. With support from the DOE's High Energy Physics (HEP), Nuclear Physics (NP), Advanced Scientific Computing Research and SciDAC Programs, prototype hardware has been designed, constructed and tested, and the software needed to use it effectively has been developed.

The LQCD project supports the development and operation of a large scale dedicated computing facility capable of sustaining over seventeen (17) teraflop/s for the study of Quantum Chromodynamics (QCD) that will play an important role in expanding our understanding of the fundamental forces of nature and the basic building blocks of matter. The project started in FY 2006 and is to be completed in FY 2009. The total cost is $9.2M. The President's FY 2006 budget allocated $2.5M to begin the project.

2. Progress relative to FY 2006 Plan

The Project has made substantial progress for FY 2006, accomplishing its milestones to date on or ahead of schedule. They have adopted good project management methods and are considering appropriate technical trade-offs. The team has been successful in obtaining contributions from FNAL laboratory management in the form of facilities to house the computing equipment at FNAL. At the time of the review, these facilities were under construction and substantial effort remained before the facilities would be acceptable for housing sensitive computing equipment. The committee urges the project team to work with FNAL staff monitor the effect of these facilities on the reliability and stability of the project equipment and develop options that could be exercised if the impact of the levels of magnetization in the building on system reliability were too high.

2.1. Findings

2.1.1. LQCD has made good progress in cross-site coordination and cooperation.
2.1.2. The JLab cluster was released to production 1 month ahead of schedule.
2.1.3. The Fermi cluster procurement is on schedule.
2.1.4. The requirements for large data support have not grown at the rate initially anticipated.
2.1.5. The scientists have reduced their requirements for large storage based on an evaluation of the tradeoffs between storage and compute investments.
2.1.6. The scope of the metafacility has been reduced.
2.1.7. The LQCD project is working to have a common runtime environment and is making progress, but this appears to be very ad hoc and informal, and is a secondary requirement to meeting site specific implementation.
2.1.8. The project management practices and controls are being effectively implemented and utilized.

2.2. Comments

2.2.1. The technical options available to the project are limited due to the computer room facilities at FNAL because the height of the raised floor does not permit its use as a cooling air plenum.
2.2.2. The facilities at FNAL are likely to lead to a low Mean Time Between Failures (MTBF) of the hardware placed there. For example, the computer room is located in a building that is magnetized, and has a substantial amount of dust in the air. Even after construction is completed and the substantial dust is cleaned up the magnetization of the structure may cause significant problems.
2.2.3. FNAL appears to be imposing procurement restrictions on the LQCD Project that go beyond DOE's substantial procurement requirements or is common at other

laboratories. Due to the rate of changes in computer equipment several other laboratories are permitted to issue RFPs for computing equipment before all the funds have arrived at the laboratory. While this should not be abused, in cases like LQCD this change, even though orders could not be placed before funds arrived, would accelerate the availability of resources to support science by as much as 3 months.

2.3. Recommendations

2.3.1. In the most recent FNAL procurement the timing of the RFP may not have allowed a full consideration of the new Socket F processor boards from AMD. In a number of other applications these new boards have provided significant benefit. The LQCD Project should carry out a more detailed analysis of the potential performance improvement of Socket F AMD processor boards for LQCD because the projected six-week delay is a small impact relative to the overall potential performance benefit.

2.3.2. Consider adding contingency to the FY08 project schedule now so that they can manage a delay in the acceptance of FY08 system, which has been named "Kaon" systems.

2.3.3. Consider further augmentations to FNAL facilities to improve the cluster environment and insure that they have appropriate housing for the FY08/09 cluster.

3. Evaluation of FY 2007 Proposed Plan

The FY 2007 plans are well defined and the committee has confidence the LQCD Project will meet the proposed schedule and performance expectations for the procurements. The plans for other aspects of the project could have more detail. In particular the committee is concerned that the metafacility and common computing environment need more coordination and planning across the sites.

3.1. Findings

3.1.1. The LQCD team has followed the 2005 review committee's advice to reduce the number of procurements.

3.1.2. The LQCD team has not decided what they're going to do for storage and I/O resources on the JLab system.

3.1.3. The LQCD team is considering a plan to merge the FY08 and FY09 procurements into one.

3.1.4. The LQCD presented a clear plan for FY07 and future procurements.

3.2. Comments

3.2.1. The review committee supports the plan to merge the FY08 and FY09 procurements.

3.2.2. While the LQCD team presented a clear plan for compute system procurements in FY07 and beyond they did not present a complete project or facility plan for FY07 or future years. A complete plan would include plans for data management, programming and user environment and physical facilities.

3.3. Recommendations

3.3.1. Document the science-based and change-based reasons and requirements for the proposed LQCD metafacility, the data environment, and a common computing environment across the three LQCD sites.

3.3.2. Provide a plan showing the planned milestones for the metafacility, common computing environment and data environment for the three project sites.

3.3.3. Request shifting the FY09 procurement funding to FY08 to support the combined FY08/FY09 procurement with a minimization of extra costs. The proposed FY 2009 procurement funding is approximately $500k and the ability to combine this funding with the larger FY 2008 hardware procurement funding would deliver the maximum capability almost 1 year earlier and avoid the fixed costs of doing the FY 2009 procurement. In addition, the larger FY 2008 procurement would give the project more leverage in negotiations with vendors.

4. Project Management

Overall, the LQCD team adopted project management principles and techniques effectively and appropriately, and implemented the project management recommendations from last year's reviews. The LQCD Project is positioned well to move forward through the next several years. The cooperation between the sites is encouraging and appears likely to continue to improve. The primary challenges remain in the execution of the project: determining the right allocation of funding towards staff support vs. hardware capacity, and improving the effectiveness of current and future systems.

4.1. Findings

4.1.1. The LQCD Project implemented the project management recommendations from last year's review. The project is now tracking external dependencies, is tracking storage costs, is forming MOUs, and is tracking the total costs to the participating laboratories.

4.1.2. An effective change control mechanism is in place with a graded approach that includes a project change control board for the decisions with the greatest impact.

4.1.3. The project's change control board was requested to accept the change from a single core Pentium to a dual core Pentium for the JLab 6N cluster. The CCB accepted this change.

4.1.4. Project costs are being managed effectively.

4.1.5. The LQCD Project has adopted project management effectively, up to and including project training of two key project staff members at FNAL.

4.1.6. The LQCD Project is considering an increase in FY07-09 manpower allocations with the cost to be supported by a decrease in hardware and a resulting decrease in delivered Tflops. This decrease is likely to require a CCB request.

4.2. Comments

4.2.1. The adoption of project management focused on the scope of the OMB 300 is resulting in coordination across the sites consistent with the stated objectives of the project.

4.2.2. The change control mechanism is being followed effectively.

4.2.3. The committee agrees with the decision by the CCB to switch to dual-core Pentiums for the JLab cluster.

4.2.4. Cluster utilization and allocation data are tracked, but not uniformly across the project. Utilization data does not appear to be presented effectively enough to support scientific analysis or optimization of workload in the future.

4.2.5. The committee supports the proposed manpower change, as the manpower contributes to overall project success and the committee had previous concerns

about the project being understaffed. This reduction in delivered Tflops is minor compared to the science risk of insufficient support.

4.2.6. Coordination of procurement planning and analysis across the sites appears to be informal and infrequent. The major project procurements could likely benefit by more coordination.

4.3. Recommendations

4.3.1. The committee recommends that system utilization be improved. The LQCD project should analyze usage data and user patterns, and reach out to other HPC facilities for workload optimization suggestions.

4.3.2. The committee recommends that the LQCD project carefully analyze FY07-FY09 staffing requirements and consider modifying the plans for further manpower in this same change that is being requested for FY07 should the analysis support this conclusion.

Intentionally Blank

APPENDIX A

Charge Memorandum

United States Government Department of Energy

memorandum

DATE: February 13, 2006

REPLY TO ATTN OF: Office of Science

SUBJECT: Annual Technical Progress Review of the Lattice Quantum Chromodynamics (QCD) Computing Initiative

TO: Michael Strayer, Associate Director, SC-30

This memorandum is to request that you organize and conduct an Annual Technical Progress Review of the ongoing Lattice Quantum Chromodynamics (QCD) Computing Initiative. This review should appropriately involve the input and participation of the science programs in the Office of High Energy Physics and Office of Nuclear Physics responsible for the effort.

A Technical, Cost, Schedule and Management Review was conducted by your Office on May 24-25, 2005, and was chaired by Dan Hitchcock. He presented the review panel's report to Ray Orbach, the Director of the Office of Science, on July 15, 2005. It would be ideal if Dan could continue as chairperson for this review in order to preserve continuity.

The Office of High Energy Physics and the Office of Nuclear Physics began this initiative in FY 2006 to develop a large-scale (additional ~ 13 TFlops, Total Estimated Cost ~ $9.2 million) Lattice QCD computing capability based on the most cost-effective technology available. On an annual basis, our offices need to be assured that the technical approach and planning for all aspects of Lattice QCD computing are optimized to maximize scientific productivity in the context of other efforts world-wide.

In particular, it is requested that your review evaluate:

- The continued significance and relevance of the LQCD project;
- The progress toward scientific and technical milestones as presented in the project's IT Exhibit 300.
- The status of the technical design and proposed technical scope for FY 2007;
- The feasibility and completeness of the proposed budget and schedule, including workforce availability; and
- The effectiveness of the management structure.

In addition, it is requested that you assess the status and plans for the continued development and evolution of the required software for Lattice QCD computing. The report should be submitted to the Office of High Energy Physics and the Office of Nuclear Physics by June 30th.

Robin Staffin
Associate Director
Office of High Energy Physics

Dennis Kovar
Associate Director
Office of Nuclear Physics

3613:q:\simongillo\Projects\LQCD\Review06\LQCD_charge_020206.doc

SC-26	SC-26	SC-26	SC-26	SC-20	SC-20	SC-20
JSimon-Gillo 2/ /06	DKovar 2/ /06	SCoon 2/ /06	GHenry 2/ /06	GCrawford 2/ /06	JMandula 2/ /06	RStaffin 2/ /06

APPENDIX B

Review Participants

Department of Energy Review of the
Lattice Quantum Chromodynamics (LQCD) Project

REVIEW COMMITTEE PARTICIPANTS

Department of Energy

Daniel A. Hitchcock, DOE/SC-21, Chairperson

Consultants

Remy Evard, ANL
William T.C. Kramer, LBNL
Mark Seager, LLNL

Observers

Sidney A. Coon, DOE/SC-26.1
John Kogut, DOE/SC-25.2

APPENDIX C

Review Agenda

Agenda for LQCD Computing Project Progress Review
FNAL, Batavia, IL
May 25-26, 2005

Day 1, Opening Session - May 25, 2006, Location: Wilson Hall, 1 East
Start of Meeting: 8:00 AM US Central

Time	Title	Speaker
7:45 AM	Setup	
8:00 AM	Executive Session	Panel
9:00 AM	Project Overview	Don Holmgren
9:15 AM	Computational Requirements	Don Holmgren
9:30 AM	2006 Acquisition at Jlab - Prototyping, Procurement, Commissioning, Performance	Chip Watson
9:50 AM	2006 Acquisition at Fermilab - Design, Prototyping and Projected Performance	Don Holmgren
10:00 AM	2006 Acquisition at Fermilab - Procurement (RFI, RFP, schedule, cost, and status)	Don Holmgren
10:15 AM	2006 Acquisition at Fermilab - Installation and Commissioning Plans	Don Holmgren
10:30 AM	Break	
10:45 AM	Proposed 2007 Acquisition (site, design, performance, schedule, and cost)	Don Holmgren
11:25 AM	Project Management	Don Holmgren, Bakul Banerjee
12:00 PM	Lunch	
1:00 PM	Open session (Q&A and talks requested by reviewers)	various
3:00 PM	Executive session	Panel
5:00 PM	Questions to LQCD project	

Day 2, May 26, 2006 . Location: Wilson Hall 1 East
Start of Meeting: 8:00 AM US Central

Time	Session
7:45 AM	Setup
8:00 AM	LQCD project responses to questions
9:00 AM	Executive session
11:00 PM	Debrief to LQCD team
12:30 PM	Adjourn

APPENDIX D

Acronyms

Acronyms

AMD Opteron	64 Bit CPU Chip from Applied Micro Devices
CKM	Cabibbo-Kobayashi-Maskawa (CKM) matrix elements related to CP violation
Dcache	Distributed Data Cache software, a joint effort of FNAL and DESY
DDR	Double Data Rate memory with speeds from 200 MHz to 333 MHz
DDR2	A new memory standard promoted by Intel. Potentially, it enables to reach higher frequencies and higher bandwidth.
FPU	Floating Point Unit
GigE	Gigabit Ethernet
IA32	Intel 32 Bit instruction architecture
ILDG	International Lattice Data Grid project
Infiniband	InfiniBand is an interconnect or I/O architecture that connects servers with remote storage and networking devices, and other servers. It can also be used inside servers for inter-processor communication. InfiniBand is a channel-based, switched fabric, point-to-point interconnect, which provides scalability and performance for a wide range of platforms and price performance points. InfiniBand provides a scalable performance range of 500 MB/s to 6 GB/s per link, meeting the needs from entry level to high-end enterprise systems
Itanium	Intel 64 Bit CPU Chip
MPI	Message Passing Interface
Myrinet	Myrinet, ANSI/VITA 26-1998, is a high-speed local area networking system designed by Myricom to be used as an interconnect between multiple machines to form computer clusters. Myrinet has much less protocol overhead than standards such as Ethernet, and therefore provides much better throughput and less latency while using the host CPU much less frequently.
PCI Express	an emerging (2004/2005) standard for high-speed graphics, likely to result in a 20% boost over 2003-era AGP 8x performance. The standard, supported by ATI and other vendors, delivers better power management, bi-directional simultaneous I/O and 4GB/s bandwidth
PowerPC	A family of RISC-based computer processors (chips) developed jointly by IBM, Apple Computer, and Motorola Corporation and used in IBM RS/6000 systems and Apple Macintosh computers
PPDG	Particle Physics Data Grid
OSG	Open Science Grid

QCDOC	QCDOC architecture has been designed to provide a highly cost-effective, massively parallel computer capably of focusing significant computing resources on relatively small but extremely demanding problems. This new design is a natural evolution of that used in our earlier QCDSP machines. The individual processing nodes are PowerPC-based and interconnected in a 6-dimension mesh with the topology of a torus. A second Ethernet-based network provides booting and diagnostic capability as well as more general I/O. The entire computer is packaged in a style that provides good temperature control and a small footprint. Central to this design is the IBM Blue Logic technology which makes possible the high-density, low-power combination of an industry standard RISC processor with 64-bit floating point, embedded DRAM, six-dimensional interprocessor communications and the wide array of predesigned functions needed to assemble a complete, functional unit.
QDP/QDP++	SciDAC Data-Parallel Programming Interface for C and C++ computer languages
QIO	SciDAC QIO/C intermediate level input-output package
QLA	SciDAC QLA linear algebra library
QMP	The QMP project is a national effort to provide a high performance message passing interface on various hardware platforms for Lattice QCD computing. This message passing interface aims to provide channel oriented communication end points to communication readers and writers with low latency and high bandwidth. QMP is tailored to the repetitive and predominantly nearest neighbor communication patterns of lattice QCD calculations.
Rambus	Rambus is a high-speed memory technology that uses a narrow 16-bit bus (Rambus channel) to transmit data at speeds up to 800MHz
SSE	SSE instructions are SIMD for single-precision floating-point numbers. SSE instructions operate on four 32-bit floats simultaneously.
VAPI	InfiniBand verbs applications programming interface
VLIW	Very Long Instruction Word, instruction sets with large-sized complex instructions encoded into one instruction.