

# **The SciDAC Lattice Gauge Theory Project**

## **Current Status and Future Prospects**

# Overview

- Hardware Status and Plans
- Software Status and Plans
- Funding Prospects
- Major Issues

# Two Hardware Tracks

- QCD on a Chip (QCDOC)
  - Columbia
  - BNL
- Optimized Clusters
  - FNAL
  - JLab
- “It is important to pursue both special purpose machines and commodity clusters. The two tracks have different risks, and it would be reckless to cut off one.” – The Wilczek Panel

# QCDOC Status and Plans

- All hardware components have been thoroughly tested.
- 128 node prototype machine is up and running.
- ASICs have been ordered for the Riken and UKQCD 1 TF development machines and 5 TF production machines.
- A substantial QCDOC will be built for the U.S. lattice gauge theory community by early fall.

# QCDOC Daughter Board



A daughter board which contains two independent QCDOC nodes.

# Current Cluster Hardware

- 48 node dual 2.0 GHz P4 Myrinet cluster began operation at FNAL in August 2002.
- 128 node single 2.0 GHz P4 Myrinet cluster began operation at JLab in September 2002.
- 128 node dual 2.4 GHz P4 Myrinet cluster began operation at FNAL in January, 2003.
- 256 node single 2.66 GHz P4 Gigabit Ethernet mesh cluster began operation at JLab in September 2003.

# Immediate Cluster Plans

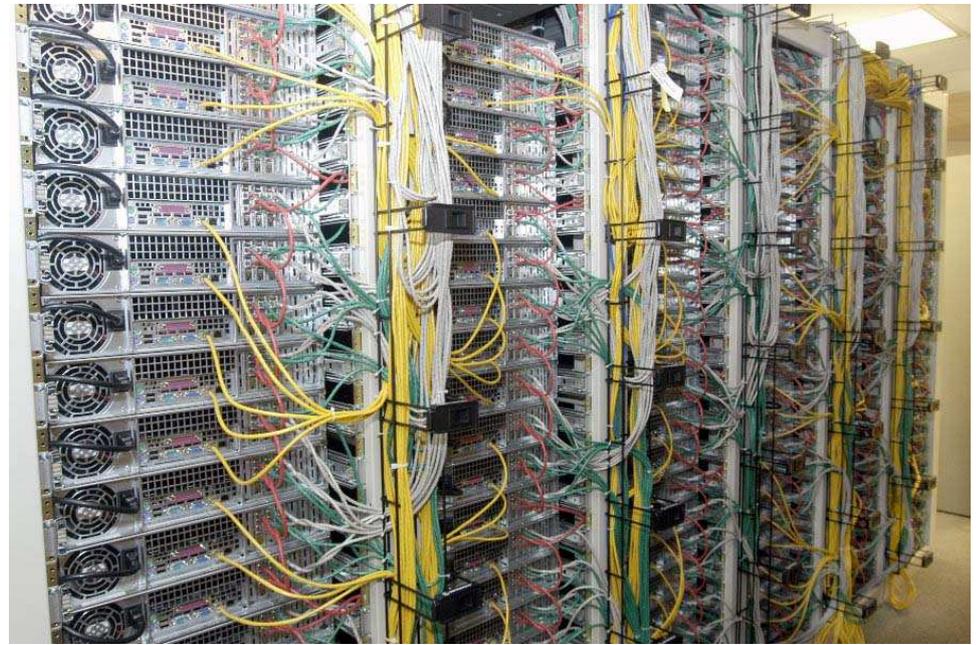
- Replace the processors on the older FNAL Myrinet cluster with 128 2.8 GHz P4E nodes in the spring of 2004.
- Construct a 32 node Infiniband cluster at FNAL in the spring of 2004, using dual 2.0 GHz P4 nodes from the older Myrinet cluster.
- Deploy either a 256 node Infiniband cluster, or a 512 node Gigabit Ethernet cluster at JLab in the late summer of 2004.
- Construct a 256 node 3.2 GHz P4E Infiniband cluster at FNAL in the fall of 2004.

# Most Recent FNAL Cluster



128 node dual 2.4GHz P4 Myrinet cluster, commissioned at FNAL in January 2003

# Most Recent JLab Cluster



256 node single 2.66 GHz P4 Gigabit Ethernet cluster, commissioned at JLab in September, 2003.

# ORNL Computing Resources

- IBM SP (Eagle)
  - 704 Power3 Processors (375 MHz)
  - 4 Processors/Node
  - 1.0 Tflops Peak
- SGI Altix (Ram)
  - 256 Intel Itanium Processors (1.5 GHz)
  - 2 TB of system memory
- IBM Power4 (Cheetah)
  - 27 p690 nodes
  - 32 1.3 GHz processors per node
  - Federated switch being installed (4 GB/s)

# SciDAC Software Effort

- The goal is to provide a uniform programming environment that will enable high efficiency use of clusters, the QCDOC and commercial supercomputers.
- QCD Applications Program Interface
  - Level 3: Highly optimized, computationally intensive subroutines.
  - Level 2: Data parallel language to enable rapid production of efficient code.
  - Level 1: Message passing and linear algebra routines.
- Work in progress on a uniform run time environment and I/O for clusters and the QCDOC.
- Participation in the International Lattice Data Grid.

# Fund Raising Activities

- Hardware Implementation Plan for 2004–2005 presented to the DOE on October 3, 2003.
- Presentation to the Physics Division of the National Science Foundation on October 10, 2003.
- The Nuclear Science Advisory Committee (NSAC) recommended that “even if external sources cannot be secured and the funding is flat, a minimal sum of \$3M/year from the nuclear science budget should be allocated to hardware investment.”
- At its meeting on February 9, 10, 2004, the High Energy Physics Advisory Panel (HEPAP) unanimously recommended that the DOE High Energy Physics Program fund computer hardware for lattice gauge theory.

# Funding Prospects

- Proposal to extend the SciDAC grant for FY 2004 and 2005 is under review.
- DOE–HEP will provide at least \$2M per year for hardware in FY 2004 and 2005. The funding level for FY 2006 and beyond has not been determined.
- We are hopeful that the DOE Nuclear Physics Program will provide funding for hardware beginning in FY 2005.
- Funding has been requested from the Advanced Scientific Computing Research Program.

# 2003 Milestones for the QCDOC

- At the 2003 Allhands Meeting we agreed to a set of milestones for the QCDOC.
- 2003 Construction criteria for 10+ Tflops QCDOC
  - Availability and performance of applications codes
  - Availability and reliability of hardware
  - Functionality of operating system
- Other Major 2003 Milestones
  - Convenient and robust user environment
  - High performance with level three inverters
- These milestones were to be achieved on a 1.5 TF Development Machine that was not funded. The milestones must therefore be revised, and most must be achieved on the 128 node prototype machine.

# Revised Milestones for the QCDOC

- QCDOC Construction Criteria
  - Availability and Performance of Applications Codes
  - Performance of Level Three Inverters
  - ASIC Clock Speed
  - Availability and Reliability of Hardware
- Other Major Milestones
  - Timelines for Efficient Lattice Evolution Software
  - Functionality of the Operating System
  - Convenient and Robust User Environment
  - Performance with Level Three Inverters

# Availability and Performance of Codes

Programs implementing the Hybrid Monte Carlo or Hybrid Molecular Dynamics algorithms must be available to evolve full QCD lattices using staggered, improved staggered, clover Wilson and domain wall fermions. The ability to run general community software must be demonstrated by showing that standard MILC code, using the QMP interface, but without specific optimization for the QCDOC, achieves at least 10% efficiency with a local volume of  $8^4$  lattice sites per node, on the full 128 node prototype machine.

# Performance of Level Three Inverters

Level-3 inverters, should achieve 40% efficiency with a local volume of  $2^4$  lattice sites per node using Wilson, clover Wilson and domain wall fermions on the full 128 node prototype machine. An improved staggered level-3 inverter should achieve 35% efficiency with a local volume of  $4^4$  lattice sites per node on the full 128 node prototype machine.

# ASIC Clock Speed

It is expected that the first two milestones will be run with the ASICs operating at a clock speed of 450 MHz, which is 90% of the design speed. The performance should be such as to guarantee that the final machine, running at the clock speed with which the milestones are passed, will achieve the target price/performance of \$1/Mflop/s for code that runs at 50% of the QCDOC's peak speed.

# Availability and Reliability of Hardware

- The ASICs have been tested extensively, and no flaws have been detected. They can be safely ordered once the first three milestones have been achieved.
- Large scale construction of backplanes, mother boards and daughter boards will not begin until reliable operation of the 2048 node development machines being constructed for RIKEN and UKQCD is established. This milestone is expected to be passed by the end of May.

# Reliable Operation of a Development Machine

The requirement for reliable operation of a 2048 node developed machine is specified as follows: The machine should run for a one week period with no more than one day lost in total for hardware maintenance and debugging on a combination of the application codes described above. (That is, at least an integrated six days of physics production running should be achieved. Uptime between the last checkpoint and a machine failure is not counted as production running.) Diagnostic support from the operating system should be sufficient that faults can be diagnosed and repaired without reference to the application code being run. During this week, 50% of the computer time should be spent on reproducibility checks, to verify that there are no undetected numerical errors.

# Timelines for Lattice Evolution Software

Timelines should be provided for the completion of the software needed for the time evolution of lattices with the Hybrid Monte Carlo or Hybrid Molecular dynamics algorithms for the Wilson, Wilson clover, domain wall, staggered and improved staggered actions. An estimate of the manpower needed for this work should also be provided.

# Functionality of the Operating System

Applications running on a node of QCDOC should have a run time environment with support for the standard C/C++ library, standard UNIX I/O routines and QMP. Applications must be able to read and write files on the host computer. The bandwidth to the host computer should be sufficient to permit a double precision  $32^3 \times 64$  lattice (1.2 GBytes) to be loaded or unloaded from the host in less than 1 minute (20 Mbytes/sec). The host computer should provide a UNIX-like environment for users to load and run programs.

# Convenient and Robust User Environment

A convenient and robust user environment must be provided. This will include the following functionality:

- A new user can easily register, log on and move code and data to the facility. Procedures are documented on the web and yield a new account within in three business days. The ssh and scp protocols are supported and 0.5 Gbyte of RAID disk space per user is routinely provided and backed up weekly. Much more space will be provided as required for particular projects.
- SciDAC specified file I/O is implemented allowing files created on a cluster to be used on QCDOC and vice-versa.

# Convenient and Robust User Environment

- MILC and SciDAC QDP(C and C++)codes compile easily.
- LQCD SciDAC batch protocols are supported and provide a uniform batch environment across both QCDOC and clusters. The batch queue and status of jobs can be viewed. Batch log files, stderr and stdoutfiles are readily available and there is a defined method (automatic or administrative procedure) to ensure that time allocations work.
- An adequate level of personal, user support is provided. This support will be provided uniformly to both cluster and QCDOC users with the actual implementation and delivery shared between BNL, FNAL and JLab.

# Performance with Level Three Inverters

CPS software, using level-3 inverters, should achieve 40% efficiency for full QCD hybrid Monte Carlo evolution with a local volume of  $2^4$  lattice sites per node using Wilson, clover Wilson and domain wall fermions on a 2048-node partition of a large QCDOC machine. MILC code, modified to call the improved staggered level-3 inverter, should achieve 30% efficiency for Hybrid Molecular Dynamics evolution on the same machine partition.

The last three milestones should be achieved by the time construction of the QCDOC is completed in the fall of 2004.

# Hardware Investment Strategy

Our goal has been to obtain \$10M per year in hardware funding beginning in 2004. In the event we receive that amount, our investment strategy for FY 2004–2006 would be:

- \$10M for a QCDOC in 2004
- \$10M for clusters in 2005
- \$10M for clusters in 2006

# Hardware Investment Strategy

Prior to the HEPAP meeting we were asked what the minimum investment by the HEP Program would have to be to produce a strong program. We responded \$3M per year. At that level our investment strategy for FY 2004–2006 would be:

- \$5M for the QCDOC in 2004 (with cash flow assistance from Columbia U.)
- \$1M for clusters in 2005
- \$3M for clusters in 2006

# Hardware Investment Strategy

We are assured of \$2M in FY 2004 and a second \$2M in FY 2005, with the possibility of higher funding in each of these years and in 2006. What should be our strategy for FY 2004 and 2005 if we receive between \$2M and \$3M in each of these years? It is assumed that all FY 2006 funds will be invested in clusters.

# Utilization of Terascale Computers

- We have told the DOE, HEPAP and NSAC that the infrastructure we are building will have a particularly strong impact on experimental programs in high energy and nuclear physics through:
  - Calculations of weak matrix elements needed for precise tests of the Standard Model.
  - Study of the properties of strongly interacting matter under extreme conditions.
  - Determination of the internal structure of hadrons.
- We need to balance
  - Lattice generation
  - Physics analysis
  - Algorithm development