

Clusters and GPUs for Lattice QCD

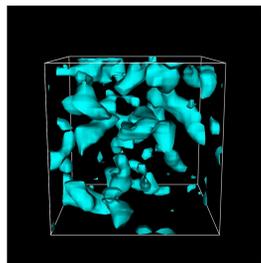
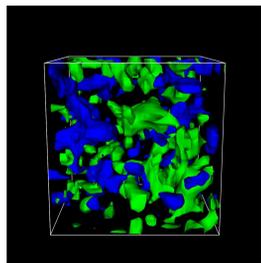
Computing Requirements

Lattice QCD codes spend much of their time inverting very large very sparse matrices. For example, a 48x48x48x144 problem, typical for current simulations, has a complex matrix of size 47.8 million x 47.8 million. The matrix has 1.15 billion non-zero elements (about one in every 2 million).

Iterative techniques like “conjugate gradient” are used to perform these inversions. Nearly all Flops performed are matrix-vector multiplies (3x3 and 3x1 complex). The matrices describe gluons, and the vectors describe quarks. Memory bandwidth limits the speed of the calculation on a single computer.

Individual LQCD calculations require many TFlop/sec-yr of computations. They can only be achieved by using large-scale parallel machines.

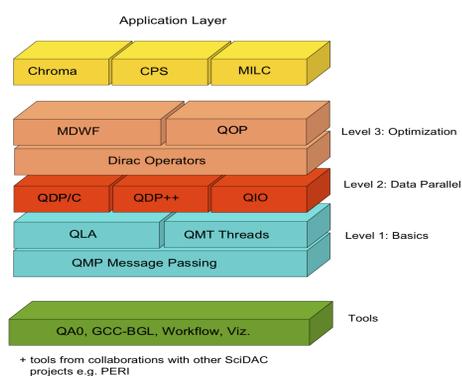
The 4-dimensional Lattice QCD simulations are divided across hundreds to many thousands of cores. On each iteration of the inverter, each core interchanges data on the faces of its 4D-sub-volume with its nearest neighbor. The codes employ MPI or other message passing libraries for these communications. Networks such as Infiniband provide the required high bandwidth and low latency.



SciDAC LQCD Software

The DOE Office of Science provided support for the development of Lattice QCD software during the SciDAC and SciDAC-2 programs and will provide support in the future under SciDAC-3 that will be crucial to exploiting exascale resources. USQCD physicists rely on these libraries and applications to perform calculations with high performance on all US DOE and NSF supercomputing hardware, including:

- Infiniband clusters
- IBM BlueGene supercomputers
- Cray supercomputers
- Heterogeneous systems such as GPU clusters
- Purpose-built systems such as the QCDOC

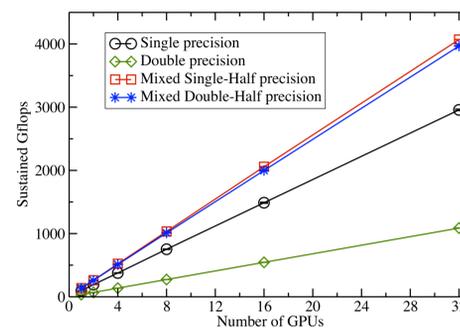


Accelerating LQCD with GPUs

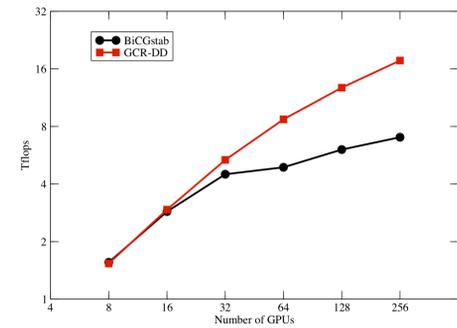
Since 2005 physicists have exploited GPUs for Lattice QCD, first using OpenGL with Cg, and more recently using CUDA. The codes achieve performance gains of 5x to 20x over the host CPUs.

Communication-avoiding algorithms (e.g. domain decomposition), and novel techniques such as using mixed precision (half-, single- and double precision) and compressed data structures, have been keys to maximizing performance gains.

Large problems require the use of multiple GPUs operating in parallel. GPU-accelerated clusters must be carefully designed to provide sufficient memory and network bandwidth.



Weak Scaling Performance Data Showing the Benefits of Mixed Precision Algorithms



Strong Scaling Performance Data Showing the Benefits of Domain Decomposition Algorithm



The screenshot shows the USQCD website with the following content:

- USQCD home | Physics program | Software | Hardware | USQCD Collaboration | Links and resources
- Fermilab Lattice Gauge Theory Computational Facility**
- Fermilab operates large clusters of computers for lattice quantum chromodynamics, as part of the national computational infrastructure for lattice QCD established by the Department of Energy. Their goal is the understanding of the strong dynamics of quarks and gluons, which is beyond the reach of the traditional perturbative methods of quantum field theory. A central goal of the groups using the computers is the accomplishment of the calculations required to extract from experiment the fundamental parameters of the Standard Model of particle physics.
- Physics Program: f_1 , f_2 , $2M_B - M_D$, $2M_B - M_D$, $v(1P - 1S)$, $T(1D - 1S)$, $T(1P - 1S)$, $T(1S - 1S)$, $T(1P - 1S)$
- LQCD Cluster Status: Sat Sep 24 08:29:40 CDT 2011
- The QCD cluster future
- The LQCD clusters



Fermilab USQCD Facilities

Fermilab, together with Jefferson Lab and Brookhaven Lab, operate dedicated facilities for USQCD, the national collaboration of lattice theorists, as part of the DOE Office of Science LQCD-ext Project.

The Fermilab “Ds” cluster has 13472 cores and uses QDR Infiniband. An earlier version with 7680 cores was #218 on the November 2010 Top500 list. It delivers 21.5 TFlop/sec sustained for LQCD calculations and was commissioned December 2010.

The Fermilab “J/Psi” cluster has 6848 cores, uses DDR Infiniband, and was #111 on the June 2009 Top500 list. It delivers 8.4 TFlop/sec sustained for LQCD calculations and was commissioned January 2009.

Fermilab has just deployed a GPU cluster with 152 nVidia M2050 GPUs in 76 hosts, coupled by QDR Infiniband. It was designed to optimize strong scaling and will be used for large GPU-count problems.

Jefferson Lab, under a 2009 ARRA grant from the DOE Nuclear Physics Office, deployed and operates GPU clusters for USQCD with over 500 GPUs. These are used for small GPU-count problems.