Report from CpF T3: Lattice Field Theory

Tom Blum, Don Holmgren & Ruth Van de Water

Snowmass on the Mississippi
July 30, 2013

Outline of written report

- 1.1 Introduction

- 1.2 Physics motivation

  - 1.2.1 Lattice field theory methodology and validation

  - 1.2.3 Lattice field theory for the energy frontier

- 1.3 Resources for lattice studies at the energy and intensity frontiers

  - 1.3.1 Computational requirements and current resources

  - 1.3.2 Projected capabilities in the next five years

  - 1.3.3 Lattice computing beyond 2018

- 1.4 Summary

Status of written report

- 1.1 Introduction ✓ First draft complete!

- 1.2 Physics motivation ✗ *In progress*

  - 1.2.1 Lattice field theory methodology and validation

  - 1.2.3 Lattice field theory for the energy frontier

- 1.3 Resources for lattice studies at the energy and intensity frontiers ✓ First draft complete!

  - 1.3.1 Computational requirements and current resources

  - 1.3.2 Projected capabilities in the next five years

  - 1.3.3 Lattice computing beyond 2018

- 1.4 Summary ✓ First draft complete!

Summary bullet 1: Scientific goals

✦ <u>The scientific impact of many future experimental measurements at the energy and intensity frontiers hinge on reliable Standard-Model predictions on the same time scale as the experiments and with commensurate uncertainties. Many of these predictions require nonperturbative hadronic matrix elements that can only be computed numerically with lattice-QCD.</u> The U.S. lattice-QCD community is well-versed in the plans and needs of the experimental high-energy program over the next decade, and will continue to pursue the necessary supporting theoretical calculations. Some of the highest priorities are improving calculations of <u>hadronic matrix elements involving quark-flavor-changing transitions</u> which are needed to interpret rare kaon decay experiments, improving calculations of the <u>quark masses m_c and m_b and the strong coupling α_s</u> which contribute significant parametric uncertainties to Higgs branching fractions, calculating the <u>nucleon axial form factor</u> which is needed to improve determinations of neutrino-nucleon cross sections relevant experiments such as LBNE, calculating the <u>light- and strange-quark contents of nucleon</u> which are needed to make model predictions for the $\mu \rightarrow e$ conversion rate at the Mu2e experiment (as well as to interpret dark-matter detection experiments in which the dark-matter particle scatters off a nucleus), and calculating the <u>hadronic light-by-light contribution to muon g – 2</u> which is needed to solidify and improve the Standard-Model prediction and interpret the upcoming measurement as a search for new physics.

Summary bullet 2: Computing needs

✦<u>The successful accomplishment of USQCD's scientific goals requires access to both capacity and capability machines, and hence support for both leadership class facilities and dedicated computing clusters.</u> Use of leadership-class facilities alone would provide insufficient computational resources needed to complete the planned calculations, and would be unsuitable for the mix of lattice-field- theory job requirements. USQCD's experience and proven track-record with purchasing, deploying, utilizing, and maintaining dedicated clusters will enable the collaboration to take advantage of future improvements in commodity clusters, such as increased core counts per processor and improved memory and networking bandwidth. USQCD's five-year computing strategy uses current vendor roadmaps to anticipate the probable evolution of high-performance computing hardware over this time period. The purchase of new dedicated lattice hardware on an annual basis, however, provides essential flexibility to accommodate changes and developments, and thereby to purchase the most cost-effective machines for lattice-field-theory calculations.

Summary bullet 3: Software needs

✦ <u>The successful utilization of future computing resources requires software that runs efficiently on new computing architectures, and hence support for postdocs and scientific staff to develop lattice-gauge-theory code</u>. Such positions cannot be supported by grants to lab and university theory groups alone. The USQCD Collaboration's libraries for lattice-field-theory calculations are publicly available and are used by most of the U.S. community. USQCD's experience and proven track-record in developing software for diverse machines such as IBM and Cray supercomputers, PC commodity clusters, and GPU-accelerated clusters, will enable the collaboration to fully exploit the computing capacity of future architectures.

Summary bullet 4: Conclusions

✦ <u>Support of USQCD through hardware and software grants, access to leadership-class computing facilities, and funding lab and university theorists, is essential to fully capitalize on the enormous investments in the DOE's high-energy physics and nuclear-physics experimental programs.</u> Given continued support of the lattice-gauge-theory effort in the U.S. and worldwide, lattice calculations can play a key role in definitively establishing the presence of physics beyond the Standard Model and in determining its underlying structure.

Computational Requirements

- Lattice QCD simulations require parallel machines
  - Codes are floating point intensive
  - Performance is limited by memory bandwidth, network latency and bandwidth
  - Applications are implemented using message passing (MPI, or custom communications code)
- The most computationally intensive work is generation of gauge-field configuration ensembles
  - Configurations in an ensemble are members of a Markov chain → jobs must be run serially
  - Time-to-solution is critical, so use the largest machines with the best strong scaling available → *capability computers*, such as IBM BlueGene or Cray
  - Jobs each use 10K to 100K+ processors
  - USQCD uses DOE Leadership Computing Facilities for these calculations (DOE ASCR funded)
    - Also NERSC, LLNL, NSF XSEDE, Japan (RIKEN BNL), UK (UKQCD BlueGene/Q)
- "Measurements" (operator expectation values) repeat a calculation with all members of an ensemble
  - Each job is independent of the others → run many jobs simultaneously
  - Time-to-solution of each job is not critical, so use *capacity computers*, such as clusters, running many "small" parallel jobs
  - Jobs each use hundreds to a few thousands of processors
  - USQCD uses dedicated LQCD capacity systems funded by DOE HEP + NP since 2006
  - *Measurements consume more that 50% of all Flops (and growing)*

Current Resources

- *Capability Hardware*

  - USQCD INCITE awards for lattice QCD simulations are consistently among the largest grants at both the ANL and ORNL LCF. At both sites substantial "zero priority" resources have been used.

  - Utilized @ ANL (by year since 2010): 187M, 182M, 143M, *290M* (2013 allocation)

  - Utilized @ ORNL (by year since 2010): 53.6M, 49.8M, 77.9M, *140M* (2013 allocation)

- *Capacity Hardware*

  - DOE HEP + NP have funded dedicated systems and support personnel at BNL, FNAL, JLab since 2006

    - *"LQCD"* (FY06-FY09) and *"LQCD-ext"* (FY10-FY14)

  - Utilized (by year since 2010): 125M, 205M, 330M, *971M* (2013 plan)

    - In 2013, 44K cores in total

  - Since late 2009, resources have included GPU-accelerated clusters

    - Currently operating 812 GPUs in total (including Fermi and Kepler models)

  - Storage: 1.8 Pbyte on tape (0.8 PB added in last year), > 1.4 Pbyte disk in Lustre parallel filesystems

Software

- DOE ASCR + HEP + NP have funded a series of lattice QCD software development grants since 2004 through SciDAC

  - Software stack includes libraries message passing, parallel I/O, site-wise linear algebra, parallel transport + computation, interfaces for optimized solvers

  - C and C++ interfaces

  - The *QUDA* library supports NVIDIA GPUs, with optimized Dirac operator solvers for all fermion actions of interest to USQCD

  - LQCD applications such as MILC, Chroma, and CPS can be built to use some or all of these SciDAC libraries

  - Libraries and applications are ported/tuned for new hardware

- Continued support for algorithm and software development is essential

Projected Capabilities (5 years)

- USQCD has submitted a follow-on 5 year proposal, *"LQCD-3"* for dedicated capacity hardware.

  - Assumed 1.5-year Moore's Law for available LCF resources, through higher core densities and further adoption of accelerators.

  - Capacity hardware plans assumes continued improvements in core count and memory bandwidth per socket, and network bandwidth. Also, hardware accelerators (GPUs and Intel MIC) are expected to influence price/performance.

- In LQCD-sustained-TF-yrs units (conversion is 6.53 M-core-hr/TF-yr), the proposed delivered computing capacity is as follows:

Year	Capacity (Cluster) TF-yrs	Leadership Class TF-yrs
2015	325	430
2016	520	680
2017	800	1080
2018	1275	1715
2019	1900	2720