Topical Groups:

■ (CompF2) Theoretical Calculations and Simulation
■ (TF05) Lattice Gauge Theory

Contact Information:

Carleton DeTar (University of Utah) [email]: detar@physics.utah.edu

Snowmass 2021 Letter of Interest
Algorithms and Software in Support of Computational HEP and NP at the Exascale and Beyond

R.C. Brower,[1] N.H. Christ,[2] C.E. DeTar,[3] R.G. Edwards,[4] and A.S. Kronfeld[5]

[1] *Department of Physics and Center for Computational Science, Boston University, Boston, MA 02215, USA*

[2] *Department of Physics, Columbia University, New York, NY 10027, USA*

[3] *Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA*

[4] *Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA*

[5] *Fermi National Accelerator Laboratory, Batavia, IL 60510 USA*

(Dated: August 29, 2020)

Lattice quantum chromodynamics (LQCD) provides crucial nonperturbative information about the strong interactions, allowing stringent tests of the Standard Model. It is also beginning to provide accurate parton distribution functions for analyzing proton collider events. The precision and predictive power of LQCD has vastly improved over the years, thanks in equal part to computer hardware and algorithmic development. The coming exascale computing era offers opportunities for even greater precision and a still greater likelihood of discovering new physics. Exascale computer architectures are complex with multiple levels of parallelism. The diversity of vendor hardware and evolving compiler layers present an enormous challenge for achieving good software performance and portability. In recognition of these challenges, the DOE Exascale Computing Project (ECP) has provided approximately $1 billion over five years to 24 application domain teams and a dozen software technology teams for exascale software and algorithmic development. However, the ECP ends in 2023, shortly after the deployment of the first exascale computers. The DOE office of Advanced Scientific Computing and Research (ASCR) charged a committee of experts to make recommendations for the exascale transition and beyond. Uninterrupted support for further exascale software and algorithmic development is essential for preserving the skilled ECP workforce and fulfilling the exascale promise for advances in high-energy and nuclear physics.

Two openly accessible exascale computers will be deployed in 2021, namely Aurora at the Argonne Leadership Computing Facility and Frontier at the Oak Ridge Computing Facility. Both will have several computational accelerators (GPUs) on each compute node. Aurora will have entirely new Intel accelerators on a Cray Slingshot network, and Frontier will have new, at least to LQCD, AMD accelerators on the same Cray network. Programmers must deal with five levels of parallelism: internode, intranode, accelerator multiprocessor, thread-level, and processor SIMD. Modern memory deployment is strongly heirarchical, which introduces another programming complication. Besides the usual host memory with associated cache and registers, special memory is shared between host and accelerator, and there are multiple types of on-board accelerator memory. Data motion is left to the skills of the programmer. Good performance requires expert care in managing that motion. Compilers are multilayer as well. Further, the organization of hardware and compiler software varies significantly among vendors. So portability among conventional clusters, existing NVIDIA architecture, and the Intel and AMD architectures is an important and challenging goal. The ECP is funding the conversion of major HPC codes in the hope that they will be capable of supporting the scientific ambitions of the various application domains at the exascale.

The three main code bases under development within the Lattice QCD application domain are the Columbia Physics System code from the Riken-Brookhaven-Columbia collaboration, the Chroma and QDP++ system from Jefferson Laboratory and collaborators, and the MILC code from the Fermilab Lattice and MILC collaborations. The three code bases emphasize different lattice-quark formulations, each with its own strengths and weaknesses, all of which are needed to cover the wide range of physical phenomena under study. Also supported is the Bielefeld-BNL HotQCD code for QCD thermodynamics studies. The first three of these codes rely on key modules in the QUDA library and/or Grid system to access accelerators.

Exascale software and algorithmic development is highly nontrivial. Altogether there are over two dozen scientists and software engineers involved in the Lattice QCD ECP effort. An effective strategy for porting codes to the new accelerators starts by porting the QUDA library and Grid system. Then the codes that rely on them will be mostly portable. That has been the primary occupation for the past couple of years for many of the members of the LQCD ECP team. Supported work involves developing both algorithms and software. Algorithmic research includes (1) development of accelerated methods for generating statistically independent gauge-field configurations, (2) devising adaptive multigrid solvers, and (3) combinatoric classification and optimization of contraction graphs for many-quark processes. Software development other than QUDA and Grid includes experimenting with software strategies for offloading work to the accelerators (OpenMP, OpenACC, DPC++, HIP), developing memory management for moving data between host CPUs and accelerators, devising code abstractions to improve portability, and incorporating researched algorithmic improvements in our software.

ECP support for the team software effort ends in 2023 soon after the first exascale machines are deployed. However, optimization work will continue, new algorithms and software will be required for new scientific goals. Thus work on algorithmic improvements and code optimization for the existing machines and preparations for the next generation of machines will still be needed. The skilled workforce developed under the ECP is at risk of disbanding once funding ceases. Indeed, much of our talent is already being snapped up by industry. To seek advice on the future of exascale computing after ECP, the DOE created an "ECP/Transition" subcommittee of the Advanced Scientific Computing Advisory Committee (ASCAC), chaired by Dr. Roscoe Giles of Boston University. That committee produced a draft report in Spring 2020 [1]. Some key recommendations are directly relevant to the future of HEP and NP computation:

- Recommendation A.3. "Collaborative applications support" "Transition ECP applications into SciDAC-like arrangements with joint funding from ASCR and application home organizations." In the past, support under the SciDAC program from the DOE offices of ASCR, HEP, and NP created crucial components of our current LQCD codes.
- Recommendation C.2: "Retain the current workforce"
- Recommendation C.3: "Strengthen ties to universities and the ecosystem" Here "ecosystem" refers to software technology experts and vendor system engineers. Universities are the pipeline for providing the young talent needed to sustain the software effort.
- Recommendation C.4: "Create career paths for scientific software professionals" For Lattice QCD, lab professionals at ANL, BNL, Fermilab, and JLab have been the backbone for expertise and continuity in LQCD software development.

Given the central role that non-perturbative QCD plays in the mission of High Energy Physics, it is urgent that new support be provided to enable the increasingly challenging algorithmic and software effort needed to advance this science and to fill the void that will be created when the Exascale Computing Project ends.

---

[1] ECP/Transition Subcommittee of the Advanced Scientific Computing Advisory Committee, "Draft ASCR/ECP Transition Report", (April, 2020).