**Accelerating Gauge Configuration Generation for Lattice QCD on Summit**

Kate Clark (NVIDIA), Balint Joo (Jefferson Lab), Mathias Wagner (NVIDIA), Evan Weinberg (Boston University), Boram Yoon (Los Alamos National Laboratory)

Jefferson Laboratory

bjoo@jlab.org

**Abstract**

We present our recent optimization of the Gauge Generation phase of lattice QCD for Summit, using the Chroma code and the QUDA Library for Lattice QCD on GPUs. The main optimizations were the integration of an algebraic multi-grid solver from QUDA into the gauge generation code in Chroma, updating the molecular dynamics time-stepping with a Force Gradient integrator and, subsequently re-tuning the molecular dynamics setup. In addition other performance oriented optimizations were added, such as switching to the Chronological predictor in QUDA, tuning the pipeline lengths of the Krylov Solvers used, and utilizing reduced precision halos exchange. Due to the larger memories available on Pascal and Volta GPUs on SummitDev and Summit, compared to the K20X GPUs in Titan we fit our problem onto substantially fewer nodes on the newer systems. In total, for our benchmark calculation the combination of hardware and software improvements have attained a wallclock speedup of ~9.1x on 8x fewer GPU devices on Summit, leading to an efficiency improvement of ~73x. Feeding the code improvements back to Titan resulted in wall-clocks speedups of 4.1x on 2x fewer devices leading to an overall improvement of ~8x. The nearly 2 orders of magnitude improvements on Summit compared to our previous running on Titan, and the nearly order of magnitude improvement on Titan, will not only multiply the real-term value of our OLCF allocations, but will also allow us to tackle problems hitherto considered too expensive and will fundamentally re-shape the structure of our calculation campaigns going forward.