Introduction

The generation of gauge field configurations is a necessary first step to any Nuclear or High Energy Physics Calculation using Lattice Gauge Theory methods. The gauge configurations sample the strong force fields in the vacuum.

Breaking the Determinant

The effect of dynamical sea quarks is to add determinant weights to $P_{eq}(U)$. These are simulated using pseudofermion terms in the action. A variety of terms can be combined to give the final forces.

Two Flavor Terms:

$$\text{det} \left[ M_{i}^{1}(U) \right] \rightarrow e^{-\phi[M_{i}^{1}(U) M_{i}(U)]^{-1}} \phi$$

Single Flavor Terms by Rational Approximation:

$$\text{det} \left[ M_{s}(U) \right] \rightarrow e^{-\phi[M_{s}^{1}(U) M_{s}(U)]^{-1/2}} \phi$$

Preconditioning Ratio Terms:

$$\text{det} \left[ M_{i}^{1} M_{i} \right] = \frac{\text{det} \left[ M_{i}^{1} M_{i} \right] \text{det} \left[ M_{i}^{1} M_{i} \right] \ldots \text{det} \left[ M_{n}^{1} M_{n} \right]}{\text{det} \left[ M_{i}^{1} M_{i} \right] \text{det} \left[ M_{i}^{1} M_{i} \right] \ldots \text{det} \left[ M_{n}^{1} M_{n} \right]}$$

Giving rise to:

$$\frac{\text{det} \left[ M_{i}^{1} \right]}{\text{det} \left[ M_{i}^{1} M_{i} \right] ^{-1}} e^{-\phi \frac{M_{i}^{1} M_{i}^{-1}}{M_{i}^{1}}} \phi$$

Force Terms and Solvers

Each pseudofermion component generates an MD force term. Evaluating these requires solving sparse linear systems with $M$. The two flavor terms can be combined to express the true flavor structure being simulated.

Two Flavor Terms:

$$\left( M_{i}^{1} M + q_{i} \right) X_{i} = \phi$$

These solves are typically done with a Multi-shift Conjugate Gradients (MCG) solver.

For light quarks, the system becomes ill-conditioned and the forces become large. We can tame the large forces by using ratio terms with longer time-steps. MCG gets all solutions in a single solve but does not scale very well on current interconnect fabrics. Two flavor terms allow scalable & multi-grid preconditioners. Our code uses solvers which have been highly optimized for GPUs, implemented in the QUDA library.

Adaptive Aggregation Multi-Grid

Adaptive Aggregation Multi-Grid Solvers reduce critical slowing down with quark mass from J.C. Osborn et al. arXiv: 1911.3775

Other Optimizations

Apart from the MG solver, we further reduced cost by implementing a force gradient integrator. We integrated the chronological predictor from QUDA into Chroma. In our solvers we optimized the pipeline-length. We implemented reduced (16-bit) precision for halo exchanges and are testing a more aggressive reduction (8-bit) to reduce bandwidth needs on the network. We upgraded the QDP-JIT software to use LLVM-6.0 (trunk) to enable Volta and POWER9 optimizations.

Results

Conclusions & Outlook

The joint effect of algorithmic and architectural optimizations, retuning of the MD structure and raw architectural performance optimizations improved the GPU hour cost of a benchmark trajectory by 73x on Summit through a 9.1x wall clock time speedup on 8x fewer devices. The improvements, when fed back onto Titan reduced the previous cost by 8x. This is a game changer for gauge generation.

Acknowledgement

This work is supported by the US Department of Energy, Office of Science, Offices of Nuclear Physics and Advanced Scientific Computing Research, through the SciDAC program under contract DE-AC05-06OR23177 under which JSA LLC operates and manages Jefferson Lab, and under the 17-SC-20-SC Exascale Computing Project.