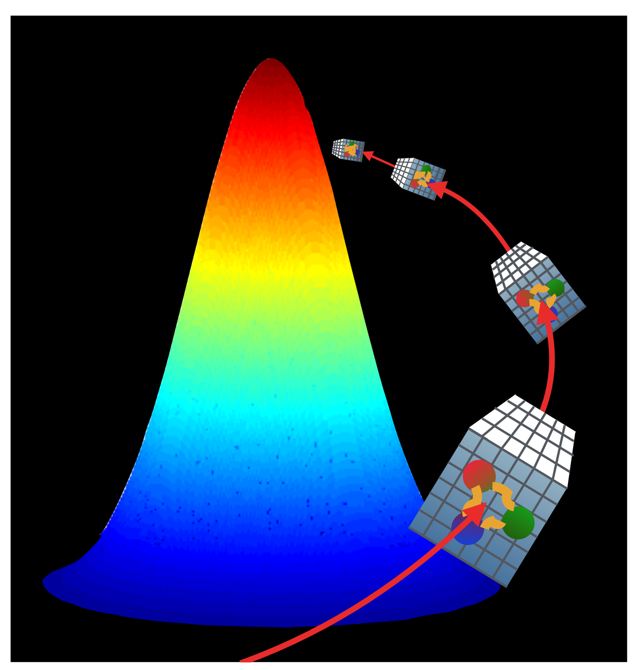
Accelerating Gauge Generation for Lattice QCD on Summit

Kate Clark (NVIDIA), Balint Joo (Jefferson Lab), Wei Sun (Jefferson Lab), Mathias Wagner (NVIDIA), Frank Winter (Jefferson Lab), Evan Weinberg (NVIDIA)

mclark@nvidia.com, bjoo@jlab.org, wsun@jlab.org mathiasw@nvidia.com, fwinter@jlab.org, weinbe2@bu.edu

Multigrid in HMC on GPUs

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.



*Gauge fields must sample the QCD Equilibrium dictated by the Action (**S**) of the theory.*

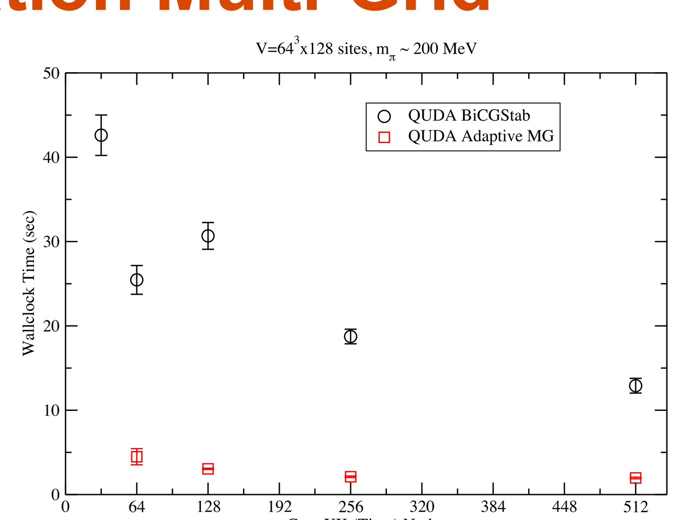
$$P_{eq}(U) \propto e^{-S(U)}$$

Hybrid Molecular Dynamics Monte Carlo Methods generate new configurations by suggesting new trial states from old ones using Molecular Dynamics (MD). The trial states are subject to a Metropolis accept/reject test.

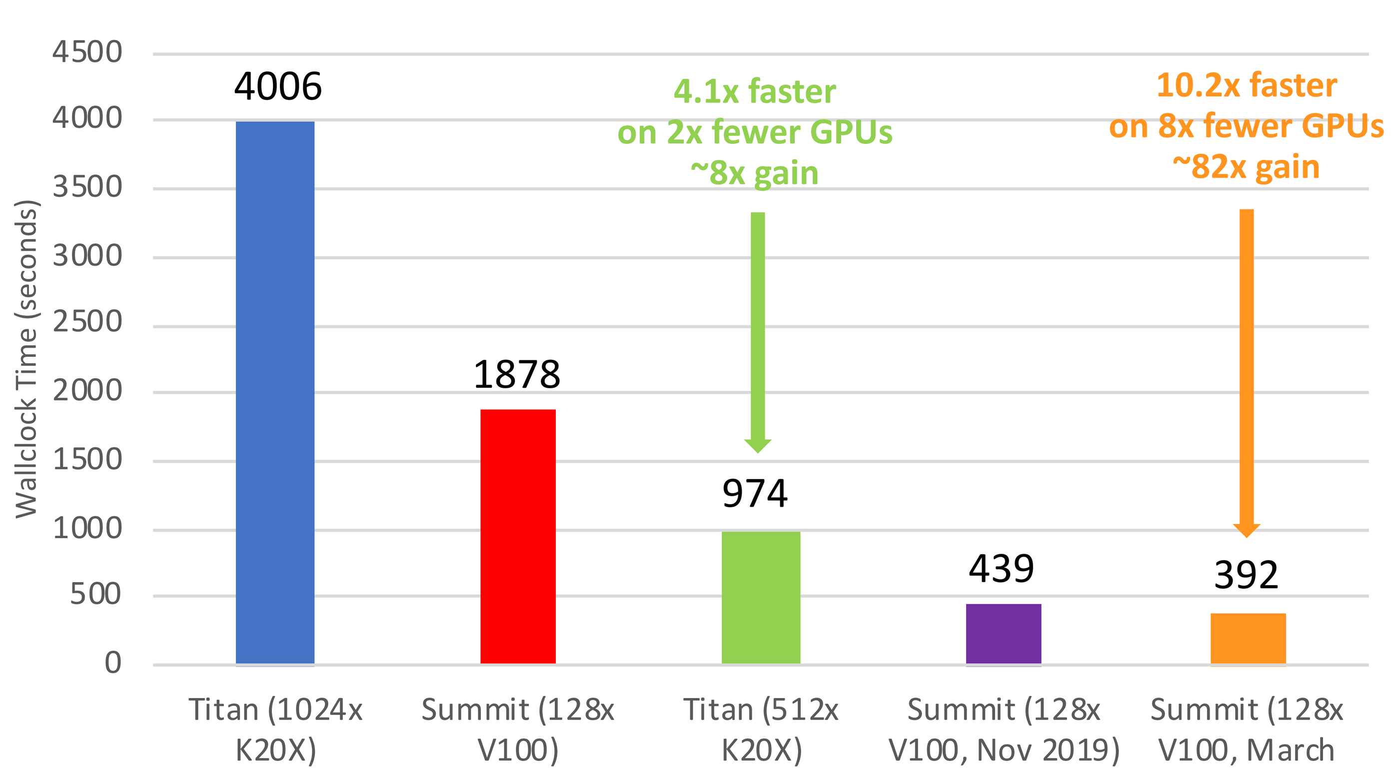
The MD defines canonically conjugate momenta (π) and integrates the (fictitious) time evolution of a Hamiltonian system with $H = T(p) + S(U)$. The integration scheme must be reversible and area-preserving to satisfy detailed balance.

Adaptive Aggregation Multi-Grid

Adaptive Aggregation Multi Grid in the QUDA Library, data from Clark et. al. SC'16 [1]. Solver outperforms BiCGStab in QUDA by 7x-10x



Integration of Adaptive Aggregation Multi-Grid solvers into a gauge generation lead to an integrated speedup of **73x.** in 2018. Since then the addition of communication avoiding GCR, pipelining and more aggressive use of reduced precision has improved this to **82x !!!**



Trajectory times for a benchmark on Summit and Titan showing overall gains. Due to the power of 2 problem size, only 4 out of 6 GPUs were used on Summit nodes from 2018.

Where to Next?

Even with the multi-grid solver, going to light quark masses some issues remain which need to be addressed

A Twisted Tale

During the MD evolution it is possible for the low-modes of the Dirac Operator to hit a near zero mode, when even the best multi-grid solver fails.

Adding small *twisted-masses* ($\mu_1, \mu_2, \mu_3 \ldots$) to the Dirac Operators used in the auxiliary determinant breakup terms instead of varying quark masses, *bounds the spectra of the operators from below and suppress zero modes.*

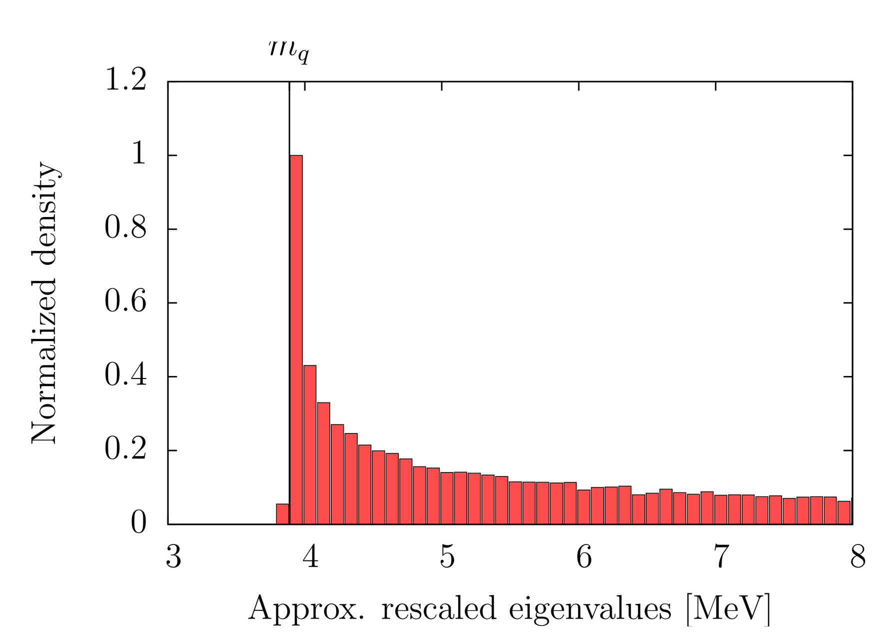
$$\frac{\det(M_1^\dagger M_1)}{\det(M_2^\dagger M_2)} \frac{\det(M_2^\dagger M_2)}{\det(M_3^\dagger M_3)} \cdots$$

$$M \rightarrow M + i\mu\gamma_5$$

$$\frac{\det(M_1^2 + \mu_1^2)}{\det(M_1^2 + \mu_2^2)} \frac{\det(M_1^2 + \mu_2^2)}{\det(M_1^2 + \mu_3^2)} \cdots$$

Adding μ_1 in the MD but not in the Metropolis acceptance step is *equivalent to reweighing to first order* in a cumulant expansion [2].

Will Multigrid work (well) for twisted operators?



Eigenvalue Spectrum of twisted mass clover operator (plot from arXiv:1610.02370 by C. Alexandrou et. al.) [3] showing the spectrum is dense just above the cutoff.

Existing results using ordinary TM-fermions showed that the density of low modes just above the spectral cut-off is high. This feeds through to the coarse operators. The authors of that study propose increasing the twisted mass on the coarsest level. In the QUDA library one can *deflate these modes,* tho deflation has some overhead cost (computing/updating the modes). An alternative approach would be to *use a direct solver* on the the coarsest level.

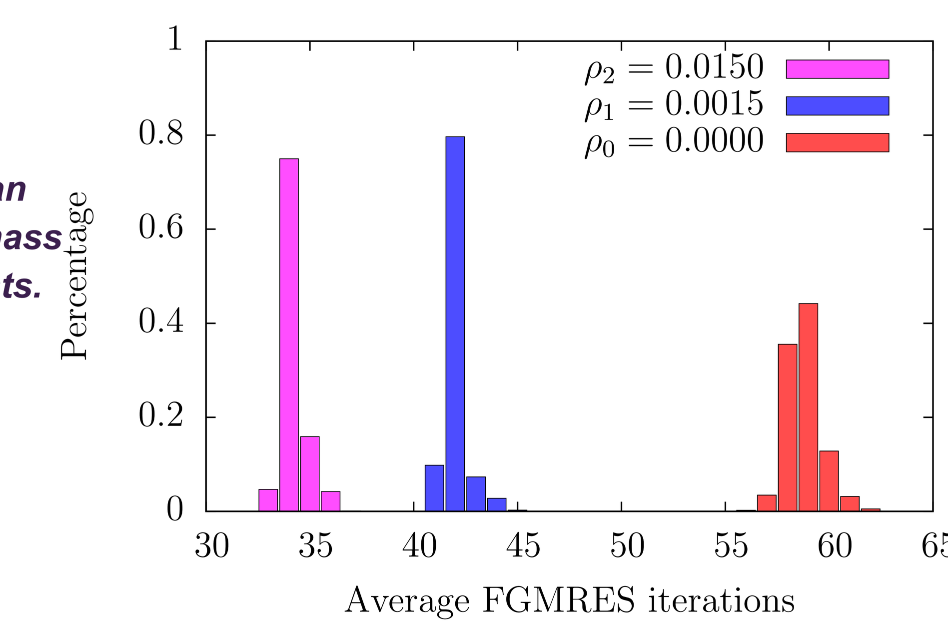
Will reweighing cause issues?

Reweighting can only be successful if the unweighted and reweighed distributions have a high overlap. Positive European experiences are discussed in [4,5].

Multi-RHS Solvers

The operators in the determinant breakup terms now differ only by their added twists. This feeds through to the coarse operators. This leads one to consider adapting multi-right hand side methods or quasi-block methods where in building the Krylov space in the Arnoldi process of FGMRES or GCR, one can reuse the linear operator for all the RHS and add on individual twists as needed.

The number of iterations in an HMC simulation of twisted mass fermions for a variety of twists. From arXiv:1610.02370 by C. Alexandrou et. al. [3]



One thing to consider is that adding a twisted mass should considerably improve the condition number, leading to lower iteration counts for different twists, potentially posing a challenge for efficient Multi-RHS solvers.

Conclusions & Outlook

Incorporating Multi-Grid solvers into HMC simulations have brought about a dramatic (as much as 82x) reduction in the cost of gauge generation. Our *current work in progress* addresses remaining issues of ***stability*** at very light quark masses while simultaneously *increasing application performance.*

References

Add QR Code here

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.