Tuning the Riemannian Manifold Hybrid Monte Carlo with Fermions

Sarah Fields, Norman Christ, Chulwoo Jung

The 41st Lattice Conference
University of Liverpool, United Kingdom

Aug 2, 2024

COLUMBIA UNIVERSITY

NSF

ECP
EXASCALE COMPUTING PROJECT

Objective

The objective of this study was to examine the methods of obtaining the necessary auxiliary determinant term in the RMHMC Hamiltonian in order to probe how the implementation of generating this term impacts the effectiveness of algorithm.

Outline

The RBC & UKQCD collaborations

Boston University
Nobuyuki Matsumoto

BNL and BNL/RBRC
Peter Boyle
Taku Izubuchi
Christopher Kelly
Shigemi Ohta (KEK)
Amarji Soni
Masaaki Tomii
Xin-Yu Tuo
Shuhei Yamamoto

University of Cambridge
Nelson Lachini

CERN
Matteo Di Carlo
Felix Erben
Andreas Jüttner (Southampton)
Tobias Tsang

Columbia University
Norman Christ
Sarah Fields
Ceran Hu
Yikai Huo
Joseph Karpie (JLab)
Erik Lundstrum
Bob Mawhinney
Bigeng Wang (Kentucky)

University of Connecticut
Tom Blum
Jonas Hildebrand
Luchang Jin
Vaishakhi Moningi
Anton Shcherbakov
Douglas Stewart
Joshua Swaim

DESY Zeuthen
Raoul Hodgson

Edinburgh University
Luigi Del Debbio
Vera Gülpers
Maxwell T. Hansen
Nils Hermansson-Truedsson
Ryan Hill
Antonin Portelli
Azusa Yamaguchi

Johannes Gutenberg University of Mainz
Alessandro Barone

Liverpool Hope/Uni. of Liverpool
Nicolas Garron

LLNL
Aaron Meyer

Autonomous University of Madrid
Nikolai Husung

University of Milano Bicocca
Mattia Bruno

Nara Women's University
Hiroshi Ohki

Peking University
Xu Feng
Tian Lin

University of Regensburg
Andreas Hackl
Daniel Knüttel
Christoph Lehner
Sebastian Spiegel

RIKEN CCS
Yasumichi Aoki

University of Siegen
Matthew Black
Anastasia Boushmelev
Oliver Witzel

University of Southampton
Bipasha Chakraborty
Ahmed Elgaziari
Jonathan Flynn
Joe McKeon
Rajnandini Mukherjee
Callum Radley-Scott
Chris Sachrajda

Stony Brook University
Fangcheng He
Sergey Syritsyn (RBRC)

Critical Slowing Down in Hybrid Monte Carlo

- Fourier acceleration attempts to tackle the issue of critical slowing down by modifying the mass term in the kinetic energy portion of the HMC Hamiltonian in a way that results in the low-modes moving at increased molecular dynamic velocities

- This is a procedure that becomes more complex when applied to QCD due to the fact that the theory is a local gauge theory

- The Riemannian manifold HMC (RMHMC) aims to achieve Fourier acceleration in QCD while maintaining gauge invariance by replacing the mass term in a HMC with a function of the SU(3) gauge-covariant Laplace operator (S. Duane and B.Pendleton Phys. Lett. B206, 101–106 (1988))

RMHMC

- The addition of a mass term that is dependent on the gauge links adds a $-\frac{1}{2}Tr\log|m[U]|$ term, thus there also needs to be a term in the Hamiltonian that produces a $\frac{1}{2}Tr\log|m[U]|$ to cancel out this unphysical term

- The Hamiltonian for the RMHMC is then:

$$H = S_G[U] + S_F[U] + \frac{1}{2}\sum_{\mu}[p_{\mu}^{\dagger}\frac{1}{m[U]}p_{\mu} + \pi_{\mu}^{\dagger}m[U]\pi_{\mu} + \phi_{\mu}^2]$$

$\uparrow$ $\frac{1}{2}Tr\log|m[U]| term$ $\uparrow$

here ϕ denotes the auxiliary fields and π denotes their corresponding momenta

- We currently input $m[U]$ and its inverse in the form of rational functions. Earlier studies of this method can be found in (T. Nguyen, *et al.* arXiv:2112.04556 [hep-lat] & C.Jung, *et al.* arXiv:2401.13226v1 [hep-lat])
- We are now studying this algorithm this on a 4 flavor ensemble with physical quark masses, $1/a = 4.0$ GeV, a volume of 32^3x64 with fermions, tuning this mass term and it's inverse for optimized acceleration and efficiency

Integration Scheme

- We utilize a Sexton and Weingarten integration scheme, where the Gauge action is separated from that of the Fermion action:

$$H = H' + S_F[U]$$
$$H' = S_G[U] + \sum_\mu [p_\mu^\dagger \frac{1}{2m[U]} p_\mu + \frac{1}{2}\pi_\mu^\dagger m[U]\pi_\mu + \frac{1}{2}\phi_\mu^2]$$

- This allows for $S_F[U]$ to be used in one integrator and then the elements of H' in another

$$T(\tau H) \approx T(\tfrac{1}{2}\tau H')T(\tau S_F[U])T(\tfrac{1}{2}\tau H')$$

$$T(\frac{1}{2}\tau H') \approx$$
$$[T(\frac{1}{4n}\tau \sum_\mu [p_\mu^\dagger \frac{1}{2m[U]} p_\mu + \frac{1}{2}\pi_\mu^\dagger m[U]\pi_\mu + \frac{1}{2}\phi_\mu^2])\, T(\frac{1}{2n}(\tau S_G[U])$$
$$\cdot T(\frac{1}{4n}\tau \sum_\mu [p_\mu^\dagger \frac{1}{2m[U]} p_\mu + \frac{1}{2}\pi_\mu^\dagger m[U]\pi_\mu + \frac{1}{2}\phi_\mu^2])]^n$$

where n is a positive integer.

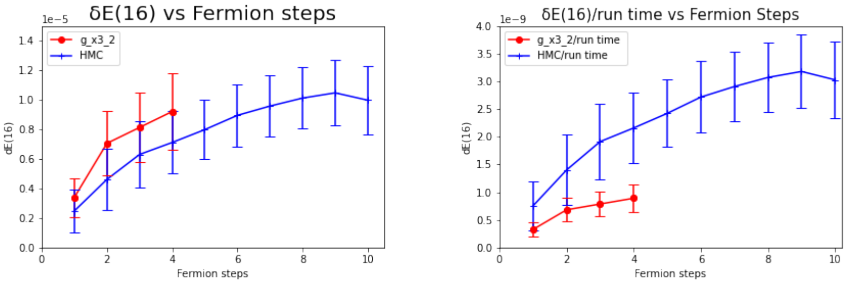
Method of Determining Movement of Modes

- When tuning the algorithm, we look at the size of change in the Wilson flowed energy as a more practical way of studying the success of attempted Fourier acceleration than obtaining the autocorrelation time
- Studying the Wilson flowed energy is useful when examining how successful the algorithms are at moving long-distance observables as this quantity examines the theory at lengths scales on the order of $\sqrt{t_w}$

The plot on the left is a plot of the average change in the Wilson-flowed energy of 12 configurations at Wilson flow time of 16 for the original optimized version of the RMHMC vs the HMC

The right plot divides $\delta E(16)$ by the run time. (the label g_x3_2 indicates an effective RMHMC mass term)

All runs were conducted on Frontier.



- ★ The RMHMC is more effective at moving long distance observables, but when you factor in added run time, this is not yet the case

Scaling Factor

- By introducing a scaling factor, λ, we can adjust how much we are sampling the auxiliary determinant term
- Recall the Hamiltonian for the RMHMC algorithm is:

$$H = S_G[U] + S_F[U] + \frac{1}{2}\sum_{\mu}[p_{\mu}^{\dagger}\frac{1}{m[U]}p_{\mu} + \pi_{\mu}^{\dagger}m[U]\pi_{\mu} + \phi_{\mu}^2]$$

- The corresponding expectation value of the gauge field momenta is:

$$\langle p \rangle \approx \sqrt{m}$$

- consequently the velocity is:

$$\langle v \rangle \approx \frac{1}{\sqrt{m}}$$

Scaling Factor

- We could divide the mass term by some constant, λ, which would make the Hamiltonian:

$$H = S_G[U] + S_F[U] + \frac{1}{2}\sum_{\mu}[p_{\mu}^{\dagger}\frac{\lambda}{m[U]}p_{\mu} + \pi_{\mu}^{\dagger}\frac{m[U]}{\lambda}\pi_{\mu} + \phi_{\mu}^{2}]$$

- Following the same logic as before, would have the expectation value of momentum and velocity:

$$\langle p \rangle \approx \frac{\sqrt{m}}{\sqrt{\lambda}} \qquad \langle v \rangle \approx \frac{\sqrt{\lambda}}{\sqrt{m}}$$

- Dividing by the scaling factor in the input mass term results in the velocity to increase by a factor of $\sqrt{\lambda}$

- To account for this, the input trajectory length can be multiplied by a factor of $\frac{1}{\sqrt{\lambda}}$. This ensures trajectories of equivalent lengths are being studied

Scaling Factor

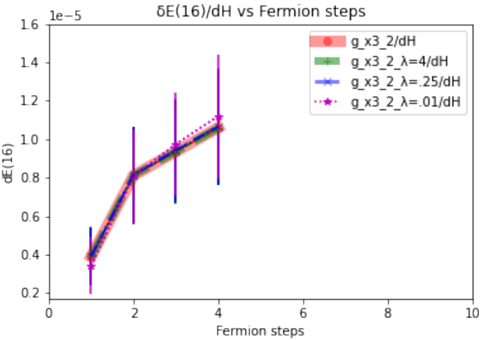
- The mass term associated with the auxiliary field is the inverse of that of the gauge field:

$$v_A \approx \sqrt{m} \qquad \Rightarrow \qquad v_A \approx \frac{\sqrt{m}}{\sqrt{\lambda}}$$

- This effect is not compensated by the manual change in the trajectory length
  - ➡ The scale factor λ has the effect of changing the amount of sampling that is done on auxiliary determinant
  - ➡ When λ is greater than 1, the velocity of the auxiliary field has been reduced $\rightarrow$ the amount of sampling of this term is also reduced
  - ➡ By introducing a scaling factor, one can tune the algorithm by determining how much the auxiliary determinant is being sampled

Scaling Factor Results

Data for various scaling factors, λ, applied to g_x3_2



Rational Function	Trajectory Length	FS 1 $\delta E(16)$	FS 2 $\delta E(16)$	FS 3 $\delta E(16)$	FS 4 $\delta E(16)$	δH
g_x3_2	0.12	3.4004E-06	7.0530E-06	8.1306E-06	9.2075E-06	0.6554
g_x3_2_$\lambda = 4$	0.06	3.3970E-06	7.0530E-06	8.1236E-06	9.1965E-06	0.6566
g_x3_2_$\lambda = .25$	0.24	3.3995E-06	7.0584E-06	8.1638E-06	9.2315E-06	0.6507
g_x3_2_$\lambda = .01$	1.20	3.1356E-06	7.4133E-06	8.9443E-06	1.0271E-05	0.7724

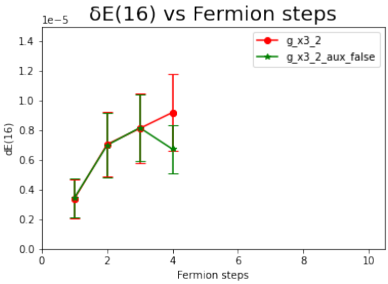
★ The scaling factor has little to no impact on $\delta E(16)$

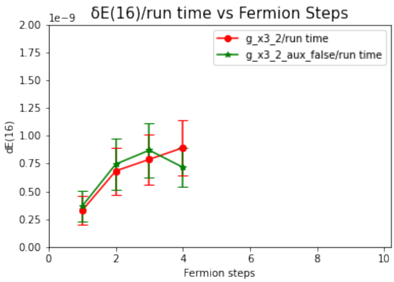
Auxiliary Term

- Since the auxiliary terms have been shown to have little to no impact on the movement of long-distance observables, we can make the auxiliary field non-dynamical
- This theoretically would reduce the run time of the algorithm, making the RMHMC more efficient

Results

Results of the change in Wilson-flowed energy of 12 configurations at Wilson time of 16 for the original optimized version of the RMHMC and the RMHMC with the non-dynamical auxiliary field approach.

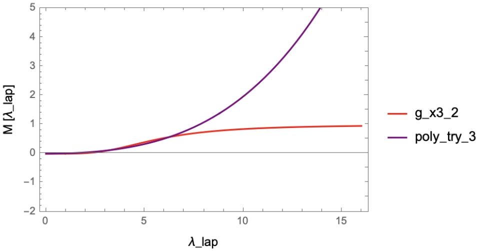


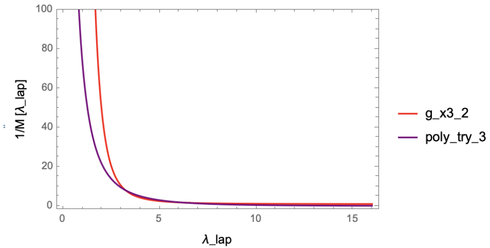


★ The run with the auxiliary fields non-dynamical was as effective as the original within error while reducing the run time by $\approx 10\%$

Polynomial Input

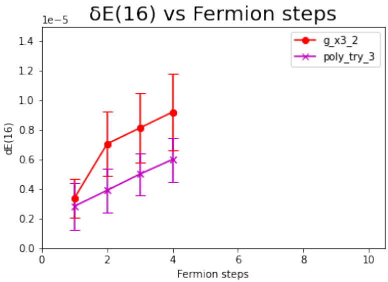
Now that we are treating the auxiliary field as pseudo-fermion field, we can benefit from simplifying the input of m[U] or 1/m[U] to a polynomial

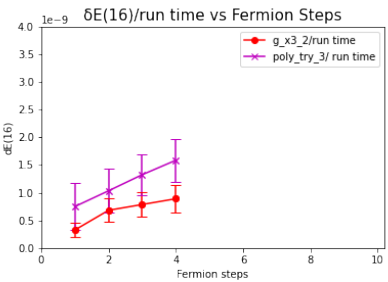




Preliminary Results

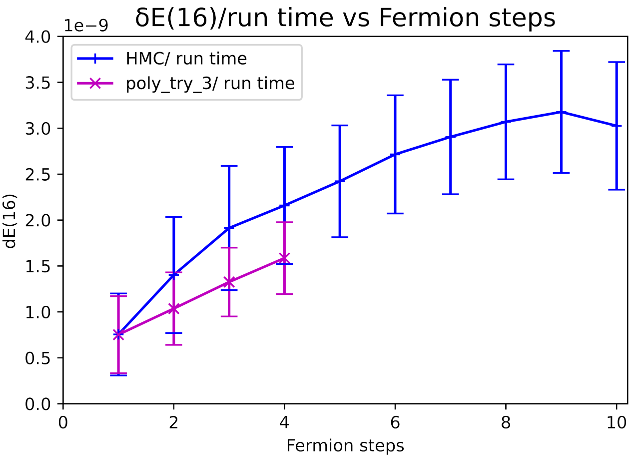
Results of the change in Wilson-flowed energy at Wilson time of 16 for the original optimized version of the RMHMC, and newer version of the RMHMC with the a polynomial $M[U]$





- ★ A polynomial mass function reduces run time significantly
- ★ The particular choice of function was not as effective at changing long distance observables as the original function, but he improved efficiency is promising.

Updated comparison of RMHMC to HMC



Summary and Outlook

- The RMHMC is effective at increasing the change of long-distance observables.
- Reducing the sampling of the auxiliary term showed little to no impact on long distance observables.
- This observation was utilized by making the auxiliary fields non-dynamical.
- Having the auxiliary terms non-dynamical had similar effects on moving long-distance observables as previous tests while reducing the run time.
- We can utilize this result to simplify the form of the input mass function or it's inverse to further improve efficiency of the algorithm.

Back Up Slides

Results

	RMHMC with Aux	RMHMC without Aux	HMC
FS 1 $\delta E(16)$	3.4022e-06	3.4618e-06	2.4859e-06
FS 2 $\delta E(16)$	7.0573e-06	7.0005e-06	4.6193e-06
FS 3 $\delta E(16)$	8.1301e-06	8.1739e-06	6.3036e-06
FS 4 $\delta E(16)$	9.2046e-06	6.7352e-06	7.1134e-06
δH	0.7760	0.9129	0.5025
Total RT (s)	10314.4915	9392.5951	3295.1259

- All trajectory runs were conducted on Frontier.

	RMHMC with Polynomial M	RMHMC without Aux	HMC
FS 1 $\delta E(16)$	2.8432e-06	3.4618e-06	2.4859e-06
FS 2 $\delta E(16)$	3.9159e-06	7.0005e-06	4.6193e-06
FS 3 $\delta E(16)$	5.0109e-06	8.1739e-06	6.3036e-06
FS 4 $\delta E(16)$	5.9949e-06	6.7352e-06	7.1134e-06
δH	0.2992	0.9129	0.5025
Total RT (s)	3780.9657	9392.5951	3295.1259

All trajectory runs were conducted on Frontier.

Sexton-Weingarten Terms

Rational Function	Inner SW terms	FS 1 $\delta E(16)$	FS 2 $\delta E(16)$	FS 3 $\delta E(16)$	FS 4 $\delta E(16)$	δH	Run Time
g_x3_2	12	3.4331E-06	7.1149E-06	7.9879E-06	8.9286E-06	0.6384	12146.8197
g_x3_2_SW_6	6	3.4330E-06	7.1136E-06	7.9880E-06	8.9296E-06	1.3656	7378.6452

Figure: The average difference in Wilson-flowed energy at Wilson time of 16 for each of the four fermion steps that make up a full RMHMC trajectory, average δH, and run time for g_x3_2 and g_x3_2_SW_6. Here FS is used to denote fermion step. Both runs were done in double precision. All averages were gathered from 10, (340-430) configurations. Here it can be seen that the reduced SW term has little to no impact on $\delta E(16)$, reduced run time, but consequently increases δH.

Iteration Break Downs

	Mixed	Double
MD time	15790.23517	10312.08841
CG iteration	165780	165780
CG time	1652.364304	1646.893077
Lap iteration	3183139	1869438
Lap time	10786.26581	6749.164435

This demonstrates that the Conjugate Gradient iterations per second is **100.35** for mixed and for double is **100.66**. The Lapalce iterations per second is **295.12** for mixed and for double is **276.99**.

	Auxiliary Term On	Auxiliary Term Off
MD time	10312.08841	9389.957807
CG iteration	165780	163207
CG time	1646.893077	1624.57916
Lap iteration	1869438	1648571.364
Lap time	6749.164435	5956.673893

The Conjugate Gradient iterations per second is **100.66** for with the auxiliary term is on and without it **100.46**. The Laplace iterations per second is **276.76** for on and for off is **276.99**.

Scaling Mixed Precision

Rational Function	Trajectory Length	FS 1 $\delta E(16)$	FS 2 $\delta E(16)$	FS 3 $\delta E(16)$	FS 4 $\delta E(16)$	δH
g_x3_2	0.120	3.4021E-06	7.0574E-06	8.1303E-06	9.2046E-06	1.6878
g_x3_2_scaled_.50	0.240	3.3984E-06	7.0612E-06	8.1651E-06	9.2339E-06	5.0289
g_x3_2_scaled_10	0.012	3.3981E-06	7.0536E-06	8.1191E-06	9.1971E-05	0.3341

Figure: The average difference in Wilson-flowed energy at Wilson time of 16 for each of the four fermion steps that make up a full RMHMC step, trajectory length, and average δH for g_x3_2, g_x3_2_scaled_2,g_x3_2_scaled_.50, and g_x3_2_scaled_10. Here FS is used to denote fermion step. Here all data was generated from the newer version of the RMHMC with mixed precision. All averages were gathered from 11, (340-440) configurations. Here it can be seen that the scaling factor has little to no impact on $\delta E(16)$, yet a notable impact on in δH and this difference grows as the scaling factor grows.

Topological Charge

Config ♯	FS 1 Topological Charge	FS 2 Topological Charge	FS 3 Topological Charge	FS 4 Topological Charge
340	4.36E-07	1.08E-06	1.77E-06	2.24E-06
350	**3.25E-07**	**9.95E-01**	**9.95E-01**	**9.91E-01**
360	2.84E-06	6.72E-06	1.06E-05	1.47E-05
370	**2.43E-01**	**2.43E-01**	**2.43E-01**	**2.43E-01**
380	1.03E-07	9.86E-07	2.25E-06	3.82E-06
390	2.06E-06	3.53E-06	4.19E-06	3.63E-06
400	7.27E-04	1.60E-03	2.40E-03	2.72E-03
410	2.70E-06	4.68E-06	6.14E-06	7.32E-06
420	1.27E-06	1.20E-06	2.00E-07	2.55E-06
430	2.88E-06	4.11E-06	3.83E-06	2.01E-06
440	1.99E-06	3.21E-06	3.89E-06	4.11E-06
450	1.34E-06	2.43E-06	2.63E-06	2.89E-06
460	1.60E-06	2.75E-06	3.24E-06	2.89E-06
470	7.97E-08	5.40E-07	1.28E-06	1.38E-06