PROCEEDINGS OF SCIENCE

Tuning the Riemannian Manifold Hybrid Monte Carlo with Fermions

Sarah Fields,[a,*] **Peter Boyle,**[b] **Norman Christ,**[a] **Chulwoo Jung**[b] **and Christoph Lehner**[c]

[a] *Department of Physics, Columbia University, New York, NY 10027, USA*
[b] *Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA*
[c] *Department of Physics, University of Regensburg, 93040 Regensburg, Germany*
E-mail: scf2157@columbia.edu, paboyle@ed.ac.uk, nhc@phys.columbia.edu, ypj@bnl.gov, christoph@lhnr.de

The Riemannian Manifold HMC (RMHMC) is designed to tackle critical slowing down by utilizing a rational function of the SU(3) gauge covariant laplace operator in place of the canonical mass term in the HMC algorithm. The RMHMC has been demonstrated to be effective at increasing the rate of change of long-distance modes. We present the results of the recent studies done to tune the algorithm and improve its efficiency.

The 41st International Symposium on Lattice Field Theory (LATTICE2024)
28 July - 3 August 2024
Liverpool, UK

*Speaker

PoS(LATTICE2024)045

1. Introduction

Decreasing the lattice spacing (a) towards the continuum limit works to reduce systematic errors in lattice QCD calculations as it allows for an improved approximation of the continuous nature of the physics being studied. A reduction in the lattice spacing introduces an increased energy scale, or shorter distance scale, which must now be taken into account. To properly accommodate this new scale, a smaller time step is required in the integration scheme. While introducing a smaller time step improves the accuracy, it consequently results in larger auto correlation times due to the fact that the slower evolving low modes are now fluctuating even less rapidly with each step. An increase in computation efforts is then required in order to control statistical errors. This phenomenon of increased auto correlation times as the continuum limit is approached is referred to as critical slowing down and causes significant challenges in lattice QCD calculations.

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 such a way that results in the low energy modes moving at increased molecular dynamic velocities. This is a procedure that becomes more complex when applied to QCD because the theory is a local gauge theory, thus making distinction between modes more challenging. The Riemannian manifold HMC (RMHMC) originally developed by Duane and Pendleton [1], 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. This can be effective because the eigenmodes of the theory are approximately covariant Laplace eigenmodes [2].

The RMHMC has been proven to be successful in achieving Fourier acceleration in studies conducted without [2] and with fermions [3]. Yet, due to its complexity, the RMHMC algorithm has been found to be computationally expensive compared to the HMC algorithm. We present the findings of initial studies done in efforts to tune this algorithm for optimized acceleration and efficiency on a 4 flavor ensemble with physical quark masses, domain wall fermions, inverse lattice spacing ($1/a$) of 4.0 GeV, and a volume of 32^3x64, smaller than what would be used in production.

2. The RMHMC algorithm

The RMHMC algorithm achieves Fourier acceleration while maintaining gauge invariance in QCD by writing the mass term in the kinetic energy Hamiltonian as a function of the SU(3) gauge-covariant Laplace operator. The addition of a mass term that is dependent on the gauge links creates an additional force of the form $F = \frac{1}{2}\,\mathrm{Tr}\{p_\mu^\dagger[\partial m^{-1}/\partial U]p_\mu\}$. We recover the original desired distribution with the addition of an auxiliary field, ϕ, with corresponding momenta, π, which generates the required $-\frac{1}{2}\,\mathrm{Tr}\log|m[U]|$ term in the action to cancel out this non-physical force. Making the final RMHMC Hamiltonian used:

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

Here $m[U]$ denotes the mass term and $S_G[U]$ and $S_F[U]$ denote the gauge and fermion action, respectively.

This mass term is then tuned in order to approximate the spectral dependence of the gauge force. Because both the mass term and its inverse appear in the Hamiltonian, the ease with which the two terms can be inverted must also be taken into account when tuning. We thus study choices

for $m[U]$ and $m[U]^{-1}$ in the form of rational functions. Earlier studies of this method can be found in [3].

We then follow the integration scheme of Sexton and Weingarten [4] where the Hamiltonian is redefined in such a way that separates the Fermion action from the gauge action

$$H = H' + S_F[U] \tag{2}$$

$$H' = S_G[U] + \frac{1}{2}\sum_{\mu}\left[p_\mu^\dagger \frac{1}{m[U]}\, p_\mu + \pi_\mu^\dagger m[U]\pi_\mu + \phi_\mu^2\right]. \tag{3}$$

This procedure permits $S_F[U]$ to be used in one integrator and then the elements of H' in another allowing for a longer time step to be used for the fermion force update. We choose T to denote some transition function that updates the coordinates or momentum and now write the integration scheme as follows:

$$T(\tau H) \approx T(\frac{1}{2}\tau H')T(\tau S_F[U])T(\frac{1}{2}\tau H'), \tag{4}$$

where the inner integration scheme is

$$\begin{aligned} T(\frac{1}{2}\tau H') \approx & \\ & [T(\frac{1}{4n}\tau \sum_{\mu}[p_\mu^\dagger \frac{1}{2m[U]} p_\mu + \frac{1}{2}\pi_\mu^\dagger m[U]\pi_\mu + \frac{1}{2}\phi_\mu^2])T(\frac{1}{2n}(\tau S_G[U])) \\ & \cdot T(\frac{1}{4n}\tau \sum_{\mu}[p_\mu^\dagger \frac{1}{2m[U]} p_\mu + \frac{1}{2}\pi_\mu^\dagger m[U]\pi_\mu + \frac{1}{2}\phi_\mu^2])]^n. \end{aligned} \tag{5}$$

Here n is a positive integer and τ is the integration step size.

3. Initial comparison of the RMHMC to the HMC

We start with an initial comparison of the most successful mass function demonstrated in the initial study of the RMHMC with fermions [3], (which hereafter will be referred to as "RMHMC_M1"), to the traditional HMC. We probe the algorithm's effectiveness in moving low modes by looking at the change in the Wilson flowed energy for a short, fixed-length trajectory. Studying the change in the Wilson flowed energy offers a more practical way of studying the success of attempted Fourier acceleration than obtaining the auto correlation time since the magnitude of the change in the Wilson flowed energy can be more easily determined than an auto correlation time. Note this Wilson flowed energy should be sensitive to the theory at length scales on the order of the square root of the Wilson flow time [5]. As our objective is to accelerate the motion of the low modes of the theory while accurately evolving all modes, we study the change in the Wilson flowed energy at a larger Wilson flow time to concentrate on the evolution of long-distance quantities.

When comparing the RMHMC to the HMC the total trajectory length for the HMC was set to the standard length of 1 and the choice of 8 integration steps was made in aim of keeping the change in the Hamiltonian, δH, within the reasonable range of 0 to 1 for the anticipated accept-reject step. The RMHMC trajectories typically contained fewer integration steps because of the limitations of the job queues on Frontier. The step size for the RMHMC runs was chosen so that δH values at the end of these trajectories were close to those of the HMC runs. (Note for the higher-order Omelyan we used, δH can be changed with minor adjustments in step size.) Since the meaning of molecular dynamics time is different between the HMC and RMHMC algorithms, when making

comparisons we instead compare the number of integration steps. Each integration step involves a single evaluation of the costly fermion-determinant force. It should also be noted that all trajectory runs in this study were conducted in double precision on 8 nodes of Frontier.

Through a comparison of the average of the change in the Wilson flowed energy at a Wilson flow time of 16 (referred to below as $\delta E(16)$) for 20 configurations we demonstrate how the original RMHMC_M1 is more successful at changing long-distance observables than the usual HMC. Fig. 1 shows the same change in $\delta E(16)$ takes only one-half the number of steps for the RMHMC. Yet, when factoring in the extra computational cost of the RMHMC algorithm the benefits of the RMHMC are diminished by its current inefficiency as the run for the HMC averaged ≈ 26.5 minutes while the RMHMC took substantially longer with a run time average of ≈ 66.5 minutes.

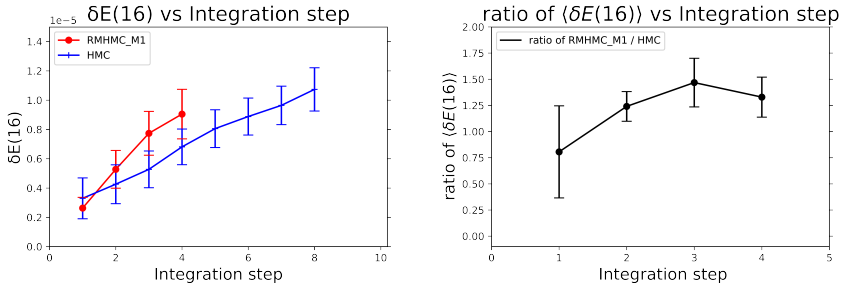

Figure 1: Left panel: The $\delta E(16)$ averaged over the same 20 independent initial configurations for the RMHMC_M1 compared to the HMC with jackknife errors. **Right panel:** The ratios of the average $\delta E(16)$ for the RMHMC_M1 divided by the average $\delta E(16)$ for the HMC for each integration step with jackknife errors. The average $|\delta H|$ for the HMC was 0.660 and the average $|\delta H|$ for the RMHMC_M1 was 0.737.

4. Sampling of the auxiliary determinant

To better understand the factors contributing to the computational cost of the RMHMC, we studied the sampling of the auxiliary determinant term.

From the Gaussian distribution of the gauge momentum in the RMHMC Hamiltonian defined in Eq. (2), the average magnitude of the gauge-field momentum is going to scale with the square root of the mass term. Consequently, the magnitude of the gauge velocity then scales with the inverse of the square root of the mass term

$$\langle |p| \rangle \approx \sqrt{m} \quad \Rightarrow \quad \langle |v| \rangle \approx \frac{1}{\sqrt{m}}. \tag{6}$$

We can study the sampling of the auxiliary determinant by introducing a tunable constant, λ. The RMHMC Hamiltonian would then be:

$$H = S_G[U] + S_F[U] + \frac{1}{2}\sum_{\mu}\left[p_\mu^\dagger \frac{\lambda}{m[U]} p_\mu + \pi_\mu^\dagger \frac{m[U]}{\lambda}\pi_\mu + \phi_\mu^2\right]. \tag{7}$$

Here the inconsequential choice to divide by λ was purely based on convenience when implementing the algorithm. Dividing the input mass term by this arbitrary scaling factor of λ in the Hamiltonian now results in the magnitude of the gauge-field velocity increasing by a factor of $\sqrt{\lambda}$

$$\langle |p| \rangle \approx \frac{\sqrt{m}}{\sqrt{\lambda}} \quad \Rightarrow \quad \langle |v| \rangle \approx \frac{\sqrt{\lambda}}{\sqrt{m}}. \tag{8}$$

To compensate for the change in the gauge-field velocity, the trajectory length is then multiplied by a factor of $\frac{1}{\sqrt{\lambda}}$. This adjustment ensures nearly equivalent gauge-field trajectories are being studied. For example, if the velocity of the gauge field is increased, we reduce the duration of the evolution resulting in gauge trajectories of equivalent length in both the original and scaled cases.

The mass term associated with the auxiliary field is the inverse of that of the gauge field resulting in the velocity of the auxiliary field, v_A, decreasing by a factor of $\frac{1}{\sqrt{\lambda}}$

$$|v_A| \approx \sqrt{m} \qquad \rightarrow \qquad |v_A| \approx \frac{\sqrt{m}}{\sqrt{\lambda}}. \tag{9}$$

This effect is not compensated but is instead enhanced by the manual change in the trajectory length. By introducing λ we can now alter the amount of sampling done on the auxiliary determinant and therefore have introduced a new tunable parameter.

We study the effects of sampling the auxiliary determinant by comparing the ratios of the average $\delta E(16)$ for the RMHMC_M1 with varying λ values divided by the average $\delta E(16)$ for the original RMHMC_M1. The ratios shown in Fig. 2 are all fairly close to one, demonstrating that all λ choices, whether they increase or decrease the amount of sampling on the auxiliary determinant, ultimately result in no notable difference on the RMHMC's ability to change long-distance observables. The most dramatic choice of λ demonstrates a ratio marginally larger than 1 for a larger number of integration steps, yet this also happens to be accompanied with larger errors indicating that additional statistics would be helpful to sharpen the conclusion.

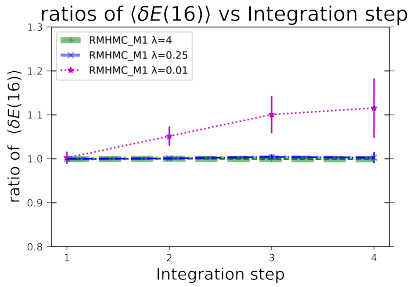

Figure 2: The ratios of the average $\delta E(16)$ for the RMHMC_M1 for varying scaling factors to that of the RMHMC_M1 for each integration step. The legend denotes the values of λ. Here it can be seen that altering the amount of sampling of the auxiliary determinant has little to no impact on the RMHMC's performance in changing the long-distance observables.

Originally the auxiliary fields and their conjugate momenta were both updated frequently since they were in the lowest level of the Sexton and Weingarten integration scheme. Now that we have shown that the frequency of sampling the auxiliary terms has little to no impact on the change in the long-distance observables over the length of a trajectory, we can replace the auxiliary field with the introduction of a pseudo-fermion-like stochastic Gaussian integral with a quadratic form given by the mass matrix m. Since the evaluation of both the operators m and m^{-1} is a substantial part of the extra costs of the RMHMC algorithm, moving the evaluation of m to the upper level of Sexton-Weingarten scheme should be an important improvement.

We then test a version of the RMHMC in which the same mass function of RMHMC_M1 is used, but now the auxiliary fields were replaced with the pseudo-fermion field approach. The version where the auxiliary fields are no longer dynamical is denoted with RMHMC_aux_ND. Fig. 3

suggests that when comparing ratios of the average $\delta E(16)$ for the RMHMC_aux_ND divided by that of the original RMHMC, this change in strategy for evaluating the auxiliary determinant may have decreased the acceleration of the evolution of $\delta E(16)$. However, given the large statistical errors in Fig. 3 and the studies reported above showing that the rate of sampling of the auxiliary determinant had little effect on $\delta E(16)$, we interpret Fig. 3 as showing a ratio consistent with one at the 1.42 σ level. This is clearly a test that requires more statistics to be definitive. The new sampling method reduced the run time by ≈10 %. Note, the improved run time is slightly offset by an increase in δH. Thus, we will assume that no significant degradation in the algorithm resulted from this change to a non-dynamical treatment of the auxiliary determinant and continue to use this "enhancement" in the remainder of these proceedings.

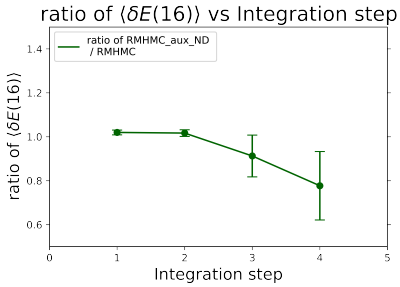

Figure 3: The ratio of the average $\delta E(16)$ for RMHMC_aux_ND divided by that of the RMHMC for each integration step with jackknife errors. The average runtime for the RMHMC_aux_ND was approximately ≈ 60 minutes, compared to the ≈ 66.5 minutes of that of the RMHMC. The average $|\delta H|$ of RMHMC_aux_ND was 0.774 and 0.727 for RMHMC. The ratio for each integration step is within 1.42σ to one, yet the large error on the last two integration steps demonstrates that additional statistics are required.

5. Polynomial forms of the input mass

Originally we were studying versions of the RMHMC algorithm in which both the RMHMC mass, $m[U]$, and its inverse, $m[U]^{-1}$ were evaluated as rational functions of the covariant Laplace operator. This was done because both $m[U]$ and $m[U]^{-1}$ appeared symmetrically in the lowest level of the Sexton-Weingarten integration scheme. In addition, rational functions provided a flexible format to match the desired shape where the mass function of RMHMC_M1 had a shape based on the gauge force density measured as a function of the spectrum of the Laplace operator [2, 3]. The improvements made to the integration scheme described in the previous section break the symmetry between the appearance of the functions $m[U]$ and $m[U]^{-1}$ and suggest that it is no longer in our best interest for these two functions to be equally easy to invert. In efforts to improve computational efficiency, we can utilize a Chebyshev polynomial approximation of one of the rational functions, allowing for one of the input functions to be simplified. We will refer to this mass function as RMHMC_M2. With utilization of a Chebyshev polynomial fit, we capture the desired shape while also reducing the complexity of the input functions in hopes to improve efficiency as half of the inversions are now less strenuous.

The ratio plot in Fig. 4 demonstrates how simplifying the input mass term results in changes in $\delta E(16)$ that are practically equivalent within error to the previously favored mass function RMHMC_M1, while producing a 50% speed up.

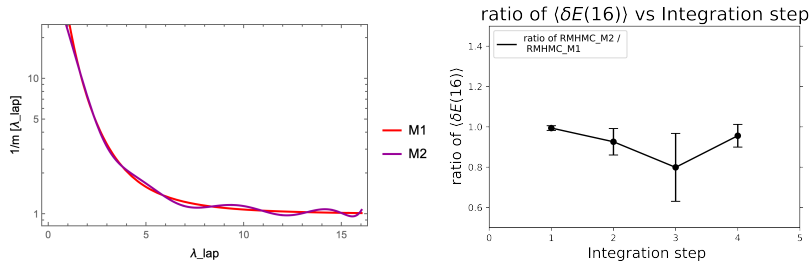

Figure 4: Left panel: Log plot of RMHMC_M2 and RMHMC_M1 versus the eigenvalue of the Laplace operator. **Right panel:** The ratio of the average of the $\delta E(16)$ of RMHMC_M2 divided by that of the RMHMC_M1 at each integration step with jackknife errors. Both runs were conducted with the new treatment of the auxiliary fields as discussed in the previous section. The average runtime for the RMHMC_M2 was ≈ 33 minutes, compared to ≈ 60 minutes of that of the RMHMC_M1. The average $|\delta H|$ for RMHMC_M2 was 0.735 and 0.774 for RMHMC_M1.

We can now compare the $\delta E(16)$ for the RMHMC_M2 to that of the HMC and the original RMHMC. Fig. 5 shows how RMHMC_M2 maintains the original RMHMC's effectiveness in changing long-distance observables within 1.57σ. The average runtime found in the RMHMC_M2 results demonstrates improved computational efficiency of the RMHMC allowing it to be more competitive when compared to the HMC in computational effort. The substantial errors demonstrate the necessity for additional statistics to better understand the scope of the RMHMC's effectiveness.

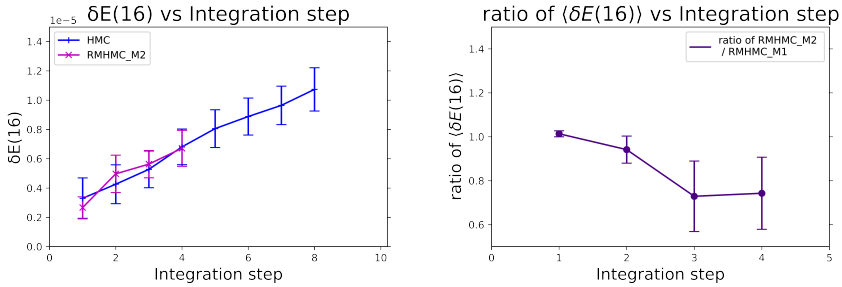

Figure 5: Left panel: The averages of the $\delta E(16)$ for the RMHMC_M2 compared to the HMC. Here the average $|\delta H|$ for the HMC was 0.660 and 0.735 for the RMHMC_M2. The average run time for RMHMC_M2 was ≈ 33 minutes and the average runtime for the HMC was ≈ 26.5 minutes. **Right panel:** The ratios of the average $\delta E(16)$ for the RMHMC_M2 divided by the average $\delta E(16)$ for the original RMHMC_M1, where the auxiliary fields were dynamical, for each integration step with jackknife errors. The two versions of the RMHMC agree within 1.57σ.

6. Summary and Discussion

Critical slowing down currently contributes greatly to the systematic errors challenging many lattice calculations. The RMHMC has shown promising results in efforts to overcome critical slowing down through achieving Fourier acceleration by introducing a mass term that is a function of the SU(3) gauge-covariant Laplace operator. Here we explore strategies for reducing the run time of the RMHMC algorithm on lattice configurations suitable except for their relatively small $32^3 \times 64$ volume for production runs.

We have suggested replacing the auxiliary fields in our original RMHMC formulation with a Gaussian, pseudo-fermion-like evaluation of the auxiliary determinant as a possible avenue for

reducing computational cost. Furthermore, it has been demonstrated that we can additionally reduce the run time by simplifying the form of the hand tuned RMHMC mass and accompanying inverse function. These two improvements together reduce the run time for a short trajectory from 60 to 33 minutes. In the right panel of Fig. 5 we show the ratio of $\delta E(16)$ for the improved algorithm divided by that for the original algorithm which shows a reduction in effectiveness at the 1.57 sigma level. As discussed above, we believe that the actual ratio should be much closer to one but more statics are needed to resolve this issue. For completeness in the left panel of Fig. 5 we show a similar ratio comparing the effectiveness with that of the original HMC. As can be seen the 30% advantage of the RMHMC shown in Fig. 1 has disappeared for these algorithmic improvements because of the loss of effectiveness suggested in Fig. 3. It should be emphasized that all comparisons of acceleration are shown as a function of the integration step size, a direct measure of computational effort in the limit that the fermion force dominates. The 2× gain in execution time achieved with the methods presented is for physical quark masses but an unrealistically small volume and it is not known how this will change as the volume is increased. Furthermore, the errors observed in this work highlight the need for substantial computational resources in production-scale algorithm development. It should also be noted that additional statistics were added since the oral presentation.

The results of this study shown in Fig. 2 suggest that the auxiliary determinant can be updated less frequently without negatively impacting the RMHMC. This result can be utilized to further improve the algorithm's efficiency by altering the integration scheme so that the auxiliary determinant is on the highest Sexton and Weingarten level. This would lead to a substantial disparity between the number of times the functions $m[U]$ and $m[U]^{-1}$ were evaluated (12x for our current implementation). We could then further benefit by using the simpler polynomial form for $m[U]^{-1}$ in the gauge-field kinetic energy which appears on the lowest level of the Sexton and Weingarten integration scheme. This change would theoretically further improve the efficiency of the algorithm as the evaluation of the function that is more costly would need to be done substantially less frequently.

It should also be noted that the RMHMC algorithm utilized in this study is not an exact algorithm as for finite time step it violates symplecticity [6]. However, we believe that the effect of this violation is most likely inconsequential for the approximate studies reported here since in earlier work as the RMHMC upheld the required $\langle e^{-\delta H}\rangle = 1$ property of proper sampling and no discrepancies have been observed between RMHMC and HMC results.

Acknowledgments

We thank our colleagues of the RBC and UKQCD collaborations for insightful discussions and ideas. This research was supported by the Exascale Computing Project (17-SC-20-SC). We are grateful for the computational resources of Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. We would also like to thank the Chien-Shiung Wu Scholarship for supporting SCF. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2036197. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. In addition NHC was supported by U.S. DOE grant No. DE-SC0011941.

References

[1] S. Duane and B.J. Pendleton, *Gauge invariant Fourier acceleration*, *Phys. Lett. B* **206** (1988) 101.

[2] T. Nguyen, P. Boyle, N.H. Christ, Y.-C. Jang and C. Jung, *Riemannian Manifold Hybrid Monte Carlo in Lattice QCD*, *PoS* **LATTICE2021** (2022) 582 [2112.04556].

[3] C. Jung and N.H. Christ, *Riemannian manifold HMC with fermions*, *PoS* **LATTICE2023** (2024) 009 [2401.13226].

[4] J.C. Sexton and D.H. Weingarten, *Hamiltonian evolution for the hybrid Monte Carlo algorithm*, *Nucl. Phys. B* **380** (1992) 665.

[5] M. Lüscher, *Properties and uses of the Wilson flow in lattice QCD*, *JHEP* **08** (2010) 071 [1006.4518].

[6] N.H. Christ, L.-C. Jin, C. Lehner, E. Lundstrum and N. Matsumoto, *An extended framework for the HMC in lattice gauge theory*, 2412.19904.