

Interpolation as a Means of Shift Selection for Multilevel Monte Carlo with Lattice Displacements

Travis Whyte,^{a,*} Andreas Stathopoulos,^a Eloy Romero^b and Kostas Orginos^{a,c}

^a*Department of Computer Science, William & Mary, Williamsburg, VA 23185, USA*

^b*Department of Physics, William & Mary, Williamsburg, VA 23185, USA*

^c*Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA*

E-mail: thwhyte@wm.edu

The calculation of disconnected diagram contributions to physical signals is a computationally expensive task in Lattice QCD. To extract the physical signal, the trace of the inverse Lattice Dirac operator, a large sparse matrix, must be stochastically estimated. Because the variance of the stochastic estimator is typically large, variance reduction techniques must be employed. Multilevel Monte Carlo (MLMC) methods reduce the variance of the trace estimator by utilizing a telescoping sequence of estimators. Frequency Splitting is one such method that uses a sequence of inverses of shifted operators to estimate the trace of the inverse of the lattice Dirac operator, however there is no a priori way to select the shifts that minimize the cost of the multilevel trace estimation. We present a sampling and interpolation scheme that is able to predict the variances associated with Frequency Splitting under displacements of the underlying space time lattice. The interpolation scheme is able to predict the variances to high accuracy and therefore choose shifts that correspond to an approximate minimum of the cost for the trace estimation. We show that Frequency Splitting with the chosen shifts displays significant speedups over multigrid deflation.

*The 39th International Symposium on Lattice Field Theory,
8th-13th August, 2022,
Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany*

*Speaker

1. Introduction

The evaluation of disconnected diagrams in Lattice QCD involves calculating the all-to-all propagator, or the trace of the inverse Lattice Dirac operator, D . Such a calculation is computationally infeasible on state of the art gauge ensembles, so the trace is estimated stochastically using the Hutchinson method [1]. To estimate the trace of a matrix, A , this method uses the estimator

$$t(A) = \frac{1}{s} \sum_{i=1}^{N_s} z_i^\dagger A z_i \quad (1)$$

for N_s number of random vectors z_i . It can be shown that the variance of the estimator is given by

$$\text{Var}[t(A)] = 2\|A\|_F^2 - 2 \sum_{i=1}^N |A_{ii}|^2. \quad (2)$$

Because the variance of the estimator depends on the Frobenius norm of the matrix A , the heaviest off-diagonal elements of the matrix will contribute the most to the variance of the estimator. In lattice QCD, the Lattice Dirac operator becomes ill conditioned at physical light quark mass and large lattice volumes, thus necessitating the development of methods to reduce the variance of the estimator $t(D^{-1})$. Probing [2] is one such variance reduction technique, which removes the heaviest elements of D^{-1} corresponding to small lattice distances. Probing was extended in [3, 4] and most recently extended to target the heaviest elements of a D^{-1} under a permutation Π_p corresponding to displacement of size p of the underlying space-time lattice [5]. Other examples of variance reduction techniques include deflation [6–8], low mode [9] and all mode [10] averaging, polynomial approximations [11].

A commonality of the previously discussed variance reduction techniques is that they are single level methods. A class of estimators exists that utilizes a telescoping sum of estimators, rather than a single estimator, to estimate the desired quantity and are thus called multilevel methods. Given a desired random variable X_L to be estimated and the sequence X_0, X_1, \dots, X_L , the unbiased estimator of X_L can be written as

$$\mathbb{E}[X_L] = \mathbb{E}[X_0] + \sum_{l=1}^L \mathbb{E}[X_l - X_{l-1}], \quad (3)$$

where the estimation is performed independently on each level l [13]. With an appropriate selection of the correlation between successive variables and number of levels, the total estimator variance, and thus the computational cost, can be reduced. Examples of multilevel methods can be found in [14] and [15]. An additional multilevel method, Frequency Splitting (FS) [16], which is the focus of this work, utilizes a telescoping series of inverses separated by a shift.

In this work, we consider the problem of selecting shifts that optimize the multilevel cost of FS for our target matrix $\Gamma\Omega_p D^{-1}$, where $\Omega_p = W\Pi_p$ with W the Wilson line and Π_p a permutation matrix corresponding to a $[0, 0, p, 0]$ -displacement that we assume is in the z direction of the lattice. Estimating the trace of $\Gamma\Omega_p D^{-1}$ is required to account for the disconnected contributions to GPDs [17] and is a computationally demanding task due to the large variance introduced by the displacements of the lattice. In Section 2 we review FS for our target problem and discuss multilevel Monte Carlo in the context of our application.

2. Frequency Splitting

Given a set of real increasing shifts $0 = \sigma_0 < \sigma_1 < \dots < \sigma_L$, the inverse lattice Dirac operator can be written as a telescoping series given by

$$D^{-1} = \sum_{l=0}^{L-1} \left((D + \sigma_l I)^{-1} - (D + \sigma_{l+1} I)^{-1} \right) + (D + \sigma_L I)^{-1}. \quad (4)$$

FS then makes use of the matrix identity $A^{-1}(A-B)B^{-1} = B^{-1} - A^{-1}$ to create products of inverses from differences of inverses, such that $(D + \sigma_l I)^{-1} - (D + \sigma_{l+1} I)^{-1} = (\sigma_{l+1} - \sigma_l)(D + \sigma_l I)^{-1}(D + \sigma_{l+1} I)^{-1}$. In lattice QCD, this was first introduced for twisted mass fermions under the name the "One End Trick" [18]. This can be generalized to our target matrix by left multiplying with $\Gamma\Omega_p$. Taking the trace, using the cyclic property and the commutation of the inverses yields

$$\mathbf{Tr}(\Gamma\Omega_p D^{-1}) = \sum_{l=0}^{L-1} (\sigma_{l+1} - \sigma_l) \mathbf{Tr} \left((D + \sigma_l I)^{-1} \Gamma\Omega_p (D + \sigma_{l+1} I)^{-1} \right) + \mathbf{Tr}(\Gamma\Omega_p (D + \sigma_L I)^{-1}). \quad (5)$$

In [16], the terms within the summation are known as the "split-even estimator". The insertion of $\Gamma\Omega_p$ in between the two inverses changes the singular spectra, which reduces the variance of the split even estimators. The total trace estimator of this quantity, is then given by

$$\begin{aligned} t(\Gamma\Omega_p D^{-1}) &= \sum_{l=0}^{L-1} \frac{(\sigma_{l+1} - \sigma_l)}{N_l} \sum_{s=0}^{N_l} z_{s,l}^\dagger (D + \sigma_l I)^{-1} \Gamma\Omega_p (D + \sigma_{l+1} I)^{-1} z_{s,l} \\ &+ \frac{1}{N_L} \sum_{s=0}^{N_L} z_{s,L}^\dagger \Gamma\Omega_p (D + \sigma_L I)^{-1} z_{s,L} \end{aligned} \quad (6)$$

with the trace estimator on each level l having variance as defined in Equation (2). If each term in Equation (6) is estimated independently, then the total multilevel computational cost is given by

$$C_{FS} = \epsilon^{-2} \left(\sum_{l=0}^L \sqrt{C_l V_l} \right)^2, \quad (7)$$

where ϵ^2 is the target variance, and C_l and V_l are the solver cost and variance of the level l , given by Equation (2), respectively [13]. The optimal number of samples for each level for a fixed set of shifts, Γ and Ω_p can be found with

$$N_l = \mu \sqrt{\frac{V_l}{C_l}}, \quad (8)$$

where the Lagrangian multiplier μ is given by $\mu = \epsilon^{-2} \left(\sum_{l=0}^L \sqrt{V_l C_l} \right)$. The total variance of the trace estimation is then given by

$$V_{FS} = \sum_{l=0}^L \frac{V_l}{N_l}. \quad (9)$$

Our goal is to find a set of shifts σ that approximately minimizes the multilevel cost of Equation (7). Because of the computational cost, we optimize only for a particular (Γ, Ω_p) combination, even though these shifts may not be optimal for other combinations. Since estimating the variance

for each level l involves solving linear systems of equations, finding the optimal shifts through sampling alone is prohibitively expensive. Next, we present our sampling and interpolation method that provides reliable predictions of the variances for many shifts with only a few samples at sampled shifts, which in turn can be used to minimize the multilevel cost.

3. Sampling and Interpolation

3.1 Sampling

While FS employs the use of trace estimators at fixed shifts, the variances V_l and V_L are generally a function of the shifts(s), Γ and Ω_p . Thus, to minimize the multilevel cost, it is necessary to estimate these quantities over a range of shifts. We explicitly denote the variances of each level as a function of shifts, α, β and ω , introducing also a more convenient \bar{V}_l for intermediate levels that does not include the squared difference of shifts and thus it is easier to interpolate,

$$V_l(\alpha, \beta, \Gamma, \Omega_p) = (\alpha - \beta)^2 \bar{V}_l(\alpha, \beta, \Gamma, \Omega_p) = (\alpha - \beta)^2 \text{Var}[t((D + \alpha I)^{-1} \Gamma \Omega_p (D + \beta I)^{-1})] \quad (10)$$

$$V_L(\omega, \Gamma, \Omega_p) = \text{Var}[t(\Gamma \Omega_p (D + \omega I)^{-1})]. \quad (11)$$

We also introduce three sets of shifts. The *sampling set*, \hat{s} , is a set of m real shifts such that $\hat{s}_0 = 0 < \hat{s}_1 < \dots < \hat{s}_{m-1}$. The sampling set is used to sample the values of Equations (10) and (11). The *evaluation set*, s , is a set of n real shifts with $s_0 = 0 < s_1 < \dots < s_{n-1} = \hat{s}_{m-1}$, on which we evaluate the interpolating polynomials to obtain a prediction of the variances \bar{V}_l and V_L at each pair of shifts, or shift, respectively. The evaluation set always contains the sampling set. The *optimal set*, σ , is the set of L shifts chosen from s such that minimizes the multilevel cost over all choices of s_i , i.e. $\sigma_0 = s_0 = 0 < \sigma_1 = s_{j_1} < \dots < \sigma_L = s_{j_L}$ with

$$\sigma = \underset{1 \leq j_1 < j_2 < \dots < j_L \leq n}{\arg \min} C_{FS}(s_0, s_{j_1}, \dots, s_{j_L}). \quad (12)$$

This optimization needs accurate values for \bar{V}_l and V_L for a wide range of shifts that it is too expensive to sample in its entirety. Instead, we interpolate a small number of sampled variances of $\bar{V}_l(\hat{s}_i, \hat{s}_j, \Gamma, \Omega_p)$ and $V_L(\hat{s}_i, \Gamma, \Omega_p)$ with $i = 0, \dots, m-1$ and $j = i, \dots, m-1$. The calculation of the variances \bar{V}_l and V_L through Equation (2) is not computationally tractable so we estimate it as

$$\text{Var}[t(A)] = \frac{1}{N_s} \sum_{k=1}^{N_s} (z_k^\dagger A z_k)^* (z_k^\dagger A z_k) - t(A)^* t(A), \quad (13)$$

where $*$ denotes the complex conjugate, with $A = (D + \hat{s}_i I)^{-1} \Gamma \Omega_p (D + \hat{s}_j I)^{-1}$ for $i = 0, \dots, m-1$ and $j = i, \dots, m-1$ in the case of Equation (10) and $A = \Gamma \Omega_p (D + \hat{s}_i I)^{-1}$ for $i = 0, \dots, m-1$ in the case of Equation (11). From Equation (13), we observe that estimating V_l and V_L involves solving systems of linear equations of the form $(D + \hat{s}_i I)x = z$ and $(D + \hat{s}_i I)^\dagger y = z$. Instead of a general random vector z , we use spin-diluted [19, 20] probing vectors. Due to the large number of inversions required, the number of z vectors used must be both practical and give a good estimate of the true variance. In [21], it was shown that accurate order of magnitude estimates for the Frobenius norm of a general matrix requires only a few noise vectors. We choose $N_s = 5$ random vectors to estimate \bar{V}_l and V_L .

Since the total number of inversions scales linearly with m , m must be chosen to keep the expense of sampling practical and provide enough data points for the interpolation to give reliable predictions. To obtain the correct trace, $\hat{s}_0 = 0$. For the last shift \hat{s}_{m-1} , we choose a large shift well within the interior of the singular spectrum of D . The behavior of $\bar{V}_l(\hat{s}_i, \hat{s}_j)$ as a function of \hat{s}_i , which is given by the red bursts in Figure 1, guides our choice of the interior sampling shifts. The variance displays a rapid decay for near zero shifts. In order to obtain accurate estimates of the variance in this region through interpolation, we choose \hat{s}_1 to be small. Due to the near exponential decay of the variance for shifts larger than 0.15, it is sufficient to choose three remaining shifts, equidistant in log space. We then choose the sampling set to be $\hat{s} = \{0.0, 0.05, 0.25, 0.50, 1.00\}$.

3.2 Interpolation

Having sampled the variances \bar{V}_l and V_L , we can then interpolate the sampled values to obtain a finer “shift space” in order to extract the shifts that approximately minimize Equation (7). In order to obtain numerically accurate results, we use piecewise cubic hermite interpolating polynomials (PCHIP) [22]. The use of PCHIP ensures continuity at the sampled points and a monotonic decrease of the predicted variances with increasing evaluation shifts. In the cases where only two sampled data points are available, which arises in the case for V_l , we use simple linear interpolation. Due to the observed near exponential decay of the variances \bar{V}_l, V_L for larger shifts, we choose to interpolate $\ln(\bar{V}_l)$ and $\ln(V_L)$. We illustrate the method of interpolation for the variance V_L . Having obtained the estimated variances at the sampled points, the interpolating polynomial q can be formed on the interval $[0, \hat{s}_{m-1}]$ using the points $\ln(V_L(\hat{s}_j))$. For details on the interpolating polynomials themselves, we defer the reader to [22]. The polynomial can be evaluated at an evaluation shift s_k and exponentiated to obtain a predicted variance $V_L(s_k) = e^{q(s_k)}$. Due to the two dimensional dependence of \bar{V}_l on the sampling set, the interpolation of \bar{V}_l must be done in successive interpolations along each dimension. Figure 1 (left) gives the algorithm for the two dimensional interpolation of \bar{V}_l and the right subgraphs illustrate the process.

Lastly, we describe the interpolation of the solver cost at each level. Due to critical slowing down at physical quark mass, we use multigrid preconditioned FGMRES as our linear solver. The cost at each level is then the number of outer iterations of FGMRES which is typically small and decreases in a step-wise function as the quark mass increases. Thus, we choose a linear interpolation between the solver costs at the sampled shifts to obtain C_l for each evaluation shift s_j .

4. Numerical Results

All calculations were performed on a lattice of size $32^3 \times 64$ with $m_q = -0.2390$. The sampling set and number of noise vectors are as discussed in Section 3. When $m_q + \hat{s}_j > 0$, we employ the use of the Generalized Hopping Parameter Expansion [23] for $V_L(\hat{s}_j)$, which is known to decrease the variance at large quark mass. Since we seek an accurate estimation of the trace at large displacements, we use the $p8k7$ probing vectors, i.e., corresponding to a lattice displacement of size $p = 8$ in the z -direction and distance of $k = 7$ coloring of the lattice, which produces 16 orthogonal vectors. The evaluation set is (in MATLAB notation) $\mathbf{s} = [\text{logspace}(-5, -2, 4) \text{ logspace}(\text{log}_{10}(0.011), 0, 76)]$.

- Input:** The sampled variances \bar{V}_l , sampled shifts, \hat{s} and evaluation shifts, s .
- Output:** Predicted variances V_l .
1. for $j = 1 : m - 1$
 2. $w = \ln(\bar{V}_l(\hat{s}_i, \hat{s}_j))$ with $i = 0, \dots, j$
 3. $q = \text{Interpolate}(\hat{s}_0, \dots, j, w)$
 4. for $k = 0 : n - 1$
 5. $\bar{V}_l(s_k, \hat{s}_j) = e^{q(s_k)}$
 6. end
 7. end
 8. $w = \ln(\bar{V}_l(\hat{s}_i, \hat{s}_j))$ with $i = j = 0, \dots, m - 1$
 9. $q = \text{Interpolate}(\hat{s}_0, \dots, m - 1, w)$
 10. for $k = 0 : n - 1$
 11. $\bar{V}_l(s_k, s_k) = e^{q(s_k)}$
 12. end
 13. for $i = 0 : n - 1$
 14. $w = \ln(\bar{V}_l(s_i, s_j))$ with $j = 0, \dots, m - 1$
 15. $q = \text{Interpolate}(\hat{s}_0, \dots, m - 1, w)$
 16. for $k = 0 : n - 1$
 17. $V_l(s_i, s_k) = (s_k - s_i)^2 e^{q(s_k)}$
 18. end
 19. end

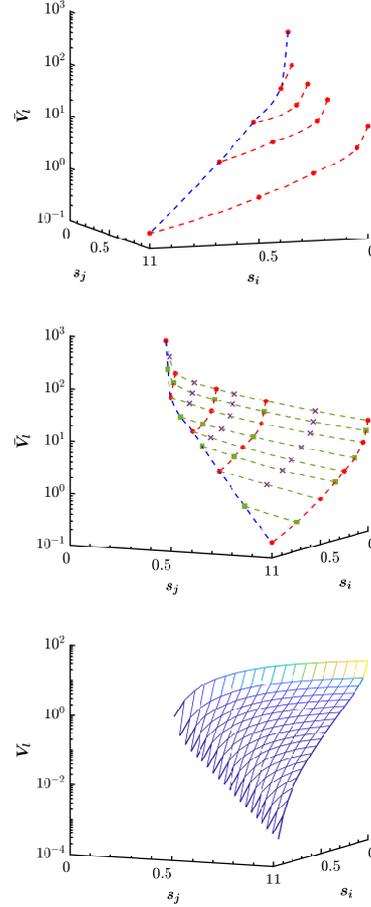


Figure 1: (Left) The algorithm for the 2D interpolation of \bar{V}_l . (Top right) The sampled variances $\bar{V}_l(\hat{s}_i, \hat{s}_j)$ (red bursts) for $\hat{s}_i, \hat{s}_j \in \{0.0, 0.05, 0.25, 0.5, 1.0\}$. The red dashed lines denote the polynomials generated at fixed \hat{s}_j at step 3 of the algorithm. The blue dashed line denotes the polynomial generated when $\hat{s}_i = \hat{s}_j$ at step 9 of the algorithm. (Right Center) Rotated 45° with respect to top subgraph and shown for an evaluation set of size $n = 9$. Green squares denote the evaluated points of polynomials generated at fixed \hat{s}_j and $\hat{s}_i = \hat{s}_j$ at steps (4 – 6) and (10 – 12). These evaluated points can then be used to form interpolating polynomials at fixed s_i , denoted by green dashes, at step 15. Purple crosses denote the evaluated points of the polynomials generated at fixed \hat{s}_i in step 17. (Right bottom) The final manifold of V_l for $n = 20$.

Due to high computational cost, we verify the accuracy of our interpolation scheme for $\Gamma = \gamma_3$ and $\gamma_5\gamma_4$, which are important in the calculation of GPDs, and Ω_p with $p = 0, \dots, 8$. Table (1) shows the predicted and estimated V_{total} and C_{FS} for $\Gamma = \gamma_3, \gamma_5\gamma_4$ and $p = 0, \dots, 8$. For each (Γ, Ω_p) pair, the shifts σ coming from that pairs' optimization were used in order to estimate the variance and multilevel cost and compare to the predicted variance and multilevel cost. We observed that the optimal cost was achieved at six shifts, while increasing the number of shifts further provided negligible benefits. We observe good agreement in our predicted values from the interpolation and the estimated values using the same set of shifts. For many of the displacements, the agreement is within one digit of accuracy.

Next we compare the variance reduction of FS using optimized shifts to that of multigrid

p	γ_3				$\gamma_5\gamma_4$			
	Predicted V_{total}	Estimated V_{total}	Predicted $C_{FS} (\times 10^5)$	Estimated $C_{FS} (\times 10^5)$	Predicted V_{total}	Estimated V_{total}	Predicted $C_{FS} (\times 10^5)$	Estimated $C_{FS} (\times 10^5)$
0	4.9504	5.2968	0.2921	0.3422	18.8176	21.6828	4.8145	5.7190
1	82.1364	99.4092	1.4293	1.7824	9.3321	9.7361	3.9570	4.4446
2	20.8019	23.7536	0.7521	0.8781	2.4040	2.5149	1.4491	1.6664
3	4.4729	4.6869	0.2371	0.2665	0.7998	0.8279	0.4895	0.5648
4	1.1335	1.1263	0.0680	0.0742	0.3110	0.3111	0.1812	0.2080
5	0.3491	0.3578	0.0215	0.0245	0.1509	0.1443	0.0796	0.0875
6	0.1469	0.1528	0.0084	0.0094	0.0581	0.0464	0.0408	0.0389
7	0.0826	0.0887	0.0041	0.0052	0.0356	0.0279	0.0253	0.0227
8	0.0410	0.0367	0.0030	0.0030	0.0320	0.0234	0.0217	0.0185

Table 1: In order to assess the accuracy of the interpolation, we introduce the quantity V_{total} , which is the total of the variances on each level l , $V_{total} = \sum_{l=0}^L V_l$. The predicted and estimated V_{total} and C_{FS} of the γ_3 and $\gamma_5\gamma_4$ operators for each displacement of size p .

deflation. The shifts used in FS, $\sigma = [0, 1e-05, 0.053, 0.146, 0.360, 0.618, 1.000]$, were obtained from the optimization using the $(\Gamma, \Omega_p) = (\gamma_3, \Omega_4)$ pair. This is because: 1) The error of the interpolation of the variance was smallest for this pair. 2) Optimizing the shifts at $p = 4$ provided additional speed up at low displacements compared to an optimization using large displacements, and 3) When shifts from an optimization with $\Gamma = \gamma_5\gamma_4$ were used, the variance of the trace estimator $\gamma_3\Omega_p D^{-1}$ was larger. In the case of multigrid deflation, we compute the smallest 400 singular triplets of the coarse grid operator. In addition, we use $p5k8$ probing vectors corresponding to a displacement-5, distance-8 coloring of the lattice, rather than the $p8k7$ coloring used in FS because the 16 probing vectors of $p8k7$ did not sufficiently complement multigrid deflation to reduce variance. The $p5k8$ coloring results in 256 probing vectors; a more challenging test for FS.

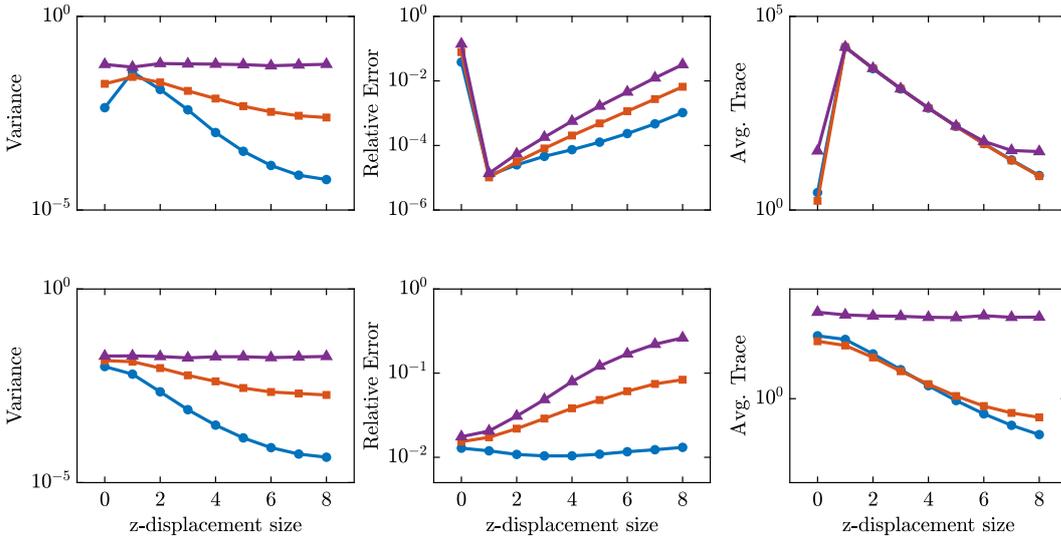


Figure 2: The estimated variance, relative error and average trace for $\Gamma = \gamma_3$ (top row) and $\Gamma = \gamma_5\gamma_4$ (bottom row) using FS + probing (circles), multigrid deflation + probing (squares) and random noise (triangles).

For each level l , we calculate the optimal number of samples from Equation (8) for a target variance of $\epsilon^2 = 0.001$. Then using Equation (9) we obtain an estimate of the total variance of FS. We can determine an equivalent number of noise vectors, and thus an equivalent computational cost, for multigrid deflation by computing the number of noise vectors through the following relation:

$$N_d = \frac{c \sum_{l=0}^L r_l N_l}{c_d r_d}, \quad (14)$$

where c is the number of colors used in FS, r_l is the number of outer iterations of FGMRES, N_l is the number of samples of the level l , respectively, and c_d and r_d are the number of colors and outer iterations for multigrid deflation. After estimating the variance of the trace estimator with multigrid deflation, we can estimate the variance achieved by multigrid deflation after N_d samples. Figure 2 displays the results of this comparison for $\Gamma = \gamma_3, \gamma_5\gamma_4$. FS has significantly smaller variance than multigrid deflation for both Γ 's at large displacements. The relative error is also similarly reduced, which is particularly important at large displacements due the trace becoming smaller in magnitude.

Lastly, we consider the effects of using a single set of shifts over multiple configurations. The cost of sampling is too expensive to optimize the shifts for every configuration within an ensemble. We therefore test the efficacy of a common set of shifts used for ten configurations separated by 100 MC steps. For each configuration we define the speedup over multigrid deflation as the ratio of estimated wallclock times for each method to reach the target variance of $\epsilon^2 = 0.001$ for the $(\Gamma, \Omega_p) = (\gamma_3, \Omega_4)$ pair. Table 2 shows consistent speedup across all ten configurations with little deviation. We conclude that the shifts coming from an optimization of one configuration can be used for other configurations within an ensemble without a penalty to performance.

Config. #	1	2	3	4	5	6	7	8	9	10
Est. Speedup	4.8436	5.4360	4.8494	4.5541	5.0838	3.4911	4.9955	4.5245	4.5861	5.7280

Table 2: Estimated speedup of FS over multigrid deflation for a target variance $\epsilon^2 = 0.001$ for (γ_3, Ω_4) .

5. Conclusion

Through a sampling and interpolation scheme, we have developed the ability to predict variances associated with the trace estimators of FS for different combinations of shifts. Thus, we can choose shifts that approximately minimize the cost of FS for a given set of probing vectors and (Γ, Ω_p) pair. FS with the chosen shifts outperforms multigrid deflation in terms of variance reduction at equal computational cost for large lattice displacements. Optimal shifts for one configuration can be reused for other configurations of the same ensemble without performance penalty.

6. Acknowledgements

This research was supported by the Exascale Computing Project (ECP), Project Number: 17-SC-20-SC, a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. Part of this work was performed using computing resources at William & Mary. The authors would like to thank David Richards for allowing compute time at JLab, and Robert Edwards for access to the gauge configurations used in this study. This research was performed with the use of the Chroma [24] library.

References

- [1] M.F. Hutchinson, *Comm. in Stat. - Sim. and Comp.* **18** 3 1059-1076 1989
- [2] Tang, Jok M. and Saad, Yousef, *N.L.A. with App.* **19** 3 485-501, 2012.
- [3] A. Stathopoulos, J. Laeuchli, K. Orginos, *SIAM J. on Sci. Comp.* **35** 5 S299-S322, 2013.
- [4] J. Laeuchli, A. Stathopoulos, *SIAM J. on Sci. Comp.* **42** 3 A1459-A1485, 2020.
- [5] H. Switzer, A. Stathopoulos, E. Romero, J. Laeuchli, K. Orginos. *arXiv:2106.01275 [hep-lat]*
- [6] A.S. Gambhir, and A. Stathopoulos and K. Orginos, *SIAM J. on Sci. Comp.* **39** 2 A532-A558, 2017
- [7] S. Baral, T. Whyte, W. Wilcox and R.B. Morgan, *Comp. Phys. Comm.* **241** 64-79, 2019.
- [8] E. Romero, A. Stathopoulos and K. Orginos, *J. of Comp. Phys.* **409** 109356, 2020.
- [9] T. DeGrand, Stefan Schaefer, *arXiv:hep-lat/0401011*
- [10] T. Blum, T. Izubuchi, E. Shintani, *arXiv:1208.4349 [hep-lat]*
- [11] Q. Liu, W. Wilcox, Ronald B. Morgan, *arXiv:1405.1763 [hep-lat]*
- [12] S. Collins, G. Bali, A. Schafer, *arXiv:0709.3217 [hep-lat]*
- [13] M. Giles, *Operations Research* **56** 3 607-617, 2008.
- [14] E. Hallman and D. Troester, *L.A. and App.* **638** 125-149, 2022.
- [15] A. Frommer, M.N. Khalil, G. Ramirez-Hidalgo, *arXiv:2108.11281 [math.NA]*
- [16] L. Giusti, T. Harris, A. Nada, S. Schaefer, *Eur. Phys. J. C* **79** 7 586, 2019.
- [17] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, F. Manigrasso, *Phys. Rev. Lett.* **126** 10, 2021.
- [18] P. Boucaud et. al., European Twisted Mass Collaboration, *Comp. Phys. Comm.* **179** 10 695-715, 2008.
- [19] W. Wilcox, *arXiv:hep-lat/9911013*
- [20] J. Foley et. al., *Comp. Phys. Comm.* **172** 3 145-162, 2005.
- [21] S. Gratton and D. Titley-Peloquin. *SIAM Journal on M.A. and App.* **39** 2 922-931, 2018.
- [22] F.N. Fritsch, R.E. Carlson, *SIAM J. on N.A. Analysis* **17** 2 238-246, 1980.
- [23] V. Gülpers, G. von Hippel, H. Wittig, *Phys. Rev. D.* **89**, 2014.
- [24] R.G. Edwards and B. Joo, *Nuclear Physics B - Proceedings Supplements* **140** 832-834, 2004.