Computational Science & Discovery

An anisotropic preconditioning for the Wilson fermion matrix on the lattice

COMPUTATIONAL
SCIENCE&DISCOVERY

An anisotropic preconditioning for the Wilson fermion matrix on the lattice

Bálint Joó[1], Robert G Edwards[1] and Michael Peardon[2]

[1] Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA
[2] School of Mathematics, Trinity College Dublin, Republic of Ireland
E-mail: bjoo@jlab.org

Received 15 October 2009, in final form 17 December 2009
Published 21 January 2010
Computational Science & Discovery **3** (2010) 015001 (20pp)
doi:10.1088/1749-4699/3/1/015001

Abstract. A preconditioning for the Wilson fermion matrix on the lattice is defined, which is particularly suited to the case when the temporal lattice spacing is much smaller than the spatial one. Details on the implementation of the scheme are given. The method is tested in numerical studies of quantum chromodynamics on anisotropic lattices.

Contents

1749-4699/10/015001+20$30.00

1. Introduction

Anisotropic discretizations [1–3] have proved to be an extremely useful tool for determining the energy spectrum of lattice field theories using the Monte Carlo method. Four-dimensional (4D) Euclidean space–time is discretized with separate grid spacings a_s and a_t for the spatial dimensions and the temporal direction with $\xi = a_s/a_t$. The spectrum of the theory is determined by measuring the correlation function between operators on the fields in a Monte Carlo calculation and performing statistical analysis to determine the rate of decay of these functions. This fall-off is related to the energies of eigenstates of the quantum-mechanical Hamiltonian. A fine resolution in the temporal direction is crucial for making an accurate determination of energies, particularly when considering massive or highly excited states. The advantage of the anisotropic lattice is that the rapid rise in numerical effort needed to reduce the lattice spacing is ameliorated by reducing the spacing in the time direction. The Hadron Spectrum collaboration is currently using these lattice regularizations in a large-scale programme [4–6] aimed at measuring the quantum chromodynamics (QCD) spectrum to high precision. Calculations by the collaboration include the full dynamics of the three lightest quark fields and, in these investigations, the dominant source of computational overhead is solving the linear system corresponding to quark propagation in a given gauge field background. Quark propagation is described by finding solutions to a lattice representation of the Euclidean-space Dirac equation:

$$(\not{D} + m)\psi = \eta, \tag{1}$$

where $(\not{D} + m)$ is the Dirac operator for a quark of mass m. The Hadron Spectrum collaboration has used anisotropic lattices with Wilson and Sheikholeslami-Wohlert (SW) [7] discretizations of the Dirac operator. These solutions, ψ, must be computed during application of the Markov process that generates the ensemble of gauge fields for Monte Carlo importance sampling as well as during the measurement phase. The vector space in which the solution is to be constructed is sufficiently large that direct methods are impractical and instead iterative solvers must be employed. While the anisotropic lattice gives substantial benefits in statistical precision, the disadvantage is that the inclusion of a fine discretization scale, a_t, increases the condition number of the linear system to be solved. This naturally leads to an increase in the number of iterations needed in the solver. A well-established means of improving efficiency is to precondition the problem.

A quick examination reveals that the dominant terms in the coefficient matrix of the linear system represent the discretization of the temporal derivative in the Dirac operator. In this work, operators that directly invert this term are constructed and used to form a preconditioner. Once the means of applying the inverse of the temporal term in the Wilson or SW operator has been defined, it is seen that further preconditioning using Schur or incomplete lower–upper (ILU) factorization can be applied. To test this idea, the condition numbers of temporally preconditioned operators are computed on several gauge field ensembles. This test uses quenched ensembles since they are substantially easier to generate with a range of different anisotropies than fully dynamical ones, and the issue of determining the parameters in the lattice action needed to give a particular physical anisotropy ξ is greatly simplified. The temporal preconditioned matrices to be inverted have condition numbers approximately 20 times smaller than their unpreconditioned counterparts and about three to four times smaller than the commonly used even–odd lattice Schur preconditioning when the anisotropy ranges from three to six. Taking into account numerical overheads, this translates to cost reductions of around 1.7–2.4. No advantage is found using temporal preconditioning in the *isotropic* case.

This paper is organized as follows. In section 2, we set up our notation, and describe in detail the basics of temporal preconditioning and how it may be combined with further even–odd preconditioning. In section 3, we detail our numerical investigations, including the tuning of our fermion anisotropy parameters and a presentation of condition numbers for various kinds of preconditioning schemes. We sum up and draw our conclusions in section 4.

2. Theory

The Wilson fermion matrix and its $\mathcal{O}(a)$-improved SW version can be defined on an anisotropic lattice with three coarse spatial directions with spacing a_s and a fine temporal spacing a_t. In practice, the anisotropy is introduced into the simulation by modifying the couplings in the lattice action that weight temporal and spatial components. Their relative weights are denoted by γ_{g} for terms in the gauge action and by γ_{f} for the fermions. These parameters are tuned carefully to ensure that the Lorentz invariance is restored in long-distance physics. If these parameters are determined using perturbative expansion, then at the tree level they are $\gamma_{\mathrm{f}} = \gamma_{\mathrm{g}} = \xi$. Since practical simulations will be performed close to the continuum limit, this relationship should hold to within approximately 25%. Following the notation of [4], we will denote the tuned values of these parameters as γ_{g}^* and γ_{f}^*, respectively.

With these definitions, the fermion matrix is defined as the sum of the following terms:

$$M = A + \mu\delta_{x,y} - D_{\mathrm{t}} - \frac{1}{\gamma_f} D_{\mathrm{s}}, \tag{2}$$

with

$$D_{\mathrm{t}} = \frac{1-\gamma_0}{2} U_0(x)\delta_{x+\hat{0},y} + \frac{1+\gamma_0}{2} U_0^\dagger(x-\hat{0})\delta_{x-\hat{0},y}, \tag{3}$$

$$D_{\mathrm{s}} = \sum_{i=1}^{3} \left(\frac{1-\gamma_i}{2} U_i(x)\delta_{x+\hat{i},y} + \frac{1+\gamma_i}{2} U_i^\dagger(x-\hat{i})\delta_{x-\hat{i},y} \right), \tag{4}$$

where γ_i for $i = 0, 1, 2, 3$ are appropriate spin matrices and μ is an anisotropic mass term corresponding to the bare mass m_0 defined as $\mu = a_t m_0 + 1 + (3/\gamma_{\mathrm{f}})$.

In the case of Wilson fermions, the term $A = 0$, while for the SW improved fermion matrix the mass term μ and the hopping terms D_{s} and D_{t} are unchanged, while the corresponding expression for A is

$$A = -\delta_{x,y} \left\{ \frac{c_{\mathrm{t}}}{2} \sum_{i=1}^{3} \sigma_{i0} F_{i0}(x) + \frac{c_{\mathrm{s}}}{2} \sum_{ij} \sigma_{ij} F_{ij}(x) \right\}, \tag{5}$$

where $\sigma_{ij} = \frac{1}{2}[\gamma_i, \gamma_j]$ is a commutator of the spin matrices and $F_{ij}(x)$ is the anti-Hermitian 'clover-leaf' term constructed from the plaquettes in the i, j-plane emanating from site x as defined in [4] and elsewhere.

The tree level tadpole improvement coefficients c_{s} and c_{t} are defined as

$$c_{\mathrm{s}} = \frac{1}{u_{\mathrm{s}}^3}\frac{1}{\gamma_{\mathrm{f}}}, \qquad c_{\mathrm{t}} = \frac{1}{2u_{\mathrm{s}}^2 u_{\mathrm{t}}} \left(\frac{\gamma_{\mathrm{g}}}{\gamma_{\mathrm{f}}} + \frac{1}{\xi} \right), \tag{6}$$

where u_{s} is the spatial tadpole factor computed from the spatial plaquettes of our gauge ensembles U_{ss} as

$$u_{\mathrm{s}} = \langle \tfrac{1}{3} U_{\mathrm{ss}} \rangle^{1/4}, \tag{7}$$

and where we have set $u_{\mathrm{t}} = 1$. The isotropic case is recovered when one sets $u_{\mathrm{s}} = u_{\mathrm{t}}$, $\gamma_{\mathrm{g}} = \gamma_{\mathrm{f}} = 1$ and so $c_{\mathrm{s}} = c_{\mathrm{t}}$ are the same tree level improved tadpole coefficients.

As in the isotropic case, these linear operators act on vectors in an $(N_{\mathrm{c}} \times N_\gamma \times N_{\mathrm{s}})$-dimensional space of complex fermion fields, where N_{c} is the number of colors in the gauge group, $N_\gamma = 4$ is the number of components in a spin-1/2 representation of the group of 4D Euclidean rotations and N_{s} is the number of sites on the 4D lattice.

The additional terms in the SW action are all diagonal in the spatial and temporal lattice coordinates. With the matrix expressed in this form, the largest terms in M are contained in A, D_{t} and μ, and the contribution from terms in D_{s} is smaller by a factor of the (bare fermion) anisotropy. This suggests that an efficient preconditioning for this matrix can be constructed if a way of applying the inverse of $A + \mu - D_{\mathrm{t}}$ is known. Note that this matrix is block diagonal in the spatial lattice coordinates, hence solving this problem requires being able to invert a one-dimensional lattice operator.

2.1. A temporal preconditioner

We now proceed to present the technique of temporal preconditioning. In the discussion below, we will initially focus on the Wilson case ($A = 0$) for simplicity. We will then comment upon the strategies available to deal with the full SW case. We begin by defining temporal spin-projection operators

$$P_{\pm} = \tfrac{1}{2}(1 \pm \gamma_0), \tag{8}$$

and a temporal hopping matrix at every spatial site on the lattice

$$T_{t,t'}(\vec{x}) = \mu\delta_{t,t'} - U_0(\vec{x}, t)\delta_{t+1,t'}, \quad t = 0 \ldots N_{\mathrm{t}} - 1, \tag{9}$$

where N_{t} is the number of lattice points in the temporal direction of the lattice. The operator T is gauge covariant with no spin structure. Further, T can have a variety of boundary conditions. We will consider primarily the cases where T is either periodic or anti-periodic where

$$T_{N_{\mathrm{t}}-1,N_{\mathrm{t}}-1}(\vec{x}) = \mu, \tag{10}$$

$$T_{0,N_{\mathrm{t}}-1}(\vec{x}) = \pm U_0(\vec{x}, N_{\mathrm{t}} - 1), \tag{11}$$

and the sign in equation (11) is positive or negative when the boundary conditions are periodic or anti-periodic, respectively.

We now define operators on all spin components,

$$C_{\mathrm{L}}^{-1} = P_+ + P_- T \quad \text{and} \quad C_{\mathrm{R}}^{-1} = P_- + P_+ T^{\dagger}, \tag{12}$$

where invertibility of these matrices has been assumed. With these definitions, the temporal hopping term can be expressed as

$$\mu - D_{\mathrm{t}} = C_{\mathrm{L}}^{-1} C_{\mathrm{R}}^{-1}. \tag{13}$$

Constructing C_{L}^{-1} and C_{R}^{-1} with the spin-projector structure given above ensures that they obey the relation

$$C_{\mathrm{L}}^{\dagger -1} = \gamma_5 C_{\mathrm{R}}^{-1} \gamma_5, \tag{14}$$

which will allow the construction of a preconditioned matrix that maintains γ_5 hermiticity. Equation (13) now shows how C_{L} and C_{R} will make a useful preconditioner for the Wilson fermion matrix on the anisotropic lattice. The fermion matrix can be written as

$$M = C_{\mathrm{L}}^{-1} C_{\mathrm{L}} M C_{\mathrm{R}} C_{\mathrm{R}}^{-1} = C_{\mathrm{L}}^{-1} \tilde{M} C_{\mathrm{R}}^{-1}, \tag{15}$$

with $\tilde{M} = C_{\mathrm{L}} M C_{\mathrm{R}}$. For Wilson fermions,

$$\tilde{M} = I - \frac{1}{\gamma_{\mathrm{f}}} C_{\mathrm{L}} D_{\mathrm{s}} C_{\mathrm{R}}, \tag{16}$$

and the preconditioned matrix is equal to the identity plus terms proportional to $1/\gamma_{\mathrm{f}}$ only.

The operation of $\tilde{M}$ on a fermion field requires the operation of $C_{\mathrm{L}} = (P_+ + P_- T)^{-1}$. Since P_+ and P_- define orthogonal projectors, this inverse can be rewritten as $C_{\mathrm{L}} = P_+ + P_- T^{-1}$ and the application of C_{L} is reduced to finding the inverse of T. If the lattice fields had Dirichlet boundary conditions, T (a lattice representation of the forward-difference operator) could be inverted easily by back-substitution, starting at the open boundary. Most lattice calculations use periodic or anti-periodic boundary conditions, however, and so back-substitution is insufficient. The Sherman–Morrison–Woodbury (SMW) [8–10] formula provides a simple means of inverting a matrix that differs in a small number of elements from another matrix whose inverse is computationally cheap to apply. The forward-difference operator with periodic boundary data, T, can be written in terms of the difference operator with open boundaries, T_0, and a correction term:

$$T = T_0 + V W^{\dagger}, \tag{17}$$

with

$$[T_0]_{t,t'}(\vec{x}) = \begin{cases} \mu\delta_{t,t'} - U_0(\vec{x},t)\delta_{t+1,t'}, & \text{if } t = 0\ldots N_\mathrm{t}-2 \\ \mu\delta_{t,t'}, & \text{if } t = N_\mathrm{t}-1, \end{cases} \tag{18}$$

and V and W are $N_\mathrm{c}N_\mathrm{t} \times N_\mathrm{c}$ column matrices where the only nonzero entries are in either the first or last sites:

$$V_t(\vec{x}) = -U_0(\vec{x},t)\delta_{t,N_\mathrm{t}-1} \tag{19}$$

and

$$W_t(\vec{x}) = \delta_{t,0}. \tag{20}$$

Since T, T_0, V and X are defined on each spatial site $\vec{x}$, we will suppress the spatial index in the subsequent discussion except where it is needed for clarity. With these definitions, $VW^\dagger$ is a rank N_c correction that adds the effects of the boundary condition back into the open-boundary-data operator. The SMW formula then gives an expression for the inverse of T defined in equation (17). Defining $X = T_0^{-1}V$ yields

$$T^{-1} = T_0^{-1} - X(I + W^\dagger X)^{-1} W^\dagger T_0^{-1}. \tag{21}$$

All that remains to be evaluated is $(I + W^\dagger X)^{-1}$, but note that this is a small (rank N_c) matrix at each spatial site whose inverse is straightforwardly computed. In practice, the algorithm proceeds as follows.

(i) Prior to use, the preconditioner is initialized. On each spatial site $\vec{x}$, the expression

   (a) $X = T_0^{-1}V$ is computed by back-substitution and
   (b) $\Lambda = (I + W^\dagger X)^{-1}$ is subsequently computed.

   All these results are stored. Λ requires storage of just $N_\mathrm{c} \times N_\mathrm{c}$ complex numbers per spatial site, while X requires $N_\mathrm{c} \times N_\mathrm{c}N_\mathrm{t}$. This is a smaller storage requirement than needed for a single fermion field. Also, we note that one can immediately compute $X\Lambda$ at this point which also requires storage of $N_\mathrm{c} \times N_\mathrm{c}N_\mathrm{t}$ per spatial site but which may overwrite the original X.

(ii) Given a particular right-hand side η, computing $\psi = T^{-1}\eta$ requires first evaluating

   (a) $\chi = T_0^{-1}\eta$ by back-substitution, then
   (b) $q = \Lambda W^\dagger \chi$. Note that $W^\dagger\chi$ is just the N_c-component vector on time-slice $t = 0$ of vector χ and so evaluation of $W^\dagger\chi$ is computationally trivial. Finally, the solution is formed:
   (c) $\psi = \chi - Xq$.

   The back-substitution process for X can formally be carried out analytically. For the case of T one has

$$X(\vec{x}, N_\mathrm{t}-1) = -\frac{1}{\mu}U_t(\vec{x}, N_\mathrm{t}-1), \tag{22}$$

$$X(\vec{x},t) = -\frac{1}{\mu^{N_t-t}} \prod_{j=N_t-t}^{N_t-1} U_t(\vec{x},j) \quad \text{for } 0 \leqslant t < N_\mathrm{t}-1, \tag{23}$$

and successive terms in X are suppressed by powers of the mass term μ, and the matrix product forms a series, which for $X(\vec{x},0)$ is the Polyakov loop. In principle, for large enough N_t, one could find some k such that for $0 \leqslant t < k$ one has $X(\vec{x},i) \approx 0$ numerically, and one may then save some numerical effort by just setting those values of $X(\vec{x},t) = 0$ and not evaluating matrix products using them, but setting them to zero also. Since the link matrices in the components of X are SU(3), one knows that their product is also SU(3), and hence one can find the norm of $X(\vec{x},i)$ as

$$\|X(\vec{x},t)\| = \sqrt{3}\mu^{-(N_\mathrm{t}-t)}, \tag{24}$$

where the factor of $\sqrt{3}$ comes from the SU(3) nature of the link matrices. We will refer to the cutting off of the computation of $X(\vec{x}, t)$ for sufficiently small values of t as the *cut-off trick*. Caution should be used, however, to ensure that there is no impact on the precision of the final result.

The inverse of $T^{\dagger}$, required for the operation of C_{R}, is formed in the same way, using appropriate redefinitions of X, W and V and using forward-substitution. We note that in step i(b) above, only an $N_{\mathrm{c}} \times N_{\mathrm{c}}$ complex matrix needs to be inverted per site. The cut-off trick also works, but now the terms are least suppressed at $t = 0$ (open end of the forward substitution) and most suppressed at $t = N_{\mathrm{t}} - 1$, building up to the Hermitian conjugate of the Polyakov loop for $X(\vec{x}, N_{\mathrm{t}} - 1)$, and the $N_{\mathrm{t}} - t$ term in equation (24) needs to be replaced with t.

Let us now comment on the case for SW fermions. Proceeding as above, the entire procedure is valid, but the preconditioned matrix changes to

$$\tilde{M} = C_{\mathrm{L}} M C_{\mathrm{R}} = I + C_{\mathrm{L}} \left(A - \frac{1}{\gamma_{\mathrm{f}}} D_{\mathrm{s}} \right) C_{\mathrm{R}}. \tag{25}$$

It then becomes tempting to extend the definition of C_{L}^{-1} and C_{R}^{-1} in such a way that

$$A + \mu - D_{\mathrm{t}} = C_{\mathrm{L}}^{-1} C_{\mathrm{R}}^{-1}, \tag{26}$$

so that the preconditioned matrix would maintain its original $I - \frac{1}{\gamma_{\mathrm{f}}} C_{\mathrm{L}} D_{\mathrm{s}} C_{\mathrm{R}}$ form, even in the case of a general, nonzero SW term, i.e. $A \neq 0$. The practical difficulty with this approach is that the $\sigma_{\mu\nu}$ terms in A couple all spin components and the forward- and backward-difference operators in D_{t} cease to be directly separable. Correspondingly, the construction of suitable C_{L} and C_{R} terms would require the inversion of a block-tridiagonal matrix, with $(N_{\gamma} N_{\mathrm{c}} \times N_{\gamma} N_{\mathrm{c}})$-sized blocks rather than just $N_{\mathrm{c}} \times N_{\mathrm{c}}$. Further, the back/forward substitutions fill in these blocks destroying the sitewise block-diagonal structure of the SW matrix, and hence the inversions of the diagonal blocks would need dense inversions of the full $N_{\gamma} N_{\mathrm{c}}$-dimensional sub-blocks. We will refer to this approach as *full SW temporal preconditioning*. The details of this approach are discussed in appendix A.

Nonetheless, even if one just uses the same C_{L} and C_{R} as for the Wilson action and suffers the contamination from the SW term in the preconditioned matrix $\tilde{M}$ in equation (25), it can be seen that the D_{t} term is still inverted and that the D_{s} terms are suppressed by a factor of $1/\gamma_{\mathrm{f}}$. Hence one can expect that this form of preconditioning is still more effective than using the unpreconditioned operator. In what follows we will refer to this approach of using the C_{L} and C_{R} preconditioners from the case of the Wilson fermion matrix to precondition the SW operator as *partial SW temporal preconditioning*.

2.2. *Combining the temporal preconditioner with other schemes*

The usual isotropic Wilson and SW operators are efficiently preconditioned by considering a Schur decomposition after first ordering lattice sites according to their 4D coordinate parity, $p_4(x) = (-1)^{x_0+x_1+x_2+x_3}$. One has

$$\begin{aligned} M &= \begin{pmatrix} M^{\mathrm{ee}} & M^{\mathrm{eo}} \\ M^{\mathrm{oe}} & M^{\mathrm{oo}} \end{pmatrix} \\ &= \begin{pmatrix} I^{\mathrm{ee}} & 0 \\ M^{\mathrm{oe}}[M^{\mathrm{ee}}]^{-1} & I^{\mathrm{oo}} \end{pmatrix} \begin{pmatrix} M^{\mathrm{ee}} & 0 \\ 0 & M^{\mathrm{oo}} - M^{\mathrm{oe}}[M^{\mathrm{ee}}]^{-1}M^{\mathrm{eo}} \end{pmatrix} \begin{pmatrix} I^{\mathrm{ee}} & [M^{\mathrm{ee}}]^{-1}M^{\mathrm{eo}} \\ 0 & M^{\mathrm{oo}} \end{pmatrix}, \end{aligned} \tag{27}$$

with $M^{\mathrm{ee(oo)}} = A^{\mathrm{ee(oo)}} + \mu$ and $M^{\mathrm{eo}} = -D_w^{\mathrm{eo}}$, $D_w^{\mathrm{eo}} = \frac{1}{\gamma_{\mathrm{f}}} D_{\mathrm{s}} + D_{\mathrm{t}}$ being the 4D Wilson Dslash operator, and the *Schur preconditioned* matrix is

$$\tilde{M} = \begin{pmatrix} M^{\mathrm{ee}} & 0 \\ 0 & M^{\mathrm{oo}} - M^{\mathrm{oe}}[M^{\mathrm{ee}}]^{-1}M^{\mathrm{eo}} \end{pmatrix}. \tag{28}$$

This kind of Schur preconditioning with a p_4 ordering is the standard even–odd preconditioning method in use for Wilson and SW fermions, and we shall refer to it as *4D Schur preconditioning* from now on.

This idea can be combined with the temporal preconditioner with a small modification; instead of ordering lattice sites by a 4D parity, a three-dimensional (3D) equivalent is used, $p_3(x) = (-1)^{x_1+x_2+x_3}$. The preconditioned matrix $\tilde{M}$ retains its form in terms of $M^{\mathrm{ee(oo)}}$ and $M^{\mathrm{eo(oe)}}$; however, these now change their meaning slightly, since with this ordering, the operators A and D_{t} connect lattice sites with the same p_3 while the spatial hopping matrix couples sites with opposite p_3 and

$$M^{\mathrm{ee(oo)}} = A^{\mathrm{ee(oo)}} + \mu - D_{\mathrm{t}}^{\mathrm{ee(oo)}}, \quad M^{\mathrm{eo(oe)}} = -\frac{1}{\gamma_{\mathrm{f}}} D_{\mathrm{s}}^{\mathrm{eo(oe)}}. \tag{29}$$

For the Wilson action, the inverse of the block matrix, $M^{\mathrm{ee}} = \mu - D_{\mathrm{t}}^{\mathrm{ee}}$, is formed using the method described in the previous section. For full temporal preconditioning in the clover case, one would need to form the inverse of $M^{\mathrm{ee}} = A^{\mathrm{ee}} + \mu - D_{\mathrm{t}}^{\mathrm{ee}}$, the difficulties of which have already been discussed, and we can refer to the result as *temporal preconditioning combined with 3D Schur preconditioning*. Alternatively, one can proceed with partial preconditioning for the clover case, using the p_3 ordering preconditioner appropriate for the Wilson action. In this case, we can refer to the result as *temporal preconditioning combined with 3D ILU preconditioning*.

First, we define the action of the left temporal preconditioner on the even three-parity sub-lattice to be $C_{\mathrm{L}}^{\mathrm{e}}$ and define correspondingly the right preconditioner and both their odd sub-lattice counterparts to be $C_{\mathrm{R}}^{\mathrm{e}}$, $C_{\mathrm{L}}^{\mathrm{o}}$, $C_{\mathrm{R}}^{\mathrm{o}}$. We also introduce the notation that for some generic term K the corresponding term $\bar{K}$ is defined as $\bar{K} = C_{\mathrm{L}} K C_{\mathrm{R}}$. Hence,

$$\bar{D}_{\mathrm{s}}^{\mathrm{eo}} = C_{\mathrm{L}}^{\mathrm{e}} D_{\mathrm{s}}^{\mathrm{eo}} C_{\mathrm{R}}^{\mathrm{o}}. \tag{30}$$

To keep the discussion below general, we will also define the operator

$$\bar{Q} = C_{\mathrm{L}} \left(A + \mu - D_{\mathrm{t}}\right) C_{\mathrm{R}}. \tag{31}$$

In the case of Wilson fermions, $A = 0$ and so $\bar{Q} = I$. For SW fermions with full SW preconditioning one also has $\bar{Q} = 1$, while with partial SW preconditioning,

$$\bar{Q} = 1 + C_{\mathrm{L}} A C_{\mathrm{R}} = 1 + \bar{A}. \tag{32}$$

$\bar{Q}$ is diagonal in terms of p_3 even–odd indices, and we may refer to its even–even (odd–odd) sub-blocks $\bar{Q}^{\mathrm{ee(oo)}}$ using $C_{\mathrm{L}}^{\mathrm{e(o)}}$, $C^{\mathrm{e(o)}}$ and $A^{\mathrm{ee(oo)}}$ as needed.

With these expressions, the even–odd preconditioners for the Wilson matrix become

$$S_{\mathrm{L}} = \begin{pmatrix} C_{\mathrm{L}}^{\mathrm{e}} & 0 \\ \frac{1}{\gamma_{\mathrm{f}}} \bar{D}_{\mathrm{s}}^{\mathrm{oe}} C_{\mathrm{L}}^{\mathrm{e}} & C_{\mathrm{L}}^{\mathrm{o}} \end{pmatrix} \quad \text{and} \quad S_{\mathrm{R}} = \begin{pmatrix} C_{\mathrm{R}}^{\mathrm{e}} & \frac{1}{\gamma_{\mathrm{f}}} C_{\mathrm{R}}^{\mathrm{e}} \bar{D}_{\mathrm{s}}^{\mathrm{eo}} \\ 0 & C_{\mathrm{R}}^{\mathrm{o}} \end{pmatrix}, \tag{33}$$

which gives

$$\tilde{M}_3 = S_{\mathrm{L}} M S_{\mathrm{R}} \tag{34}$$

$$= \begin{pmatrix} \bar{Q}^{\mathrm{ee}} & -\frac{1}{\gamma_{\mathrm{f}}}[1 - \bar{Q}^{\mathrm{ee}}]\bar{D}_{\mathrm{s}}^{\mathrm{eo}} \\ -\frac{1}{\gamma_{\mathrm{f}}} \bar{D}_{\mathrm{s}}^{\mathrm{oe}}[1 - \bar{Q}^{\mathrm{ee}}] & \bar{Q}^{\mathrm{oo}} - \frac{1}{\gamma_{\mathrm{f}}^2} \bar{D}_{\mathrm{s}}^{\mathrm{oe}}[2 - \bar{Q}^{\mathrm{ee}}]\bar{D}_{\mathrm{s}}^{\mathrm{eo}} \end{pmatrix}. \tag{35}$$

If $\bar{Q} = 1$, as is the case of Wilson fermions or clover fermions with full temporal preconditioning, this matrix reduces to

$$\tilde{M}_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 - \frac{1}{\gamma_{\mathrm{f}}^2} \bar{D}_{\mathrm{s}}^{\mathrm{oe}} \bar{D}_{\mathrm{s}}^{\mathrm{eo}} \end{pmatrix}, \tag{36}$$

so that the preconditioned matrix differs from the identity only by terms proportional to $1/\gamma_{\mathrm{f}}^2$. In the case of partial SW preconditioning, where $\bar{Q} = 1 + \bar{A}$, the preconditioned matrix is

$$\tilde{M}_3 = \begin{pmatrix} 1 + \bar{A}^{\mathrm{ee}} & \frac{1}{\gamma_{\mathrm{f}}} \bar{A}^{\mathrm{ee}} \bar{D}_{\mathrm{s}}^{\mathrm{eo}} \\ \frac{1}{\gamma_{\mathrm{f}}} \bar{D}_{\mathrm{s}}^{\mathrm{oe}} \bar{A}^{\mathrm{ee}} & 1 + \bar{A}^{\mathrm{oo}} - \frac{1}{\gamma_{\mathrm{f}}^2} \bar{D}_{\mathrm{s}}^{\mathrm{oe}} [1 - \bar{A}^{\mathrm{ee}}] \bar{D}_{\mathrm{s}}^{\mathrm{eo}} \end{pmatrix}. \tag{37}$$

We can see from equation (37) that in contrast to full preconditioning in equation (36), we now have nonzero off diagonal elements (in even–odd space) that are only suppressed by γ_{f}. Furthermore, the diagonal elements contain components proportional to a_{t} in the clover terms $\bar{A}$. These terms can counter potential suppression by γ_{f}^2 in the odd–odd checkerboarded term of $\tilde{M}_3$.

To complete this discussion, we note that effective use of the preconditioner requires the inverses S_{L}^{-1} and S_{R}^{-1} to be applied so that solutions with unpreconditioned matrices may be obtained. Using the definitions of $C_{\mathrm{L(R)}}^{-1}$ restricted to the even and odd sites, respectively, we have

$$S_{\mathrm{L}}^{-1} = \begin{pmatrix} \left(C_{\mathrm{L}}^{\mathrm{e}}\right)^{-1} & -\frac{1}{\gamma_{\mathrm{f}}} \bar{D}^{\mathrm{eo}} \left(C_{\mathrm{L}}^{\mathrm{o}}\right)^{-1} \\ 0 & \left(C_{\mathrm{L}}^{\mathrm{o}}\right)^{-1} \end{pmatrix}$$

and

$$S_{\mathrm{R}}^{-1} = \begin{pmatrix} \left(C_{\mathrm{R}}^{\mathrm{e}}\right)^{-1} & 0 \\ -\frac{1}{\gamma_{\mathrm{f}}} \left(C_{\mathrm{R}}^{\mathrm{e}}\right)^{-1} \bar{D}_{\mathrm{s}}^{\mathrm{eo}} & \left(C_{\mathrm{R}}^{\mathrm{o}}\right)^{-1} \end{pmatrix}. \tag{38}$$

2.3. Numerical cost of the ILU scheme

We consider the partial temporally preconditioned scheme combined with ILU even–odd preconditioning to be potentially the most attractive, since it is simple in terms of implementation and is equivalent to the full 3D Schur preconditioned scheme in the case of Wilson fermions. However, applying the preconditioners does incur some numerical overhead. The overhead depends to some degree on details of the implementation of the method. We will consider two implementations below.

First we consider the naive implementation of the method, with no gauge fixing and in a spinor basis where γ_0 is not diagonal. This could be the case in a general code, using a chiral spin-basis such as the Chroma software system [11]. Neglecting the cost of preparing the sources and recovering the solutions (using S_{L}^{-1} and S_{R}^{-1}), we can compare costs of the usual 4D Schur preconditioning and the partially temporally preconditioned ILU scheme, which we will denote as $\mathcal{C}(\tilde{M}_{\mathrm{4DSchur}})$ and $\mathcal{C}(\tilde{M}_{\mathrm{ILU}})$, respectively, by counting the floating point operations (FLOPs) in the respective preconditioned linear operators.

Relegating the actual counting of FLOPs to appendix B, we merely state here that the ratio of floating point costs,

$$R = \frac{\mathcal{C}(\tilde{M}_{\mathrm{ILU}})}{\mathcal{C}(\tilde{M}_{\mathrm{4DSchur}})}, \tag{39}$$

is

$$R^{\mathrm{W}} = \frac{3432 N_{\mathrm{t}} - 576}{2668 N_{\mathrm{t}}} \approx 1.286 - \frac{0.216}{N_{\mathrm{t}}} \tag{40}$$

for Wilson fermions and

$$R^{\mathrm{C}} = \frac{6012 N_{\mathrm{t}} - 1152}{3732 N_{\mathrm{t}}} \approx 1.611 - \frac{0.309}{N_{\mathrm{t}}} \tag{41}$$

for clover fermions, respectively, resulting typically in about a 29% overhead for Wilson and a 61% overhead for clover from the ILU scheme in terms of FLOPs as compared to the standard 4D Schur even–odd scheme. This must be matched by the gain in terms of condition number from the preconditioner for it to remain competitive.

Concurrent with writing this paper, some clever optimization techniques were brought to our attention by the authors of [12, 13] arising from work with general purpose graphics processing units (GPGPUs). These techniques save both memory bandwidth and FLOPs. The first technique we consider is to fix the gauge prior to the inversion process, to the axial gauge (temporal gauge). The effect of this operation is that all the links in the time direction, that are not on the temporal boundary are transformed to the unit matrix: $U_0(\vec{x}, t) = 1$ for $t = 0 \dots N_\mathrm{t} - 2$. This can save FLOPs in the back(forward) substitutions with T_0, where in each step one can save an SU(3) matrix-color vector multiply. Further, one can save on memory requirements since the block vector X is simplified to

$$X(\vec{x}, t) = \frac{1}{\mu^{N_\mathrm{t}-t}} U_0(\vec{x}, N_\mathrm{t} - 1), \tag{42}$$

with the other terms in the partial Polyakov loops now being the identity. Correspondingly, instead of storing all of X, one can easily compute any component of it from the boundary link matrix that one stores anyway. Fixing to the axial gauge is a straightforward operation that can be amortized over either one and especially over several solves.

The second trick that can prove useful is to employ the Dirac–Pauli spin basis in which γ_0 is diagonal. This simplifies the projector operators P_+ and P_- so that they select the top or bottom two spin components of four spinors respectively. This can save FLOPs on spinor reconstruction (which now no longer needs to be done in the time direction) but also when one has a sum of the form

$$S = P_+\psi + P_-\chi, \tag{43}$$

one has no arithmetic to perform, since the spin components filtered by the projectors can be written directly into the correct components of S without requiring any addition. We enumerate in explicit detail the savings from these two implementation techniques in appendix C. It is shown there that employing both of these techniques can save roughly 22–25% in terms of FLOPs over the naive implementation.

When compared to the 4D Schur preconditioned scheme, which also benefits from these improvements, we find that the techniques result in relative overhead ratios of

$$R^\mathrm{W} = \frac{2664N_\mathrm{t} - 48}{2044N_\mathrm{t}} \approx 1.30 - \frac{0.023}{N_\mathrm{t}} \tag{44}$$

for Wilson fermions and

$$R^\mathrm{C} = \frac{4476N_\mathrm{t} - 96}{3108N_\mathrm{t}} \approx 1.440 - \frac{0.031}{N_\mathrm{t}} \tag{45}$$

for clover fermions. In particular, the relative overhead for the clover operator appears to be substantially reduced compared to the naive implementation (44% as opposed to the previous 61%). The foregoing discussion does not make use of the cut-off trick, which can be used to further reduce the floating point costs of the preconditioned operators as discussed earlier.

3. Numerical investigation

3.1. Strategy and choice of parameters

Wishing to investigate the efficacy of the preconditioning strategies as functions of both lattice anisotropy and quark mass in as realistic a setting as possible, we have opted to measure the condition numbers of the various operators in three quenched ensembles. These ensembles were chosen to have target (renormalized) anisotropies of $\xi = 1.0, 3.0$ and 6.0, respectively, thus ranging from the fully isotropic to the highly anisotropic.

Table 1. Parameters of the quenched ensembles used in this study. In the first column ξ^{T} refers to the desired target anisotropies, with the final bare gauge anisotropies γ_{g}^{*} being shown in the fourth column. Apart from (†) the temporal extent grows from $N_{\mathrm{t}} = 16$ with the anisotropy. (†) was used merely for checking the pion mass at the isotropic parameter set since the $N_{\mathrm{t}} = 16$ case was too short for measuring the pion mass. We also show our measurements for the spatial tadpole coefficient u_{s}.

ξ^{T}	β	$V\ (N_{\mathrm{s}}^3 \times N_{\mathrm{t}})$	γ_{g}^{*}	u_{s}
1	6.0	$16^3 \times 16$	1	0.8780
1	6.0	$16^3 \times 48$(†)	1	0.8780
3	6.1	$16^3 \times 48$	2.464	0.8279
6	6.1	$16^3 \times 96$	4.7172	0.8195

We chose a wide range of quark masses, to give pion masses in the range of about 450–750 MeV. We note that this necessitated tuning the fermion anisotropies γ_{f} so as to make the renormalized anisotropies ξ (as fixed by the pion dispersion relation) the same as our target anisotropies ξ, a subject we will discuss in more detail below.

During our study, we opted to keep the physical temporal extent of the lattice fixed in the time direction. This approach means that as we increased the anisotropy, we likewise increased the number of lattice points in the time direction. We used $N_{\mathrm{t}} = 16$, 48 and 96 for the anisotropies of $\xi = 1.0$, 3 and 6, respectively. This increase of the temporal resolution may have an effect on the condition numbers of our operators. We felt, however, that this is typically the approach one would use in a real calculation, rather than keeping the temporal extent fixed, and that we should absorb this effect in our efficiency estimates, in order to give a realistic measure of the performance of the preconditioning.

3.2. Code and computers

We coded the temporal preconditioned Wilson–clover operators in the Chroma [11] software system. We implemented both 3D ILU and 3D Schur even–odd preconditionings in space, in combination with the temporal preconditioning. In the case of the 3D Schur even–odd preconditioning, we used an inner conjugate gradients solve to invert M^{ee} in the Schur complement. We also used the unpreconditioned and 4D Schur even–odd preconditioned operators already present in the Chroma suite to measure reference results. Our implementations used the naive implementation technique discussed in section 2.3. In order to tune the fermion anisotropies, we carried out some hadron spectroscopy calculations, in particular the measurement of the pseudoscalar correlation functions at various momenta. In order to measure the condition numbers of the square operators, we used the Ritz-minimization technique of Bunk [14]. Both these sets of measurements are standard within the Chroma distribution. Fitting of our spectroscopy results used the so-called 4H code developed by UKQCD. Our calculations were carried out on the Jefferson Lab 6n and 7n clusters.

3.3. Selecting the gauge action parameters

We chose our isotropic reference case to be a quenched dataset with the Wilson gauge action at $\beta = 6.0$, as it is well known in the literature to have a lattice spacing of 0.1 fm [15]. In the anisotropic case, using $\beta = 6.1$ has approximately the same spatial lattice spacing at $\xi = 3$ and 6 [16]. We chose the bare gauge anisotropies from the formula suggested by Klassen [16]. Our gauge production parameters are summarized in table 1.

3.4. Tuning the fermion parameters

In this study, we opted to use clover fermions, with tree-level tadpole improved clover coefficients as defined in (6). In the anisotropic cases, we needed to tune the fermion anisotropy as well as our quark masses to fall within our desired range.

Our tuning exercise then consisted of choosing trial values, of γ_{f} for a selection of values for m_0, and computing the pion dispersion relations for each $(m_0, \gamma_{\mathrm{f}})$ pair. Since anisotropic tuning is not the main subject

Table 2. At each of our target anisotropies ξ^{T}, we show the bare values of m_0, our tuned bare fermion anisotropy γ_{f}^{*} and the resulting measured renormalized anisotropies ξ as determined from the pion dispersion relation. We also show the resulting pion masses in lattice units ($E(\vec{p}=0)$) and in physical units (m_π). The $\xi = 1.0$ results were determined on the $16^3 \times 48$ lattice.

ξ^{T}	m_0	γ_{f}^{*}	ξ	$E(\vec{p}=0)$	m_π (MeV)	No. configuration used
1	−0.359	1	1	0.383(3)	766(6)	100
1	−0.379	1	1	0.296(2)	597(4)	100
1	−0.392	1	1	0.225(2)	450(4)	59
3	−0.13	2.95	3.06(4)	0.105(2)	641(15)	46
3	−0.132	2.96	3.08(4)	0.097(1)	597(12)	46
3	−0.135	2.96	3.03(4)	0.082(2)	498(15)	46
6	−0.058	5.43	5.99(10)	0.0578(9)	693(16)	50
6	−0.061	5.63	5.96(10)	0.042(1)	504(17)	41

of our paper, we were content to do this very roughly and were satisfied by a renormalized anisotropy within 10% of our target. To compute the dispersion relation, we extracted the ground state energies (E_0) of the pion for several initial momenta $\vec{p}$, by fitting the pion correlation function,

$$C(t,\vec{p}) = \sum_{\vec{x}} \mathrm{e}^{\mathrm{i}\vec{x}\cdot\vec{p}} \langle \mathcal{O}^{\dagger}(\vec{x},t)\mathcal{O}(\vec{0},0)\rangle \overset{t\to\infty}{\to} 2A\mathrm{e}^{-E_0(N_{\mathrm{t}}/2)} \cosh\left(E_0\left(\frac{N_{\mathrm{t}}}{2}-t\right)\right), \tag{46}$$

where $\mathcal{O}$ is a suitable pion interpolating operator,

$$\mathcal{O}(\vec{x},t) = \bar{\psi}(\vec{x},t)\Gamma\psi(\vec{x},t). \tag{47}$$

We used spatial momenta ranging in magnitude from $|\vec{p}^2| = 0$ to $|\vec{p}^2| = (3/2\pi N_{\mathrm{s}})^2$. We averaged the correlation function over the momenta, which resulted in equal values of $|\vec{p}^2|$. We used the interpolating operator $\Gamma = \gamma_5$ for the zero momentum fits, whereas for the finite momentum fits we used $\Gamma = \gamma_4\gamma_5$ as we found the signal to be cleaner.

Once E_0 was determined for all values of $|\vec{p}|^2$, we fitted them to the dispersion relation formula

$$E^2 = \frac{1}{\xi^2}p^2 + \hat{m}^2, \tag{48}$$

where $\hat{m} = E_0(|\vec{p}|^2 = 0)$, by a straight-line fit, to extract ξ. To compute our estimate for the pion mass, we then computed the mass in units of the spatial lattice spacing, $a_{\mathrm{s}}m_{\mathrm{latt}} = \xi\hat{m}$, and converted this number to physical units assuming $a_{\mathrm{s}} = 0.1$ fm.

The correlation functions themselves were constructed using Gaussian gauge invariant source smearing [17] and stout link smearing [18]. No smearing was performed at the sink. Our tuning calculations were carried out using 40–100 configurations for each (m_0, γ_{f}) pair. We used the bootstrap method to estimate our errors on the masses and fitted anisotropies with 200 bootstrap samples in each case. When estimating the physical pion mass, we added the bootstrap errors on ξ and $\hat{m}$ in quadrature, rather than under the bootstrap. The results of our tuning are shown in table 2, wherein we show the fitted fermion anisotropies and our estimates of the mass of the pion.

3.5. Results

With the tuned fermion parameters in hand, we computed the condition numbers for the various kinds of preconditionings in our three quenched ensembles. We did not make use of the cut-off trick in our temporally preconditioned operators. This time in the isotropic case, we used the $N_{\mathrm{t}} = 16$ ensemble, since we did not need long time extents for fitting. Specifically we computed condition numbers for the unpreconditioned (Unprec) and the 4D Schur even–odd preconditioned operator (4D Schur) to use as a standard to compare against,

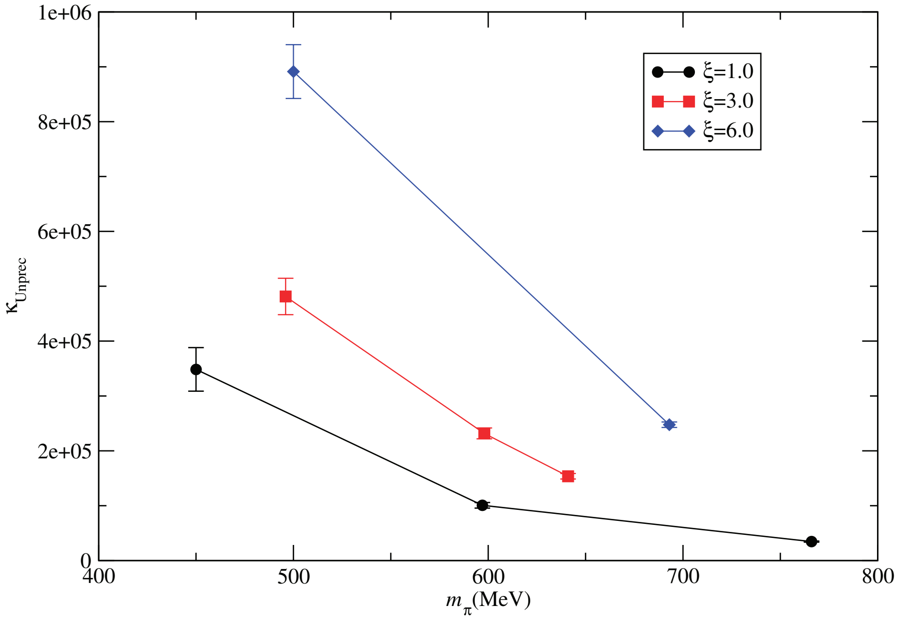

Figure 1. Condition numbers of the unpreconditioned clover operator against the measured pion mass, for all three of our anisotropy values: $\xi = 1.0$ (black circles), $\xi = 3.0$ (red squares) and $\xi = 6.0$ (blue diamonds).

as well as the temporally preconditioned operators combined with both 3D ILU even–odd preconditioning (TPrec + ILU) and Schur style preconditioning in three dimensions (Tprec + 3D Schur). We used 19 configurations from each ensemble to measure the condition numbers.

In figure 1, we show how the condition number of the unpreconditioned operator varies with pion (quark) mass and anisotropy. We can see, as one would expect, that the condition numbers increase with decreasing pion (quark) mass as well as with increasing anisotropy. In figure 2, we plot the ratios of the condition numbers of the preconditioned operators to the condition number of the unpreconditioned operator. We separate the results into three graphs for the three values of ξ used. Separation in terms of $\gamma_{\rm f}$ is not practical, since for different masses one can have different values of $\gamma_{\rm f}$ for the same renormalized anisotropy ξ. The ratios give a nice clean signal, as it appears that from configuration to configuration the ratios do not fluctuate very much (although the condition numbers themselves do). Hence the average of the ratios is very stable.

First we can see in the leftmost graph of figure 2 that in the isotropic case, the greatest gain comes from the 4D Schur preconditioning, whose condition number is about 15% of the unpreconditioned (Unprec.) case. Some gain over the unpreconditioned case can still be achieved using temporal preconditioning combined with either 3D Schur (TPrec + 3D Schur) or partial temporal preconditioning with ILU in 3D (TPrec + ILU). However, the temporally preconditioned cases are not as efficacious as the 4D Schur preconditioning in the isotropic case. Looking at the middle and rightmost graphs of figure 2, we see that in the anisotropic cases the temporally preconditioned operators fare much better, with condition numbers that are around 4–6% of the unpreconditioned one.

We note that the mass dependence in these ratios appears to be very mild, for a given value of ξ. Also with the temporal preconditioning, we see an improvement in the condition number ratios as ξ is increased. This is presumably due to the suppression factors of $1/\gamma_{\rm f}$ and $1/\gamma_{\rm f}^2$ in the temporally preconditioned operators. We expect that this is because of the clover term in the preconditioner, which has components in both the spatial and temporal directions, and the temporal components will counteract the suppression factors $\gamma_{\rm f}$ and $\gamma_{\rm f}^2$ to some degree. It is, however, encouraging to see that the partial preconditioning combined with 3D ILU even–odd preconditioning in space is nearly as good as in the 3D Schur preconditioned case. This suggests that not dealing with the clover term in the preconditioner is not in fact catastrophic. This is welcome news as the 3D ILU preconditioning is considerably simpler to implement and involves fewer FLOPs than the fully preconditioned 3D Schur approach.

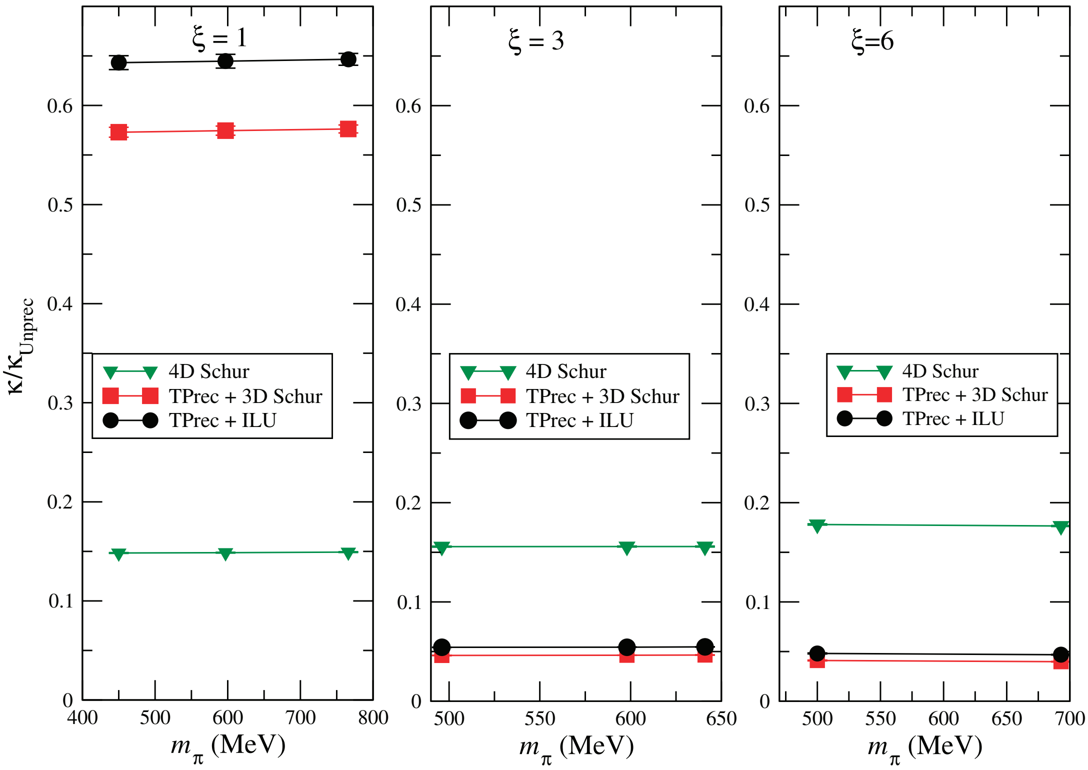

Figure 2. Ratios of the condition numbers of the preconditioned operators to the unpreconditioned operator at the same mass and anisotropy (lower is better conditioned). The three graphs are for anisotropies of $\xi = 1$ (left), $\xi = 3$ (middle) and $\xi = 6$ (right). The colors are 4D Schur preconditioning (green triangles), temporal preconditioning + 3D ILU (black circles) and temporal preconditioning + 3D Schur preconditioning (red squares).

We replot the condition number data in figure 3, this time plotting the ratio of the condition number of the 4D Schur operator to those of the temporally preconditioned operators, to see if we gain in condition with respect to the 'standard preconditioning'. The ratios are defined as

$$\frac{\kappa_{\text{4D Schur}}}{\kappa_{\text{TPrec.} + \text{ILU}}} \quad \text{and} \quad \frac{\kappa_{\text{4D Schur}}}{\kappa_{\text{TPrec.} + \text{3D Schur}}}, \tag{49}$$

and hence values larger than one indicate that the temporally preconditioned operator is better conditioned than the Schur4D, whereas values less than one indicate that the Schur4D is better conditioned.

We see, as before, in the leftmost graph of figure 3, that in the isotropic case neither of the temporal preconditioned approaches can beat the 4D Schur preconditioned approach. The condition numbers of the temporally preconditioned operators are about four times those of the Schur 4D case. Looking at the middle and rightmost graphs, one sees that the TPrec. + ILU preconditioning gives a decrease in the condition number of a factor of about 2.8–3.8 (depending on ξ), while the TPrec + 3D Schur preconditioning gives a decrease of a factor of about 3.3–4.4 compared to the standard 4D Schur even–odd preconditioning (again depending on ξ).

At this point, we should recall that the numerical overhead of the ILU style preconditioning is between 44 and 61%, depending on implementation. Using the most conservative overhead of 61%, this gives an overall gain in *total cost* for this scheme, compared to the 4D Schur case, of 2.8/1.61–3.8/1.61, which equates to between 1.74 and 2.36 for anisotropies of $\xi = 3$–6. With the most optimistic overhead estimate (44%), these gains can grow to 1.94–2.64.

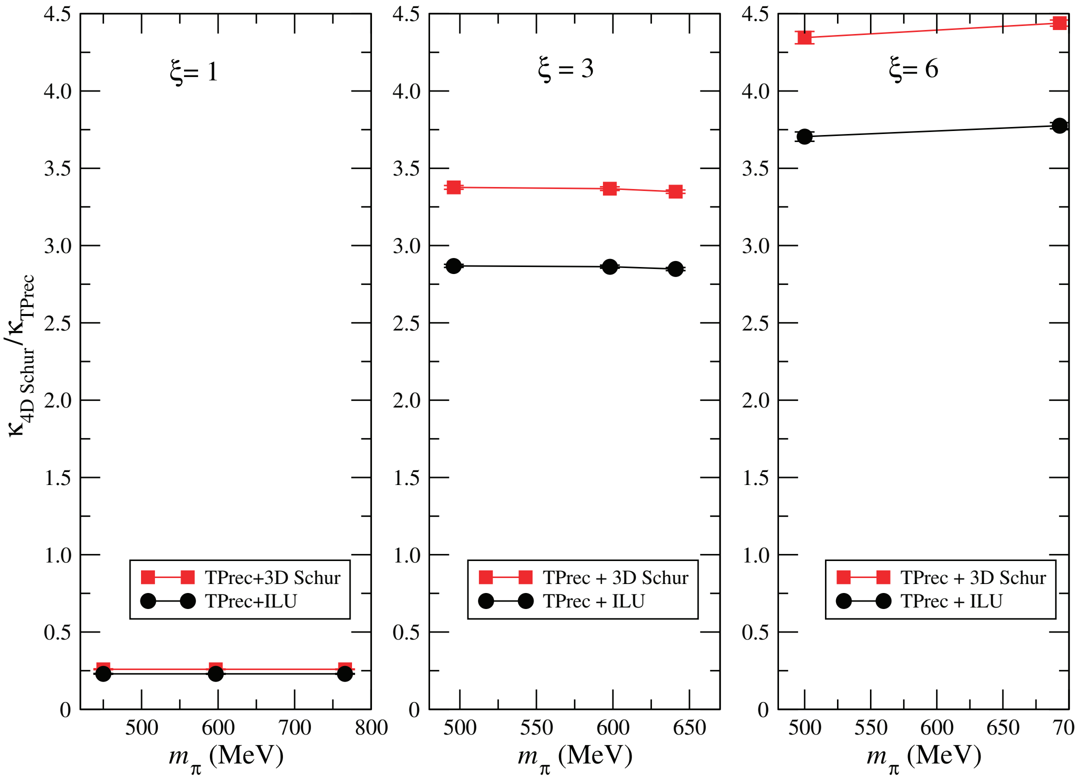

Figure 3. Ratios of the condition numbers of the 4D Schur even–odd preconditioned operator to the two temporally preconditioned cases (higher is better conditioned). The three graphs are for anisotropies of $\xi = 1$ (left), $\xi = 3$ (middle) and $\xi = 6$ (right). The colors are 4D Schur/TPrec + ILU (black circles) and 4D Schur/TPrec + 3D Schur (red squares).

4. Conclusions

We have demonstrated the technique of temporal preconditioning for anisotropic discretizations of the Wilson and SW Dirac operators in lattice QCD. We discussed the implementation of the technique in an efficient manner, and its combination with further even–odd preconditioning techniques, in particular the 3D ILU and 3D Schur even–odd approaches.

In the case of Wilson fermions, the 3D ILU and 3D Schur approaches are in fact identical and we have shown that the only terms in the preconditioned matrix that differ from the identity are suppressed by two powers of the anisotropy, ξ^2. For SW fermions, the presence of the clover term complicates the implementation since there are explicit terms that are diagonal in spatial indices and that connect the null spaces of the forward and backward projectors. The 3D ILU approach remains straightforward; however, the resulting preconditioned matrix has nonzero off-diagonal elements, which are only suppressed by factors of ξ rather than ξ^2.

We implemented the method in Chroma, a general code base for lattice QCD. We have also considered optimization techniques suggested from work in the realm of GPGPUs. Appendix B shows that the number of FLOPs in the temporally preconditioned operator with ILU even–odd preconditioning is about 30 and 61% higher than the corresponding 4D Schur preconditioned operator for the Wilson and SW operators, respectively, in the most generic case. Specialized implementation can reduce the floating point overhead of the clover action to about 44%. These costs may further be ameliorated as needed through judicious use of the cut-off trick. We have not used the cut-off trick in the numerical results of this paper.

We have carried out numerical tests in a variety of quenched ensembles, both isotropic and anisotropic, to investigate the efficacy of the techniques discussed, using Wilson–clover fermions over a range of quark masses.

Our main conclusion is that the technique works well for anisotropic cases. In our studies, with anisotropies of $\xi = 3$ and 6, the temporally preconditioned clover operator with ILU preconditioning had condition numbers between a factor of about 2.8–3.3 times smaller than the usual 4D Schur preconditioned case, whereas the temporally preconditioned operator with 3D Schur even–odd preconditioning had condition numbers that were about 3.3–4.4 times smaller than the usual 4D Schur preconditioned case. Combined with the various overhead estimates, this gives the temporally preconditioned, ILU preconditioned clover operator a cost advantage of a factor of 1.74–2.36 over the more standard 4D Schur preconditioned approach, depending on implementation. The 4D Schur preconditioner, however, appeared to be the best conditioned in the isotropic case.

Due to the decreases in condition number observed using the temporal preconditioned methods, it becomes attractive to extend the preconditioning scheme to the hybrid molecular dynamics–Monte Carlo algorithm such as Hybrid Monte Carlo [19]. New terms will arise in the Hamiltonian, due to the determinants of the preconditioning matrices. We leave the full discussion of these ramifications to a future publication.

Acknowledgments

This work was done using the Chroma software suite on clusters at Jefferson Laboratory using time awarded under the USQCD Initiative. MP was supported by Science Foundation Ireland under research grants 04/BRG/P0266 and 07/RFP/PHYF168. MP is grateful for the generous hospitality of the theory center at TJNAF, during which time some of this research was carried out. We would like to thank Mike Clark and Ron Babich for insightful discussions and pointing out the utility of the Dirac basis and the axial gauge to this technique. Authored by Jefferson Science Associates, LLC under US DOE Contract No. DE-AC05-06OR23177. The US Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce this manuscript for US Government purposes.

Appendix A. Full preconditioning with clover

In this appendix, we consider the question of how to deal with the inversion of the term $(A + \mu - D_{\mathrm{t}})$, in the case of full temporal preconditioning.

We begin with

$$A(\vec{x},t) + \mu - D_{\mathrm{t}}(\vec{x},t) = A(\vec{x},t) + \mu - P_{-}U_t(\vec{x},t)\delta_{\vec{x},t+1;\vec{x},t'} - P_{+}U_t^{\dagger}(\vec{x},t-1)\delta_{\vec{x},t-1;\vec{x},t}. \tag{A.1}$$

Inverting $A + \mu - D_t$ can be done with a single-step Woodbury procedure,

$$A + \mu - D_t = T_0 + VW^{\dagger}, \tag{A.2}$$

where

$$V = \begin{pmatrix} -U^{\dagger}(\vec{x}, N_{\mathrm{t}}-1)P_{+} \\ 0 \\ 0 \\ \vdots \\ 0 \\ -U(\vec{x}, N_{\mathrm{t}}-1)P_{-} \end{pmatrix}, \quad W^{\dagger} = (P_{-}, 0, 0, 0, \ldots P_{+}), \tag{A.3}$$

and, suppressing spatial indices, the matrix T_0 is now tridiagonal:

$$T_0 = \begin{pmatrix} A(0)+\mu & -U(0)P_{-} & 0 & \ldots & 0 \\ -U^{\dagger}(0)P_{+} & A(1)+\mu & -U(1)P_{-} & 0 & \ldots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & & -U^{\dagger}(N_{\mathrm{t}}-3)P_{+} & A(N_{\mathrm{t}}-2)+\mu & -U(N_{\mathrm{t}}-2)P_{-} \\ 0 & 0 & \vdots & -U^{\dagger}(N_{\mathrm{t}}-2)P_{+} & A(N_{\mathrm{t}}-1)+\mu \end{pmatrix}. \tag{A.4}$$

This matrix, while easy to apply, is not immediately straightforward to invert, because of its projector structure. In our numerical work, to gauge the efficacy of this approach, we used an inner conjugate gradients algorithm to invert this matrix.

Appendix B. FLOP count for the ILU scheme

Let us recount the number of FLOPs needed to apply the usual 4D Schur preconditioned operator:

$$\tilde{M}_{\text{4DSchur}} = A^{\text{oo}} - \tfrac{1}{4} D^{\text{oe}} (A^{-1})^{\text{ee}} D^{\text{eo}}, \tag{B.1}$$

where the mass term has been absorbed into the diagonal part of A and D^{oe} denotes the 4D hopping matrix. The parameters γ_{f} have been absorbed into a rescaling of the gauge links. We will use the notation $\mathcal{C}(Q)$ to denote the floating point cost of some generic term Q. Applying the A^{oo} term and A^{ee} both require 522 FLOPs per site, on $V/2$ sites (single checkerboard) each, giving $\mathcal{C}(A^{\text{oo}}) = \mathcal{C}(A^{\text{ee}}) = 522(V/2)$ FLOPs.

The D^{oe} and D^{eo} terms require 1320 FLOPs each per site on $V/2$ sites (single checkerboard) each. This number is arrived at by considering the spin projection operators as having no FLOPs, since only sign flips and exchanges of real and imaginary components are involved. Each SU(3) matrix–vector (matvec) operation requires 66 FLOPs, corresponding to three inner products between the three rows of the matrix, and the column vector. Each inner product involves three complex multiplies (six FLOPs each) and two complex adds (two FLOPs each), or $3 \times 6 + 2 \times 2 = 22$ FLOPs, giving a total of $3 \times 22 = 66$ for the three inner products in the complete matvec operation.

Spin reconstruction (recons) takes 12 FLOPs per site, coming from a single complex add for each of two spin-color components (six complex adds in total), again not counting sign flips and real complex component interchanges. Finally, 24 FLOPs are required to sum two (now reconstructed) four spinors (sumvec4) to evaluate the sum over directions (four spin-color components, so 12 complex components in total, with two FLOPs per component).

In N_{d} dimensions

$$\mathcal{C}(D^{\text{oe}}) = 2N_{\text{d}}(N_\gamma^* \text{matvec} + \text{recons}) + \text{sumvec4} \times (2N_{\text{d}} - 1), \tag{B.2}$$

where the factor of $2N_{\text{d}}$ comes from doing forward and backward projections and reconstructions in N_{d} dimensions, and N_γ^* is the number of spin components left after spin projection. So, in four dimensions, with $N_\gamma = 4$ we have $N_{\text{d}} = 4$, $N_\gamma^* = 2$ and one has

$$\mathcal{C}(D^{\text{oe}}) = 8(2 \times 66 + 12) + 24 \times 7 = 1320 \text{ FLOPs}. \tag{B.3}$$

Correspondingly, per site the 4D Schur operator takes up

$$\mathcal{C}(\tilde{M}_{\text{4DSchur}}) = 2 \times (1320 + 522) + 48 = 3732 \text{ FLOPs}, \tag{B.4}$$

where the last factor of 48 comes from the AXPY operation to apply the factor of $\frac{1}{4}$ and subtracting the two terms from each other. Hence, applying the operator costs $3732(V/2)$ FLOPs in total, which it is convenient to re-express as $3732N_{\text{t}}(V_{\text{s}}/2)$, with N_{t} the extent of the time direction and V_{s} the number of spatial coordinates per timeslice.

Let us now consider the ILU preconditioned operator

$$\tilde{M}_{\text{ILU}} = \begin{pmatrix} 1 + \bar{A}^{\text{ee}} & \bar{A}^{\text{ee}} \bar{D}_{\text{s}}^{\text{eo}} \\ \bar{D}^{\text{oe}} \bar{A}^{\text{ee}} & 1 + \bar{A}^{\text{oo}} - \bar{D}_{\text{s}}^{\text{oe}} \left[1 - \bar{A}^{\text{ee}}\right] \bar{D}_{\text{s}}^{\text{eo}} \end{pmatrix}, \tag{B.5}$$

where we have absorbed the factors of ξ into the $\bar{D}$ terms. Applying $\tilde{M}_{\text{ILU}}$ to some vector ψ, to result in $\chi = \tilde{M}_{\text{ILU}}\psi$, implies that

$$\chi_{\text{e}} = \psi_{\text{e}} + \bar{A}^{\text{ee}} \left(\bar{D}^{\text{eo}} \psi_{\text{o}} + \psi_{\text{e}}\right), \tag{B.6}$$

$$\chi_{\text{o}} = \psi_{\text{o}} + \bar{A}^{\text{oo}} \psi_{\text{o}} + \bar{D}^{\text{oe}} \left\{\bar{A}^{\text{ee}} \left[\bar{D}^{\text{eo}} \psi_{\text{o}} + \psi_{\text{e}}\right] - \bar{D}^{\text{eo}} \psi_{\text{o}}\right\}, \tag{B.7}$$

and one can reuse several terms between $\chi_{\rm e}$ and $\chi_{\rm o}$. Thus $\tilde{M}_{\rm ILU}$ can be efficiently applied by computing in sequence

$$\chi_{\rm e} = \psi_{\rm e}, \tag{B.8}$$

$$\chi_{\rm o} = \psi_{\rm o}, \tag{B.9}$$

$$t_1 = \bar{D}^{\rm eo}\psi_{\rm o}, \tag{B.10}$$

$$t_2 = \bar{A}^{\rm ee}\left(t_1 + \psi_{\rm e}\right), \tag{B.11}$$

$$\chi_{\rm e} = \chi_{\rm e} + t_2, \tag{B.12}$$

$$\chi_{\rm o} = \chi_{\rm o} + \bar{A}^{\rm oo}\psi_{\rm o} + \bar{D}^{\rm oe}\left(t_2 - t_1\right), \tag{B.13}$$

and apart from five vector adds, we need to apply $\bar{A}$ and $\bar{D}$ twice each (once per checkerboard). Since $\bar{A} = C_{\rm L}AC_{\rm R}$ and $\bar{D} = C_{\rm L}DC_{\rm R}$, we have

$$\begin{aligned} \mathcal{C}(\tilde{M}_{\rm ILU}) &= 5 \times 24 \times (V/2) + 2\left[\mathcal{C}(\bar{D}_{\rm s}) + \mathcal{C}(\bar{A})\right] \\ &= 120(V/2) + 2\left[2\mathcal{C}(C_{\rm L}) + 2\mathcal{C}(C_{\rm R}) + \mathcal{C}(A) + \mathcal{C}(D_{\rm s})\right], \end{aligned} \tag{B.14}$$

where the 5 × 24 × $(V/2)$ term is to account for the five vector adds, each on a single checkerboard.

We still have $\mathcal{C}(A) = 522$ per site and by substituting $N_{\rm d} = 3$ into the discussion for $\mathcal{C}(D)$ one finds that applications of the 3D $D_{\rm s}$ operator cost 984 FLOPs per site. The preconditioners $C_{\rm L}$ and $C_{\rm R}$ have the same cost each, namely $\mathcal{C}(C)$, and so

$$\begin{aligned} \mathcal{C}(\tilde{M}_{\rm ILU}) &= 120(V/2) + 2\left[4\mathcal{C}(C) + \mathcal{C}(A) + \mathcal{C}(D_{\rm s})\right] \\ &= 120(V/2) + 8\mathcal{C}(C) + (2 \times 522 + 2 \times 984)(V/2) \\ &= 8\mathcal{C}(C) + 3132(V/2). \end{aligned} \tag{B.15}$$

Applying C requires a back (forward) substitution for each *spatial site* of the appropriate checkerboard. We consider the back-substitution here, but the working is similar (and the FLOP count is identical) for the forward substitution. The back-substitution needs to be performed on only two spin-color components, after spin projections with either $1 + \gamma_0$ or $1 - \gamma_0$ for $C_{\rm L}$ or $C_{\rm R}$, followed by an appropriate reconstruction. After the back-substitution, the Woodbury procedure involves, for each spatial site, working with length $N_{\rm t}$ block vectors, a matrix multiplication by the a pre-computed SU(3) matrix and an SU(3) subtraction. As usual we do not count any FLOPs for the projection part of the spin projection steps. This gives

$$\mathcal{C}(C) = N_\gamma^*\left[(V_{\rm s}/2)\left(\mathcal{C}(B) + \mathcal{C}(W)\right)\right] + (V/2)(12 + 48), \tag{B.16}$$

where $V_{\rm s}$ denotes the number of spatial sites, $\mathcal{C}(B)$ denotes the cost for the back-substitution, $\mathcal{C}(W)$ denotes the rest of the Woodbury process, the $12(V/2)$ FLOPs come from the spin reconstruction on one checkerboard and the $48(V/2)$ FLOPs come from adding the P_+ and P_- terms in $C_{\rm L}$ and $C_{\rm R}$; one vector addition costing 24 FLOPs and an overall scaling by $\frac{1}{2}$ to normalize the projectors costing another 24 FLOPs.

We now need to consider the back-substitution on a single spin-color component. The first step is just a scaling by the diagonal $\psi_{N_{\rm t}-1} = \frac{1}{\mu}\chi_{N_{\rm t}-1}$ or six FLOPs. Then there follow $N_{\rm t} - 1$ steps of $\psi_i = \frac{1}{\mu}[\chi_i - U_i\psi_{i+1}]$, each one comprised of an SU(3) matrix vector multiply (66 FLOPs) and a subtraction and a scaling by μ (six FLOPs each). In total,

$$\mathcal{C}(B) = 6 + (N_{\rm t} - 1)\left(66 + 6 + 6\right) = 78N_{\rm t} - 72 \text{ FLOPs}. \tag{B.17}$$

The remainder of the Woodbury procedure involves computing the $X\Lambda W^\dagger$ term. This can be achieved by precomputing $X\Lambda W^\dagger$ for each value of t at initialization, and then this process costs only $N_{\rm t}$ matrix vector operations (66 FLOPs each), and finally we need to subtract the result of this matrix multiply from the result of the back/forward substitution (six FLOPs) for each value of N_t and so

$$\mathcal{C}(W) = (66+6)N_\mathrm{t} = 72N_\mathrm{t}\ \text{FLOPs} \tag{B.18}$$

and

$$\begin{aligned}\mathcal{C}(C) &= N_\gamma^*(V_\mathrm{s}/2)\,[78N_\mathrm{t} - 72 + 72N_\mathrm{t}] + (V/2)[12+48]\\ &= N_\gamma^*(V_\mathrm{s}/2)[150N_\mathrm{t} - 72] + 60(V/2)\\ &= (V_\mathrm{s}/2)(300N_\mathrm{t} - 144) + 60N_\mathrm{t}(V_\mathrm{s}/2)\\ &= (V_\mathrm{s}/2)(360N_\mathrm{t} - 144)\ \text{FLOPs},\end{aligned}$$

where we have used $V_\mathrm{s}V_\mathrm{t} = V$, and $N_\gamma^* = 2$ so

$$\begin{aligned}\mathcal{C}(\tilde{M}_\mathrm{ILU}) &= 8\mathcal{C}(C) + 3132(V/2)\\ &= 8(V_\mathrm{s}/2)\,(360N_\mathrm{t} - 144) + 3132N_\mathrm{t}(V_\mathrm{s}/2)\\ &= (V_\mathrm{s}/2)\,(2880N_\mathrm{t} - 1152) + 3132N_\mathrm{t}(V_\mathrm{s}/2)\\ &= (6012N_\mathrm{t} - 1152)\,(V_\mathrm{s}/2)\ \text{FLOPs}.\end{aligned} \tag{B.19}$$

Comparing the costs of the two preconditioned operators, we have for clover fermions

$$R = \frac{\mathcal{C}(\tilde{M}_\mathrm{ILU})}{\mathcal{C}(\tilde{M}_\mathrm{4DSchur})} = \frac{6012N_\mathrm{t} - 1152}{3732N_\mathrm{t}} \approx 1.611 - \frac{0.309}{N_\mathrm{t}} \tag{B.20}$$

and so we consider two limiting cases: in the first instance when $N_\mathrm{t} = 1$ we have $R = 1.302$ to 3 decimal places (d.p.) and when N_t is sufficiently large that the term involving it is negligible we have $R = 1.611$ to 3 d.p. In a typical case, when $N_\mathrm{t} = 128$–256 one has $R \approx 1.61$ to 2 d.p.

Similarly we can consider the case for just Wilson fermions. In this case

$$\tilde{M}_\mathrm{4DSchur} = \mu - \frac{1}{4\mu}D^\mathrm{oe}D^\mathrm{eo} \tag{B.21}$$

and the cost is

$$\mathcal{C}(\tilde{M}_\mathrm{4DSchur}) = (48 + 2\times 1320)N_\mathrm{t}(V_\mathrm{s}/2) = 2668N_\mathrm{t}(V_\mathrm{s}/2)\ \text{FLOPs}, \tag{B.22}$$

where the 48 FLOPS come from the subtraction and scaling (AXPY) operation.

The temporally preconditioned scheme needs only the evaluation of

$$\tilde{M}_\mathrm{ILU} = I - \bar{D}_\mathrm{s}^\mathrm{oe}\bar{D}^\mathrm{eo} \tag{B.23}$$

and so the cost in FLOPs is

$$\begin{aligned}\mathcal{C}(\tilde{M}_\mathrm{ILU}) &= 24(V/2) + 4\mathcal{C}(C) + 2\mathcal{C}(D_\mathrm{s})\\ &= 24N_\mathrm{t}(V_\mathrm{s}/2) + 4(V_\mathrm{s}/2)\,(360N_\mathrm{t} - 144) + 2\times 984N_\mathrm{t}(V_\mathrm{s}/2)\\ &= (V_\mathrm{s}/2)\,(24N_\mathrm{t} + 1440N_\mathrm{t} - 576 + 1968N_\mathrm{t})\\ &= (V_\mathrm{s}/2)\,(3432N_\mathrm{t} - 576)\ \text{FLOPs}.\end{aligned} \tag{B.24}$$

We have

$$R = \frac{3432N_\mathrm{t} - 576}{2668N_\mathrm{t}} \approx 1.286 - 0.216\frac{1}{N_\mathrm{t}}, \tag{B.25}$$

giving R in the range of $R \approx 1.07$–1.286.

Use of the cut-off trick removes the multiplication by components of X and subtraction of the term Xq in steps 2(b) and 2(c) of the Sherman–Morrison–Woodbury process. If the cut-off value of t is k, one saves k timeslices, from $t = 0$ to $t = k - 1$, per spatial site in the ILU clover operator evaluating T^{-1} or $(T^{\dagger})^{-1}$, on each of $N_{\gamma}^{*} = 2$ spin components. On one spin component one saves 66 + 6 FLOPs, and so per spatial site one saves $2 \times (66 + 6) = 144$ FLOPs per timeslice: altogether $144k$ FLOPs. One does this every C_{L} and C_{R}, of which there are four in the Wilson case and eight in the ILU preconditioned clover case, each of which is evaluated on a single checkerboard ($V_{\mathrm{s}}/2$ sites). Correspondingly the cut-off trick on k timeslices saves $576k(V_{\mathrm{s}}/2)$ FLOPs for the ILU Wilson and $1152k(V_{\mathrm{s}}/2)$ FLOPs for the ILU clover operator.

Appendix C. Dirac basis and axial gauge

The use of the Dirac basis and the axial gauge was advocated in [12, 13] to save FLOPs and memory bandwidth in implementations of the Dirac operator on GPU systems. The use of these techniques is also beneficial in the case of temporal preconditioning. We analyze below the benefits for temporal preconditioning in terms of FLOPs savings, but note also that savings in memory bandwidth resulting from the use of these techniques can help the performance of the implementations.

The use of the Dirac basis, where γ_0 is diagonal, can save the cost of spinor reconstruction in the time direction, saving some 48 FLOPs per site in the 4D D operators—12 each in the forward and backward time directions, respectively, from not having to reconstruct, and another 12 each when accumulating since now only half vectors need to be accumulated, rather than the reconstructed four vectors. This corresponds to a saving of $48N_{\mathrm{t}}(V_{\mathrm{s}}/2)$ FLOPs, per D operator, or a total of $96N_{\mathrm{t}}(V_{\mathrm{s}}/2)$ when considering the full 4D Schur preconditioned operator (where D is applied twice).

Correspondingly, this trick can save the $(V/2)(12 + 48) = 60N_{\mathrm{t}}(V_{\mathrm{s}}/2)$ FLOPs term from the cost of applying C_{L} and C_{R} since no spinor reconstruction is needed and one does not need to expend FLOPs when adding the results of the P_{+} and P_{-} projectors, as the projectors will simply write their results to different components of the final four spinor. In the case of temporal preconditioned clover fermions, where the preconditioner is used eight times overall, this results in a saving of $480N_{\mathrm{t}}(V_{\mathrm{s}}/2)$ FLOPs. In the case of unimproved fermions, where the preconditioner is used only four times, one can save $240N_{\mathrm{t}}(V_{\mathrm{s}}/2)$ FLOPs.

Use of the axial gauge allows one to save further FLOPs. In this case all the links in the temporal direction (apart from the boundary) are transformed to have value $U_i = 1$. In the case of the 4D D operator, this saves 2 SU(3) matrix vector multiplies per spin component, coming from the forward and backward temporal link matrices on the non-boundary sites. To make counting easier, we will assume that the number of boundary sites is negligible and assume the saving for every site. Hence we count $2N_{\gamma}^{*} \times 66 = 264$ FLOPs saved per site. Two applications of D are needed for the 4D Schur preconditioned operators, so this trick saves roughly 264 FLOPs per site or $528N_{\mathrm{t}}(V_{\mathrm{s}}/2)$ FLOPs in total.

Correspondingly, the $N_{\mathrm{t}} - 1$ back(forward) substitution steps change from $\psi_i = \frac{1}{\mu}[\chi - U_i \psi_{i+1}]$ to merely $\psi_i = \frac{1}{\mu}[\chi - \psi_{i+1}]$ and one saves $N_{\mathrm{t}} - 1$ matrix multiplies in applying the preconditioner for each spin component, leading to a cost saving of $66N_{\gamma}^{*}(N_{\mathrm{t}} - 1) = 132N_{\mathrm{t}} - 132$ FLOPs per spatial site.

In the case of clover fermions, where eight applications of the preconditioner are needed, the saving is $(1056N_{\mathrm{t}} - 1056)V_{\mathrm{s}}/2$ FLOPs, whereas in the case of unimproved Wilson fermions where four applications are needed, the saving is $(528N_{\mathrm{t}} - 528)(V_{\mathrm{s}}/2)$.

Combining the savings from the use of the Dirac basis and the use of the axial gauge, one obtains, for the cost of the temporally preconditioned clover operator,

$$\mathcal{C}(\tilde{M}_{\mathrm{ILU}}) = (4476N_{\mathrm{t}} - 96)(V_{\mathrm{s}}/2), \tag{C.1}$$

which is a saving of roughly 25% over the previous cost in equation (B.19). The cost of the 4D Schur preconditioned operator also reduced,

$$\mathcal{C}(\tilde{M}_{\mathrm{4DSchur}}) = 3108N_{\mathrm{t}}(V_{\mathrm{s}}/2), \tag{C.2}$$

giving the overhead from the preconditioning as

$$R = 1.440 - \frac{0.031}{N_t}. \tag{C.3}$$

In the unimproved Wilson operator, one has

$$\mathcal{C}(\tilde{M}_{\mathrm{ILU}}) = (2664 N_t - 48)(V_s/2), \tag{C.4}$$

corresponding to a saving of about 22% over the previous case in equation (B.24).

The cost of the 4D Schur preconditioned operator becomes

$$\mathcal{C}(\tilde{M}_{\mathrm{4DSchur}}) = 2044 N_t (V_s/2), \tag{C.5}$$

giving

$$R = 1.30 - \frac{0.023}{N_t}. \tag{C.6}$$

Thus, the overhead of temporal preconditioning is between roughly 41–43% for clover and 27–30% for the unimproved Wilson case.

References

[1] Morningstar C J and Peardon M J 1997 Efficient glueball simulations on anisotropic lattices *Phys. Rev.* D **56** 4043–61

[2] Umeda T *et al* 2003 Two flavors of dynamical quarks on anisotropic lattices *Phys. Rev.* D **68** 034503

[3] Morrin R, Cais A O, Peardon M, Ryan S M and Skullerud J-I 2006 Dynamical QCD simulations on anisotropic lattices *Phys. Rev.* D **74** 014505

[4] Lin H-W *et al* 2009 First results from 2 + 1 dynamical quark flavors on an anisotropic lattice: light-hadron spectroscopy and setting the strange-quark mass *Phys. Rev.* D **79** 034502

[5] Edwards R G, Joo B and Lin H-W 2008 Tuning for three-flavors of anisotropic clover fermions with Stout-link smearing *Phys. Rev.* D **78** 054501

[6] Lin H-W, Edwards R G and Joo B 2007 Parameter tuning of three-flavor dynamical anisotropic clover action *PoS*(LATTICE2007) 119 (arXiv:0709.4680 [hep-lat])

[7] Sheikholeslami B and Wohlert R 1985 Improved continuum limit lattice action for QCD with Wilson fermions *Nucl. Phys.* B **259** 572

[8] Sherman J and Morrison W J. 1949 Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix *Ann. Math. Stat.* **20** 621

[9] Sherman J and Morrison W J 1950 Adjustment of an inverse matrix corresponding to a change in one element of a given matrix 1950 *Ann. Math. Stat.* **21** 124–7

[10] Woodbury M A 1950 *Inverting Modified Matrices (Statistical Research Group, Memorandum Report 42)* (Princeton, NJ: Princeton University Press)

[11] Edwards R G and Joo B 2005 The Chroma software system for lattice QCD *Nucl. Phys. Proc. Suppl.* **140** 832

[12] Clark M A, Babich R, Barros K, Brower R C and Rebbi C 2009 Solving Lattice QCD systems of equations using mixed precision solvers on GPUs arXiv:0911.3191 [hep-lat]

[13] Barros K, Babich R, Brower R, Clark M A and Rebbi C 2008 Blasting through lattice calculations using cuda *PoS* (LATTICE2008) 045

[14] Bunk B 1997 Computing the lowest eigenvalues of the fermion matrix by subspace iterations *Nucl. Phys. Proc. Suppl.* **53** 987–9

[15] Michael C and Shanahan H 1996 Masses and decay constants of the light mesons in the quenched approximation using the tadpole–improved SW-clover action *Nucl. Phys. Proc. Suppl.* **47** 337–44

[16] Klassen T R 1998 The anisotropic Wilson gauge action *Nucl. Phys.* B **533** 557–75

[17] Allton C R *et al* 1993 Gauge invariant smearing and matrix correlators using Wilson fermions at $\beta = 6.2$ *Phys. Rev.* D **47** 5128–37

[18] Morningstar C and Peardon M J 2004 Analytic smearing of $SU(3)$ link variables in lattice QCD *Phys. Rev.* D **69** 054501

[19] Duane S, Kennedy A D, Pendleton B J and Roweth D 1987 Hybrid Monte Carlo *Phys. Lett.* B **195** 216–22