# Lawrence Berkeley National Laboratory

**Title**
Multigrid for chiral lattice fermions: Domain wall

**Authors**
Brower, Richard C
Clark, MA
Weinberg, Evan
et al.

# Multigrid for chiral lattice fermions: Domain wall

Richard C. Brower[*]

*Boston University, Boston, Massachusetts 02215, USA*

M. A. Clark[†] and Evan Weinberg[‡]

*NVIDIA Corporation, Santa Clara, California 95051, USA*

Dean Howarth[§]

*Physics Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA,*
*Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
*and Department of Physics and Center for Computational Science, Boston University, Boston,*
*Massachusetts 02215, USA*

The phenomena of *critical slowing down* in the iterative solution of the Dirac equation presents a major challenge to further applications of lattice field theory in the approach to the continuum solution. We propose a new multigrid approach for chiral fermions, applicable to both the 5D domain wall or 4D overlap operator. The central idea is to directly coarsen the 4D Wilson kernel, giving an effective domain wall or overlap operator on each level. We provide here an explicit construction for the Shamir domain wall formulation with numerical tests for the 2D Schwinger prototype, demonstrating near ideal multigrid scaling. The framework is designed for a natural extension to 4D lattice QCD chiral fermions, such as the Möbius, Zolotarev or Borici domain wall discretizations or directly to a rational expansion of the 4D overlap operator. For the Shamir operator, the effective overlap operator is isolated by the use of a Pauli-Villars preconditioner in the spirit of the Kähler-Dirac spectral map used in a recent staggered multigrid algorithm [R. C. Brower, E. Weinberg, M. A. Clark, and A. Strelchenko, Phys. Rev. D **97**, 114513 (2018)].

## I. INTRODUCTION

The intersection of the preexascale era and ever-evolving theoretical insights continues to boost the predictive power of computational approaches to quantum field theories, most substantively for lattice quantum chromodynamics [1]. The increased computational resources enable larger lattice volumes and finer lattice spacings, which on the one hand reveal multiple scales, but on the other hand further increases the cost of the iterative solution of the lattice Dirac equation. This increase in cost is best known as *critical slowing*: a superlinear increase in the ill conditioning of the Dirac matrix as the fermion mass approaches the chiral limit [2].

A robust solution to this problem was recognized over 30 years ago [3]: coupling the fine-grid representation of the linear solver to coarser grids, i.e., grids on multiple scales, echoing the features of Wilson's real space renormalization group [4]. In the context of linear solvers, this manifests as a recursive multigrid (MG) preconditioner. The original investigations of a gauge-invariant MG algorithm in the 1990s produced encouraging results for weak (or smooth) background gauge fields near the continuum limit [5,6]. Given the computational resources at the time, lattice sizes were too small and gauge fields were too rough for this MG algorithm to achieve useful improvements.

The successful extension to strong background chromodynamic fields was enabled by the development of fully recursive adaptive geometric MG method for the Wilson-Dirac fermion discretization [7,8]. While this was and continues to be of great use, it is only one piece of the broader problem: there are two other discretizations, referred to as staggered [9] and domain wall [10] fermions, that are used extensively in high energy applications that more faithfully represent *chiral* symmetry on the lattice. The extension of adaptive geometric MG to these discretizations has proven to be more difficult, perhaps related to the improved lattice chiral symmetry. There has been

[*]brower@bu.edu
[†]mclark@nvidia.com
[‡]eweinberg@nvidia.com
[§]howarth1@llnl.gov

progress on MG algorithms for domain wall fermions, two-level for the normal operator [11,12] and fully recursive for the $\Gamma_5$-Hermitian operator [13]. There has also been progress on a non-Galerkin algorithm for the closely related overlap operator [14,15]. Recently an adaptive geometric multigrid algorithm for staggered fermions was discovered based on a novel preconditioner inspired by the Kähler-Dirac spin structure [16,17].

Here we propose a new approach to the domain wall discretization which leverages, at least on a heuristic level, features developed from both the Wilson and staggered MG methods. We hope that a comparison of methods will lead to new optimizations across the full set of discretizations. The design strategy of our domain wall MG algorithm consists of trying to separate the 4D physical subspace of low modes found in the effective 4D overlap operator from the larger 5D domain wall vector space. This procedure is conveniently enumerated in three steps:

  (i) Approximate Pauli-Villars preconditioning of the domain wall operator [18].
  (ii) Wilson kernel MG projection on the domain wall and Pauli-Villars factors [8].
  (iii) Truncated projection/prolongation restricted to the domain wall boundary.

The salient features of each step are as follows: (i) the exact Pauli-Villars inverse $D_{\mathrm{PV}}^{-1}$, which is a perfect map from the DW spectrum onto overlap, is well approximated by the application of the Pauli-Villars adjoint, $D_{\mathrm{PV}}^{\dagger}$. (ii) A Galerkin MG coarsening is applied to the 4D Wilson kernel on each extradimensional slice separately for both the domain wall and Pauli-Villars factors. The null space projection is formulated entirely from the 4D Wilson kernel and does not scale with the size of the extra dimension. (iii) Finally, within the multigrid cycle, the residual coarsening and error interpolation is restricted to the domain wall, which in turn allows the extent of the extra dimension of the coarse-level operator to be reduced.

We again follow the successful development strategy for the Wilson [7] and staggered [19] MG algorithms by using the two-flavor lattice Schwinger model [20,21] as a prototype for exploration and testing. The reader is referred to Fig. 6 and the accompanying Table II for a concise summary of the performance of our domain wall algorithm for the 2D Schwinger model. In Sec. II, the underlying motivation and formalism is given. For simplicity, the discussion is

restricted to the Shamir domain wall operator [22]. This is followed in Sec. III by the details of the implementation and benchmarks for our prototype 2D Schwinger model. Care is taken to present the formalism in a dimension-agnostic form to accommodate extensions from 2D to 4D gauge theories. In Sec. V we conclude by noting that our core developments not only apply to the 4D Shamir formulation presented here but also to the Möbius [23,24], Borici [25,26], and Zolotarev [27,28] formulations, as well as directly to the overlap operator approximation to the sign function [14,29–31].

## II. DOMAIN WALL FORMALISM

All lattice discretizations of the Dirac operator seek to rapidly approach the continuum Dirac operator,

$$D\psi(x) = \gamma_\mu(\partial_\mu - iA_\mu(x))\psi(x) + m\psi(x), \qquad (1)$$

as the lattice spacing vanishes. The continuum operator is a first derivative, an anti-Hermitian operator, thus the spectrum is imaginary indefinite except for a small real shift for $m > 0$. It obeys an exact chiral symmetry at zero mass ($m = 0$). The Wilson discretization,

$$
\begin{aligned}
D_W(U, m)_{x,y} = &-\frac{1 - \gamma_\mu}{2} U_\mu(\vec{x})\delta_{x+\mu,y} \\
&-\frac{1 + \gamma_\mu}{2} U_\mu^\dagger(\vec{x} - \hat{\mu})\delta_{x-\mu,y} + (d + m)\delta_{x,y},
\end{aligned}
$$

$$(2)$$

introduces an anti-Hermitian "naïve" first difference and adds a Hermitian second difference (or so-called Wilson term) to lift the doublers to the cutoff scale $\pi/a$ at the expense of explicitly violating lattice chiral symmetry at $\mathbb{O}(a)$ in lattice spacing. The Wilson lattice operator then requires fine tuning of the bare quark in order to restore chiral symmetry in the continuum limit.

The chiral overlap [14,32] and domain wall [22] fermions, beyond their remarkable physical properties, have the feature that the Wilson kernel can be repurposed. In the domain wall (DW) approach, the Wilson kernel is present in an extra dimension separating 4D domain walls by a lattice of length $L_s$. Suppressing the four-dimensional indices, the DW operator is given by

$$
D_{\mathrm{DW}}(m)_{s's} =
\begin{bmatrix}
D_W(M_5) + 1 & P_- & 0 & \cdots & -mP_+ \\
P_+ & D_W(M_5) + 1 & P_- & \cdots & 0 \\
0 & P_+ & D_W(M_5) + 1 & \cdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & P_- \\
-mP_- & 0 & \cdots & P_+ & D_W(M_5) + 1
\end{bmatrix},
\qquad (3)
$$

where $P_\pm = \frac{1}{2}(1 \pm \gamma_5)$. The indices $s, s' = 1, \cdots L_s$ label 4D blocks in the extra fifth dimension (or $d + 1$ dimension). The bulk mass, $M_5 < -1$, is tachyonic. The physical bare mass parameter is encoded by the boundary parameter $m$.

In the limit of $L_s \to \infty$ an exact lattice chiral symmetry appears up to an explicit fermion mass gap given by

$$m_q = \frac{m}{1-m} \simeq m \quad \text{as} \quad m \to 0. \tag{4}$$

The result is that propagators between the domain walls are described by the effective 4D overlap operator proposed by Neuberger [14,33] with the deformed chiral algebra of the Ginsparg-Wilson identity [34],

$$\gamma_5 D_{\text{ov}}^{-1} + D_{\text{ov}}^{-1} \gamma_5 = \mathcal{O}(a), \tag{5}$$

at zero quark mass. The explicit spectral map from the domain wall to the overlap operator will be presented in Sec. II B, following closely the notation in [18] which we will refer to as BNO. This spectral map between domain wall and overlap operators plays a central role in our DW MG algorithm.

### A. Standard approaches and shortcomings

The domain wall operator encodes chiral symmetry in a subtle and indirect fashion. The full spectrum of the domain wall operator in Eq. (3), illustrated in Fig. 1 for two dimensions, does not have the expected small eigenvalues of the continuum as you approach the chiral limit, but instead has $\mathcal{O}(L_s)$ small eigenvalues. This can easily be seen in the free limit ($U = 1$) which at zero momentum has the spectrum,

$$\lambda_n = m^{1/L_s} e^{i(2n+1)\pi/L_s} \quad \text{for} \quad n = 0, 1, ..., L_s - 1. \tag{6}$$
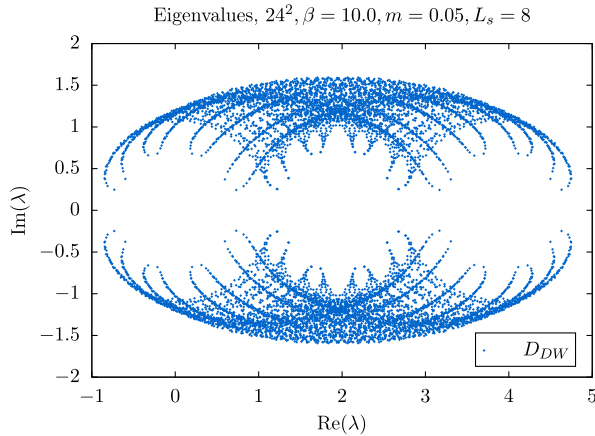


FIG. 1.   The spectrum of the domain wall operator lacks the continuum chiral low modes; instead there are $2^{d/2} L_s$ of them ($2^{d/2}$ coming from the spin degrees of freedom) in a circle around zero in the complex plane.

In the exact chiral limit, first taking $L_S \to \infty$ followed by $m \to 0$, this operator has no zero modes. Instead they form a unit circle around the origin in the complex plane. This feature persists when gauge fields are turned on as illustrated in Fig. 1 for two dimensions with Abelian gauge fields.

The DW operator features further issues. Unlike with staggered or Wilson fermions, the DW operator is dramatically non-normal ($[D_{\text{DW}}(m), D_{\text{DW}}^\dagger(m)] \neq 0$), even in the exact free field. The spectrum does not satisfy the half-plane eigenvalue condition with positive real values ($\text{Re}[\lambda] > 0$). These two properties defeat reliable convergence properties of traditional Krylov solvers. For example, consider a normal indefinite matrix, whose spectrum fits in a circle of radius $r$ centered at a complex point $c \in \mathbb{C}$. For generalized minimum residual (GMRES) methods, one can show the relative residual on iteration $n$ of a GMRES method is bounded by $|r/c|^n$ [35].

A standard method to solve the domain wall linear system is to replace

$$D_{\text{DW}}(m) \ \Psi = b \tag{7}$$

with the normal system,

$$D_{\text{DW}}^\dagger(m) \ D_{\text{DW}}(m) \ \Psi = D_{\text{DW}}^\dagger(m) \ b. \tag{8}$$

This system has multiple benefits but also complications.

The normal operator encodes a single low mode with, in the free-field limit, eigenvalues $\mathcal{O}(m^2)$. The low modes are "bound" to the domain wall, as can be visually inspected by looking at the profile of the eigenvectors (the singular vectors of the domain wall operator) in the bulk dimension. Both of these properties form a stark contrast with $D_{\text{DW}}$; the normal operator transforms the physics of the domain wall operator into a single chiral fermion below the cutoff.

From a numerical standpoint, the normal operator is Hermitian positive definite and can be solved efficiently by traditional Krylov methods, e.g., conjugate gradient. Further, it is amenable to deflation with eigenpairs generated via an efficient Lanczos process. Also as a Hermitian positive definite matrix, the solver for this normal operator can be implemented as a traditional MG algorithm [11,12]. However, a numerical implementation of the coarsened normal operator, a distance-two stencil, requires a nontrivial increase in computation and communication relative to a distance-one stencil [36–39]. This owes in large part to a far more complicated gather pattern due to around-the-corner terms, which the original fine level original normal operator avoids by being the product of two distance-one operators.

One solution to this issue is to recognize that this normal operator is the square of a distance-one "$\gamma_5$" Hermitian operator,

$$D_{\text{DW}}^\dagger(m) \ D_{\text{DW}}(m) = (\Gamma_5 D_{\text{DW}}(m))^2, \tag{9}$$

where $\Gamma_5 = \gamma_5 R$ is a product of $\gamma_5$ and the reflection in the extra dimension by $R_{ss'} = \delta_{(s+s'-1)\%L_s,0}$. This operator has an indefinite real spectrum similar to the imaginary spectra in the continuum and staggered operator on the lattice. Appealingly, the operator also has a single chiral mode with eigenvalues of $\mathcal{O}(m)$. Its distance-1 stencil and well-behaved spectra led to its use in the hierarchically deflated conjugate residual algorithm of [13]. However, this operator itself has its own frustrations: while it can be coarsened, it develops spurious small eigenvalues similar to with a naïve approach to staggered fermions, which was shown in [19] to harm a fully recursive algorithm.

As a last concern, this approach does not generalize beyond Shamir domain wall fermions: the fully general Möbius formulation has a Wilson kernel inverse as part of its definition of $\Gamma_5$. Our new approach seeks to avoid these difficulties base on the spectral map form the domain wall and to the overlap representations in BNO, combined with methods borrowed from prior multigrid algorithms for Wilson [7] and staggered [19] discretizations.

### B. Spectral map from domain wall to overlap

One may view the Pauli-Villars operator, $D_{\mathrm{PV}} \equiv D_{\mathrm{DW}}(1)$, as a left preconditioner of the domain wall operator in the linear system

$$D_{\mathrm{PV}}^{-1} D_{\mathrm{DW}}(m)\Psi = D_{\mathrm{PV}}^{-1} b. \qquad (10)$$

We will show that the Pauli-Villars operator is an ideal, albeit expensive, preconditioner and that even a simple approximation to the inverse dramatically accelerates convergence. This is accomplished via the generalized eigenmode problem, $D_{\mathrm{DW}}(m)\Psi_\lambda = \lambda D_{\mathrm{PV}}\Psi_\lambda$, which separates the low chiral generalized eigenvectors ($\lambda \simeq 0$) bound to the walls at $s = 1$ and $s = L_s$,

$$\psi_x = \frac{1}{2}(1-\gamma_5)\Psi_{x,1} + \frac{1}{2}(1+\gamma_5)\Psi_{x,L_s}, \qquad (11)$$

from the high bulk modes at the cutoff: $\lambda = \mathcal{O}(\pi/a)$.

To see this explicitly, it is convenient as in BNO to first move both walls to $s = 1$ by introducing a cyclic permutation of the negative chiral modes at $s = L_s$ to $s = 1$ by

$$\mathcal{P}_{s's} = \begin{bmatrix} P_- & P_+ & \cdots & 0 \\ 0 & P_- & P_+ & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_+ \\ P_+ & 0 & \cdots & P_- \end{bmatrix}_{s's}, \qquad (12)$$

with $P_\pm = \frac{1}{2}(1 \pm \gamma_5)$. This defines a unitary transformation of our domain wall operators,

$$D_{\mathrm{DW}}(m) \to \mathcal{P}^\dagger D_{\mathrm{DW}}(m)\mathcal{P}. \qquad (13)$$

Following the derivation via a LDU transformation in BNO, the preconditioned matrix in this *permuted chiral basis*,

$$K_{\mathrm{DW}}(m) = (D_{\mathrm{PV}}\mathcal{P})^{-1} D^{\mathrm{DW}}(m)\mathcal{P}$$

$$= \begin{bmatrix} D_{\mathrm{ov}}(m) & 0 & 0 & \cdots & \cdots & 0 \\ -(1-m)\Delta_2 & 1 & 0 & 0 & \cdots & 0 \\ -(1-m)\Delta_3 & 0 & 1 & 0 & \cdots & 0 \\ -(1-m)\Delta_4 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ -(1-m)\Delta_{L_s} & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

$$(14)$$

is mapped into a block diagonal form [18] in the extra dimension. This remarkable identity is the central observation for our preconditioned multigrid algorithm. The effective overlap operator block, $K_{1,1}^{\mathrm{DW}}(m) = D_{\mathrm{ov}}(m)$, has all the non-trivial low eigenvalues. The additional extra heavy modes are mapped exactly to unit eigenvalues (or in physical units at the $1/a$), irrespective of $L_s$, lattice spacing (here scaled to $a = 1$), and gauge interactions. A representative example of the complete spectrum is illustrated in Fig. 2.

Parenthetically, we should acknowledge that this general mechanism to isolate the low domain wall modes in an effective overlap block is well known and it is the fundamental insight to chiral lattice fermions [32]. In particular the Monte Carlo sampling of the path integral must divide by the Pauli-Villars determent to give a finite determinant ratio, $\det[K] = \det[D_{\mathrm{DW}}(m)]/\det[D_{\mathrm{DW}}(1)]$ in the continuum. As explained by Kaplan and Schmaltz in [40] in an elegant exposition based on a *kinematic super symmetry* cancellation between the bulk fermion and the bosonic Pauli-Villars pseudofermions, broken only by
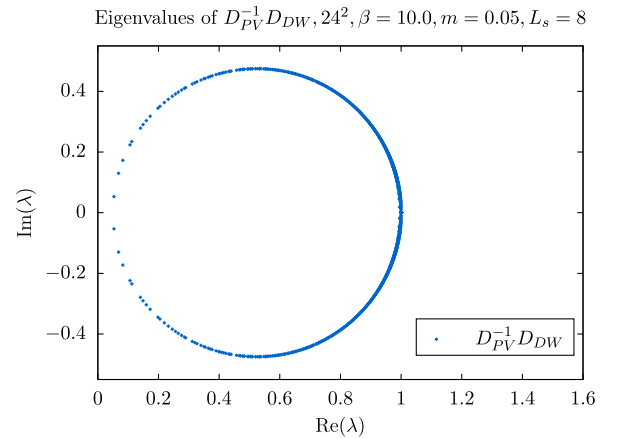


FIG. 2. A representative spectrum of the preconditioned domain wall operator, $D_{\mathrm{PV}}^{-1} D^{\mathrm{DW}}(m)$. Deviations from an exact circle exist but are qualitatively negligible. Note the similarity to the Kähler-Dirac preconditioned staggered operator in Fig. 3.

domain wall boundary, to give rise to boundary chiral modes via the Callan-Harvey descent relations [41]. The block structure of $K_{\mathrm{DW}}(m)$ lends itself to a structured block-inverse given by

$$
K_{\mathrm{DW}}^{-1}(m) = \begin{bmatrix} D_{\mathrm{ov}}^{-1}(m) & 0 & 0 & \cdots & \cdots & 0 \\ (1-m)\Delta_2 & 1 & 0 & 0 & \cdots & 0 \\ (1-m)\Delta_3 & 0 & 1 & 0 & \cdots & 0 \\ (1-m)\Delta_4 & 0 & 0 & 1 & \cdots & 0 \\ & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ (1-m)\Delta_{L_s} & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}. \quad (15)
$$

The identification of the overlap propagator $G_{11} \equiv D_{\mathrm{ov}}^{-1}(m)$ agrees with the practice of computing DW propagators by solving the linear system, $D_{\mathrm{DW}}(m)G = D_{\mathrm{PV}}b$, from an arbitrary source on the wall $b \sim \delta_{1,s}$. The heavy modes $(0,0,\ldots,1,\ldots,0)$ are static. On the other hand, the non-zero elements in the first column show that the chiral modes bleed exponentially into the interior by factors, $\Delta_{s+1} = T\Delta_s = T^s/(1+T^L)$ in terms of the transfer matrix: $T = (1-H)/(1+H)$. At finite $L_s$ the overlap operator is

$$
\begin{aligned}
D_{\mathrm{ov}}(m) &= \frac{1+m}{2} + \frac{1-m}{2}\gamma_5\epsilon_L[H], \\
&= m + (1-m)D_{\mathrm{ov}}(0),
\end{aligned} \quad (16)
$$

with

$$
\epsilon_L[H] = \frac{(1-H)^L - (1+H)^L}{(1-H)^L + (1+H)^L} \quad (17)
$$

and

$$
H = \gamma_5 D_W(M_5)/(2 + D_W(M_5)) \quad (18)
$$

in the Shamir implementation. In the limit $L_s \to \infty$ this becomes the exact sign function: $\epsilon_L[H] \to \hat{\gamma}_5 = \mathrm{sign}[H]$. The same spectral transformation into this sparse structure in Eq. (15) applies to other implementations of domain wall fermions (Möbius, Borici, Zolotarev, etc). The modifications include a variation of the Hermitian kernel $H$ and the functional $\epsilon_L[H]$ that converge, $\epsilon_L[H] \to \mathrm{sign}[H]$ to the sign function as $L_s \to \infty$.

We can trace the sparse block structure to the mass dependence on the dyadic structure at the boundaries relating the Pauli-Villars operator to the domain wall operator,

$$
D_{\mathrm{PV}} = D_{\mathrm{DW}}(m) + (1-m) \begin{bmatrix} P_+ \\ 0 \\ 0 \\ \vdots \\ P_- \end{bmatrix} \otimes [P_- \; 0 \; 0 \; \cdots \; P_+], \quad (19)
$$

or $D_{\mathrm{PV}} \equiv D_{\mathrm{DW}}(m) + (1-m)UV^\dagger$. After applying the Sherman-Morrison-Woodbury formula [42],

$$
D_{\mathrm{PV}}^{-1}D_{\mathrm{DW}}(m) = 1 - D_{\mathrm{DW}}^{-1}(m)U\frac{(1-m)}{I + (1-m)V^T D_{\mathrm{DW}}^{-1}(m)U}V^T, \quad (20)
$$

and again considering the chiral basis, $V^T \to V^T\mathcal{P} = [1\,0\,0\,\cdots\,0]$, we see a clear projection onto the first column in $K_{s,s'}^{\mathrm{DW}}(m)$, reproducing the sparse structure in Eq. (14).

### 1. Free-field limit

The analysis of the free-field ($U = 1$) limit for the domain wall and Pauli-Villars operators gives valuable insight and guidance to our MG construction, particularly when examining the low spectra well below the UV cutoff scale, $\pi/a$. Qualitative and even quantitative features survive the introduction of the gauge fields generated by lattice Monte Carlo methods.

We begin by transforming the free Wilson kernel in Eq. (3) to momentum space,

$$
\begin{aligned}
\tilde{D}_W(p_\mu) &= am + \gamma_\mu \sin(ap_\mu) + 2\sin^2(ap_\mu/2), \\
&= am + \sum_\mu (1 - e^{-i\gamma_\mu ap_\mu}),
\end{aligned} \quad (21)
$$

where we introduce the expression on the right to emphasize the well-known feature of circular arcs in the complex spectrum of the Wilson operator. This gaping separates the doubler modes from the continuum modes as evident in Fig. 4 even with nonzero gauge fields turned on. The lattice spacing **a** has been introduced to identify physical low modes ($|p| \ll \pi/a$) relative to UV cutoff: $\mathcal{O}(\pi/a)$. The low spectrum, $\lambda_\pm \simeq m \pm i\sqrt{p^2} + ap^2/2$, is the continuum Dirac spectrum plus the $\mathcal{O}(ap^2)$ Wilson term.

This Fourier analysis had a straightforward generalization to the Pauli-Villars operator, $D_{\mathrm{PV}}$, as the boundary conditions are antiperiodic in the fifth dimension,

$$
\tilde{D}_{\mathrm{PV}}(p_\mu, p_5) = -e^{-i\gamma_5 ap_5} + \tilde{D}_W(p_\mu) + (1 + M_5). \quad (22)
$$

The boundary conditions restrict the bulk momentum $p_5$ to half integer modes: $p_5 = \pi(2n+1)/L_s$. After setting the mass to its free-field tachyonic value $M_5 = -1$, the low momentum expansion again has the familiar Wilson form
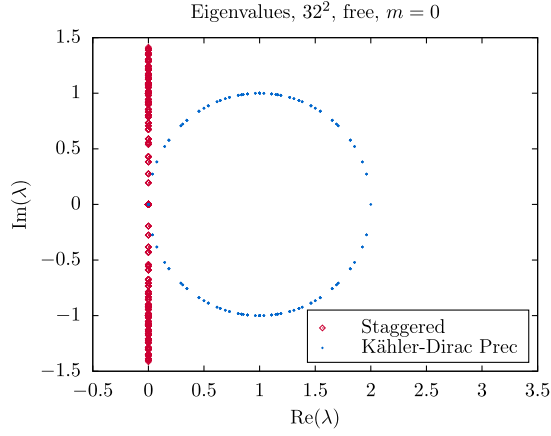
FIG. 3. The spectrum of the two-dimensional free, massless staggered operator before (vertical line) and after Kähler-Dirac preconditioning.

$(i\gamma_\mu p_\mu + (a/2)p^2)$ with the first "eye" displaced to a circle centered at $\lambda = 1$. Not surprisingly the free domain wall operator, $D_{\mathrm{DW}}(m)$, which differs by a fermion-mass dependence on the boundary (reducing to Dirichlet boundary conditions for $m = 0$), has a qualitatively similar spectra as seen in Fig. 1 even with nonzero gauge fields. Both have no small eigenvalues below the cutoff for free fields. (With interactions small eigenvalues occur when the topological charge changes.)

Turning to the normal equation, we see a dramatic difference. While the domain wall normal operator has chiral modes at $m = 0$, as we show in the Appendix A, the normal equation for the Pauli-Villars operator is positive definite with a large gap from zero with singular values of order the cutoff:

$$D_{\mathrm{PV}}D_{\mathrm{PV}}^\dagger = 1 + \mathcal{O}(p^4). \tag{23}$$

The first correction is $\mathcal{O}(a^2 p^4)$ in physical units. Indeed more generally for $M_5 = -1$ one can prove for any finite $L_s$ in the free limit that $D_{\mathrm{PV}}^\dagger D_{\mathrm{PV}}$ is bounded from below by 1, i.e., at the lattice cutoff scale: $1/a^2$. This feature suggest the usefulness of our approximate preconditioning,

$$D_{\mathrm{PV}}^{-1}D_{\mathrm{DW}} = D_{\mathrm{PV}}^\dagger (D_{\mathrm{PV}}D_{\mathrm{PV}}^\dagger)^{-1}D_{\mathrm{DW}},$$
$$\simeq D_{\mathrm{PV}}^\dagger D_{\mathrm{DW}}, \tag{24}$$

which avoids the expensive need to invert the Pauli-Villars operator. Even after gauge fields are included, we note in Fig. 5 this approximation conforms well at small eigenvalues including the $\mathcal{O}(ap^2)$ for the parabolic curvature in the complex plane.

To explore this further, we summarize results from Appendix A, comparing the low momentum spectra for the exact overlap map, $D_{\mathrm{PV}}^{-1}D_{\mathrm{DW}}$, with the approximate map, $D_{\mathrm{PV}}^\dagger D_{\mathrm{DW}}$. At $m = 0$, the low spectrum for $D_{\mathrm{PV}}^{-1}D_{\mathrm{DW}}$ is given by $\lambda_\pm^0 = \pm i\sqrt{p^2} + ap^2 + \mathcal{O}(a^2)$. By using the shift identity, $\lambda^0 \to \lambda = am + (1 - am)\lambda^0$ as implicit in Eq. (A6), the low spectrum for nonzero mass is

$$\lambda_\pm = am \pm i(1 - am)\sqrt{p^2 + a(1 - am)p^2} + \mathcal{O}(a^2). \tag{25}$$

which after a rescaling, $\lambda_\pm \to \lambda_\pm/(1 - am) = am_q \pm i\sqrt{p^2 + ap^2}$, is a Wilson-like dispersion relation. The general expansion in powers of momentum $p^2$ has a rather remarkable independence on $L_s$. As noted in Appendix A, a direct evaluation of the free overlap kernel in Eq. (A6) for any finite $L_s \geq 2$ results in a series in $p^2$ which, up to $\mathcal{O}((a^2 p^2)^k)$, is independent of $L_s \geq k$. This invariance of the low momentum expansion with respect to size of the
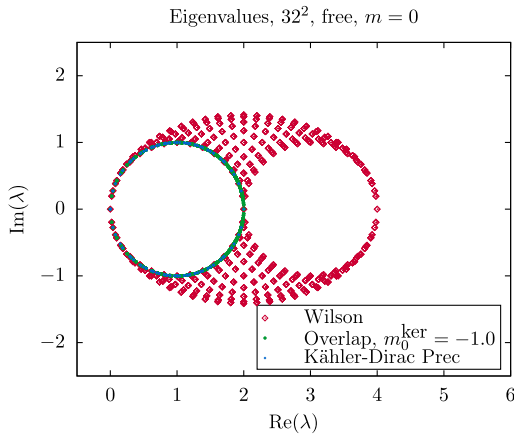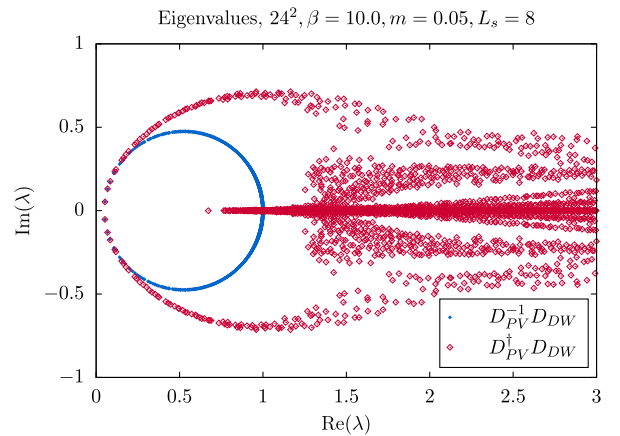


FIG. 4. Comparison of the 2D free effective overlap and Kähler-Dirac preconditioned operator with almost identical circles centered at $\lambda = 1$ overlaid with 2D free Wilson operator with two doublers at $\lambda = 2$ and one at $\lambda = 4$.



FIG. 5. The spectrum of our target multigrid operator, $D_{\mathrm{PV}}^\dagger D_{\mathrm{DW}}$, compared with the effective overlap spectrum, $D_{\mathrm{PV}}^{-1}D_{\mathrm{DW}}$. For clarity of presentation we truncate the $x$ axis; the spectrum of $D_{\mathrm{PV}}^\dagger D_{\mathrm{DW}}$ extends out to $\mathrm{Re}(\lambda) \approx 25$.

extra dimension may explain the efficacy of reducing the size of extra dimension $L_s$ on the coarse level iterations as documented in Sec. IV B.

Finally we compare the low mass spectra to our approximate preconditioned operator, $D_{PV}^\dagger D_{DW}(m)$,

$$\lambda_\pm = am \pm i(1 - am)\sqrt{p^2 + ap^2} + \mathcal{O}(a^2), \quad (26)$$

relative to the exact preconditioner in Eq. (25). The only difference to quadratic order occurs at dimension 6 with a contribution $\mathcal{O}(a^2 m p^2)$, helping to explain why a Wilson-like spectra in the overlap sector is preserved in Fig. 5, including the parabolic curvature to $\mathcal{O}(ap^2)$.

### 2. Similarity with the Kähler-Dirac preconditioned staggered operator

It is interesting to compare the Wilson and the overlap spectra with the preconditioned spectrum for the staggered MG algorithm in Ref. [19]. The staggered lattice operator has the unique property, shared by the continuum, of being an exactly anti-Hermitian operator plus a constant mass shift as illustrated for $m = 0$ by the vertical (red) spectra in Fig. 3. Both the staggered and continuum operators are normal operators.

At first these similarities between the staggered operator and the continuum may seem to be ideal for multigrid, but this turned out to be major obstacle to extending the Galerkin projection method used successfully for the Wilson MG algorithm [15] to the staggered operator. The solution found in Ref. [19] was to first precondition by *dividing* the anti-Hermitian staggered operator by the spin-taste Kähler-Dirac block, deforming the spectrum into one resembling the overlap spectrum or the first "eye" of Wilson spectra in Fig. 3.

More specifically, this required writing the staggered operator as an "even/odd" $2^d$ block operator (i.e., $2^2$ squares in 2D, $2^4$ hypercubes in 4D) decomposed as a sum of block-local terms "$B$" and block-hopping terms "$C$" as described in Eq. 2.11 of [19]. Each of these terms are separately anti-Hermitian operators with an indefinite spectra. In this formalism, the preconditioned operator is simply the block Jacobi preconditioned form, $B^{-1}D_{stag}$. This maps the spectrum onto an exactly unitary circle in the free case, as shown in Fig. 3, resembling the exact overlap operator at $L_s = \infty$.

The structural similarity to the Pauli-Villars preconditioner is striking. In this case, we could have also started with a pair of anti-Hermitian indefinite operators, "$i\Gamma_5 D_{PV}$" for the Pauli-Villars operator and "$i\Gamma_5 D_{DW}(m)$" for the domain wall operator. We can now formulate the Pauli-Villars preconditioned domain wall operator as

$$D_{DW}(m) \to D_{PV}^{-1}D_{DW}(m) \equiv (i\Gamma_5 D_{PV})^{-1}(i\Gamma_5 D_{DW}(m)),$$
$$(27)$$

where we have made explicit that this preconditioning takes an imaginary indefinite spectrum to a unitary circle, identical in form to Fig. 2. Beyond the practical consequence of this observation, it is intriguing to ask why this is the case. It may hint of a unifying principle for our multigrid algorithm common to all three major fermion discretizations in the chiral limit, which is worthy of additional investigation.

### C. Outline of our three step multigrid implementation

#### 1. Pauli-Villars preconditioning

For the first step we need to consider the ideal Pauli-Villars preconditioner. It is worth reemphasizing the challenge and importance of preconditioning the domain wall operator. The domain wall operator on a $d + 1$ dimensional lattice increases the number of eigenvalues from $N_{ev} = 2^{d/2} \times N_c \times L^d$ by a factor of $L_s$. The Pauli-Villars inverse spectra transform is an ideal preconditioner, putting the "bulk" $N_{ev} \times (L_s - 1)$ eigenvalues exactly at the cutoff $1/a$ in physical units.

Due to the Pauli-Villars operator having a maximally indefinite spectrum, it is most optimally solved via the normal operator. Although the Pauli-Villars normal operator is extremely well conditioned with a positive real spectrum starting at $1/a^2$, its use as a preconditioner is still prohibitively expensive. Instead we consider an approximate Pauli-Villars preconditioner,

$$D_{PV}^{-1} = D_{PV}^\dagger [D_{PV}D_{PV}^\dagger]^{-1} \simeq D_{PV}^\dagger, \quad (28)$$

as motivated by low-order expansion of the Pauli-Villars normal operator given in Eq. (23). This approximate operator importantly preserves the property that the spectrum is confined to the right complex half-plane,

$$(D_{PV}^\dagger D_{PV})D_{PV}^{-1}D_{DW}|\lambda\rangle = re^{i\theta}|\lambda\rangle, \quad (29)$$

with $-\pi/2 < \theta < \pi/2$ for all eigenvalues $\lambda = r \exp[i\theta]$. This is proven in Appendix B based on the positive definite spectra of the normal operator factor and the right-half plane spectrum of $D_{PV}^{-1}D_{DW}$. This does imply Krylov solvers such as BiCGStab can be directly applied to this approximate operator. While Fig. 5 is consistent with this property, the qualitatively strong match between the low eigenvalues of the two operators suggests we can make a much stronger statement. Indeed the two operators are nearly identical, with deviations confined to larger eigenvalues in the approach to the cutoff scale $\pi/a$. This is again motivated by the free field limit where we prove for $M_5 = -1$

$$D_{PV}D_{PV}^\dagger = 1 + \mathcal{O}(p^4), \quad (30)$$

and as a result the spectrum of the approximation is valid up to $\mathcal{O}(p^4)$ corrections. Indeed in the free theory, the additive

operator to 1 (or $1/a^2$ in physical units) is positive definite for all momenta. Further details on this can be found in Appendix A.

We note that a point of future investigation could be approximating $(D_{\mathrm{PV}}D_{\mathrm{PV}}^{\dagger})^{-1}$ by a low-order polynomial in the normal operator as opposed to truncating it to 1. Given the success of the truncation to 1, it is unclear if higher order polynomials would be worth the additional computational burden.

### 2. Wilson kernel MG projection

In the second step, we introduce a coarsening projection using the familiar Galerkin projecting developed for Wilson MG acting independently on the Wilson kernel for each of the $s = 1, 2, \ldots L_s$ slices,

$$\hat{D}^W_{\hat{x},\hat{x}'} = \mathbb{P}^{\dagger}_{\hat{x},x} D^W_{x,x'} \mathbb{P}_{x',\hat{x}'} \quad \text{or} \quad \hat{D}_W = \mathbb{P}^{\dagger} D_W \mathbb{P}. \quad (31)$$

Here color and spin indices are implicit, and on the right we have further followed the convention of Wilson MG by suppressing the indices of the $d$-dimensional space-time lattice. The projection preserves $\gamma_5$ so that $\gamma_5 \mathbb{P} = \mathbb{P}\sigma_3$ and

$$\mathbb{P}^{\dagger}(1 \pm \gamma_5)\mathbb{P} = 1 \pm \sigma_3. \quad (32)$$

With the normalization convention of the restrictor, $\mathbb{P}^{\dagger}\mathbb{P} = \mathbb{I}$, giving the identity operator on the coarse vector space and the fact that $D_{\mathrm{DW}}(m)$ is a linear functional of the Wilson kernel, we have

$$\hat{D}_W = \mathbb{P}^{\dagger} D_W \mathbb{P} \Rightarrow \hat{D}_{\mathrm{DW}}(m) = \mathbb{P}^{\dagger} D_{\mathrm{DW}}(m)\mathbb{P}, \quad (33)$$

with the implicit redefining on the right of the restrictor as diagonal in $s$ space: $\mathbb{P} \to \mathbb{P}\delta_{s,s'}$. This notational *slight of hand* is common practice in the physics literature with tensor expressions. For example in Eq. (32) we also implicitly redefined $\gamma_5$ as diagonal in color and $d$-dimensional space-time.

Of course this factorization also applies to the Pauli-Villars term and to generalized domain wall formulations such as Möbius, Zolotarev etc. However this factorization does not apply to nonlinear functional of the kernel such as the preconditioned product

$$(\mathbb{P}^{\dagger} D_{\mathrm{PV}} \mathbb{P}^{\dagger})^{-1} \mathbb{P}^{\dagger} D_{\mathrm{DW}}(m)\mathbb{P}^{\dagger} \neq \mathbb{P}^{\dagger} D_{\mathrm{PV}}^{-1} D_{\mathrm{DW}}(m)\mathbb{P}^{\dagger} \quad (34)$$

or the overlap operator. For these the kernel projection does not commute with the operator. To make this clear we explicitly write the coarsened form of the domain wall and Pauli-Villars operators,

$$\hat{D}_{\mathrm{DW}}(m) = \begin{bmatrix} \hat{D}_W(M_5) + 1 & \hat{P}_- & 0 & \cdots & -m\hat{P}_+ \\ \hat{P}_+ & \hat{D}_W(M_5) + 1 & \hat{P}_- & \cdots & 0 \\ 0 & \hat{P}_+ & \hat{D}_W(M_5) + 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \hat{P}_- \\ -m\hat{P}_- & 0 & \cdots & \hat{P}_+ & \hat{D}_W(M_5) + 1 \end{bmatrix}, \quad (35)$$

where $\hat{P}_{\pm} = \frac{1}{2}(1 \pm \sigma_3)$. This coarsened operator has identical algebraic structure as original fine-level Domain wall operator. So when applying a coarse Pauli-Villars preconditioner $\hat{D}_{\mathrm{PV}}^{-1}$, the exact same algebraic manipulations for the BNO factorization carry over. Again the bulk modes are moved to the cutoff and "chiral" modes are confined to the boundary to form an effective coarse $\hat{D}_{\mathrm{ov}}(m)$ operator. This recursive construction is the key element for our DW multigrid construction.

### 3. Truncated projection/prolongation

Lastly, in the third step, we define a convention for residual coarsening and error correction prolongation. One straightforward approach would be to coarsen and prolongate across all $L_s$ slices. Instead, we find it is possible and in fact advantageous to only restrict and prolong the boundary contribution for the residual coarsening and the error correction, respectively. We assert this convention

here and use it in general going forward, however we quantitatively study only acting on the boundary as opposed to the entire bulk in Sec. IV A.

In the convention where the boundary is at the $s = 1$ slice, we define the projection of residual and prolongation of the error on the boundary by

$$\hat{r}_s = \begin{cases} \mathbb{P}^{\dagger} r_1 & \text{for } s = 1 \\ 0 & \text{for } s > 1 \end{cases} \quad (36)$$

and

$$e_s = \begin{cases} \mathbb{P}\hat{e}_1 & \text{for } s = 1 \\ 0 & \text{for } s > 1 \end{cases}. \quad (37)$$

The $s > 1$ elements are restored efficiently by the outer solver and as a smoother correction. This is inspired by the *Sec 3.2 Overlap to bulk domain wall reconstruction*

procedure in BNO [18]. It may be surprising, but this is still effective despite the fact we are using $\hat{D}_{\text{PV}}^\dagger$ in place of $\hat{D}_{\text{PV}}^{-1}$. We discuss this further below, but in brief, this can be motivated as reasonable again by equivalence, $D_{\text{PV}}^\dagger \sim D_{\text{PV}}^{-1}$, up to the quadratic order in momenta in the free theory.

All of these elements come with a variety of potential parameters for optimization on each level. For example since we are only transferring the residual and error correction on a single slice we are also able to reduce the extra dimension on the coarse levels. We will denote $L_s$ for the first coarsening as $\hat{L}_s$, on the second coarsening $\hat{\hat{L}}_s$, and so on. We will use a reduced $\hat{L}_s$ and $\hat{\hat{L}}_s$ going forward, however we quantitatively study varying the coarser $L_s$ in Sec. IV B.

### 4. Summary

The full MG algorithm for the linear system $D_{\text{DW}}\,x = b$, formulated as an extension to a $K$ cycle (i.e., a multigrid cycle where each coarse solve is wrapped in a Krylov solver) is as follows:

(1) Left precondition the system by $D_{\text{PV}}^\dagger$, giving the new linear system

$$D_{\text{PV}}^\dagger D_{\text{DW}}\, x = D_{\text{PV}}^\dagger b. \qquad (38)$$

(2) Perform an MG-preconditioned iterative solve (via generalized conjugate residual (GCR), flexible GMRES (FGMRES), etc.) using the operator $D_{\text{PV}}^\dagger D_{\text{DW}}$. The MG preconditioner within the $K$ cycle is defined as follows.

  (i) Relax on the current residual with $D_{\text{PV}}^\dagger D_{\text{DW}}$, known as the presmoother.

  (ii) Project the residual on the wall using Eq. (36).

  (iii) Using a Krylov solver, approximately solve the coarse level system:

$$\hat{D}_{\text{PV}}^\dagger \hat{D}_{\text{DW}}\,\hat{e} = \hat{r}, \qquad (39)$$

    using, as defined in Eq. (35), the projected Wilson kernel

$$\hat{D}_W(U, M_5) = \mathbb{P}^\dagger D_W(U, M_5)\mathbb{P}. \qquad (40)$$

  (iv) Prolong the error with Eq. (37) and correct the solution: $x = x + e$.

  (v) Postsmooth on the accumulated error from the previous two steps with $D_{\text{PV}}^\dagger D_{\text{DW}}$.

(3) Repeat until the desired tolerance on $\|b - D_{\text{DW}}x\|$.

Needless to say, there are several knobs to tune, even as far as MG algorithms go. We have explored a few of these parameters in a preliminary form in the 2D two-flavor Schwinger model and have left unexplored further until testing for 4D domain wall methods discussed briefly in the conclusion.

## III. NUMERICAL TESTS WITH THE 2D SCHWINGER MODEL

We now turn to testing our domain wall MG algorithm on the two-flavor Schwinger model [20,21,43]. As with the cases of Wilson [7] and staggered MG [19], the Schwinger model is a useful framework for the development and testing of algorithms for QCD [44]. As a low-dimensional prototype model it has the advantage of enabling the rapid exploration of a wide variety of alternative features in a serial laptop code. This can be used to demonstrate validity of an MG algorithm and guide the subsequent application to QCD and software tuning at scale on modern GPU accelerated systems. The importance of this two-step approach can not be over emphasized.

For our investigations in the interacting case, we have fixed $M_5 = -1.05$ relative to the correct free-field value $M_5 = -1$ and $L_s = 16$ as a representative large value. Our current performance of the DW MG algorithm outlined above is illustrated in Fig. 6 and the accompanying Table II. The one exception to these parameters is our study of the continuum limit, where we explore the addition of deflation on the coarsest level.

Before describing the details, it is apparent that our basic algorithm vastly improves scaling in the approach to the continuum and chiral limit, nearly eliminating critical slowing down. Our study of additional deflation in the continuum limit, given in Table I, suggests that the lack of perfect scaling shown in Fig. 6 can be improved with deflation. This is supported by ongoing investigations of deflation of the coarsest level with both twisted mass and highly improved staggered quarks (HISQ) fermions in 4D QCD and inspiration from [13].

### A. Algorithmic details and analysis

A benefit of our MG algorithm is its setup is the same as the setup for the traditional $\gamma_5$-preserving MG algorithm for
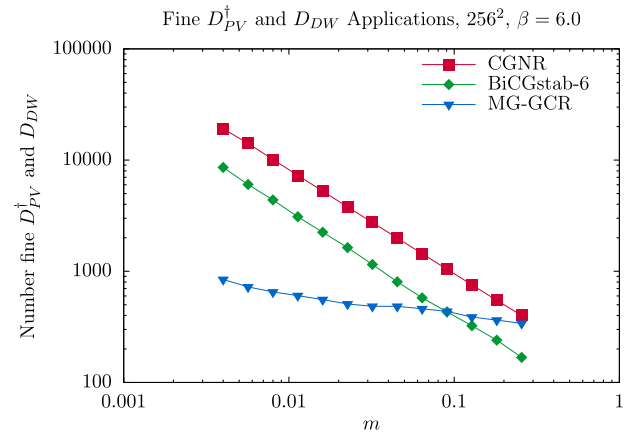


FIG. 6. The number of domain wall operator ($D_{\text{DW}}, D_{\text{DW}}^\dagger, D_{\text{PV}}^\dagger$) applications required for a solve to a tolerance of $10^{-10}$ using CGNR on $D_{\text{DW}}^\dagger D_{\text{DW}}$, BiCGStab-6 on $D_{\text{PV}}^\dagger D_{\text{DW}}$, and multigrid for a representative fixed $\beta$ and volume. Note that this is a log-log plot.

TABLE I.   The behavior of our MG algorithm in the approach to the continuum limit at constant physics and physical volume. We see an improvement in convergence as we approach the continuum limit, noting that stable convergence depends on an inexpensive deflation of the coarsest level. We chose to deflate 128 eigenvectors on the coarsest level.

| $L$ | $m$ | $\beta$ | Deflation | Fine $D_{PV}, D_{DW}$ | Intermediate average iterations | Coarsest average iterations |
|---|---|---|---|---|---|---|
| 64 | 0.01 | 3.0 | N | 820 | 2.88(07) | $2.62(06) \times 10^3$ |
| 128 | 0.005 | 12.0 | N | 484 | 2.85(17) | $4.71(42) \times 10^3$ |
| 256 | 0.0025 | 48.0 | N | 460 | 3.32(23) | $1.06(11) \times 10^4$ |
| 64 | 0.01 | 3.0 | Y | 820 | 2.62(09) | 51.2(3.9) |
| 128 | 0.005 | 12.0 | Y | 412 | 2.29(14) | 68.6(0.9) |
| 256 | 0.0025 | 48.0 | Y | 412 | 2.00(09) | 93.2(2.0) |

Wilson fermions. In Eq. (35) we see that the coarsened domain wall operator takes a Galerkin-projected Wilson operator as a kernel. For this reason we expect the setup to be roughly the same cost as one five-dimensional domain wall solve. We have tuned the Wilson operator to the critical mass for near-null vector generation. Near-null vectors are generated by relaxing on the homogeneous normal system with a Gaussian-distributed initial guess followed by a chiral doubling via $\frac{1}{2}(1 + P_{\pm})$ to preserve a $\sigma_3$ Hermiticity on the coarser levels [7,19].

For the solve, the main structure of the $K$ cycle is unchanged relative to our previous work. Fixed parameters related to the setup and the $K$ cycle (target tolerance on each level, etc.) are described in Table II. We have standardized on $L_s = 16$ throughout this investigation, though explorations for other values of $L_s$ are discussed in Sec. IV.

Inspired by the work of [13], we tested adding a deflation step to the coarsest level, resulting in a four-level algorithm. This is important for addressing critical slowing down on

TABLE II.   Relevant fixed parameters we use for our $K$ cycle. For consistency, we use the same setup parameters throughout the procedures described in this paper. For the setup we have tuned the Wilson operator to the critical mass.

|  | Parameter |  |
|---|---|---|
| Setup | Setup operator | Normal operator, $D_W D_W^\dagger$ |
|  | Setup solver | CG |
|  | Max iterations | 250 |
|  | Max residual tolerance per null vector | $10^{-4}$ |
|  | Number of null vectors, level 1 ($n_{vec}^1$) | 8 |
|  | Size of aggregate block, level 1 | $4^2$ |
|  | Number of null vectors, level $l > 1$ ($n_{vec}^l$) | 12 |
|  | Size of aggregate block, level $l > 1$ | $2^2$ |
|  | Number of levels $l_{max}$ | 3 |
| Solver, level 1 | Operator | $D_{PV}^\dagger D_{DW}$ |
|  | Restart length of GCR | 16 |
|  | Relative residual tolerance | $10^{-10}$ |
|  | GCR iterations for pre-, postsmooth | 0, 8 |
| Solver, level 2 | Operator | $\hat{D}_{PV}^\dagger \hat{D}_{DW}$ |
|  | $\hat{L}_s$ | 4 |
|  | Max iterations | 16 |
|  | Restart length of GCR | 16 |
|  | Relative residual tolerance | 0.25 |
|  | GCR iterations for pre-, postsmooth | 0, 8 |
| Solver, level 3 | Operator | $(\hat{\hat{D}}_{PV}^\dagger \hat{\hat{D}}_{DW})^\dagger (\hat{\hat{D}}_{PV}^\dagger \hat{\hat{D}}_{DW})$ |
|  | $\hat{\hat{L}}_s$ | 4 |
|  | Solver | CGNR |
|  | Relative residual tolerance | 0.05 |
|  | Maximum iterations | 1024 |
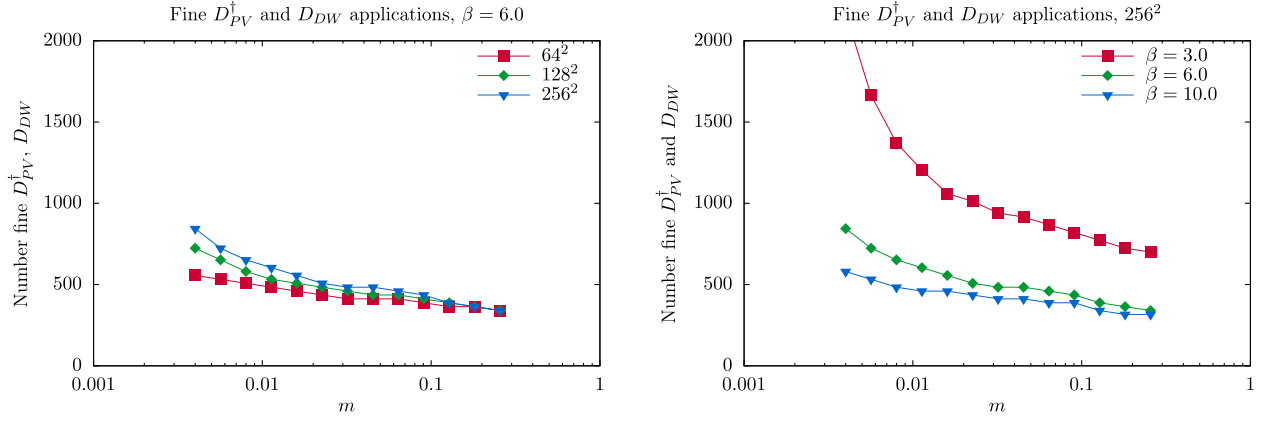
FIG. 7.    The number of applications of the fine domain wall operator (separately $D_{\mathrm{DW}}$ and $D_{\mathrm{PV}}^{\dagger}$) within an MG-preconditioned solve as a function of mass. On the left, we consider fixed $\beta = 6.0$, and on the right, fixed volume $256^2$.

the coarsest level, as has been also seen in 4D studies for Wilson and staggered fermions. For conciseness of presentation we do not explore this past the results given in Table I.

To study the viability of our MG algorithm, we consider sweeps in input fermion mass $m$ at both fixed $\beta = 6.0$, varying the 2D lattice volume and at fixed volume $256^2$, varying the bare coupling $\beta$. An ideal MG algorithm shifts critical slowing down to the coarsest level, corresponding to mass-independent behavior on the finest and intermediate levels. Further, it should be insensitive to the volume at constant physics. We are interested in an algorithm that works for all reasonable values of $\beta$, however in practice it only needs to works for large $\beta$ approaching the continuum at fixed physical correlation lengths well below the UV cutoff.

### B. Elimination of critical slowing down

In Fig. 7 we consider the number of fine domain wall operator applications as a function of the input fermion mass, which is proportional to the number of outer GCR

iterations. On the left, we see that for fixed $\beta$ critical slowing down has been largely eliminated, albeit there is still a small increase at vanishing mass. On the right, we see that for a fixed volume, the MG algorithm shows an improved fermion mass independence in the approach to the continuum limit. The algorithm is unsuccessful for our coarsest $\beta = 3.0$, however this corresponds to a gauge correlation length of $l_{\sigma} \approx 2.4$, which is pushing into an unphysical regime. Similar effects have been noticed in staggered MG [19]. As noted in Table I, preliminary investigations of deflation on the coarsest level lead to a further reduction in iteration count and, by extension, operator application count on the fine level, though not a complete elimination of the increase as a function of decreasing mass.

#### 1. Intermediate level

In Fig. 8 we consider the average number of GCR iterations on the intermediate level as a function of the input fermion mass. We see that the average iteration count is roughly independent of the coupling $\beta$ and the volume,
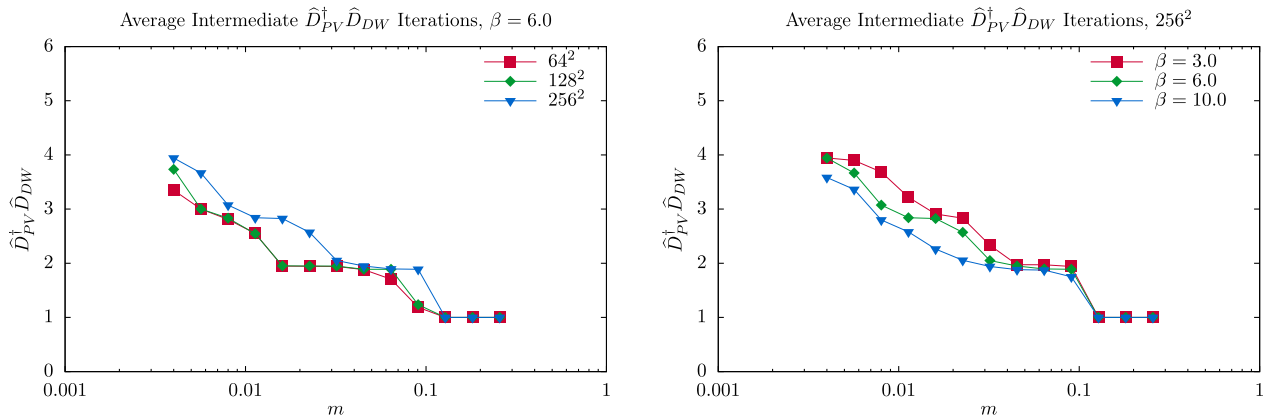


FIG. 8.    The average number of iterations for the inner Krylov solve as a function of mass. On the left, we consider fixed $\beta = 6.0$, and on the right, fixed volume $256^2$. Note that this is a log-linear plot.
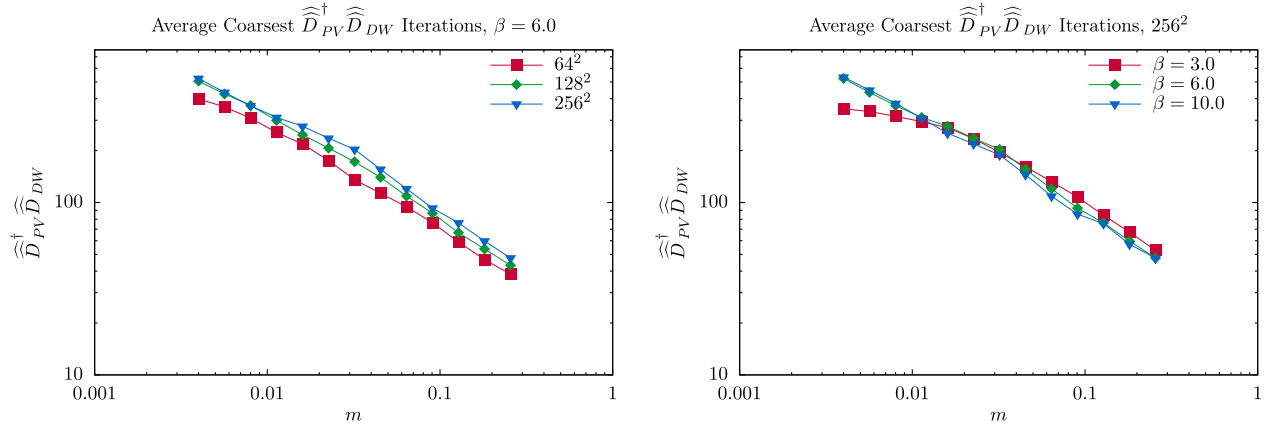
FIG. 9. The average number of iterations of CGNR on the coarsest level for, on the left, fixed $\beta$ and on the right, fixed volume. Note that this is a log-log plot.

which is encouraging. There is a weak mass dependence to the iteration count, however we note that the growth in iteration count appears to be weaker than power law. This is a significant improvement over the power-law dependence that is traditional of critical slowing down. Preliminary investigations of deflation on the coarsest level lead to a reduction in iteration count on the intermediate level, though not a complete elimination of the increase as a function of decreasing mass.

### 2. Coarsest level

In the previous two paragraphs we have demonstrated the elimination of critical slowing down from the finest and the intermediate level. This is because critical slowing down has been shifted to the coarsest level. In Fig. 9 we consider the average number of conjugate gradient on the normal residual (CGNR) iterations on the coarsest level as a function of the input fermion mass. In contrast to plots for the fine and coarse levels, here we present this data on a log-log plot to examine power behavior. In both the left and right panels, we see behavior consistent with power-law divergence of the iteration count independent of volume and $\beta$, showing that critical slowing down has been successfully shifted to the coarsest level. As has been seen in studies with twisted clover and HISQ fermions in 4D, this final critical slowing down can be efficiently eliminated by deflation.

### 3. Comparison with direct solve

In the previous paragraphs we have demonstrated a MG algorithm which shifts critical slowing down from the finest level down to the coarsest level. In Fig. 6, we can see a stark contrast in behavior between our MG algorithm and CGNR directly on the domain wall operator. While the number of fine domain wall operator applications scales with only weak mass dependence in the case of our MG algorithm, there is a strong power-law dependence present for CGNR.

This is the critical slowing down which has shifted to the coarsest level in our MG algorithm.

We also considered applying the BiCGStab-$l$ Krylov solver directly to $D_{\mathrm{PV}}^{\dagger} D_{\mathrm{DW}}$. We can do this because, in contrast to $D_{\mathrm{DW}}$ in isolation, $D_{\mathrm{PV}}^{\dagger} D_{\mathrm{DW}}$ obeys the half-plane condition as proven in Appendix B. While this approach unsurprisingly still demonstrates critical slowing down, there is a marked reduction in fine operator applications. This could be of immediate use for computing domain wall propagators for four-dimensional QCD.

## IV. DISCUSSION

The MG algorithm described above, while generally successful, does introduce several components that are worth understanding better and very likely can lead to further improved performance. Even if we were to use the full effective overlap operator, $D_{\mathrm{PV}}^{-1} D_{\mathrm{DW}}$, the assumption that we need only prolong and restrict the boundary mode when going between levels is nontrivial. That this continues to be adequate with our approximation, $D_{\mathrm{PV}}^{\dagger} D_{\mathrm{DW}}$ is even more surprising. Also the benefit of reducing $L_s$ on the coarsened levels begs a better understanding. Here is our initial effort to explore these issues.

As we did in our prior paper on the staggered multigrid [19], we begin by studying the spectrum of our approximate operator and its "coarsened" version to glean some intuition. While we find a strong overlap of physical low eigenvalues between the fine and the coarsened operator, we also find that the coarsened operator includes additional spurious small eigenvalues. Unlike the naïve formulation of MG for staggered fermions, these modes do not appear to undermine the success of the algorithm. To probe this phenomena, we consider the local colinearity and the oblique projector as monitors for the quality of our MG preconditioner. We see that transferring only the boundary component of the fine residual and the coarse error
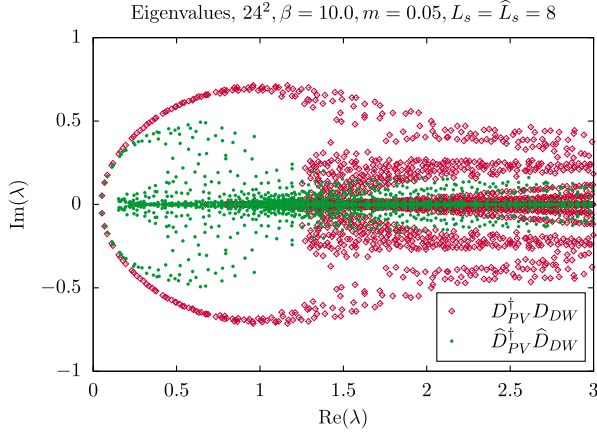
FIG. 10. A representative spectrum of $D_{\mathrm{PV}}^{\dagger}D_{\mathrm{DW}}$ and its coarsened version $\hat{D}_{\mathrm{PV}}^{\dagger}\hat{D}_{\mathrm{DW}}$ at fixed $L_s = \hat{L}_s = 8$.

correction is essential for the success of our multigrid algorithm, and present a physical argument for why this "cures" the problem introduced by the spurious small eigenvalues. In the free field limit presented in detail in Appendix A, we note that the low momentum modes to order $\mathcal{O}(p^2)$ are fixed for any $L_s \geq 2$, which maybe an indication of the underlying mechanism. We investigated reducing $L_s$ on the coarser levels, finding that this reduction leads to an improved algorithm relative to using the fine $L_s$ on the coarsened levels.

### A. Boundary-only transfer operator

In Fig. 10, we see that the spectrum of the coarsened operator $\hat{D}_{\mathrm{PV}}^{\dagger}\hat{D}_{\mathrm{DW}}$ relative to the spectra for the fine operator $D_{\mathrm{PV}}^{\dagger}D_{\mathrm{DW}}$ introduces a large number of nearly real low modes. The problem resembles the spurious small eigenvalues that plagued the direct application of Galerkin

projection to the staggered operator prior to the Kähler-Dirac preconditioning.

In this instance we posit that the saving grace for the domain wall operator is that the low modes of $D_{\mathrm{PV}}^{-1}D_{\mathrm{DW}}$ are bound to the chiral walls, while higher modes bleed more dominantly into the bulk as suggested by Eq. (15). Based on this we posit that projecting only the boundary modes between levels acts as a filter against the bulk spurious modes.

In Sec. II C we noted two ways to formulate the transfer operator between levels. The method we utilize across this paper is only restricting and prolonging the boundary mode as given in Eqs. (36) and (37), respectively. Another formulation would be to repeat the 2D prolongator/restrictor across the bulk dimension.

As a quantitative approach to this study we follow the investigations of the staggered MG paper [19]. Consider the normalized right eigenpairs of the fine operator, $(\lambda, v_\lambda)$. Given these we inspect both the *local colinearity*, a measure of preserving the low eigenspace in the least-squares sense, defined by

$$\|(1 - PR)v_\lambda\|_2,\tag{41}$$

and the *oblique projector*, defined by

$$\|(1 - P(\hat{D}_{\mathrm{PV}}^{\dagger}\hat{D}_{\mathrm{DW}})^{-1}RD_{\mathrm{PV}}^{\dagger}D_{\mathrm{DW}})v_\lambda\|_2,\tag{42}$$

which quantifies the reduction or enhancement of a given error component for a magnitude less than or greater than one. While an error enhancement is not inherently a problem, a very large error enhancement requires a prohibitively expensive compensation at the smoother step.

In Fig. 11, we consider the local colinearity and the oblique projector for a given configuration. On the

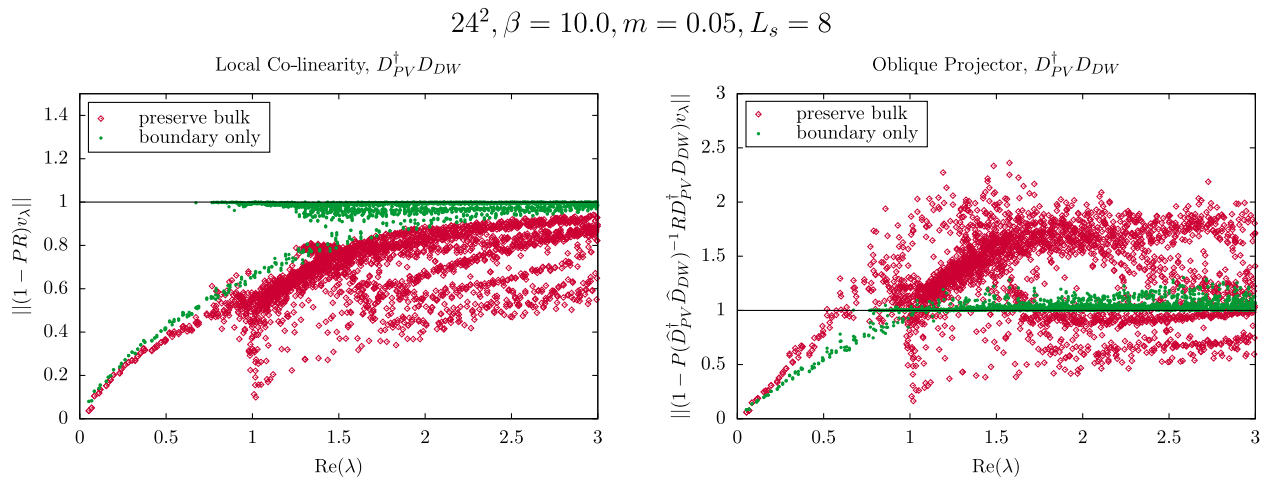$$24^2, \beta = 10.0, m = 0.05, L_s = 8$$



FIG. 11. On the left, a measurement of local colinearity, and on the right, the effect of the oblique projector on eigenvectors. Both figures compare a prolongator and restrictor which transfer the boundary and bulk vs just the boundary. The parameters for the MG aggregation are given in Table II.
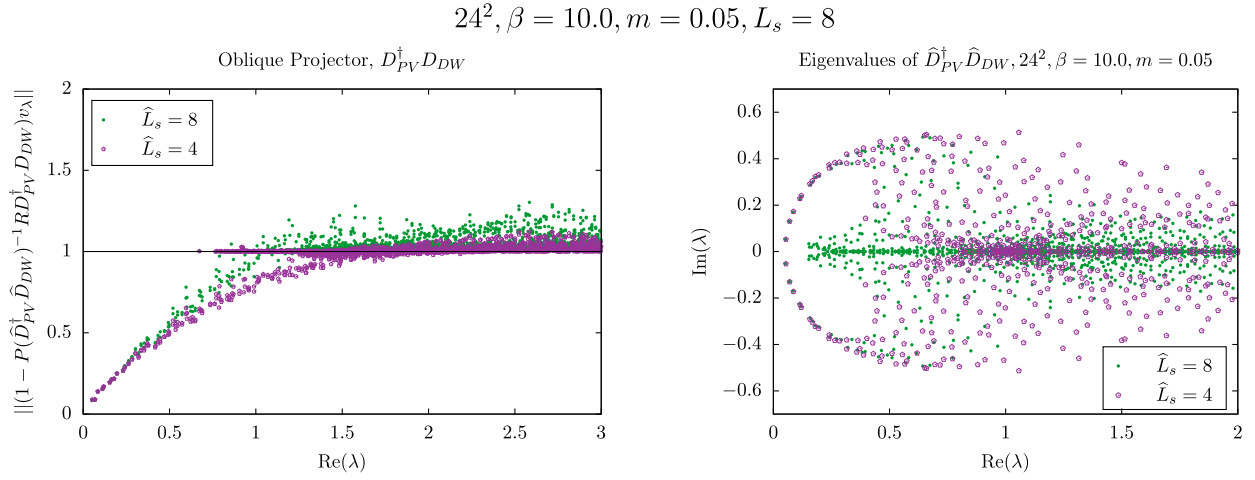
$$24^2, \beta = 10.0, m = 0.05, L_s = 8$$



FIG. 12. On the left, the effect of the oblique projector for a boundary-only transfer operator for two value of the coarse $\hat{L}_s$: the original 8, and a reduced 4. On the right, a comparison of the coarse spectrum for each of these choices of $\hat{L}_s$.

left-hand side, we see that the local colinearity is smaller in magnitude when prolongating and restricting the entire boundary and bulk compared with just transferring the boundary. This is not surprising as the boundary-only transfer operator by construction has a smaller span than the full transfer operator. On the other hand, the oblique projector using the boundary-only transfer operator is generally smaller in magnitude than the full transfer operator. This is a good indicator that the coarsening prescription given in the previous section combined with a transfer operator acting only on the boundary modes leads to an improved MG algorithm.

### B. Reduction of coarse level $L_s$

An additional benefit of using a transfer operator which only acts on the boundaries is it gives us the flexibility to tune $L_s$ between levels. We will investigate this by two avenues. First, we will consider the behavior of the local colinearity and the oblique objector for two different values of the coarse $L_s$. Next, we will perform an explicit test of domain wall MG for a large value of the outer $L_s$, with a range of fixed smaller $L_s$ values on all coarser levels. We illustrate the oblique projector on the left-hand side of Fig. 12. We see that a reduced coarse $\hat{L}_s = 4$ behaves at least as well as maintaining a constant $\hat{L}_s = 8$. In addition, for larger values of the fine eigenvalue (i.e., higher momentum and bulk modes), the error enhancement is further suppressed. This may be related to the spectrum of the reduced $\hat{L}_s$ operator on the right-hand side of Fig. 12. We see that this operator does suffer from fewer spurious small eigenvalues, leading to a reduced risk of error enhancement. Following this investigation of the oblique projector and the spectrum on a small configuration, we have studied the performance of a solve on a $128^2$ configuration with a representative $\beta = 6.0$. Unlike our previous studies, we have chosen a large $L_s = 32$ for the

outermost solve. For simplicity we chose a fixed reduced $L_s$ for both the intermediate and coarsest level.

We see in Fig. 13 that reducing the $L_s$ between the fine and the intermediate level leads to a perfectly well-behaved preconditioner. This is impressive for two reasons. One, a reduction in $L_s$ does lead to an enhancement in chiral symmetry breaking, suggesting that the intermediate and coarsest levels would not accurately capture the low modes of the fine level. This does not appear to be a problem. Two, a reduced $L_s$ leads to enhanced stability for very small masses. There also may be a side benefit of fewer spurious small modes for the coarsest operator with reduced $L_s$.

We are encouraged that reducing $L_s$ on the intermediate and coarsest level leads to an improved algorithm, and it is what informed the formulation studied in Sec. III. Of course, there is a wider parameter space we could explore in
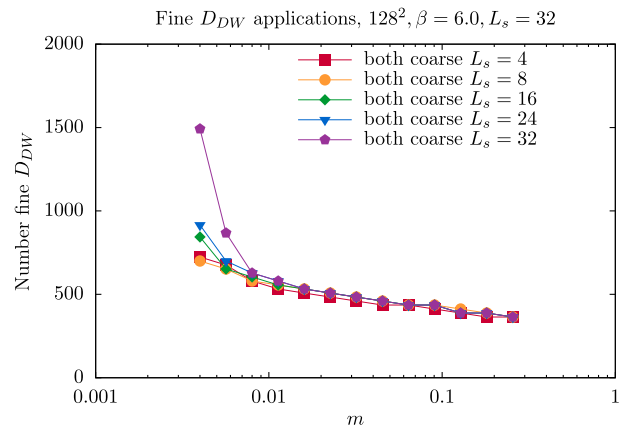


FIG. 13. The number of applications of the fine domain wall operator ($D_{DW}$ and $D_{PV}^\dagger$ each count as one) per MG-preconditioned solve as a function of mass for fixed $\beta = 6.0$ and volume $128^2$. We vary the intermediate and coarse $L_s$ from 4 to 32. $M_5$ remains fixed.

this study: varying the length of the extra dimension separately for each level, tuning $M_5$, tuning $m$ to approximately constant $m + m_{\rm res}$, deflating the coarsest level to avoid a large iteration count. In light of our success simply reducing $L_s$ and making no other changes, we defer such in depth investigations to a study in four dimensions.

## V. CONCLUSION

We have presented a new approach to formulating an MG solver for domain wall fermions. This is a critical step to realizing the full benefit of this chiral formulation which is theoretically superior but more computationally demanding than the Wilson and staggered discretizations. For clarity of exposition our formalism was restricted to the Shamir version of domain wall and for easy of development and testing restricted to domain wall fermions for the 2D two-flavor Schwinger model. Neither of these choices are fundamental. These results convince us that this is a solution to a long sought fully recursive MG algorithm for domain wall fermions that can eliminate critical slowing down approaching the continuum limit for small fermion masses.

As was the case with prototyping MG solvers for the Wilson [7] and staggered [19] discretizations, the next step is to develop software and optimize performance for the Dirac solver for lattice QCD. We anticipate a range of algorithmic embellishments and software methods to optimize such an algorithm for use on exascale-trajectory machines in both the weak- and strong-scaling regimes.

One salient feature is important to emphasize. The projection and prolongation only requires finding the near null space of the 4D Wilson kernel, saving computational overhead and memory occupancy relative to the naïve cost of the 5D domain wall operator. The fact that there is no expansion of the null space relative to Wilson MG due to the *heavy flavors* in the extra dimension is very good news. We foresee that this approach will be effective even for work flows that require few solves per gauge field, e.g., hybrid Monte Carlo gauge evolution.

This points to another benefit. The basic method presented here for the Shamir implementation applies equally well to Möbius, Zolotarev, Borici, etc., formalism via the domain wall/Pauli-Villars factorization in BNO. The only requirement is that each factor is a linear functional of the Wilson kernel. As a consequence it should also be straight forward to generalize our algorithm directly to the overlap operator itself. To appreciate the full landscape consider the large class of chiral fermion methods presented by Edwards, Joo, Kennedy, Orginos, and Wenger in Ref. [31]. In all of these the sign function $\text{sign}[H]$ in the overlap operator

$$D_{\rm ov} = \frac{1+m}{2} + \frac{1-m}{2}\gamma_5\text{sign}[H] \qquad (43)$$

must be approximated in a variety of ways as functional of the Wilson operator: $\text{sign}[H] \simeq \epsilon[D_W]$ using, for example, rational approximations such as Padè expansions, partial fractions, continued fractions, etc. The coarsening step would then act by projecting the 4D Wilson kernel into a near null space *before* building the $\text{sign}[H]$ function,

$$\epsilon[D_W] \to \epsilon[\widehat{D_W}] = \epsilon[\mathbb{P}^\dagger D_W \mathbb{P}]. \qquad (44)$$

As with Eq. (34), the kernel projection does not commute with the effective chiral operator: $\epsilon[\mathbb{P}^\dagger D_W \mathbb{P}] \neq \mathbb{P}^\dagger \epsilon[D_W]\mathbb{P}$. From this vantage point, our domain wall MG algorithm implementation is a special case using the domain wall/Pauli-Villars factorization in Eq. (35) to identify an effective overlap operator.

Clearly there is a larger landscape of MG algorithms for chiral fermion operators to explore. Optimizations will depend on the specific applications and target architectures. We anticipate that alternative implementations of the MG solver for domain wall fermions in four dimensions [11,12] may contribute to further optimizations. We leave the detailed study of these generalizations and optimizations for MG for domain wall and overlap chiral fermions to future investigations.

## APPENDIX A: LOW MOMENTUM EXPANSIONS FOR DOMAIN WALL FERMIONS

The free theory, setting $U = 1$, can be expanded and diagonalized in momentum space for finite $L_s$, giving valuable guidance to our MG construction. First consider our approximation, $D_{\rm PV}^\dagger \simeq D_{\rm PV}^{-1}$, which we will prove to valid up to error $\mathcal{O}(p^4)$ at finite $L_s$. Indeed the Pauli-Villars operator is especially simple because it is antiperiodic in $L_s$, and thus even with interacting fields can be diagonalized via Fourier modes giving

$$D_{\rm PV}(U, p_5) = \gamma_5\sin(p_5) + 1 - \cos(p_5) + M_5 + D_W(U, 0), \qquad (A1)$$

in terms of $d \times d$ space-time blocks for the Wilson operator $D_W(U, 0)$. Setting $M_5 = -1$,

$$D_{\rm PV}(U, p_5) = -e^{-ip_5\gamma_5} + D_W(U, 0), \qquad (A2)$$

we note that the first term by in isolation gives a circle of eigenvalues in the complex plane. Moreover, as is apparent in Fig. 1, this basic pattern persists even with nontrivial gauge fields. In the free-field limit, we can further diagonalize in space-time Fourier modes, $p_\mu$, giving

$$\tilde{D}_{\mathrm{PV}}(p_\mu, p_5) = i \sum_{M=1}^{5} \gamma_M \sin(p_M) + \sum_{M=1}^{5} 2 \sin^2(p_M/2) + M_5,$$

(A3)

where the summation includes $\gamma_5$ and $p_5$. The free normal operator is

$$\tilde{D}_{\mathrm{PV}}^\dagger \tilde{D}_{\mathrm{PV}} = \sum_{M=1}^{5} \sin^2(p_M) + \left( \sum_{M=1}^{5} 2\sin^2(p_M/2) + M_5 \right)^2,$$

(A4)

diagonal in spin structure. In the case of $M_5 = -1$, the low momentum expansion of this operator is $\tilde{D}_{\mathrm{PV}}^\dagger \tilde{D}_{\mathrm{PV}} = M_5^2 + \mathcal{O}(p_M^4)$. This gives

$$\tilde{D}_{\mathrm{PV}}^\dagger = \tilde{D}_{\mathrm{PV}}^{-1} + \mathcal{O}(p^4).$$

(A5)

Indeed more generally for $M_5 = -1$ one can prove that $D_{\mathrm{PV}}^\dagger \tilde{D}_{\mathrm{PV}}$ is bounded below by 1, i.e., at the lattice cutoff scale $1/a^2$. With nontrivial gauge fields the approximation $\tilde{D}_{\mathrm{PV}}^\dagger \approx \tilde{D}_{\mathrm{PV}}^{-1}$ holds qualitatively particularly when $M_5$ is appropriately tuned.

The next step is to compare the low eigenspectra of the preconditioned operator $D_{\mathrm{PV}}^{-1} D_{\mathrm{DW}}(m)$ and the approximation $D_{\mathrm{PV}}^\dagger D_{\mathrm{DW}}(m)$. From Fig. 5 we note the near exact coincidence of small eigenvalues even in the presence of gauge fields. We can elucidate this in the free-field limit.

We cannot simultaneously diagonalize $D_{\mathrm{PV}}$ and $D_{\mathrm{DW}}(m)$ in the bulk dimension due to the difference in boundary conditions, which complicates the analysis. In this case we take advantage of low-momentum perturbation theory in the space-time dimensions. Given $\tilde{D}_{\mathrm{DW}}(p_\mu, m)_{ss'}$, we expand both it and the Pauli-Villars factor at low $d$-momenta, take the appropriate products and solve the characteristic polynomial for the eigenvalues. The pairing of low modes gives a complex square root singularity in the complex plane, leading to the circular structure of the low spectrum as evident in Fig. 5. The results from this approach are given in the text in Eqs. (25) and (26). Comparing the exact vs the approximate low spectra we see they are identical up to a six-dimensional operator $\mathcal{O}(a^2 m p^2)$.

We additionally note that the low momentum expansion of the effective overlap operator,

$$D_{\mathrm{ov}}(m) = \frac{1+m}{2} + \frac{1-m}{2} \gamma_5 \epsilon_L[H],$$

(A6)

where

$$\epsilon_L[H] = \frac{(1-H)^L - (1+H)^L}{(1-H)^L + (1+H)^L}$$

(A7)

has universal coefficients even at finite $L_s$ up to $\mathcal{O}(p^n)$ for $L_s > n$. As a result, the eigenvalues are also equivalent in perturbation theory up to that order. For example, consider for $d = 2$ the roots of the polynomial, $\lambda^2 - \lambda \mathrm{tr} + \det = 0$ in terms of the trace and determinant. For the Shamir kernel, Evaluating this for the Shamir kernel, $H = \gamma_5 D_W(M_5)/(2 + D_W(M_5))$, this is given by

$$\det = \lambda_+ \lambda_- = p_x^2 + p_y^2 - (4/3)[(p_x^2 + p_y^2)^2 - p_x^2 p_y^2] \quad \text{(A8)}$$

and

$$\mathrm{tr} = \lambda_+ + \lambda_- = 2(p_x^2 + p_y^2),$$

(A9)

up to $\mathcal{O}(p^2)$, again universal for any $L_s \geq 2$, resulting in the eigenvalues

$$\lambda_\pm = \pm i \sqrt{p^2(1 - p^2) + 2 p_x^2 p_y^2/3} + p^2,$$

(A10)

to the same order. To include nonzero mass we can use the mass shift identity for eigenvalues, $\lambda^0 \to \lambda = m + (1-m)\lambda^0$, from Eq. (3) to get

$$\lambda_\pm = m + p^2 \pm i(1-m)\sqrt{p^2(1-p^2) + 2 p_x^2 p_y^2/3} - m p^2,$$

(A11)

which up to rescaling by $1/(1-m)$ gives $\lambda_\pm \sim m_q + p^2 \pm i\sqrt{p^2}$ in agreement with Eq. (25) in the text up to a dimension-three scaling operator of $\mathcal{O}(m p^2)$. Higher-order expansions in 2D and 4D effective overlap operators are easily shown in MATHEMATICA to follow this expansion invariance up to $\mathcal{O}(p^{L_s})$. We believe the fact that the quadratic form is independent of $L_s \geq 2$ even for our approximate $D_{\mathrm{PV}}^\dagger D_{\mathrm{DW}}(m)$ effective overlap operator is at the core the effectiveness of using small $L_s$ for MG iterations on the coarse level.

Finally we note that this expansion is almost certainly a divergent asymptotic series and as such does not by itself lend itself to fix the coefficients in improved higher-order polynomial approximation

$$[D_{\mathrm{PV}} D_{\mathrm{PV}}^\dagger]^{-1} \simeq c_0 + c_1 D_{\mathrm{PV}} D_{\mathrm{PV}}^\dagger + c_2 (D_{\mathrm{PV}} D_{\mathrm{PV}}^\dagger)^2 \cdots$$

(A12)

to $[D_{\mathrm{PV}} D_{\mathrm{PV}}^\dagger]^{-1}$ in Eq. (28) beyond the zeroth order. We see this first-order equivalence is given by $c_i = (1, 0, 0, 0 \cdots)$.

We can take thus further by expanding $[D_{PV}D_{PV}^\dagger]^{-1} = 1/(1 + A_{PV})$ in $A_{PV} = D_{PV}D_{PV}^\dagger - 1$ to show that each term in Eq. (28) corrects the approximation by another two powers of $p$. For example just the second-order expansion, $c_i = (2, -1, 0, 0, \ldots)$, giving the equivalence $D_{PV}^{-1} = D_{PV}^\dagger(2 - D_{PV}D_{PV}^\dagger) + \mathcal{O}(p_\mu^4)$. This polynomial on its own cannot be used in our MG algorithm because $2 - D_{PV}D_{PV}^\dagger$ is not positive definite so coefficient in a polynomial truncation must be weighted appropriately over the entire spectrum.

## APPENDIX B: PROOF THAT THE SPECTRUM HAS A POSITIVE REAL PART

In the full interacting case with gauge fields we consider our approximation $D_{PV}^\dagger D_{DW}$ as $(D_{PV}^\dagger D_{PV})D_{PV}^{-1}D_{DW}$ and note from Eq. (14) that $D_{PV}^{-1}D_{DW}$ has the eigenspectra of the finite-$L_s$ overlap kernel,

$$D_{ov}^{(L_s)}(m) = \frac{1+m}{2} + \frac{1-m}{2}\gamma_5 \epsilon_{L_s}[H_5], \qquad (B1)$$

plus additional 1 eigenvalues. We recall that for even $L_s$, $\epsilon_{L_s}[H_5] \in [-1, 1]$. Applying the Cauchy-Schwarz $\gamma_5 \epsilon_{L_s}[H_5]$ inequality for a normalized vectors, $\langle u|u\rangle = \langle v|v\rangle = 1$, the matrix element is bounded in magnitude,

$$|\langle u|\gamma_5 \epsilon_{L_s}[H_5]|v\rangle| = \sqrt{\langle v|\epsilon_{L_s}[H_5]\epsilon_{L_s}[H_5]|v\rangle} \le 1. \quad (B2)$$

Equivalently this is just a statement of the unitarity bound on matrix elements. By extension, the spectrum of $D_{ov}(m)$ lives inside a circle of radius $\frac{1-m}{2}$ centered at $(\frac{1+m}{2}, 0)$ in the complex plane. Thus both $D_{ov}(m)$ and $D_{PV}^{-1}D_{DW}$ have positive definite real part for $m > 0$.

Let us now consider the eigenvalue problem for the operator $D_{PV}^\dagger D_{DW}$ for *any* right eigenvector $|\lambda\rangle$,

$$D_{PV}^\dagger D_{DW}|\lambda\rangle = re^{i\theta}|\lambda\rangle, \qquad (B3)$$

where we have generally written the eigenvalue $\lambda$ as $re^{i\theta}$. We wish to prove the right-half plane condition $\theta \in (-\pi/2, \pi/2)$. We note we can rewrite the left-hand side of the above system as

$$(D_{PV}^\dagger D_{PV})D_{PV}^{-1}D_{DW}|\lambda\rangle = re^{i\theta}|\lambda\rangle. \qquad (B4)$$

Left multiplying by $(D_{PV}^\dagger D_{PV})^{-1}$ and taking the full matrix element with $\langle\lambda|$ we have

$$\langle\lambda|D_{PV}^{-1}D_{DW}|\lambda\rangle = re^{i\theta}\langle\lambda|(D_{PV}^\dagger D_{PV})^{-1}|\lambda\rangle. \qquad (B5)$$

Since $D_{PV}^{-1}D_{DW}$ satisfies the right half-plane condition and $D_{PV}^\dagger D_{PV}$ is a Hermitian positive operator, this proves the right half-plane condition $\theta \in (-\pi/2, \pi/2)$ for $m > 0$.

[1] K. G. Wilson, Confinement of quarks, Phys. Rev. D **10**, 2445 (1974).

[2] T. Blum, R. Van de Water, D. Holmgren, R. Brower, S. Catterall et al., Report of the Snowmass 2013 Computing Frontier working group on Lattice Field Theory–Lattice field theory for the energy and intensity frontiers: Scientific goals and computing needs, arXiv:1310.6087.

[3] R. C. Brower, C. Rebbi, and E. Vicari, Non-Abelian Projective Multigrid for Lattice Gauge Theory, Phys. Rev. Lett. **66**, 1263 (1991).

[4] K. G. Wilson, The renormalization group: Critical phenomena and the Kondo problem, Rev. Mod. Phys. **47**, 773 (1975).

[5] R. C. Brower, R. G. Edwards, C. Rebbi, and E. Vicari, Projective multigrid for Wilson fermions, Nucl. Phys. **B366**, 689 (1991).

[6] A. Hulsebos, J. Smit, and J. C. Vink, Multigrid inversion of the staggered fermion matrix, Nucl. Phys. B, Proc. Suppl. **20**, 94 (1991).

[7] J. Brannick, R. C. Brower, M. A. Clark, J. C. Osborn, and C. Rebbi, Adaptive Multigrid Algorithm for Lattice QCD, Phys. Rev. Lett. **100**, 041601 (2008).

[8] R. Babich, J. Brannick, R. Brower, M. Clark, T. Manteuffel, S. McCormick, J. Osborn, and C. Rebbi, Adaptive Multigrid Algorithm for the Lattice Wilson-Dirac Operator, Phys. Rev. Lett. **105**, 201602 (2010).

[9] J. Kogut and L. Susskind, Hamiltonian formulation of Wilson's lattice gauge theories, Phys. Rev. D **11**, 395 (1975).

[10] D. B. Kaplan, A method for simulating chiral fermions on the lattice, Phys. Lett. B **288**, 342 (1992).

[11] S. D. Cohen, R. Brower, M. Clark, and J. Osborn, Multigrid algorithms for domain-wall fermions, Proc. Sci., LAT-TICE2011 (**2011**) 030.

[12] P. A. Boyle, Hierarchically deflated conjugate gradient, arXiv:1402.2585.

[13] A. Yamaguchi and P. A. Boyle, Hierarchically deflated conjugate residual, Proc. Sci., LATTICE2016 (**2016**) 374.

[14] H. Neuberger, Vectorlike gauge theories with almost massless fermions on the lattice, Phys. Rev. D **57**, 5417 (1998).

[15] J. Brannick, A. Frommer, K. Kahl, B. Leder, M. Rottmann, and A. Strebel, Multigrid preconditioning for the overlap operator in lattice QCD, Numer. Math. **132**, 463 (2016).

[16] P. Becher and H. Joos, The Dirac-Kähler equation and fermions on the lattice, Z. Phys. C **15**, 343 (1982).

[17] G. T. Bodwin and E. V. Kovács, Equivalence of Dirac-Kähler and staggered lattice fermions in two dimensions, Phys. Rev. D **38,** 1206 (1988).

[18] R. C. Brower, H. Neff, and K. Orginos, The Möbius domain wall fermion algorithm, Comput. Phys. Commun. **220,** 1 (2017).

[19] R. C. Brower, E. Weinberg, M. A. Clark, and A. Strelchenko, Multigrid algorithm for staggered lattice fermions, Phys. Rev. D **97,** 114513 (2018).

[20] J. Schwinger, Gauge invariance and mass. II, Phys. Rev. **128,** 2425 (1962).

[21] A. V. Smilga, Critical amplitudes in two-dimensional theories, Phys. Rev. D **55,** R443 (1997).

[22] Y. Shamir, New domain-wall fermion actions, Phys. Rev. D **62,** 054513 (2000).

[23] R. C. Brower, H. Neff, and K. Orginos, Möbius fermions: Improved domain wall chiral fermions, Nucl. Phys. B, Proc. Suppl. **140,** 686 (2005).

[24] R. Brower, H. Neff, and K. Orginos, Möbius fermions, Nucl. Phys. B, Proc. Suppl. **153,** 191 (2006).

[25] A. Borici, Truncated overlap fermions: The link between overlap and domain wall fermions, NATO Sci. Ser. C **553,** 41 (2000), https://arxiv.org/abs/hep-lat/9912040.

[26] A. Borici, Truncated overlap fermions, Nucl. Phys. B, Proc. Suppl. **83,** 771 (2000).

[27] T.-W. Chiu, Optimal Lattice Domain-Wall Fermions, Phys. Rev. Lett. **90,** 071601 (2003).

[28] T.-W. Chiu, Locality of optimal lattice domain-wall fermions, Phys. Lett. B **552,** 97 (2003).

[29] R. G. Edwards and U. M. Heller, Exact chiral symmetry for domain wall fermions with finite Ls, Nucl. Phys. B, Proc. Suppl. **94,** 737 (2001).

[30] R. G. Edwards and U. M. Heller, Domain wall fermions with exact chiral symmetry, Phys. Rev. D **63,** 094505 (2001).

[31] R. G. Edwards, B. Joo, A. D. Kennedy, K. Orginos, and U. Wenger, Comparison of chiral fermion methods, Proc. Sci., LAT2005 (**2006**) 146.

[32] H. Neuberger, Exactly massless quarks on the lattice, Phys. Lett. B **417,** 141 (1998).

[33] Y. Kikukawa and T. Noguchi, Low energy effective action of domain-wall fermion and the Ginsparg-Wilson relation, Nucl. Phys. B Proc. Suppl. **83,** 630 (2000).

[34] P. H. Ginsparg and K. G. Wilson, A remnant of chiral symmetry on the lattice, Phys. Rev. D **25,** 2649 (1982).

[35] J. Liesen and P. Tichý, Convergence analysis of Krylov subspace methods, GAMM-Mitteilungen **27,** 153 (2004).

[36] M. Benson, Iterative solution of large scale linear systems, M.Sc. Thesis, Lakehead University, Mathematics Report, 1973.

[37] E. Chow, Parallel implementation and practical use of sparse approximate inverse preconditioners with a priori sparsity patterns, Int. J. High Perform. Comput. Appl. **15,** 56 (2001).

[38] H. D. Sterck, U. M. Yang, and J. J. Heys, Reducing complexity in parallel algebraic multigrid preconditioners, SIAM J. Matrix Anal. Appl. **27,** 1019 (2006).

[39] E. Treister and I. Yavneh, Non-Galerkin multigrid based on sparsified smoothed aggregation, SIAM J. Sci. Comput. **37,** A30 (2015).

[40] D. B. Kaplan and M. Schmaltz, Supersymmetric Yang-Mills theories from domain wall fermions, Chin. J. Phys. **38,** 543 (2000), https://arxiv.org/abs/hep-lat/0002030.

[41] C. G. Callan, Jr. and J. A. Harvey, Anomalies and fermion zero modes on strings and domain walls, Nucl. Phys. **B250,** 427 (1985).

[42] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. (Johns Hopkins University Press, Baltimore, 2013).

[43] D. H. Adams, Theoretical Foundation for the Index Theorem on the Lattice with Staggered Fermions, Phys. Rev. Lett. **104,** 141602 (2010).

[44] P. M. Vranas, Chiral symmetry restoration in the Schwinger model with domain wall fermions, Phys. Rev. D **57,** 1415 (1998).