

Adaptive aggregation-based domain decomposition multigrid for twisted mass fermions

Constantia Alexandrou,^{1,2} Simone Bacchio,^{1,3} Jacob Finkenrath,² Andreas Frommer,³
Karsten Kahl,³ and Matthias Rottmann³

¹*Department of Physics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus*

²*Computation-based Science and Technology Research Center, The Cyprus Institute, P.O. Box 27456, 1645 Nicosia, Cyprus*

³*Fakultät für Mathematik und Naturwissenschaften, Bergische Universität Wuppertal, 42119 Wuppertal, Germany*

(Received 17 October 2016; published 12 December 2016)

The adaptive aggregation-based domain decomposition multigrid method [A. Frommer *et al.*, SIAM J. Sci. Comput. **36**, A1581 (2014)] is extended for two degenerate flavors of twisted mass fermions. By fine-tuning the parameters we achieve a speed-up of the order of a hundred times compared to the conjugate gradient algorithm for the physical value of the pion mass. A thorough analysis of the aggregation parameters is presented, which provides a novel insight into multigrid methods for lattice quantum chromodynamics independently of the fermion discretization.

DOI: [10.1103/PhysRevD.94.114509](https://doi.org/10.1103/PhysRevD.94.114509)

I. INTRODUCTION

Lattice quantum chromodynamics (QCD) is computationally demanding due to the fact that a linear system of very large dimension needs to be solved repeatedly. The linear system to be solved is given by

$$D(U, \mu)v = b \quad (1)$$

where $D(U, \mu)$ is the Dirac operator. The Dirac operator depends on the gauge configuration U and the twisted mass (TM) parameter μ , which determines the quark mass. The right-hand side (rhs) b is a known spinor source and v is the solution to the equation. In order to collect enough statistics for the observable under study, Eq. (1) needs to be solved many times, with both varying gauge configurations and the rhs. In fact the so-called point-to-all quark propagator is obtained by solving Eq. (1) for twelve different rhs. Several of these point-to-all propagators are calculated for different space-time positions on each gauge configuration and a large number of configurations is needed to obtain sufficient statistical accuracy. The time for the inversion of the Dirac operator depends on its condition number, which gets worse as the mass of the light quarks or equivalently μ decreases. Therefore it is crucial to use algorithms which are less sensitive to the condition number and adapt them to the specific features of the discretization. This is particularly important for inversions using gauge configurations simulated with up- and down-quark masses fixed to their physical value, where improved solvers are indispensable for speeding up the computations and enabling the accumulation of enough statistics. Nowadays, such simulations are intensively being pursued by a number of lattice QCD collaborations and a lot of effort has been devoted in

speeding up the solvers. A very successful approach has been based on multigrid ideas in preconditioning standard Krylov subspace solvers. There is a number of variant formulations of highly optimized multigrid solvers, which yield improvements of more than an order of magnitude in the case of the Wilson Dirac discretization, as reported in Refs. [1–5]. The most known implementations of multigrid algorithms, which are also available as open source software, are the following: (i) a two-level multigrid approach based on Lüscher's inexact deflation [1] provided in the software openQCD [6]; (ii) a multigrid approach developed in Refs. [7–10], referred to as multigrid with generalized conjugate residual (MG-GCR) which is part of the USQCD package QOPQDP [11]; and (iii) an aggregation-based domain decomposition multigrid approach [4], referred to as DD- α AMG, recently made publicly available in the DDalphaAMG library [12].

Although these solvers have been developed for clover Wilson fermions they can be extended to other fermion discretization schemes as has been done for example for the overlap operator using DD- α AMG [13] or for domain-wall fermions using MG-GCR [14].

In this work we focus on the TM Wilson Dirac operator. This discretization scheme has the advantage that all observables are automatically $\mathcal{O}(a)$ improved when tuned to maximal twist [15], where a is the lattice spacing. This formulation is thus particularly suitable for hadron structure studies, since the probe such as the axial current needs no further improvement in contrast to clover improved fermions. Furthermore, the presence of a finite twisted mass term bounds the spectrum of DD^\dagger from below by a positive term. This avoids exceptional configurations and, at the same time, gives an upper bound to the condition number, improving the convergence of numerical methods used

in lattice QCD. Adding a small twisted mass term to the Wilson Dirac operator is also utilized as a “trick” within simulations with the Wilson Dirac operator. For example, it is exploited in the case of quenched calculations using Wilson fermions [16,17], in the simulation of clover improved Wilson fermions where it is combined with twisted mass reweighting [18,19] or in mass reweighting [20]. Currently, the twisted mass formulation is the discretization scheme adopted by the European Twisted Mass Collaboration for studying a wide range of observables. All these simulations employing twisted mass fermions will substantially benefit from an efficient multigrid method, especially for the analysis at the physical value of the light quark masses.

In this work we extend and optimize the DD- α AMG method to the case of the twisted mass Wilson Dirac operator. Therefore we exploit the Γ_5 -preservation within DD- α AMG which complements the twisted mass formulation. Our analysis shows that firstly the optimal parameters for the method are different from those in the Wilson Dirac case, and secondly that, with appropriate parameter choices, we can gain speed-ups of more than 1 order of magnitude over methods previously used. The latter is similar to what has been observed in the case of clover improved fermions. Determining the appropriate parameter set is crucial since the tuning is a highly nontrivial task, and a complete analysis of how the method behaves for the different values of the parameters is still missing in the literature. We demonstrate in this study how the parameter space can effectively be reduced by tuning first the aggregation parameters. In doing so we also illustrate the robustness of the method. Moreover, our tuning approach can be easily applied to other aggregation-based multigrid solvers and fermion discretization schemes.

The paper is structured as follows: In Sec. II we introduce the twisted mass operator and discuss its properties with respect to multigrid preconditioning. In Sec. III we present our numerical results for twisted mass fermions, including the novel investigation of the aggregation parameters. Finally, in Appendix A we give an overview of the implementation of the method and how the DDalphaAMG library is extended to the case of twisted mass fermions [21].

II. DD- α AMG FOR TWISTED MASS FERMIONS

In this section we discuss the extension of the DD- α AMG solver to the case of the twisted mass fermion discretization scheme of the Dirac operator. For completeness we first review the twisted mass discretization scheme before we outline the various components of the multigrid preconditioner, pointing out the modifications necessary when going from the Wilson formulation to the twisted mass formulation, along the lines of Ref. [4].

We work on a four-dimensional hypercubic lattice defined by

$$\mathcal{V} = \{x = (x_0, x_1, x_2, x_3), 1 \leq x_0 \leq N_T, 1 \leq x_1, x_2, x_3 \leq L\} \quad (2)$$

with N_T being the number of points in the temporal direction and L the number of points in the spatial directions x, y and z . Here and in what follows, the lattice spacing a is set to unity. The lattice volume is given by $V = N_T \cdot L^3$. Fermion fields are defined on the sites of the lattice, and each has four spin and three color degrees of freedom. The overall space is thus

$$\mathcal{V}_s = \mathcal{V} \times \mathcal{S} \times \mathcal{C} \quad (3)$$

where \mathcal{S} denotes the spin space and \mathcal{C} the color space.

A. Wilson twisted mass fermions

The Wilson Dirac operator $D_W = D(m)$ with the clover term can be written as

$$\begin{aligned} (D_W \psi)(x) &= \left((m+4)I_{12} - \frac{c_{sw}}{32} \sum_{\mu, \nu=0}^3 (\gamma_\mu \gamma_\nu) \otimes (Q_{\mu\nu}(x) - Q_{\nu\mu}(x)) \right) \psi(x) \\ &\quad - \frac{1}{2} \sum_{\mu=0}^3 ((I_4 - \gamma_\mu) \otimes U_\mu(x)) \psi(x + \hat{\mu}) \\ &\quad - \frac{1}{2} \sum_{\mu=0}^3 ((I_4 + \gamma_\mu) \otimes U_\mu^\dagger(x - \hat{\mu})) \psi(x - \hat{\mu}), \end{aligned}$$

with m being the mass parameter and c_{sw} the parameter of the clover term. The gauge links $U_\mu(x)$ are $SU(3)$ matrices, and the set $\{U_\mu(x) : x \in \mathcal{L}, \mu = 0, 1, 2, 3\}$ is referred to as a gauge configuration. The γ -matrices act on the spin degrees of freedom of the spinor field $\psi(x)$ and fulfil the anticommutation relation $\{\gamma_\mu, \gamma_\nu\} = 2 \cdot I_4 \delta_{\mu\nu}$ for $\mu, \nu = 0, 1, 2, 3$.

The Wilson Dirac operator satisfies the relation $\Gamma_5 D_W = D_W^\dagger \Gamma_5$ referred to as Γ_5 -Hermiticity, where $\Gamma_5 = I_V \otimes \gamma_5 \otimes I_C$ acts on the space \mathcal{V}_s defined in Eq. (3) as a linear transformation of the spin degrees of freedom at each lattice site. In this paper we use a representation for $\gamma_5 = \gamma_0 \gamma_1 \gamma_2 \gamma_3$, which is diagonal in the spin space

$$\gamma_5 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}. \quad (4)$$

The clover term, $Q_{\mu\nu}$, is given by

$$\begin{aligned}
Q_{\mu\nu}(x) &= U_\mu(x)U_\nu(x+\hat{\mu})U_\mu(x+\hat{\nu})^\dagger U_\nu(x)^\dagger \\
&+ U_\nu(x)U_\mu(x-\hat{\mu}+\hat{\nu})^\dagger U_\nu(x-\hat{\mu})^\dagger U_\mu(x-\hat{\mu}) \\
&+ U_\mu(x-\hat{\mu})^\dagger U_\nu(x-\hat{\mu}-\hat{\nu})^\dagger U_\mu(x-\hat{\mu}-\hat{\nu})U_\nu(x-\hat{\nu}) \\
&+ U_\nu(x-\hat{\nu})^\dagger U_\mu(x-\hat{\nu})U_\nu(x-\hat{\nu}+\hat{\mu})U_\mu(x)^\dagger.
\end{aligned}$$

In the case of the clover improved Wilson Dirac operator, the c_{sw} -term reduces the discretization error from $\mathcal{O}(a)$ to $\mathcal{O}(a^2)$ when c_{sw} is properly chosen.

The twisted mass formulation is a lattice regularization that allows automatic $\mathcal{O}(a)$ improvement by tuning only one parameter, namely the bare untwisted quark mass needs to be tuned to the so-called critical mass. This formulation is particularly appropriate to hadron structure studies since the renormalization of local operators is significantly simplified with respect to the standard Wilson regularization. In the continuum, the twisted mass formulation is equivalent to the standard QCD action in a different basis. A mass term of the form $i\mu\tau_3 \otimes \Gamma_5$ can be added to the standard quark mass, where $\tau_3 = \text{diag}(1, -1)$ is the third Pauli matrix acting on a two-dimensional flavor space [15,22]. By performing a suitable chiral transformation of the quark fields the mass term can be rewritten in the standard form. The twisted mass theory becomes nontrivial on the lattice since Wilson fermions explicitly break chiral symmetry. We denote the twisted mass parameter by $\mu \in \mathbb{R}$ and apply the twist in the flavor space of the up (u) and down (d) quark. Then the twisted mass term acts with a positive shift given by $i\mu\Gamma_5$ on the u-quark operator and with a negative shift given by $-i\mu\Gamma_5$ on the d-quark operator. The twisted mass term breaks the isospin symmetry between the u and the d quark explicitly, which vanishes in the continuum limit. In the flavor space the operator applied to a spinor field is given as

$$\begin{bmatrix} D(\mu) & 0 \\ 0 & D(-\mu) \end{bmatrix} \begin{bmatrix} \psi_u \\ \psi_d \end{bmatrix} = \begin{bmatrix} (D_W(\psi_u)(x) + i\mu\Gamma_5\psi_u(x)) \\ (D_W(\psi_d)(x) - i\mu\Gamma_5\psi_d(x)) \end{bmatrix}, \quad (5)$$

where

$$D(\mu) = D_W + i\mu\Gamma_5 \quad (6)$$

is what we refer to as the twisted mass Wilson Dirac operator defined on the space \mathcal{V}_s . Adding a clover term reduces the isospin breaking induced by the twist.

Due to Γ_5 -Hermiticity, the symmetrized Wilson Dirac operator $H_W = \Gamma_5 D_W$ is Hermitian (and indefinite), such that we have

$$H_W = V\Lambda V^\dagger,$$

where the diagonal matrix Λ contains the eigenvalues λ_j of H_W (which are all real) and the unitary matrix V of the

corresponding eigenvectors. The symmetrized twisted mass operator $H(\mu) = \Gamma_5 D(\mu) = H_W + i\mu$ thus satisfies

$$H_W + i\mu = V\Lambda V^\dagger + i\mu = V e^{i\Theta} \sqrt{\Lambda^2 + \mu^2} V^\dagger, \quad (7)$$

where the diagonal matrix $e^{i\Theta}$ contains the complex phases θ_j of the eigenvalues $\lambda_j + i\mu$ and $\sqrt{\Lambda^2 + \mu^2}$ their absolute values. For the nonsymmetrized twisted mass operator $D(\mu)$, analogously to the Wilson case [4], this gives the singular value decomposition

$$D(\mu) = \Gamma_5 V e^{i\Theta} \sqrt{\Lambda^2 + \mu^2} V^\dagger = U \sqrt{\Lambda^2 + \mu^2} V^\dagger \quad (8)$$

with $U = \Gamma_5 V e^{i\Theta}$ and V being unitary. The smallest singular value $\sqrt{\lambda_i^2 + \mu^2}$ is thus not smaller than μ , which shows that a nonzero value of μ protects the twisted mass operator $D(\mu)$ from being singular, unlike the Wilson Dirac operator where this can happen for small quark masses m .

Similarly, for the squared twisted mass operator

$$\begin{aligned}
D^\dagger(\mu)D(\mu) &= (D^\dagger - i\mu\Gamma_5)(D + i\mu\Gamma_5) \\
&= D^\dagger D + \mu^2 = H_W^2 + \mu^2 \quad (9)
\end{aligned}$$

we have

$$H_W^2 + \mu^2 = V(\Lambda^2 + \mu^2)V^\dagger, \quad (10)$$

the eigenvalues of which are bounded from below by μ^2 .

Attaining automatic $\mathcal{O}(a)$ improvement on the lattice can be accomplished by tuning the partial-conserved axial current (PCAC) pion mass to 0. This corresponds to setting the angle Θ of the axial transformation $e^{i\Theta\tau_3\Gamma_5}$ to $\pi/2$. Then the renormalized light quark mass is directly proportional to the twisted mass parameter μ with $m_R = Z_P^{-1}\mu$ and Z_P being the pseudoscalar renormalization constant. The breaking of the isospin symmetry in the twisted mass formulation results in the neutral pion being lighter than the charged pion. This slows down or even prohibits simulations for light quark masses close to the physical value. By adding the clover term to the action, the critical twisted mass value for the light quarks is significantly reduced compared to simulations without this term. This allows simulations at the physical point with a value of the lattice spacing around $a = 0.1$ fm or even larger [23].

A special property of the twisted mass operator is that at maximal twist the region just above μ^2 is densely populated with the eigenvalues of the squared operator, cf. Eq. (9). We illustrate this in Fig. 1, which displays a histogram of the (scaled) ensemble averaged moduli of the eigenvalues of the nonsquared symmetrized even-odd reduced (or preconditioned) twisted mass Dirac operator \hat{H} . The eigenvalues of the operator are measured on an

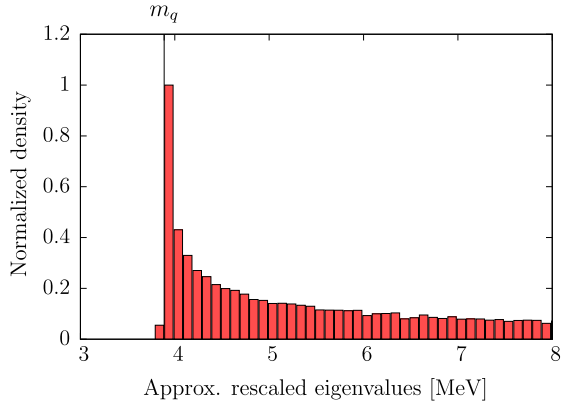


FIG. 1. The density of the approximated rescaled eigenvalues of the Hermitian even-odd reduced twisted mass Dirac operator measured on gauge configurations of the physical ensemble *cA2.09.48* (see Sec. III). The quark mass is given by $m_q = 3.89$ MeV in the $\overline{\text{MS}}$ -scheme.

ensemble simulated at a physical value of the light quark mass doublet, which we refer to as the *physical ensemble*. \hat{H} is obtained using an even-odd ordering of the lattice sites such that

$$D = \begin{bmatrix} D_{oo} & D_{oe} \\ D_{eo} & D_{ee} \end{bmatrix}$$

where D_{oo} and D_{ee} are diagonal in spinor space. \hat{H} is then given as $\hat{H} = \Gamma_5 \hat{D}$ with

$$\hat{D} = (D_{ee} - D_{eo} D_{oo}^{-1} D_{oe}). \quad (11)$$

For $c_{sw} = 0$, i.e. without the clover term, the spectrum of \hat{D} is directly connected to the spectrum of the full operator D , e.g. in the case of the small eigenvalues we have

$$\frac{\lambda_D}{m+4} = 1 - \sqrt{1 - \lambda_{\hat{D}}/(m+4)} \approx \frac{1}{2} \frac{\lambda_{\hat{D}}}{m+4} \quad (12)$$

with λ_D and $\lambda_{\hat{D}}$ being an eigenvalue of D and \hat{D} , respectively. Although this relation does not hold exactly for the Hermitian even-odd reduced twisted mass Dirac operator, we found that numerically this relation still holds approximately for the eigenvalues close to μ . We find that the largest relative deviation of the smallest eigenvalue to the approximated cutoff 2μ is given by $|\lambda_{\min} - 2\mu|/2\mu < 0.0005$ MeV. Thus we rescale the spectrum by a factor 2 and relate it to the energy scale of the $\overline{\text{MS}}$ -scheme defined at 2 GeV. The eigenvalue density is shown in Fig. 1. In contrast to the spectrum of the Wilson Dirac operator shown in Refs. [1,24], the density of the eigenvalues increases close to the physical quark mass.

Since $D(\mu)$ is non-normal, the left and right eigenvectors differ. For the twisted mass Dirac operator, the left

eigenvectors of the u quarks, represented by $D(\mu)$, are connected to the right eigenvectors of the d quarks, represented by $D(-\mu)$. If $\varphi_{j,L}^u$ and $\varphi_{j,R}^u$ are left and right eigenvectors of $D(\mu)$, respectively, with corresponding eigenvalue λ_j^u , then due to Γ_5 -Hermiticity, we have

$$\begin{aligned} \lambda_j^u &= \varphi_{j,L}^{u\dagger} D(\mu) \varphi_{j,R}^u = (\varphi_{j,R}^{u\dagger} (D(\mu))^\dagger \varphi_{j,L}^u)^\dagger \\ &= (\varphi_{j,R}^{u\dagger} \Gamma_5 D(-\mu) \Gamma_5 \varphi_{j,L}^u)^\dagger \\ &= (\varphi_{j,L}^{d\dagger} D(-\mu) \varphi_{j,R}^d)^\dagger = \text{conj}(\lambda_j^d). \end{aligned} \quad (13)$$

Thus, eigenpairs of $D(-\mu)$ are connected to the eigenpairs of $D(\mu)$ by the transformations $\lambda_j^d = \text{conj}(\lambda_j^u)$, $\varphi_{j,L}^d = \Gamma_5 \varphi_{j,L}^u$ and $\varphi_{j,R}^d = \Gamma_5 \varphi_{j,R}^u$.

B. Multigrid preconditioning for twisted mass fermions

Our task is to solve the linear system

$$D(\mu)\psi = (D + i\mu\Gamma_5)\psi = b \quad (14)$$

for the Wilson twisted mass operator $D(\mu)$. The idea is to precondition a flexible iterative Krylov solver at every iteration step using a multigrid preconditioner.

Let ψ denote the current approximate solution to Eq. (14); then the corresponding error is given by $\epsilon = D^{-1}(\mu)b - \psi$, and the residual r satisfies

$$r = b - D(\mu)\psi = D(\mu)\epsilon. \quad (15)$$

In lattice QCD, preconditioners like the Schwarz alternating procedure (SAP [25]) efficiently reduce error components belonging to ultraviolet (UV) modes, i.e. error components belonging to eigenvectors to large eigenvalues. Thus the error ϵ is then dominated by infrared (IR) modes, i.e. eigenvectors corresponding to small eigenvalues. For larger volumes, the increasing number of IR modes slows down the preconditioned Krylov method.

Multigrid methods deal efficiently with both IR and UV modes, independently of the volume size. Here we focus on the DD- α AMG preconditioner. A generic preconditioning step is described via its error propagation

$$\epsilon \leftarrow (I - MD)^k (I - PD_c^{-1}(\mu)RD)(I - MD)^j \epsilon. \quad (16)$$

Therein M denotes the smoother, P denotes the interpolation which maps from a coarser space with fewer degrees of freedom to \mathcal{V}_s , and R denotes the restriction, the counterpart of P , which maps from \mathcal{V}_s to the coarser space. Having P and R , the coarse grid operator $D_c(\mu)$, which can be seen as a coarse version of $D(\mu)$, is defined by the Galerkin condition

$$D_c(\mu) = RD(\mu)P. \quad (17)$$

Furthermore, $I - MD$ denotes the error propagator of the smoother. The powers j and k denote its repeated application, i.e. we have j presmoothing and k postsmoothing steps. This error propagator is supposed to act on the UV modes; it reduces the error components belonging to UV modes. $I - PD_c^{-1}(\mu)RD$ denotes the error propagator of the coarse grid correction which acts complementary to the error propagator of the smoother. It reduces the error components belonging to IR modes. As an iteration prescription, the coarse grid correction itself is given by

$$\psi \leftarrow \psi + PD_c^{-1}(\mu)Rr. \quad (18)$$

This two-level approach can be recursively expanded to a multilevel approach by using again a two-level approach for the inversion of $D_c(\mu)$. Then on each new level ℓ , transfer operators R_ℓ and P_ℓ are employed in the construction of the next level operator $D_{\ell+1}(\mu) = R_\ell D_\ell(\mu)P_\ell$ generalizing Eq. (17), and a smoother M_ℓ is assigned. DD- α AMG uses a K-cycle approach [26], which means that at each level, a multigrid preconditioned Krylov method is performed until a prescribed decrease of the residual is obtained. Our numerical tests showed that the inversion of the coarsest Dirac operator D_L can be evaluated at low accuracy, which reduces the computational effort. This is not the case for the inexact deflation approach from Ref. [1] as explained in Ref. [4].

C. Coarse grid

The coarse grid correction, given in Eq. (18), should approximately remove the error corresponding to the IR-mode components on the fine grid. An ideal, yet naive, choice is to span the prolongation operator P by the right eigenvectors corresponding to small eigenvalues of $D(\mu)$ and the restriction operator R by the left ones, since then the coarse grid correction entirely removes the (right) IR modes present in the error. Such a choice is very similar to exact deflation of the small eigenvalues in that the coarse grid correction solves directly the linear system of the IR modes. In practice, the computational effort for generating such an ‘‘exact’’ coarse grid projection is more than a magnitude higher than in aggregation-based approaches, see e.g. Tables II and III in Sec. III A 2, and it scales with $\mathcal{O}(V^2)$. Instead, the DD- α AMG solver uses a coarse grid correction, which is based on aggregation and the property of *local coherence* described in Ref. [1]. Local coherence implies that it is possible to approximate the subspace of the IR modes by aggregates over just a small set of $\mathcal{O}(20)$ test vectors v_i . In this section, we discuss how to construct the aggregation-based prolongation, the restriction and the coarse grid operators from the test vectors, while in Sec. II F we describe the strategy for generating the test vectors.

The coarse grid is obtained by mapping an aggregate \mathcal{A}_j of the fine grid to a single site of the coarse grid, where

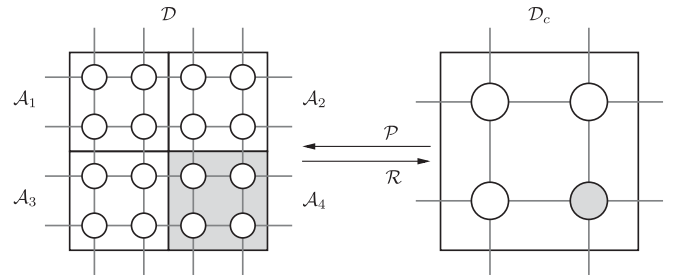


FIG. 2. Block aggregation of the Dirac operator and symbolic representation of the movement from the fine operator to the coarse operator and vice versa.

$$\mathcal{A}_j = \mathcal{V}_j \times \mathcal{S}_j \times \mathcal{C}_j. \quad (19)$$

We use static blocks with a fixed size given by

$$\mathcal{V}_j = \mathcal{T}_j \times \mathcal{Z}_j \times \mathcal{Y}_j \times \mathcal{X}_j \quad (20)$$

such that these blocks decompose the lattice \mathcal{V}_s as shown in Fig. 2. The number of sites on the coarse lattice \mathcal{V}_c is then given by $N_b = V/V_b$ with V_b being the block volume.

Now, a projection P and a restriction R between the fine grid and the coarse grid

$$\mathcal{V} \times \mathcal{S} \times \mathcal{C} \xrightleftharpoons[P]{R} \mathcal{V}_c \times N_b, \quad (21)$$

can be constructed by using the test vectors v_i . For the prolongation P the test vectors are decomposed into blocks over the aggregates as

$$(v_1 | \dots | v_{N_b}) = \begin{pmatrix} \text{block} \\ \text{block} \\ \vdots \\ \text{block} \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \text{block} & & & \\ & \text{block} & & \\ & & \ddots & \\ & & & \text{block} \end{pmatrix} \begin{matrix} \mathcal{A}_1 \\ \mathcal{A}_2 \\ \vdots \\ \mathcal{A}_{N_b} \end{matrix}. \quad (22)$$

For the restriction operator, R^\dagger is constructed similarly from a possibly different set of test vectors w_i . For numerical reasons, the parts of all test vectors over a given aggregate are orthonormalized.

It is possible to extend the Γ_5 -Hermiticity of the Wilson Dirac operator to the coarse grid operator. This can be done by decomposing the aggregation in the spin space \mathcal{S} with an aggregate

$$\mathcal{A}_{j,+} = \mathcal{V}_j \times \mathcal{S}_{0,1} \times \mathcal{C}, \quad (23)$$

which collects the two upper spin components 0 and 1, and an aggregate

$$\mathcal{A}_{j,-} = \mathcal{V}_j \times \mathcal{S}_{2,3} \times \mathcal{C}, \quad (24)$$

which collects the two lower spin components 2,3. In the chiral basis, Γ_5 acts with $+1$ to the components of the aggregate $\mathcal{A}_{j,+}$ and with -1 on the other aggregates. Now, the coarse grid $\Gamma_{5,c}$ can be defined by $\Gamma_{5,c} = I_{\mathcal{V}_c} \otimes \tau_3 \otimes I_{N_v}$, where τ_3 acts on the different spin aggregates. This type of aggregation was proposed in Ref. [8] for the MG-GCR method and it is used as well in the DD- α AMG method, where it is termed *standard aggregation*. The Γ_5 -compatible prolongation P satisfies

$$\Gamma_5 P = P \Gamma_{5,c} \quad (25)$$

which can be represented as

$$\begin{pmatrix} I_{6 \cdot V_b} & & & \\ & -I_{6 \cdot V_b} & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix} \begin{pmatrix} \text{[Diagram: 6x6 block]} \\ \text{[Diagram: 6x6 block]} \\ \text{[Diagram: 6x6 block]} \\ \text{[Diagram: 6x6 block]} \end{pmatrix} \\ = \begin{pmatrix} \text{[Diagram: 6x6 block]} \\ \text{[Diagram: 6x6 block]} \\ \text{[Diagram: 6x6 block]} \\ \text{[Diagram: 6x6 block]} \end{pmatrix} \begin{pmatrix} I_{N_v} & & & \\ & -I_{N_v} & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix}. \quad (26)$$

With the standard aggregation the coarse grid operator with a finite twisted mass term is given by

$$D_c(\mu) = R(D + i\mu\Gamma_5)P = RDP + i\mu\Gamma_{5,c}RP. \quad (27)$$

As pointed out in Ref. [4], the natural choice for the restriction operator is $R = (\Gamma_5 P)^\dagger$ since in this way the restriction approximates the subspace of the small *right* eigenvectors of the Wilson Dirac operator for the case in which the interpolation approximates the small *left* eigenvectors. Using this formulation, the resulting coarse grid operator $RD(\mu)P$ is equivalent to $H_c + i\mu$ with $H_c = P^\dagger H P$, which is a complex shifted maximally indefinite operator [cf. Eq. (7)]. Due to the commutativity relation between P and Γ_5 , given in Eq. (25), and already noted in Ref. [9], the coarse grid corrections obtained by $R = (\Gamma_5 P)^\dagger$ and $R = P^\dagger$ are identical. DD- α AMG thus uses $R = P^\dagger$, and the coarse grid twisted mass operator there is defined by

$$D_c(\mu) = P^\dagger D P + i\mu\Gamma_{5,c}. \quad (28)$$

$D_c(\mu)$ preserves several important properties of the fine grid operator $D(\mu)$. It preserves the sparse structure in that

only neighboring aggregates are coupled. In the square coarse grid operator

$$D_c^\dagger(\mu)D_c(\mu) = D_c^\dagger D_c + \mu^2, \quad (29)$$

the eigenvalues are again bounded from below by μ^2 , and there is a $\Gamma_{5,c}$ -symmetry which reproduces the connection between the u- and d-quark operators as

$$\begin{aligned} D_c^\dagger(\mu) &= P^\dagger (D + i\mu\Gamma_5)^\dagger P \\ &= P^\dagger (\Gamma_5 D \Gamma_5 - i\mu\Gamma_5) P = \Gamma_{5,c} D_c(-\mu) \Gamma_{5,c}. \end{aligned} \quad (30)$$

Although the dimension of the coarse grid operator is reduced, it can develop a large number of small eigenvalues close to μ . This can critically slow down the convergence of a standard Krylov solver to be used on the coarsest grid such that the time spent in the coarsest operator inversions dominates by far the overall inversion time even though only poor accuracy is required. We therefore decrease the density of small eigenvalues of the coarsest grid operator by increasing the twisted mass parameter by a factor of δ given by

$$D_c(\mu, \delta) = D_c + i\delta\mu \cdot \Gamma_{5,c} \quad (31)$$

with $\delta \geq 1$.¹

We analyze the effect of δ in detail in Sec. III C 2. As compared to the standard Wilson Dirac operator, for the twisted mass Dirac operator it turns out that the overall execution time is minimized if one relaxes even further the accuracy of the coarsest grid solve as is described in more detail in Sec. III C 2 as well.

D. Smoother

In the DD- α AMG approach, a red-black SAP is used as a smoother. This domain decomposition method was introduced to lattice QCD in Ref. [25], where it was used as a preconditioner. The lattice is partitioned into alternated “red (r)” and “black(b)” lattice blocks in a checkerboard manner, and the subdomains are obtained as the full color-spin space over the respective lattice block [see Eq. (20)]. Reordering the Dirac twisted mass operator such that the red blocks come first, we obtain

$$D + i\mu\Gamma_5 = \begin{pmatrix} D_{rr} + i\mu\Gamma_5^r & D_{rb} \\ D_{br} & D_{bb} + i\mu\Gamma_5^b \end{pmatrix}, \quad (32)$$

where D_{rr} and D_{bb} are block diagonal matrices filled with the respective subdomains, while D_{rb} and D_{br} connect the neighboring blocks. Note that the operator D has only next

¹This trick is used in heavy $N_f = 2$ twisted mass simulations at the charm quark mass and we thank Björn Leder for this suggestion.

neighbor interactions and thus blocks of a specific color do not couple to the same color. The eigenmodes of the blocks have a higher cutoff than the full operator, given by $p_\nu = \pi/L_b$ in the free case for $\mu = 0$ due to the Dirichlet boundary conditions.

If now the role of the smoother is to reduce the UV modes, a natural choice for the operator M in the error propagation [see (16)] is given by the inverse of the block operator, resulting in

$$\begin{aligned} E_{\text{SAP}} &= (I - M_{\text{SAP}}D(\mu)) \\ &= (I - B_{rr}(\mu)D(\mu))(I - B_{bb}(\mu)D(\mu)), \end{aligned} \quad (33)$$

where $B_{rr}(\mu)$ and $B_{bb}(\mu)$ are the block inverses defined as

$$\begin{aligned} B_{rr} &= \begin{pmatrix} (D_{rr} + i\mu\Gamma_5^r)^{-1} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \\ B_{bb} &= \begin{pmatrix} 0 & 0 \\ 0 & (D_{bb} + i\mu\Gamma_5^b)^{-1} \end{pmatrix}. \end{aligned} \quad (34)$$

In practice, the (approximate) inversion of the blocks on the diagonal of B_{rr} and B_{bb} is performed with small computational cost by a few steps of an iterative method like the minimal residual method [4,27]. Note that we fix the block size to coincide with the aggregates on each level of the multigrid hierarchy.

E. Krylov subspace methods

A Krylov subspace method preconditioned by the chosen multigrid approach has to be a flexible algorithm, since the smoother as well as the solver on the coarse grid system are nonstationary processes. Flexible solvers, which have been employed in multigrid preconditioning for lattice QCD are the flexible biconjugate gradient (CG) stabilized method ([28]), generalized conjugate residual ([29]) and flexible generalized minimal residual [(F)GMRES, [30]] solver. In the DD- α AMG approach a FGMRES solver is used for the inversion of the fine grid operator and in the K cycle for the inversion of the coarser operators except for the coarsest. The latter is inverted by even-odd preconditioned GMRES, i.e. GMRES is run for the even-odd reduced system $\hat{D}\phi_o = \eta_o - D_{oe}D_{ee}^{-1}\eta_e$ with \hat{D} from Eq. (11) and then ϕ_e in the solution $\begin{bmatrix} \phi_e \\ \phi_o \end{bmatrix}$ of $D\begin{bmatrix} \phi_e \\ \phi_o \end{bmatrix} = \begin{bmatrix} \eta_e \\ \eta_o \end{bmatrix}$ is retrieved as $\phi_e = D_{ee}^{-1}(\eta_e - D_{eo}\phi_o)$. Even-odd preconditioning is also used in the smoother when inverting the blocks. In both cases a speed-up of up to 50% compared to the full operator can be achieved. This can be explained by the fact that the small eigenvalues are increased by a factor of 2 as seen from Eq. (12).

In general, the accuracy of the coarse grid inversions may be very much relaxed as compared to the target accuracy

of the fine grid inversion. Indeed, for DD- α AMG with the K-cycle strategy, optimal results are obtained when requiring the (approximate) inversions of the coarser operators to reduce the residual by just 1 order of magnitude.

F. Setup phase

In the setup phase we have to compute a sufficient number of approximate low modes v_i , which when chopped into aggregates approximate the IR modes well due to the local coherence property. We employ the setup algorithm from Ref. [4] for the twisted mass operator.

The vectors $\{v_i\}$ are generated by using a variant of block inverse iteration

$$\begin{aligned} v_i^{(k)} &= D^{-1}(\mu)v_i^{(k-1)}, \\ i &= 1, \dots, N_v, k = 1, \dots, n_{\text{setup}}, \\ v_i^{(0)} &\text{ chosen randomly,} \end{aligned} \quad (35)$$

where the $v_i^{(k)}$ vectors converge to the eigenvectors with eigenvalues of smallest modulus. In practice, in order to maintain numerical stability, after each iteration k the vectors $v_i^{(k)}$ spanning the space of approximate IR modes have to be orthonormalized, and $D^{-1}(\mu)$ is replaced by a multigrid iteration with prolongation operator constructed from the current set of approximate low modes $v_i^{(k)}$. For $k = 1$, where a multigrid hierarchy is not yet available, we just apply some steps of the SAP smoother. This approach results in a self-adapting procedure where the multigrid hierarchy is improved while using it to expose the small eigenmodes. Typically, a small number of setup iterations n_{setup} is sufficient.

III. NUMERICAL RESULTS FOR TWISTED MASS FERMIONS

The DD- α AMG approach uses a wide range of parameters, and its time to solution depends on a good choice of the parameters. The purpose of this section is therefore threefold. We first provide a set of default parameter choices, which in our extensive numerical testing turned out to yield good overall performance. Secondly, we show that for physically relevant configurations, appropriately chosen parameters yield speed-ups of about 2 orders of magnitude compared to standard methods for twisted mass fermions. We also show numerically that these parameters can be kept fixed over a whole statistical ensemble without a notable decrease in performance. Finally, we present an analysis on the dependence of several parameters which implies a general strategy for obtaining good parameters for a given ensemble and computer. All the presented numerical results have been obtained on SuperMUC phase 2 at the Leibniz Supercomputing Centre, an Intel Haswell Xeon computer on which we used up to 4096 cores. We have also

performed runs on JURECA at the Julich Supercomputing Centre and on Cy-Tera at the Computation-based Science and Technology Research Center Cyprus obtaining compatible results. The stopping criterion of the overall iteration was fixed such that the residual is reduced by a factor of 10^9 .

A. Default parameters

Table I summarizes our default parameter set used for DD- α AMG. The parameter tuning was done for the ensemble *cA2.09.48* [23] with a lattice volume of $V = 96 \times 48^3$ and lattice spacing of $a = 0.0931(10)$ fm as determined from the nucleon mass. This ensemble was generated at a pion mass close to the physical one, namely $m_\pi = 0.131$ MeV [23].

1. Discussion

As explained in Sec. II F, in each iteration of the adaptive setup routine the currently available multigrid hierarchy

TABLE I. The parameter set used in DD- α AMG, obtained by parameter tuning for the TM fermion ensemble *cA2.09.48* [23].

	Parameter		Optimal
Multigrid	Number of levels	n_ℓ	3
	Number of setup iterations	n_{setup}	5
	Number of test vectors on level 1	$N_{v,1}$	28
	Number of test vectors on level 2	$N_{v,2}$	28
	Size of lattice blocks for aggregates on level 1	$V_{b,1}$	4^4
	Size of lattice blocks for aggregates on level ℓ , $\ell > 1$	$V_{b,2}$	2^4
	Solver	Mixed precision (F)GMRES	
Relative residual tolerance (restarting criterion)			10^{-6}
Smoother (SAP)	Red-black multiplicative SAP		
	Size of lattice-blocks on level 1		4^4
	Size of lattice-blocks on level ℓ , $\ell > 1$		2^4
	Number of postsmoothing steps		4
	MINRES iterations to invert the blocks		3
K-cycle	With single precision FGMRES		
	Restart length		5
	Number of maximal restarts		2
	Relative residual tolerance (stopping criterion)		10^{-1}
Coarsest grid	Solved by even-odd preconditioned GMRES		
	Twisted mass parameter	μ_{coarse}	$5.2 \cdot \mu$
	Restart length		100
	Number of maximal restarts		5
	Relative residual tolerance (stopping criterion)		10^{-1}

is used to perform one iteration with the multigrid preconditioner on each test vector. If not stated otherwise, we used the parameters from Table I for the preconditioner in the setup phase and in the solve phase.

As the setup iteration proceeds the test vectors become more rich in low mode components. This also leads to a more ill-conditioned coarse operator D_c and higher iteration counts on the coarse grid [cf. Ref. [4]], i.e. one can observe a higher cost per setup iteration as the setup proceeds. This can be seen in Table III where the setup times for $n_{\text{setup}} = 3$ and $n_{\text{setup}} = 5$ are stated. Indeed a factor of 5 in computing time between both setups can be observed. The suggested value $n_{\text{setup}} = 5$ from Table I thus yields a good value when several inversions [$\mathcal{O}(100)$ and more] with the same operator are desired. The relatively large setup time can be neglected in this case. We did not find that more setup iterations than $n_{\text{setup}} = 5$ yield substantial further improvement in the solve time. On the other hand, when solving for a few rhs, a good balance of setup and solve time has to be found. Therefore a smaller number of setup iterations like $n_{\text{setup}} = 3$ might be more suitable.

For larger pion masses one can, in principle, improve the generation of the subspace of small eigenmodes in the setup by using a smaller mass parameter m or a smaller twisted mass parameter μ , which helps to probe the small eigenmodes more rapidly within the setup phase. However, at least at the physical point, we do not find a significant improvement by using different mass parameters.

We use the same subspace for the u quark with $+\mu$ and the d quark with $-\mu$, i.e. we run the setup phase only once for both quarks. Although the eigenspace changes, numerically we do not find a large difference. This also saves computing time for many applications where the square operator has to be inverted, e.g. during the integration within the hybrid Monte Carlo (HMC) algorithm [31,32].

2. Comparison with CG

To put the improvements for twisted mass simulations in perspective, we start with an experiment for Wilson Dirac fermions, thus complementing the results from Ref. [5] for our target machine. We use the default parameter set from Ref. [5] and a configuration from ensemble VII [33] with a pion mass of $am_\pi = 0.05786$, a lattice spacing of $a = 0.0071$ fm and a lattice volume of $V = 64 \times 48^3$.

Table II presents a comparison of the inversion times of the CG solver, the CG solver with exact deflation (CG-eDe) [27,34] and the DD- α AMG solver. For CG and CG-eDe we use the publicly available software package *tmLQCD* [35] that is commonly used for simulations with the twisted mass operator and provides a count of floating point operations per second (Flop/s) for the CG solver. This makes it possible to rescale results obtained on different systems.

The conjugate gradient solver requires a positive definite Hermitian matrix, which is obtained by solving the linear

TABLE II. Results for clover Wilson fermions.^a Comparison of CG (tmLQCD), eigCG (tmLQCD + ARPACK) with 800 eigenvectors and DD- α AMG with parameters from Ref. [5]. The results were for a $48^3 \times 64$ lattice from ensemble VII of Ref. [33] with $m_\pi = 0.1597(15)$ GeV.

Solver	Setup time (core hrs)	Inversion time (core hrs)	Total iteration count of the fine grid solve	Total iteration count of coarse grid solvers
CG		174.8	26 937	
CG-eDe	1527.4	5.4	649	
DD- α AMG	13.3	0.9	16	2988

^aThe timings for CG and eigCG have been normalized to 1.0 Gflop/s per core (the average of standard performance ~ 0.7 Gflop/s and optimal performance ~ 1.3 Gflop/s); the rhs of the equation $D\psi = b$ has been randomly generated and in all the cases the propagator ψ has been computed to a relative precision of 10^{-9} .

TABLE III. Results for TM fermions.^a Comparison of CG (tmLQCD), eigCG (tmLQCD + ARPACK) with 1600 eigenvectors and DD- α AMG. The results were computed for the cA2.09.48 ensemble [23] with $m_\pi = 0.131$ GeV.

Solver	Setup time (core hrs)	Inversion time (core hrs)	Total iteration count of the fine grid solve	Total iteration count of coarse grid solvers
CG		338.6	34 790	
CG-eDe	6941.1	9.8	695	
DD- α AMG for $n_{\text{setup}} = 3$	7.7	2.5	28	16 619
DD- α AMG for $n_{\text{setup}} = 5$	38.3	1.5	15	11 574

^aThe timings for CG and eigCG have been normalized to 1.0 Gflop/s per core (the average of standard performance ~ 0.7 Gflop/s and optimal performance ~ 1.3 Gflop/s); the rhs of the equation $D\psi = b$ has been randomly generated and in all the cases the propagator ψ has been computed to a relative precision of 10^{-9} .

system with the squared even-odd reduced operator given by $\hat{D}^\dagger \hat{D}x = b'$ with $b' = \hat{D}^\dagger b$. This squares the condition number of the involved matrix.

The CG-eDe and the DD- α AMG solver involve a setup phase, which has to be done for each new configuration once before the linear system is solved. In the case of the CG-eDe solver, $\mathcal{O}(100)$ lowest eigenmodes of the squared even-odd reduced Dirac operator are calculated (here we use 800 for the case of Wilson case and 1600 for the twisted mass case). The number of eigenvectors in CG-eDe is obtained by optimizing the time to solution (setup + inversion time) for computing the $\mathcal{O}(1000)$ rhs. Indeed the setup phase is extremely expensive which makes CG-eDe inefficient for a small number of rhs. The low mode computation is done by using the publicly available package ARPACK with tmLQCD [35].

Table II shows that in the case of Wilson Dirac fermions, the DD- α AMG solver speeds up the time to solution by roughly a factor of 200 compared to CG and roughly by a factor 5 compared to the CG-eDe solver. When including the setup time, DD- α AMG is roughly a factor of 12 faster than CG for one rhs, while CG-eDe is not competitive due to its computationally demanding setup.

In the twisted mass fermion case, we are able to achieve the same speed-ups as for the Wilson fermion case. This is not straightforward as it requires the coarsest grid twisted mass μ_{coarse} to be chosen as different from the fine grid

twisted mass μ . We choose $\mu_c = 5.2\mu$ (cf. Table I) for reasons that are explained in Sec. III C 2. The results for twisted mass fermions are shown in Table III.

Indeed, we are able to achieve a speed-up in inversion time of roughly a factor of 220 compared to the CG and roughly by a factor 6 compared to the CG-eDe solver, with the time for the setup being almost a factor of 100 less as compared to CG-eDe. For $n_{\text{setup}} = 3$ we used the parameters from Table I except for the number of test vectors on the fine level $N_{v,1}$ being 20 instead of 28. For $n_{\text{setup}} = 5$ we used a further optimized set which is given in the last row of Table IV. This yields another factor of 1.5 speed-up in the inversion time at the expense of increasing the setup time.

In Fig. 3 we show the speed-up in total time (setup + solve) compared to one CG solve for DD- α AMG and CG-eDe as a function of the number of rhs n_{rhs} . The difference between the two blue curves in the figure on the right-hand side is due to results from a different number of setup iterations (3 and 5) and a different number of test vectors (20 and 28), where three setup iterations are optimal for few rhs (blue dotted line) and five setup iterations are optimal for many rhs (blue line). In summary, the results for twisted mass fermions show that for one rhs DD- α AMG is roughly 30 times faster than the CG for one rhs, 120 times faster for twenty rhs and 220 times faster for a thousand rhs.

TABLE IV. Summary of δ parameters yielding the optimal solve time for various block sizes $V_{b,1}$. The shown numbers were computed for the *cA2.09.48* ensemble and a relative residual tolerance of 10^{-9} . The number of test vectors was chosen as in Fig. 6(c) according to the block size $V_{b,1}$. All other parameters were fixed to the values in Table I.

Block size $V_{b,1}$	Test vectors $N_{v,1}$	δ for the coarsest $\mu\mu_{\text{coarse}} = \delta\mu$	Setup time (core hrs)	Inversion time (core hrs)
$3^3 \times 4$	24	7.8	31.1	3.19(3)
2×4^3	24	6.6	27.5	2.71(5)
$3^2 \times 4^2$	24	8.6	23.5	2.47(2)
3×4^3	28	5.4	28.0	2.04(5)
4^4	28	5.2	22.2	1.75(4)
$4^3 \times 6$	36	4.5	37.4	1.74(6)
$4^3 \times 8$	40	4.0	40.7	1.73(5)
$4^2 \times 6^2$	40	4.1	37.7	1.59(6)
4×6^3	44	4.0	38.3	1.52(4)

B. Stability of optimal parameters

Empirically we find that the solver performance is stable for the tuned parameter set. Within the ensemble *cA2.09.48* we do not find any configurations where the iteration

counts or the time to solution differ by more than 5%, as can be seen in Fig. 4. This behavior is also corroborated by the performance of the multigrid solver during the force computation in the HMC algorithm, where it shows very stable iteration counts for simulations at the physical point; see Fig. 5.

C. Analysis of parameter settings and tuning strategy

As demonstrated in the previous section, well-tuned parameters are important for good performance, and they tend to be stable at least for configurations from a given ensemble. It is therefore advisable to invest some effort in obtaining good parameters. Since the parameter space is large, it cannot be searched exhaustively, and there is thus a need for a strategy of how to tune the method in practice. Our suggestion for twisted mass simulations is to keep the default parameters given in Table I, but tune the aggregation parameters and the twisted mass parameter on the coarsest level, μ_{coarse} . This is justified by the analysis that we explain in the subsequent sections as follows: In Sec. III C 1, we present a novel analysis of the aggregation parameters without tuning μ_{coarse} , i.e. we fix $\delta = 1$; in Sec. III C 2, we show the benefits obtained by increasing μ_{coarse} and we repeat the previous analysis; in Sec. III C 3,

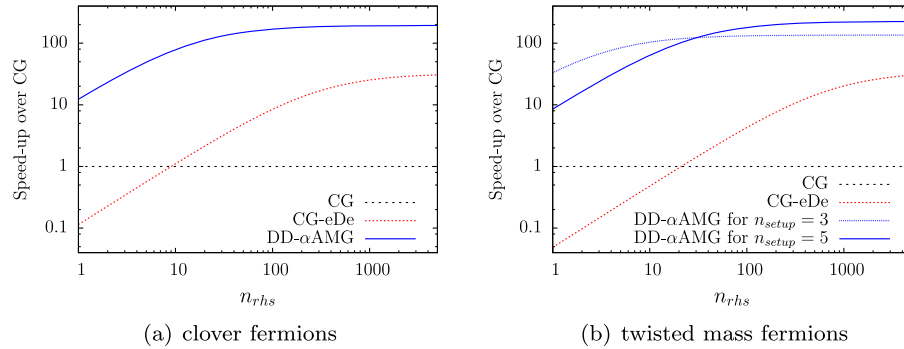


FIG. 3. Speed-up over CG using the results of Tables II and III.

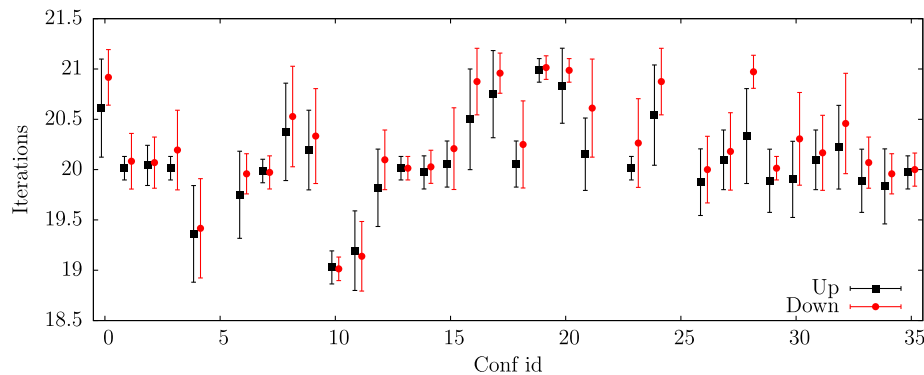


FIG. 4. We depict the average iteration counts on computing quark propagators for several configurations of the ensemble *cA2.09.48*. For the black square points the setup is generated with the same μ while for the red circle points the setup is generated with a TM parameter μ with opposite sign.

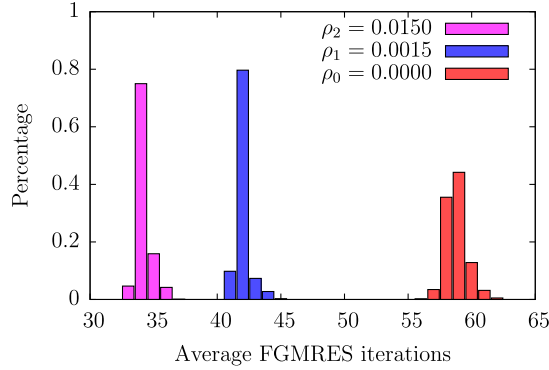


FIG. 5. We show the (F)GMRES iteration counts averaged over the trajectory when the solver is used for computing the force terms in an HMC simulation. The data is for the ensemble *cA2.09.64* with statistics of 2000 trajectories and the squared operator $\hat{D}^\dagger \hat{D} + \rho_i^2$ is inverted, i.e. two inversions are performed with DD- α AMG inverting $\hat{D} + i\rho_i \hat{\Gamma}_5$ and $\hat{D}^\dagger - i\rho_i \hat{\Gamma}_5$ where $\hat{\Gamma}_5$ is the restriction of Γ_5 to the odd lattice sites. The parameter ρ_i sets the mass for the Hasenbusch preconditioning [36]; the integration scheme is equivalent to the one presented in Ref. [23] for the ensemble *cA2z.09.48*. As a setup strategy within the HMC, we produce an initial set of test vectors using three setup iterations. Right before every inversion for ρ_0 we update the setup using one setup iteration at ρ_0 . The resulting preconditioner is used for all other Hasenbusch masses until the next update.

we demonstrate that also other multigrid approaches can benefit from an increasing μ_{coarse} . All the tests are performed on one configuration averaging the time to solution for the u and d propagator. This choice is motivated by the stability of the solver presented in the previous section.

1. Aggregation parameters

Aggregation parameters are the number of the test vectors $N_{v,\ell}$ and the size of the lattice blocks $V_{b,\ell}$ on each level ℓ . They should be tuned simultaneously since they define the size of the coarser Dirac operator and consequently the size of the projected subspace. In the present analysis, the solver is restricted to a three-level implementation. We do not find an improvement in the time to solution by using a fourth level, which is also the result found in Ref. [5] for the Wilson operator with similar lattice sizes. On the other hand, the inversion time increases when just a two-level multigrid method is used.

To optimize the aggregation parameters on the first level, we fix the values of the parameters on the second level for which we find a block size of $V_{b,2} = 2^4$ and a number of test vectors around 28 to work well. Fixing these parameters by setting $N_{v,2} = \max(28, N_{v,1})$ we analyze how the time needed to solve for one rhs and for the setup depends on the block size $V_{b,1}$ and the number of test vectors $N_{v,1}$. The results are depicted in Fig. 6 for the *cA2.09.48* ensemble and a scaling parameter $\delta = 1$ on all levels has been used.

We find that every block size has an optimal number of test vectors as shown in Figs. 6(a) and 6(b) that minimizes the time to solution. By fitting the data to a polynomial of order 2 in $N_{v,1}$ we estimate the minima and find that the optimal number of test vectors grows approximately linearly for block sizes $V_{b,1} < 8 \cdot 4^3$ with the block volume

$$N_{v,1} = \alpha + \beta V_{b,1}, \quad (36)$$

which is shown in Fig. 6(c). This indicates a nontrivial connection between the fine grid size and the dimension of the coarse grid operator given by

$$\dim(D_{c,1}) = 2 \left(\frac{\alpha}{V_{b,1}} + \beta \right) V. \quad (37)$$

This implies that the optimal size of the coarse grid operator increases linearly with the volume V at fixed aggregation block size. For larger block sizes the behavior deviates from this linear dependence, however. For this case the minimum of the time to solution is already reached for a smaller number of test vectors. We can interpret this complex behavior as giving some insight into the nontrivial link between coarse grid size and local coherence.

The time to solution of the multigrid method is dominated by the solves with the coarse grid operator. We use a K cycle and a three-level approach where, due to fixing the blocks of the coarsest grid, it follows that $\dim(D_{c,1}) \propto \dim(D_{c,2})$. The numerical effort needed for a matrix-vector multiplication involving $D_{c,1}$ depends linearly on $\dim(D_{c,1})$. Allowing for an additional second order term we model the time for an inversion as

$$t_{\text{solve}} = \gamma' + \delta' \dim(D_{c,1}) + \epsilon' \dim(D_{c,1})^2. \quad (38)$$

By using the dependence on $V_{b,1}$ of Eq. (37) we can rewrite t_{solve} with

$$t_{\text{solve}} = \gamma + \delta V_{b,1}^{-1} + \epsilon V_{b,1}^{-2}. \quad (39)$$

Within this model, connecting $N_{v,1}$ and $V_{b,1}$ according to Eq. (36) and Fig. 6(c), the time to solution can be fitted very well up to $V_{b,1} \approx 8^2 \times 4^2$ as shown in Fig. 6(d). In Fig. 7, we display the data obtained for the ensemble *cA2.09.64* with a lattice size of 128×64^3 . The other lattice parameters are the same as those for *cA2.09.48*. In contrast to the discussed analysis of *cA2.09.48*, the coarse grid scale factor is set to $\delta = 5$ on the coarsest grid. The full analysis on the dependence of the algorithm on δ is reserved for the next section. For both lattice sizes the inversion times reach the minimum for the same block volumes. The asymptotes γ from Eq. (39) are given by 1.64(13) for the *cA2.09.48* ensemble and by 6.3(5) for the *cA2.09.64* ensemble, which is an increase of $V^{5/4}$. As mentioned above, for larger

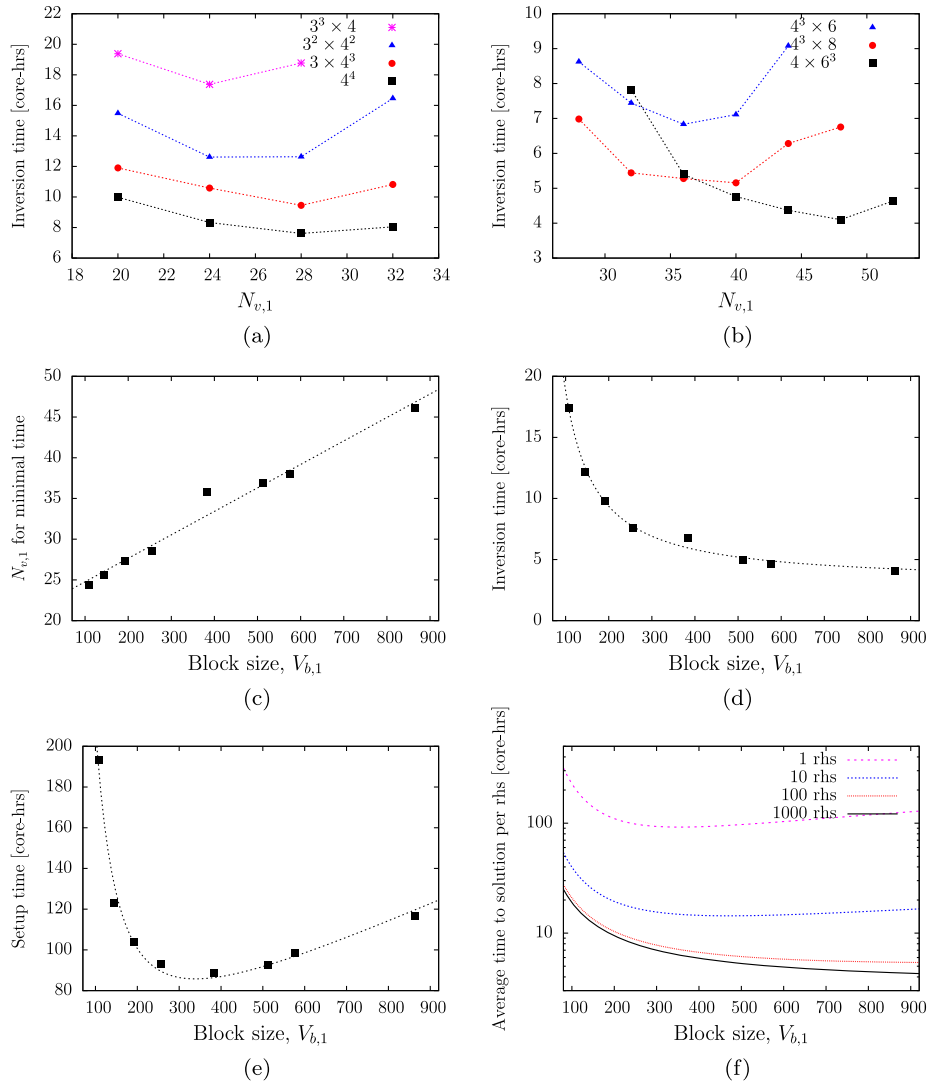


FIG. 6. Analysis of the aggregation parameters for the *cA2.09.48* ensemble. For each set of raw data presented in (a) and (b) the position and value of the minimum has been extracted with a parabolic interpolation and displayed in (c) and (d), respectively. In (d)–(f), $N_{v,1}$ and $V_{b,1}$ are connected according to the minima found in (c), i.e. (d) shows the inversion time and (e) the setup time, both for the minima from (c). The estimated average total time per right-hand side $(t_{\text{setup}} + n_{\text{rhs}}t_{\text{solve}})/n_{\text{rhs}}$ is shown in (f). The fitting functions are explained in the text.

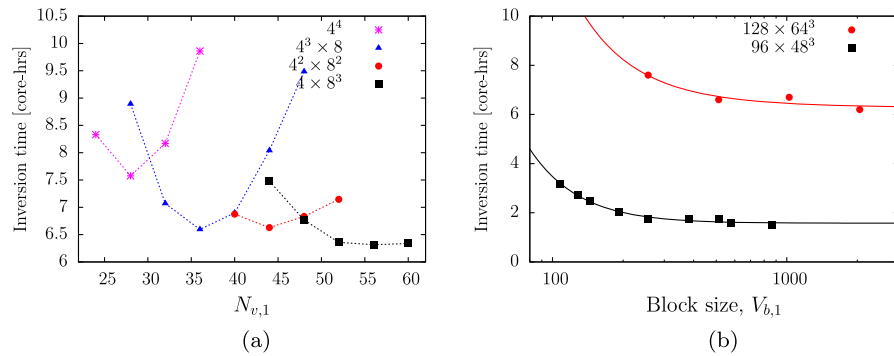


FIG. 7. Analysis of aggregation parameters for *cA2.09.64* and $\delta = 5$ on the coarsest grid. The sets of raw data are presented in (a); from each set, the value of the minimum has been extracted with a parabolic interpolation and plotted with red points in (b). The black points are for the ensemble *cA2.09.48* and $\delta = 5$ on the coarsest grid.

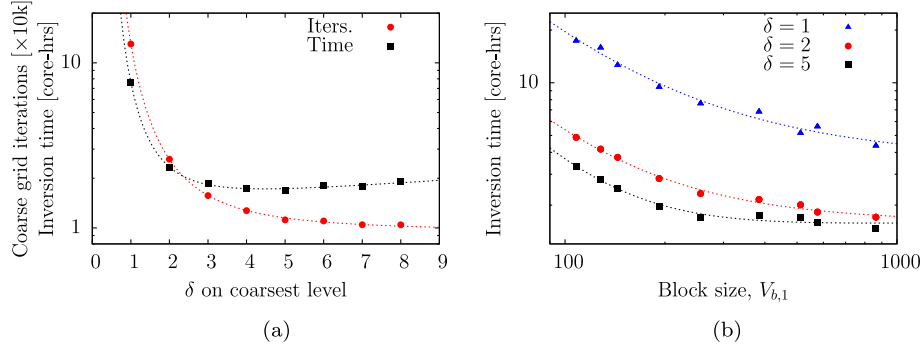


FIG. 8. The number of coarse grid iterations as a function of the δ parameter and the inversion time for different δ -values depending on the block size. The results depicted in (a) are for a block size of $V_{b,1} = 4^4$ and compare the number of iterations on the coarse grid with the fine grid inversion time. The number of iterations has been scaled down by a factor 10000. The behavior for the coarse grid iterations is proportional to $1/\delta^2$, while the inversion time has an additional term linear in δ due to the increase in the fine grid iteration count. The results depicted in (b) show the inversion time for different values of δ and different block sizes. The behavior is similar to the one observed in Fig. 6(d).

block sizes, the linear connection of $N_{v,1}$ and $V_{b,1}$ does not hold and the solution time is increased. This is also observed for the time spent in the setup. We also observe that with increasing lattice volume the region where the linear dependence holds is shifted.

During the setup procedure the solver is applied on $N_{v,1}$ test vectors for several iterations. Here, the test vectors are orthonormalized at the end of each iteration and used for building the multigrid hierarchy. We model the setup time by allowing for a linear and quadratic term as

$$t_{\text{setup}} = \zeta + \eta N_{v,1} t_{\text{solve}} + \theta \dim(D_{c,1}), \quad (40)$$

where the amount of computation which does not involve the solver scales with $\dim(D_{c,1})$ at leading order. Figure 6(e) shows that the measured timings are indeed very well described by the ansatz given in Eq. (40). We remark that for block sizes $V_b \sim 4^4$ up to $6^2 \times 4^2$ the time to solution shows a relatively large plateau, which makes the timings for the multigrid solver relatively stable. Further fine-tuning in this region leads only to small improvements.

The optimal choice for $N_{v,1}$ and $V_{b,1}$ depends on the number of rhs n_{rhs} . The total time is given by

$$t_{\text{total}} = t_{\text{setup}} + n_{\text{rhs}} t_{\text{solve}}, \quad (41)$$

and we find a nontrivial dependence of $V_{b,1}$ on n_{rhs} for an optimal time to solution, as depicted in Fig. 6(f). This motivated our suggestion to consider two different values for n_{setup} , depending on the number of rhs.

2. Tuning the coarse grid scale factor δ

At physical quark masses, the density of the low-lying eigenvalues for the twisted mass operator increases compared to the Wilson Dirac operator, as explained in

Sec. II A. Densely populated low eigenvalues slow down the iteration of the Krylov subspace solvers on the coarsest grid and thus of the whole multigrid method. This is much more pronounced for the twisted mass Dirac operator than it is for the Wilson Dirac operator. For the latter, Table II reports a total of around 3000 coarse grid iterations for the considered ensemble, whereas data depicted in Fig. 8(a) show that for $\delta = 1$ we have roughly 40 times more coarse grid iterations in the twisted mass case, an unexpectedly large increase.

When using a larger twisted mass value on the coarsest grid operator, given by $\mu_{\text{coarse}} = \delta\mu$ with $\delta \geq 1$, we make the small eigenvalues less dense in the low-lying part of the spectrum. This speeds up the inversion time on the coarsest grid. We also analyze the effect of the scaling factor δ as a function of the block size. The results are depicted in Fig. 8 and we find that the optimal aggregation parameters determined in the previous section do not depend on the chosen coarse grid scale factor δ . In Table IV we present the final results for the *cA2.09.48* ensemble, summarizing the δ parameters which minimize the inversion time for various block sizes and the corresponding optimal numbers of test vectors. By using a scaling factor $\delta = 2$ the iteration count on the coarsest grid is already reduced by a factor of 5, which results in a speed-up of the fine grid inversion time by roughly a factor of 3. Optimal performance is achieved for a relatively large plateau around $\delta \sim 5$. Obviously, a large scaling factor δ causes a distinct violation of the Galerkin condition, i.e. $P^\dagger D(\mu)P \neq D_c(\delta\mu)$. However, we only find a minor increase in the iteration count of the fine grid solver from 14 to 16 iterations.

3. The scale factor δ and different multigrid approaches

In addition to the studies of the delta parameter in DD- α AMG we tested the influence of a delta parameter

in the QUDA implementation of MG-GCR [37] and in openQCD [6].² The latter is restricted to a two-level approach. For both approaches we found that the iteration counts on the coarsest grid were reduced by using a scale factor $\delta > 1$, also resulting in reduced inversion and setup times. For $\delta \approx 5$ we observe speed-ups similar to those for DD- α AMG for the *cA2.09.48* ensemble.

For MG-GCR this behavior is to be expected since both approaches share the same type of Γ_5 -respecting interpolation [cf. Eq. (25)]. However, the interpolation within the openQCD multigrid solver is not Γ_5 respecting, resulting in a nondiagonal summand $P^\dagger \Gamma_5 P \delta \mu$ on the coarse grid instead of $\Gamma_{5,c} \delta \mu$ in Eq. (31).

We indeed find cases where the openQCD solver and DD- α AMG show different behaviors. When going to the *cA2.09.64* ensemble with a bigger volume, we observe that the optimal δ for the openQCD solver increases while it remains constant for DD- α AMG. Furthermore, when solving for $D(-\mu)$ with a setup built for $D(\mu)$ from the *cA2.09.48* ensemble, we find that the optimal δ increases by a factor of 4 within the openQCD solver whereas it remains constant for DD- α AMG.

For the Wilson fermion case it was reported in Ref. [5] that the two-level openQCD solver shows about the same performance as the three-level DD- α AMG approach. For the twisted mass ensemble *cA2.09.48*, the openQCD solver is roughly a factor of 4 slower than DD- α AMG.

IV. CONCLUSIONS AND OUTLOOK

The DD- α AMG solver is extended to the case of $N_f = 2$ twisted mass fermions. By carefully tuning the parameters we show that one can achieve speed-ups similar to those obtained in the case of Wilson fermions, to be precise, a factor of 200 compared to the CG. This is mainly achieved by adapting the twisted mass parameter on the coarsest grid operator such that computing the coarsest grid correction becomes less time consuming. Using a factor $\delta \sim 5$ to increase $\mu_{\text{coarse}} = \delta \cdot \mu$ on the coarsest level decreases the coarse grid iteration count by roughly a factor around 10 and improves the time to solution by a factor of around 4 for the case of ensembles generated at the physical pion mass.

An optimal set of parameters for different applications of the solver is presented and the strategy for tuning the aggregation parameters and the factor δ for the twisted mass parameter is discussed in detail. Using the optimal set of parameters, inversions are more than 2 orders of magnitude faster as compared to the standard CG. Within the HMC, DD- α AMG achieves a speed-up of an order of magnitude compared to the standard CG.

The DDalphaAMG library for twisted mass fermions is publicly available, and an interface to the tmlQCD

²We use for the tests a modified openQCD version, which is optimized for twisted mass fermions, i.e. where the even-odd reduced twisted mass Dirac operator is implemented.

software package is provided. The technical details are summarized in Appendix A.

For the future we plan to extend multigrid methods to the heavy doublet sector of the twisted mass formulation.

ACKNOWLEDGMENTS

This project has received funding from the Horizon 2020 research and innovation program of the European Commission under the Marie Skłodowska-Curie Grant No. 642069. S. B. is supported by this program. We thank Giannis Koutsou for fruitful discussions and for helping us access the configurations. We also thank Artur Strebel and Simon Heybrock for guidance with the DD- α AMG code and Björn Leder for his suggestion to shift the twisted mass on the coarse grid to speed up the coarse grid solver. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. for funding the project *pr74yo* by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre; the computing time granted by the John von Neumann Institute for Computing (NIC); time provided on the supercomputer JURECA at Jülich Supercomputing Centre (JSC) through Grant No. *ecy00*; the High Performance Computing Center in Stuttgart for providing computation time on the high performance computing system Hazel Hen for calculating the eigenvalues through Grant No. *GCS-Nops (44066)*; and the computational resources on Cy-Tera at the Cyprus Supercomputing Centers through Grant No. *Ispre258s1*.

Gauss Centre for Supercomputing

APPENDIX: OVERVIEW OF THE DDALPHAAMG LIBRARY

The DDalphaAMG solver library, available at Ref. [12], has been recently released under GNU General Public License. This software package includes an implementation of the DD- α AMG for clover Wilson fermions as described in Ref. [4]. The implementation is of production code quality; it includes a hybrid MPI/openMP parallelization, state-of-the-art mixed precision and odd-even preconditioning approaches and also SSE3 optimizations. Implementation details can be found in Ref. [38].

Based on the DDalphaAMG code we developed a version, which supports twisted mass fermions, available in Ref. [21]. We added the following features to the library: $N_f = 2$ twisted mass fermions and twisted boundary conditions are supported, and different twisted mass shifts on the even and odd sites can be applied, which are required for the Hasenbusch mass preconditioning in the HMC when even-odd preconditioning is used.

User documentation for the library can be found in `src/DDalphaAMG.h`, and a sample code that

illustrates how to use the library interface functions can be found in `tests/DDalphaAMG_sample.c`. Moreover, the library has been integrated in the tmLQCD program [35]; therefore we also adjusted

the library interface. The interfaced code is available in Ref. [39]. Details on the employment of DDalphaAMG within tmLQCD are available in Ref. [40].

-
- [1] M. Lüscher, Local coherence and deflation of the low quark modes in lattice QCD, *J. High Energy Phys.* **07** (2007) 081.
- [2] M. Lüscher, Deflation acceleration of lattice QCD simulations, *J. High Energy Phys.* **12** (2007) 011.
- [3] J. C. Osborn, R. Babich, J. Brannick, R. C. Brower, M. A. Clark, S. D. Cohen, and C. Rebbi, Multigrid solver for clover fermions, *Proc. Sci.*, LATTICE2010 (2010) 037.
- [4] A. Frommer, K. Kahl, S. Krieg, B. Leder, and M. Rottmann, Adaptive aggregation-based domain decomposition multigrid for the lattice Wilson Dirac operator, *SIAM J. Sci. Comput.* **36**, A1581 (2014).
- [5] A. Frommer, K. Kahl, S. Krieg, B. Leder, and M. Rottmann, An adaptive aggregation-based domain decomposition multilevel method for the lattice wilson dirac operator: multilevel results, [arXiv:1307.6101](https://arxiv.org/abs/1307.6101).
- [6] M. Lüscher, OpenQCD software, version 1.2 and 1.4, <http://luscher.web.cern.ch/luscher/openQCD>.
- [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] M. A. Clark, J. Brannick, R. C. Brower, S. F. McCormick, T. A. Manteuffel, J. C. Osborn, and C. Rebbi, The removal of critical slowing down, *Proc. Sci.*, LATTICE2008 (2008) 035.
- [9] R. Babich, J. Brannick, R. C. Brower, M. A. Clark, S. D. Cohen, J. C. Osborn, and C. Rebbi, The role of multigrid algorithms for LQCD, *Proc. Sci.*, LAT2009 (2009) 031.
- [10] R. Babich, J. Brannick, R. C. Brower, M. A. Clark, T. A. Manteuffel, S. F. McCormick, J. C. Osborn, and C. Rebbi, Adaptive Multigrid Algorithm for the Lattice Wilson-Dirac Operator, *Phys. Rev. Lett.* **105**, 201602 (2010).
- [11] USQCD and J. C. Osborn, QCD operations API on top of the QDP/C library (QOPQDP), <http://usqed-software.github.io/qopqdp>.
- [12] M. Rottmann, A. Strebel, S. Heybrock, S. Bacchio, and B. Leder, DDalphaAMG, <https://github.com/DDalphaAMG/DDalphaAMG>.
- [13] 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 (2015).
- [14] S. D. Cohen, R. C. Brower, M. A. Clark, and J. C. Osborn, Multigrid algorithms for domain-wall fermions, *Proc. Sci.*, LATTICE2011 (2011) 030.
- [15] R. Frezzotti and G. C. Rossi, Chirally improving Wilson fermions. 1. O(a) improvement, *J. High Energy Phys.* **08** (2004) 007.
- [16] K. Jansen, M. Papinutto, A. Shindler, C. Urbach, and I. Wetzorke, Quenched scaling of Wilson twisted mass fermions, *J. High Energy Phys.* **09** (2005) 071.
- [17] A. M. Abdel-Rehim, R. Lewis, and R. M. Woloshyn, Spectrum of quenched twisted mass lattice QCD at maximal twist, *Phys. Rev. D* **71**, 094505 (2005).
- [18] M. Lüscher and F. Palombi, Fluctuations and reweighting of the quark determinant on large lattices, *Proc. Sci.*, LATTICE2008 (2008) 049.
- [19] M. Lüscher and S. Schaefer, Lattice QCD with open boundary conditions and twisted-mass reweighting, *Comput. Phys. Commun.* **184**, 519 (2013).
- [20] J. Finkenrath, F. Knechtli, and B. Leder, One flavor mass reweighting in lattice QCD, *Nucl. Phys.* **B877**, 441 (2013); Corrigendum, *Nucl. Phys.* **B880**, 574 (2014).
- [21] S. Bacchio and J. Finkenrath, DDalphaAMG, twisted mass version, <https://github.com/sbacchio/DDalphaAMG>.
- [22] R. Frezzotti and G. C. Rossi, Chirally improving Wilson fermions. II. Four-quark operators, *J. High Energy Phys.* **10** (2004) 070.
- [23] A. Abdel-Rehim *et al.*, Simulating QCD at the physical point with $N_f = 2$ Wilson twisted mass fermions at maximal twist, [arXiv:1507.05068](https://arxiv.org/abs/1507.05068).
- [24] L. Del Debbio, L. Giusti, M. Lüscher, R. Petronzio, and N. Tantalo, Stability of lattice QCD simulations and the thermodynamic limit, *J. High Energy Phys.* **02** (2006) 011.
- [25] M. Lüscher, Solution of the Dirac equation in lattice QCD using a domain decomposition method, *Comput. Phys. Commun.* **156**, 209 (2004).
- [26] Y. Notay and P. S. Vassilevski, Recursive Krylov-based multigrid cycles, *Numer. Linear Algebra Appl.* **15**, 473 (2008).
- [27] Y. Saad, Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems, *Math. Comp.* **42**, 567 (1984).
- [28] J. A. Vogel, Flexible BiCG and flexible Bi-CGSTAB for nonsymmetric linear systems, *Applied Mathematics and Computation* **188**, 226 (2007).
- [29] S. C. Eisenstat, H. C. Elman, and M. H. Schultz, Variational iterative methods for nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.* **20**, 345 (1983).
- [30] Y. Saad, A flexible inner-outer preconditioned GMRES algorithm, *SIAM J. Sci. Comput.* **14**, 461 (1993).
- [31] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo, *Phys. Lett. B* **195**, 216 (1987).
- [32] Steven A. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, Hybrid molecular dynamics algorithms for the numerical simulation of quantum chromodynamics, *Phys. Rev. D* **35**, 2531 (1987).
- [33] G. S. Bali, S. Collins, B. Glässle, M. Göckeler, J. Najjar, R. H. Rödl, A. Schäfer, R. W. Schiel, W. Söldner, and

- A. Sternbeck, Nucleon isovector couplings from $N_f = 2$ lattice QCD, *Phys. Rev. D* **91**, 054501 (2015).
- [34] H. Neff, N. Eicker, T. Lippert, J. Negele, and K. Schilling, On the low fermionic eigenmode dominance in QCD on the lattice, *Phys. Rev. D* **64**, 114509 (2001).
- [35] K. Jansen and C. Urbach, tmLQCD: a program suite to simulate Wilson twisted mass lattice QCD, *Comput. Phys. Commun.* **180**, 2717 (2009).
- [36] M. Hasenbusch and K. Jansen, Speeding up lattice QCD simulations with clover improved Wilson fermions, *Nucl. Phys.* **B659**, 299 (2003).
- [37] R. Babich, K. Barros, R. Brower, M. Clark, J. Foley, J. Giedt, S. Gottlieb, B. Joo, C. Rebbi, G. Shi, and A. Strelchenko, QUDA: a library for QCD on GPUs QUDA, version 0.8, <http://lattice.github.io/quda/>.
- [38] M. Rottmann, Ph.D. thesis, Bergische Universität Wuppertal, 2016.
- [39] S. Bacchio and J. Finkenrath, tmLQCD interfaced to DDalphaAMG, <https://github.com/Finkenrath/tmLQCD>.
- [40] S. Bacchio, C. Alexandrou, J. Finkenrath, A. Frommer, K. Kahl, and M. Rottmann, DDalphaAMG for twisted mass fermions, *Proc. Sci.*, LATTICE2016 (2016) 259.