Investigating a two-level algorithm for fermionic observables

Lorenzo Barca● and Stefan Schaefer●

John von Neumann-Institut für Computing NIC, Deutsches Elektronen-Synchrotron DESY,
Platanenallee 6, 15738 Zeuthen, Germany

Jacob Finkenrath●

Department of Physics, Bergische Universität Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany
and CERN, Esplanade des Particules 1, 1211 Geneva 23, Switzerland

We investigate the combination of a two-level sampling algorithm with distillation techniques to compute disconnected fermionic correlation functions. The method relies on a factorization of the quark propagator into domain-local contributions that depend only on the gauge fields within overlapping temporal regions, enabling independent submeasurements of each term through a two-level sampling strategy. The statistical uncertainty follows the expected inverse scaling with the number of second-level submeasurements, up to exponential boundary effects, in contrast to the inverse square-root scaling of standard sampling. This leads to an exponential reduction of statistical errors at negligible additional computational cost compared to the gain in statistical precision. The method is tested on pure gauge ensembles, providing a controlled benchmark for its forthcoming application to dynamical QCD studies.

I. INTRODUCTION

One ubiquitous challenge in lattice QCD simulations is the exponential degradation of the signal-to-noise ratio in correlation functions [1,2]. A solution to this problem would eliminate one of the most severe limitations currently hindering the precision and scope of lattice QCD calculations. For hadron structure observables, such as nucleon form factors [3] and semileptonic decay amplitudes of heavy-light mesons [4], the exponential degradation of the signal-to-noise ratio restricts the accessible source-sink separations, limiting control over excited-state contamination. At the same time, high-precision determinations of correlation functions at long Euclidean distances are crucial for phenomenologically important observables, such as the lattice QCD determination of the hadronic contributions to the muon $g-2$ [5]. The signal-to-noise problem is particularly severe in the study of glueballs, which suffer from notoriously poor signal-to-noise, see Refs. [6,7] for recent reviews, and are the central focus of this work.

In particular, the lattice computation of glueball observables is exacerbated because of disconnected contributions:

while the signal decays exponentially with the distance between the operators, the statistical noise remains constant with standard sampling techniques. Nonetheless, lattice simulations have provided robust predictions for the glueball spectrum in pure SU(N) Yang-Mills theories [8–13], as well as recent first investigations of their internal structure [14], offering valuable insights into the mass hierarchy, quantum numbers, and properties of these states in the absence of quarks. However, the identification of glueballs in nature remains elusive, as glueball states are expected to mix with conventional mesons and appear in decay channels involving multi-particle final states [15–17]. Experimental searches for glueballs are still ongoing, particularly in the radiative decays of charmonium states such as the J/ψ. Notably, recent results include the first identification of a glueball candidate with pseudoscalar quantum numbers [18].

Consequently, realistic and unambiguous determinations of glueball properties in full lattice QCD require correlation functions built from a comprehensive set of interpolating operators, including pure gluonic, mesonic, hybrid, and multihadron interpolators. This broad operator basis is essential to capture relevant decay channels, mixing patterns, and coupled-channel dynamics. However, such studies remain extremely challenging: the factorial growth of Wick contractions, the need for all-to-all quark propagators, and the poor signal-to-noise ratio of disconnected contributions at early time separations significantly limit the achievable precision.

Due to these challenges, robust investigations of glueballs in dynamical QCD remain very scarce. The few existing dynamical studies, such as Refs. [19–23], rely on relatively small variational bases, in some cases consisting only of pure-gluonic operators, which limit their ability to fully resolve the physics underlying glueball-meson mixing. Recent works have expanded the variational basis by exploiting distillation [24,25], enabling the inclusion of multihadron, fermionic, and pure-gluonic interpolators [26–29]. Nonetheless, their precision remains limited by the poor signal-to-noise ratio of disconnected contributions at early times or by the still restricted number of operators in the basis.

Alternative sampling methods, like the multilevel algorithm [30–33], can be used to reduce the error in a more efficient way than standard sampling. In [34,35], we demonstrate that a two-level sampling can reduce more efficiently the statistical error of pure gauge glueball two-point functions at large distances. A key idea behind the two-level algorithm is to exploit locality by splitting the temporal extent of the lattice into active regions separated by frozen regions, defined in [35], and performing N_1 independent submeasurements in each active region. When the source and sink operators lie in different active regions, the resulting factorization leads to a variance scaling ideally as $1/N_1^2$, up to corrections which saturate the two-level error reduction at large N_1 and which depend exponentially on the distance of the operators from the frozen regions.

However, the application of multilevel sampling with fermions in full QCD is not straightforward. After integrating out the fermionic degrees of freedom, the fermion observables are written as traces of products of quark propagators times the determinants of the quark propagators, which depend on the gauge fields over the full space time and therefore hinder the application of multilevel algorithms. By decomposing the quark propagator in the different regions that comprise the lattice extent, it is possible to make the fermionic observables amenable for multilevel integration. In quenched QCD, an important step forward has been made in [36,37] by rewriting the quark propagator as a series of terms with a well-defined hierarchical structure. This factorization of the quark propagator enables a two-level integration of the fermionic observables.

In full QCD, the application of multilevel integration requires more advanced techniques due to the presence of the fermion determinant as the fermionic weight. A factorization of the fermion determinant via a multiboson approximation [38] makes the fermion weight amenable for a local integration, as demonstrated in [39–41].

As a step toward the computation of the glueball spectrum in full QCD, we combine, for the first time, distillation techniques with a two-level algorithm in quenched QCD. This combined approach is particularly valuable for glueball studies: distillation tackles the factorial growth of Wick contractions arising from the multiparticle operator basis required in comprehensive dynamical analyses, while the two-level sampling mitigates the poor signal-to-noise ratio of disconnected contributions in isosinglet observables. A key question we investigate is whether two-level sampling can improve the signal already at sufficiently short operator separations, where it is already needed but where our pure-gauge study revealed limitations from boundary effects. This work thus provides a benchmark for future applications in full QCD, where a reliable determination of glueball properties requires several correlation functions with high statistical precision, and more broadly paves the way for next-generation studies of hadron correlation functions facing similar signal-to-noise challenges.

II. LATTICE QCD CORRELATION FUNCTIONS AND SIGNAL-TO-NOISE RATIO

We are interested in estimating two-point correlation functions of the form

$$\langle \mathrm{O}(\vec{p}, x_0)\bar{\mathrm{O}}(\vec{p}, y_0)\rangle = \frac{1}{\mathcal{Z}}\int [dU][dq][d\bar{q}]e^{-S[q,\bar{q},U]} \times \mathrm{O}(\vec{p}, x_0)\bar{\mathrm{O}}(\vec{p}, y_0), \quad (1)$$

where O denotes interpolating operators constructed to have overlap with specific states (e.g., glueballs), $\mathcal{Z}$ is the QCD partition function, and $S = S_f + S_g$ is the QCD action. After integrating out the fermionic fields, the correlation function in the quenched approximation becomes

$$\langle \mathrm{O}(\vec{p}, x_0)\bar{\mathrm{O}}(\vec{p}, y_0)\rangle = \frac{1}{\mathcal{Z}_g}\int [dU]e^{-S_g[U]}\langle \mathrm{O}(\vec{p}, x_0)\bar{\mathrm{O}}(\vec{p}, y_0)\rangle_F, \quad (2)$$

where $\langle\cdots\rangle_F$ denotes the Wick contractions expressed in terms of traces of products of Dirac propagators. In this work we focus on single-meson operators O_Γ and $\pi\pi$ interpolators projected to isospin $I = 0$, i.e., singlets under SU(2) isospin transformations. After projecting to zero total momentum, single-meson operators read

$$\mathrm{O}_\Gamma(\vec{0}, x_0) = \sum_{\vec{x}} \bar{q}(x)\Gamma q(x), \quad (3)$$

with $x = (\vec{x}, x_0)$, and $\Gamma = I, \gamma_5, \gamma_\mu, \gamma_4\gamma_5, \gamma_i\gamma_j$. Since in this work we consider only zero momentum, we drop the explicit momentum labels for simplicity.

In particular, the Wick contractions of the two-point functions of $\mathrm{O} = \mathrm{O}_\Gamma$ contain *disconnected quark-line* contributions, which read

$$\langle O(x_0)\bar{O}(y_0)\rangle_{F,\text{disc}} = \sum_{\vec{x},\vec{y}} \langle O(x)\rangle_F \langle \bar{O}(y)\rangle_F. \quad (4)$$

For instance, in the case of single-meson operators, the explicit expressions of the Wick contractions contain quark loops, which read

$$\mathcal{O}_\Gamma(x_0) := \sum_{\vec{x}} \langle O_\Gamma(x)\rangle_F = -\sum_{\vec{x}} \text{Tr}[D^{-1}(x,x)\Gamma]. \quad (5)$$

For scalar observables such as $\mathcal{O}_I$, the gauge average of the one-point function has a nonvanishing vacuum expectation value (VEV), which is constant in time.

The signal is obtained after subtracting this VEV from the two-point correlator,

$$C(x_0 - y_0)_{\text{disc}} = \langle \mathcal{O}(x_0)\bar{\mathcal{O}}(y_0)\rangle - \langle \mathcal{O}(x_0)\rangle\langle\bar{\mathcal{O}}(y_0)\rangle. \quad (6)$$

Correlation functions are estimated in lattice QCD [42] by generating a finite number N of gauge configurations U according to the QCD probability distribution, and computing ensemble averages via importance sampling. The physical signal decays exponentially with the operator separation,

$$C(x_0 - y_0)_{\text{disc}} \sim e^{-E|x_0 - y_0|}, \quad (7)$$

while the variance of disconnected quark-line contributions remains constant in time [2]. Consequently, the signal-to-noise ratio deteriorates exponentially [1],

$$|C(x_0 - y_0)_{\text{disc}}|/\sigma_C(x_0 - y_0) \sim \sqrt{N} e^{-E|x_0 - y_0|}. \quad (8)$$

Therefore, the analysis is restricted to the distances where the signal is still much larger than the noise.

In principle, statistical precision can be improved by increasing the number of gauge configurations N, but this becomes exponentially expensive at large separations. Maintaining a constant relative error requires $N \propto \exp(2E|x_0 - y_0|)$.

It is therefore essential to both develop algorithms that mitigate the exponential signal-to-noise degradation and design operators with improved overlap to the physical states. For the former, multilevel algorithms offer a powerful way to reduce the exponential signal-to-noise degradation by exploiting the locality of the action and of the observables. In particular, Refs. [34,35] show that, for pure-gauge observables, two-level sampling achieves a variance scaling as $1/N_1^2$ (with N_1 the number of second-level submeasurements), up to boundary effects that depend exponentially on the distance of the operators from the frozen regions.

As for the latter, quark-field smearing techniques such as *distillation* [24] can be combined with link smearing methods [43] to reduce excited-state contamination. A variational basis can then be constructed to solve a generalized eigenvalue problem (GEVP) and use the resulting eigenvectors to project the operators toward the desired states [44–46].

In this work we employ standard distillation, both to enhance quark-field smearing and to efficiently handle the increasing number of Wick contractions relevant for future multiparticle analyses. Within the distillation framework [24], the disconnected contractions in Eq. (4) can be expressed as

$$\langle O_\Gamma(x_0)\bar{O}_{\Gamma'}(y_0)\rangle_{F,\text{disc}} = \text{Tr}[\tau(x_0,x_0)\Gamma]\text{Tr}[\tau(y_0,y_0)\Gamma'], \quad (9)$$

where τ are the perambulators, whose expressions are

$$\tau_{nm}(z_0,z_0)_{\alpha\beta} = v_n^\alpha(z_0)^\dagger D^{-1} v_m^\beta(z_0), \quad (10)$$

with $z_0 = x_0, y_0$. Here v_n denotes the nth eigenvector of the three-dimensional Laplacian operator

$$\nabla^2(t)_{\vec{x},\vec{y}} = -6\delta_{\vec{x},\vec{y}} + \sum_{k=1}^{3}[U_k(\vec{x},t)\delta_{\vec{x}+\hat{k},\vec{y}} + U_k^\dagger(\vec{x}-\hat{k},t)\delta_{\vec{x}-\hat{k},\vec{y}}], \quad (11)$$

constructed from the gauge fields U. Each eigenvector is used as a spatial source for all Dirac components, so that the same set of spatial eigenvectors is employed for the inversions in every spin channel. This can be viewed as an embedding of the eigenvectors to the Dirac subspace. In Eq. (9), the sum over the eigenvector indices $n, m = 1, \ldots, N_v$ is implied. By truncating the sum to a small N_v, the quark fields are effectively smeared, producing extended interpolating operators built from the low modes of the Laplacian operator.

III. TWO-LEVEL SAMPLING FOR FERMIONIC OBSERVABLES IN QUENCHED QCD

A. Factorization of the quark propagator

In the quenched approximation, the two-point functions in Eq. (2) depend on the pure-gauge action $S_g[U]$ and on the Wick contractions $\langle O(x_0)\bar{O}(y_0)\rangle_F$. The action $S_g[U]$ is local in the gauge fields, being constructed from Wilson plaquettes.

For fermionic observables, the Wick contractions are traces of products of quark propagators [see, e.g., Eq. (4)], and each propagator depends on the gauge fields over the full spacetime. A factorization of the propagator, and hence of the fermionic correlation function, into domain-local pieces, first developed in Ref. [36], is therefore required for multilevel sampling.

We decompose the temporal extent of the lattice Λ into four regions, $\Lambda = \Lambda_0 \oplus \Lambda_1 \oplus \Lambda_2 \oplus \Lambda_3$ (Fig. 1), and
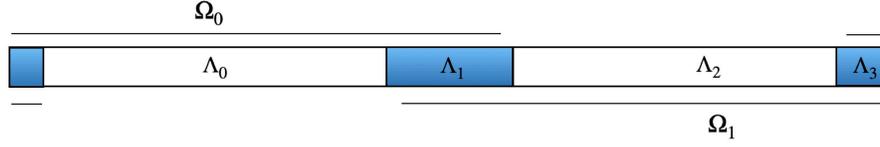
FIG. 1.   Domain decomposition of the temporal lattice extent used in this analysis. The temporal extent is factorized in four regions: $\Lambda = \Lambda_0 \oplus \Lambda_1 \oplus \Lambda_2 \oplus \Lambda_3$, and we consider the two overlapping domains $\Omega_0 = \Lambda_3 \oplus \Lambda_0 \oplus \Lambda_1$, and $\Omega_1 = \Lambda_1 \oplus \Lambda_2 \oplus \Lambda_3$.

introduce the overlapping domains $\Omega_0 = \Lambda_3 \oplus \Lambda_0 \oplus \Lambda_1$ and $\Omega_1 = \Lambda_1 \oplus \Lambda_2 \oplus \Lambda_3$.

With this domain decomposition, the factorized expression of a quark propagator traveling from x to x with $x_0 \in \Lambda_2$ reads [36,37,47]

$$D^{-1}(x, x)$$
$$= D_{\Omega_1}^{-1}(x, x) + D_{\Omega_1}^{-1}(x, \bullet) D_{\Lambda_{b0}} D_{\Omega_0}^{-1} D_{\Lambda_{b2}} (1 - w)^{-1} D_{\Omega_1}^{-1}(\bullet, x),$$
$$\tag{12}$$

where $\bullet$ represents all intermediate points, while $D_{\Lambda_{b0}}$ and $D_{\Lambda_{b2}}$ denote the hopping terms from all frozen regions to the active regions Λ_0 and Λ_2, and

$$w = P D_{\Omega_1}^{-1} D_{\Lambda_{b0}} D_{\Omega_0}^{-1} D_{\Lambda_{b2}} \tag{13}$$

with P being the projector to the surface of the boundaries. Employing the Neumann series,

$$\frac{1}{1-w} = \sum_{n=0}^{\infty} w^n, \tag{14}$$

which converges exponentially with n, yields

$$D^{-1}(x, x) = D_{\Omega_1}^{-1}(x, x) + D_{\Omega_1}^{-1}(x, \bullet) D_{\Lambda_{b0}} D_{\Omega_0}^{-1} D_{\Lambda_{b2}}$$
$$\times \sum_{n=0}^{\infty} (P D_{\Omega_1}^{-1} D_{\Lambda_{b0}} D_{\Omega_0}^{-1} D_{\Lambda_{b2}})^n D_{\Omega_1}^{-1}(\bullet, x). \tag{15}$$

A similar expression follows for the propagator from y to y with $y_0 \in \Lambda_0$. The first term in Eq. (15) describes propagation confined to Ω_1; the remaining terms account for propagation into Ω_0 and back. These contributions are exponentially suppressed at large n. Heuristically [1,48], on a given gauge field

$$\|D^{-1}(y, x)\| \sim \exp\left(-\frac{1}{2} m_\pi |y - x|\right), \tag{16}$$

with the norm $\|A\| = \sqrt{\text{Tr}\{AA^\dagger\}}$, so contributions mediated by distant regions decay with the distance. Accordingly, the norm of the nth term in the sum of Eq. (15) is suppressed as $\exp(-m_\pi \Delta n)$, where Δ is the thickness of the overlaps Λ_1, Λ_3.

Using Eq. (15), a two-level integration can be applied to the two-point function in Eq. (4). Retaining only the leading (fully local) term in Eq. (15), the two-level estimator for this first approximation to the isosinglet single-meson two-point functions with $\Gamma = \Gamma'$ reads

$$C_{\Gamma,(0)}^{2\text{lvl}}(x_0 - y_0) = \frac{1}{\mathcal{Z}_B} \int [dU_B] e^{-S_g[U_B]} [\mathcal{O}_\Gamma^{\Omega_1}(x_0)][\bar{\mathcal{O}}_\Gamma^{\Omega_0}(y_0)],$$
$$\tag{17}$$

where the subscript (0) denotes the first term in the factorization, and the local (second-level) averages of the quark loops are

$$[\mathcal{O}_\Gamma^{\Omega_1}(x_0)] = \frac{1}{\mathcal{Z}_g[U_1]} \int [dU_1] \sum_{\vec{x}} \text{Tr}[D_{\Omega_1}^{-1}(x, x)\Gamma], \tag{18}$$

with an analogous expression for $[\bar{\mathcal{O}}_\Gamma^{\Omega_0}(y_0)]$. If both loops reside in the same domain Ω_r, a two-level separation is not possible and the standard (global) average must be used,

$$C_{\Gamma,(0)}^{1\text{lvl}}(x_0 - y_0) = \frac{1}{\mathcal{Z}_g} \int [dU] e^{-S_g[U]} \mathcal{O}_\Gamma^{\Omega_r}(x_0) \bar{\mathcal{O}}_\Gamma^{\Omega_r}(y_0). \tag{19}$$

This one-level estimator is also used for global corrections (see below).

The statistical behaviour differs between the two schemes when observables are estimated with Markov-chain Monte Carlo [e.g., Hybrid Monte Carlo (HMC)]. While the variance of the standard estimator scales as $1/N_1$, the leading contribution to the two-level variance scales as $1/N_1^2$ (up to exponential boundary terms), mirroring the pure-gauge case [35].

A key difference relative to pure-gauge observables is that $D_{\Omega_1}^{-1}(x, x)$ is only an approximation to the full propagator $D^{-1}(x, x)$. To correct for this, we also compute the standard two-point function with the solver on the full lattice,

$$C_\Gamma^{1\text{lvl}}(x_0 - y_0) = \frac{1}{\mathcal{Z}_g} \int [dU] e^{-S_g[U]} \mathcal{O}_\Gamma(x_0) \bar{\mathcal{O}}_\Gamma(y_0), \tag{20}$$

and define the leading correction

$$\delta_\Gamma^{(0)}(x_0, y_0) = C_\Gamma^{1\text{lvl}}(x_0, y_0) - C_{\Gamma,(0)}^{1\text{lvl}}(x_0, y_0). \tag{21}$$

The improved two-level estimator is then

$$C_\Gamma^{2\text{lvl}}(x_0, y_0) = C_{\Gamma,(0)}^{2\text{lvl}}(x_0, y_0) + \delta_\Gamma^{(0)}(x_0, y_0). \tag{22}$$

Beyond the leading contribution, we also compute the next term in the Neumann series in Eq. (15), which provides a second-order approximation to the propagator. This first correction accounts for propagation from one active region into the other and back, and can itself be treated with a two-level integration (see Sec. III B). Including this contribution improves the accuracy of the approximation near the frozen boundaries while preserving the variance reduction of the two-level scheme. Higher-order terms can likewise be evaluated and would further extend the two-level error reduction in regimes where $\delta_\Gamma^{(n)}(x_0, y_0)$, corresponding to the globally added one-level correction to the nth approximation of the quark propagator, begins to dominate.

Alternatively, a multiboson approximation [38] could be used to treat all residual corrections; we find the direct correction via the full solver already efficient and do not pursue the multiboson approach here.

B. Analysis of disconnected two-points functions with one- and two-level estimators

The procedure to generate the gauge configurations follows the pure-gauge setup of Ref. [35]. First, N_0 gauge configurations are generated by updating the gauge fields over the full spacetime; these are the "boundary" fields U_B that appear in Eq. (17). Second, for each stored U_B we generate a sequence of N_1 gauge fields updated only in the active temporal regions Λ_0 and Λ_2, while Λ_1 and Λ_3 are frozen. In total, $N_0 \times N_1$ configurations are produced.

On these configurations we compute N_v Laplacian eigenvectors and estimate traces of quark loops using the full and the domain-restricted Wilson-Dirac propagators,

$$\mathcal{O}_{\Gamma,ij}(x_0) = \sum_{n=1}^{N_v} \text{Tr}[v_n^\dagger(x_0) D_{ij}^{-1} v_n(x_0) \Gamma], \quad (23)$$

$$\mathcal{O}_{\Gamma,ij}^{\Omega_r}(x_0) = \sum_{n=1}^{N_v} \text{Tr}[v_n^\dagger(x_0) D_{\Omega_r,ij}^{-1} v_n(x_0) \Gamma], \quad (24)$$

with $D_{\Omega_r}^{-1} = D_{\Omega_0}^{-1}$ for $x_0 \in \Omega_0$ and $D_{\Omega_r}^{-1} = D_{\Omega_1}^{-1}$ for $x_0 \in \Omega_1$. Here $i = 1, \ldots, N_0$ and $j = 1, \ldots, N_1$ label the gauge fields U_{ij}. The disconnected two-point function with the full propagator (standard sampling) is

$$C_\Gamma^{1\text{lvl}}(x_0, y_0) = \frac{1}{N_0 N_1} \sum_{ij} \mathcal{O}_{\Gamma,ij}(x_0) \bar{\mathcal{O}}_{\Gamma,ij}(y_0). \quad (25)$$

Using the approximated propagators instead, the one- and two-level estimators corresponding to Eqs. (17) and (19) read

$$C_{\Gamma,(0)}^{1\text{lvl}}(x_0, y_0) = \frac{1}{N_0 N_1} \sum_{ij} \mathcal{O}_{\Gamma,ij}^{\Omega_1}(x_0) \bar{\mathcal{O}}_{\Gamma,ij}^{\Omega_0}(y_0), \quad (26)$$

$$C_{\Gamma,(0)}^{2\text{lvl}}(x_0, y_0) = \frac{1}{N_0} \sum_i [\mathcal{O}_{\Gamma,i}^{\Omega_1}(x_0)][\bar{\mathcal{O}}_{\Gamma,i}^{\Omega_0}(y_0)]. \quad (27)$$

Here the square brackets denote second-level averages in the active regions [cf. Eqs. (17) and (18)], i.e. for the quark loops in Ω_1,

$$[\mathcal{O}_{\Gamma,i}^{\Omega_1}(x_0)] = \frac{1}{N_1} \sum_{j=1}^{N_1} \mathcal{O}_{\Gamma,ij}^{\Omega_1}(x_0). \quad (28)$$

The only difference between Eqs. (25) and (26) is the propagator approximation; the corresponding correction is

$$\delta_\Gamma^{(0)}(x_0, y_0) = C_\Gamma^{1\text{lvl}}(x_0, y_0) - C_{\Gamma,(0)}^{1\text{lvl}}(x_0, y_0). \quad (29)$$

Adding this to the two-level estimator of the first approximation gives the corrected result

$$C_\Gamma^{2\text{lvl}}(x_0, y_0) = C_{\Gamma,(0)}^{2\text{lvl}}(x_0, y_0) + \delta_\Gamma^{(0)}(x_0, y_0). \quad (30)$$

For the two-level method to be efficient, the fluctuations of $\delta_\Gamma^{(0)}$ (computed using one-level sampling) must remain much smaller than those of $C_{\Gamma,(0)}^{2\text{lvl}}$; otherwise the correction hinders the two-level gain. Since the norm of the nth term in Eq. (15) is expected to be suppressed as $\exp(-m_\pi n\Delta)$, the onset of saturation can be anticipated from the thickness Δ of the frozen regions and N_1. In order to improve the approximation and the statistical precision, additional terms of the Neumann series in Eq. (15) can be included in the two-level treatment.

Separating the $n = 0$ term, the first correction to the fully factorized loop is

$$\mathcal{O}_\Gamma^{(1)}(x_0) = \sum_{n=1}^{N_v} \text{Tr}[v_n^\dagger D_{\Omega_1}^{-1} D_{\Lambda_{b0}} D_{\Omega_0}^{-1} D_{\Lambda_{b2}} D_{\Omega_1}^{-1} v_n \Gamma], \quad (31)$$

$$\mathcal{O}_\Gamma^{(1)}(y_0) = \sum_{n=1}^{N_v} \text{Tr}[v_n^\dagger D_{\Omega_0}^{-1} D_{\Lambda_{b2}} D_{\Omega_1}^{-1} D_{\Lambda_{b0}} D_{\Omega_0}^{-1} v_n \Gamma], \quad (32)$$

for $x_0 \in \Omega_1$, $y_0 \in \Omega_0$ and for b we implicitly include contributions from the frozen regions. The corresponding first-order correction to the two-point function is estimated as

$$C_{\Gamma,(1)}^{1\text{lvl}}(x_0, y_0) = \frac{1}{N_0 N_1} \sum_{ij} [\mathcal{O}_{\Gamma,ij}^{(1)}(x_0) \bar{\mathcal{O}}_{\Gamma,ij}^{\Omega_0}(y_0) + \mathcal{O}_{\Gamma,ij}^{\Omega_1}(x_0) \bar{\mathcal{O}}_{\Gamma,ij}^{(1)}(y_0) + \mathcal{O}_{\Gamma,ij}^{(1)}(x_0) \bar{\mathcal{O}}_{\Gamma,ij}^{(1)}(y_0)]. \quad (33)$$

Physically, this correction accounts for contributions where one quark loop stays in one region and the other quark loop propagates to the other region and comes back, plus the contribution where both quark loops propagate to the other region and come back to the same region.

To perform a two-level integration, contributions supported in Ω_0 and Ω_1 must be separated and averaged independently. Numerically, we first average terms in Ω_0 and Ω_1 over gauge updates restricted to those domains, then average over boundary fields. To this end, we insert Laplacian eigenvectors and/or stochastic vectors at the boundary of the frozen regions to cut the fermionic lines in Ω_0 and Ω_1. This allows the second-level and independent sampling of the propagators in these regions. In particular, we approximate the identity via

$$1 \approx P_v + P_\perp \equiv \sum_{k=1}^{N_{\tilde{\eta}}} \tilde{\eta}_k (\tilde{\eta}_k)^\dagger, \qquad (34)$$

where

$$P_v = \sum_{\beta=0}^{3} \sum_{i=1}^{N_v^b} v_i^\beta (v_i^\beta)^\dagger, \qquad P_\perp = \frac{1}{N_{\eta^\perp}} \sum_{k=1}^{N_{\eta^\perp}} \eta_k^\perp (\eta_k^\perp)^\dagger. \quad (35)$$

In these expressions, P_v is the low-mode Laplacian projector, already used within distillation to compute smeared interpolating operators at source and sink, and $P_\perp$ is used to estimate stochastically the remaining part of the identity matrix. The latter is done by introducing stochastic vectors η_k, in this work Z_2 noise vectors, and computing

$$\eta_k^\perp = (1 - \mathbb{P}_v) \eta_k. \qquad (36)$$

Hence, the total number of *deflated* vectors in Eq. (34) is $N_{\tilde{\eta}} = 4N_v^b + N_{\eta^\perp}$. Notice that the number of eigenvectors used at the boundary of the frozen regions (N_v^b) can differ from the number of eigenvectors used to smear the interpolating operator (N_v). Choosing $N_v^b \leq N_v$ allows reuse of solution vectors $D_{\Omega_r}^{-1} v_n$ in both the leading term and the first correction.

For each *deflated* stochastic vector $\tilde{\eta}_k$, located at the time slice of the boundary, we compute

$$\mathcal{O}_{\Gamma,k}^{(1)}(x_0) = \sum_{n=1}^{N_v} \mathrm{Tr}\{v_n^\dagger D_{\Omega_1}^{-1} D_{\Lambda_{10}} D_{\Omega_0}^{-1} D_{\Lambda_{12}} \tilde{\eta}_k \tilde{\eta}_k^\dagger D_{\Omega_1}^{-1} v_n \Gamma\}, \quad (37)$$

and the analogous expression for $y_0 \in \Lambda_2$. As $N_{\tilde{\eta}} \to \infty$ this reproduces Eqs. (31) and (32). This stochastic representation enables a two-level average of Eq. (33). Using cyclicity of the trace, the two-level estimator of the first term becomes

$$\langle \mathcal{O}_\Gamma^{(1)}(x_0) \bar{\mathcal{O}}_\Gamma^{\Omega_0}(y_0) \rangle^{2\mathrm{lvl}} = \frac{1}{N_0} \sum_{i=1}^{N_0} \sum_{k=1}^{N_{\tilde{\eta}}} [\psi_k^{\Omega_1}(x_0)^\dagger]_i [\psi_k^{\Omega_0}(y_0)]_i, \qquad (38)$$

where

$$[\psi_k^{\Omega_1}(x_0)^\dagger]_i = \frac{1}{N_1} \sum_{j=1}^{N_1} (\tilde{\eta}_k^\dagger D_{\Omega_1}^{-1} v_n(x_0) \Gamma v_n^\dagger(x_0) D_{\Omega_1}^{-1} P_{\Lambda_{10}})_{|U_{ij}}, \qquad (39)$$

$$[\psi_k^{\Omega_0}(y_0)]_i = \frac{1}{N_1} \sum_{j=1}^{N_1} (D_{\Lambda_{10}} D_{\Omega_0}^{-1} D_{\Lambda_{12}} \tilde{\eta}_k \bar{\mathcal{O}}_\Gamma^{\Omega_0}(y_0))_{|U_{ij}}. \qquad (40)$$

Here U_{ij} denotes the gauge field used in the evaluation, and $\bar{\mathcal{O}}_\Gamma^{\Omega_0}(y_0)$ is a complex number for each U_{ij}. Furthermore, $P_{\Lambda_{10}}$ is a projector to the boundary Λ_{10}, which is used to reduce the dimensionality of the solution vector for disk storage.

The solution vectors $\chi_n = D_{\Omega_1}^{-1} v_n(x_0)$ are reused both here and in Eq. (24); the only additional inversions are those involving boundary sources, $D_{\Lambda_{10}} D_{\Omega_0}^{-1} D_{\Lambda_{12}} \tilde{\eta}_k$. These boundary inversions are independent of the specific source and sink positions x_0 and y_0, so they can be computed once and reused throughout. Placing the deflated vectors next to $D_{\Omega_0}^{-1}$ in Eq. (37) permits recycling of the low-mode solutions $D_{\Omega_r}^{-1} v_n$.

Similar expressions for the two-level estimators of the second and third terms in Eq. (33). The third term in Eq. (33) is analogous and is expected to be exponentially suppressed relative to the first two contributions, since it contains first-order corrections at both source and sink. We confirm this numerically and discuss it in the next section.

In particular, in the next section we present numerical results for the scalar ($\Gamma = I$) one-point functions from Eq. (24) and compare them to the unbiased one-level (full-solver) results in Eq. (23), thereby benchmarking the propagator factorization. We then compare standard and two-level estimates of the two-point functions for several choices of Γ.

Two important differences with respect to the pure-gauge study are:

(i) The two-level algorithm is applied to the leading and/or next-to-leading terms of the propagator factorization, with the residual global correction added from the one-level estimator; and

(ii) The thickness Δ of the frozen regions is larger than in the pure-gauge case. The increased thickness leads to a suppression of higher-order terms in the convergence series of the propagator and hence correlation functions. In addition, a sufficiently thick frozen region is required in the multiboson

factorization of the fermion determinant in dynamical QCD studies to ensure rapid convergence of the associated Neumann series [39,40,47].

As a consequence of the increased thickness, however, the two-level error reduction at fixed separation $|x_0 - y_0|$ is degraded: the two-level variance saturates at the level of fluctuations propagating from both the boundaries, which do not scale with $1/N_1$ [35]; when the frozen regions are thicker, the effective distance to those boundaries is smaller (at fixed $|x_0 - y_0|$), and saturation sets in earlier. A central goal of this study is to determine whether, despite these effects, two-level sampling remains more efficient than standard sampling for the fermionic observables considered here.

IV. NUMERICAL RESULTS

In this section we present our numerical results and address several key questions regarding the performance of the factorized two-point function and its corrections:

(i) Does the statistical error of the approximated two-point function in Eq. (27) decrease as $1/N_1$ with the number of level-1 updates?

(ii) Once the correction is included using standard sampling, does the error of the full two-point function in Eq. (30) still follow a $1/N_1$ scaling?

(iii) For which values of N_1 does the correction term obtained from standard sampling become significant and start to dominate the total error?

(iv) Is it sufficient to evaluate only the leading correction in Eq. (33) with a two-level sampling [e.g., via Eq. (38) for the first term] to further reduce the statistical uncertainty?

(v) Given that larger frozen regions are expected to impact two-level efficiency, do we obtain reliable signals for the effective masses?

In the following sections we describe the simulation setup and present results addressing these raised questions.

A. Details of the simulations

We simulate the four-dimensional SU(3) Yang-Mills theory using the Wilson gauge action with periodic boundary conditions. Gauge configurations are generated at $\beta = 6.0$, corresponding to $a = 0.0876$ fm [49], on a volume $V/a^4 = 64 \times 16^3$. To perform the two-level integration, we use molecular dynamics updates as in Ref. [35]. In the simulation, the frozen regions are $x_0/a \in \{0, 1, 61, 62, 63\}$ for Λ_3 and $x_0/a \in \{29, 30, 31, 32, 33\}$ for Λ_1, while the remaining time slices constitute Λ_0 and Λ_2.

We generate $N_0 = 101$ level-0 (boundary) configurations and for each of them a sequence of $N_1 = 1000$ level-1 (local) updates; the resulting fields are denoted U_{ij} with $i = 1, \ldots, N_0$ and $j = 1, \ldots, N_1$. On each U_{ij} we compute the full and the domain-restricted Wilson-Dirac

propagators using a clover-improved Wilson action with $\kappa = 0.13393$ and $c_{\text{SW}} = 1.769231$, for which the lightest pseudoscalar mass is found at $m_\pi \approx 760$ MeV. The mass is tuned so that the lowest noninteracting $\pi\pi$ energy lies close to the pure-gauge scalar glueball mass (at a similar coupling, $\beta = 5.99$, one finds $m_G^{0^{++}} \approx 1560$ MeV [8]). We use $N_v = 10$ Laplacian eigenvectors to smear the quark fields in the meson interpolating operators and invert the Dirac operator on these eigenvector sources.

To compute the approximated propagators we adopt a modified version of OPENQCD [50,51] imposing Dirichlet boundary conditions at the boundaries of the frozen regions. For instance, $D_{\Omega_0, ij}^{-1}(y_0, y_0)$ is computed on U_{ij} with inversions restricted to Ω_0, thereby neglecting contributions from Λ_2 (cf. Fig. 1).

B. One-point functions

For a first illustration of the convergence of the Neumann series adopted for the quark propagator, we compute scalar quark loops on the full set of $N_0 \times N_1$ configurations using the full propagator and the domain-restricted propagators [Eqs. (23) and (24)].

We also evaluate the first correction with standard sampling [Eqs. (31) and (32)], which is then used to improve the approximation to the second order. Figure 2 compares the scalar one-point functions (which have non-vanishing VEVs) obtained with the full propagator (green) and with the first (blue) and second (orange) approximations; orange points are x shifted for visibility. The full-propagator loops fluctuate around a constant, as expected. The first approximation to the scalar quark loop [cf. Eq. (28) with $\Gamma = I$] agrees with the full result when the loop is sufficiently far from the other active region, but deviations appear as the loop approaches the boundaries. Data at $x_0/a \in \{0, 30, 31, 32, 62, 63\}$ are always computed with the full propagator (hence identical in the comparison) and are noisier because the gauge links on those time slices are updated less than in the dynamical regions; this difference is more visible as N_1 increases. Including the first correction leads to a significant improvement of the approximation toward the exact loop near the boundaries, consistent with Eq. (16). In particular, within our level of accuracy, the results obtained with the second-order approximation are consistent, with the full scalar quark loops. Similar findings are obtained for the two-point functions, as discussed in the next section. The remaining effects will be taken into account in the final estimates of the two-point functions.

C. Two-point functions

Using Eqs. (25)–(27) with $N_0 = 101$ and N_1 values up to 1000 for selected time slices, we compare one- and two-level estimates of the disconnected two-point functions. Our goals are: (i) verify the expected $1/N_1^2$ variance scaling
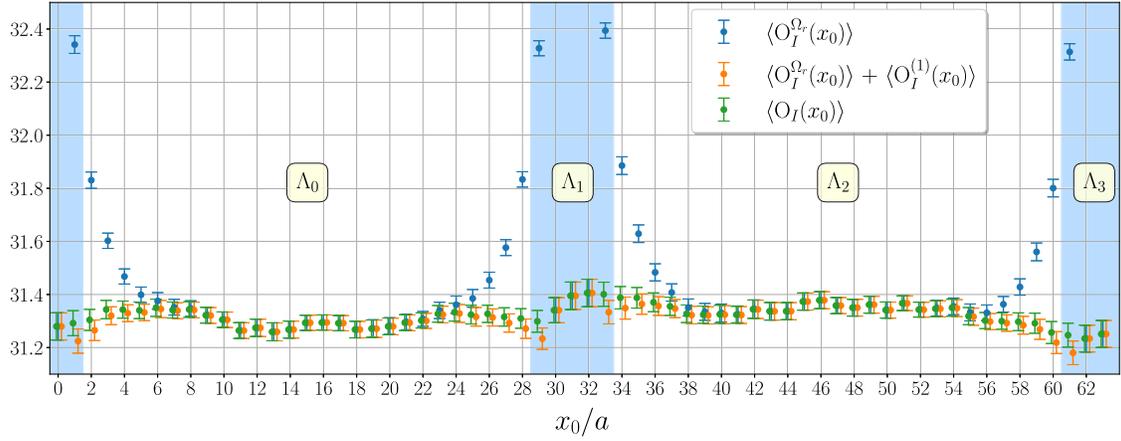
FIG. 2.   Scalar one-point functions obtained with the full propagator (green), the leading approximation (blue), and the leading approximation plus first correction (orange). Blue vertical bands mark the frozen regions Λ_1 and Λ_3. Estimates follow Eqs. (23) and (24).

of the *approximated* two-level estimator (versus $1/N_1$ for one level), and (ii) study the dependence on the distance of the operators from the frozen regions, as in the pure-gauge analysis [35]. Because Eqs. (26) and (27) are leading-order in the factorisation, we compute the global difference in Eq. (29) with one-level sampling and add it to the two-level result [Eq. (30)].

Figure 3 displays the error for the scalar disconnected correlator $C_I(x_0 = 40a, y_0)$ from the approximated and corrected two-level estimators at different values of $N_1 = 1$, 10, 200, 1000. For the red data points, $N_1 = 1000$ submeasurements are performed only in the range $15 \leq x_0/a \leq 42$ and $15 \leq y_0/a \leq 42$, while $N_1 = 200$ is used for the remaining time slices, resulting in a visible change in statistical precision around $y_0 = 15a$ for the red

circles. When both loops lie in the same domain, the error is approximately constant in y_0 and decreases as $1/\sqrt{N_1}$; when they lie in different active domains, the variance follows $1/N_1^2$ up to boundary-induced corrections that decay exponentially with the distance from the frozen regions. We observe clear $1/N_1^2$ scaling up to $N_1 \simeq 200$. For larger N_1, after adding the global correction computed with $N_0 = 101$, $N_1 = 200$, the total error no longer decreases: the correction's own uncertainty dominates. Indeed, the first correction in Eq. (33) [implicitly added by Eq. (29) along with the remaining terms] carries an error of that order (yellow points). Consequently, if only the leading approximation is treated with the two-level scheme, the corrected error cannot be reduced below the noise level of the correction unless the global difference is recomputed
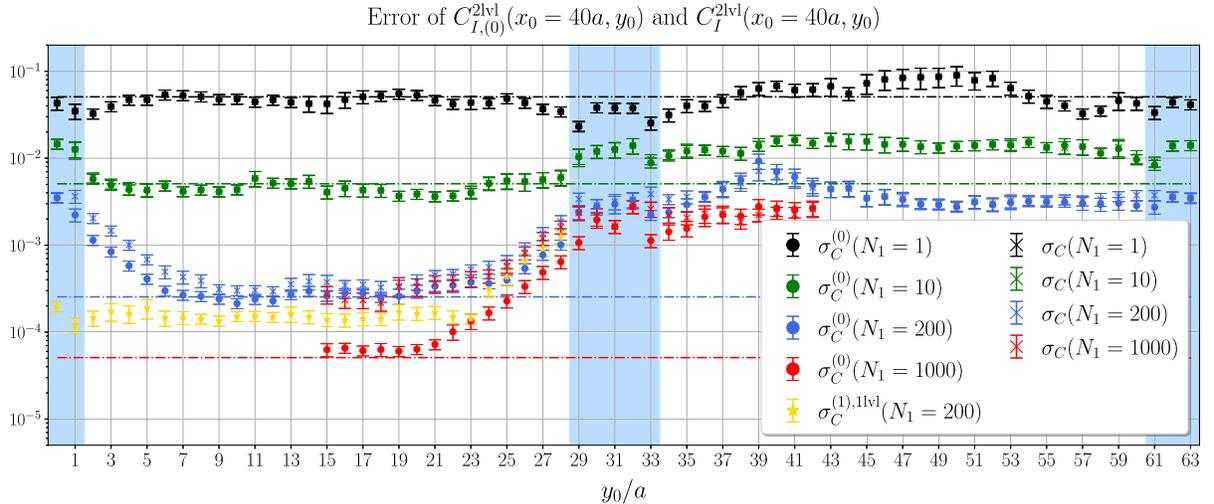

FIG. 3.   Statistical errors of the two-level estimators for the disconnected scalar two-point function before (circles) and after (crosses) adding the global correction of Eq. (29). Circles and crosses correspond to Eqs. (27) and (30), respectively. Blue vertical bands indicate the frozen regions Λ_1 and Λ_3. One loop is fixed at $x_0 = 40a$ while the other is varied. Dashed lines show ideal $1/N_1$ error scaling normalized to $\sigma_C^{(0)}(N_1 = 1)$ for $N_1 = 1$, 10, 200, 1000. Yellow points: statistical error of the one-level estimate of the first correction [Eq. (33)] with $N_0 = 101$, $N_1 = 200$.
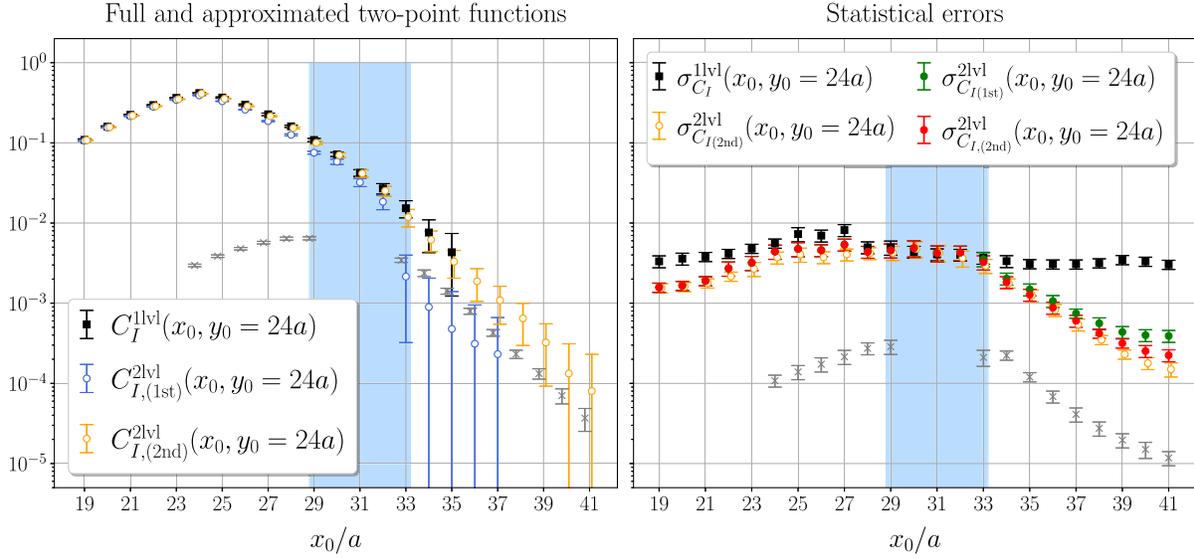
FIG. 4.   (Left) Signals and errors of the scalar disconnected two-point functions with the full solver (black) from standard sampling, with first approximation (blue), and with second approximation (orange) of the quark propagator from two-level sampling. The source is fixed at $y_0 = 24a$ and the sink position x_0 is varied. (Right) Statistical errors: empty markers are approximated results and full markers are corrected results. The grey crosses show the signal (left plot) and error (right plot) of the third term in Eq. (33) with one-level sampling using $N_0 = 101$ and $N_1 = 200$ configurations.

with higher statistics. A more efficient strategy is therefore to evaluate the first correction itself using a two-level approach, as discussed next.

a. Analysis with first correction Going beyond the leading approximation and including the second order into the multilevel scheme reduces the size of the uncertainty of the corresponding global correction. We therefore estimate the first correction to the two-point function with both one- and two-level sampling [Eqs. (33), (38)] using up to $N_{\eta^\perp} = 100$ stochastic vectors, deflated by $N_v^b = 10$ Laplacian eigenvectors as discussed in the previous section. Figure 4 shows the signal and errors of the two-point functions at fixed $y_0 = 24a$ and different x_0. The unbiased one-level (full-solver) signal is lost around $x_0 \approx 36a$ with $N_0 = 101$ and $N_1 = 200$, whereas the two-level estimators retain a clear signal at larger separations due to the exponential error reduction with boundary distance. Adding the two-level first correction $C_{I,(1)}^{2lvl}(x_0, y_0)$ to the leading approximation $C_{I,(1st)}^{2lvl}(x_0, y_0) := C_{I,(0)}^{2lvl}(x_0, y_0))$ yields a *second* approximation $C_{I,(2nd)}^{2lvl}(x_0, y_0))$ whose signal nearly coincides within the errors with the full one even near the boundaries, where higher-order terms matter most, see left plot in Fig. 4. The residual global difference (computed with $N_1 = 200$ as for the first approximation) is tiny; the corrected two-level second approximation agrees within errors with the pure two-level second approximation result.

With $N_0 = 101$ and $N_1 = 200$ (one level), the error of the third term in Eq. (33) is about an order of magnitude smaller than the first two and its statistical uncertainty

decreases exponentially with the distance from the boundaries (grey crosses in the right panel of Fig. 4); hence, we do not evaluate this term with two-level sampling. At fixed y_0, we find that the variance of the second term in Eq. (33) is approximately constant in x_0, while the first term decays exponentially with the distance to the boundaries.

b. Weighted average. Since correlators at different source times y_0 have different variances at fixed $\Delta t = x_0 - y_0$ (especially with two-level sampling), we form a weighted average [35],

$$\bar{C}^X(\Delta t) = \sum_{y_0} w(y_0 + \Delta t, y_0) C^X(y_0 + \Delta t, y_0), \quad (41)$$

with $X = 1lvl, 2lvl$. The weight functions are chosen to be proportional to the inverse of the variance, i.e., $w(x_0, y_0) = \mathcal{N}\sigma^2(x_0, y_0)^{-1}$, with the normalization $\mathcal{N}$ such that $\sum_{y_0} w(y_0 + \Delta t, y_0) = 1$.

D. Noise-to-signal ratio

Using the weighted correlators, we observe the following scaling for the variance,

$$\frac{\sigma_{\bar{C}}^2(\Delta t)}{\bar{C}^2(\Delta t)} \approx \frac{\tilde{a}_0^2}{N_0 N_1} e^{2m^\Gamma \Delta t} \quad \text{for } \Delta t \lesssim \Delta \tilde{t}, \quad (42)$$

$$\frac{\sigma_{\bar{C}}^2(\Delta t)}{\bar{C}^2(\Delta t)} \approx \frac{\tilde{c}_0^2}{N_0 N_1^2} e^{2m^\Gamma \Delta t} + \frac{2\tilde{c}_1^2}{N_0 N_1} e^{m^\Gamma \Delta t} + \frac{\tilde{c}_2^2}{N_0} \quad \text{for } \Delta t \gtrsim \Delta \tilde{t}, \quad (43)$$
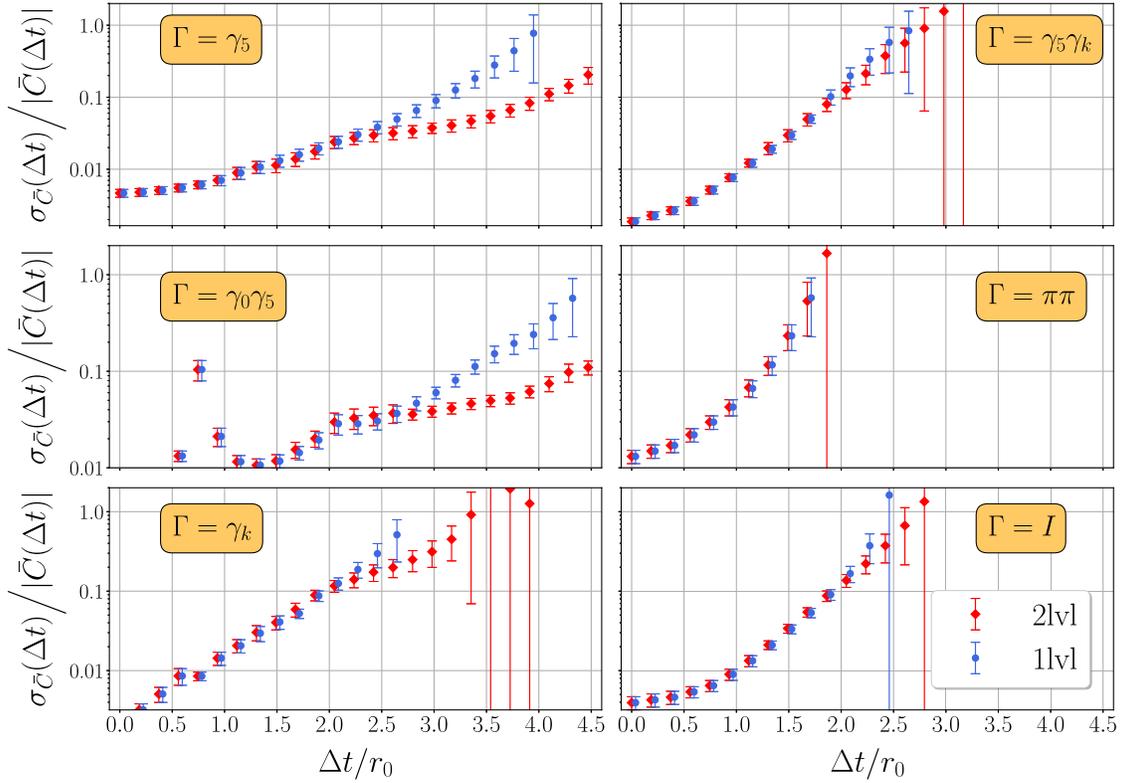
FIG. 5. Noise-to-signal ratios of weighted disconnected two-point functions for $O_\Gamma = \bar\psi\Gamma\psi$ and $\Gamma = I, \gamma_5, \gamma_\mu, \gamma_i\gamma_j\epsilon_{ijk}, \gamma_5\gamma_k$. One-level: $N_0 = 101$, $N_1 = 200$; two-level: $N_0 = 101$, $N_1 = 200$.

where m^Γ is the mass gap of the disconnected two-point functions in the channel Γ. The noise-to-signal scaling in Eq. (43) is typical with standard sampling algorithms. Figure 5 shows the noise-to-signal ratios for the disconnected correlators considered here, including the $I = 0$ of the $\pi\pi$ channel.

In particular, we can see the two different scalings in Eqs. (42) and (43): at short distance, the noise-to-signal ratio grows exponentially as fast as the one-level error. After some intermediate distance $\Delta\tilde{t}$, which depends on the channel, the two-level noise-to-signal error does not grow as fast as the one-level error, thus enabling the analysis of the correlation functions for larger distances. This scaling is similar to what observed in the pure-gauge study [35].

However, compared to the pure-gauge study, the two-level regime sets in at larger separations (in units of r_0). For instance, in the channel $\Gamma = I$, the crossover distance with $N_1 = 1000$ and similar N_0 and m_I is found at $\Delta\tilde{t} \approx r_0$ in the pure-gauge study, see top plot in Fig. (4) of Ref. [35], whereas in this quenched study we find it at $\Delta\tilde{t} \approx 2.3r_0$ with $N_1 = 200$. This is because the frozen regions are thicker: at fixed Δt, the effective distance to the boundaries is smaller and boundary fluctuations are larger. Adapting the crossover estimate of Ref. [35] to frozen regions thicker by $4a$, the transition occurs at

$$\exp\left(-m^\Gamma(\Delta\tilde{t} - 4a)\right) \lesssim \frac{1}{\sqrt{N_1}}. \qquad (44)$$

For instance, the two level starts to outperform the standard sampling at earlier time separations in the $\Gamma = I$ channel, compared to the $\Gamma = \gamma_5$, because we find that $m^I > m^{\gamma_5}$ at $m_\pi = 760$ MeV. This is a qualitative criterion because correlators/variances are not single exponential at short-intermediate distances, but it captures that heavier channels enter the two-level regime earlier.

E. Effective masses of singlet observables in quenched QCD

A preliminary study of the connected and disconnected contributions to the isosinglet two-point functions shows that the connected pieces do not suffer from a severe signal-to-noise problem using standard sampling techniques, whereas most of the statistical noise originates from the disconnected terms. We therefore compute the connected pieces with standard sampling and add the improved two-level estimators of the disconnected contributions. For $N_f = 3$, the full isosinglet weighted correlators are

$$\bar{C}_\Gamma(\Delta t) = \bar{C}_\Gamma(\Delta t)_{\text{conn}} - 3\bar{C}_\Gamma(\Delta t)_{\text{disc}}, \qquad (45)$$
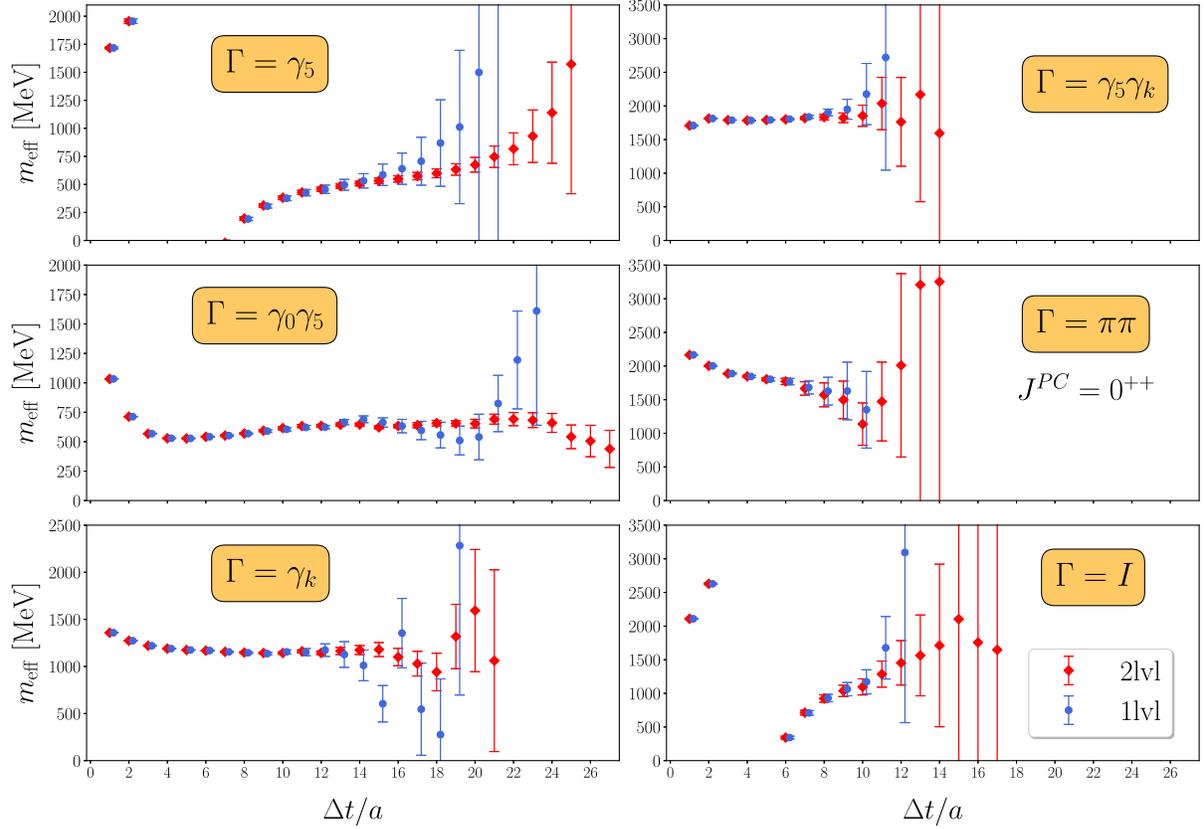
FIG. 6.   Effective masses of singlet two-point functions with $O_\Gamma = \bar{\psi}\Gamma\psi$ with $\Gamma = \gamma_5, \gamma_4\gamma_5, \gamma_k, \gamma_5\gamma_k, \gamma_i\gamma_j\epsilon_{ijk}$, and I.

with disconnected terms as defined in Sec. II. From the weighted averages [Eq. (41)] we form effective masses

$$m_{\text{eff}}(\Delta t) = \ln\left(\frac{\bar{C}(\Delta t)}{\bar{C}(\Delta t + a)}\right). \tag{46}$$

Figure 6 shows m_{eff} for $O_\Gamma = \bar{\psi}\Gamma\psi$ with $\Gamma = I, \gamma_5, \gamma_\mu, \gamma_i\gamma_j\epsilon_{ijk}, \gamma_i\gamma_5$, comparing one-level and two-level estimators with $N_0 = 101$ and $N_1 = 200$. The two-level results exhibit substantially improved signal-to-noise for the full isosinglet correlation functions. We refrain from performing mass fits, since reliable determinations would require a GEVP analysis with a sufficiently large operator basis, and quenched observables are subject to well-known artifacts. At large distances in the scalar channel, for example, the dominant state is $\eta'\pi$, whose energy lies below the f_0 because the η' remains light in the absence of fermion loops; see the partially quenched discussion in Ref. [52]. The disconnected pseudoscalar channel is likewise affected [53,54].

F. Cost budget

We now estimate the computational efficiency in terms of Dirac inversions, which constitute the dominant cost, and of the required ensemble size for a given target precision. While the two-level error decreases as $1/N_1$ in amplitude

(variance $\sim 1/N_1^2$) until boundary saturation, the one-level error falls only as $1/\sqrt{N_1}$. Thus, for fixed precision the required number of submeasurements differs.

In standard sampling, using distillation on $N = N_0 N_1^2$ configurations, the cost in inversions scales as $4N_v N_t \times N_0 N_1^2$ (factor 4 for spin indices). In the two-level approach, we first factorize the propagators and stochastically estimate boundary pieces using $N_{\tilde{\eta}} = 4N_v^b + N_{\eta^\perp}$ vectors [e.g., $D_{\Lambda_{10}} D_{\Omega_0}^{-1} D_{\Lambda_{12}} \tilde{\eta}_k$ in Eq. (40)], adding $2N_{\tilde{\eta}}$ solves per configuration. These boundary solves can be reused across sources/ sinks positions. If $N_v^b \leq N_v$, low-mode solutions $D_{\Omega_\ell}^{-1} v_n$ are reused for both the leading term and the first correction, so the $4N_v^b$ solutions are not an extra cost. To correct the approximation globally (when done by difference with the full solver), one computes the observable with the full propagator: this costs $4N_v N_t$ additional inversions per configuration.

Collecting terms, the additional inversions required by the two-level method are $4N_v N_t + 2N_{\tilde{\eta}}$ per configuration, i.e., a total of $(2 \cdot 4N_v N_t + 2N_{\tilde{\eta}})N_0 N_1$ inversions for leading plus first-correction two-level estimates (counting both domains). By contrast, the standard method requires $4N_v N_t \times N_0 N_1^2$ inversions to reach the same precision at large distances.

For the present parameters $N_v = 10, N_t = 64, N_0 = 101, N_1 = 1000, N_{\eta^\perp} = 50$, the standard sampling would require $\sim 2.6 \times 10^{11}$ inversions, versus $\sim 2.7 \times 10^8$ for the two-level

method. Numerically, at $x_0 = 27a$ in the scalar channel, using $N_v^b = N_v = 10$ accounts for roughly 50% of the first correction [black circles in Fig. 7(a)]; the remaining piece is captured with $N_{\eta^\perp} = 50$ deflated stochastic vectors. Hence, the extra inversions to reach the second approximation are modest, while the two-level sampling delivers an almost $N_1 \sim 10^3$ gain in precision at large distances.

In Fig. 7(b), we show the signal (left plot) and statistical errors (right plot) of the corrected second-order

approximation of the disconnected contribution to the scalar two-point functions with $N_{\eta^\perp} = 1$, 4, 10, 50, 100. We find that $N_{\eta^\perp} = 50$ is already sufficient to obtain an accurate stochastic determination of the signal, see left plot in Fig. 7(b), and to saturate the introduced stochastic noise in the two-level estimation of the first correction, at least up to $x_0 = 41a$ for $y_0 = 24a$, $N_0 = 101$, and $N_1 = 200$, see right plot in Fig. 7(b).

(a)

(b)
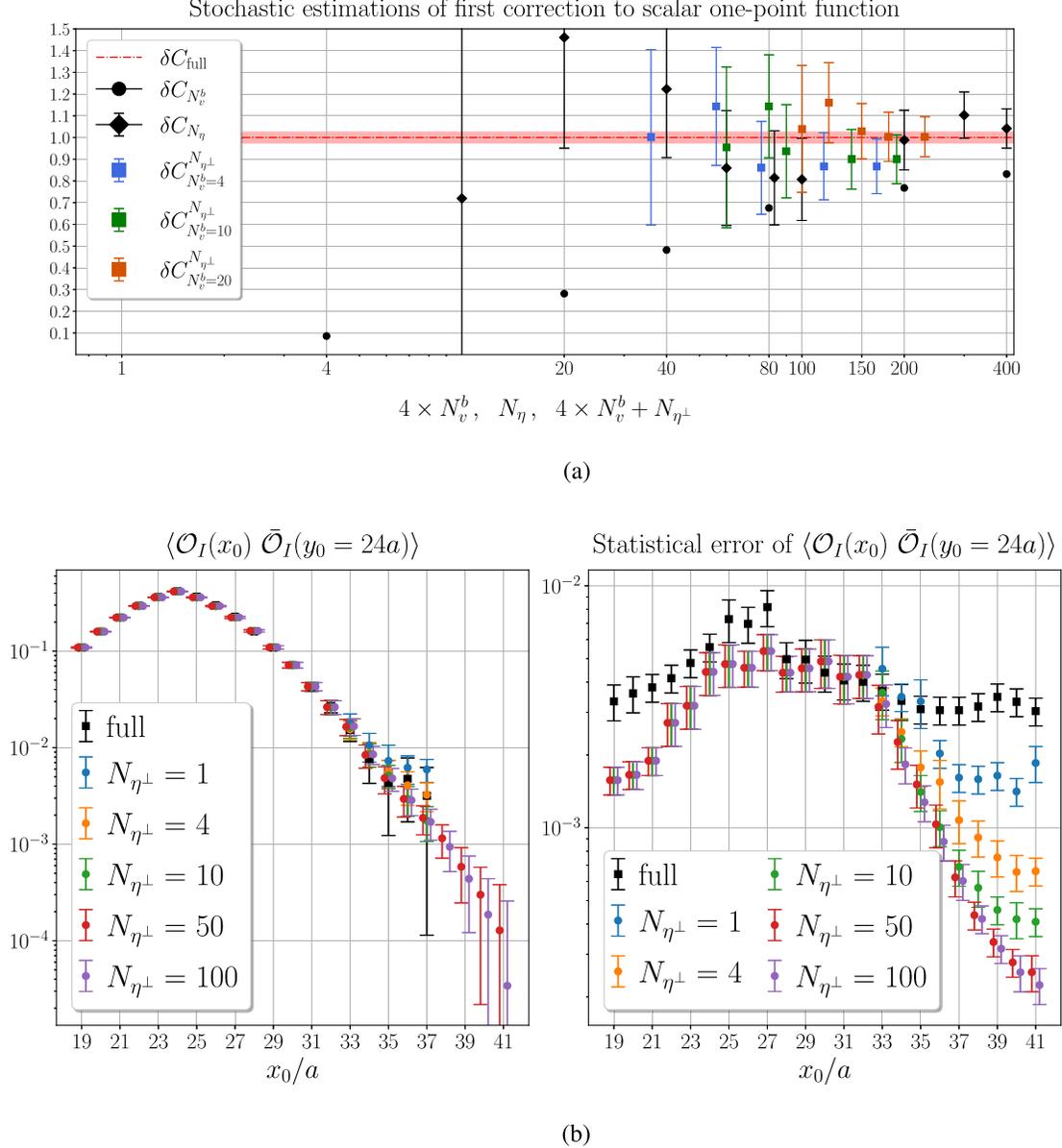
FIG. 7. Analysis of the saturation of stochastic noise for the first correction to one-point function estimator (a) and the two-point function estimator (b). (a) Signal for the normalized estimators of the first correction to the scalar one-point function at $x_0 = 27a$, δC, defined as the ratio of the stochastic estimator (Eq. (37)) to its standard determination (Eqs. (31) and (32)). The stochastic estimation uses $\tilde{\eta}_k$ taken exclusively as Laplacian eigenvectors (black circles), exclusively as stochastic vectors (black diamonds), or as deflated stochastic vectors (coloured squares). (b) Comparison of the signal for the estimator of the disconnected contribution to the scalar two-point functions between the exact estimator using standard sampling and the deflated stochastic estimator with $N_v^b = 10$ using two-level sampling.

Finally, we briefly comment on the memory require-ments. As in standard distillation, the dominant memory footprint is set by the storage of the Laplacian eigenvectors and perambulators. The two-level sampling strategy does not require storing additional large data structures. The 3D vectors entering Eqs. (39) and (40) can be accumulated on the fly and only need to be stored temporarily to allow for intermediate averaging over the second-level submeasure-ments. As a consequence, the overall memory cost of distillation combined with two-level sampling is very similar to that of standard distillation.

V. CONCLUSIONS

In this study, we investigate the application of two-level and standard sampling techniques to decrease the error of disconnected fermionic observables in quenched QCD, which hinder the reliable estimation of effective masses of singlet observables. To enable two-level sampling for quark loops, we factorized the quark propagator into contributions depending on two overlapping regions—comprising dynamical and frozen domains, thereby allowing independent sampling of the loop contributions. Compared to our previous pure-gauge study [35], we employed thicker frozen regions at the second level to ensure that correction terms are exponentially suppressed. In this work, we compute the first correction both with standard and two-level sampling—a key novelty of the present study—while the remaining higher-order corrections are evaluated globally using standard sampling by taking the difference with the correlation functions obtained from full propagators. We find that the two-level sampling outperforms standard sampling by significantly reducing statistical uncertainties at large distances, thereby enabling the analysis of correlation functions at larger distances. The high statistics of $N_0 = 101$ and $N_1 = 1000$ enable us to study the scaling behavior of the two-level error reduction, and we find that it is consistent with that observed in the pure gauge case. In particular, the constant term in the variance decreases as $1/N_1^2$, accompanied by exponential corrections originating from the frozen regions. As the thickness Δ of the frozen regions is larger than in the pure-gauge setup, where the dynamical domains were separated by only a single time slice, the two-level algorithm begins to outperform at larger separations, reducing the two-level gain in the error. We also find that efficient scaling of the two-level error reduction at large N_1 requires evaluating higher-order terms of the propagator series with a two-level scheme. In practice, for $N_0 = 101$ and $N_1 \approx 200$, the first correction becomes the dominant contribution to the statistical error of the disconnected piece of the scalar two-point functions, see Fig. 3. Computing this correction stochastically with a two-level sampling leads to a clear additional reduction in the variance, and we find that $N_{\eta^\perp} = 50$ stochastic vectors are already sufficient to saturate the induced stochastic noise for the distances

and for the statistics used in this work ($N_0 = 101$, $N_1 = 200$). As N_1 increases further, the higher-order terms in the two-point function expansion start to dominate the residual error as they decrease like $\sim 1/\sqrt{N_1}$ at fixed N_0. Consequently, these terms must also be computed with a two-level scheme if one aims to maintain the two-level error reduction at larger N_1. It is particularly promising that computing *only* the first approximation of the disconnected two-point function with a two-level scheme already reduces the long-distance error by a factor of ≈ 200 with $N_1 = 200$, as one can see from Fig. 3, showing that the leading term in the propagator series captures the dominant source of error. This study confirms the high efficiency of the two-level compared to the standard sampling and paves the way for applying this sampling technique for dynamical studies.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

[1] G. Parisi, The strategy for computing the hadronic mass spectrum, Phys. Rep. **103**, 203 (1984).

[2] G. Peter Lepage, The analysis of algorithms for lattice field theory, in *Theoretical Advanced Study Institute in Elementary Particle Physics* (TASI89, 1989).

[3] Lorenzo Barca, Current-enhanced excited states in lattice QCD three-point functions, Phys. Rev. D **112**, L091503 (2025).

[4] Alexei Bazavov *et al.*, D-meson semileptonic decays to pseudoscalars from four-flavor lattice QCD, Phys. Rev. D **107**, 094516 (2023).

[5] Alexei Bazavov *et al.*, Hadronic vacuum polarization for the muon g-2 from lattice QCD: Long-distance and full light-quark connected contribution, Phys. Rev. Lett. **135**, 011901 (2025).

[6] Davide Vadacchino, A review on Glueball hunting, Proc. Sci. LATTICE2022 (**2023**) 236, https://doi.org/10.48550/arXiv.2305.04869.

[7] Colin Morningstar, Update on glueballs, Proc. Sci. LATTICE2024 (**2025**) 004, https://doi.org/10.48550/arXiv.2502.02547.

[8] Andreas Athenodorou and Michael Teper, The glueball spectrum of SU(3) gauge theory in $3 + 1$ dimensions, J. High Energy Phys. 11 (2020) 172.

[9] Biagio Lucini, Antonio Rago, and Enrico Rinaldi, Glueball masses in the large N limit, J. High Energy Phys. 08 (2010) 119.

[10] Y. Chen *et al.*, Glueball spectrum and matrix elements on anisotropic lattices, Phys. Rev. D **73**, 014516 (2006).

[11] Biagio Lucini, Michael Teper, and Urs Wenger, Glueballs and k-strings in SU(N) gauge theories: Calculations with improved operators, J. High Energy Phys. 06 (2004) 012.

[12] Colin J. Morningstar and Mike J. Peardon, The glueball spectrum from an anisotropic lattice study, Phys. Rev. D **60**, 034509 (1999).

[13] G. S. Bali, K. Schilling, A. Hulsebos, A. C. Irving, Christopher Michael, and P. W. Stephenson, A comprehensive lattice study of SU(3) glueballs, Phys. Lett. B **309**, 378 (1993).

[14] Ryan Abbott, Daniel C. Hackett, Dimitra A. Pefkou, Fernando Romero-López, and Phiala E. Shanahan, Lattice evidence that scalar glueballs are small, Phys. Rev. Lett. **136**, 041901 (2026).

[15] Alexey A. Petrov, Glueball-meson molecules, Phys. Lett. B **843**, 138030 (2023).

[16] Eberhard Klempt and Andrey V. Sarantsev, Singlet-octet-glueball mixing of scalar mesons, Phys. Lett. B **826**, 136906 (2022).

[17] V. Crede and C. A. Meyer, The experimental status of glueballs, Prog. Part. Nucl. Phys. **63**, 74 (2009).

[18] Beijiang Liu, Discovery of a glueball-like particle X(2370) at BESIII, Proc. Sci. ICHEP2024 (2025) 490.

[19] Andreas Athenodorou, Jacob Finkenrath, Adam Lantos, and Michael Teper, Glueball Spectrum with four light dynamical fermions, arXiv:2308.10054.

[20] Long-Cheng Gui, Jia-Mei Dong, Ying Chen, and Yi-Bo Yang, Study of the pseudoscalar glueball in J/ψ radiative decays, Phys. Rev. D **100**, 054511 (2019).

[21] Wei Sun, Long-Cheng Gui, Ying Chen, Ming Gong, Chuan Liu, Yu-Bin Liu, Zhaofeng Liu, Jian-Ping Ma, and Jian-Bo Zhang, Glueball spectrum from $N_f = 2$ lattice QCD study on anisotropic lattices, Chin. Phys. C **42**, 093103 (2018).

[22] E. Gregory, A. Irving, B. Lucini, C. McNeile, A. Rago, C. Richards, and E. Rinaldi, Towards the glueball spectrum from unquenched lattice QCD, J. High Energy Phys. 10 (2012) 170.

[23] Christopher M. Richards, Alan C. Irving, Eric B. Gregory, and Craig McNeile, Glueball mass measurements from improved staggered fermion simulations, Phys. Rev. D **82**, 034501 (2010).

[24] Michael Peardon, John Bulava, Justin Foley, Colin Morningstar, Jozef Dudek, Robert G. Edwards, Balint Joo, Huey-Wen Lin, David G. Richards, and Keisuke Jimmy Juge, A novel quark-field creation operator construction for hadronic physics in lattice QCD, Phys. Rev. D **80**, 054506 (2009).

[25] Colin Morningstar, John Bulava, Justin Foley, Keisuke J. Juge, David Lenkner, Mike Peardon, and Chik Him Wong, Improved stochastic estimation of quark propagation with Laplacian heaviside smearing in lattice QCD, Phys. Rev. D **83**, 114505 (2011).

[26] Juan Andrés Urrea-Niño, Roman Höllwieser, Francesco Knechtli, Tomasz Korzec, Jacob Finkenrath, and Michael Peardon, S-wave flavor-singlet meson mixing in QCD with light and charm quarks, Phys. Rev. D **112**, 074502 (2025).

[27] Francesco Knechtli, Tomasz Korzec, Michael Peardon, and Juan Andrés Urrea-Niño, Optimizing creation operators for charmonium spectroscopy on the lattice, Phys. Rev. D **106**, 034501 (2022).

[28] Xiangyu Jiang, Wei Sun, Feiyu Chen, Ying Chen, Ming Gong, Zhaofeng Liu, and Renqiang Zhang, η-glueball mixing from Nf = 2 lattice QCD, Phys. Rev. D **107**, 094510 (2023).

[29] Ruairí Brett, John Bulava, Daniel Darvish, Jacob Fallica, Andrew Hanlon, Ben Hörz, and Colin Morningstar, Spectroscopy from the lattice: The scalar glueball, AIP Conf. Proc. **2249**, 030032 (2020).

[30] G. Parisi, R. Petronzio, and F. Rapuano, A measurement of the string tension near the continuum limit, Phys. Lett. **128B**, 418 (1983).

[31] Martin Lüscher and Peter Weisz, Locality and exponential error reduction in numerical lattice gauge theory, J. High Energy Phys. 09 (2001) 010.

[32] Harvey B. Meyer, Locality and statistical error reduction on correlation functions, J. High Energy Phys. 01 (2003) 048.

[33] Harvey B. Meyer, The Yang-Mills spectrum from a two level algorithm, J. High Energy Phys. 01 (2004) 030.

[34] Lorenzo Barca, Francesco Knechtli, Michael J. Peardon, Stefan Schaefer, and Juan Andrés Urrea-Niño, Performance of two-level sampling for the glueball spectrum in pure gauge theory, Proc. Sci. LATTICE2023 (**2024**) 030, https://doi.org/10.48550/arXiv.2312.11372.

[35] Lorenzo Barca, Stefan Schaefer, Francesco Knechtli, Juan Andrés Urrea-Niño, Sofie Martins, and Michael Peardon, Exponential error reduction for glueball calculations using a two-level algorithm in pure gauge theory, Phys. Rev. D **110**, 054515 (2024).

[36] Marco Cè, Leonardo Giusti, and Stefan Schaefer, Domain decomposition, multi-level integration and exponential noise reduction in lattice QCD, Phys. Rev. D **93,** 094507 (2016).

[37] Marco Cè, Leonardo Giusti, and Stefan Schaefer, Domain decomposition and multilevel integration for fermions, Proc. Sci. LATTICE2016 (**2016**) 263, https://doi.org/10.48550/arXiv.1612.06424.

[38] Martin Luscher, A new approach to the problem of dynamical quarks in numerical simulations of lattice QCD, Nucl. Phys. **B418,** 637 (1994).

[39] Marco Cè, Leonardo Giusti, and Stefan Schaefer, A local factorization of the fermion determinant in lattice QCD, Phys. Rev. D **95,** 034503 (2017).

[40] Marco Cè, Leonardo Giusti, and Stefan Schaefer, Local multiboson factorization of the quark determinant, EPJ Web Conf. **175,** 11005 (2018).

[41] Mattia Dalla Brida, Leonardo Giusti, Tim Harris, and Michele Pepe, Multi-level Monte Carlo computation of the hadronic vacuum polarization contribution to $(g_\mu - 2)$, Phys. Lett. B **816,** 136191 (2021).

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

[43] M. Albanese *et al.*, Glueball masses and string tension in lattice QCD, Phys. Lett. B **192,** 163 (1987).

[44] Benoit Blossier, Michele Della Morte, Georg von Hippel, Tereza Mendes, and Rainer Sommer, On the generalized eigenvalue method for energies and matrix elements in lattice field theory, J. High Energy Phys. 04 (2009) 094.

[45] Lorenzo Barca, Gunnar Bali, and Sara Collins, Toward N to $N\pi$ matrix elements from lattice QCD, Phys. Rev. D **107,** L051505 (2023).

[46] Lorenzo Barca, Gunnar Bali, and Sara Collins, Nucleon sigma terms with a variational analysis from lattice QCD, Phys. Rev. D **111,** L031505 (2025).

[47] Leonardo Giusti, Marco Cè, and Stefan Schaefer, Multiboson block factorization of fermions, EPJ Web Conf. **175,** 01003 (2018).

[48] Martin Luscher, Computational strategies in lattice QCD, in *Les Houches Summer School: Session 93: Modern perspectives in Lattice QCD: Quantum Field Theory and High Performance Computing* (2010), pp. 331–399, https://doi.org/10.48550/arXiv.1002.4232.

[49] Silvia Necco and Rainer Sommer, The $N(f) = 0$ heavy quark potential from short to intermediate distances, Nucl. Phys. **B622,** 328 (2002).

[50] Martin Lüscher, openQCD, http://luscher.web.cern.ch/luscher/openQCD/.

[51] Martin Luscher and Stefan Schaefer, Lattice QCD with open boundary conditions and twisted-mass reweighting, Comput. Phys. Commun. **184,** 519 (2013).

[52] William A. Bardeen, A. Duncan, E. Eichten, Nathan Isgur, and H. Thacker, Chiral loops and ghost states in the quenched scalar propagator, Phys. Rev. D **65,** 014509 (2001).

[53] Claude W. Bernard and Maarten F. L. Golterman, Partially quenched gauge theories and an application to staggered fermions, Phys. Rev. D **49,** 486 (1994).

[54] Stephen R. Sharpe, Chiral perturbation theory and weak matrix elements, Nucl. Phys. B, Proc. Suppl. **53,** 181 (1997).