

Exploratory calculation of $K_L \rightarrow \mu^+ \mu^-$ decay from lattice QCD at physical pion mass

Peter Boyle^{1,2}, En-Hung Chao³, Norman Christ³, Ceran Hu³, Luchang Jin⁴, and Yidi Zhao³

¹Physics Department, Brookhaven National Laboratory, Upton, New York 11973, USA

²School of Physics and Astronomy, The University of Edinburgh, Edinburgh EH9 3FD, United Kingdom

³Physics Department, Columbia University, New York City, New York 10027, USA

⁴Physics Department, University of Connecticut, Storrs, Connecticut 06269-3046, USA



(Received 23 September 2025; accepted 2 April 2026; published 30 April 2026)

We compute the complex, long-distance two-photon-exchange amplitude which contributes to the rare $K_L \rightarrow \mu^+ \mu^-$ decay from lattice QCD. We use a $24^3 \times 64$ physical-pion-mass gauge field ensemble at an inverse lattice spacing of 1.023 GeV and a QED_∞-based formalism. Our implementation strategies for all five non-SU(3)-flavor-suppressed diagram topologies are given in detail. We achieve a 25% statistical precision on the dispersive part of this long-distance amplitude. This calculation is carried out with 2 + 1 quark flavors and therefore requires the addition of counterterms to compensate for the absence of the Glashow-Iliopoulos-Maiani mechanism. These counterterms are not included in the current calculation and will be the subject of a second paper. Although a direct comparison to experiment cannot yet be made because of those omitted counterterms, the present exploratory calculation allows one to identify principal sources of statistical uncertainty in this calculation. The precision of our results is limited by the reconstruction of the physical contribution of the η intermediate state, for which various strategies are tested and compared.

DOI: [10.1103/22hn-ymsd](https://doi.org/10.1103/22hn-ymsd)

I. INTRODUCTION

The decay of a long-lived kaon into a pair of charged muons is a rare strangeness-changing neutral current process which is forbidden at first order in the standard model. Despite its rarity, clean signals have been obtained from experiment. In particular, the state-of-the-art measurement from the E871 experiment at the Brookhaven National Laboratory [1] has achieved a percent-level precision with the branching ratio [2]

$$\text{Br}(K_L \rightarrow \mu^+ \mu^-) = 6.84(11) \times 10^{-9}. \quad (1)$$

Because this decay is forbidden at tree level, it should provide a precision test of the standard model that is sensitive to physics from a high-energy scale. At one loop in the electroweak interaction, this decay results from the exchange of two W bosons and from two W bosons and one Z boson, entering at $\mathcal{O}(G_F^2)$, where G_F is the Fermi decay constant. This is a short-distance (SD) contribution that can be accurately calculated in electroweak and QCD

perturbation theory. Currently, the best estimate from perturbation theory for this SD part includes the effects from the charm quark calculated to next-to-next-to-leading order, which alone leads to a branching ratio of $\text{Br}(K_L \rightarrow \mu^+ \mu^-)_{\text{SD}} = 0.79(12) \times 10^{-9}$ [3].

However, a direct comparison with experiment is obscured by a significant contribution to this decay of order $G_F \alpha^2$, where α is the fine-structure constant. This third-order electroweak process involves the exchange of a single W boson and two photons, a process which we will abbreviate as LD2 γ . An accurate determination of this two-photon-exchange process is required if $K_L \rightarrow \mu^+ \mu^-$ decay is to be used to test the high-energy structure of the standard model at one loop. This two-photon-exchange amplitude is a long-distance (LD) process involving energies at the 1 GeV scale and below. This amplitude is nonperturbative and its first-principles determination requires a lattice quantum chromodynamics (LQCD) calculation.

Unlike the $\mathcal{O}(G_F^2)$ SD part, which contributes only to the dispersive part of the decay amplitude and will be referred as the SD G_F^2 amplitude below, the two-photon-exchange contribution is the sum of dispersive and absorptive parts. Throughout this work, we adopt the phase convention where the latter is purely imaginary and its algebraic sign will be specified later in the context of our lattice calculation. With this choice, the dispersive part of the LD2 γ decay amplitude is purely real.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

The absorptive part of the $\text{LD2}\gamma$ decay amplitude—and therefore its imaginary part—can be determined directly by using the optical theorem and the known $K_L \rightarrow \gamma\gamma$ decay rate [4]. This imaginary part alone results in a branching ratio of $\text{Br}(K_L \rightarrow \mu^+\mu^-)_{\text{abs}} = 6.59(5) \times 10^{-9}$, representing 96% of the experimental $K_L \rightarrow \mu^+\mu^-$ decay rate. Since to a good approximation the $\text{LD2}\gamma$ process gives the entire imaginary part of the decay amplitude, the magnitudes of both the real and imaginary parts of the $K_L \rightarrow \mu^+\mu^-$ decay amplitude $\mathcal{A}_{K_L\mu\mu}$ are known from experiment,

$$|\text{Re}\mathcal{A}_{K_L\mu\mu}^{\text{expt}}| = 1.53(14) \times 10^4 \text{ MeV}^3, \quad (2)$$

$$|\text{Im}\mathcal{A}_{K_L\mu\mu}^{\text{expt}}| = 7.10(3) \times 10^4 \text{ MeV}^3, \quad (3)$$

using conventions defined below [Eq. (21)].

In contrast, the real part of the $\text{LD2}\gamma$ amplitude cannot be determined from other measurable physical processes. Over many decades there have been attempts to determine the real part from phenomenological models [5–7]. Very recently, an analysis of the $K_L \rightarrow \gamma^*\gamma^*$ transition form factor based on dispersive techniques has been performed [8], where the authors find a mild tension of at least 1.7σ between the standard model and experiment from the real part of the $K_L \rightarrow \mu^+\mu^-$ decay amplitude. However, the relative sign between the $\text{LD2}\gamma$ and SDG_F^2 amplitudes cannot be resolved by any of these approaches. In fact, from the available phenomenological estimates, the real part of the combined $\text{LD2}\gamma$ and SDG_F^2 amplitudes has a different sign depending on whether the interference between these two amplitudes is constructive or destructive. Thus, a calculation from first principles which determines the sign of the $\text{LD2}\gamma$ amplitude relative to the SDG_F^2 term will be especially valuable in a comparison between the standard model and experiment.

Fortunately, with major advances in computer technology and substantial algorithmic developments, this complex $\text{LD2}\gamma$ process is now accessible to a first-principles, LQCD calculation. That such a complex process might be accessible to a LQCD calculation is suggested by the success of LQCD calculations of the hadronic light-by-light (HLbL) scattering contribution to the anomalous magnetic moment ($g-2$) of the muon. In this process the most important Feynman diagram contains a muon propagator joined to a closed quark loop by three exchanged photons [9–13], a process similar in complexity to the $\text{LD2}\gamma$ process being studied here.

As for the entire calculation of $g-2$ for the muon, the HLbL calculation is naturally performed in Euclidean space and is thus well suited to LQCD. In contrast, the $\text{LD2}\gamma$ process involves a physical Minkowski-space decay and is described by a complex amplitude. It is not a process that is usually studied using LQCD methods. However, since the imaginary part comes from the two on-shell photons, particles within the perturbative part of the process, a

LQCD calculation of the full complex amplitude is possible [14,15] and LQCD calculations of the similar but less complex $\pi^0 \rightarrow e^+e^-$ decay have been successfully performed [16].

An important purpose of this earlier work was to develop methods that could be applied to carry out the calculation presented here. Since QED can be accurately treated in an infinite-volume spacetime continuum using perturbation theory, we combine that formulation of QED with the finite-volume lattice treatment of QCD, a method referred to as QED_∞ . This detailed approach to the calculation of the $\text{LD2}\gamma$ amplitude was proposed and the accompanying systematic errors analyzed in detail in Ref. [17]. Note that a salient feature of QED_∞ is that, when the photon propagator is treated in this way for the case of exponentially localized hadronic correlation functions, finite-volume effects are exponentially suppressed in contrast to the power-law suppression seen in other finite-volume QED regularizations such as QED_L [10]. We expect this substantial advantage to apply here as well, although for the contribution of the $\pi\pi\gamma$ states, here argued to be negligible, further study would be needed to see if an advantage remains for QED_∞ .

The usual Wick rotation that allows Minkowski-space quantities to be computed using a Euclidean metric faces two obstacles in case of the $\text{LD2}\gamma$ amplitude. The first is the presence of propagator poles in the perturbative part of the amplitude that obstruct the usual Wick rotation of the contour over which an explicit integration of an internal energy variable is performed. As in the case of $\pi^0 \rightarrow e^+e^-$ decay, the contribution of these poles can also be obtained from Euclidean-space QCD amplitudes, allowing a LQCD calculation of the entire complex amplitude.

The second difficulty arises from the large-time behavior of the hadronic Green's functions when these functions are analytically continued in time during the Wick rotation of the integral over time that accompanies the Wick rotation of the internal loop energy integral. This difficulty can be treated by evaluating the decay amplitude resulting from the Schrödinger evolution over a finite time T_{max} and then taking the limit $T_{\text{max}} \rightarrow \infty$. Intermediate states in the decay process with energy less than the kaon mass, such as a single intermediate off-shell pion ($K_L \rightarrow \pi^0 \rightarrow \mu^+\mu^-$), result in a Wick-rotated time contour made up of individual segments which separately diverge exponentially at large T_{max} , although their sum remains finite.

Specifically, the Wick-rotated time-integration contour can be divided into two parts. The first is the usual integral in Euclidean time, $t_0 = i\tau$ with $0 \leq \tau \leq T_{\text{max}}$ (contour $\mathcal{C}_E$ in Fig. 1) which, because of these states, grows exponentially as $T_{\text{max}} \rightarrow \infty$. The second is the circular arc portion of the complex t_0 integral with a radius T_{max} joining the $t_0 = iT_{\text{max}}$ end point of the rotated contour on the imaginary, Euclidean-time axis to the original $t_0 = T_{\text{max}}$ end point on the real, Minkowski-time axis (contour $\mathcal{C}_{\text{arc}}$ in

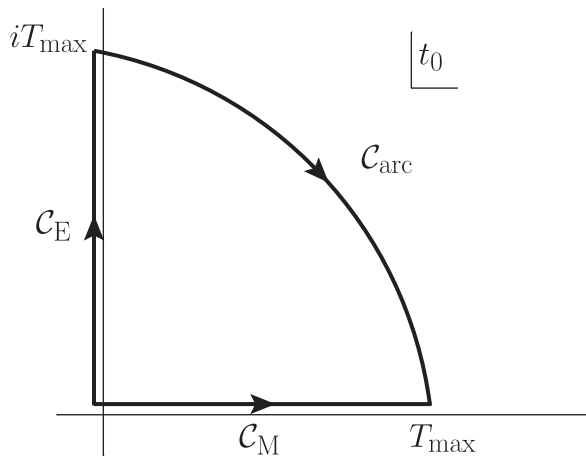


FIG. 1. The Wick-rotated time-integration contour described in the text ($C_E \cup C_{\text{arc}}$) and the original Minkowski contour (C_M). Cauchy's theorem implies that the integrals along both contours will be the same.

Fig. 1). While normally neglected as $T_{\text{max}} \rightarrow \infty$, when states $|n\rangle$ with $E_n < M_K$ are present, this semicircular contour also grows exponentially as $T_{\text{max}} \rightarrow \infty$, canceling the exponential growth in the Euclidean-time integration.

In a finite-volume calculation, such QCD energy eigenstates $|n\rangle$ are discrete. Their contributions are simple exponential functions of t_0 . As a result, the contributions from the integrals over these semicircular contours are simple exponentials in T_{max} . These terms can be directly evaluated from the product of matrix elements that can be separately calculated using LQCD and then used to remove the exponentially growing terms from a LQCD calculation in which such exponential growth appears.

In Ref. [17], the authors identify the intermediate states which are less energetic than the kaon and require the special treatment described above. For physical kinematics, the most important such intermediate states are a single pion (π^0) at rest and a finite-volume two-pion state with nonzero spatial momentum combined with a photon ($\pi\pi\gamma$) whose total energy in the kaon rest system is less than M_K . The π^0 state is straightforward to deal with as in previous calculations of the $K_L - K_S$ mass difference, ΔM_K [18,19].

The $\pi\pi\gamma$ intermediate state offers more serious challenges. First, with our QED_∞ treatment of electromagnetism, there can be one or more finite-volume $\pi\pi$ states with nonzero spatial momentum in the kaon rest system with a recoiling infinite-volume photon carrying the opposite momentum. (In fact, because of momentum nonconservation in QED_∞ , the $\pi\pi$ and photon momenta need not be equal and opposite, although such terms are suppressed by powers of $1/L$ and likely unimportant [17].) For free-particle kinematics $L \geq 11.8$ fm is required if the energy of a $\pi\pi\gamma$ state is to fall below M_K . If momentum nonconservation is considered then such effects will enter only for $L > 6$ fm. Thus, for our $L = 4.7$ fm volume no exponentially growing terms resulting from finite-volume

$\pi\pi\gamma$ states will occur. For the future studies of larger volumes, one may need to deal with multiple $\pi\pi\gamma$ states that each give exponentially growing behavior.

Once the exponential growing effects of states with energy less than M_K have been properly managed, the result is the desired Minkowski-space $\text{LD2}\gamma$ decay amplitude computed in finite volume. If the finite-volume hadronic amplitude being computed using LQCD contains no multiparticle intermediate states with energy below M_K , then the resulting position-space amplitude is exponentially localized and our QED_∞ result will have finite-volume errors that decrease exponentially in the linear size of that volume. However, this condition is not obeyed by the $\pi\pi\gamma$ state discussed above.

The two pions and photon in the $\pi\pi\gamma$ state with energy below M_K can be described as making up a “propagating” state with only power-law-suppressed finite-volume effects. Unlike the propagating two-pion intermediate states which enter the calculation of ΔM_K where a correction for the finite-volume effects of the $\pi\pi$ intermediate states is known [20], a correction for the potentially large finite-volume effects arising from the three-particle $\pi\pi\gamma$ states for our coordinate-space-based framework is at present unknown. However, a recent paper of Tuo and Feng [21] provides a valuable analysis of this problem. Fortunately, for the limited energy at or below M_K such three-particle states are expected to be phase-space suppressed and contribute little to the $\text{LD2}\gamma$ amplitude being studied.

In Ref. [17] a detailed phenomenological estimate of the size of the low-energy $\pi\pi\gamma$ state contribution is performed, concluding that these effects are at or below the 10% level. Thus, in this paper we will neglect the effects of these states and include a 10% systematic error resulting from this approximation. There are Green's functions in the calculation presented here in which the effects of a $\pi\pi\gamma$ might be seen. Our failure to observe such effects is consistent the expectation that they can indeed be neglected.

In this work, we apply the proposed computational strategy to a $24^3 \times 64$, inverse lattice spacing $1/a = 1.023$ GeV physical-pion-mass domain wall fermion ensemble (“24ID”) with $N_f = 2 + 1$ dynamical quark flavors from the RBC/UKQCD Collaboration. In the present work, we include three quark flavors in the valence sector. In the absence of the Glashow-Iliopoulos-Maiani (GIM) mechanism, the computed amplitude will contain unphysical, logarithmically divergent terms [22], which must be corrected by the addition of explicit counterterms. This issue will be addressed briefly in Sec. II B, but a complete treatment of those counterterms will be the subject of a future publication.

In spite of potentially large discretization effects due to the coarse lattice spacing, such an exploratory study allows us to develop and demonstrate our numerical implementation strategies and recognize possible challenges to overcome.

One such difficulty that is studied at length is the contribution of an intermediate η meson. Although we do

not have an exponentially growing contribution in the limit of large Euclidean time for the case of the η , owing to the typical temporal extent of the lattice and the expected unfavorable signal-to-noise ratio from a lattice calculation with quark-line disconnected diagrams, any practical truncation in the time integration will leave a large unphysical exponentially falling η contribution $[\propto \exp[-(M_\eta - M_K)T_{\max}]/(M_\eta - M_K)]$, since the mass of the η is very close to that of the kaon. This unphysical η contribution needs to be carefully removed with its physical part preserved.

As will be argued in Sec. II, with exact isospin symmetry, the electromagnetic (EM) current can be written as a linear combination of vector currents transforming as operators with isospin $I = 0$ and $I = 1$. Consequently, the product of two EM currents can be decomposed into combinations with $I = 0$ and 2 and $I = 1$. Since the $I = 1$ combination does not couple to the η , this term can be precisely determined and used to search for the effects of a possible $\pi\pi\gamma$ intermediate state.

This paper is structured as follows. In Sec. II, the notation and conventions used will be specified and the theoretical framework reviewed. The simulation details and numerical results will be discussed in Sec. III. In particular, strategies are developed and compared to reconstruct the physical contribution of the noisier $I = 0, 2$ operator product, a contribution which dominates our error budget. In Sec. IV, we address the relative sign between the computed $\text{LD2}\gamma$ amplitude and the earlier result for the SDG_F^2 amplitude. Finally, we conclude in Sec. V with a discussion of future plans and the limited conclusions that

can be drawn from our calculation performed at a single, relatively coarse lattice spacing.

The implementation details of the Wick contractions and the (approximate) low Dirac eigenmodes which allow us to improve the quality of the simulated data will be deferred to Appendixes A and B. The subtleties in obtaining $\mathcal{O}(a^2)$ -accurate results while treating the unphysical intermediate states are explained in Appendix C. A supplementary investigation of the effects of three-pion intermediate states is reported in Appendix D. Except where stated otherwise, all numerical quantities in this paper are expressed in lattice units.

II. THEORETICAL FRAMEWORK

In our calculation, we neglect CP violation and therefore treat all Cabibbo-Kobayashi-Maskawa (CKM) matrix elements as real. Likewise, we also use a K_L state with the definite $CP = -1$, which then requires the final muon pair to be in a parity odd and charge conjugation even 1S_0 state, where both the muon and antimuon have the same helicity. Figure 2, reproduced from Ref. [17], shows the combination of perturbative muon and photon propagators and QCD four-point function that must be combined to obtain the $\text{LD2}\gamma$ decay amplitude. Here and in the remainder of the paper we will work in the rest system of the kaon.

Although Euclidean-space lattice QCD is being used to compute this two-photon decay process, the result of this calculation is presented as a conventional Minkowski-space amplitude obtained from an expansion of the expression

$$\mathcal{A}_{\text{LD2}\gamma}(2\pi)^3\delta^3(\mathbf{k}^+ + \mathbf{k}^- - \mathbf{P}_K) = \left\langle \mu^+(\mathbf{k}^+)\mu^-(\mathbf{k}^-) \left| \text{T} \left\{ \exp \left[-i \int_{-\infty}^{\infty} dt H_{\text{EM}}(t) \right] H_W \right\} \right| K_L(\mathbf{P}_K) \right\rangle \quad (4)$$

to fourth order in the EM Hamiltonian H_{EM} . The QCD energy and momentum eigenstates in Eq. (4) obey the normalization condition $\langle \mathbf{p}' | \mathbf{p} \rangle = 2E(\mathbf{p})(2\pi)^3\delta^3(\mathbf{p}' - \mathbf{p})$

and H_W is the effective $\Delta S = \pm 1$ weak Hamiltonian. Both muons have the same helicity, $h = \pm \frac{1}{2}$.

A. Euclidean-space evaluation of the two-photon-exchange amplitude

In this section we apply the strategy developed in Ref. [17] and present the Euclidean-space Green's functions needed to compute the $\text{LD2}\gamma$ decay amplitude. We will neglect the effects of the low-energy $\pi\pi\gamma$ intermediate states but remove the unphysical, exponentially growing Euclidean-space contribution of the π^0 .¹

¹In Appendix D, a supplementary study on the effects of the three-pion intermediate states is presented. These intermediate states are often neglected in both phenomenological applications and in lattice calculations because of phase-space suppression. Our findings in Appendix D provide numerical evidence to support the latter statement.

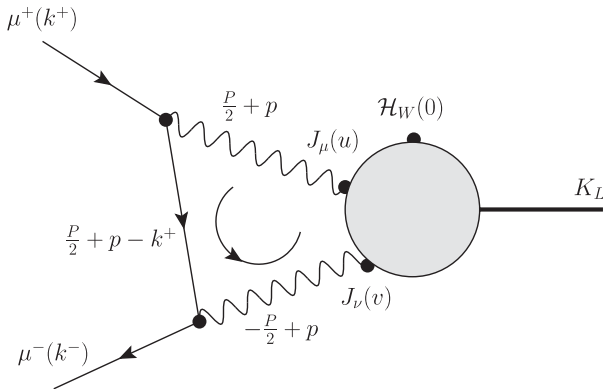


FIG. 2. Feynman graph representing the $\text{LD2}\gamma$ amplitude computed.

This calculation can be expressed concretely as the difference of two amplitudes. The first, $\mathcal{A}^{\text{lat}}$, results from the lattice calculation of the Euclidean-space four-point function with an upper limit δ_{max} imposed on the time difference between the smaller time v_0 and the time at which the weak operator $\mathcal{H}_W(0)$ is evaluated. Here we break the symmetry between the two currents by assuming the time order $u_0 \geq v_0$. Since at the highest level the two currents cannot be distinguished, this strategy can be viewed as a simple relabeling of the currents that allows us to identify the two operators between which the unphysical states propagate without affecting the final result. Of course at the level of our code the two currents can be distinguished by both their connection to the muon line and their connection to the quarks. Thus, one current might be distinguished as that connected to the photon whose other vertex is the point at which the μ^+ is emitted. However, because of the charge conjugation symmetry of the $\mu^+ \mu^-$ final state and our exact treatment of this QED portion of the amplitude, the case where this current is instead coupled to the photon whose second vertex is that at

which the μ^- is emitted is identical. Considering the different possible coupling to the quarks, there are many diagrams where the topology of the diagrams distinguishes the two currents and we must include cases where the current at v_0 enters each topology. In the case where both currents are coupled to the same quark line, charge conjugation will again imply that their order on this directed fermion line does not matter. However, this is only true statistically for the QCD portion of the graph so we evaluate both orderings to improve the final signal.

The second amplitude $\mathcal{A}^{\text{sub}}$ can be computed from three-point functions and is also a function of δ_{max} . This second amplitude provides the subtraction that removes the unphysical exponentially growing part of $\mathcal{A}^{\text{lat}}$.

The complete physical LD2 γ amplitude can then be written² as

$$\mathcal{A}_{\text{LD2}\gamma} = \mathcal{A}^{\text{lat}} - \mathcal{A}^{\text{sub}}, \quad (5)$$

where

$$\begin{aligned} \mathcal{A}^{\text{lat}}(\delta_{\text{max}}, \delta_{\text{min}}, R_{\text{max}}, t_{\text{sep}}) &= \int_{-\delta_{\text{min}}}^{\delta_{\text{max}}} dv_0 \int_V d^3 \mathbf{v} \int_{v_0}^{R_{\text{max}}+v_0} du_0 \int_V d^3 \mathbf{u} e^{M_K(u_0+v_0)/2} \\ &\times L_{\mu\nu}(\mathbf{u} - \mathbf{v}) \mathcal{N} e^{M_K t_{\text{sep}}} \langle T \{ J_\mu(u) J_\nu(v) \mathcal{H}_W(0) K_L(-t_{\text{sep}}) \} \rangle, \end{aligned} \quad (6)$$

$$\begin{aligned} \mathcal{A}^{\text{sub}}(\delta_{\text{max}}, \delta_{\text{min}}, R_{\text{max}}, t_{\text{sep}}) &= \sum_n \int_V d^3 \mathbf{v} \int_0^{R_{\text{max}}} dw_0 \int_V d^3 \mathbf{u} e^{M_K w_0/2} \left[\frac{e^{-(E_n - M_K)\delta_{\text{max}}}}{M_K - E_n} \right] \\ &\times L_{\mu\nu}(\mathbf{u} - \mathbf{v}, w_0) \mathcal{N} e^{M_K t_{\text{sep}}} \langle T \{ J_\mu(\mathbf{u}, w_0) J_\nu(\mathbf{v}, 0) \} | n \rangle \langle n | T \{ \mathcal{H}_W(0) K_L(-t_{\text{sep}}) \} \rangle. \end{aligned} \quad (7)$$

Here the operators $J_\mu(u)$ and $J_\nu(v)$ are the EM currents, $\mathcal{H}_W(0)$ is the weak Hamiltonian, $K_L(-t_{\text{sep}})$ a kaon interpolating operator and $\mathcal{N}$ a factor which normalizes that operator. The calculation is performed in the spatial volume V with a time extent T .

We refer to the factor $L_{\mu\nu}(\mathbf{u} - \mathbf{v})$ as the leptonic kernel. It is obtained from the muon and photon propagators shown in Fig. 2 after the integral over the loop momentum p is evaluated using an appropriately deformed contour for the loop energy p_0 so that the latter integral, performed over the difference of the time coordinates $u_0 - v_0$, converges. This is discussed at greater length in Refs. [16,17] and the leptonic kernel $L_{\mu\nu}(\mathbf{u} - \mathbf{v})$ is given explicitly in Eq. (20) in Ref. [17].

Equation (6) provides the details of the spacetime integrations performed when the lattice four-point function is evaluated. Both the positions $\mathbf{u}$ and $\mathbf{v}$ are integrated over V and the kaon interpolating operator is separated in time from the weak Hamiltonian by t_{sep} . We must take care that the most negative value of v_0 , $-\delta_{\text{min}}$, is sufficiently large to capture the complete contribution to the decay amplitude but

also small enough that the difference $t_{\text{sep}} - \delta_{\text{min}}$ is sufficiently large to guarantee that only a kaon state can be produced by $K_L(-t_{\text{sep}})$. The quantity R_{max} is the largest time separation allowed between the two currents and must be chosen large enough to capture the full contribution to the decay amplitude. A failure to observe a plateau as R_{max} is increased may be a sign of an unexpected contribution from a low-energy $\pi\pi\gamma$ state. A graphic illustration for the relevant time separations mentioned so far is given in Fig. 3.

Equation (7) provides the unphysical part of Eq. (6) which grows exponentially with increasing δ_{max} . That equation can be obtained from Eq. (6) in three steps. First, a sum over intermediate states $|n\rangle\langle n|$ with $E_n < M_K$

²The structure of Eqs. (6), (7), and (5) appears similar to that of Eqs. (16), (18), and (19) in Ref. [17]. However, the details are different since in the theoretical equation (16) in Ref. [17] all contributions from states with energy below M_K have been removed, while in the more practical lattice equation (6) no states have been removed and only an upper limit on the time δ_{max} has been imposed.

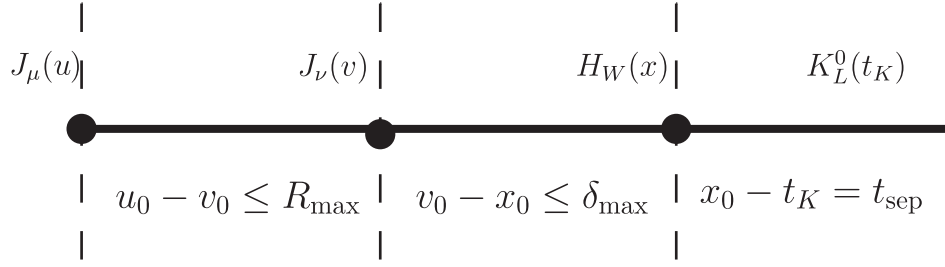


FIG. 3. Relevant separations involved in our formalism. Here we denote the position of the weak Hamiltonian with x instead of using the position 0 as in Eqs. (6) and (7). This more general location for $\mathcal{H}_W(x)$ is also used for clarity in some other places in the paper, e.g., Appendix A.

is inserted between $J_\nu(v)$ and $\mathcal{H}_W(0)$. Second, the integral over v_0 is performed over the interval $[0, \delta_{\max}]$ while the difference $w_0 = u_0 - v_0$ is held fixed. Third, the factor resulting from that v_0 integration $(e^{-(E_n - M_K)\delta_{\max}} - 1)/(M_K - E_n)$ is altered by dropping the physical “ -1 ” term in the numerator. In our treatment of the states with $E_n < M_K$ only the intermediate π^0 state is introduced.

The result, given in Eq. (7), is the unphysical term which grows exponentially as $\delta_{\max}$ is increased. This term must be subtracted from the lattice result given in Eq. (6) to obtain the finite physical result. Their combination, given in Eq. (5), should show a plateau as $\delta_{\max}$ is increased.

A similar approach can be used to more accurately treat the η intermediate state. If the η state is included in the right-hand side of Eq. (7), the unphysical term that results will decrease slowly with increasing $\delta_{\max}$. For practical values of $\delta_{\max}$ this unphysical term would be too large to neglect so that a similar subtraction is required.

The detailed treatment of Eqs. (6) and (7) for the case of a lattice calculation when the integral over time is replaced by a sum over discrete time values while avoiding $\mathcal{O}(a)$ discretization errors is given in Appendix C.

Henceforth, we will in addition use a more convenient definition of the $R_{\max}$ -regulated leptonic kernel on the lattice which respects the Bose symmetry of the decay amplitude

$$K_{\mu\nu}(r, R_{\max}) \equiv \begin{cases} 2L_{\mu\nu}(r)e^{M_K r_0/2}, & \text{if } 0 < r_0 \leq R_{\max}, \\ L_{\mu\nu}(\mathbf{r}, 0), & \text{if } r_0 = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

B. Decay amplitude from the lattice

In the continuum, the strangeness-changing $\Delta S = 1$ process can be represented by a weak effective Hamiltonian obtained by using QCD perturbation theory and the renormalization group to run down from the 100 GeV scale to a scale μ below the charm-quark mass [23]

$$\mathcal{H}_W^{\text{eff}} = \frac{G_F}{\sqrt{2}} V_{us}^* V_{ud} \sum_{i=1}^{10} C_i(\mu) Q_i^{\overline{\text{MS}}} + \text{H.c.} \quad (9)$$

where G_F is the Fermi decay constant, the V_{ij} are elements of the CKM matrix, and the $C_i(\mu)$ are Wilson coefficients at the

scale μ . The operators $Q_i^{\overline{\text{MS}}}$ are the ten linearly dependent four-quark $\Delta S = 1$ operators, which make up the traditional physical basis and are defined using QCD perturbation theory in the $\overline{\text{MS}}$ renormalization scheme. Both the Wilson coefficients and the operators $Q_i^{\overline{\text{MS}}}$ depend on the $\overline{\text{MS}}$ renormalization scale μ .

At energy scales of 1.5 GeV and above where the quark masses and chiral condensate might be neglected, three-flavor QCD becomes symmetric under $SU(3)_L \times SU(3)_R$ chiral flavor group and it is convenient to work with what is known as the “chiral basis,” a set of seven linearly independent four-quark operators each of which transforms in a specific representation of $SU(3)_L \times SU(3)_R$. Specific equations defining the seven operators in the chiral basis and for the ten operators in the physical basis can be found in Ref. [24] where the physical basis used here is referred to as “basis II.” In order to increase the accuracy of perturbation theory, as in previous RBC/UKQCD work, we will consider the three-flavor operators Q_i and their Wilson coefficients C_i at energies above the charm scale. When doing so we keep within the three-flavor theory and do not include a charm quark.

In order to express the weak Hamiltonian $\mathcal{H}_W^{\text{eff}}$, written in the physical basis with Wilson coefficients obtained from QCD and from electroweak perturbation theory, in terms of operators defined directly using LQCD we apply the same renormalization procedure as appears in the RBC/UKQCD Collaboration paper describing a $K \rightarrow \pi\pi$ calculation performed using the same 24ID ensemble [25]. In a series of steps the physical-basis $\overline{\text{MS}}$ operators given in Eq. (9) is transformed into an operator written in terms of the seven chiral-basis operators defined in the lattice formulation of QCD for the specific 24ID ensemble used in our calculation. These seven lattice operators Q_i^{lat} are multiplied by chiral-basis, lattice-renormalized Wilson coefficients C_i^{lat} .

Schematically, the Wilson coefficients multiplying the seven chiral-basis weak operators defined on the lattice can be obtained from those computed in the continuum using the physical basis in the $\overline{\text{MS}}$ scheme at a given high renormalization scale μ_h by equating the lattice and continuum operators,

$$\sum_{i=1}^{10} C_i(\mu_h) Q_i^{\overline{\text{MS}}} = \sum_{i=1}^{10} C_i(\mu_h) \left\{ \sum_{j=1}^7 \left[T + \frac{\alpha_s(\mu_h)}{4\pi} \Delta T \right]_{ij} \left\{ \sum_{k=1}^7 Z_{jk}^{\overline{\text{MS}} \rightarrow \text{RI}}(\mu_h) \right. \right. \right. \quad (10)$$

$$\left. \left. \left. \left\{ \sum_{\ell=1}^7 S_{k\ell}^{\mu_h \rightarrow \mu_l} \left\{ \sum_{m=1}^7 Z_{\ell m}^{\text{RI} \rightarrow \text{lat}}(\mu_l, a) Q_m^{\text{lat}} \right\} \right\} \right\} \right\} \right\} = \sum_{i=1}^7 C_i^{\text{lat}} Q_i^{\text{lat}}. \quad (11)$$

Equating the coefficients of the seven operators Q_i^{lat} in these two equations then determines the Wilson coefficients C_i^{lat} .

Reading from left to right on the right side of Eq. (10) we begin with the ten physical-basis $\overline{\text{MS}}$ operators with their Wilson coefficients defined at the high $\overline{\text{MS}}$ energy scale $\mu_h = 4.006$ GeV, chosen sufficiently large that perturbation theory should be reliable. The first matrix $T + \frac{\alpha_s(\mu_h)}{4\pi} \Delta T$, found in Ref. [24], converts those $\overline{\text{MS}}$ operators to the chiral basis. The second matrix $Z(\mu_h)^{\overline{\text{MS}} \rightarrow \text{RI}}$, computed in perturbation theory in Ref. [24], converts those $\overline{\text{MS}}$ chiral-basis operators into the regularization-independent (RI) symmetric momentum subtraction (SMOM) $(\not{q}, \not{q})$ scheme [25] at the same high-energy scale μ_h .

Next the step-scaling matrix $S^{\mu_h \rightarrow \mu_l}$ is defined by

$$S_{k\ell}^{\mu_h \rightarrow \mu_l} = \sum_{n=1}^7 Z(\mu_h, a')^{\text{RI} \rightarrow \text{lat}}_{k,n} \left[(Z(\mu_l, a')^{\text{RI} \rightarrow \text{lat}})^{-1} \right]_{n,m}. \quad (12)$$

This matrix expresses the seven chiral-basis operators defined at the perturbative scale μ_h in terms of those same operators defined at a lower-energy scale μ_l , accessible from our coarse 24ID ensemble. The two factors in Eq. (12) are computed nonperturbatively using a lattice with an inverse lattice spacing $1/a' = 3.1$ GeV in Ref. [25]. The intermediate lattice energy scale $1/a'$ is chosen to be larger than the scale of the 24ID ensemble to reduce the discretization errors associated with the higher RI scale μ_h .

The final matrix $Z_{\ell m}^{\text{RI} \rightarrow \text{lat}}(\mu_l, a)$ expresses the SMOM $(\not{q}, \not{q})$ operators defined at the scale μ_l in terms of the lattice operators whose matrix elements we compute on the 24ID ensemble. Thus, $1/a = 1.023$ GeV, the inverse lattice spacing of the 24ID ensemble.

Since the precision of our calculation is limited at the 10% level by our lack of control of the finite-volume errors associated with the $\pi\pi\gamma$ intermediate state, we chose to compute the matrix elements of only the most important “current-current” physical-basis operators, Q_1^{lat} and Q_2^{lat} . These should be the lattice operators with the largest

Wilson coefficients. The renormalization procedure outlined above provides the Wilson coefficients for the seven chiral-basis lattice operators Q_i^{lat} . Because of the linear dependence of the ten physical-basis operators there is ambiguity in determining the coefficients of the ten physical-basis operators in terms of the seven C_i^{lat} . We arbitrarily resolve this ambiguity by choosing the ten physical-basis lattice coefficients of $\{Q_i^{\text{lat}}\}_{1 \leq i \leq 10}$ to be those for which the coefficients of the three operators Q_4 , Q_9 and Q_{10} vanish.

We find

$$\begin{aligned} C_1^{\text{lat}} &= -0.312(14) - i1.34(33) \times 10^{-5}, \\ C_2^{\text{lat}} &= 0.718(14) + i1.68(33) \times 10^{-5}, \\ C_3^{\text{lat}} &= 0.018(14) + i7.5(3.2) \times 10^{-6}, \\ C_5^{\text{lat}} &= 0.006(11) - i8.7(2.3) \times 10^{-6}, \\ C_6^{\text{lat}} &= -0.009(9) + i3.27(17) \times 10^{-5}, \\ C_7^{\text{lat}} &= 8.40(1) \times 10^{-5} + i1.318(5) \times 10^{-7}, \\ C_8^{\text{lat}} &= -1.074(3) \times 10^{-4} - i3.183(9) \times 10^{-7}, \end{aligned} \quad (13)$$

where the uncertainties are purely statistical.³ We will ignore the imaginary parts of the Wilson coefficients in this calculation because we are computing the CP even decay amplitude.

The Wilson coefficients for the QCD and electroweak penguin operators Q_{3-10}^{lat} are much smaller in magnitude than the current-current operators, which are the only ones included in this work. However, because of the partial cancellation between the contributions of Q_1 and Q_2 that we found (see Table III), it is important to also estimate the sizes of the operators Q_{3-10}^{lat} to support the neglect of them as done in this work. A possible target in the near future would be Q_3 , as it is the largest of those neglected and similar computing strategies established in this work could be easily applied.

To evaluate the matrix elements required in Eqs. (6) and (7), we use a Coulomb-gauge-fixed wall source for the kaon interpolating operator to improve the overlap with the ground state. We use a local operator for the EM current, which is multiplicatively renormalized,

$$J_\nu^{\text{lat}}(v) = Z_V (\mathcal{Q}_u \bar{u} \gamma_\nu u(v) + \mathcal{Q}_d \bar{d} \gamma_\nu d(v) + \mathcal{Q}_s \bar{s} \gamma_\nu s(v)), \quad (14)$$

with physical EM charges

$$\mathcal{Q}_u = \frac{2}{3}, \quad \mathcal{Q}_d = -\frac{1}{3}, \quad \mathcal{Q}_s = -\frac{1}{3}. \quad (15)$$

³We are grateful to Masaaki Tomii for making the data generated for Ref. [25] available to us for the determination of the Wilson coefficients listed here.

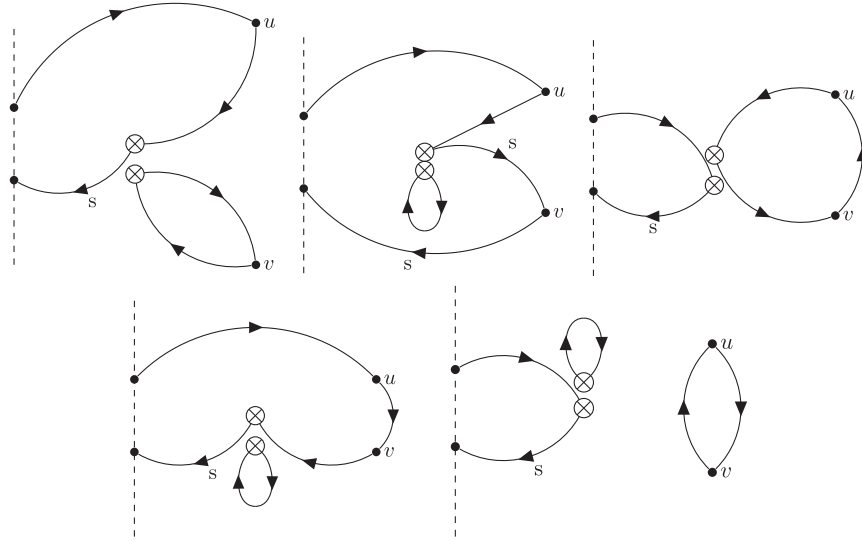


FIG. 4. The five types of Wick-contraction topology evaluated in this paper. From left to right, top to bottom: type 1, type 2, type 3, type 4 and type 5. The vertical dashed line represents the Coulomb-gauge-fixed kaon interpolating operator, the pair of circled crosses the four-quark weak operator and the vertices labeled u and v the two EM currents.

We neglect $SU(3)$ -flavor-suppressed diagrams, i.e., those with a self-contracted loop (“tadpole”) at one of the EM currents. The contribution of these diagrams is expected to be small as is the case in the $\pi^0 \rightarrow e^+e^-$ calculation [16].

We classify the Feynman diagrams representing the fermion Wick contractions needed to evaluate the four-point correlator appearing in Eq. (6) into two categories, the “connected” diagrams (types 1, 2, 3, 4), where all four operators in the correlation function are connected to each other by quark lines, and the “disconnected” diagrams (type 5) otherwise. The computed contraction classes are shown in Fig. 4 with one representative diagram per topology. Note that each type includes several different flavor combinations and either of the operators Q_1 and Q_2 . A complete list of Wick contractions can be found in Appendix A.

Finally, we should recognize the ultraviolet divergences present in type-1 and type-3 contractions when evaluated in our $2+1$ flavor theory, in spite of the renormalization procedure that has already been applied to the weak Hamiltonian. The treatment of these divergences is complicated by our use of nonconserved, local electromagnetic currents which will not obey the standard Ward identities for internal momenta close to the lattice scale. Since our ensemble has a lattice spacing too large to allow the introduction of a valence charm quark, the physical effects of the charm quark in this $2+1$ flavor calculation must be represented by counterterms which can be determined without reference to perturbation theory as suggested in Sec. V B of Ref. [26]. The determination of these counterterms and their matrix elements which enter this decay will be the subject of a future paper.

C. Isospin decomposition

Since the π^0 and η intermediate states behave very differently in our calculation it is convenient to have a mathematical technique to distinguish the amplitudes to which each contributes. Decomposing the product of the two EM currents according to their total isospin symmetry allows us to do that. In this way the statistical fluctuation from the unphysical π^0 does not interfere with the analysis of the isospin component where the π^0 is absent. For three flavors with the physical charge factors in Eq. (15), we can decompose the electromagnetic current J_μ into an isovector part J_μ^3 and an isoscalar part J_μ^8 ,

$$J_\mu = \frac{1}{2}J_\mu^3 + \frac{1}{6}J_\mu^8, \quad (16)$$

where

$$J_\mu^3 = \bar{\psi}\lambda^3\psi, \quad J_\mu^8 = \sqrt{3}\bar{\psi}\lambda^8\psi. \quad (17)$$

Here the λ^i are the conventional Gell-Mann matrices and $\psi = (u, d, s)^T$. We can then decompose the product of the two EM currents into two isospin components, one with $I = 1$ and the other with $I = 0$ or 2 ,

$$J_\mu(u)J_\nu(v) = \mathcal{J}_{\mu\nu}^{I=1}(u, v) + \mathcal{J}_{\mu\nu}^{I=0,2}(u, v), \quad (18)$$

where

$$\mathcal{J}_{\mu\nu}^{I=1}(u, v) = \frac{1}{12}(J_\mu^3(u)J_\nu^8(v) + J_\mu^8(u)J_\nu^3(v)), \quad (19)$$

$$\mathcal{J}_{\mu\nu}^{I=0,2}(u, v) = \frac{1}{4} \left(J_{\mu}^3(u) J_{\nu}^3(v) + \frac{1}{9} J_{\mu}^8(u) J_{\nu}^8(v) \right). \quad (20)$$

The amplitudes in Eqs. (6) and (7) can be decomposed according to this strategy leading to the isospin-specific amplitudes $\mathcal{A}_{I=1}^{\text{lat/sub}}$ and $\mathcal{A}_{I=0,2}^{\text{lat/sub}}$. The π^0 intermediate state contributes only to $\mathcal{A}_{I=1}^{\text{lat/sub}}$ while the η contributes only to $\mathcal{A}_{I=0,2}^{\text{lat/sub}}$.

III. NUMERICAL RESULTS

Here we discuss our numerical results for the $I = 1$ and $I = 0, 2$ amplitudes on the $24^3 \times 64$, physical quark mass ensemble (24ID) generated by the RBC-UKQCD Collaboration in separate subsections. This ensemble is generated with $2 + 1$ flavor Möbius domain wall fermions, whose simulation parameters can be found in Sec. II of Ref. [25]. The inverse lattice spacing for this ensemble is $1/a = 1.023$ GeV, and the pseudoscalar meson masses are $M_\pi = 142$ and $M_K = 515$ MeV. The renormalization factor for the vector current is $Z_V = 0.72672(35)$, determined in Ref. [27].

The present calculation makes extensive use of propagator data generated as part of the RBC/UKQCD HLbL project [28]. In particular, 64 Coulomb-gauge-fixed wall-source and 512 randomly distributed point-source propagators were made available for this work for all of the 110 configurations analyzed. These existing propagators reduce the computation cost of the current project, where all but the type-3 diagrams and a portion of the type-5 diagrams can be calculated without performing further inversions. See Appendix A for detailed descriptions of the computational strategy for each diagram.

In both the $I = 1$ and $I = 0, 2$ cases, we compute the total contribution to Eqs. (6), partially summed over the earliest time v_0 up to δ_{max} for $0 \leq \delta_{\text{max}} \leq T/2$, and (7) for several values of the parameters t_{sep} and R_{max} . Depending on the diagram, we collect data for 3–6 values of t_{sep} ranging from 6 to 16 lattice units and perform an error-weighted average over the results for all values of t_{sep} that are computed in order to improve the quality of the signal.

Having different values of R_{max} serves two purposes. First, we need to tune R_{max} to achieve a balance between omitting large values of u_0 that contribute to our amplitude and those that contribute only noise. Since R_{max} is the largest distance separating the two EM currents, we expect the propagation of a ρ meson to dominate the R_{max} behavior. For physical kinematics, the mass of the ρ meson in lattice units is about 0.8, hence one would expect a rapid saturation of the u_0 integral. However, having a value for R_{max} that is too large mainly introduces noise into the calculation.

Second, as discussed in Ref. [17], in a QED $_{\infty}$ calculation the inequality of QCD and QED volumes leads to the

nonconservation of momentum which allows a residual exponentially growing contribution in R_{max} which is power-law suppressed in the volume. Having results for different, sufficiently large values of R_{max} should help us determine if this effect is small enough to be neglected, as claimed in Ref. [17].

Our implementation of type-3 diagrams utilizes the all-to-all technique described in Appendix A 3 with 1000 Z-Möbius-approximated low modes with reduced extent in the fifth direction $L_s = 12$ (see Ref. [29] and references therein) and 16 hits for every time-diluted wall source with $\mathbb{Z}_2 \times \mathbb{Z}_2$ noise. For the tadpole term appearing in type 4 and type 5 [Eq. (A64)], we use a similar technique but with 64 spin-color diluted volume-source random vectors per configuration.

For the $I = 0, 2$ case, the contribution of the intermediate η at rest needs special care. Although formally converging at the $\delta_{\text{max}} \rightarrow \infty$ limit as $M_\eta > M_K$, the unphysical contribution, Eq. (7), falls off very slowly due to the small separation between M_η and M_K . As the $I = 0, 2$ part of the amplitude involves disconnected diagrams, the signal deteriorates well before the integral over the slowly falling η contribution has become saturated. In Sec. III B, we discuss several strategies to reconstruct the physical η contribution.

The numerical results on the decay amplitude $\mathcal{A}$ in this section are related to the physical decay rate $\Gamma_{K_L \rightarrow \mu^+ \mu^-}$ via

$$\Gamma_{K_L \rightarrow \mu^+ \mu^-} = \frac{\beta}{8\pi M_K} \left| e^4 V_{ud} V_{us}^* \frac{G_F}{\sqrt{2}} \mathcal{A} \right|^2, \quad \beta = \sqrt{1 - \frac{4m_\mu^2}{M_K^2}}, \quad (21)$$

where e is the fundamental electric charge. All the numerical values for the constants and parameters appearing above will be taken from Ref. [2]. We compute all the five types of diagram on the same 110 configurations separated by at least ten molecular-dynamics units in order to form the isospin components with the correlation between different types of diagram consistently accounted for.

A. $I = 1$ amplitude

We have obtained a very clear signal for the $I = 1$ channel since it contains only connected diagrams. The unphysical pion subtraction needs to be done with care. As explained in detail in Appendix C to avoid $\mathcal{O}(a)$ errors while integrating, the trapezoid rule is carefully followed at the boundary of the integration domain when evaluating the sum over v_0 in Eq. (6). We then subtract the exponentially growing term given by Eq. (7) (see Fig. 5). This continuum-determined subtraction is meaningful when the upper limit δ_{max} for the v_0 integral is large. If δ_{max} is one or two lattice spacings this continuum expression loses meaning. In order to plot the integrated expression determining the result in

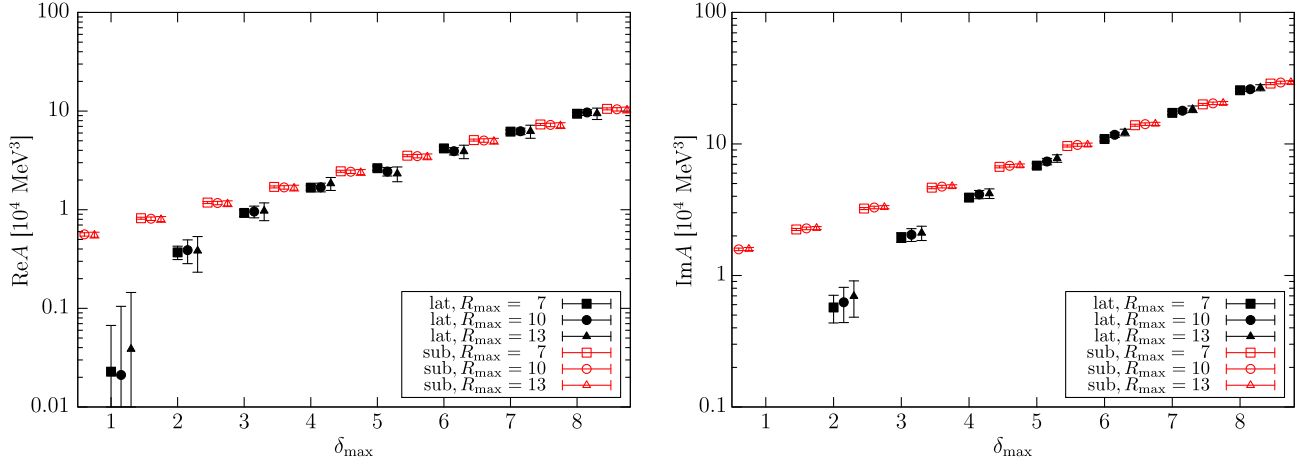


FIG. 5. The real (left) and imaginary (right) parts of the $I = 1$ components of Eqs. (6) (“lat”) and (7) (“sub”) as a function of $\delta_{\max}$, the upper limit on the integral over the time separation between the closest EM current and the weak operator.

Eq. (5) even for small $\delta_{\max}$ we adopt a reasonable but *ad hoc* subtraction for each v_0 . This subtraction under the sum sign, given in Eq. (C8), is chosen so that when summed to a large value of $\delta_{\max}$ an $O(a^2)$ -accurate result is obtained. Note that this subtraction is needed only when $v_0 \geq 0$.

The results for the partially integrated amplitudes are given in Fig. 6 as a function of the upper limit $\delta_{\max}$ for both the real and imaginary parts at three different values of $R_{\max}$ (7, 10 and 13). We see that plateaus are formed at small $\delta_{\max} > 0$, which indicates the insensitivity to long-distance physics of this observable. All results with different values of $R_{\max}$ are consistent with each other within the quoted statistical errors, suggesting that our smallest $R_{\max} = 7$ should be sufficiently large to capture the nonzero hadronic contributions. Within the quoted error, we do not see an obvious change in the central value as $R_{\max}$ is increased, confirming the expectation that for the current spatial volume there are no propagating two-pion states that must

be accommodated, a central assumption of our computational strategy.

We also do not see any sign of residual linearly or exponentially growing behavior in the plotted results as $\delta_{\max}$ increases. This suggests that only the pion state gives an unphysical contribution. In particular there is no suggestion of any contribution from a propagating three-pion state that is visible above the 10% statistical errors present for both the real and the imaginary parts when $R_{\max} = 7$ and $\delta_{\max} = 4$. We will use these values to obtain our final results.

B. $I = 0, 2$ amplitude

1. Type-5 diagrams

The $I = 0, 2$ channel involves disconnected diagrams (type 5), which require substantially more measurements than the other diagrams in order to achieve the same

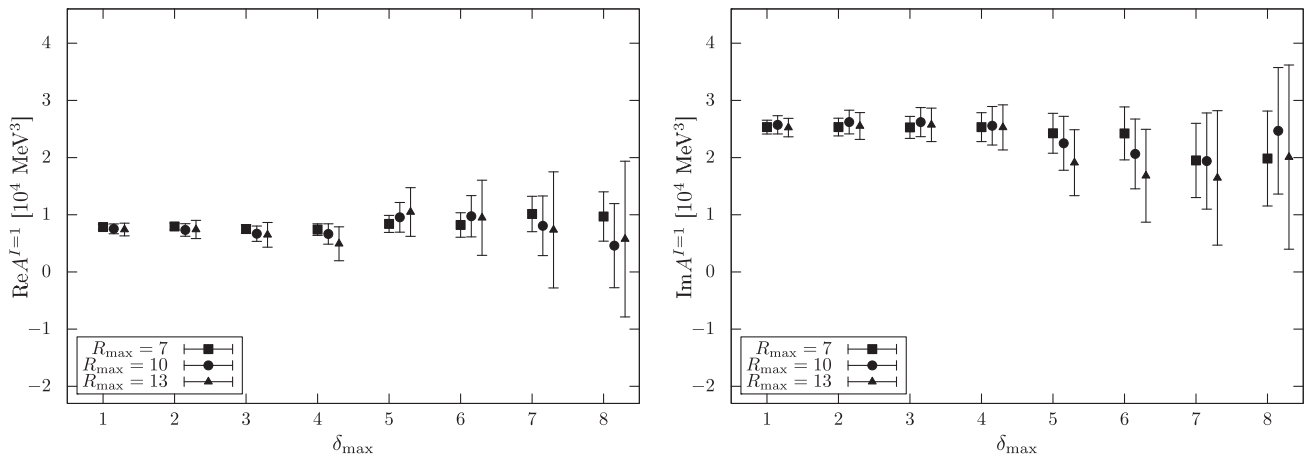


FIG. 6. The real (left) and imaginary (right) parts of the $I = 1$ amplitude from Eq. (5) as a function of $\delta_{\max}$.

statistical precision. We use a low-mode-improved method to calculate the type 5, diagram 1 (T5D1) as described in Appendix A 5, where the low-mode part is constructed using the 1000 eigenmodes of the Z-Möbius-approximated Dirac operator with the smallest eigenvalues.

The high-mode correction [Eq. (A89)] that must be added to the low-mode contribution is obtained from 2048 measurements per configuration using point-source propagators whose sources are located at the position of one of the EM currents and whose sink positions are the location of the second EM current. These propagators are evaluated with a “sloppy” stopping condition (a relative error of 10^{-4}) when performing a conjugate gradient inversion of the Dirac operator. With the Möbius-accelerated domain wall fermion solver [30], the computational cost of these sloppy inversions is 3 times less than if a higher precision of 10^{-10} were chosen. One quarter of these sloppy measurements are performed using the source positions used for the 512 higher-precision (relative error of 10^{-8}) propagators that were available. The difference between the results using these higher-precision and sloppy propagators is used to correct for the bias present in the sloppy inversions. With this strategy, the bias correction is about an order of magnitude smaller than the statistical fluctuations [31], which makes this an efficient strategy.

By themselves, the type-5 diagrams exhibit exponentially growing behavior due to the π^0 pole which is present in the disconnected amplitude and is canceled after the proper $I = 0, 2$ combination with type-3 and type-4 diagrams has been included. With the current statistics, we see that the final, combined $I = 0, 2$ amplitude has a rather clear signal without an unphysical π^0 contribution. (See the data points labeled “ η unsub.” in Fig. 10.)

2. Treatment of the η intermediate state

We propose two classes of methods (three different methods in total) to remove the slowly decreasing unphysical contribution of the η intermediate state. The first class is referred to as “direct subtraction” and corresponds to a single method (“method 1”), which is exactly the same procedure as used to treat the unphysical π^0 state in the $I = 1$ case.

In the second class the weak Hamiltonian is modified by subtracting a quark-bilinear operator whose on-shell matrix element is zero but whose coefficient can be chosen to make the off-shell $\langle \eta | \mathcal{H}_W | K_L \rangle$ matrix element vanish,

$$\mathcal{H}'_W = \mathcal{H}_W + c_s(\bar{s}d + \bar{d}s). \quad (22)$$

Such a modification of $\mathcal{H}_W$ will not change the physical result, but with the η contribution removed will permit truncating the v_0 integral at an early time before the statistical noise has become unacceptably large.

The on-shell matrix element of the operator $\bar{s}d + \bar{d}s$ vanishes because it is the divergence of the vector current

J_μ^{sd} which generates the flavor rotation [32]

$$\begin{pmatrix} d \\ s \end{pmatrix} \rightarrow 1 + \epsilon i \tau_2 \begin{pmatrix} d \\ s \end{pmatrix}, \quad \text{where } \tau_2 \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad (23)$$

is the conventional Pauli matrix. Using q to represent the two-component column vector of quark fields on the right-hand side of Eq. (23), we can write

$$J_\mu^{sd} = \bar{q} \gamma^\mu \tau_2 q. \quad (24)$$

The usual proof that the on-shell matrix element of the operator $\bar{s}d + \bar{d}s$ vanishes exploits the Ward identity obtained by replacing $\mathcal{H}_W(x)$ in the hadronic Green's function given in Eq. (6) with the current J_μ^{sd} and taking the divergence

$$\begin{aligned} & \frac{\partial}{\partial x_\lambda} \langle 0 | T \{ J_\mu^a(u) J_\nu^b(v) J_\lambda^{sd}(x) K_L(t_K) \} | 0 \rangle \\ &= (m_s - m_d) \langle 0 | T \{ J_\mu^a(u) J_\nu^b(v) (\bar{s}d + \bar{d}s)(x) K_L(t_K) \} | 0 \rangle \\ &\quad - i \langle 0 | T \{ \delta^4(x - u) [Q^{sd}, J_\mu^a(u)] J_\nu^b(v) K(t_K) \\ &\quad + \delta^4(x - v) J_\mu^a(u) [Q^{sd}, J_\nu^b(v)] K_L(t_K) \} | 0 \rangle, \end{aligned} \quad (25)$$

where we have adopted the Hilbert-space expression for the Ward identity since the two δ function terms can be easily written in terms of commutators with the conserved charge operator $Q^{sd} \equiv \int d^3 \mathbf{x} J_0^{sd}(\mathbf{x}, x_0)$. We have omitted a third δ function term involving the kaon interpolating operator $K(t_K)$ since in our calculation $t_{\text{sep}} = x_0 - t_K$ is always strictly greater than zero. Here we have introduced the labels a and b on the two currents to enable a separation of the EM current into $I = 0$ and $I = 1$ components.

If the two J^i 's are taken to be the full EM currents with equal charges for the down and the strange quarks, then $[Q^{sd}, J_\mu^{\text{em}}] = 0$ and we will arrive at the usual conclusion that the combination $\bar{s}d + \bar{d}s$ is a total derivative (“method 2”). However, if one considers products of the individual $I = 0$ and $I = 1$ components separately in the product of the two EM currents in Eq. (25) (“method 3”), one needs to include additional contact terms due to the nonvanishing commutators $[Q^{sd}, J_\mu^i]$ when $i = 3$ or 8 . Simple algebra leads to

$$\begin{aligned} -i \frac{1}{3} [Q^{sd}, J_\mu^8](u) &= i [Q^{sd}, J_\mu^3](u) = (\bar{s} \gamma_\mu d + \bar{d} \gamma_\mu s)(u) \\ &\equiv \tilde{K}_\mu^*(u). \end{aligned} \quad (26)$$

The rightmost term in Eq. (26) can be viewed as an interpolating operator for a $|K^{0*}\rangle + |\bar{K}^{0*}\rangle$ state.

The Ward identity appearing in Eq. (25) is a continuum equality and may not be obeyed by operators defined in lattice field theory. However, it is likely that if the continuum current J_μ^{sd} is replaced by a conserved

Möbius-fermion lattice current [33], that an analogous lattice Ward identity could be derived, with the $\tilde{K}_\mu^*$ contact term possibly becoming nonlocal. We will neglect such refinements as being of $O(a^2)$.

The use of this Ward identity leads to the second class of η -intermediate-state treatments, referred to as “ $\bar{s}d$ -subtraction” methods, which consist in modifying each weak operator Q_i in our master equations (6) and (7) according to

$$Q_i^{\text{sub}}(0) = Q_i(0) + c_{s,i}(\bar{s}d(0) + \bar{d}s(0)), \quad (27)$$

where

$$c_{s,i} \equiv -\frac{\langle \eta | Q_i(0) | K_L \rangle}{\langle \eta | \bar{s}d(0) + \bar{d}s(0) | K_L \rangle}. \quad (28)$$

In the case that the full EM currents are used, no contact terms appear in the Ward identity and we do not change the LD2 γ amplitude when replacing the operator Q_i with Q_i^{sub} which will ensure that the η intermediate state does not contribute.

For the case where we are separating the product of EM currents into $I = 1$ and $I = 0, 2$ components, Eqs. (27) and (28) are unchanged but the $c_{s,i}(\bar{s}d(0) + \bar{d}s(0))$ term added to Q_i in Eq. (27) is no longer equal to a total divergence. We can still use this strategy to remove the η intermediate state but must compute and explicitly remove the matrix element of the contact term that appears in Eq. (25). The contribution of the contact term can be expressed as a product of the usual leptonic kernel $L_{\mu\nu}$, the coefficient c_{si} and a three-point hadronic Green’s function involving the kaon interpolating operator and the product of two currents obtained from the sum of the commutators $[Q^{sd}(u_0), J_\mu^a(u)]J_\nu^b(v) + J_\mu^a(u)[Q^{sd}(v_0), J_\nu^b(v)]$, for an appropriate combination of the indices a and b . The weak Hamiltonian $H_W(x)$ does not enter this simpler Green’s function. In both cases, Eq. (28) determines the coefficients $c_{s,i}$ so that there will be no physical nor unphysical contributions from an η intermediate state.

In both classes of η -intermediate-state treatments, we need to compute matrix elements related to the η , which are very noisy because of the presence of quark-line disconnected diagrams. To address this difficulty, we perform analyzes using smaller η propagation times where the noise is less pronounced which requires that we also include the more massive η' state. Thus, we will work with two interpolating operators: an SU(3) octet for the η and an SU(3) singlet operator for the η' . Extracting the η mass from a two-point function now becomes a 2×2 generalized eigenvalue problem (GEVP) [34]. For this purpose, we consider two operators

$$O_l \equiv \frac{i}{\sqrt{2}}(\bar{u}\gamma_5 u + \bar{d}\gamma_5 d), \quad (29)$$

$$O_s \equiv i\bar{s}\gamma_5 s. \quad (30)$$

The GEVP consists in finding two pairs of eigenvalues and two-dimensional vectors (λ_n, V_n) for $n = 1, 2$ with $\lambda_1 \leq \lambda_2$ which solve the linear system

$$C(t)V_n(t, t_0) = \lambda_n(t, t_0)C(t_0)V_n(t, t_0), \quad (31)$$

where

$$C(t) = \begin{pmatrix} \langle O_l(t)O_l^\dagger(0) \rangle & \langle O_s(t)O_l^\dagger(0) \rangle \\ \langle O_l(t)O_s^\dagger(0) \rangle & \langle O_s(t)O_s^\dagger(0) \rangle \end{pmatrix}. \quad (32)$$

The solutions $V_{1,2}$ and $\lambda_{1,2}$ provide the linear combinations of operators with optimal overlap with the η and η' states and give their masses respectively.

We obtain the following values for the effective masses of the η and η' in lattice units from 519 configurations, which do not include the 110 configurations used for the four-point function calculations,

$$M_\eta = 0.569(19), \quad M_{\eta'} = 0.996(60). \quad (33)$$

These results were obtained by fitting the time dependence of the eigenvalues $\lambda(t, t-1)$ over the range $t \in [4, 8]$ for M_η and $t \in [2, 4]$ for $M_{\eta'}$. The effective-mass plot for the η and η' is given in Fig. 7. The 519 configurations used in this calculation were separated by four molecular-dynamics time units and we have tested for autocorrelations by also binning the results into bins of 1, 2 and 3 configurations. The consistency among results shown in Fig. 7 for different bin sizes suggests that the autocorrelation is small.

Note that, although the signal is lost for the η' beyond the fitted region, including the η' in the GEVP analysis helps stabilize the effective mass obtained for the η at larger time

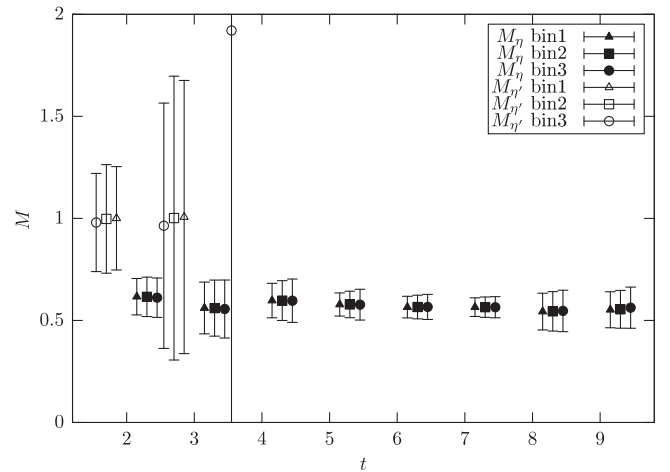


FIG. 7. The effective masses for η and η' measured by solving the GEVP with different bin sizes.

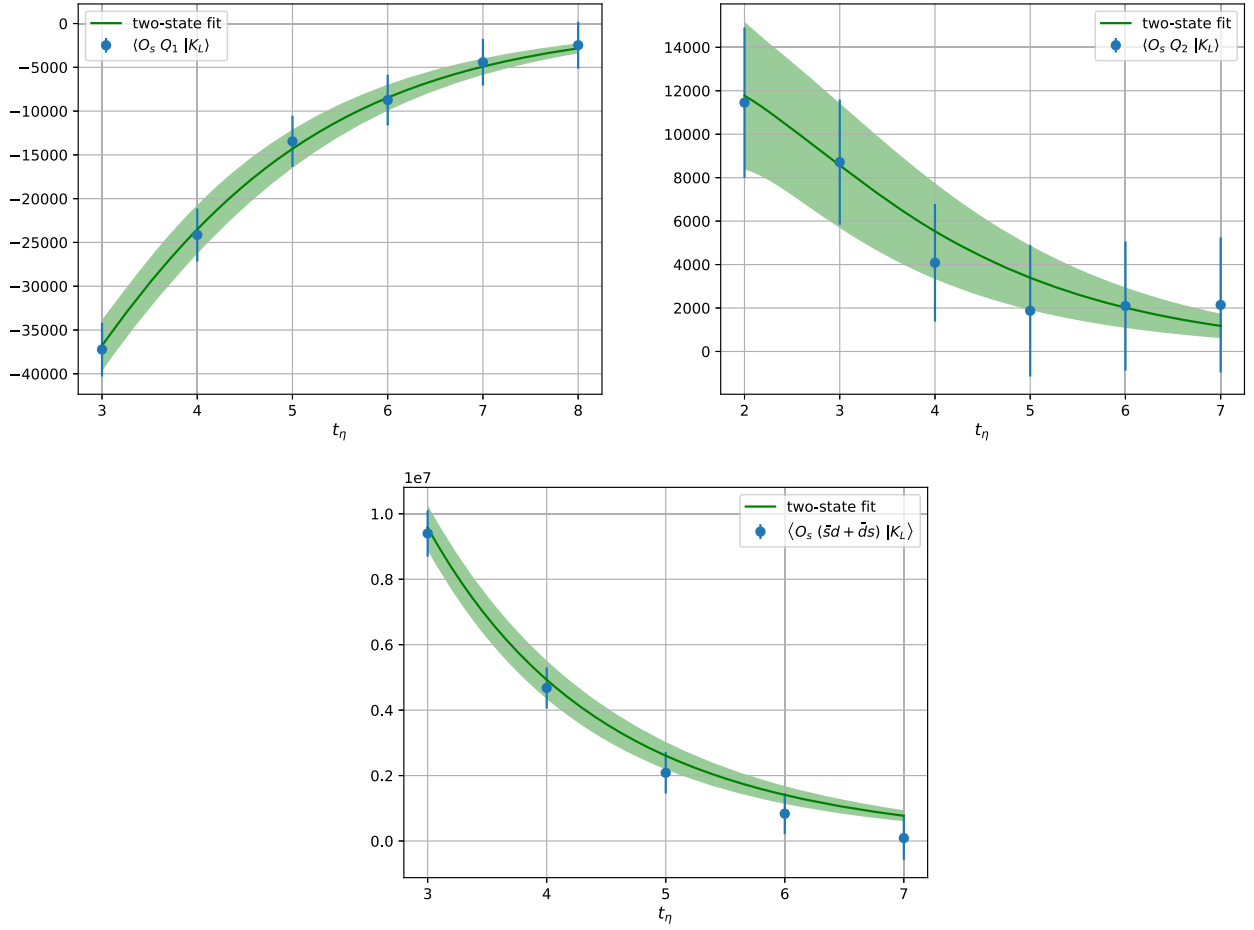


FIG. 8. Fitting $\langle O_s(t_\eta) O(0) | K_L \rangle$ with $O = Q_1, Q_2$ and $\bar{s}d + \bar{d}s$ to the ansatz given in Eq. (34) with $t_K = 8$.

separations. The masses obtained are then used to extract the matrix elements $\langle \eta | Q_{1,2} | K_L \rangle$ and $\langle \eta | (\bar{s}d + \bar{d}s) | K_L \rangle$ from the three-point correlation functions $\langle O_{l,s}(t_\eta) Q_{1,2}(0) K_L(-t_K) \rangle$ and $\langle O_{l,s}(t_\eta) (\bar{s}d(0) + \bar{d}s(0)) K_L(-t_K) \rangle$. Recognizing that it is better to take advantage of the much smaller statistical fluctuations when O_s is used alone as the interpolating operator for the η ground state, we perform a two-state fit to dependence on time separation t_η between the operator and the sink for each three-point function after projecting to the kaon ground state at a reasonably large t_K ,

$$A(t_\eta) = A_\eta e^{-M_\eta t_\eta} + A_{\eta'} e^{-M_{\eta'} t_\eta}, \quad (34)$$

to extract the desired matrix elements A_η and $A_{\eta'}$.

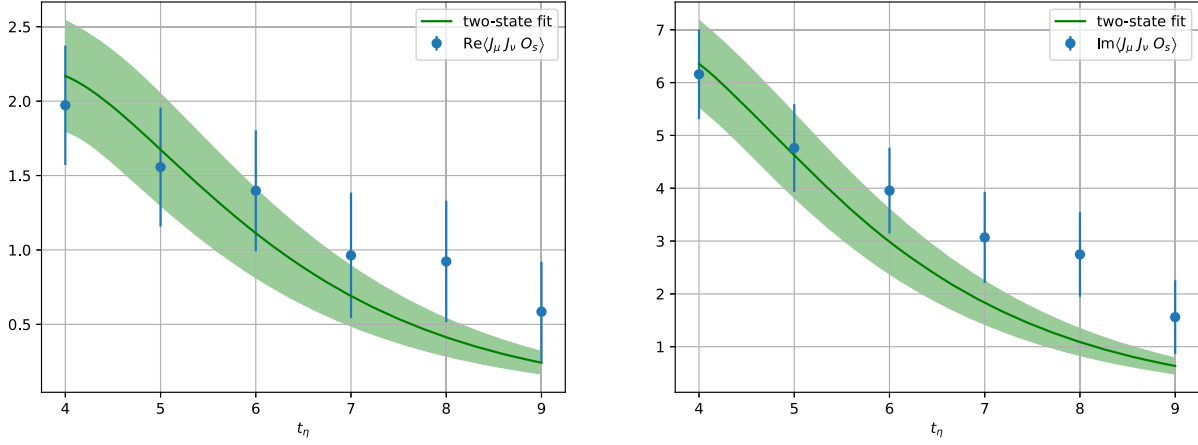
This procedure is repeated at several different values of t_K . The final quoted value for each three-point function is determined from a correlated fit to the outcomes from all these different two-state fits to a constant.⁴ The same fitting

strategy Eq. (34) is also applied to determine the real and imaginary parts of $\int d^4 r K_{\mu\nu}(r, R_{\max}) \langle 0 | J_\mu(r) J_\nu(0) | \eta \rangle$ from $\int d^4 r K_{\mu\nu}(r, R_{\max}) \langle 0 | J_\mu(r) J_\nu(0) O_s(-t_\eta) \rangle$. This matrix element is needed for the direct subtraction method 1, with the difference that one needs to consider only the separation between the η interpolating operator and the earliest EM current. The fitting qualities are illustrated in Figs. 8 and 9 and the results for the matrix elements summarized in Table I. All the three-point functions are measured on the same 110 configurations as the four-point functions so that correlations between the statistical fluctuations arising from the η intermediate state can be exploited. These results lead to the following values for the coefficients $c_{s,i}$ [Eq. (28)] needed for $\bar{s}d$ -subtraction method 2 and method 3:

$$c_{s,1} = 0.0097(32), \quad c_{s,2} = -0.0020(11). \quad (35)$$

The results from the three different treatments of the η intermediate state and the raw four-point function data before any treatment for the η are plotted in Fig. 10 for

⁴This fitting strategy differs from that used for the four-point functions where the average over t_{sep} is performed before fitting. However, this difference is not important.

FIG. 9. Fitting $W(R_{\max} = 7)$ (cf. Table I) to the ansatz in Eq. (34).

comparison. All three methods are very consistent with each other in the plateau region starting around $\delta_{\max} = 3$, but the signal begins to deteriorate rapidly when $\delta_{\max} \geq 6$. The consistency among these different treatments of the η is a strong and nontrivial check for our numerical implementation, as a sizable upward shift in the central values compared to the statistical error after the removal of the η is observed. This large observed shift also suggests that the physical contribution of the η could be important and the pion-pole dominance approximation commonly considered in the literature might not be accurate [7,35]. However, such a statement needs to be supported by a calculation which includes a continuum limit.

C. Combined results

We combine our $I = 1$ and $I = 0, 2$ results from the three different treatments for the η , including correlations, and compare them to the SDG_F^2 result and the experimental

TABLE I. Results for the η -related matrix elements needed for the various proposed η -subtraction schemes. All the values are given in lattice units and the states are normalized to unity. Here $W(R_{\max}) \equiv (m_\mu \alpha_{\text{QED}}^2)^{-1} \sum_{r \in \Lambda} K_{\mu\nu}(r, R_{\max}) \langle 0 | J_\mu(r) J_\nu(0) | \eta \rangle$ with everything in lattice units.

Quantity	Value	t_η	t_K	$\chi^2/\text{d.o.f.}$
$\langle \eta Q_1(0) K_L \rangle$	-0.0118(18)	[3, 8]	[6, 17]	0.86
$\langle \eta Q_2(0) K_L \rangle$	0.00243(89)	[2, 7]	[6, 17]	0.59
$\langle \eta \bar{s}d(0) + \bar{d}s(0) K_L \rangle$	1.22(32)	[3, 7]	[6, 19]	0.67
$\text{Re}\{W(7)\}$	0.0103(37)	[4, 9]	...	0.85
$\text{Re}\{W(10)\}$	0.0094(38)	[4, 9]	...	0.45
$\text{Re}\{W(13)\}$	0.0122(48)	[4, 9]	...	0.67
$\text{Im}\{W(7)\}$	0.0269(75)	[4, 9]	...	1.56
$\text{Im}\{W(10)\}$	0.0283(81)	[4, 9]	...	1.70
$\text{Im}\{W(13)\}$	0.0268(85)	[4, 9]	...	1.14

values in Fig. 11. Since the phase of our lattice result for $\mathcal{A}_{K_L \mu \mu}$ depends on the unphysical phase of the kaon source, we adjust that phase so that the absorptive part of $\mathcal{A}_{K_L \mu \mu}$ is purely imaginary as stated earlier, and that the imaginary part of the decay amplitude is positive. Because only the magnitudes of the experimental values are known, we arbitrarily choose their signs to agree with our lattice results.

Using the adopted phase of our kaon source we directly evaluate the matrix element of the effective Hamiltonian given in Eq. (110) in Ref. [36] in lattice calculation. The result, also shown in Fig. 11 and Table II, is given by a consistent calculation of the SDG_F^2 amplitude that can be directly combined with the $\text{LD2}\gamma$ result computed here. A more detailed discussion of our evaluation of the standard model result for the SDG_F^2 amplitude and its sign is given in Sec. IV, where we conclude that the interference between the $\text{LD2}\gamma$ and the SDG_F^2 dispersive parts is destructive, at least for this calculation in which necessary renormalization counterterms have not yet been included.

For each of the three treatments of the η shown in Fig. 11, we obtain the value by performing a correlated fit to the data for $\delta_{\max} \in [4, 6]$. With this chosen fit range, we expect that the intermediate-state contributions are well saturated, as can be seen from the statistical consistency of the results shown in Fig. 10 as $\delta_{\max}$ increases from 3 to 8. Here all three methods are very consistent with comparable statistical errors, but the values obtained for the imaginary part are about 3σ above the experimental result deduced from the $K_L \rightarrow \gamma\gamma$ process. For our final results for each isospin channel and their total, tabulated in Table II, we quote the central values and statistical errors from method 3 and take the standard deviation among the three different methods as the systematic error from the treatment of the η intermediate state. Another potential source of systematic error is the choice of the fit range for $\delta_{\max}$, which we do not include in our analysis in the present calculation as it is subdominant compared to other sources of uncertainty, such as the

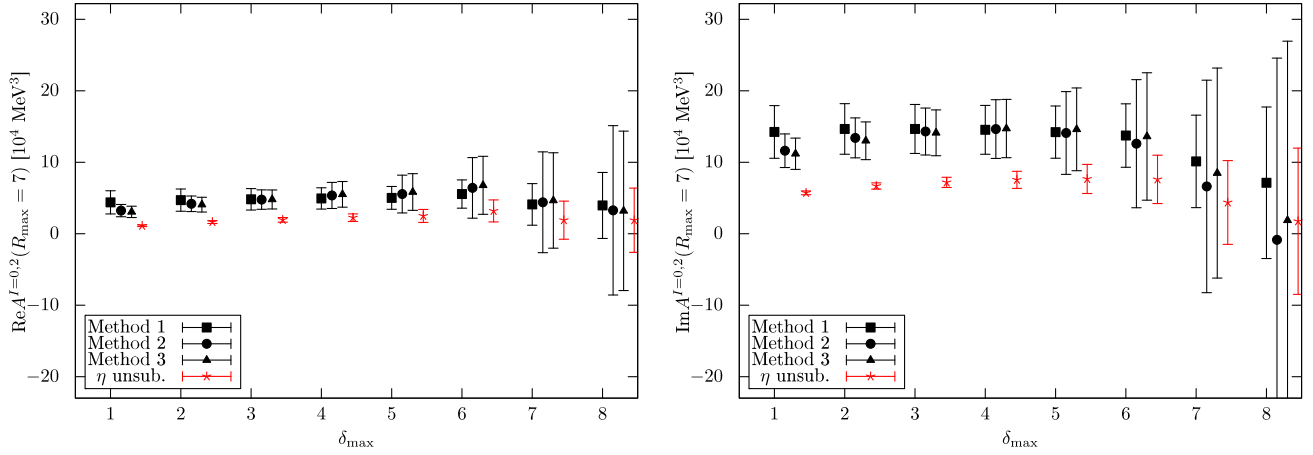
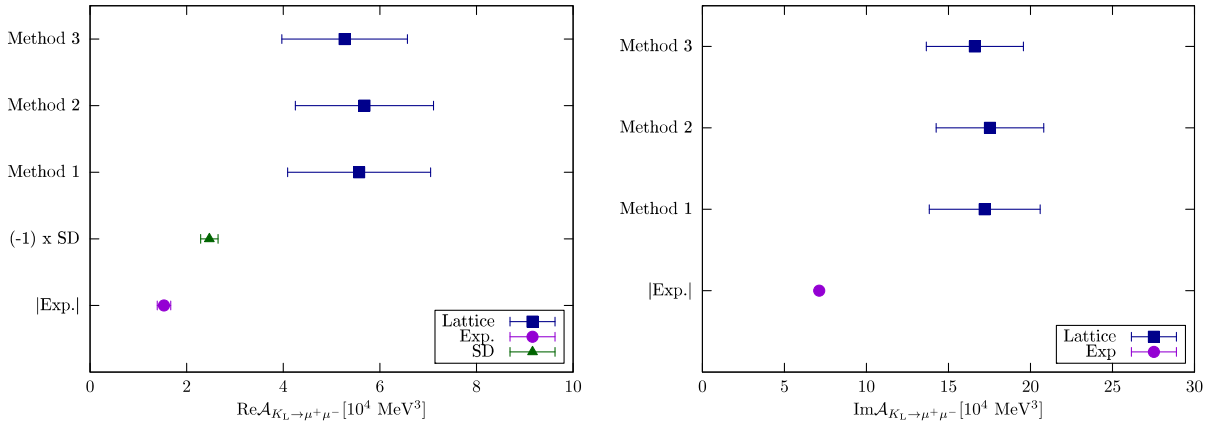
FIG. 10. The $I = 0, 2$ amplitude as a function of the temporal cutoff $\delta_{\max}$.FIG. 11. Comparison between the real and imaginary parts of the LD2 γ decay amplitude computed on the 24ID ensemble to the SDG $_F^2$ part as well as the magnitudes of the real and imaginary parts of the decay amplitude extracted from experiment. See the text for the sign conventions used here.

TABLE II. Summary table of our results on the 24ID ensemble compared to the prediction for the short-distance contribution [3] and the experimental values. Here we have adopted the phase convention that the absorptive part of the LD2 γ amplitude comes only from the imaginary part of the decay amplitude and that this imaginary part is positive, which determines our choice for the phase of the kaon interpolating operator (see Sec. IV for details). The values for $I = 0, 2$ and the total are quoted from method 3, where the rightmost errors are systematic, estimated from the standard deviation among the results from the three different treatments of the η . Of course, a physically meaningful comparison between the experimental and theoretical results in this table must await the inclusion of the omitted renormalization counterterms in the lattice result.

	$\text{Re} \mathcal{A} \times 10^{-4} \text{ (MeV}^3\text{)}$	$\text{Im} \mathcal{A} \times 10^{-4} \text{ (MeV}^3\text{)}$
$I = 1$	0.69(9)	2.64(21)
$I = 0, 2$	5.06(1.41)(0.09)	14.31(3.11)(0.22)
Total (LD2 γ)	5.27(1.30)(0.17)	16.61(2.96)(0.38)
SDG $_F^2$	-2.47(18)	...
Expt.	$\pm 1.53(14)$	$\pm 7.10(3)$

intrinsic statistical noise from the η intermediate state given the current statistics. For completeness, we also list the values for the contributions from Q_1 and Q_2 to the total LD2 γ in Table III. We can notice a large cancellation between the two contributions.

Finally, it is interesting to remark that the statistical error of the data from method 1 of the η treatment started out much larger but increases more slowly than those from method 2 and method 3 in Fig. 10; however, under a correlated fit over different $\delta_{\max}$, the overall statistical errors for all three methods are comparable (see Fig. 11). In fact, as the physical contribution of η is explicitly reconstructed, the error on the three-point functions needed for the subtraction—which remains identical as we increase $\delta_{\max}$ —dominates at small $\delta_{\max}$. As $\delta_{\max}$ increases, the uncertainty from the four-point function eventually becomes larger than the former. This competition in statistical error and the correlation between data points are different in the two other cases. As the correlated-fit results in Fig. 11 show, all three methods give uncertainties

TABLE III. Individual contributions to the total amplitude in Table II with their Wilson coefficients included, estimated from method 3 with $R_{\max} = 7$ and values taken at $\delta_{\max} = 4$.

	Q_1	Q_2
$\text{Re } \mathcal{A} \times 10^{-4} \text{ (MeV}^3\text{)}$	13.18(90)	-7.57(87)
$\text{Im } \mathcal{A} \times 10^{-4} \text{ (MeV}^3\text{)}$	36.82(2.06)	-20.59(1.97)

with similar sizes. With the present study, we are unable to conclude which of the three methods is the most advantageous.

Given the exploratory nature of this first lattice calculation, a detailed discussion of systematic errors would be premature. However, there are at least four likely sources of systematic error. The 24ID ensemble used here with an inverse lattice spacing of $1/a = 1.023 \text{ GeV}$ leads to uncertain but possibly large discretization errors. For a similar calculation of the electromagnetic difference between the π^+ and π^0 masses [37] the result obtained from the 24ID ensemble was 20% below the experimental value. The finite-volume errors associated with low-energy $\pi\pi\gamma$ states was estimated to be at or below 10% [17] and errors as large as 10% might result from our omission of all but two of the seven weak operators present in the $\Delta S = 1$ effective weak Hamiltonian. Finally, the 30% statistical error present in our treatment of the η intermediate state could well obscure systematic errors of the same size, in spite of the agreement between the three different methods we tried.

IV. INTERFERENCE BETWEEN THE LD2 γ AND SDG $_F^2$ AMPLITUDES

We begin this section by discussing our lattice QCD evaluation of the SDG $_F^2$ amplitude based on the earlier standard model prediction presented in Refs. [3,23,36], including both its magnitude and sign, using the same conventions adopted when computing the LD2 γ amplitude which is the subject of this paper. The second part of this section examines a series of issues that require special attention when computing the sign of LD2 γ amplitude.

A. Lattice QCD evaluation of the SDG $_F^2$ amplitude

In order to determine the relative sign between the SD part and the LD part, we follow closely the conventions of Buchalla and Buras [36]. Their effective Hamiltonian describing the $O(G_F^2)$ contribution to $K_L \rightarrow \mu^+\mu^-$ provides an accessible summary of their calculation which we can easily exploit to determine the sign of this SDG $_F^2$ amplitude,

$$\mathcal{H}_W^{\text{SD}} = -\frac{G_F}{\sqrt{2}} \frac{\alpha_{\text{em}}}{2\pi \sin^2 \theta_W} \left[(V_{cs}^* V_{cd} Y_{\text{NL}} + V_{ts}^* V_{td} Y(x_t)) \times (\bar{s}d)_{V-A} (\bar{\mu}\mu)_{V-A} + \text{H.c.} \right]. \quad (36)$$

The quantities Y_{NL} and $Y(x_t)$ are known functions which have been computed with the next-to-next-to-leading order charm-quark effects included in Ref. [3]. Here $\alpha_{\text{em}} = \alpha_{\text{em}}(M_Z)$ and $\sin^2 \theta_W$ are evaluated at the electroweak scale [38] while x_t is the ratio of the squares of the masses of the top quark and W boson, m_t^2/M_W^2 .

We can now evaluate the matrix element of $\mathcal{H}_W^{\text{SD}}$ given in Eq. (36) between the same perturbative $\mu^+\mu^-$ and non-perturbative $|K_L\rangle$ states as were used to compute the $\mathcal{A}_{K_L\mu\mu}$ in Sec. III, giving the relative sign of these two amplitudes physical meaning.

While experimentally inaccessible, in the conventional theoretical framework of our calculation we can also determine the relative sign between these two $K_L \rightarrow \mu^+\mu^-$ amplitudes and f_K which appears in the QCD matrix element $\langle 0 | [(\bar{s}d)_{V-A} + \text{H.c.}] | K_L \rangle$ that enters as a factor in the matrix element of $\mathcal{H}_W^{\text{SD}}$. For this relative sign with respect to f_K to be meaningful, we must describe explicitly the perturbative $\mu^+\mu^-$ state used in our calculation of $\mathcal{A}_{K_L\mu\mu}$.

Our $\mu^+\mu^-$ state can be specified as follows. For the perturbative QED sector, we choose to work in the Weyl basis for the γ matrices and use the representation for the spinor fields with momentum parallel to the z direction,

$$u^\pm(p) = \begin{pmatrix} \sqrt{E - p_z} \phi_\pm \\ \sqrt{E + p_z} \phi_\pm \end{pmatrix}, \quad v^\pm(p) = \begin{pmatrix} \sqrt{E - p_z} \chi_\pm \\ -\sqrt{E + p_z} \chi_\pm \end{pmatrix}, \quad (37)$$

where

$$\phi^+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \phi^- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \chi^+ = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \chi^- = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (38)$$

With this choice, the SDG $_F^2$ matrix element becomes

$$\mathcal{A}_{\text{SDG}_F^2} = \langle \mu^+(\mathbf{k}^+) \mu^-(\mathbf{k}^-) | \mathcal{H}_W^{\text{SD}}(0) | K_L \rangle \quad (39)$$

$$= 2m_\mu \frac{G_F}{\sqrt{2}} \frac{\alpha_{\text{em}}(M_Z)}{2\pi \sin^2 \theta_W} \text{Re}(V_{cs}^* V_{cd} Y_{\text{NL}} + V_{ts}^* V_{td} Y(x_t)) \times \langle 0 | [\bar{s}(0) \gamma^0 \gamma^5 d(0) + \text{H.c.}] | K_L \rangle, \quad (40)$$

where the μ^+ destruction operator appears to the left of that for the μ^- in the matrix elements defined in Eqs. (4) and (39). Here the rightmost (renormalized) matrix element is to be computed in a QCD-only background. This matrix element is proportional to the decay constant of the kaon, which is necessarily computed, including its sign in our

conventions, when Eq. (40) is evaluated. We find

$$\langle 0 | [\bar{s}(0) \gamma^0 \gamma^5 d(0) + \text{H.c.}] | K_L \rangle = -1.173(8) \times 10^5 \text{ MeV}^2 \quad (41)$$

$$= \sqrt{2} M_K f_K. \quad (42)$$

Here we have defined the phase of f_K [2] as that obtained from a matrix element with a kaon state whose phase is chosen to make the imaginary part of the amplitude $\mathcal{A}_{\text{LD2}\gamma}$ positive.

B. Using lattice QCD to compute analytically continued Minkowski-space Green's functions

Consider a general Minkowski-space QCD Green's function $G(x^{[N]})$ that depends on N spacetime positions $x^{[N]} \equiv \{x_j\}_{1 \leq j \leq N}$. If sums over intermediate QCD energy eigenstates are inserted between the N operators appearing in that Green's function then the individual terms in the sums over intermediate states depend exponentially on the time components of those positions. As a result the entire Green's function is an analytic function of the times $\{(x_j)^0\}_{1 \leq j \leq N}$ if small negative imaginary parts are added to those N times to ensure that the sums over intermediate states converge.

If each of these N time variables is analytically continued from their original real values to negative imaginary values $(x_j)^0 \rightarrow \exp\{-i\phi\} x_j^0$ with $\phi: 0 \rightarrow \frac{\pi}{2}$ and $(x_j)^0 = \exp\{-i\frac{\pi}{2}\} (x_{E,j})^4$ then we can define the result as the Euclidean Green's function $G_E(x_E^{[N]})$ depending on the N Euclidean four-vectors $x_E^{[N]} \equiv \{x_{E,j}\}_{1 \leq j \leq N}$ with $x_{E,j} = (\vec{x}_{E,j}, (x_{E,j})^4) = (\vec{x}_j, i(x_j)^0)$. Of course, $G_E(x_E^{[N]})$ could have been obtained from a Schrödinger-picture theory by constructing time-dependent operators $\mathcal{O}_E(x_E^{[N]})$ using a Heisenberg representation in which $\exp\{-iH_{\text{QCD}}x^0\}$ is replaced by $\exp\{-H_{\text{QCD}}x^4\}$. If that Schrödinger-picture theory is relativistically invariant then the $G(x^{[N]})$ will be Lorentz covariant while $G_E(x_E^{[N]})$ will be $O(4)$ covariant.

In the calculation reported here such an analytic continuation has been performed to transform a Minkowski-space Green's function obeying our Minkowski-space conventions into the Euclidean-space Green's functions appearing in Eqs. (6), (7) and (40). We must be careful that the conventions used in our Euclidean-space lattice calculation are those that would result from the analytic continuation that has been performed.

The Green's functions that must be reproduced in Euclidean space involve four operators, the electromagnetic current $J^\mu(x)$, the effective weak Hamiltonian density $\mathcal{H}_W(x)$, the $\Delta S = \pm 1$ axial current $(J_{sd}^5)^\mu(x) \equiv \bar{s}(x) \gamma^\mu \gamma^5 d(x) + \text{H.c.}$

and the kaon interpolating $K_L(t)$. The operator $K_L(t)$ appears linearly in all of the Green's functions of interest and does not carry a physically meaningful phase so we need not be concerned that the operator $K_L(t)$ may not even have a familiar Minkowski-space expression. The Minkowski- and Euclidean-space representations for the operators $J^\mu(x)$, $\mathcal{H}_W(x)$ and $(J_{sd}^5)^\mu(x)$ can be easily compared knowing how the Dirac matrix conventions are related.

Recall that the Minkowski-space Dirac operator for a particle of mass m in our conventions can be reexpressed after the analytic continuation to Euclidean-space as follows:

$$D = i\gamma^\mu \frac{\partial}{\partial x^\mu} - m \quad (43)$$

$$= i \left[\gamma^0 \frac{\partial}{\partial(-ix^4)} + \sum_{j=1}^3 \gamma^j \frac{\partial}{\partial x^j} \right] - m \quad (44)$$

$$= - \left[\gamma^0 \frac{\partial}{\partial x^4} - i \sum_{j=1}^3 \gamma^j \frac{\partial}{\partial x^j} \right] - m, \quad (45)$$

which takes the form used in Euclidean space,

$$D_E = \sum_{i=\mu}^4 \gamma_E^\mu \frac{\partial}{\partial x_E^\mu} + m \quad (46)$$

if we define

$$\gamma_E^4 = \gamma^0, \quad \gamma_E^j = -i\gamma^j. \quad (47)$$

Since only the three spatial γ matrices are changed when writing the analytically continued Green's function in Euclidean notation, both γ^0 and γ^5 need not be altered. Adopting this prescription for the γ matrices used in our lattice calculation, we can reproduce the desired Minkowski-space matrix elements from the following formulas:

$$J^0 = J_E^4, \quad J^j = iJ_E^j, \quad (48)$$

$$(J_{sd}^5)^0 = (J_{sd,E}^5)^4, \quad (J_{sd}^5)^j = i(J_{sd,E}^5)^j, \quad (49)$$

$$\begin{aligned} & \bar{\psi}_a \gamma^\mu (1 - \gamma^5) \psi_b \bar{\psi}_c \gamma_\mu (1 - \gamma^5) \psi_d \\ &= (\bar{\psi}_E)_a \gamma_E^\mu (1 - \gamma_E^5) (\psi_E)_b (\bar{\psi}_E)_c \gamma_E^\mu (1 - \gamma_E^5) (\psi_E)_d. \end{aligned} \quad (50)$$

Here the ψ_i represent different spinor fields. The Euclidean and Minkowski Grassmann spinor variables are unchanged

except for their spacetime labels—no rotation of spinor indices is needed. The validity of these statements can be readily checked by analytically continuing some simple position-space Green’s functions in the time.

The Euclidean-space γ matrix conventions derived above are very close to those used in lattice QCD, for example those found in Refs. [39,40]. The conventions in those two references are mathematically similar to those in the GRID [41,42] lattice code which we use. The GRID convention is shared with the Columbia Physics System [43], with CHROMA [44], and is as documented in the USQCD QDP++ documentation on spin conventions [45]. However, in all these cases the γ_E^5 matrix that is introduced is the negative of the γ_E^5 matrix that results from the analytic continuation in the time that we are making. Thus, when using the result for the matrix element of the axial current $(J_{sd}^5)^0$ we must introduce an extra minus sign to obtain the matrix element of interest, in apparent conflict with the projection operators defined in Eq. (8.11) of Ref. [39]. The only other sign that must be recognized when interpreting our lattice matrix element as providing Minkowski-space results is the product $J^i J^j$ which appears in Eqs. (6) and (7) where a minus sign must be introduced, since our lattice calculation contains the factors $(\gamma_E^i)(\gamma_E^j) = -(\gamma^i)(\gamma^j)$ and the factors $+(\gamma^i)(\gamma^j)$ are what should occur in our Minkowski-space calculation.

V. CONCLUSION

The formalism developed in Ref. [17] allows for a first-principles determination of the long-distance two-photon contribution to the $K_L \rightarrow \mu^+ \mu^-$ decay amplitude from lattice QCD. Although our method does not accurately treat the contribution from low-energy $\pi\pi\gamma$ states, in Ref. [17] we show that these effects are at or below the 10% level. In addition to the possibility for the systematic improvement of lattice QCD results, another appealing aspect of this *ab initio* treatment of QCD is the ability to determine the relative sign of the two-photon-exchange amplitude and the amplitude arising from the one-loop short-distance electroweak contribution.

In this paper we present our numerical implementation of the formalism of Ref. [17] and the first result using a $24^3 \times 64$ Möbius domain wall fermion ensemble at physical pion mass and an inverse lattice spacing of $1/a = 1.023$ GeV. We have included all Feynman diagrams with the exception of SU(3)-flavor-suppressed tadpole terms in which an electromagnetic current attaches to a closed quark loop. However, we should emphasize that, while extensive, the calculation presented in the paper is not complete and cannot yet be directly compared with experiment. Additional counterterms needed to represent the effects of the GIM mechanism have not been included. This second part of the lattice QCD calculation of the

two-photon contribution to $K_L \rightarrow \mu^+ \mu^-$ will be presented in a second paper.

In addition to determining the needed renormalization counterterms, the natural next target for a more accurate result is the $48^3 \times 96$ physical-pion-mass ensemble with $1/a = 1.73$ GeV (“48I”) from the RBC/UKQCD Collaboration [46], a calculation that is now also underway. Since this ensemble is significantly larger than the current 24ID ensemble, the reconstruction of the η contribution, which is at present the largest source of statistical error, should be improved by the average over a larger volume and, because of the smaller lattice spacing, the increased number of data points in the region where the η signal can be seen.

ACKNOWLEDGMENTS

We thank our RBC and UKQCD Collaboration colleagues for discussion and ideas, in particular Chulwoo Jung for technical support. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility using NERSC Awards No. HEP-ERCAP0023253, No. 0026351, and No. 0031326. P.B. was supported in part by US DOE Contract No. DESC0012704(BNL) and the Scientific Discovery through Advanced Computing (SciDAC) program LAB 22-2580. E.-H.C. and N.C. were supported in part by the U.S. Department of Energy (DOE) Grant No. DE-SC0011941. L.J. acknowledges support by DOE Office of Science Early Career Award No. DE-SC0021147 and DOE Award No. DE-SC0010339.

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.

APPENDIX A: COMPUTATION STRATEGIES

We give the implementation details of the Wick contractions to get $\mathcal{A}^{\text{lat}}$ [Eq. (6)] considered in this work. For the ones needed for $\mathcal{A}^{\text{sub}}$ [Eq. (7)] and the η -removal described in Sec. III B, we simply follow the strategies reported in Ref. [47]. We denote the whole set of lattice sites by Λ and the hyperplane corresponding to the time slice t by $\Lambda_0(t)$. The coordinate-space variables u and v indicate the locations of the EM currents (with time order $v_0 \leq u_0$), x that of the weak Hamiltonian and t_K that of the kaon wall, in accordance with Sec. II B.

While computing the amplitude, Eq. (6), translational invariance allows us to set the position of one of the three local operators to be the reference point and sum over the

two other positions. The reference point corresponds to the source location of point-source propagators needed for the contraction diagram to be evaluated. Calculations will be repeated for many different reference point to improve

statistics. For each contraction class, we evaluate the associated “reduced amplitudes” defined as follows on each gauge configuration. For type-1 and type-2 diagrams, we find it convenient to set the reference point to x ,

$$\bar{\mathcal{A}}_i^{\mathcal{D}}(\delta_{\max}, R_{\max}, t_{\text{sep}}) = \sum_{\substack{u, v \in \Lambda, \\ d \leq \delta_{\max}, u_0 - v_0 \leq R_{\max}}} \delta_{v_0 - x_0, d} e^{M_K(v_0 - t_K)} K_{\mu\nu}(u - v, R_{\max}) \mathcal{C}_{i,\mu\nu}^{\mathcal{D}}(u, v, x, t_K), \quad (\text{A1})$$

and to v for type-3 and type-4 diagrams and type-5 diagram 2 (T5D2),

$$\bar{\mathcal{A}}_i^{\mathcal{D}}(\delta_{\max}, R_{\max}, t_{\text{sep}}) = \sum_{\substack{u, x \in \Lambda, \\ d \leq \delta_{\max}, u_0 - v_0 \leq R_{\max}}} \delta_{v_0 - x_0, d} e^{M_K(v_0 - t_K)} K_{\mu\nu}(u - v, R_{\max}) \mathcal{C}_{i,\mu\nu}^{\mathcal{D}}(u, v, x, t_K). \quad (\text{A2})$$

The case of T5D1 is a bit different as we use a low-mode-improved version of Eq. (A2), which will be discussed in detail in Sec. A 5. In the above, the subscript i is the index for the corresponding weak operator and the superscript $\mathcal{D}$ indicates the associated diagram. The contraction functions $\mathcal{C}^{\mathcal{D}}$ will be given in each dedicated subsection of this appendix. The isospin components as defined in Sec. II C can be written as linear combinations of the reduced amplitudes, Eqs. (A1) and (A2),

$$\mathcal{A}_i^{\text{lat}}(\delta_{\max}, R_{\max}, t_{\text{sep}}) = \bar{\mathcal{N}} \sum_{\substack{i=1,2 \\ \mathcal{D} \in \text{diagrams}}} b_i^{\text{I}}(\mathcal{D}) C_i \langle \bar{\mathcal{A}}_i^{\mathcal{D}}(\delta_{\max}, R_{\max}, t_{\text{sep}}) \rangle_U, \quad (\text{A3})$$

where the charge factors $b_i^{\mathcal{D}}$ are given in Table IV, the C_i 's are the Wilson coefficients, $\bar{\mathcal{N}}$ is the normalization factor to get the kaon ground state in the relativistic normalization and $\langle \cdot \rangle_U$ is the gauge-ensemble average.

The $R_{\max}$ -regulated kernel $K_{\mu\nu}(u - v, R_{\max})$ is defined in Eq. (8). We will keep the $R_{\max}$ dependence implicit in the remainder of this appendix. The variable t_{sep} is defined as the distance between the weak operator and the kaon wall,

$$t_{\text{sep}} \equiv x_0 - t_K. \quad (\text{A4})$$

In our calculations, we aim at computing each diagram at several values of $(t_{\text{sep}}, R_{\max})$. For each $R_{\max}$, our computer program constructs a different kernel with points with $u_0 - v_0 > R_{\max}$ set to zero.

In the rest of this appendix, we will use the notations

$$\Gamma_{\mu}^5 \equiv \gamma_5 \gamma_{\mu}, \quad \Gamma_{\mu}^L \equiv \gamma_{\mu} (1 - \gamma_5). \quad (\text{A5})$$

The Lorentz indices μ and ν are associated with the currents at u and v respectively. A light- (respectively, strange-) quark point-source propagator with source at y and sink at x is denoted by $L(x, y)$ [respectively, $H(x, y)$]. $L(x, t_K)$ [respectively, $H(x, t_K)$] denotes a light- (respectively, strange-) quark Coulomb-gauge-fixed wall-source propagator which carries zero spatial momenta and has its source located at time slice t_K . Finally, for the naming convention, a diagram “a” is related to its counterpart “b” by swapping the locations of the two currents, i.e., $(u, \mu) \leftrightarrow (v, \nu)$. For simplicity, we will make the coordinate-space arguments in some of the expressions implicit, while keeping them consistent with Eqs. (A1) and (A2).

1. Type 1

For the type-1 topology, there are four diagrams: T1D1a, T1D1b, T1D2a, T1D2b, as shown in Fig. 12. The contraction functions entering Eq. (A1) are

TABLE IV. The charge factors defined in Eq. (A3).

$b_i^{I=1}$	T1D1	T1D2	T2D1	T3D1	T4D1	T4D2	T5D1	T5D2
Q_1	...	-1/6	1/6	1/6	-1/6	...	...	...
Q_2	...	-1/6	-1/6	-1/6	1/6	...	...	...
$b_i^{I=0,2}$	T1D1	T1D2	T2D1	T3D1	T4D1	T4D2	T5D1	T5D2
Q_1	-2/9	-1/18	-1/18	5/18,	5/18	1/9	-5/9	-1/9
Q_2	-2/9	-1/18	1/18	-5/18	-5/18	-1/9	5/9	1/9

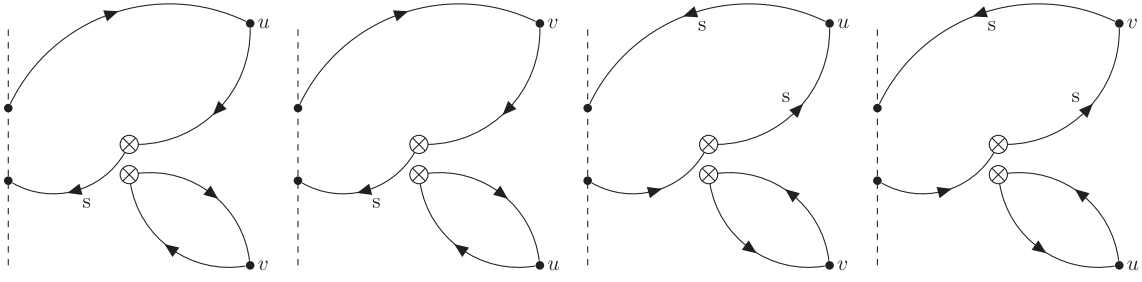


FIG. 12. Type-1 diagrams. From left to right: T1D1a, T1D1b, T1D2a and T1D2b.

$$\mathcal{C}_{1,\mu\nu}^{\text{T1D1a}} = \text{Tr}[\Gamma_\rho^L L(x, v) \gamma_\nu L(v, x)] \text{Tr}[\Gamma_\rho^L \{L(x, u) \gamma_\mu L(u, t_K) \gamma_5 H(t_K, x) + \text{H.c.}\}], \quad (\text{A6})$$

$$\mathcal{C}_{1,\mu\nu}^{\text{T1D2a}} = \text{Tr}[\Gamma_\rho^L L(x, v) \gamma_\nu L(v, x)] \text{Tr}[\Gamma_\rho^L \{L(x, t_K) \gamma_5 H(t_K, u) \gamma_\mu H(u, x) + \text{H.c.}\}], \quad (\text{A7})$$

$$\mathcal{C}_{2,\mu\nu}^{\text{T1D1a}} = \text{Tr}_C \left[\text{Tr}_S[\Gamma_\rho^L L(x, v) \gamma_\nu L(v, x)] \text{Tr}_S[\Gamma_\rho^L \{L(x, u) \gamma_\mu L(u, t_K) \gamma_5 H(t_K, x) + \text{H.c.}\}] \right], \quad (\text{A8})$$

$$\mathcal{C}_{2,\mu\nu}^{\text{T1D2a}} = \text{Tr}_C \left[\text{Tr}_S[\Gamma_\rho^L L(x, v) \gamma_\nu L(v, x)] \text{Tr}_S[\Gamma_\rho^L \{L(x, t_K) \gamma_5 H(t_K, u) \gamma_\mu H(u, x) + \text{H.c.}\}] \right], \quad (\text{A9})$$

plus those for the diagram b's, obtained with $(u, \mu) \leftrightarrow (v, \nu)$. In the above, Tr_C (respectively, Tr_S) indicates tracing with regard to the color (respectively, spin) indices, whereas Tr is the trace over both spin and color indices.

The advantage of adopting the parametrization (A1) should become clear at this point: by taking the reference point at x , we only need a pair of wall-source propagators and a pair of point-source propagators for both the light and strange quarks per reference point. To get the reduced amplitude (A1), we convolve the four-point functions with the leptonic kernel. With this observation, the sum over u for all $v \in \Lambda$ can be done efficiently via fast Fourier transform. We rewrite the four-point functions in blocks to compute them for different $(t_{\text{sep}}, R_{\text{max}})$ with reduced overhead. We define the “bubble” blocks as

$$G_{\nu\rho}(v, x) \equiv \text{Tr}_S[\Gamma_\rho^L A_\nu], \quad A_\nu \equiv L^\dagger(v, x) \Gamma_\nu^5 L(v, x), \quad (\text{A10})$$

and the “triangle” blocks as

$$F_{\mu\rho}^1(u, x, t_{\text{sep}}) = (-1) \times \text{Tr}_S \left[\Gamma_\rho^L (B_1(u) + B_1^\dagger(u)) \right], \\ B_1 \equiv L^\dagger(u, x) \Gamma_\mu^5 L(u, t_K) H^\dagger(x, t_K), \quad (\text{A11})$$

$$F_{\mu\rho}^2(u, x, t_{\text{sep}}) = \text{Tr}_S \left[\Gamma_\rho^L (B_2(u) + B_2^\dagger(u)) \right], \\ B_2(u) \equiv L(x, t_K) H^\dagger(u, t_K) \Gamma_\mu^5 H(u, x). \quad (\text{A12})$$

Our program loops over the kernels first. In each iteration, we need the kernel-dependent quantities, obtained by convolving the above quantities with different regulated kernels

$$\hat{G}_{\nu\rho}(v, x) \equiv (L_{\mu\nu} * G_{\mu\rho})(v, x), \quad (\text{A13})$$

$$\hat{F}_{\nu\rho}(v, x, t_{\text{sep}}) \equiv (L_{\mu\nu} * F_{\mu\rho})(v, x, t_{\text{sep}}), \quad (\text{A14})$$

where the convolution $*$ is defined via

$$(K_{\mu\nu} * F_{\mu\rho})(v) \equiv \sum_{u \in \Lambda} K_{\mu\nu}(u - v) F_{\mu\rho}(u) \\ = \mathcal{F}^{-1}[\mathcal{F}[K_{\mu\nu}](-p) \mathcal{F}[F_{\mu\rho}](p)], \quad (\text{A15})$$

with $\mathcal{F}$ denoting a Fourier transform. Noting that the bubble blocks do not depend on t_{sep} , we can reuse the bubble blocks while looping over t_{sep} 's. Taking the color trace properly, we get the reduced amplitudes

$$\bar{\mathcal{A}}_1^{\text{T1D1a}} = (-1) \times \sum_{v \in \Lambda | d \leq \delta_{\text{max}}} \delta_{v_0 - x_0, d} e^{M_K(v_0 - t_K)} \text{Tr}_C \left[\hat{F}_{\nu\rho}^1(v, x, t_{\text{sep}}) \right] \text{Tr}_C[G_{\nu\rho}(v, x)], \quad (\text{A16})$$

$$\bar{\mathcal{A}}_1^{\text{T1D1b}} = (-1) \times \sum_{v \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(v_0-t_K)} \text{Tr}_C[F_{\nu\rho}^1(v, x, t_{\text{sep}})] \text{Tr}_C[\hat{G}_{\nu\rho}(v, x)], \quad (\text{A17})$$

$$\bar{\mathcal{A}}_2^{\text{T1D1a}} = (-1) \times \sum_{v \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(v_0-t_K)} \text{Tr}_C[\hat{F}_{\nu\rho}^1(v, x, t_{\text{sep}}) G_{\nu\rho}(v, x)], \quad (\text{A18})$$

$$\bar{\mathcal{A}}_2^{\text{T1D1b}} = (-1) \times \sum_{v \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(v_0-t_K)} \text{Tr}_C[F_{\nu\rho}^1(v, x, t_{\text{sep}}) \hat{G}_{\nu\rho}(v, x)], \quad (\text{A19})$$

and the D2 diagrams are obtained by substituting F^1 with F^2 .

2. Type 2

The diagrams for this contraction class are T2D1a and T2D1b as shown in Fig. 13. The four-point functions for this contraction class are

$$\mathcal{C}_{1,\mu\nu}^{\text{T2D1a}} = \text{Tr}[\Gamma_\rho^L L(x, x)] \text{Tr}[\Gamma_\rho^L \{L(x, u) \gamma_\mu L(u, t_K) \gamma_5 H(t_K, v) \gamma_\nu H(v, x) - \text{H.c.}\}], \quad (\text{A20})$$

$$\mathcal{C}_{2,\mu\nu}^{\text{T2D1a}} = \text{Tr} \left[\Gamma_\rho^L L(x, x) \Gamma_\rho^L \{L(x, u) \gamma_\mu L(u, t_K) \gamma_5 H(t_K, v) \gamma_\nu H(v, x) - \text{H.c.}\} \right], \quad (\text{A21})$$

plus the b diagrams obtained with $(u, \mu) \leftrightarrow (v, \nu)$. The parametrization equation (A1) is chosen for the same reason as in the type-1 case.

Similarly, we can write the quantities to compute in terms of sub-blocks to reduce computation overhead. We define

$$F_\mu(u, x, t_{\text{sep}}) \equiv \gamma_5 L^\dagger(u, x) \Gamma_\mu^5 L(u, t_K), \quad (\text{A22})$$

$$G_\nu(v, x, t_{\text{sep}}) \equiv H^\dagger(v, t_K) \Gamma_\nu^5 H(v, x), \quad (\text{A23})$$

and their counterparts convolved with the kernel

$$\hat{F}_\mu(v, x, t_{\text{sep}}) \equiv (K_{\mu\nu} * F_\mu)(v, x, t_{\text{sep}}), \quad (\text{A24})$$

$$\hat{G}_\nu(v, x, t_{\text{sep}}) \equiv (K_{\mu\nu} * G_\mu)(v, x, t_{\text{sep}}). \quad (\text{A25})$$

With

$$T_\rho^1(x) \equiv \Gamma_\rho^L L(x, x), \quad (\text{A26})$$

$$T_\rho^2[B, C](z, x, t_{\text{sep}}) \equiv \Gamma_\rho^L [B_\nu(z, x, t_{\text{sep}}) C_\nu(z, x, t_{\text{sep}}) - \text{H.c.}], \quad (\text{A27})$$

we then have

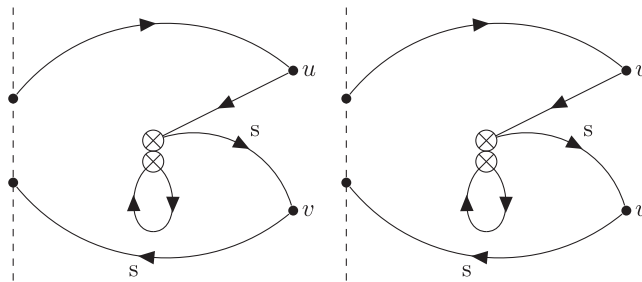


FIG. 13. Type-2 diagrams. From left to right: T2D1a and T2D1b.

$$\bar{\mathcal{A}}_1^{\text{T2D1a}} = \sum_{v \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(v_0-t_K)} \text{Tr}[T_\rho^1(x)] \text{Tr}[T_\rho^2(\hat{F}, G; v, x, t_{\text{sep}})], \quad (\text{A28})$$

$$\bar{\mathcal{A}}_1^{\text{T2D1b}} = \sum_{v \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(v_0-t_K)} \text{Tr}[T_\rho^1(x)] \text{Tr}[T_\rho^2(F, \hat{G}; v, x, t_{\text{sep}})], \quad (\text{A29})$$

$$\bar{\mathcal{A}}_2^{\text{T2D1a}} = \sum_{v \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(v_0-t_K)} \text{Tr}[T_\rho^1(x) T_\rho^2(\hat{F}, G; v, x, t_{\text{sep}})], \quad (\text{A30})$$

$$\bar{\mathcal{A}}_2^{\text{T2D1b}} = \sum_{v \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(v_0-t_K)} \text{Tr}[T_\rho^1(x) T_\rho^2(F, \hat{G}; v, x, t_{\text{sep}})]. \quad (\text{A31})$$

3. Type 3

The type-3 diagrams are T3D1a and T3D1b, depicted in Fig. 14. The corresponding four-point functions are

$$\mathcal{C}_{1,\mu\nu}^{\text{T3D1a}} = \text{Tr}[\Gamma_\rho^L L(x, u) \gamma_\mu L(u, v) \gamma_\nu L(v, x)] \text{Tr}[\Gamma_\rho^L \{L(x, t_K) \gamma_5 H(t_K, x) - \text{H.c.}\}], \quad (\text{A32})$$

$$\mathcal{C}_{2,\mu\nu}^{\text{T3D1a}} = \text{Tr}[\Gamma_\rho^L L(x, u) \gamma_\mu L(u, v) \gamma_\nu L(v, x) \Gamma_\rho^L \{L(x, t_K) \gamma_5 H(t_K, x) - \text{H.c.}\}], \quad (\text{A33})$$

plus the b diagrams with $(u, \mu) \leftrightarrow (v, \nu)$. We adopt the parametrization (A2) for the type-3 diagrams with reference point at v so that the “bubble” formed by the kaon wall and the weak operator is summed over the whole spacetime volume, presumably leading to smaller noise for this part. The reduced amplitudes are then given by

$$\bar{\mathcal{A}}_1^{\text{T3D1a}} = \sum_{u,x \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(t_{\text{sep}}+d)} \text{Tr}[\Gamma_\rho^L F(x, v)] \text{Tr}[\Gamma_\rho^L G(x, t_K)], \quad (\text{A34})$$

$$\bar{\mathcal{A}}_1^{\text{T3D1b}} = \sum_{u,x \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(t_{\text{sep}}+d)} \text{Tr}[\Gamma_\rho^L F^\dagger(x, v)] \text{Tr}[\Gamma_\rho^L G(x, t_K)], \quad (\text{A35})$$

$$\bar{\mathcal{A}}_2^{\text{T3D1a}} = \sum_{u,x \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(t_{\text{sep}}+d)} \text{Tr}[\Gamma_\rho^L F(x, v) \Gamma_\rho^L G(x, t_K)], \quad (\text{A36})$$

$$\bar{\mathcal{A}}_2^{\text{T3D1b}} = \sum_{u,x \in \Lambda | d \leq \delta_{\max}} \delta_{v_0-x_0,d} e^{M_K(t_{\text{sep}}+d)} \text{Tr}[\Gamma_\rho^L F^\dagger(x, v) \Gamma_\rho^L G(x, t_K)], \quad (\text{A37})$$

where

$$F(x, v) \equiv S(x, v) \gamma_5 L^\dagger(x, v), \quad (\text{A38})$$

$$G(x, t_K) \equiv L(x, \tilde{t}_K) H^\dagger(x, \tilde{t}_K) \gamma_5 - \text{H.c.}, \quad (\text{A39})$$

$$S(x, v) \equiv L(x, u) \gamma_\mu L(u, v) \gamma_\nu K_{\mu\nu}(u - v). \quad (\text{A40})$$

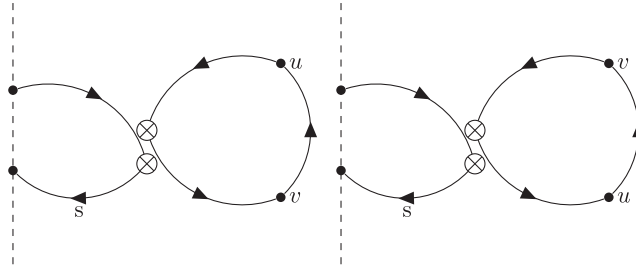


FIG. 14. Type-3 diagrams. From left to right: T3D1a and T3D1b.

Compared to the previous cases, extra propagators are needed in addition to the available point- and wall-source ones due to the summation over two of the coordinate-space variables. Instead of using the sequential-propagator technique, we adopt the all-to-all propagator technique [48] to reduce the overhead on the inversion of the Dirac operator while computing with different $R_{\max}$ -regulated leptonic kernels. In our implementation, we use a low-mode-improved stochastic estimator, where the Z-Möbius-approximated low-mode part of the Dirac operator is implemented exactly from the eigenvectors of the preconditioned Dirac operator and high-mode part reconstructed stochastically. The method works as follows. Let ξ be a random variable satisfying the usual orthonormality condition on the entire spacetime

$$\mathbb{E}[\xi(x)\xi^\dagger(y)] = \delta_{x,y}, \quad (\text{A41})$$

where $\mathbb{E}$ is the expectation value over the probability distribution of ξ . With $\{\xi_j\}_{1 \leq j \leq N_{\text{hits}}}$ a set of realizations of ξ and two sets of spin-color vectors $\{v_i\}_{1 \leq i \leq N_{\text{ev}}}$ and $\{w_i\}_{1 \leq i \leq N_{\text{ev}}}$, we can construct an unbiased estimator for the light-quark propagator

$$\begin{aligned} \hat{L}(x, y) &= \sum_{i=1}^{N_{\text{ev}}} v_i(x) \otimes [w_i]^*(y) \\ &\quad - \frac{1}{N_{\text{hits}}} \sum_{i=1}^{N_{\text{ev}}} \sum_{j=1}^{N_{\text{hits}}} \sum_{z \in \Lambda} \langle w_i(z), \xi_j(z) \rangle v_i(x) \otimes \xi_j^*(y) \\ &\quad + \frac{1}{N_{\text{hits}}} \sum_{j=1}^{N_{\text{hits}}} \sum_{z \in \Lambda} L(x, z) \xi_j(z) \otimes \xi_j^*(y). \end{aligned} \quad (\text{A42})$$

For our setup, we construct the vectors $\{v_i\}_{1 \leq i \leq N_{\text{ev}}}$ and $\{w_i\}_{1 \leq i \leq N_{\text{ev}}}$ from the (approximated) low modes of the preconditioned Dirac operator as explained in Appendix B. Those will be denoted with a superscript “l” [Eqs. (B9) and (B10)]. We exploit the following identity valid for two spin-color vectors V and W and a spin-color matrix B at a given site

$$\text{Tr}[(V \otimes W^*)B] = \langle B^\dagger W, V \rangle, \quad (\text{A43})$$

where $\langle \cdot, \cdot \rangle$ is the usual scalar product defined for spin-color vectors.

Define

$$\Phi(u, v) \equiv \gamma_\mu L(u, v) \gamma_\nu K_{\mu\nu}(u - v), \quad (\text{A44})$$

and

$$\mathcal{G}_\rho^l(x, t_K) \equiv \text{Tr}[\Gamma_\rho^L G(x, t_K)], \quad (\text{A45})$$

$$\mathcal{G}_\rho^2(x, t_K) \equiv \Gamma_\rho^L G(x, t_K). \quad (\text{A46})$$

With Eqs. (A42) and (A43), we can express the reduced amplitudes (A34)–(A37) as

$$\bar{\mathcal{A}}_1^{\text{T3D1a}} = \sum_{d \leq \delta_{\max}} e^{M_K(t_{\text{sep}}+d)} \hat{\mathcal{C}}_1^a(d, t_{\text{sep}}, v), \quad (\text{A47})$$

$$\bar{\mathcal{A}}_2^{\text{T3D1a}} = \sum_{d \leq \delta_{\max}} e^{M_K(t_{\text{sep}}+d)} \hat{\mathcal{C}}_2^a(d, t_{\text{sep}}, v), \quad (\text{A48})$$

$$\bar{\mathcal{A}}_1^{\text{T3D1b}} = \sum_{d \leq \delta_{\max}} e^{M_K(t_{\text{sep}}+d)} \hat{\mathcal{C}}_1^b(d, t_{\text{sep}}, v), \quad (\text{A49})$$

$$\bar{\mathcal{A}}_2^{\text{T3D1b}} = \sum_{d \leq \delta_{\max}} e^{M_K(t_{\text{sep}}+d)} \hat{\mathcal{C}}_2^b(d, t_{\text{sep}}, v), \quad (\text{A50})$$

where

$$\begin{aligned} \hat{\mathcal{C}}_q^a(d, t_{\text{sep}}, v) &= \sum_{i=1}^{N_{\text{ev}}} \langle M_i(v), N_{q,i}^a(d, t_{\text{sep}}, v) \rangle \\ &\quad - \frac{1}{N_{\text{hits}}} \sum_{i=1}^{N_{\text{ev}}} \sum_{j=1}^{N_{\text{hits}}} \langle w_i^l(z), \xi_j(z) \rangle \langle P_j(v), N_{q,i}^a(d, t_{\text{sep}}, v) \rangle \\ &\quad + \frac{1}{N_{\text{hits}}} \sum_{j=1}^{N_{\text{hits}}} \langle P_j(v), Q_{q,j}^a(d, t_{\text{sep}}, v) \rangle, \end{aligned} \quad (\text{A51})$$

$$\begin{aligned} \hat{\mathcal{C}}_q^b(d, t_{\text{sep}}, v) &= \sum_{i=1}^{N_{\text{ev}}} \langle M_i(v), N_{q,i}^b(d, t_{\text{sep}}, v) \rangle^* \\ &\quad - \frac{1}{N_{\text{hits}}} \sum_{i=1}^{N_{\text{ev}}} \sum_{j=1}^{N_{\text{hits}}} \langle w_i^l(z), \xi_j(z) \rangle^* \langle P_j(v), N_{q,i}^b(d, t_{\text{sep}}, v) \rangle^* \\ &\quad + \frac{1}{N_{\text{hits}}} \sum_{j=1}^{N_{\text{hits}}} \langle P_j(v), Q_{q,j}^b(d, t_{\text{sep}}, v) \rangle^*, \end{aligned} \quad (\text{A52})$$

$$M_i(v) \equiv \sum_{u \in \Lambda} \Phi^\dagger(u, v) w_i^l(u), \quad (\text{A53})$$

$$N_{q,i}^a(d, t_{\text{sep}}, v) \equiv \sum_{x \in \Lambda} \delta_{v_0-x_0, d} \gamma_5 L^\dagger(x, v) \mathcal{G}_\rho^q(x, t_K) \Gamma_\rho^L v_i^l(x), \quad (\text{A54})$$

$$N_{q,i}^b(d, t_{\text{sep}}, v) \equiv \sum_{x \in \Lambda} \delta_{v_0-x_0, d} \gamma_5 L^\dagger(x, v) [\mathcal{G}_\rho^q(x, t_K) \Gamma_\rho^L]^\dagger v_i^l(x), \quad (\text{A55})$$

$$P_j(v) \equiv \sum_{u \in \Lambda} \Phi^\dagger(u, v) \xi_j(u), \quad (\text{A56})$$

$$Q_{q,j}^a(d, t_{\text{sep}}, v) \equiv \sum_{x \in \Lambda} \delta_{v_0-x_0, d} \gamma_5 L^\dagger(x, v) \mathcal{G}_\rho^q(x, t_K) \Gamma_\rho^L \Omega_j(x), \quad (\text{A57})$$

$$Q_{q,j}^b(d, t_{\text{sep}}, v) \equiv \sum_{x \in \Lambda} \delta_{v_0-x_0, d} \gamma_5 L^\dagger(x, v) [\mathcal{G}_\rho^q(x, t_K) \Gamma_\rho^L]^\dagger \Omega_j(x), \quad (\text{A58})$$

$$\Omega_j(x) \equiv \sum_{z \in \Lambda} L(x, z) \xi_j(z). \quad (\text{A59})$$

Among the above quantities, we have a clear separation between the kernel-dependent quantities and the t_{sep} -dependent ones. This factorization helps reduce the

overhead for getting results at various $(t_{\text{sep}}, R_{\text{max}})$. With the available precomputed point-source propagators, we can construct the kernel-dependent quantities $M_i(v)$ and $P_j(v)$ for a fixed reference point v . Although the contraction cost scales linearly with the numbers of vectors used, i.e., N_{ev} and N_{hits} , with our typical number of low modes of $O(1000)$, the calculation is still dominated by the inversion of the Dirac operator on random vectors to get $\Omega_j(x)$ [Eq. (A59)].

4. Type 4

The diagrams for this contraction class, T4D1a, T4D1b, T4D2a and T4D2b, are depicted in Fig. 15. The corresponding contraction functions entering Eq. (A2) are defined as

$$\mathcal{C}_1^{\text{T4D1a}} = \text{Tr} \left[\Gamma_\rho^L L(x, x) \right] \text{Tr} \left[\Gamma_\rho^L \{ L(x, v) \gamma_\nu L(v, u) \gamma_\mu L(u, t_K) \gamma_5 H(t_K, x) - \text{H.c.} \} \right], \quad (\text{A60})$$

$$\mathcal{C}_2^{\text{T4D1a}} = \text{Tr} \left[\Gamma_\rho^L L(x, x) \Gamma_\rho^L \{ L(x, v) \gamma_\nu L(v, u) \gamma_\mu L(u, t_K) \gamma_5 H(t_K, x) - \text{H.c.} \} \right], \quad (\text{A61})$$

$$\mathcal{C}_1^{\text{T4D2a}} = \text{Tr} \left[\Gamma_\rho^L L(x, x) \right] \text{Tr} \left[\Gamma_\rho^L \{ L(x, t_K) \gamma_5 H(t_K, u) \gamma_\mu H(u, v) \gamma_\nu H(v, x) - \text{H.c.} \} \right], \quad (\text{A62})$$

$$\mathcal{C}_2^{\text{T4D2a}} = \text{Tr} \left[\Gamma_\rho^L L(x, x) \Gamma_\rho^L \{ L(x, t_K) \gamma_5 H(t_K, u) \gamma_\mu H(u, v) \gamma_\nu H(v, x) - \text{H.c.} \} \right], \quad (\text{A63})$$

plus the b diagrams with $(\mu, u) \leftrightarrow (\nu, v)$.

We use the parametrization equation (A2) with the reference point at v (or with u and v swapped for the b diagrams) so that we can compute these diagrams with several kernels with different R_{max} without involving sequential propagators. With this parametrization, we need to know the tadpole $\text{Tr}[L(x, x)]$ at the weak Hamiltonian for all $x \in \Lambda$. This can be achieved using a similar low-mode-improved stochastic estimator (A42) discussed in the type-3 case,

$$\text{Tr}[L(x, x)] = \mathbb{E} \left[\text{Tr} \left[\hat{L}(x, x) \right] \right]. \quad (\text{A64})$$

For this quantity, we use full spacetime volume sources for the random vector instead, which is supposed to be more

efficient due to the exponentially suppressed off-diagonal terms in the propagator. We can decompose the amplitudes into blocks with explicit t_{sep} or kernel dependence to reduce the computational overhead.

Define

$$B_1(v, t_K) = (-1) \times \sum_{u \in \Lambda} K_{\mu\nu}(u-v) \Gamma_\mu^5 L^\dagger(u, v) \Gamma_\nu^5 L(u, t_K), \quad (\text{A65})$$

$$B_2(v, t_K) = (-1) \times \sum_{u \in \Lambda} H^\dagger(u, t_K) K_{\mu\nu}(u-v) \Gamma_\mu^5 H(u, v) \Gamma_\nu^5, \quad (\text{A66})$$

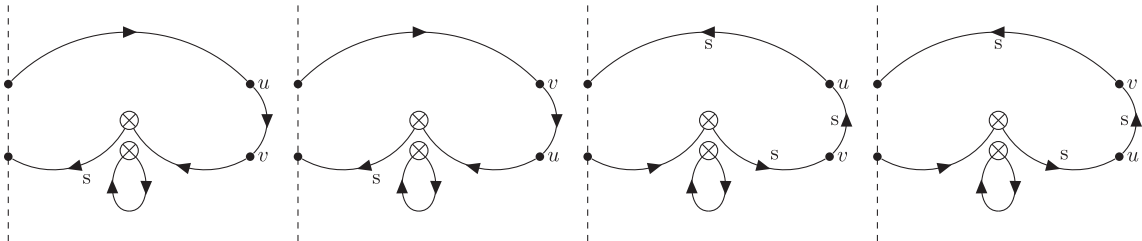


FIG. 15. Type-4 diagrams. From left to right: T4D1a, T4D1b, T4D2a, T4D2b.

$$B_3(u, v_0, t_K) = (-1) \times \sum_{\vec{v} \in \Lambda_0} K_{\mu\nu}(u-v) \Gamma_\mu^5 L^\dagger(v, u) \Gamma_\nu^5 L(v, t_K), \quad (\text{A67})$$

$$B_4(u, v_0, t_K) = (-1) \times \sum_{\vec{v} \in \Lambda_0} H^\dagger(v, t_K) K_{\mu\nu}(u-v) \Gamma_\mu^5 H(v, u) \Gamma_\nu^5, \quad (\text{A68})$$

$$G_1(x) = \text{Tr}[\Gamma_\rho^L L(x, x)] \Gamma_\rho^L, \quad (\text{A69})$$

$$G_2(x) = \Gamma_\rho^L L(x, x) \Gamma_\rho^L, \quad (\text{A70})$$

and for $i = 1, 2$,

$$E_i^{(1)}(x, t_K) = H^\dagger(x, t_K) \gamma_5 G_i(x), \quad (\text{A71})$$

$$E_i^{(2)}(x, t_K) = G_i(x) \gamma_5 H(x, t_K), \quad (\text{A72})$$

$$F_i^{(1)}(x, t_K) = \gamma_5 G_i(x) L(x, t_K), \quad (\text{A73})$$

$$F_i^{(2)}(x, t_K) = L^\dagger(x, t_K) G_i(x) \gamma_5. \quad (\text{A74})$$

We can then write the reduced amplitudes of the type-4 “a” diagrams as

$$\bar{\mathcal{A}}_i^{\mathcal{D}} = \sum_{d \leq \delta_{\max}} \sum_{x \in \Lambda} \delta_{v_0 - x_0, d} e^{M_K(t_{\text{sep}} + d)} \tilde{\mathcal{C}}_i^{\mathcal{D}}(x, v, t_{\text{sep}}), \quad (\text{A75})$$

with, for $i = 1, 2$,

$$\tilde{\mathcal{C}}_i^{\text{T4D1a}}(x, v, t_{\text{sep}}) = \text{Tr} \left[E_i^{(2)}(x, t_K) B_1^\dagger(v, t_K) L^\dagger(x, v) - E_i^{(1)}(x, t_K) L(x, v) B_1(v, t_K) \right], \quad (\text{A76})$$

$$\tilde{\mathcal{C}}_i^{\text{T4D2a}}(x, v, t_{\text{sep}}) = \text{Tr} \left[F_i^{(1)}(x, t_K) B_2(v, t_K) H^\dagger(x, v) - F_i^{(2)}(x, t_K) H(x, v) B_2^\dagger(v, t_K) \right]. \quad (\text{A77})$$

Similarly, for the b diagrams, we write

$$\bar{\mathcal{A}}_i^{\mathcal{D}} = \sum_{d \leq \delta_{\max}} \sum_{v_0=0}^{T-1} \sum_{x \in \Lambda} \delta_{v_0 - x_0, d} e^{M_K(t_{\text{sep}} + d)} \tilde{\mathcal{C}}_i^{\mathcal{D}}(x, u, t_{\text{sep}}), \quad (\text{A78})$$

where

$$\tilde{\mathcal{C}}_i^{\text{T4D1b}}(x, u, t_{\text{sep}}) = \text{Tr} \left[E_i^{(1)}(x, t_K) L(x, u) B_3(u, v_0, t_K) - E_i^{(2)}(x, t_K) B_3^\dagger(u, v_0, t_K) L^\dagger(x, u) \right], \quad (\text{A79})$$

$$\tilde{\mathcal{C}}_i^{\text{T4D2b}}(x, u, t_{\text{sep}}) = \text{Tr} \left[F_i^{(2)}(x, t_K) H(x, u) B_4^\dagger(u, v_0, t_K) - F_i^{(1)}(x, t_K) B_4(u, v_0, t_K) H^\dagger(x, u) \right], \quad (\text{A80})$$

for $i = 1, 2$.

5. Type 5

The type-5 diagrams, T5D1 and T5D2 are depicted in Fig. 16. The reduced amplitudes (A2) for these diagrams are

$$\bar{\mathcal{A}}_i^{\text{T5D1}} = \sum_{d \leq \delta_{\max}} \sum_{u \in \Lambda} e^{M_K(t_{\text{sep}} + d)} F_i(t_{\text{sep}}, d, v_0) K_{\mu\nu}(u-v) \Pi_{\mu\nu}^{\text{T5D1}}(u, v), \quad (\text{A81})$$

$$\bar{\mathcal{A}}_i^{\text{T5D2}} = \sum_{d \leq \delta_{\max}} \sum_{u \in \Lambda} e^{M_K(t_{\text{sep}} + d)} F_i(t_{\text{sep}}, d, v_0) K_{\mu\nu}(u-v) \Pi_{\mu\nu}^{\text{T5D2}}(u, v), \quad (\text{A82})$$

where

$$F_i(t_{\text{sep}}, d, v_0) \equiv \sum_{\substack{\vec{x} \in \Lambda_0(x_0), \\ x_0 - t_K^{-1} t_{\text{sep}} = d, \\ v_0 - t_0 = d}} \text{Tr}[\mathcal{G}_i(x) \{H(x, t_K) \gamma_5 L(t_K, x) - \text{H.c.}\}], \quad (\text{A83})$$

$$\mathcal{G}_1 = \text{Tr}[\Gamma_\rho^L L(x, x)] \Gamma_\rho^L, \quad (\text{A84})$$

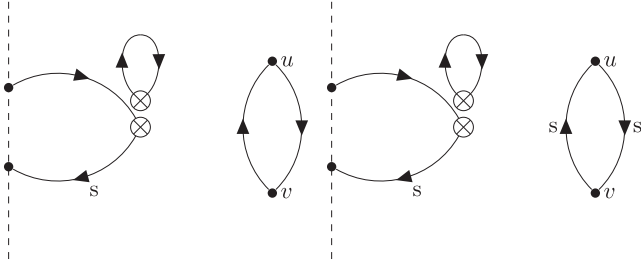


FIG. 16. T5D1 (left) and T5D2 (right). The point source is put at v if a point-source propagator is used.

$$\mathcal{G}_2 = \Gamma_\rho^L L(x, x) \Gamma_\rho^L, \quad (\text{A85})$$

$$\Pi_{\mu\nu}^{\text{T5D1}}(u, v) \equiv \text{Tr}[L(u, v) \gamma_\nu L(v, u) \gamma_\mu], \quad (\text{A86})$$

$$\Pi_{\mu\nu}^{\text{T5D2}}(u, v) \equiv \text{Tr}[H(u, v) \gamma_\nu H(v, u) \gamma_\mu]. \quad (\text{A87})$$

Because of the quark-disconnected structure, these diagrams are intrinsically noisy. For T5D2, we find it sufficient to sample the reduced amplitude at many different

reference points v as done for the type-3 and type-4 diagrams. T5D1 is much noisier due to the light-quark propagators. To improve the statistics, we adapt the low-mode averaging method [49] for our calculation of Π^{T5D1} . With the available (Z-Möbius-approximated) low modes, we decompose a propagator into a low-mode (l) and a high-mode (h) part, defined according to

$$L(x, y) = L^l(x, y) + L^h(x, y),$$

$$L^l(x, y) \equiv \sum_{i=1}^{N_{\text{ev}}} v_i^l(x) \otimes [w_i^l]^*(y), \quad (\text{A88})$$

where the vectors v_i^l and w_i^l are given in Eqs. (B9) and (B10). We then decompose Π^{T5D1} into

$$\begin{aligned} \Pi_{\mu\nu}^{\text{T5D1}}(u, v) &= \Pi_{\mu\nu}^l(u, v) + \Pi_{\mu\nu}^h(u, v), \\ \Pi_{\mu\nu}^l(u, v) &= \text{Tr}[L^l(u, v) \gamma_\nu L^l(v, u) \gamma_\mu]. \end{aligned} \quad (\text{A89})$$

For the contribution of the low-mode part, we can do an exact volume average via

$$\begin{aligned} \mathcal{A}_i^{\text{T5D1,l}} &\equiv \frac{1}{V} \sum_{d \leq \delta_{\text{max}}} \sum_{u, v \in \Lambda} e^{M_K(t_{\text{sep}}+d)} F_i(t_{\text{sep}}, \delta, v_0) K_{\mu\nu}(u-v) \Pi_{\mu\nu}^l(u, v) \\ &= \frac{1}{V} \sum_{d \leq \delta_{\text{max}}} \sum_{u, v \in \Lambda} \sum_{i,j=1}^{N_{\text{ev}}} e^{M_K(t_{\text{sep}}+d)} F_i(t_{\text{sep}}, \delta, v_0) K_{\mu\nu}(u-v) G_\nu^{(i,j)}(v) G_\mu^{(j,i)}(u). \end{aligned} \quad (\text{A90})$$

Here V is the spacetime volume and

$$G_\nu^{(i,j)}(u) \equiv \langle w_i^l(u), \gamma_\nu v_j^l(u) \rangle. \quad (\text{A91})$$

We recognize the sum over u in Eq. (A90) is a convolution with the leptonic kernel, which can be done efficiently with a fast Fourier transform. On the other hand, while the contribution of high-mode part Π^h to the reduced amplitude can be sampled using stochastic methods, we find it more cost efficient to do it with point-source propagators at many different source positions v .

APPENDIX B: PRECONDITION OF THE DIRAC OPERATOR AND THE Z-MÖBIUS LOW MODES

For part of the calculations reported in Appendix A, Z-Möbius-approximated low modes are exploited in order to improve sampling efficiency. We give some details about the computation of these low modes in this appendix.

Using the notation of Ref. [46], we solve the preconditioned system with the Möbius domain wall Dirac operator D_{MDWF} ,

$$D_{\text{prec}} \psi = D_- \eta, \quad D_{\text{prec}} \equiv D_- D_{\text{MDWF}}. \quad (\text{B1})$$

With the even-odd preconditioning

$$D_{\text{prec}} = \begin{pmatrix} M_{ee} & M_{eo} \\ M_{oe} & M_{oo} \end{pmatrix}, \quad (\text{B2})$$

we can Schur decompose the inverse of D_{prec} as

$$D_{\text{prec}}^{-1} = \begin{pmatrix} 1 & -M_{ee}^{-1} M_{eo} \\ 0 & 1 \end{pmatrix} \mathcal{M} \begin{pmatrix} 1 & 0 \\ 0 & D_{oo}^\dagger \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -M_{oe} M_{ee}^{-1} & 1 \end{pmatrix}, \quad (\text{B3})$$

where

$$\mathcal{M} \equiv \begin{pmatrix} M_{ee}^{-1} & 0 \\ 0 & (D_{oo}^\dagger D_{oo})^{-1} \end{pmatrix}, \quad (\text{B4})$$

$$D_{oo} \equiv M_{oo} - M_{oe} M_{ee}^{-1} M_{eo}. \quad (\text{B5})$$

We use the SchurDiagTwo scheme defined in the GRID library [41], which further re-expresses Eq. (B3) in terms of

$$D'_{oo} \equiv D_{oo} M_{oo}^{-1}. \quad (\text{B6})$$

We define the low-mode part of the five-dimensional D_{MDWF} by replacing $\mathcal{M}$ in the definition of D_{prec} with the first N_{ev} orthonormal eigenvectors $\{h_i\}$ of $(D'_{oo})^\dagger D'_{oo}$ with associated eigenvalues $\{\lambda_i\}_{1 \leq i \leq N_{\text{ev}}}$,

$$\mathcal{M}^1 = \begin{pmatrix} 0 & 0 \\ 0 & M_{oo}^{-1} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \sum_{i=1}^{N_{\text{ev}}} \lambda_i^{-1} h_i \otimes h_i^* \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & [M_{oo}^{-1}]^\dagger \end{pmatrix}. \quad (\text{B7})$$

Denote V^{45} and U^{54} the maps projecting to and from the physical four-dimensional space on the fifth-dimension boundary, we define the low-mode part of the physical, four-dimensional propagator as

$$L^1(x, y) = \sum_{i=1}^{N_{\text{ev}}} v_i^1(x) \otimes [w_i^1]^*(y), \quad (\text{B8})$$

where

$$v_i^1 = V^{45} \lambda_i^{-1} \begin{pmatrix} -M_{ee}^{-1} M_{eo} M_{oo}^{-1} h_i \\ M_{oo}^{-1} h_i \end{pmatrix}, \quad (\text{B9})$$

$$w_i^1 = [U^{54}]^\dagger D_-^\dagger \begin{pmatrix} -[M_{ee}^{-1}]^\dagger M_{oe}^\dagger D'_{oo} h_i \\ D'_{oo} h_i \end{pmatrix}. \quad (\text{B10})$$

For this work, the eigenvectors involved are not generated for the original D_{MDWF} but for its Z-Möbius-approximated counterpart (see Ref. [29] and references therein) to reduce computational cost as it has a shorter extent in the fifth dimension.

APPENDIX C: ELIMINATING THE UNPHYSICAL INTERMEDIATE-STATE CONTRIBUTION WITHOUT INTRODUCING $\mathcal{O}(a)$ ERRORS

In this appendix we describe how the continuum equations (6) and (7) of Sec. II are implemented in our lattice calculation in a way that avoids introducing $\mathcal{O}(a)$ errors. Since our gauge and fermion actions introduce finite lattice spacing errors only at $\mathcal{O}(a^2)$ and above, we must take care that any spacetime integrals we perform, such as the integral over v_0 in Eq. (6), do not introduce errors of $\mathcal{O}(a)$. This attention to possible $\mathcal{O}(a)$ errors when removing the contributions from excited states is part of the larger question, to be addressed in a later paper, of correcting unphysical short-distance contributions present in our three-flavor calculation. For these corrections as well we must be careful that they do not inadvertently introduce $\mathcal{O}(a)$ themselves.

Consider the case of a single state $|n\rangle$ with energy $E_n < M_K$ and represent Eqs. (6) and (7) schematically by the two equations

$$A^{\text{lat}}(\delta_{\text{max}}) = \int_{-\delta_{\text{min}}}^{\delta_{\text{max}}} f(v_0) dv_0, \quad (\text{C1})$$

$$A^{\text{sub}}(\delta_{\text{max}}) = A_n \frac{e^{-(E_n - M_K)\delta_{\text{max}}}}{M_K - E_n}, \quad (\text{C2})$$

where for large v_0 the integrand in Eq. (C1) behaves as

$$f(v_0) \sim A_n e^{-(E_n - M_K)v_0}. \quad (\text{C3})$$

We can avoid introducing errors of $\mathcal{O}(a)$ if we use the trapezoidal rule when replacing the integral over v_0 in Eq. (C1) by a sum over lattice times,

$$A^{\text{lat}}(\delta_{\text{max}}) = +\frac{1}{2}f(\delta_{\text{max}}) + \sum_{v_0=-\delta_{\text{min}}}^{\delta_{\text{max}}-1} f(v_0). \quad (\text{C4})$$

Here v_0 , δ_{max} and δ_{min} are now integers corresponding to discrete values of the time variable v_0 . Note, we have not replaced the term $f(-\delta_{\text{min}})$ evaluated at the most negative value of v_0 by a term reduced by a factor of 1/2 because we assume that δ_{min} is sufficiently large that $f(-\delta_{\text{min}}) = 0$. The lattice version of Eq. (5) can then be written,

$$A_{\text{LD2}\gamma}(\delta_{\text{max}}) = +\frac{1}{2}f(\delta_{\text{max}}) + \sum_{v_0=-\delta_{\text{min}}}^{\delta_{\text{max}}-1} f(v_0) - A_n \frac{e^{-(E_n - M_K)\delta_{\text{max}}}}{M_K - E_n}. \quad (\text{C5})$$

Equation (C5) provides the physical result for LD2 γ amplitude by removing the unphysical exponentially growing contribution from the intermediate state $|n\rangle$ from the $\mathcal{O}(a^2)$ -accurate sum over lattice four-point Green's function results. For the case that more than a single state $|n\rangle$ has energy below M_K , the rightmost subtraction term on the right-hand side of Eq. (C5) must be replaced by a sum over all such states.

In Eq. (C5) the upper limit δ_{max} must be chosen sufficiently large that the unphysical, exponentially falling terms associated with states with energies larger than M_K can be neglected. In the case of the η meson, the noise associated with the disconnected diagrams does not allow such a large choice for δ_{max} and the unphysical, exponentially falling contribution of the η should also be included in the subtraction term on the right-hand side of Eq. (C5).

While Eq. (C5) is all that is needed to obtain a physical result, it may also be useful to have an expression for the four-point Green's function as a function of v_0 from which these unphysical terms have been subtracted. However, for v_0 near zero, specifying such a subtracted Green's function is somewhat arbitrary since the subtraction term $A^{\text{sub}}(\delta)$ is meaningful only when δ is large compared to the lattice spacing. For example when $v_0 = 0$ we encounter the product $J_\nu(\mathbf{v}, 0) \mathcal{H}_W(\mathbf{0}, 0)$. If written in this order we can

insert a π^0 intermediate state while if the factors J and H are reversed we cannot.

For the purposes of plotting the subtracted Green's function as a function of v_0 , we make the choice

$$f^{\text{sub}}(v_0) = f(v_0) - w(v_0)A_n e^{-(E_n - M_K)v_0} \quad (\text{C6})$$

where the subtraction weight function $w(v_0)$ is given by

$$\omega(v_0) = \begin{cases} e^{-(E_n - M_K)v_0} & \text{if } v_0 \geq 1, \\ \frac{1}{2} & \text{if } v_0 = 0, \\ 0 & \text{if } v_0 < 0. \end{cases} \quad (\text{C7})$$

While this choice of $w(v_0)$ may be somewhat arbitrary, the subtracted Green's function $f^{\text{sub}}(v_0)$ is a reasonable choice and does obey the necessary requirement that when summed over v_0 the result agrees with the $\mathcal{O}(a^2)$ -correct result in Eq. (C5),

$$A_{\text{LD}2\gamma}(\delta_{\text{max}}) = \sum_{v_0=-\delta_{\text{min}}}^{\delta_{\text{max}}} f^{\text{sub}}(v_0) \quad (\text{C8})$$

for sufficiently large δ_{max} , up to terms of $\mathcal{O}(a^2)$.

APPENDIX D: SYSTEMATIC ERROR FROM THE THREE-PION INTERMEDIATE STATE

In the framework discussed in Sec. II, we ignored possible unphysical contributions from three-pion state in Eq. (7). Depending on the geometry of the lattice, one might need to subtract the unphysical contributions from the excited three-pion states in addition to the ground state as they could be lighter than the kaon. Treating the pions as noninteracting, unphysical, exponentially growing contributions from 3π -excited states start to appear at a spatial lattice extent $L \gtrsim 11$ fm with physical pion and kaon. Nonetheless, already as part of our strategy to avoid the complications associated with $\pi\pi(\gamma)$ intermediate state, such a large volume should not be included. As a consequence, the unphysical contribution from the excited 3π states should not be an issue in our proposed lattice calculation. On the other hand, as seen in Sec. III A, subtracting the unphysical π^0 contribution seems to result in a stable plateau in the cutoff δ_{max} , suggesting that the unphysical contribution from the 3π ground state is negligible under the achievable precision. However, just like the case of the η intermediate state described in Sec. III B, as the energy of the 3π state is fairly close to the mass of the kaon, the amplitude might still be far from saturated at the cutoff. The objective of this section is to estimate the size of the systematic error related to this issue

with the 3π state. To this end, we write schematically the spectral decomposition of the contribution from the 3π channel to the $K_L \rightarrow \mu^+\mu^-$ amplitude as

$$\mathcal{A}_{3\pi} \equiv \left[\int_{3m_\pi}^{E_{\text{cut}}} dE_{3\pi} + \int_{E_{\text{cut}}}^{\infty} dE_{3\pi} \right] \frac{A(E_{3\pi})}{E_{3\pi} - M_K - i\varepsilon}. \quad (\text{D1})$$

In the above, E_{cut} is a cutoff energy which we cannot probe with a restricted spatial lattice extent. The $i\varepsilon$ prescription is used here to indicate that we are focusing on the $K_L \rightarrow 3\pi \rightarrow \gamma^*\gamma^*$ process where special care is required: analogous to the case of the π^0 , the unphysical 3π states appear under Wick rotation with the time ordering where the weak Hamiltonian is at an earlier time than the two EM currents [Eq. (7)]. A typical gauge ensemble with physical pion mass which we would include in our continuum trajectory would have $m_\pi L \approx 4$. On such an ensemble, the first excited (noninteracting) three-pion state produced by the initial kaon, comprising two back-to-back pions and a pion at rest, has an energy around 660 MeV. This provides a natural energy cutoff E_{cut} in Eq. (D1). If computed in Euclidean space, the second integral in the square brackets in Eq. (D1) can be saturated to about 60%–70% if the data quality allows to reach an Euclidean time of 1.2–1.6 fm. Assuming that the latter is feasible, the main source of systematic error with the proposed lattice calculation should come from the first integral in the square brackets of Eq. (D1).⁵ To proceed further, we make the hypothesis that there is no steep enhancement for $A(E_{3\pi})$ in the interval $E_{3\pi} \in [3m_\pi, E_{\text{cut}}] = [420, 660]$ MeV and that it can be approximated by a constant $A(E_{3\pi}) \approx A(M_K)$. Then,

$$\begin{aligned} \mathcal{S} &\equiv \int_{3m_\pi}^{E_{\text{cut}}} dE_{3\pi} \frac{A(M_K)}{E_{3\pi} - M_K - i\varepsilon} \\ &= A(M_K) \left[\ln \left(\frac{E_{\text{cut}} - M_K}{M_K - 3m_\pi} \right) + i\pi \right] \end{aligned} \quad (\text{D2})$$

should give a realistic estimate for the systematic error related to the 3π intermediate state in our lattice calculation.

Through out this appendix, we will use the convention of Eq. (4) for the amplitudes. We will use the shorthand notation $|3\pi\rangle \equiv |\pi^{a_1}(\mathbf{q}_1)\pi^{a_2}(\mathbf{q}_2)\pi^{a_3}(\mathbf{q}_3)\rangle$, where the superscripts indicate the charges, which are kept generic for the moment. In Minkowski space, this contribution reads

⁵In the case of our 24ID study, this range corresponds to $\delta_{\text{max}} \in [6, 8]$.

$$\begin{aligned}
\mathcal{A}_{3\pi} &= \int d\Pi_{3\pi} \int d^4r L_{\mu\nu}(r) e^{iP \cdot r/2} \frac{\delta^2}{e^2 \delta A_\mu(r) \delta A_\nu(0)} \left[\langle 0 | e^{iS_{\text{int}}} | 3\pi \rangle \right] \\
&\times \frac{\langle 3\pi | \mathcal{H}_W(0) | K_L \rangle}{E_{3\pi} - M_K - i\varepsilon} (2\pi)^3 \delta^3 \left(\sum_{i=1}^3 \mathbf{q}_i \right) \\
&= -i \frac{m_\mu e^2}{4096\pi^{10} M_K} \int \frac{dE_1 dE_2 d^2\Omega_1 d^2\Omega_2 |\mathbf{q}_1| |\mathbf{q}_2|}{E_3(E_1 + E_2 + E_3 - M_K - i\varepsilon)} \langle 3\pi | \mathcal{H}_W(0) | K_L \rangle \\
&\times \int d^4p \frac{\epsilon^{\mu\nu\alpha\beta} p_\alpha P_\beta \widetilde{J}_\mu \widetilde{J}_\nu(p)}{[(P-p)^2 + i\varepsilon](p^2 + i\varepsilon)[(p-k^+)^2 - m_\mu^2 + i\varepsilon]}, \tag{D3}
\end{aligned}$$

where k^+ is the momentum of the outgoing μ^+ , $P = (M_K, \mathbf{0})$ is the four-momentum of the kaon in the rest frame, $E_i = \sqrt{m_{\pi^i}^2 + \mathbf{q}_i^2}$ and Ω_i are the energy and the solid angle of the spatial momentum of the i th pion,

$$d\Pi_{3\pi} = \frac{1}{(2\pi)^9} \frac{d^3\mathbf{q}_1}{2E_1} \frac{d^3\mathbf{q}_2}{2E_2} \frac{d^3\mathbf{q}_3}{2E_3} |3\pi\rangle \langle 3\pi| \tag{D4}$$

is the Lorentz-invariant phase-space measure of the three pions and

$$\widetilde{J}_\mu \widetilde{J}_\nu(p) \equiv \int d^4r e^{ipr} \frac{\delta^2}{\delta A_\mu(r) \delta A_\nu(0)} \left[\langle 0 | e^{iS_{\text{int}}} | 3\pi \rangle \right], \tag{D5}$$

with S_{int} the action from the (effective) interaction Lagrangian and A_μ a photon field.

The experimental decay amplitude $A_{+-0}^L \equiv \langle \pi^+ \pi^- \pi^0 | \mathcal{H}_W(0) | K_L \rangle$ admits the following parametrization [50]⁶:

$$\begin{aligned}
A_{+-0}^L &= (\alpha_1 + \alpha_3) - (\beta_1 + \beta_3) \bar{y} + (\zeta_1 - 2\zeta_3) \left(\bar{y}^2 + \frac{1}{3} \bar{x}^2 \right) \\
&+ (\xi_1 - 2\xi_3) \left(\bar{y}^2 - \frac{1}{3} \bar{x}^2 \right), \tag{D6}
\end{aligned}$$

where

$$\begin{aligned}
\bar{y} &\equiv \frac{s_3 - s_0}{m_{\pi^+}^2}, \quad \bar{x} \equiv \frac{s_2 - s_1}{m_{\pi^+}^2}, \quad s_i \equiv (P_K - p_i)^2, \\
s_0 &\equiv \frac{1}{3} (M_K^2 + m_{\pi^0}^2 + m_{\pi^+}^2 + m_{\pi^-}^2). \tag{D7}
\end{aligned}$$

Note that the coefficients $\alpha_1, \dots, \xi_3$ in Eq. (D6) fitted in Ref. [50] result in a $K_L \rightarrow \pi^+ \pi^- \pi^0$ decay rate of 1.57×10^{-18} GeV—consistent with the experimental value [2]—if the physical pion masses are used. In the remainder of our

discussion, we will assume isospin symmetry for simplicity. In that limit, Eq. (D6) with the same fitted coefficients gives a decay rate of 1.30×10^{-18} GeV, suggesting a correction for the strong isospin-breaking effects at the 10% level on the decay amplitude.

As there is no direct experimental data for the quantity Eq. (D5), we estimate it in chiral perturbation theory. Because of charge and parity symmetries, one needs to include the Wess-Zumino (WZ) term [52,53] for the coupling of an odd number of pions to two photons. For simplicity, we will only focus on the pion-pole contribution from the leading $O(p^4)$ WZ term and the $O(p^2)$ chiral Lagrangian as displayed in Fig. 17. The logarithmic divergence appearing in the $2\gamma\mu$ triangle diagram in Fig. 17 can be renormalized by adding a $\pi^0 \mu^+ \mu^-$ counterterm in the standard way [54]. Following the conventions of Ref. [55], where the pion decay constant takes the experimental value of $f_\pi \approx 93.3$ MeV, we get

$$\begin{aligned}
\mathcal{A}_{3\pi}^{\pi^0\text{-pole}} &= i \frac{m_\mu M_K e^4}{6144\pi^8 f_\pi^3} \int \frac{dE_1 dE_2 dE_3 \langle \pi^+ \pi^- \pi^0 | \mathcal{H}_W | K_L \rangle}{E_1 + E_2 + E_3 - M_K - i\varepsilon} \\
&\times \frac{(2M_K^2 - 6M_K E_3 + m_\pi^2)}{M_K^2 - m_\pi^2} A_\mu(M_K^2), \tag{D8}
\end{aligned}$$

where, for a lepton l ,

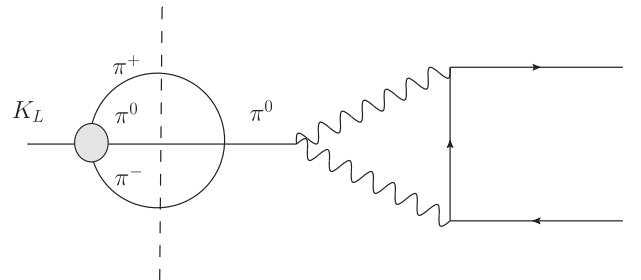
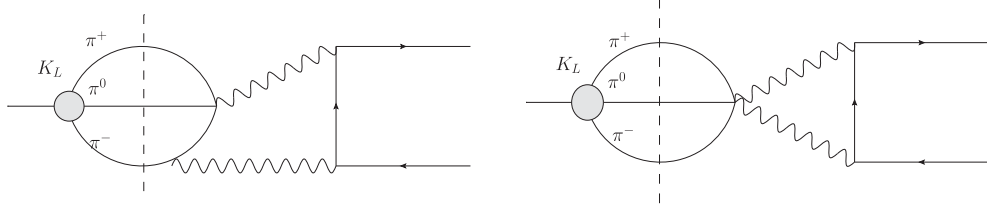


FIG. 17. Three-pion contribution from the four-pion and WZ $\pi^0 \gamma \gamma$ vertices.

⁶Here the notation is due to Ref. [51], where the third pion is the “odd” pion, i.e., the first two of them satisfy the Bose-symmetry constraint.

FIG. 18. Three-pion contribution from the WZ $\pi^0\pi^+\pi^-\gamma$ and $\pi^0\pi^+\pi^-\gamma\gamma$ vertices.

$$A_I(q^2) \equiv \frac{2i}{\pi^2 q^2} \int d^4k \frac{q^2 k^2 - (q \cdot k)^2}{k^2 (q-k)^2 [(p-k)^2 - m_l^2]} \times \tilde{F}(k^2, (q-k)^2), \quad (\text{D9})$$

with $\tilde{F}$ the $\pi^0 \rightarrow \gamma^* \gamma^*$ transition form factor normalized to unity when both photons are on shell. Assuming lepton universality, we can fix the scale-dependent low-energy constant $\chi(\mu)$ using the $\pi^0 \rightarrow e^+ e^-$ process. Within chiral perturbation theory,

$$\text{Re} A_I(q^2)|_{\text{ChPT}} = \frac{\text{Li}_2[-y_I(q^2)] + \frac{1}{4} \ln^2[y_I(q^2)] + \frac{\pi^2}{12}}{\sigma_I(q^2)} + 3 \ln\left(\frac{m_l}{\mu}\right) - \frac{5}{2} + \chi(\mu), \quad (\text{D10})$$

where

$$y_I(q^2) \equiv \frac{1 - \sigma_I(q^2)}{1 + \sigma_I(q^2)}, \quad \sigma_I(q^2) = \sqrt{1 - \frac{4m_l^2}{q^2}}. \quad (\text{D11})$$

From Ref. [56],

$$\chi(\mu = 0.77 \text{ GeV}) = 2.69(10). \quad (\text{D12})$$

Altogether, we get

$$\mathcal{S} = (1.84 + i2.32) \times 10^{-17} \text{ GeV}, \quad (\text{D13})$$

leading to a ratio to the experimentally extracted dispersive $K_L \rightarrow \mu^+ \mu^-$ decay amplitude of $|\mathcal{S}/\text{Re} \mathcal{A}_{K_L \mu \mu}^{\text{exp}}| = 1.3 \times 10^{-4}$. As stated earlier, this estimate is sensitive to the strong isospin-breaking effects as they affect the $K_L \rightarrow \pi^+ \pi^- \pi^0$, but we expect the correction to be minor. Unless the assumption of $A(E_{3\pi}) \approx A(M_K)$ which motivates Eq. (D2) suffers from further enhancements of $\mathcal{O}(100)$ in the range $E_{3\pi} \in [3m_\pi, E_{\text{cut}}]$, the missing 3π contributions from our proposed lattice framework should be less than a percent. Similar calculations can be performed for the case with a $\pi^0 \pi^0 \pi^0$ intermediate state, which lead to an estimate at the same order of magnitude. Finally, we should note that the diagrams in Fig. 18, which are both needed to make a gauge-invariant amplitude, also enter at the same order in chiral perturbation theory. We found the total contribution from those diagrams to be about an order of magnitude smaller than the pion-pole diagram in Fig. 17 and decided not to report the entire calculation in the present paper to avoid clutter.

-
- [1] D. Ambrose *et al.* (E871 Collaboration), Improved branching ratio measurement for the decay $K_L^0 \rightarrow \mu^+ \mu^-$, *Phys. Rev. Lett.* **84**, 1389 (2000).
 - [2] R. L. Workman *et al.* (Particle Data Group Collaboration), Review of particle physics, *Prog. Theor. Exp. Phys.* **2022**, 083C01 (2022).
 - [3] M. Gorbahn and U. Haisch, Charm quark contribution to $K_L \rightarrow \mu^+ \mu^-$ at next-to-next-to-leading, *Phys. Rev. Lett.* **97**, 122002 (2006).
 - [4] B. R. Martin, E. De Rafael, and J. Smith, Neutral kaon decays into lepton pairs, *Phys. Rev. D* **2**, 179 (1970).
 - [5] G. D'Ambrosio, G. Isidori, and J. Portoles, Can we extract short distance information from $B(K_L \rightarrow \mu^+ \mu^-)$, *Phys. Lett. B* **423**, 385 (1998).
 - [6] M. Knecht, S. Peris, M. Perrottet, and E. de Rafael, Decay of pseudoscalars into lepton pairs and large N(c) QCD, *Phys. Rev. Lett.* **83**, 5230 (1999).
 - [7] G. Isidori and R. Unterdorfer, On the short distance constraints from $K_{L,S} \rightarrow \mu^+ \mu^-$, *J. High Energy Phys.* **01** (2004) 009.
 - [8] M. Hoferichter, B.-L. Hoid, and J. R. de Elvira, Improved standard-model prediction for $K_L \rightarrow \ell^+ \ell^-$, *J. High Energy Phys.* **04** (2023) 071.
 - [9] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, and C. Lehner, Lattice calculation of hadronic light-by-light contribution to the muon anomalous magnetic moment, *Phys. Rev. D* **93**, 014503 (2016).
 - [10] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, C. Lehner, and C. Tu (RBC, UKQCD Collaborations),

- Hadronic light-by-light contribution to the muon anomaly from lattice QCD with infinite volume QED at physical pion mass, *Phys. Rev. D* **111**, 014501 (2025).
- [11] N. Asmussen, E.-H. Chao, A. Gérardin, J. R. Green, R. J. Hudspith, H. B. Meyer, and A. Nyffeler, Hadronic light-by-light scattering contribution to the muon $g - 2$ from lattice QCD: Semi-analytical calculation of the QED kernel, *J. High Energy Phys.* **04** (2022) 040.
- [12] E.-H. Chao, R. J. Hudspith, A. Gérardin, J. R. Green, H. B. Meyer, and K. Ottnad, Hadronic light-by-light contribution to $(g - 2)_\mu$ from lattice QCD: A complete calculation, *Eur. Phys. J. C* **81**, 651 (2021).
- [13] E.-H. Chao, R. J. Hudspith, A. Gérardin, J. R. Green, and H. B. Meyer, The charm-quark contribution to light-by-light scattering in the muon $(g - 2)$ from lattice QCD, *Eur. Phys. J. C* **82**, 664 (2022).
- [14] N. H. Christ, X. Feng, L. Jin, C. Tu, and Y. Zhao, Calculating the two-photon contribution to $\pi^0 \rightarrow e^+ e^-$ decay amplitude, *Proc. Sci. LATTICE2019* (2020) 097 [arXiv:2001.05642].
- [15] N. H. Christ, X. Feng, L. Jin, C. Tu, and Y. Zhao, Lattice QCD calculation of the two-photon contributions to $K_L \rightarrow \mu^+ \mu^-$ and $\pi^0 \rightarrow e^+ e^-$ decays, *Proc. Sci. LATTICE2019* (2020) 128.
- [16] N. Christ, X. Feng, L. Jin, C. Tu, and Y. Zhao, Lattice QCD calculation of $\pi^0 \rightarrow e^+ e^-$ decay, *Phys. Rev. Lett.* **130**, 191901 (2023).
- [17] E.-H. Chao and N. Christ, Calculating the two-photon exchange contribution to $K_L \rightarrow \mu^+ \mu^-$ decay, *Phys. Rev. D* **110**, 054514 (2024).
- [18] Z. Bai, N. H. Christ, T. Izubuchi, C. T. Sachrajda, A. Soni, and J. Yu, $K_L - K_S$ mass difference from lattice QCD, *Phys. Rev. Lett.* **113**, 112003 (2014).
- [19] Y. Huo, N. H. Christ, and B. Wang, Enhanced lattice QCD studies on ε_K and ΔM_K , *Proc. Sci. LATTICE2024* (2025) 239.
- [20] N. H. Christ, X. Feng, G. Martinelli, and C. T. Sachrajda, Effects of finite volume on the $K_L - K_S$ mass difference, *Phys. Rev. D* **91**, 114510 (2015).
- [21] X.-Y. Tuo and X. Feng, Finite-volume formalism for physical processes with an electroweak loop integral, *Phys. Rev. D* **112**, 034512 (2025).
- [22] G. Isidori, G. Martinelli, and P. Turchetti, Rare kaon decays on the lattice, *Phys. Lett. B* **633**, 75 (2006).
- [23] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, Weak decays beyond leading logarithms, *Rev. Mod. Phys.* **68**, 1125 (1996).
- [24] C. Lehner and C. Sturm, Matching factors for Delta $S = 1$ four-quark operators in RI/SMOM schemes, *Phys. Rev. D* **84**, 014001 (2011).
- [25] T. Blum, P. A. Boyle, D. Horying, T. Izubuchi, L. Jin, C. Jung, C. Kelly, C. Lehner, A. Soni, and M. Tomii (RBC, UKQCD Collaborations), $\Delta I = 3/2$ and $\Delta I = 1/2$ channels of $K \rightarrow \pi\pi$ decay at the physical point with periodic boundary conditions, *Phys. Rev. D* **108**, 094517 (2023).
- [26] N. H. Christ, X. Feng, A. Portelli, and C. T. Sachrajda (RBC, UKQCD Collaborations), Prospects for a lattice computation of rare kaon decay amplitudes: $K \rightarrow \pi \ell^+ \ell^-$ decays, *Phys. Rev. D* **92**, 094512 (2015).
- [27] J. Tu, Lattice QCD simulations towards strong and weak coupling limits, Ph.D. thesis, Columbia U., 2020.
- [28] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, and C. Lehner, Hadronic light-by-light scattering contribution to the muon anomalous magnetic moment from lattice QCD, *Phys. Rev. Lett.* **124**, 132002 (2020).
- [29] G. McGlynn, Algorithmic improvements for weak coupling simulations of domain wall fermions, *Proc. Sci. LATTICE2015* (2016) 019.
- [30] H. Yin and R. D. Mawhinney, Improving DWF simulations: The force gradient integrator and the Möbius accelerated DWF solver, *Proc. Sci. LATTICE2011* (2011) 051 [arXiv:1111.5059].
- [31] E.-H. Chao, Progress on $K_L \rightarrow \mu^+ \mu^-$ from lattice QCD, in *12th International Workshop on the CKM Unitarity Triangle* (2024), arXiv:2403.18885.
- [32] C. W. Bernard, T. Draper, A. Soni, H. D. Politzer, and M. B. Wise, Application of chiral perturbation theory to $K \rightarrow 2\pi$ decays, *Phys. Rev. D* **32**, 2343 (1985).
- [33] P. A. Boyle, Conserved currents and results from $2 + 1f$ dynamical Möbius DWF simulations at the physical point, *Proc. Sci. LATTICE2014* (2015) 087.
- [34] B. Blossier, M. Della Morte, G. von Hippel, T. Mendes, and R. Sommer, On the generalized eigenvalue method for energies and matrix elements in lattice field theory, *J. High Energy Phys.* **04** (2009) 094.
- [35] D. Gomez Dumm and A. Pich, Long distance contributions to the $K_L \rightarrow \mu^+ \mu^-$ decay width, *Phys. Rev. Lett.* **80**, 4633 (1998).
- [36] G. Buchalla and A. J. Buras, The rare decays $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ and $K_L \rightarrow \mu^+ \mu^-$ beyond leading logarithms, *Nucl. Phys. B* **412**, 106 (1994).
- [37] X. Feng, L. Jin, and M. J. Riberdy, Lattice QCD calculation of the pion mass splitting, *Phys. Rev. Lett.* **128**, 052003 (2022).
- [38] G. Buchalla and A. J. Buras, Two loop large m_t electroweak corrections to $K \rightarrow \pi \nu \bar{\nu}$ for arbitrary Higgs boson mass, *Phys. Rev. D* **57**, 216 (1998).
- [39] I. Montvay and G. Münster, *Quantum fields on a lattice*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 1997).
- [40] C. Gattringer and C. B. Lang, *Quantum Chromodynamics on the Lattice*, Vol. 788 (Springer, Berlin, 2010).
- [41] P. A. Boyle, G. Cossu, A. Yamaguchi, and A. Portelli, Grid: A next generation data parallel C++ QCD library, *Proc. Sci. LATTICE2015* (2016) 023.
- [42] A. Yamaguchi, P. Boyle, G. Cossu, G. Filaci, C. Lehner, and A. Portelli, Grid: OneCode and FourAPIs, *Proc. Sci. LATTICE2021* (2022) 035 [arXiv:2203.06777].
- [43] C. Jung, Overview of Columbia Physics System (CPS), *Proc. Sci. LATTICE2013* (2014) 417.
- [44] R. G. Edwards and B. Joo (SciDAC, LHPC, and UKQCD Collaborations), The Chroma software system for lattice QCD, *Nucl. Phys. B, Proc. Suppl.* **140**, 832 (2005).
- [45] QDP++, <https://github.com/usqcd-software/qdpxx/tree/master>.
- [46] T. Blum *et al.* (RBC, UKQCD Collaborations), Domain wall QCD with physical quark masses, *Phys. Rev. D* **93**, 074505 (2016).
- [47] Y. Zhao, Lattice calculation of the $\pi^0 \rightarrow e^+ e^-$ and the $K_L \rightarrow \gamma\gamma$ decays, Ph.D. thesis, Columbia U., 2022.

- [48] J. Foley, K. Jimmy Juge, A. O’Cais, M. Peardon, S. M. Ryan, and J.-I. Skullerud, Practical all-to-all propagators for lattice QCD, *Comput. Phys. Commun.* **172**, 145 (2005).
- [49] L. Giusti, P. Hernandez, M. Laine, P. Weisz, and H. Wittig, Low-energy couplings of QCD from current correlators near the chiral limit, *J. High Energy Phys.* **04** (2004) 013.
- [50] J. Bijnens, P. Dhonte, and F. Borg, $K \rightarrow 3 \pi$ decays in chiral perturbation theory, *Nucl. Phys.* **B648**, 317 (2003).
- [51] S. Weinberg, New test for $\Delta I = \frac{1}{2}$ in K^+ decay, *Phys. Rev. Lett.* **4**, 87 (1960); *Phys. Rev. Lett.* **4**, 585(E) (1960).
- [52] J. Wess and B. Zumino, Consequences of anomalous Ward identities, *Phys. Lett.* **37B**, 95 (1971).
- [53] E. Witten, Global aspects of current algebra, *Nucl. Phys.* **B223**, 422 (1983).
- [54] M. J. Savage, M. E. Luke, and M. B. Wise, The rare decays $\pi^0 \rightarrow e^+e^-$, $\eta \rightarrow e^+e^-$ and $\eta \rightarrow \mu^+\mu^-$ in chiral perturbation theory, *Phys. Lett. B* **291**, 481 (1992).
- [55] S. Scherer and M. R. Schindler, *A Primer for Chiral Perturbation Theory*, Vol. 830 (Springer Berlin, Heidelberg, 2012).
- [56] M. Hoferichter, B.-L. Hoid, B. Kubis, and J. Ludtke, Improved standard-model prediction for $\pi^0 \rightarrow e^+e^-$, *Phys. Rev. Lett.* **128**, 172004 (2022).