Hadronic vacuum polarization: Comparing lattice QCD and data-driven results in systematically improvable ways

Michel Davier,[1] Zoltán Fodor,[2,3,4,5] Antoine Gérardin[●],[6] Laurent Lellouch[●],[6,*] Bogdan Malaescu[●],[7,†]
Finn M. Stokes[●],[5] Kálmán K. Szabó,[3,5] Balint C. Toth,[3] Lukas Varnhorst,[3] and Zhiqing Zhang[●][1]

[1]*IJCLab, Université Paris-Saclay et CNRS/IN2P3, F-91405 Orsay, France*
[2]*Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[3]*Department of Physics, University of Wuppertal, D-42119 Wuppertal, Germany*
[4]*Institute for Theoretical Physics, Eötvös University, H-1117 Budapest, Hungary*
[5]*Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52428 Jülich, Germany*
[6]*Aix Marseille Univ, Université de Toulon, CNRS, CPT, IPhU, Marseille, France*
[7]*LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, 75252, France*

The precision with which hadronic vacuum polarization (HVP) is obtained determines how accurately important observables, such as the muon anomalous magnetic moment a_μ or the low-energy running of the electromagnetic coupling α, are predicted. The two most precise approaches for determining HVP are dispersive relations combined with $e^+e^- \to$ hadrons cross section data and lattice QCD. However, the results obtained in these two approaches display significant tensions, whose origins are not understood. Here we present a framework that sheds light on this issue and—if the two approaches can be reconciled—allows them to be combined. Via this framework, we test the hypothesis that the tensions can be explained by modifying the R-ratio in different intervals of center-of-mass energy $\sqrt{s}$. As ingredients, we consider observables that have been precisely determined in both approaches. These are the leading hadronic contributions to a_μ, to the so-called intermediate window observable, and to the running of α between spacelike virtualities 1 and 10 GeV^2 (for which only a preliminary lattice result exists). Our tests take into account all uncertainties and correlations, as well as uncertainties on uncertainties in the lattice results. For instance, using this framework we show that results obtained in the two approaches can be made to agree, for all three observables, by modifying the ρ peak in the experimental spectrum. More specifically, we show that this requires a common ∼5% increase in the contributions of the peak to each of the three observables. This result is robust against the presence or absence of the running of α in the comparison. However, such an increase is much larger than the uncertainties on the measured R-ratio. We also discuss a variety of generalizations of the methods used here, as well as the limits in the information that can be extracted from the R-ratio via a finite set of observables.

DOI: 10.1103/PhysRevD.109.076019

I. INTRODUCTION

A virtual, propagating photon polarizes the vacuum into quarks and gluons. Known as hadronic vacuum polarization (HVP), this effect is important when processes involving electromagnetism at the quantum level are studied with high precision. For a photon of small virtuality, predicting this effect requires being able to describe the strong interaction in its nonperturbative regime. At present, there are two approaches for making precise predictions of this polarization. These are based, on the one hand, on large-scale numerical simulations in lattice quantum chromodynamics (QCD) and, on the other, on the exploitation of $e^+e^- \to$ hadron data.

HVP has been recently the center of much attention, because of its importance in the standard model prediction of the anomalous magnetic moment of the muon a_μ. This quantity is currently being measured at the Fermi National Accelerator Laboratory (Fermilab) [1] and was previously measured at the Brookhaven National Laboratory [2]. While the data-driven determination of the leading-order (LO) HVP contribution to a_μ [3–5], $a_\mu^{\text{LO-HVP}}$, yields a standard model prediction 4.2σ below the world average of

[*]laurent.lellouch@cnrs.fr; https://www.cpt.univ-mrs.fr/~lellouch
[†]malaescu@in2p3.fr; http://lpnhe.in2p3.fr/malaescu

the measurement of the total a_μ [1,2], the most precise lattice calculation of this contribution [6] reduces this difference to 1.5σ, with comparable uncertainties [7]. Recently, a new experimental measurement from Fermilab became available [8]. The combination of the earlier world average with this measurement, based on the data from runs 2 and 3, is larger than the theoretical prediction based on dispersive integrals from the Muon $g - 2$ Theory Initiative White Paper (WP) [5] by 5.1σ and 1.7σ above the prediction based on lattice QCD by the BMW Collaboration [6].

As discussed below and in Ref. [6], this reduction in the tension between prediction and experiment comes at the expense of being 2.0σ above the data-driven prediction. Moreover, for the so-called intermediate-distance-window contribution to $a_\mu^{\text{LO-HVP}}$ [9], this excess rises to 3.8σ ([6] and below) and even 4.2σ for a weighted average of independent lattice determinations of this quantity [6,10–13] (see Sec. IV A). A similar comparison was performed in Ref. [11], without including the intermediate-time-window results of Refs. [12,13] that appeared as Ref. [11] was being published. That work also includes the first lattice computation of the so-called short-distance window, as well as a comparison of that result with the one obtained in the data-driven approach [14].

The above discussion does not take into account the measurement of the $e^+e^- \to \pi^+\pi^-(\gamma)$ cross section, from threshold to 1.2 GeV, reported by the CMD-3 Collaboration in the preprint [15]. This is because an explanation has not yet been found for the fact that the cross section that they obtain is significantly larger than all previous, modern measurements and, in particular, the one published by the same collaboration in Ref. [16].

Another place where HVP plays an important role is in the scale dependence of the electromagnetic coupling α. The fact that the lattice predicts a larger value for $a_\mu^{\text{LO-HVP}}$ has an impact on the value of α at the scale of the Z-boson mass M_Z. This has been investigated in Refs. [17–20]. The overall conclusion is that the lattice excess in a_μ, alone, does not imply a change in $\alpha(M_Z^2)$ large enough to have an impact on precision electroweak tests. This is confirmed by direct lattice calculations of the running of α [6,21,22]. In particular, Ref. [6] suggests that the faster running of α observed in lattice calculations is concentrated for space-like values of momentum scale below a few GeV^2.

Beyond signaling tensions between the lattice and data-driven approaches for specific physical observables, the discrepancies discussed above also contain information about the agreement or disagreement between the primary quantities used to compute these HVP related observables in each of the two approaches. On the lattice this primary observable is the quark electromagnetic-current, two-point function with vanishing three-momentum, computed as a function of Euclidean time t. In the data-driven approach, it is instead the cross section for e^+e^- annihilation into hadrons, measured as a function of center-of-mass (c.m.) energy $\sqrt{s}$, normalized by the tree-level cross section for $e^+e^- \to \mu^+\mu^-$ in the massless limit, i.e., the R-ratio $R(s)$.[1]

To help pinpoint the possible sources of the disagreement between the two approaches, and possibly correct them, one needs to be able to make sharp statements about the agreement or disagreement between the lattice correlation function and the experimental R-ratio in different regions of t and $\sqrt{s}$. Because the current correlator is proportional to the Laplace transform of $\sqrt{s}R(s)$ [23], it is straightforward to determine that correlator, and any observable that can be obtained from it, once $R(s)$ is measured. Thus, it is relatively simple to identify regions of t in which the lattice and data-driven approaches disagree, possibly pointing to a problem with lattice computations at those length scales.

Determining the R-ratio in specific intervals of $\sqrt{s}$ from a lattice computation of the current correlator requires performing an inverse Laplace transform. Such lattice results will have uncertainties and be obtained at a finite number of points, making this inverse a notoriously ill-posed problem.

A number of interesting methods have been advocated for relating the two approaches. For instance, Ref. [24] proposes a modification of the Backus-Gilbert method that involves reconstructing, not the spectral function itself, but rather the spectral function convoluted, at individual values of $\sqrt{s}$, with a Gaussian whose narrowness is limited by the statistical and systematic uncertainties on the lattice results. The first application of this method was performed in Ref. [25]. In Ref. [26] the authors propose a physics- and data-constrained dispersive representation of the pion electromagnetic form factor. Then, in Ref. [27] they use it to study the consequences of a value of $a_\mu^{\text{LO-HVP}}$, such as the one obtained on the lattice in Ref. [6], on important observables impacted by this form factor. Another approach proposes to gain information about the R-ratio in different regions of $\sqrt{s}$, via well-chosen linear combinations of so-called time-window observables computed on the lattice [14]. Yet another approach proposes to use spectral-width sum rules to constrain $R(s)$, in narrow regions of $\sqrt{s}$, using the lattice correlator [28].

More generally, a variety of methods have been developed for extracting real-time properties (e.g., spectral functions) from Euclidean correlation functions computed in lattice QCD. These include approaches based on the Backus-Gilbert method with various regularizations [29–33]; smeared spectral functions using Chebyshev polynomials [34]; the maximum entropy method [35–41]; a Bayesian reconstruction

---

[1]The determination of $R(s)$ actually mixes experimental measurements and perturbative QCD predictions, depending on the $\sqrt{s}$ range, as discussed in Appendix A. In the following it will generally be referred to as "experimental R-ratio" and, more specifically, as "measured R-ratio" for the $\sqrt{s}$ regions where measurements are being used.

approach [42,43]; Bayesian, Generalized Randomized Hamiltonian Monte Carlo [44]; sum rules or other parametrizations [45–48], also combined with Bayesian approaches [49]; a modified lattice correlator which emphasizes different parts of the spectral function at large times [50]; recent developments based on special properties of certain conformal maps [51]; Gaussian processes [52]; and machine learning [53–63].[2] Comparisons of various approaches, together with their regularization, have also been performed [67–70].

Here we propose to begin more modestly in terms of R-ratio reconstruction. However, we do build a quantitative measure of comparison between the two approaches into our formalism and include it in our results.

To provide a relevant comparison between lattice and data-driven results, the lattice results used must have subpercent uncertainties. At such precisions, leading-order strong-isospin breaking and QED corrections are required. Moreover, if one is interested in probing the spectral function above the $c\bar{c}$ threshold, one must include the contributions of at least the four lightest quark flavors. For information above the $b\bar{b}$ threshold, a fifth quark flavor must be included. Of course, all continuum and infinite-volume limits must be taken in a controlled fashion. In addition, as the ranges of $\sqrt{s}$ that one is interested in reconstructing become narrower, the function with which the lattice results must be convoluted becomes more and more oscillatory. Thus, not only do the uncertainties have to be small, but also the statistical and systematic correlations between the lattice quantities have to be well known.

In the present paper we focus on the results published in Ref. [6], which satisfy all of the above requirements. These comprise $a_\mu^{\text{LO-HVP}}$ and the intermediate-time-window observable $a_{\mu,\text{win}}^{\text{LO-HVP}}$. We also call upon the preliminary result for the hadronic contribution to the running of α between spacelike virtualities 1 and 10 GeV^2, $\delta(\Delta\alpha_{\text{had}}^{(5)}) \equiv \Delta\alpha_{\text{had}}^{(5)}(-10 \text{ GeV}^2) - \Delta\alpha_{\text{had}}^{(5)}(-1 \text{ GeV}^2)$ [6], which provides interesting complementary information.

Here we use a limited form of the more general approach we propose, focusing on different intervals of $\sqrt{s}$ in the $e^+e^- \to$ hadron spectrum, such as the ρ-peak region, the low-mass region below that peak, different high-mass regions, etc. The idea is to test the extent to which the lattice results are consistent with a modification of the spectrum that leads to a common rescaling of the observables of interest in the chosen region and to determine the required rescaling factor. The simplest allowed modification would consist of directly rescaling the experimental R-ratio within the chosen region by that same factor. This would be a rather crude distortion of the experimental

spectral function that can be viewed as a first approximation to a more physical modification. In fact, because the constraints that we consider pertain only to integrals of the R-ratio, there remains a significant amount of freedom in the shape of the corresponding modification. Therefore, many other modifications are possible, some of which may actually be physical. This point is discussed in more detail in Appendix E 1.

Beyond complete and precise lattice results for the HVP quantities discussed above, our proposed reconstruction and comparison approach also requires state-of-the-art determinations of the contributions to those quantities from different $\sqrt{s}$ intervals in the data-driven approach, including a reliable quantification of correlations. For this we use the Davier-Hoecker-Malaescu-Zhang (DHMZ) methodology implemented in the HVPtools software [3,71–73].

Another feature of our analysis is that we include "uncertainties on the uncertainties" for the lattice results. That is, we determine statistical and systematic uncertainties on the covariance matrix between the different lattice observables considered. This is important, because reconstruction methods are very sensitive to uncertainties and correlations in the input data, a corollary of the ill-posed nature of inverse methods. The same should eventually be done for the correlations between the corresponding data-driven observables. This is left for future work.

The remainder of the paper is organized as follows. Section II briefly presents the lattice correlation function and how it can be obtained from the R-ratio. The observables that we use in the comparison of the lattice and data-driven approaches are defined in Sec. III, including the window observables put forward in Ref. [9]. In Sec. IV, we introduce the comparison methodology for testing the lattice correlation function with data-driven results. More importantly, we present our method for determining the size of the modification of the experimental R-ratio, in a given $\sqrt{s}$ interval, that would be required to reconcile the lattice and data-driven values of the observables of interest. Section V presents results of applications of these methods to the comparison of the two approaches. This is followed by our conclusions in Sec. VI, in which we also summarize the main results of our study.

In addition, we provide a number of appendixes. Appendix A presents the data-driven determination of the relevant observables and of their correlations, and Appendix B provides a summary of the lattice determination of these quantities, including uncertainties on the covariance matrix. In Appendix C we discuss, in more detail, the method used in this paper for determining the size of the possible modifications of the experimental R-ratio and its possible pitfalls. Appendix D is dedicated to showing the stability of this method with respect to the averaging procedure upon which it calls. Appendix E presents various generalizations of this method, which can provide increasingly refined reconstructions of the R-ratio required to accommodate lattice results, as more of

---

[2]Some of these methods have also been applied in the context of the reconstruction of parton distribution functions or of distribution amplitudes of hadrons, from lattice QCD [64–66].

these results become available and/or their precision increases. Finally, in Appendix F, some of the limitations on the number of observables that can be studied in the data-driven approach are discussed.

II. THE HVP CORRELATOR FROM THE LATTICE AND THE R-RATIO

Recent lattice calculations of leading-order HVP effects are based on the following Euclidean-time t, quark, electromagnetic-current two-point function with vanishing three-momentum and averaged over spatial components [23],

$$C(t) = \frac{1}{3e^2} \sum_{i=1}^{3} \int d^3x \langle J_i(\vec{x}, t) J_i(0) \rangle. \qquad (1)$$

Here, $e > 0$ is the unit of electric charge, $\frac{J_\mu}{e} = \frac{2}{3}\bar{u}\gamma_\mu u - \frac{1}{3}\bar{d}\gamma_\mu d - \frac{1}{3}\bar{s}\gamma_\mu s + \frac{2}{3}\bar{c}\gamma_\mu c - \frac{1}{3}\bar{b}\gamma_\mu b + \frac{2}{3}\bar{t}\gamma_\mu t$, and the angle brackets stand for the QCD + QED expectation value up to and including order e^2. For the quantities considered here, the top-quark contribution is either negligible or excluded by definition. However, all other flavors are required. Thus, the lattice results that we use [6] are obtained including the first five quark flavors.

The two-point function of Eq. (1) is not generically studied in the data-driven approach because quantities of phenomenological interest are usually obtained directly from the R-ratio. The latter is defined in terms of the cross section for e^+e^- annihilation into hadrons and is a function of the c.m. energy $\sqrt{s}$,

$$R(s) \equiv \frac{\sigma(e^+e^-(s) \to \text{hadrons}(+\gamma))}{4\pi\alpha^2(s)/(3s)}, \qquad (2)$$

where the denominator is the Born cross section for $e^+e^- \to \mu^+\mu^-$ in the massless limit. In that denominator, it is the electromagnetic coupling at c.m. energy squared s that is used to remove unwanted vacuum polarization effects in the numerator. In particular, in comparisons of $R(s)$ and $C(t)$, this means that it is the one-photon-irreducible part of the latter that is relevant. We denote it $C_{1\gamma I}(t)$. Moreover, the cross section in the numerator of Eq. (2) includes the final-state radiation of photons. While these photons imply that effects of higher order in α are included, they are kept to obtain an infrared-safe cross section at next-to-leading order (NLO) in α. Moreover, in applications, the contributions of these photons can consistently be taken into account in higher-order calculations.

From $R(s)$, it is straightforward to obtain $C_{1\gamma I}(t)$, after invoking the optical theorem that relates the $e^+e^- \to$ hadron cross section to the spectral function associated with $C_{1\gamma I}(t)$. The result is proportional to the Laplace transform of $\sqrt{s}R(s)$ [23],

$$C_{1\gamma I}(t) = \frac{1}{24\pi^2} \int_0^\infty ds \sqrt{s} R(s) e^{-|t|\sqrt{s}}. \qquad (3)$$

The integral diverges at short distances like t^{-3}, up to logarithms. However, in physical quantities, $C_{1\gamma I}(t)$ is multiplied by weights that vanish faster than t^3 as $t \to 0$.

III. OBSERVABLES OF INTEREST

In principle one could try to make a direct comparison of e.g., the function $t^4 C_{1\gamma I}(t)$,[3] computed on the lattice and in the data-driven approach [via Eq. (3)]. However, computing $t^4 C_{1\gamma I}(t)$ as a function of Euclidean time t, on the lattice, necessarily introduces a model dependence to account for the fact that, in each simulation, the correlator is obtained for different discrete values of t, separated by as much as a tenth of a femtometer. Moreover, $t^4 C_{1\gamma I}(t)$ is affected by different systematics in different t regions and the models have to be adapted to account for these differences. In addition, the statistical and systematic fluctuations in $t^4 C_{1\gamma I}(t)$ will be strongly correlated at neighboring times. These correlations not only restrict the usefulness of obtaining the function at multiple time points, but also complicate error propagation in derived quantities. Furthermore, they pose challenges in comparing it with its data-driven counterpart, which is also susceptible to such correlations.

Thus, to compare the two approaches, we focus here on quantities, derived from $C_{1\gamma I}(t)$, that are of direct phenomenological interest and that are likely to be computed by several lattice collaborations. The fact that our framework allows the use of such quantities, without necessitating the computation of $t^4 C_{1\gamma I}(t)$ as a function of t, is actually one of its main virtues.

The LO-HVP contribution to the anomalous magnetic moment of the muon can be obtained from $C_{1\gamma I}(t)$ via [23]

$$a_\mu^{\text{LO-HVP}} = \alpha^2 \int_0^\infty dt K(t) C_{1\gamma I}(t), \qquad (4)$$

where the kernel is

$$K(t) = \int_0^\infty \frac{dQ^2}{m_\mu^2} \omega\left(\frac{Q^2}{m_\mu^2}\right) \left[t^2 - \frac{4}{Q^2}\sin^2\left(\frac{t\sqrt{Q^2}}{2}\right)\right], \qquad (5)$$

$\omega(r) = [r + 2 - \sqrt{r(r+4)}]^2/\sqrt{r(r+4)}$ and α is the fine-structure constant in the Thomson limit. In fact, it is a discretized version of Eq. (4) that is the basis of the lattice calculation of $a_\mu^{\text{LO-HVP}}$.

When the data-driven approach is used, $a_\mu^{\text{LO-HVP}}$ is computed directly from $R(s)$ as an integral over s (see Appendix A). A comparison with lattice results gives a first means of confronting the two approaches. However, in the

---

[3]$C_{1\gamma I}(t)$ has a divergence proportional to $1/t^3$ at $t = 0$, whose coefficient is regularization-scheme dependent. To compare the correlation function obtained in the two approaches requires eliminating the divergence. That can be done, for instance, by considering $t^4 C_{1\gamma I}(t)$ instead of $C_{1\gamma I}(t)$ itself.

case of a disagreement, it is important to understand more precisely how different Euclidean times t, in $C_{1\gamma I}(t)$ computed on the lattice, contribute to this tension. Indeed, different length scales on the lattice are subject to different statistical and systematic uncertainties, and knowing which ones agree or disagree with the data-driven approach may help point to aspects of the lattice calculations that require more attention.

To shed light on this issue, particularly useful quantities are the so-called Euclidean-time-window contributions to $a_\mu^{\text{LO-HVP}}$ [9,14]. These are restrictions of the integral in Eq. (4) to time intervals of interest, using a smoothed-out Heaviside function that limits edge effects when dealing with discrete times, as is the case on the lattice. This function is taken to be $\theta(t; \Delta) \equiv \frac{1}{2}[1 + \tanh(t/\Delta)]$, where $\Delta > 0$ determines the time range over which the function transitions from 0 to 1. The desired, restricted integrals are then defined as

$$a_{\mu,\text{win}}^{\text{LO-HVP}}(t_i, t_f) = \alpha^2 \int_0^\infty dt\, W(t; t_i, t_f, \Delta) K(t) C_{1\gamma I}(t), \quad (6)$$

where the window function $W(t; t_i, t_f, \Delta)$ takes on different forms depending on the values of t_i and t_f,

$$
W(t; t_i, t_f, \Delta)
$$
$$
=
\begin{cases}
1, & \text{when } t_i = 0 \text{ and } t_f = \infty, \\
1 - \theta(t - t_f; \Delta), & \text{when } 0 = t_i < t_f < \infty, \\
\theta(t - t_i; \Delta) - \theta(t - t_f; \Delta), & \text{when } 0 < t_i < t_f < \infty, \\
\theta(t - t_i; \Delta), & \text{when } 0 < t_i < t_f = \infty,
\end{cases}
$$
$$(7)$$

where we take $\Delta = 0.15$ fm, as advocated in [9]. Note that $W(t; t_i, t_f, \Delta)$ is defined such that $a_{\mu,\text{win}}^{\text{LO-HVP}}(0, \infty) = a_\mu^{\text{LO-HVP}}$. Also, in the following, we call $a_{\mu,\text{win}}^{\text{LO-HVP}}$ the Euclidean-time-window contribution to $a_\mu^{\text{LO-HVP}}$ from the so-called "intermediate" time interval $[0.4, 1.0]$ fm, i.e., Eqs. (6) and (7) with $t_i = 0.4$ and $t_f = 1.0$ fm. Here we use a version of Eq. (5) in which the upper limit of the integral is $Q_{\text{max}}^2 = 3$ GeV^2 for reasons explained in Ref. [21].

These time-window observables are particularly useful because they are straightforward to compute in both the lattice and data-driven approaches.

On the lattice, their computation requires modifying the kernel in Eq. (4) used to compute $a_\mu^{\text{LO-HVP}}$ that is using an appropriate discretization of Eq. (6). Of course, the difficulty of the continuum and infinite-volume extrapolations, as well as the quality of the statistical signal, will depend acutely on the chosen window.

With the R-ratio, one simply substitutes Eq. (3) into Eq. (6), yielding

$$a_{\mu,\text{win}}^{\text{LO-HVP}}(t_i, t_f) = \left(\frac{\alpha m_\mu}{3\pi}\right)^2 \int_0^\infty \frac{ds}{s^2} \hat{K}(s; t_i, t_f) R(s), \quad (8)$$

with

$$\hat{K}(s; t_i, t_f) = \frac{3s^{5/2}}{8m_\mu^4} \int_0^\infty dt\, e^{-t\sqrt{s}} W(t; t_i, t_f, \Delta) K(t). \quad (9)$$

A first quantitative comparison of the two approaches can proceed through the comparison of individual windows, as was done in Refs. [6,14].[4]

In addition to time-window observables, one can further compare the lattice and data-driven approaches via the HVP function $\hat{\Pi}(Q^2)$ at different values of the spacelike momentum q, with $Q^2 = -q^2 \geq 0$ [21,22]. Alternatively, one can consider the difference of this function at two values of Q^2 to isolate an interval of interest. This has already be done in Ref. [6]. Instead of directly considering $\hat{\Pi}(Q^2)$, which is convention dependent, here we focus on the hadronic running of the electromagnetic coupling in the on-shell scheme, to which it is related. Thus, we work here with the hadronic running of the fine-structure constant in the five-flavor theory in the spacelike region, $\Delta\alpha_{\text{had}}^{(5)}(-Q^2)$. In terms of the lattice correlation function of Eq. (1), this quantity is given by

$$\Delta\alpha_{\text{had}}^{(5)}(-Q^2) = e^2 \int_0^\infty dt \left[t^2 - \frac{4}{Q^2}\sin^2\left(\frac{t\sqrt{Q^2}}{2}\right)\right] C_{1\gamma I}(t) \quad (10)$$

and, in terms of the R-ratio, by

$$\Delta\alpha_{\text{had}}^{(5)}(-Q^2) = \frac{\alpha Q^2}{3\pi} \int_0^\infty ds \frac{R(s)}{s(s + Q^2)}. \quad (11)$$

So as to simplify notations in the following, we denote quantities computed on the lattice a_j^{lat} and their counterparts computed in the data-driven approach a_j^{R}. Here the j indexes different moment integrals, such as the total $a_\mu^{\text{LO-HVP}}$ and window contributions to $a_\mu^{\text{LO-HVP}}$, as well as the running of α between two spacelike virtualities.

IV. COMPARISON METHODOLOGY

In the following two subsections, we present in more detail the methodologies that we will use to compare and test the lattice and data-driven approaches in Sec. V.

---

[4]An evaluation of the $a_{\mu,\text{win}}^{\text{LO-HVP}}$ integral based on hadronic τ decay spectral functions, corrected for isospin-breaking effects, has been performed in Ref. [74].

A. Testing the lattice with R-ratio results

Using the notation introduced in the last paragraph of the previous section, the first step is to compare the observables of interest, a_j^{lat} and a_j^{R}, one by one, and to determine their level of compatibility. Depending on how localized these observables are in Euclidean time or in c.m. energy, one can isolate distance, virtuality, or energy scales that may be problematic in each approach. Since a more effective way of isolating possibly problematic regions of the experimental R-ratio is presented below in Sec. IV B, the observables $a_j^{\text{lat/R}}$ that we focus on are chosen to emphasize features of the lattice correlator. Detailed results for the comparison of individual observables obtained from the lattice and the experimental R-ratio are given in Sec. V A.

To go beyond a comparison of individual observables, we propose a more global comparison of the two approaches that includes not just one observable at a time, but many. Thus, we define the following χ^2 function:

$$
\chi^2 = \sum_{j,k}[a_j^{\text{lat}} - a_j][C_{\text{lat}}^{-1}]_{jk}[a_k^{\text{lat}} - a_k]
$$
$$
+ \sum_{j,k}[a_j^{\text{R}} - a_j][C_{\text{R}}^{-1}]_{jk}[a_k^{\text{R}} - a_k], \qquad (12)
$$

where j, k run over the observables considered in the comparison and where the a_j are parameters. This function requires knowledge of the covariance matrices $C_{\text{lat/R}}$ between the observables obtained in each approach. Correlations between lattice and experimental R-ratio results are assumed to vanish, which is the case in the examples considered below. This more global measure of the compatibility of the two approaches is also important because it serves as a baseline for the improvements in the agreement brought by modifications of the experimental R-ratio considered in Sec. IV B.

For absolute uncertainties,[5] the minimum of the χ^2 of Eq. (12), as a function of the parameters a_i, is attained at

$$
a_i = \sum_{j,k}[(C_{\text{lat}}^{-1} + C_{\text{R}}^{-1})^{-1}]_{ij}[[C_{\text{lat}}^{-1}]_{jk}a_k^{\text{lat}} + [C_{\text{R}}^{-1}]_{jk}a_k^{\text{R}}], \quad (13)
$$

and takes the minimum value

$$
\chi^2_{\text{min}} = \sum_{j,k}[a_j^{\text{lat}} - a_j^{\text{R}}][(C_{\text{lat}} + C_{\text{R}})^{-1}]_{jk}[a_k^{\text{lat}} - a_k^{\text{R}}]. \quad (14)
$$

We take the p-value associated with the minimum value of χ^2 and the number of degrees of freedom (*d.o.f.*) to be a measure of the overall agreement of the two approaches. Here, the *d.o.f.* is simply the number of observables

considered in the comparison. This measure of agreement makes sense only if $C_{\text{lat/R}}$ are well known, because the latter determine how much independent information is available in the lattice and R-ratio inputs and because the p-value obtained will depend strongly on them. This p-value may dilute some of the tensions observed in individual observables. However, in the presence of strong correlations, its value will better reflect the significance of having many observables disagree. Detailed results for this combined comparison of lattice and experimental R-ratio observables, via Eq. (14), are given in Sec. V A.

Note that if the p-value corresponding to this minimum is acceptable, then the values of a_j at the minimum of χ^2 correspond to a weighted average of the experimental R-ratio and lattice integrals. However, this averaging constrains the shape of the underlying spectral functions only in a very limited fashion. These aspects are discussed in more detail in Appendix E.

B. Testing the experimental R-ratio with lattice results: Rescaling

While fully reconstructing the R-ratio from the HVP correlator computed on the lattice is an ill-posed problem, interesting information about $R(s)$ can still be derived.

The method proposed here is applicable to any observable a_j related to the HVP. It is meant to help isolate c.m. energy regions in the experimental R-ratio that may be responsible for tensions with the chosen lattice observables. Thus, we split up the observables obtained from the R-ratio, via e.g., Eqs. (8) and (11), into contributions from different $\sqrt{s}$ intervals I_b,

$$
a_j^{\text{R}} = \sum_b a_{jb}^{\text{R}}. \qquad (15)
$$

Here the sum over b (i.e., over all of the intervals) covers the full support of $R(s)$. In the case of $a_\mu^{\text{LO-HVP}}$, window observables, and the running of α between two spacelike momenta, the a_{jb}^{R} are obtained by restricting the integrals over s in Eqs. (8) and (11), respectively, to the interval I_b.

When computing the Euclidean-time-window observables, the edges of the corresponding intervals are smoothed out to account for the fact that, at finite lattice spacing, the density of points in $C_{1\gamma I}(t)$ is not that large. This is not necessary for $R(s)$, because the density of measured points in s is significant, allowing us to consider sharp intervals I_b, with an interpolation between the two points bracketing each boundary of the interval.

As discussed in the Introduction, in this paper we investigate whether a tension between lattice and data-driven results can be explained by a change in the experimental R-ratio that is consistent with a rescaling of the contributions, from one or more intervals in $\sqrt{s}$ to the observables of interest. Thus, we wish to solve the following system of equations for γ_b:

---

[5]This can be generalized for also treating relative uncertainties that are scaled with the fitted quantities, e.g., by explicitly including the corresponding dependencies on the fitted quantities in the χ^2 definition and/or using an iterative procedure [75,76] (see also Sec. IV B and Appendix C).

$$a_j^{\text{lat}} = \sum_b \gamma_b a_{jb}^{\text{R}}, \qquad (16)$$

taking into account all uncertainties and correlations among the lattice results, on the one hand, and the experimental R-ratio ones, on the other. If a given $\delta_b \equiv (\gamma_b - 1)$ is set to zero *a priori*, the R-ratio integrals remain unchanged in the corresponding interval.

Here we consider the special case where a subset of the $\sqrt{s}$ intervals A give contributions to the observables that are rescaled by a common factor γ while, for the complementary subset of intervals B, the contributions are kept unchanged.[6] In such a case, Eq. (16) becomes

$$a_j^{\text{lat}} = \sum_{b \in A} \gamma a_{jb}^{\text{R}} + \sum_{b \in B} a_{jb}^{\text{R}}, \qquad (17)$$

which implies

$$\gamma = \frac{a_j^{\text{lat}} - \sum_{b \in B} a_{jb}^{\text{R}}}{\sum_{b \in A} a_{jb}^{\text{R}}} \equiv \tilde{\gamma}_j. \qquad (18)$$

With more than one independent observable a_j, this becomes an overconstrained system, typically characterized by a χ^2 function,

$$\chi^2 = \sum_{j,k} [\gamma - \tilde{\gamma}_j][(C_{\text{lat}}^{\tilde{\gamma}} + C_{\text{R}}^{\tilde{\gamma}})^{-1}]_{jk}[\gamma - \tilde{\gamma}_k], \qquad (19)$$

where $C_{\text{lat}}^{\tilde{\gamma}}$ and $C_{\text{R}}^{\tilde{\gamma}}$ are the covariance matrices of the $\tilde{\gamma}_j$ rescaling coefficients, originating from the lattice and the dispersive uncertainties, respectively. They are obtained through a linear propagation of uncertainties, from the lattice QCD and experimental R-ratio a_j integrals through Eq. (18).

The rescaling factor γ optimizing the constraints from Eq. (18) can be determined by minimizing the χ^2 function from Eq. (19) with respect to γ. This minimization yields an average of the input $\tilde{\gamma}_j$ values, with weights given by the inverse of the sum of covariance matrices that appears in Eq. (19). Although very commonly used, this approach can yield biased results (see e.g., Ref. [75] and the discussion in Appendix C). In particular, such biases are caused by the uncertainties on the uncertainties and on their correlations, present both for the experimental measurements used in the dispersive approaches and for the lattice QCD results. In order to avoid these undesirable effects, we employ the weighted average of Eq. (C1), in which we set the off-diagonal elements of $C_{\text{lat}}^{\tilde{\gamma}}$ and $C_{\text{R}}^{\tilde{\gamma}}$ to zero, followed by a full propagation of the uncertainties with their correlations (see Appendix C for a detailed discussion).

We also consider a method based on a generalization of the χ^2 of Eq. (12), instead of the weighted averages

discussed above. This method also allows us to study similar modifications of the experimental R-ratio. It is discussed in Appendix E 3.

C. Possible extensions of the methodology

In Appendix E, we consider comparisons and tests of the lattice and data-driven approaches that go beyond the rescaling of the experimental R-ratio integrals in a $\sqrt{s}$ interval, by a single, common parameter γ. Moreover, in Appendix F, we consider the possible advantages and limitations associated with the use of additional observables in the data-driven approach.

V. COMPARISON RESULTS

A. Testing the lattice with R-ratio results

As a first step, we compare individually $a_\mu^{\text{LO-HVP}}$, the intermediate-time-window observable $a_{\mu,\text{win}}^{\text{LO-HVP}}$, and $\delta(\Delta\alpha_{\text{had}}^{(5)})$. The kernels for these quantities are shown as functions of t and $\sqrt{s}$ in Fig. 1. As is well known, the kernels for $a_\mu^{\text{LO-HVP}}$ are highly peaked for small values of $\sqrt{s}$ and large values of t. While the Euclidean-time kernel for $a_{\mu,\text{win}}^{\text{LO-HVP}}$ is strongly localized in t in the interval ranging approximately from 0.35 to 1.51 fm,[7] its c.m. counterpart emphasizes a $\sqrt{s}$ region from threshold to around 3.2 GeV.[8] Finally, the kernels for $\delta(\Delta\alpha_{\text{had}}^{(5)})$ extend over all t, with an oscillation of period $\sim$1 fm, and over a large range of $\sqrt{s}$, going from threshold to well beyond 5 GeV. The upshot is that the observables which we consider here probe the behavior of the lattice correlator $C_{1\gamma I}(t)$ at very different distance scales, while the c.m. energies at which the R-ratio is probed have more overlap. Thus, the comparison of these observables tells us more about specific features of the lattice correlator than they do about those of the timelike spectrum.

The lattice [6] and data-driven results for the observables of interest are shown in Table I, together with their pairwise differences. While the preliminary lattice [6] and the data-driven results for $\delta(\Delta\alpha_{\text{had}}^{(5)})$ are compatible within 1.4σ,[9] there is a 2.0σ tension between the results for

---

[6] In practice, for the cases we consider here, these subsets contain one or two $\sqrt{s}$ intervals.

[7] This interval is the range of t values within which the kernel for $a_{\mu,\text{win}}^{\text{LO-HVP}}$ takes on values greater than 1% of its maximum value. The same exercise on the window function of Eq. (7), with $t_i = 0.4$ and $t_f = 1$ fm, gives the interval [0.05,1.35] fm. Thus, the region of the lattice correlator that is probed via $a_{\mu,\text{win}}^{\text{LO-HVP}}$ is shifted upward relative to the one given by the window function alone.

[8] This upper limit is the value of $\sqrt{s}$ below which the kernel takes on values larger than 1% of its maximum value.

[9] This compatibility is less good if the lattice results of Ref. [22] are considered. Indeed, using those results and their correlations we obtain $\delta(\Delta\alpha_{\text{had}}^{(5)}) = 49.28(0.45) \times 10^{-4}$, which is 2.6(1.2)% or 2.3σ larger than the data-driven result presented in Table I.
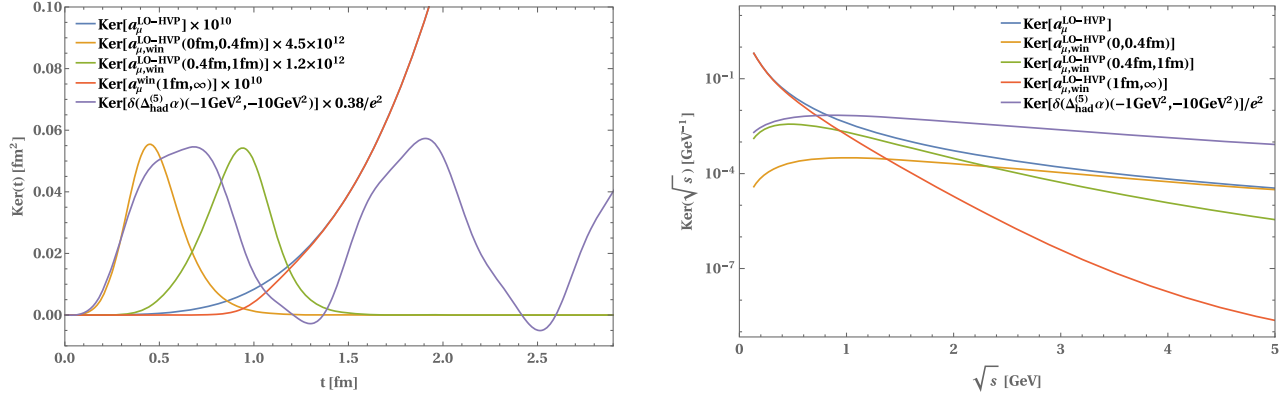
FIG. 1.   Kernels for $a_\mu^{\text{LO-HVP}}$, $a_{\mu,\text{win}}^{\text{LO-HVP}}$, and $\delta(\Delta\alpha_{\text{had}}^{(5)})$ as a function of Euclidean time (left) and of c.m. energy $\sqrt{s}$ (right). In addition, to clarify the arguments laid out later in the section, we have added the kernels of the short-distance window [Eq. (7) with $t_i = 0$ and $t_f = 0.4$ fm] and the long-distance one as well [Eq. (7) with $t_i = 1$ fm and $t_f = \infty$]. Some of the kernels are rescaled for better visibility. $a_{\mu,\text{win}}^{\text{LO-HVP}}$, the short-distance and long-distance windows are a contribution to $a_\mu^{\text{LO-HVP}}$ so that their kernels (green, orange, and red curves, respectively) are contributions to $a_\mu^{\text{LO-HVP}}$'s kernel (blue curve).

$a_\mu^{\text{LO-HVP}}$. For $a_{\mu,\text{win}}^{\text{LO-HVP}}$, the tension is much more significant, at 3.8σ. This number rises to 4.2σ when the lattice result of Ref. [6] is replaced by a weighted average of independent lattice determinations of this quantity [6,10–13]. All of these points are illustrated in Fig. 2. This large discrepancy points to a significant inconsistency between the approaches that must be resolved before giving a standard model prediction for $a_\mu^{\text{LO-HVP}}$.

Because the window observable probes intermediate Euclidean-time distances, its computation is particularly well suited to the lattice approach. The observable is significantly less noisy statistically than $a_\mu^{\text{LO-HVP}}$, less sensitive to finite-volume effects, and less sensitive to short-distance discretization effects, making its computation in lattice QCD particularly reliable.

In principle, the disagreement with the data-driven approach could be due to a common feature in the lattice calculations of Fig. 2 that would lead to an overlooked systematic error. For instance, one may worry that all the calculations are based on the time-momentum representation (TMR) of the current-current correlator given in Eq. (1). However, Ref. [83] explicitly checked that the connected contributions of the ud and s quarks to the window observable, obtained using a Lorentz-covariant coordinate-space representation, agree with those computed in the TMR to the level of 0.7%, in simulations with pion and kaon masses of ~350 and ~450 MeV, respectively. Alternatively, one may question the universality of the continuum limit. However, as discussed in the caption of Fig. 2, the lattice results for $[a_{\mu,\text{win}}^{\text{LO-HVP}}]_{\text{iso}}^{ud}$ [6,10–13], included in the average, are obtained with different discretizations, thus testing this universality. Moreover, two of these analyses were blinded [12,13], strengthening this test. There are four additional calculations of the leading u and d quark

TABLE I.   Comparison of lattice and data-driven results for the three main observables of interest in this paper. Each row contains the results and comparisons for the observable described in the first column. The second column provides the corresponding lattice result from Ref. [6] and the third, the data-driven one computed here, using the methods of Ref. [3]. The fourth column displays the absolute difference between the two results, the fifth their relative difference, the sixth their difference in units of combined standard deviations, and the last column, the corresponding p-value. The values in parentheses correspond to the total uncertainties of the corresponding quantities. Note that the powers of 10 in the observable column only apply to columns 2–4. The three remaining columns have units specified in the corresponding column label.

Observable	Lattice [6]	Data driven	Diff.	% Diff.	σ	p-value (%)
$a_\mu^{\text{LO-HVP}} \times 10^{10}$	707.5(5.5)	694.0(4.0)	13.5(6.8)	1.9(1.0)	2.0	4.7
$a_{\mu,\text{win}}^{\text{LO-HVP}} \times 10^{10}$	236.7(1.4)	229.2(1.4)	7.5(2.0)	3.2(0.8)	3.8	0.01
$[\Delta\alpha_{\text{had}}^{(5)}(-10 \text{ GeV}^2) - \Delta\alpha_{\text{had}}^{(5)}(-1 \text{ GeV}^2)] \times 10^4$	48.67(0.32)[a]	48.02(0.32)	0.65(0.45)	1.3(0.9)	1.4	15.

[a]This result's continuum limit does not include the logarithmically enhanced discretization uncertainties discovered subsequently in Ref. [82], nor was this quantity the focus of Ref. [6]. However, we include it in the present study to illustrate how using a quantity which is complementary to $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ can provide important information.
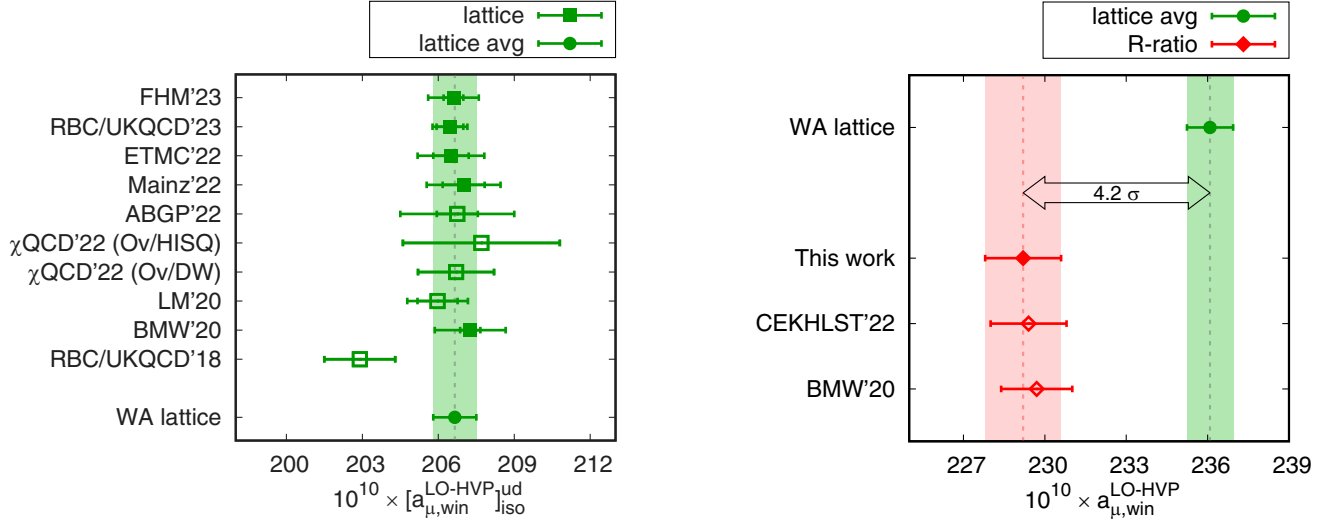
FIG. 2. Comparison of lattice and data-driven results for the contribution $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$ to $a_{\mu}^{\mathrm{LO\text{-}HVP}}$, from the Euclidean time interval [0.4,1.0] fm [see Eqs. (6) and (7) and subsequent text]. Left: Comparison of lattice results for the isospin-symmetric, u and d connected contribution to the intermediate-window observable, noted $[a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}]_{\mathrm{iso}}^{ud}$ (green squares). For each group, only the most recent results are shown. BMW'20 [6], LM'20 [77], ABGP'22 [78], and FHM'23 [13] are obtained with different varieties of staggered fermions; Mainz'22 [10] with $O(a)$-improved-Wilson, ETMC'22 [11] with twisted-mass, and RBC/UKQCD'23 [12] with domain-wall fermions; χQCD'22 [79] with overlap valence quarks on either HISQ or domain-wall configurations. The filled squares correspond to fully independent results, while the open ones are obtained using subsets of configurations from other calculations. LM'20 relies on a subset of the configurations used in FHM'23 and χQCD'22 on subsets of those used in FHM'23 and RBC/UKQCD'23. The green band corresponds to a weighted average of the fully independent (filled squares) lattice results for $[a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}]_{\mathrm{iso}}^{ud}$. Following the procedure discussed in Appendix C, the mean is performed without correlations in the determination of the weights, while the uncertainty propagation and χ^2 evaluation take them into account. These are obtained by assuming 100% correlation between the total systematic errors of the different calculations. In this way we find $[a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}]_{\mathrm{iso}}^{ud} = (206.6 \pm 0.9) \times 10^{-10}$ with a correlated $\chi^2/d.o.f. = 1.7/4$. Note that a phenomenological estimate of this quantity was reported in Refs. [80,81]. Right panel: Comparison of the world average (WA) of lattice results (green filled circle and band) with a number of R-ratio results for the intermediate-window observable $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$ (red diamonds), including the one determined in this paper (filled red diamond and band) and by CEKHLST'22 [14]. The WA for $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$ is obtained from that for $[a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}]_{\mathrm{iso}}^{ud}$, by adding all other quark, QED, and strong-isospin-breaking contributions from Ref. [6]. The resulting average is $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}} = (236.1 \pm 0.9) \times 10^{-10}$. The difference of this result with the data-driven one of the present work is 4.2σ, shown as a horizontal arrow.

contributions [9,77–79]. All nine calculations agree, as shown in the left panel of Fig. 2.

From the perspective of the R-ratio as a function of $\sqrt{s}$, the range covered by the intermediate window is more spread out. Though it is not exclusively emphasized, this region includes the ρ peak, in which some disagreement between the different measurements is apparent [3,15]. However, the difference between the KLOE [84–86] and the BABAR [87] measurements of the $e^+e^- \to \pi^+\pi^-(\gamma)$ cross section is taken into account as an additional systematic uncertainty, as described in Ref. [3]. So, unless a large, unknown systematic uncertainty is present in the measurements of the $e^+e^- \to$ hadrons cross section [84–87], the strong tension with the lattice calculation of the intermediate window cannot be explained within the data-driven approach. The very recent measurement of the $e^+e^- \to \pi^+\pi^-(\gamma)$ cross section by the CMD-3 Collaboration [15] helps bring the data-driven results for

$a_{\mu}^{\mathrm{LO\text{-}HVP}}$ and $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$ in line with the lattice results for those quantities. However, it does so at the expense of a dramatic tension with previous measurements that is not yet understood.

As a preamble to the following discussion about the consequences, for the lattice correlator, of the tensions and agreement in the values of $a_{\mu}^{\mathrm{LO\text{-}HVP}}$, $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$, and $\delta(\Delta\alpha_{\mathrm{had}}^{(5)})$ computed in the two approaches, it is important to remember that the current correlator is a smooth, monotonically decreasing function of t that behaves as t^{-3} for small t, up to logarithms, and falls off exponentially with exponent $2M_\pi t$ at large t in infinite volume.

While the significance of the difference between the lattice and data-driven determinations of $a_{\mu}^{\mathrm{LO\text{-}HVP}}$ is smaller than in $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$, the absolute difference itself is about twice as large. Thus, about half the discrepancy

in $a_\mu^{\text{LO-HVP}}$ must come from the value of the lattice correlator for t below ~ 0.35 fm and/or above ~ 1.51 fm. An important caveat is that the uncertainty on this second half of the discrepancy is around 100% (see Table I), making this contribution to the discrepancy not significantly different from 0.

Combining this observation with the results for $\delta(\Delta\alpha_{\text{had}}^{(5)})$, which agree in the two approaches, is more complicated. Indeed, the kernel of this observable is oscillatory in t and has support at all times. However, the kernel of $\delta(\Delta\alpha_{\text{had}}^{(5)})$ also has significant overlap with the one of $a_{\mu,\text{win}}^{\text{LO-HVP}}$. This suggests that $\delta(\Delta\alpha_{\text{had}}^{(5)})$ will receive an enhancement similar to the one of $a_{\mu,\text{win}}^{\text{LO-HVP}}$, from the region delimited by the kernel of the window observable. Of course, that will not be true if all of the enhancement in $a_{\mu,\text{win}}^{\text{LO-HVP}}$ comes from times $1.2 \lesssim t \lesssim 1.5$ fm, where the running-of-α kernel is suppressed. Assuming, for the moment, that $\delta(\Delta\alpha_{\text{had}}^{(5)})$ is enhanced in the window, agreement on that quantity in the two approaches then requires that the lattice correlator be suppressed for some values of t outside that window.

This discussion leads to a situation where $a_\mu^{\text{LO-HVP}}$ suggests that the lattice correlator is enhanced, compared to the data-driven one, for t below ~ 0.35 fm and/or above ~ 1.5 fm, while $\delta(\Delta\alpha_{\text{had}}^{(5)})$ suggests the opposite. To try to reconcile the two statements, it is useful to look at contributions to each of the two observables from the short-distance (SD) window [defined via the window function of Eq. (7), with $t_i = 0$ and $t_f = 0.4$ fm], the intermediate-distance (ID) one (denoted simply as window in this paper and with $t_i = 0.4$ and $t_f = 1$ fm), and the long-distance (LD) one (with $t = 1$ fm to ∞). Using the data from Ref. [88] for a rough estimate of the ratio of the contributions of the SD:ID:LD windows, we find 10%:33%:57% for $a_\mu^{\text{LO-HVP}}$ and 70%:29%:1% for $\delta(\Delta\alpha_{\text{had}}^{(5)})$. Of course $a_{\mu,\text{win}}^{\text{LO-HVP}}$ receives a 100% contribution from the ID window.

Thus, if the excess in $\delta(\Delta\alpha_{\text{had}}^{(5)})$ from the ID window were to be compensated by a suppression of the correlator in the LD window alone, which contributes only 1%, that suppression would have to be significant. However, such a significant suppression would also reduce $a_\mu^{\text{LO-HVP}}$, since the contribution of the LD window to that observable is dominant. To compensate that suppression would require a significant enhancement of the lattice correlator in the SD window because that window only contributes at the level of 10% to $a_\mu^{\text{LO-HVP}}$. But such an SD enhancement would make the lattice determination of $\delta(\Delta\alpha_{\text{had}}^{(5)})$ significantly larger than that from the experimental R-ratio, given that the SD window is responsible for about 70% of the value of that observable. This discussion suggests that the SD window

is not responsible for the additional enhancement in the lattice computation of $a_\mu^{\text{LO-HVP}}$ over that of $a_{\mu,\text{win}}^{\text{LO-HVP}}$ in the comparison with the data-driven results. It also suggests that an LD suppression of the lattice correlator cannot be responsible for mitigating a possible enhancement due to the ID-window contribution to $\delta(\Delta\alpha_{\text{had}}^{(5)})$.

To summarize, we know for sure that the lattice correlator is enhanced in the ID-window region compared to the one obtained from the data-driven approach. In addition, the arguments made above suggest that the lattice correlator may be slightly suppressed in the part of the SD window that has overlap with the $\delta(\Delta\alpha_{\text{had}}^{(5)})$ kernel and enhanced in the LD one. Note that a suppression of the correlator in the full SD window cannot be significant because that would contradict the agreement of the lattice [11] and the data-driven [14] SD-window results.[10]

We now consider combined comparisons of the lattice and experimental R-ratio results for the above observables via the minimization of the χ^2 function defined in Eq. (12). Beyond giving a more global measure of the compatibility of the two approaches, the resulting p-values also provide baselines against which to compare those obtained when we study possible rescalings of integrals of the R-ratio in chosen $\sqrt{s}$ intervals in the following section. These comparisons require knowledge of the covariance matrices, among the observables considered, in each of the lattice and data-driven approaches. While the correlations between $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$, determined in lattice QCD, are obtained as described in Appendix B, the determination of the running of α is impacted by uncertainties that are, to good approximation, independent of those of the former. On the other hand, for the dispersive approach, the correlations among the uncertainties of the contributions to all moment integrals are derived as discussed in Appendix A.

We first consider a combined comparison of $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. The resulting $\chi^2/d.o.f.$ range from $14.4_{-2.1}^{+3.0}/2$ to $18.8_{-1.7}^{+2.0}/2$ when the two most extreme of the four systematic variations of the lattice covariance matrices, given in Appendix B,[11] are considered. The central values correspond to the nominal lattice sample. The uncertainties are statistical, obtained from the $\pm 1\sigma$ quantile deviations from the median of the bootstrap

---

[10]The RBC/UKQCD Collaboration also computed the SD window on the lattice, but only the connected, light-quark contribution to it, which means that their result cannot be compared directly to the data-driven result of Ref. [14]. Nevertheless, their result for this contribution is in full agreement with that of Ref. [11].

[11]These two results correspond to the covariance matrices labeled 2 and 3, respectively.

distributions for the lattice covariance matrices.[12] Half the difference between the central values, i.e., $4.4/2 = 2.2$, can be viewed as the lattice systematic uncertainty on these results. It is of comparable size to the statistical one. Unlike what has been just discussed for the lattice results, statistical and systematic uncertainties on the uncertainties associated with the data-driven approach are not yet included here, but should be in the future.

The corresponding p-values are $7^{+10}_{-6} \times 10^{-4}$ and $8^{+13}_{-6} \times 10^{-5}$, respectively. These are very small and significantly smaller than the p-value for $a_\mu^{\text{LO-HVP}}$ alone in Table I. Nevertheless, the first of the above p-values can be as much as 17 times larger than the one in given in Table I for $a_{\mu,\text{win}}^{\text{LO-HVP}}$ and the second as much as a factor of 5 smaller, within 1 standard deviation. In any case, the probability that the lattice and data-driven results for $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ agree simultaneously is very small.

The disagreement reduces to some extent when $\delta(\Delta\alpha_{\text{had}}^{(5)})$ is added to the mix. In the lattice calculation, which is still preliminary, this quantity is negligibly correlated with $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. While here the χ^2 remain essentially unchanged, ranging from $14.4^{+3.0}_{-2.1}$ to $18.8^{+2.0}_{-1.7}$, the d.o.f. increases from 2 to 3, leading to improved p-values that range from $2.3^{+4.0}_{-1.9} \times 10^{-3}$ and $3.0^{+4.2}_{-2.1} \times 10^{-4}$. That is, the inclusion of $\delta(\Delta\alpha_{\text{had}}^{(5)})$ and, to a lesser extent, $a_\mu^{\text{LO-HVP}}$ is observed to dilute the disagreement between the lattice and data-driven approaches for $a_{\mu,\text{win}}^{\text{LO-HVP}}$. This dilution is expected because the disagreement between the two approaches is much smaller for both $\delta(\Delta\alpha_{\text{had}}^{(5)})$ and $a_\mu^{\text{LO-HVP}}$. However, in the presence of strong correlations, this combined measure will better reflect the significance of having agreement or disagreement among many observables.

Very similar results are obtained in the approach of Sec. IV B, when the γ rescaling coefficient in Eq. (16) is set to unity. This is not surprising because, with $\gamma = 1$, Eqs. (14) and (19) indicate that the χ^2 and d.o.f. are the same in the two approaches, up to possible nonlinearities in the propagation of uncertainties. In particular, the minimal/maximal $\chi^2/d.o.f.$ values are stable, within 0.2 or less, with respect to the choice of the split in various $\sqrt{s}$ regions. When sampling the $\chi^2/d.o.f.$ values for the various bootstrap replicas of the lattice covariance matrix, with either two or three moment integrals, the median of the resulting distribution is close to the nominal values

(within 0.15 or less), while the variance and $\pm 1\sigma$ quantiles of the distribution are smaller than 1.5 and the $\pm 2\sigma$ quantiles are smaller than 3.0 (similar to the uncertainty values given earlier in this section). It is hence observed that the level of the tension between the dispersive- and lattice-based moment integrals is relatively stable with respect to the choice of the split in various $\sqrt{s}$ regions and/or statistical and systematic variations of the lattice covariance matrix.

Of course, other quantities related to hadronic vacuum polarization that are computed in the lattice and data-driven approaches can also be compared individually or added with correlations to the χ^2 to further sharpen the comparison of the two approaches.

B. Testing the experimental R-ratio with lattice results: Rescaling

In this section we present the results obtained when comparing the lattice QCD and dispersive results, employing the methodology discussed in Sec. IV B and Appendix C. With this methodology, we determine the size of modifications to the experimental R-ratio that lead to a rescaling of the contributions, from a given $\sqrt{s}$ interval to the observables of interest. As in the previous subsection, the study is performed for either two moment integrals ($a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$) or three of them [i.e., including, in addition, $\delta(\Delta\alpha_{\text{had}}^{(5)})$]. Results are given for various scenarios, i.e., $\sqrt{s}$ regions where the dispersive integrals are computed and the rescaling factors are determined. These include $\sqrt{s}$ regions below or above 0.63, 0.96, 1.1, 1.8, and 3.0 GeV, respectively, as well as rescaling inside or outside the [0.63, 0.92] GeV interval. While the first set of intervals covers various low- and high-mass regions, with different contributions to the dispersive integrals and different methodologies used for deriving them, the last one covers the region dominated by the ρ-resonance peak and the region complementary to it. The intervals are displayed in the left panel of Fig. 3. All uncertainties and correlations are accounted for, as discussed in the previous subsection. There we needed those associated with the moment integrals of the R-ratio over the full range of $\sqrt{s}$. Here we need those corresponding to the contributions to moment integrals from the various $\sqrt{s}$ intervals, whose determinations are also described in Appendix A.

Table II and Fig. 3 also present the results for the rescaling factor, $\gamma_1 \equiv 1 + \delta_1$, for the various $\sqrt{s}$ regions I_1 discussed in the previous paragraph, using either two or three moment integrals. The results are determined via the averages defined by the χ^2 function of Eq. (19), in which γ is replaced by γ_1 and with the weights of the input $\tilde{\gamma}_j$ values obtained by setting all correlations to zero, for reasons discussed in Appendix D. Out of the four systematic uncertainty variations of the lattice covariance matrices discussed in Appendix B 2, results are presented here for

---

[12]As discussed in Appendix B, it is important to note that the statistical covariance of the lattice data is determined from only 48 binned samples, which is smaller than the number of independent samples in the full dataset. However, as seen below, the statistical uncertainties on the lattice covariance matrix that arise due to this conservative binning are sufficiently small for the purposes of the present study. In future work, we will aim to improve this evaluation.
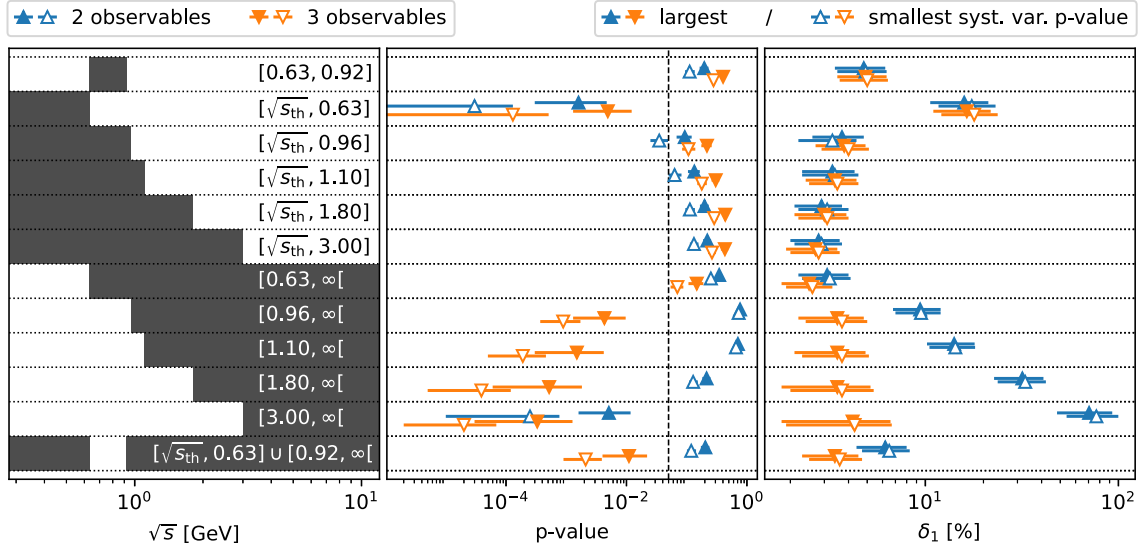
FIG. 3.　Plot illustrating Table II. The first panel displays, in dark gray, the various $\sqrt{s}$ intervals considered here. The first row corresponds to the ρ peak. In the second panel, we plot the p-values that indicate the compatibility of the rescaling hypothesis, in the given interval, with the lattice and data-driven results for $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ (two observables, blue triangles) and with the results for the additional observable $\delta(\Delta\alpha_{\text{had}}^{(5)})$ (three observables, orange upside-down triangles). These p-values are obtained from the χ^2 defined in Eq. (19), using the appropriate $d.o.f.$. The error bars are the statistical uncertainties resulting from those on the lattice covariance matrices determined in Appendix B (see also footnote 12). The corresponding uncertainty distributions are plotted in Figs. 4 and 6 (for the third and eleventh row of this plot, respectively). The filled/empty points correspond to the largest/smallest p-value obtained from the four systematic variations of the lattice covariance matrix of Eqs. (B31)–(B34). The vertical dashed line corresponds to 5%. We consider solutions with p-values above and close to that line to be compatible with the rescaling hypothesis in the corresponding interval and those far below to be incompatible. The rightmost panel displays the rescaling percentages δ_1, corresponding to the p-values on the same row that have the same plot symbol. Here the uncertainties are the leading ones, given by the nominal values of the lattice and R-ratio covariance matrices, i.e., not including the small, additional uncertainty associated with the statistical uncertainty on the lattice covariance matrices.

the two which induce the smallest/largest χ^2 values in the fit for each of the intervals. Indeed, depending on the fit configuration, the smallest χ^2 values are generally obtained for matrices "0" and "2," while the largest ones are for matrix "3."

More specifically, Table II and Fig. 3 report the nominal results (obtained using the nominal lattice covariance matrices) for the rescaling percentage δ_1 with its uncertainty propagated from the covariance matrices of the lattice QCD and dispersive results[13] in Table II, the $\chi^2/d.o.f.$ obtained by injecting γ_1 into Eq. (19), and the corresponding p-value.

Using either two or three moment integrals, the rescaling of the integrals of the R-ratio by a common factor, in the $\sqrt{s}$ regions below 0.96, 1.1, 1.8, or 3.0 GeV, allows a good agreement between the lattice and R-ratio approaches. In those scenarios, the rescalings range from approximately 2% to 5%, the lower number

corresponding to the larger intervals. In addition, a rescaling by around 5% in the ρ-peak interval, [0.63, 0.92] GeV, also restores the agreement.[14] However, the tension persists when the rescaling is performed in the region below 0.63 GeV.

When using only the two observables $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$, agreement is also restored through rescalings above either 0.63, 0.96, 1.1, 1.8 GeV or outside the [0.63, 0.92] GeV interval, while tensions persist for a rescaling performed above 3.0 GeV.

───────────

[13]It has also been checked that nonlinear effects in the uncertainty propagation are small, by comparing the outcome of the uncertainty propagation when performing $+1\sigma$ variations, on the one hand, and -1σ variations, on the other.

[14]As a function of $\sqrt{s}$, the lattice SD and ID windows correspond to weighting the R-ratio by kernels depicted by the orange and green curves in the right-hand panel of Fig. 1. In Ref. [11] the authors compare the data-driven results for those windows [6,14], as well as the one for $a_\mu^{\text{LO-HVP}}$ [3,4] (with a kernel given by the blue curve in that same figure), to their result for the SD window, theirs and others' [6,10,11] for the ID one, as well as the result of Ref. [6] for $a_\mu^{\text{LO-HVP}}$. On the basis of that comparison, the authors argue that "qualitatively, … the accurate lattice results for the time windows and for the full HVP term could be compatible with an overall few-percent enhancement of the e^+e^- cross section data in the 2π channels at center-of-mass energies below 1 GeV," anticipating our quantitative results.

TABLE II. Results for various rescaling scenarios, using two observables ($a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$) or three [adding $\delta(\Delta\alpha_{\text{had}}^{(5)})$], as indicated in column 1. In these scenarios it is assumed that the integrals of the experimental R-ratio in the $\sqrt{s}$ interval I_1 are rescaled by a common $\gamma_1 \equiv 1 + \delta_1$. This interval is given in column 2 and the *nominal* rescaling percentage δ_1 (obtained through the weighted average discussed in Sec. IV B, using the nominal values of the lattice covariance matrices obtained in Appendix B), in column 4, with the corresponding nominal uncertainty propagated from the covariance matrices of the lattice QCD and dispersive results ($C_{\text{lat}}^{\tilde\gamma}$ and $C_R^{\tilde\gamma}$) indicated between "()". Column 3 indicates which of the four systematic uncertainty variations of the lattice covariance (Lat. cov.) matrices give the best and worst fit qualities. Column 5 gives the nominal $\chi^2/d.o.f.$ for the scenario [obtained by injecting γ_1 into Eq. (19)], and column 6 gives the corresponding p-value. The values indicated between "[]" correspond to the $\pm 1\sigma$ quantiles based on the lattice bootstrap replicas discussed in Appendix B (see also footnote 12). The last column indicates the shift on $\Delta\alpha_{\text{had}}^{(5)}(M_Z^2)$ induced by the δ_1 rescaling.

No. of observables	I_1 (GeV)	Lat. cov.	δ_1 (%)	$\chi^2/d.o.f.$	p-value (%)	$\delta_1 \times \Delta\alpha_{\text{had}}^{(5)}(M_Z^2)[I_1] \times 10^4$
2	$[\sqrt{s_{\text{th}}}, 0.63]$	0	$15.9(5.3)[^{+0.9}_{-0.8}]\%$	$10.0[^{+2.4}_{-1.9}]/1$	$0.16[^{+0.31}_{-0.13}]\%$	0.80
2	$[\sqrt{s_{\text{th}}}, 0.63]$	3	$17.4(5.7)[^{+0.6}_{-0.5}]\%$	$17.4[^{+2.2}_{-1.9}]/1$	$0.003[^{+0.010}_{-0.004}]\%$	0.88
2	$[0.63, \infty[$	0	$3.1(0.9)[^{+0.05}_{-0.05}]\%$	$0.9[^{+0.1}_{-0.1}]/1$	$34.6[^{+3.2}_{-3.2}]\%$	8.49
2	$[0.63, \infty[$	3	$3.2(0.9)[^{+0.02}_{-0.02}]\%$	$1.3[^{+0.1}_{-0.1}]/1$	$25.2[^{+2.8}_{-2.2}]\%$	8.71
3	$[\sqrt{s_{\text{th}}}, 0.63]$	0	$16.4(5.4)[^{+0.9}_{-0.7}]\%$	$10.6[^{+2.2}_{-1.7}]/2$	$0.49[^{+0.73}_{-0.36}]\%$	0.83
3	$[\sqrt{s_{\text{th}}}, 0.63]$	3	$17.9(5.8)[^{+0.6}_{-0.5}]\%$	$17.8[^{+2.1}_{-1.9}]/2$	$0.013[^{+0.038}_{-0.016}]\%$	0.91
3	$[0.63, \infty[$	0	$2.5(0.7)[^{+0.08}_{-0.07}]\%$	$3.8[^{+0.6}_{-0.5}]/2$	$14.7[^{+4.2}_{-4.0}]\%$	6.68
3	$[0.63, \infty[$	3	$2.6(0.7)[^{+0.04}_{-0.04}]\%$	$5.3[^{+0.5}_{-0.4}]/2$	$7.0[^{+1.8}_{-1.6}]\%$	6.96
2	$[\sqrt{s_{\text{th}}}, 0.96]$	0	$3.7(1.1)[^{+0.1}_{-0.1}]\%$	$2.8[^{+0.5}_{-0.4}]/1$	$9.3[^{+2.8}_{-2.5}]\%$	1.32
2	$[\sqrt{s_{\text{th}}}, 0.96]$	3	$3.9(1.1)[^{+0.06}_{-0.06}]\%$	$4.4[^{+0.4}_{-0.5}]/1$	$3.5[^{+1.4}_{-1.0}]\%$	1.39
2	$[0.96, \infty[$	0	$9.4(2.6)[^{+0.04}_{-0.04}]\%$	$0.09[^{+0.01}_{-0.009}]/1$	$77.0[^{+1.2}_{-1.3}]\%$	22.59
2	$[0.96, \infty[$	3	$9.5(2.5)[^{+0.02}_{-0.02}]\%$	$0.12[^{+0.01}_{-0.01}]/1$	$72.9[^{+1.5}_{-1.1}]\%$	22.75
3	$[\sqrt{s_{\text{th}}}, 0.96]$	0	$3.8(1.1)[^{+0.09}_{-0.09}]\%$	$3.1[^{+0.4}_{-0.3}]/2$	$21.7[^{+4.3}_{-4.2}]\%$	1.36
3	$[\sqrt{s_{\text{th}}}, 0.96]$	3	$4.0(1.1)[^{+0.06}_{-0.05}]\%$	$4.5[^{+0.4}_{-0.4}]/2$	$10.7[^{+3.2}_{-2.3}]\%$	1.42
3	$[0.96, \infty[$	2	$3.5(1.3)[^{+0.2}_{-0.2}]\%$	$10.9[^{+2.2}_{-1.6}]/2$	$0.43[^{+0.54}_{-0.30}]\%$	8.35
3	$[0.96, \infty[$	3	$3.7(1.3)[^{+0.1}_{-0.1}]\%$	$14.1[^{+1.5}_{-1.2}]/2$	$0.089[^{+0.083}_{-0.052}]\%$	8.91
2	$[\sqrt{s_{\text{th}}}, 1.1]$	0	$3.3(1.0)[^{+0.1}_{-0.1}]\%$	$2.2[^{+0.4}_{-0.3}]/1$	$13.4[^{+3.2}_{-2.9}]\%$	1.40
2	$[\sqrt{s_{\text{th}}}, 1.1]$	3	$3.4(1.0)[^{+0.05}_{-0.04}]\%$	$3.5[^{+0.3}_{-0.4}]/1$	$6.3[^{+1.9}_{-1.3}]\%$	1.46
2	$[1.1, \infty[$	0	$14.1(3.9)[^{+0.07}_{-0.08}]\%$	$0.1[^{+0.02}_{-0.02}]/1$	$70.9[^{+1.6}_{-1.6}]\%$	33.01
2	$[1.1, \infty[$	3	$14.3(3.8)[^{+0.04}_{-0.04}]\%$	$0.2[^{+0.02}_{-0.02}]/1$	$65.8[^{+1.8}_{-1.4}]\%$	33.31
3	$[\sqrt{s_{\text{th}}}, 1.1]$	0	$3.4(1.0)[^{+0.07}_{-0.07}]\%$	$2.4[^{+0.3}_{-0.3}]/2$	$30.3[^{+4.5}_{-4.4}]\%$	1.44
3	$[\sqrt{s_{\text{th}}}, 1.1]$	3	$3.5(1.0)[^{+0.04}_{-0.04}]\%$	$3.5[^{+0.3}_{-0.3}]/2$	$17.8[^{+3.8}_{-2.8}]\%$	1.49
3	$[1.1, \infty[$	2	$3.5(1.4)[^{+0.2}_{-0.2}]\%$	$13.0[^{+2.9}_{-2.0}]/2$	$0.15[^{+0.27}_{-0.12}]\%$	8.14
3	$[1.1, \infty[$	3	$3.7(1.4)[^{+0.1}_{-0.1}]\%$	$17.1[^{+1.9}_{-1.6}]/2$	$0.019[^{+0.027}_{-0.014}]\%$	8.70
2	$[\sqrt{s_{\text{th}}}, 1.8]$	0	$2.9(0.8)[^{+0.1}_{-0.1}]\%$	$1.7[^{+0.3}_{-0.3}]/1$	$19.8[^{+3.4}_{-3.2}]\%$	1.63
2	$[\sqrt{s_{\text{th}}}, 1.8]$	3	$3.1(0.9)[^{+0.03}_{-0.03}]\%$	$2.5[^{+0.2}_{-0.3}]/1$	$11.3[^{+2.4}_{-1.8}]\%$	1.69
2	$[1.8, \infty[$	0	$31.8(9.1)[^{+0.6}_{-0.6}]\%$	$1.5[^{+0.2}_{-0.2}]/1$	$21.4[^{+3.4}_{-3.2}]\%$	70.17
2	$[1.8, \infty[$	3	$32.9(9.2)[^{+0.4}_{-0.3}]\%$	$2.3[^{+0.2}_{-0.2}]/1$	$12.8[^{+2.5}_{-1.8}]\%$	72.62
3	$[\sqrt{s_{\text{th}}}, 1.8]$	0	$3.0(0.9)[^{+0.05}_{-0.05}]\%$	$1.7[^{+0.2}_{-0.2}]/2$	$43.7[^{+4.9}_{-5.2}]\%$	1.65
3	$[\sqrt{s_{\text{th}}}, 1.8]$	3	$3.1(0.9)[^{+0.03}_{-0.03}]\%$	$2.5[^{+0.2}_{-0.3}]/2$	$28.6[^{+4.5}_{-3.5}]\%$	1.70
3	$[1.8, \infty[$	2	$3.5(1.7)[^{+0.2}_{-0.1}]\%$	$15.1[^{+3.6}_{-2.4}]/2$	$0.052[^{+0.130}_{-0.046}]\%$	7.79
3	$[1.8, \infty[$	3	$3.7(1.7)[^{+0.08}_{-0.08}]\%$	$20.3[^{+2.4}_{-2.0}]/2$	$0.0039[^{+0.0081}_{-0.0034}]\%$	8.18
2	$[\sqrt{s_{\text{th}}}, 3.0]$	0	$2.8(0.8)[^{+0.06}_{-0.06}]\%$	$1.5[^{+0.2}_{-0.2}]/1$	$22.0[^{+3.4}_{-3.2}]\%$	2.03
2	$[\sqrt{s_{\text{th}}}, 3.0]$	3	$2.9(0.8)[^{+0.03}_{-0.03}]\%$	$2.3[^{+0.2}_{-0.2}]/1$	$13.2[^{+2.5}_{-1.9}]\%$	2.10
2	$[3.0, \infty[$	0	$70.5(22.4)[^{+3.6}_{-3.2}]\%$	$7.8[^{+1.7}_{-1.4}]/1$	$0.51[^{+0.66}_{-0.35}]\%$	143.42
2	$[3.0, \infty[$	3	$76.9(23.9)[^{+2.4}_{-2.2}]\%$	$13.4[^{+1.6}_{-1.5}]/1$	$0.025[^{+0.052}_{-0.024}]\%$	156.38

(Table continued)

TABLE II. *(Continued)*

No. of observables	I_1 (GeV)	Lat. cov.	δ_1 (%)	$\chi^2/d.o.f.$	p-value (%)	$\delta_1 \times \Delta\alpha_{had}^{(5)}(M_Z^2)[I_1] \times 10^4$
3	$[\sqrt{s_{th}}, 3.0]$	0	$2.7(0.8)[^{+0.05}_{-0.05}]\%$	$1.7[^{+0.3}_{-0.2}]/2$	$43.1[^{+5.8}_{-6.1}]\%$	1.97
3	$[\sqrt{s_{th}}, 3.0]$	3	$2.8(0.8)[^{+0.03}_{-0.03}]\%$	$2.7[^{+0.3}_{-0.3}]/2$	$26.3[^{+4.3}_{-3.7}]\%$	2.04
3	$[3.0, \infty[$	2	$4.2(2.4)[^{+0.09}_{-0.08}]\%$	$16.0[^{+3.9}_{-2.6}]/2$	$0.033[^{+0.094}_{-0.030}]\%$	8.59
3	$[3.0, \infty[$	3	$4.3(2.4)[^{+0.06}_{-0.05}]\%$	$21.7[^{+2.7}_{-2.2}]/2$	$0.0020[^{+0.0049}_{-0.0018}]\%$	8.80
2	$[0.63, 0.92]$	0	$4.8(1.4)[^{+0.1}_{-0.1}]\%$	$1.7[^{+0.3}_{-0.2}]/1$	$19.6[^{+3.4}_{-3.2}]\%$	1.42
2	$[0.63, 0.92]$	3	$4.9(1.4)[^{+0.06}_{-0.05}]\%$	$2.5[^{+0.3}_{-0.3}]/1$	$11.2[^{+2.4}_{-1.8}]\%$	1.47
2	$[\sqrt{s_{th}}, 0.63] \cup [0.92, \infty[$	0	$6.2(1.8)[^{+0.1}_{-0.1}]\%$	$1.6[^{+0.2}_{-0.2}]/1$	$20.4[^{+3.4}_{-3.2}]\%$	15.33
2	$[\sqrt{s_{th}}, 0.63] \cup [0.92, \infty[$	3	$6.5(1.8)[^{+0.07}_{-0.07}]\%$	$2.4[^{+0.2}_{-0.2}]/1$	$11.9[^{+2.4}_{-1.8}]\%$	15.88
3	$[0.63, 0.92]$	0	$4.9(1.4)[^{+0.08}_{-0.08}]\%$	$1.8[^{+0.2}_{-0.2}]/2$	$40.2[^{+4.0}_{-4.1}]\%$	1.45
3	$[0.63, 0.92]$	3	$5.0(1.4)[^{+0.05}_{-0.05}]\%$	$2.6[^{+0.2}_{-0.2}]/2$	$27.9[^{+4.0}_{-3.1}]\%$	1.50
3	$[\sqrt{s_{th}}, 0.63] \cup [0.92, \infty[$	0	$3.4(1.1)[^{+0.2}_{-0.1}]\%$	$9.0[^{+1.8}_{-1.3}]/2$	$1.1[^{+1.1}_{-0.7}]\%$	8.30
3	$[\sqrt{s_{th}}, 0.63] \cup [0.92, \infty[$	3	$3.6(1.1)[^{+0.08}_{-0.08}]\%$	$12.4[^{+1.3}_{-1.1}]/2$	$0.21[^{+0.18}_{-0.12}]\%$	8.80

Adding the third observable $\delta(\Delta\alpha_{had}^{(5)})$ increases the constraints and tensions arise for a rescaling in any of the intervals that do not include the ρ peak.

It is worth noting that the values of the rescaling percentage δ_1 for the various configurations presented in Table II are much larger than the (sub)percent-level uncertainties of the experimental R-ratio in the corresponding $\sqrt{s}$ regions. Therefore, such shifts are not to be interpreted as a plausible way of solving the tension between lattice QCD and the dispersive approaches, but rather as a way of comparing various hypotheses for the possible source(s) of the tension.

Also given in Table II are the $\pm 1\sigma$ statistical quantiles obtained for δ_1, the $\chi^2/d.o.f.$, and the corresponding p-value, when considering their distributions obtained from the bootstrap replicas of the lattice covariance matrices computed in Appendix D (see also footnote 12). Examples of such distributions are shown in Figs. 4–7. In the case of Fig. 3, which summarizes our results, those uncertainties are shown only for the p-values.

For δ_1, in many scenarios the $\pm 1\sigma$ quantiles based on the lattice bootstrap replicas are smaller than the primary uncertainty propagated from the covariance matrices of the lattice QCD and dispersive results, by about an order of magnitude or more. The changes associated with the four systematic variations of the lattice covariance matrix are also small, though somewhat larger than the bootstrap ones. This is not necessarily surprising because they are uncertainties on the primary uncertainty. In fact, the latter are of only a few percent. Furthermore, the same statistical and systematic uncertainties on the $\chi^2/d.o.f.$ and on the corresponding p-values do not change the fact that a given fit can be considered "good" or "bad." These observations indicate that the statistical and systematic

variations of the lattice covariance matrix do not change the conclusions qualitatively.

As mentioned in the previous subsection, the same bootstrap resampling can present some slight differences between the mean, median, and nominal values of the corresponding distributions. In particular, the asymmetric, non-Gaussian tails (present especially for the p-values) impact the mean values (see e.g., Figs. 4 and 6). Still, these differences are well within the $\pm 2\sigma$ quantiles (most often within the $\pm 1\sigma$ quantiles) and do not impact the conclusions of the study. The asymmetric tails also induce some differences between the uncertainties obtained from asymmetric variances,[15] those from the $\pm 1\sigma$ quantiles, and those from the $\pm 2\sigma$ ones (divided by 2). However, here also the conclusions of the study are not impacted by the slight differences among these three estimates of uncertainties.

In order to display the implications of observed normalization shifts for electroweak precision observable (EWPO) fits [17–20], in Table II we indicate the impact of the δ_1 rescaling on $\Delta\alpha_{had}^{(5)}(M_Z^2)$.[16] As expected, the largest impacts on $\Delta\alpha_{had}^{(5)}(M_Z^2)$ are observed for modifications of the R-ratio in the high-$\sqrt{s}$ intervals. However, while such modifications may have reasonable p-values with constraints from only a_μ^{LO-HVP} and $a_{\mu,win}^{LO-HVP}$, when one

---

[15]These variances are computed from the values in the distribution that are above and below the nominal one.

[16]While the uncertainty of the δ_1 rescaling factors is determined from the fits, we do not attempt a correlated propagation to (the uncertainty of) the corresponding shift of $\Delta\alpha_{had}^{(5)}(M_Z^2)$, nor a simple rescaling of the uncertainty of the latter. This follows the rescaling approach discussed in Ref. [20]. There it was pointed out that, if such rescaling is necessary for the nominal values of the hadronic spectra, it is not obvious how the corresponding uncertainties should be computed.
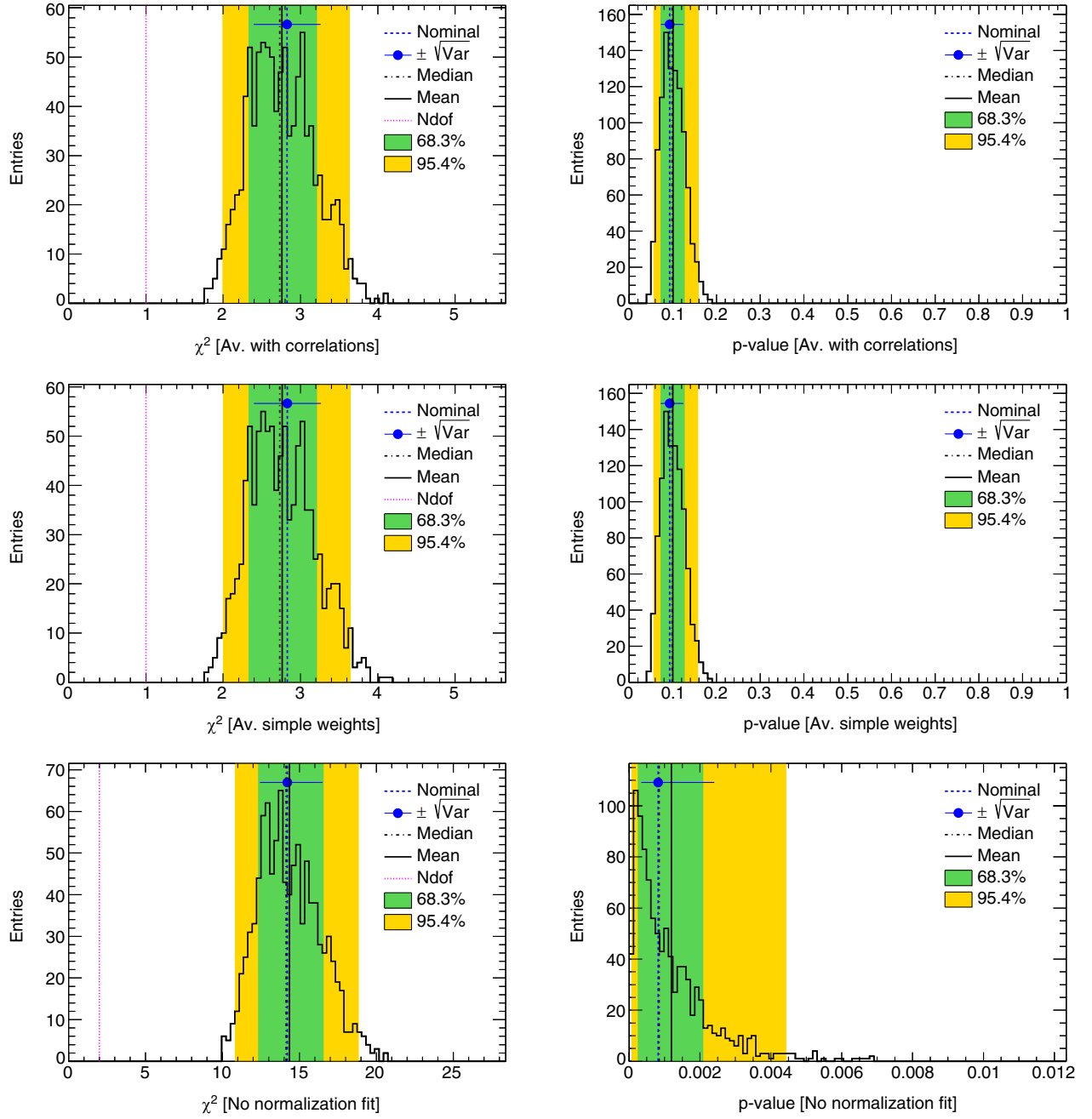
FIG. 4. Distributions of χ^2 (left) and p-value (right) obtained by sampling the bootstrap replicas of the lattice covariance matrix. These figures show the effects of different approaches to determining γ_1 and its uncertainty in the low-mass region below 0.96 GeV. Two moment integrals are considered here, with the lattice covariance matrix 0. In the first row, we consider normalization fits using the full covariance matrices for determining the averaging weights—their comparison with the two plots in the following row is discussed in Appendix D. In the middle row, we consider normalization fits with the averaging weights proportional to the inverse of the $\tilde{\gamma}_j$ uncertainties squared, as considered in the present section. In the final row, we set the γ_b normalization coefficients to unity, i.e., we perform no rescaling—these are discussed in Sec. V A. The blue horizontal error bar indicates the asymmetric variance of the distribution, computed on each side of the nominal value (blue point and blue dashed vertical line). The median (dash-dotted black line), mean (black continuous line), 68.3% and 95.4% quantiles (colored bands) of the distribution are also indicated. The number of degrees of freedom in the χ^2 calculation is indicated by the dotted pink line. Note that the scales in the last row of plots (corresponding to no rescaling) are very different from those of the two previous rows (in which rescaling is allowed) to account for the fact that the χ^2 and p-values with no rescaling are very poor.
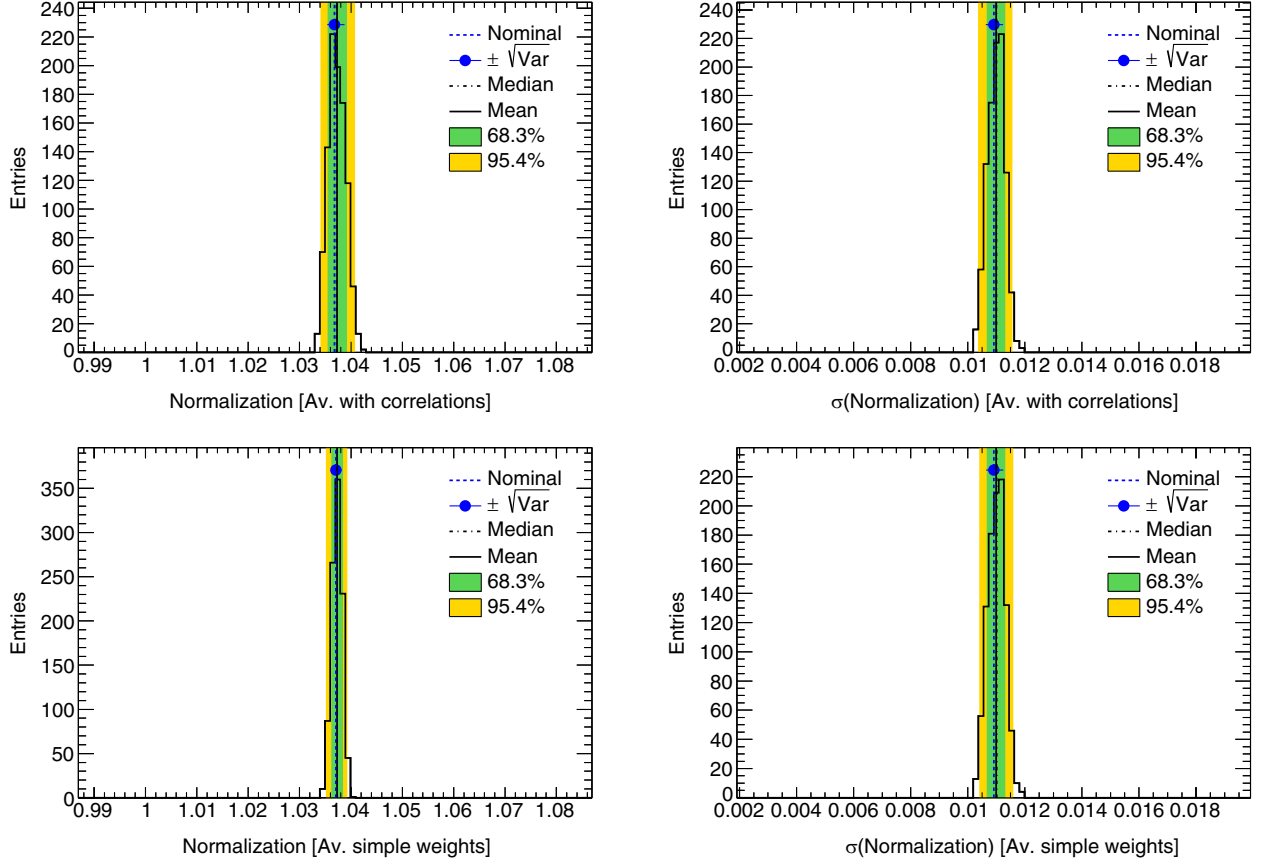
FIG. 5.   Distributions of the normalization rescaling factor γ_1 (left) and its uncertainty propagated from the covariance matrices of the lattice QCD and dispersive results (right), obtained by sampling the bootstrap replicas of the lattice covariance matrix. These figures show the effects of different approaches to determining γ_1 and its uncertainty in the low-mass region below 0.96 GeV. Two moment integrals are considered here, with the lattice covariance matrix 0. In the first row, we consider normalization fits using the full covariance matrices of the $\tilde{\gamma}_j$ for determining the averaging weights. In the second row, we consider normalization fits with the averaging weights proportional to the inverse of the $\tilde{\gamma}_j$ uncertainties squared. The blue horizontal line indicates the asymmetric variance of the distribution, computed on each side of the nominal value (dashed blue vertical line). The median (dash-dotted black line), mean (black continuous line), 68.3% and 95.4% quantiles (colored bands) of the distribution are also indicated.

also includes the hadronic contribution to the five-flavor running of the electromagnetic coupling, between -10 and -1 GeV^2, the fit quality becomes very poor and the common rescaling of the contributions to the observables in these high-$\sqrt{s}$ intervals is strongly reduced too. Thus, in the presence of the $\delta(\Delta\alpha_{\mathrm{had}}^{(5)})$ constraint, the corresponding modifications of the R-ratio will not significantly impact the conclusions of EWPO fits.

As they are for the comparisons of Sec. V A, the conclusions of the studies performed here are stable against the statistical and systematic variations of the lattice covariance matrix. As shown in Appendix D, they are also stable with respect to the averaging methodology that is employed. Moreover, very similar results are obtained using the χ^2 approach to the rescaling described in Appendix E 3.

VI. SUMMARY OF RESULTS AND CONCLUSIONS

We have presented a framework that enables a comparison of the primary ingredients that are used for determining HVP in the lattice QCD and data-driven approaches. The framework makes use of observables computable in both formalisms. These primary ingredients are for the lattice approach, the zero three-momentum, current-current correlator of Eq. (1), studied as a function of Euclidean time t; for the data-driven approach, the experimental R-ratio is studied as a function of timelike c.m. energy, $\sqrt{s}$ [Eq. (2)]. The observables are integrals over t and $\sqrt{s}$, weighted with appropriately chosen kernels. We alternatively call these observables or moment integrals.
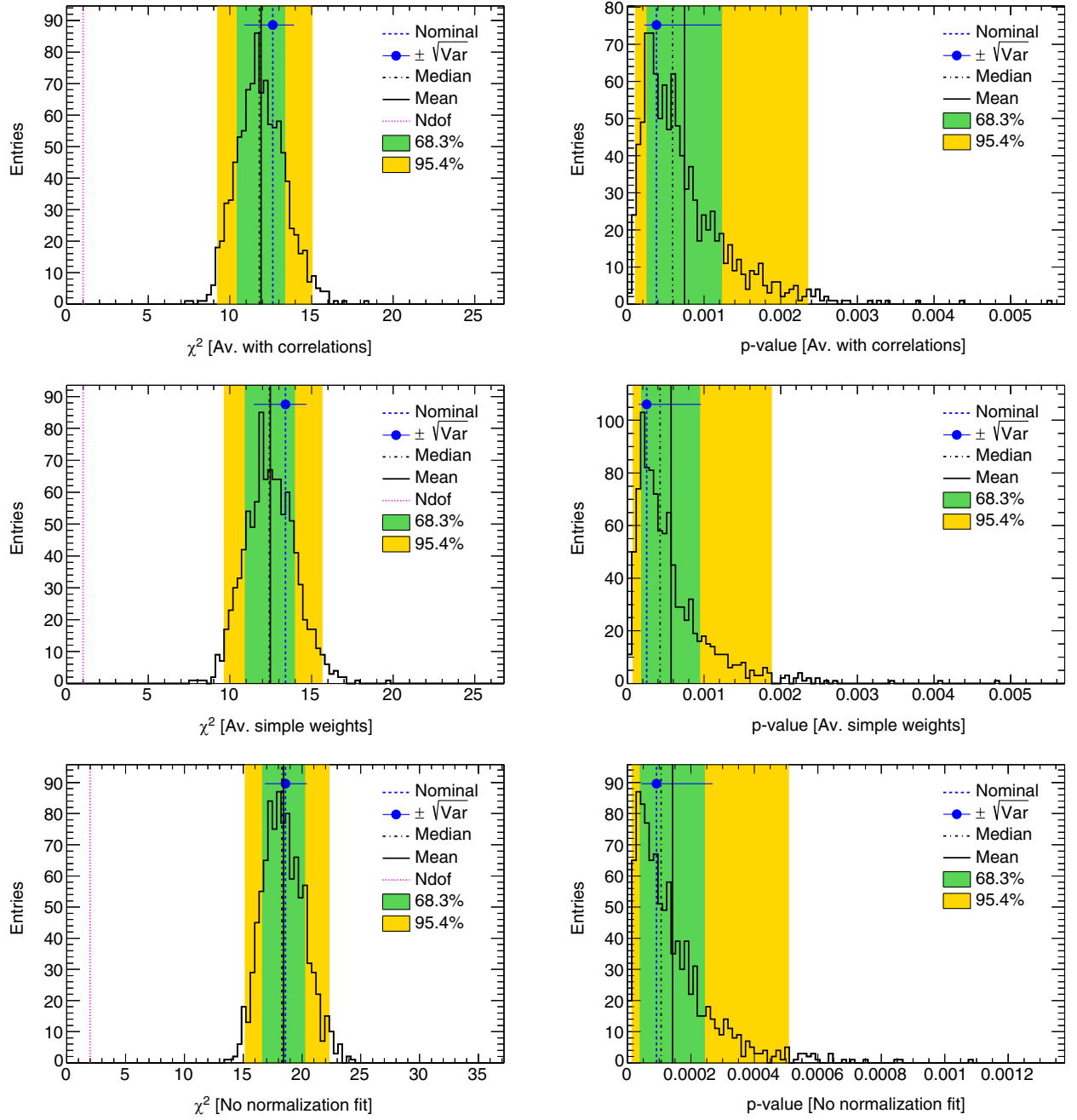
FIG. 6. Distributions of χ^2 (left) and p-value (right), obtained by sampling the bootstrap replicas of the lattice covariance matrix. These figures show the effects of different approaches to determining γ_1 and its uncertainty in the high-mass region above 3.0 GeV. Two moment integrals are considered here, with the lattice covariance matrix 3. In the first row, we consider normalization fits using the full covariance matrices for determining the averaging weights—their comparison with the two plots in the following row is discussed in Appendix D. In the middle row, we consider normalization fits with the averaging weights proportional to the inverse of the $\tilde{\gamma}_j$ uncertainties squared, as considered in the present section. In the final row, we set the γ_b normalization coefficients to unity, i.e., we perform no rescaling—these are discussed in Sec. VA. The blue horizontal error bar indicates the asymmetric variance of the distribution, computed on each side of the nominal value (blue point and blue dashed vertical line). The median (dash-dotted black line), mean (black continuous line), 68.3% and 95.4% quantiles (colored bands) of the distribution are also indicated. The number of degrees of freedom in the χ^2 calculation is indicated by the dotted pink line. Note that the scales in the last row of plots (corresponding to no rescaling) are very different from those of the two previous rows (in which rescaling is allowed) to account for the fact that the χ^2 and p-values with no rescaling are very poor.
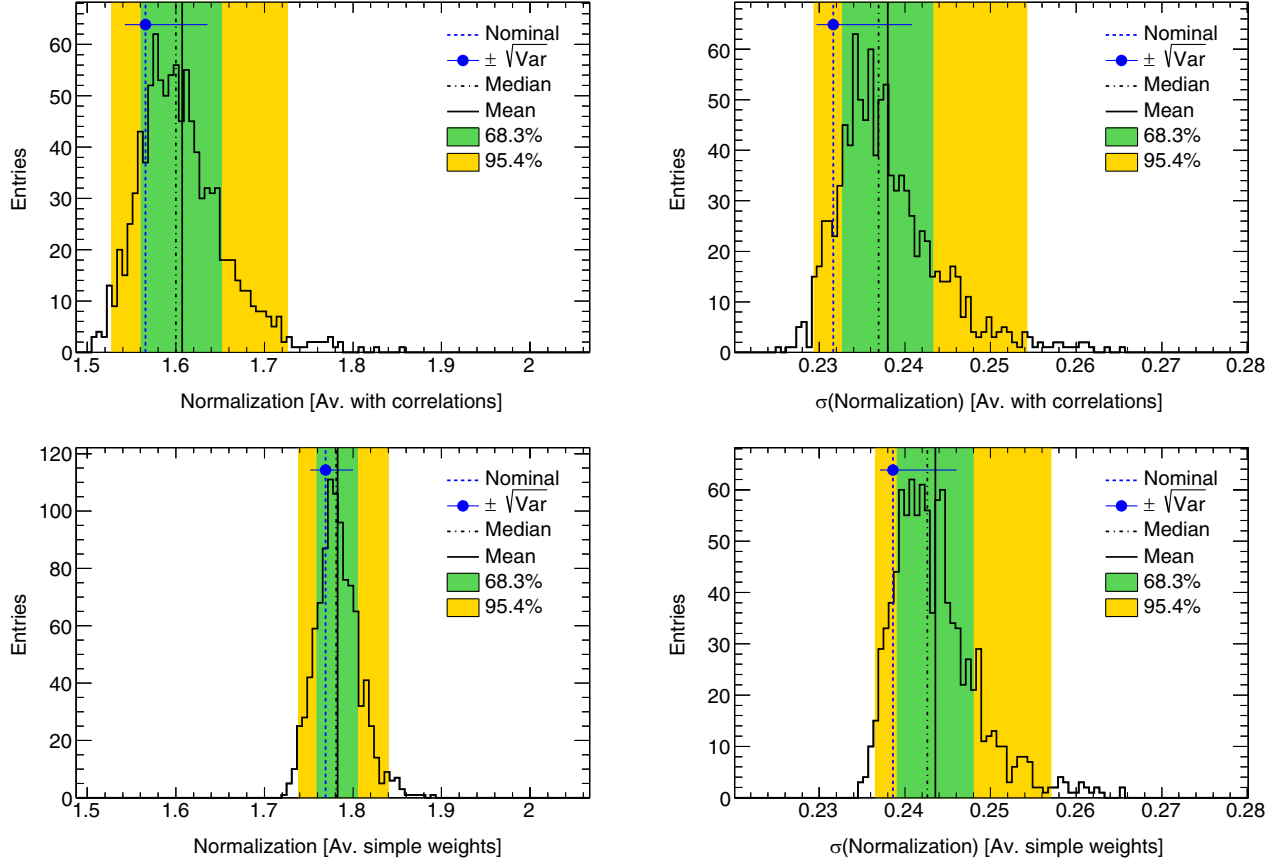
FIG. 7. Distributions of the normalization rescaling factor γ_1 (left) and its uncertainty propagated from the covariance matrices of the lattice QCD and dispersive results (right), obtained by sampling the bootstrap replicas of the lattice covariance matrix. These figures show the effects of different approaches to determining γ_1 and its uncertainty in the high-mass region above 3.0 GeV. Two moment integrals are considered here, with the lattice covariance matrix 3. In the first row, we consider normalization fits using the full covariance matrices of the $\tilde{\gamma}_j$ for determining the averaging weights. In the second row, we consider normalization fits with the averaging weights proportional to the inverse of the $\tilde{\gamma}_j$ uncertainties squared. The blue horizontal line indicates the asymmetric variance of the distribution, computed on each side of the nominal value (dashed blue vertical line). The median (dash-dotted black line), mean (black continuous line), 68.3% and 95.4% quantiles (colored bands) of the distribution are also indicated.

The kernels can be chosen to be localized on the *t*-axis, as are the $a_\mu^{\text{LO-HVP}}$ time windows proposed in Ref. [9]. Then by comparing the corresponding observables obtained in the two approaches, one can aim to isolate distance scales in the lattice correlator that may lie at the origin of the observed disagreement in the values, for instance, of $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. Such comparisons are also interesting for HVP observables that have limited support in photon virtuality. These are typically related to the Adler function and the hadronic running of the electromagnetic coupling. They provide a complementary way to isolate relevant scales in the lattice correlator. More generally, our method applies to any observable related to HVP.

Our framework also allows us to gain information about the R-ratio in chosen regions of c.m. energy, from these same observables. The main challenge here comes from the fact that this requires solving a notoriously ill-posed, inverse problem that arises because the lattice correlator is a Laplace transform of the R-ratio [Eq. (2)]. It also comes from the fact that observables which are localized in Euclidean time are much less so in c.m. energy. We solve the second problem by partitioning the $\sqrt{s}$-axis from threshold to infinity into intervals and computing the contributions of each such interval to the observables of interest. The first challenge—the reconstruction of a continuous function from a finite number of its weighted integrals does not have *a priori* a unique solution—is addressed by limiting the information that we try to extract while monitoring the possible pitfalls associated with such a reconstruction.

In the present work, we focus on simple scenarios. We consider three observables related to HVP: $a_\mu^{\text{LO-HVP}}$, $a_{\mu,\text{win}}^{\text{LO-HVP}}$, and the hadronic running of the electromagnetic coupling between spacelike virtualities -10 and -1 GeV^2, $\delta(\Delta\alpha_{\text{had}}^{(5)})$. This choice is motivated by the fact that these quantities have been determined on the lattice to the

subpercent level, with all relevant corrections [6]. It is important to note that the result for $\delta(\Delta\alpha_{\text{had}}^{(5)})$ is not the main focus of Ref. [6] and should be understood as preliminary. Nevertheless, we include it because it imposes complementary constraints on the spectral function and demonstrates the flexibility of our approach.

Then, we use these observables to

(I) isolate regions of t in which the lattice correlator leads to values of the observables that are inconsistent with those obtained in the data-driven approach;

(II) determine intervals of $\sqrt{s}$ in which a change in the experimental R-ratio could explain the observed discrepancies between the lattice and R-ratio approaches, by allowing a common rescaling of the contributions from these intervals to the observables of interest.

Because our approach is based on the minimization of χ^2 functions, it allows us to quantify the compatibility of the observables in the two approaches, as well as the consistency of the observables with our hypothesis that a change of the experimental R-ratio, in a particular $\sqrt{s}$ interval, could explain the observed discrepancies. Clearly, the very simple model of item II tells us little about the shape of the deformations of the timelike spectral function that may be required. Moreover, it does not incorporate, *a priori*, physical constraints, such as those imposed by chiral perturbation theory at small c.m. energies, by perturbative QCD at large ones, or by analyticity and positivity at all energies. However, it is not necessarily incompatible with some of them either. In addition, it can help isolate $\sqrt{s}$ intervals in which the experimental R-ratio may have to be modified to explain the lattice results.

Within a given approach, the observables obtained are not independent. Thus, it is critical to take into account correlations between them. This is necessary to correctly quantify the compatibility of the lattice and data-driven observables with one another. It is also particularly important when considering the inverse problem because its solutions are very sensitive to these correlations. In fact, because of the exponential enhancement of fluctuations, it is also important to consider the stability of the results against uncertainties on the corresponding covariance matrices. We take into account all uncertainties and correlations and, for the lattice results, we also consider the effect of uncertainties on the corresponding covariance matrices. The consideration of such an effect in the data-driven approach is left for future work.

Using the methods described in Sec. IVA, we find the results presented and discussed in Sec. VA. They represent our attempt to isolate regions in the lattice correlator, on the Euclidean time axis, that do not agree with the data-driven approach.

As already observed in Ref. [6], the lattice result for $a_\mu^{\text{LO-HVP}}$ is 2.1σ above the one obtained from the data-driven approach (corresponding to a p-value of 5×10^{-2}), and the one for $a_{\mu,\text{win}}^{\text{LO-HVP}}$, 3.8σ above (p-value of 1×10^{-4}), while the

preliminary lattice result for $\delta(\Delta\alpha_{\text{had}}^{(5)})$ is only 1.4σ above (p-value of 0.15). Moreover, the lattice result for the leading, isospin-limit, up-and-down quark contribution to the window observable has been independently confirmed by nine collaborations [6,10–13], of which two also confirm the results for some of the other contributions [10,12]. This increases the tensions with the data-driven results to a conservative 4.2σ. Similar observations were made in [25].

The discrepancies discussed in the previous paragraph are confirmed when correlations are taken into account via Eq. (12). Considering first $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ together we find p-values which range approximately from 2×10^{-5} to 1.7×10^{-3}, taking into account the maximum systematic variation and 1σ statistical fluctuation in the lattice covariance matrices. These are substantially smaller than the one for $a_\mu^{\text{LO-HVP}}$ alone.

Adding to the comparison the preliminary lattice result for $\delta(\Delta\alpha_{\text{had}}^{(5)})$, and its data-driven counterpart, reduces the overall tension to some degree. The p-values now range from approximately 10^{-4} to 6×10^{-3}, still making an overall agreement very unlikely.

Combining all of this information with the shape of the observable's kernels, in Sec. VA we argue that our results point to an excess of the lattice correlator in at least part of the intermediate-time range from 0.4 to 1.5 fm, a probable excess at time separations above 1.5 fm, and possibly a slight suppression in part of the short-time range below ∼0.6 fm, which must be limited in light of the agreement of the short-distance-window lattice results of Refs. [11,89] with the data-driven one of Ref. [14].

Then, we turn to the implications that the comparison of the two approaches may have on the experimental R-ratio. Again, the tests considered take into account all systematic errors and correlations, as well as uncertainties on uncertainties in the lattice results.

Using the averaging procedure of Sec. IV B, we find that the lattice predictions for $a_\mu^{\text{LO-HVP}}$ and the corresponding $a_{\mu,\text{win}}^{\text{LO-HVP}}$ [but excluding $\delta(\Delta\alpha_{\text{had}}^{(5)})$], can be reproduced within uncertainties via a modification of the experimental R-ratio in various intervals of $\sqrt{s}$. More specifically, if the lower end of these intervals begins at threshold, the intervals must include the ρ peak. Indeed, for intervals whose upper bound is anywhere between $\sqrt{s} = 0.96$ and 3.0 GeV, the modifications have to be such that they increase the contributions to both the data-driven $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ in those intervals by approximately 2%–5%, the lower value corresponding to the larger intervals. These conclusions allow scenarios, such as those advocated in [90], in which the R-ratio would have to receive modifications above 1 GeV to explain the lattice results for $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$.

Interestingly, a modification of only the ρ-peak region is also compatible with the lattice results, but requires around

a 5% increase of the contributions to the two observables from that region. It is worth noting that such an increase is comparable to the one claimed by CMD-3 in a very similar region [15]. However, without having worked out the correlations between CMD-3 and the other $e^+e^- \to$ hadron measurements, it is impossible to make a more quantitative statement. Moreover, in the presence of large, unexplained discrepancies between measurements, and thus of large unknown systematics, one cannot claim to control such correlations and uncertainties.

Somewhat surprisingly, a scenario in which the experimental R-ratio is modified in an interval from threshold to infinity from which the ρ peak is excluded is also permitted, with p-values above $\sim 10\%$ in the two-observable case— i.e., beginning right below the $c\bar{c}$ resonance region. Moreover, modifications of the experimental R-ratio, in intervals that extend to infinity and begin at $\sqrt{s}$ as small as 0.63 GeV and as large as 1.8 GeV, are also allowed. In that case, the contributions to the data-driven $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ must be increased anywhere between approximately 2% and 40%. On the other hand, an explanation of the discrepancy between the two approaches via a modification of the experimental R-ratio only above 3 GeV is excluded with high probability.

In order to see if some of the above scenarios can be eliminated, we consider adding the observable $\delta(\Delta\alpha_{\text{had}}^{(5)})$. This observable has a footprint along the $\sqrt{s}$-axis that is significantly more extended toward larger c.m. energies than those of $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. We find that the addition of this observable eliminates, with significant probability, all scenarios in which the modification of the experimental R-ratio excludes the ρ peak. Nevertheless, the scenario including only the ρ-peak interval, and the one in which the interval begins below the ρ peak and extends to infinity, both survive this additional constraint. Moreover, these possible modifications of the experimental R-ratio imply a rescaling of the data-driven results, for all three observables, by a common factor between about 1.02 and 1.06. In particular, the scenario including only the ρ peak is barely changed by the addition of the constraint from $\delta(\Delta\alpha_{\text{had}}^{(5)})$, making it a robust explanation for the discrepancy between the lattice and data-driven approaches.[17]

---

[17]These conclusions allow a scenario such as the one discussed in Ref. [25]. In that work, neglecting arguably small QED, strong-isospin breaking, and b-quark effects, the authors compute an observable in lattice QCD which corresponds to convoluting the R-ratio with a Gaussian kernel and compare it to a data-driven determination based on the compilation of [4]. In particular, they consider a Gaussian centered at 0.80 GeV with a width at half maximum of 1.26 GeV, thus smearing the R-ratio over a $\sqrt{s}$ interval ranging from threshold to 2.06 GeV, when 95.5% of the area under the Gaussian is considered. In that interval in $\sqrt{s}$, they find that their lattice result for the Gaussian-convoluted R-ratio exceeds the data-driven one by 2.8% or 3σ.

Not surprisingly, very similar conclusions are obtained using the fitted-moment formalism given in Appendix E 3. There, however, the natural interpretation of the results is slightly different. In that approach, the contributions of different $\sqrt{s}$ intervals to the various observables become fit parameters, as are the rescaling factors. Thus, under the assumption that the proposed rescaling scenario is correct, as should also be verified by the corresponding $\chi^2/d.o.f.$, the fitted contributions to the observables can be understood as a combination of lattice and data-driven results for these contributions.

It is important to remember that all of the modifications to the experimental R-ratio discussed above are not consistent with the uncertainties of the timelike spectrum. This is made obvious by the very small p-values, discussed above, that correspond to the minimum of the χ^2 of Eq. (12), which measures the combined compatibility of the lattice and data-driven results for $a_\mu^{\text{LO-HVP}}$, $a_{\mu,\text{win}}^{\text{LO-HVP}}$, and $\delta(\Delta\alpha_{\text{had}}^{(5)})$.

As discussed in Appendix E, the methods presented above can be generalized to include more and more information about HVP from the lattice and data-driven approaches. This additional information will allow us to refine our understanding of the discrepancies between the two approaches, both in terms of the Euclidean correlation function and the experimental R-ratio. Our methods can also be generalized to include more theoretical constraints, such as those discussed earlier in this section. Moreover, the methods presented in Appendix E supply means of obtaining combinations of lattice and data-driven results for various quantities of interest related to HVP. The validity of these combinations rests on that of the models used to reconcile the two approaches.

Finally, in Appendix F we argue that the number of linearly independent moments that can be obtained with reasonable precision from the experimental R-ratio is less than 10 for the types of observables considered here, which should be computable with subpercent precision in the lattice approach.

Note added.—Recently, a preprint was uploaded on arXiv [91], which discusses the impact of tensions in $e^+e^- \to \pi^+\pi^-(\gamma)$ measurements on evaluating the HVP contribution to $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. This is particularly relevant given recent measurements by CMD-3 of $e^+e^- \to \pi^+\pi^-$ [15] and *BABAR*'s experimental study of higher-order radiative effects in $e^+e^- \to \pi^+\pi^-\gamma$ and $e^+e^- \to \mu^+\mu^-\gamma$ [92]. This study suggests that the measurement performed by the KLOE experiment is impacted by the higher-order radiative effects discussed in [92], at a level larger than the systematic uncertainty that is currently assigned to this effect. Furthermore, employing a combination of *BABAR*, CMD-3, and hadronic τ-decay data would decrease the tensions between data-driven and lattice results for $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. Resolving discrepancies

between CMD-2 [16] and CMD-3 [15] measurements would also be necessary before drawing final conclusions.

APPENDIX A: DATA-DRIVEN DETERMINATION OF THE OBSERVABLES AND THEIR CORRELATIONS

Computing the contributions to the moments integrals from the low-energy region, below 1.8 GeV, requires combining the spectra measured by various experiments. These measurements are performed either through a scan of the energy in the center of mass of the collider, as is the case of e.g., the CMD-2 and SND experiments, or using the initial-state radiation (ISR) technique, in the case of KLOE and *BABAR* (see e.g., Refs. [16,84–87,95–97] for the $\pi^+\pi^-$ channel, dominant in the case of $a_\mu^{\text{LO-HVP}}$). In the latter case, selecting events with a hard photon emitted from the initial state allows such studies to cover a broad range of invariant masses for the hadronic final-state system. Actually, *BABAR* covers the full energy range of interest, due to selecting events with the hard photon reconstructed in the detector and to the high energy in the center of mass of the collider. While the KLOE 2008 [84] and KLOE 2012 [86] measurements consider events with the hard ISR photon along the beam, the KLOE 2010 measurement uses large-angle ISR photons [85]. Furthermore, in measurements where the ISR luminosity is evaluated *in situ* based on $e^+e^- \to \mu^+\mu^-$ events [86,87,97], several systematic uncertainties are reduced. Still, among the currently available results, only the *BABAR* study uses the reconstructed extra photons (performing a measurement reconstructed at NLO and inclusive of any number of additional photons), necessary

at the subpercent-level precision aimed for here. The differences between the sampled event topologies, the detector technologies, and especially the analysis methodologies result in systematic tensions between the *BABAR* and KLOE measurements [3,73], but also between the KLOE measurements themselves (see Ref. [5] and references therein). Since these tensions are still to be understood and fixed, they require special care in the current combinations.

The channel-by-channel combinations in the DHMZ approach (implemented in the HVPtools software) [3,71–73] are performed in fine energy ranges, using a spline-based procedure. It takes into account the information on the uncertainties and their correlations between bins/points and between experiments. The DHMZ procedure also accounts for correlations between the various different exclusive channels when summing their contributions, which induces a necessary enhancement of the total uncertainty. Furthermore, the local tensions between the measurements are taken into account based on the computation of a $\chi^2/d.o.f.$ in fine energy ranges, used for the rescaling of the uncertainties. However, in the presence of systematic tensions, such as the ones between the *BABAR* and the KLOE measurements in the $\pi\pi$ channel, the local rescaling of the uncertainties is not sufficient. For this reason, in the most recent DHMZ study [3], these tensions are taken into account by treating half of the difference between integrals without either *BABAR* or KLOE as an extra uncertainty. This yields the dominant uncertainty on $a_\mu^{\text{LO-HVP}}$, amounting to 2.8×10^{-10}. The fidelity of the full analysis chain (polynomial interpolation, averaging, integration) has been successfully checked through a closure test, based on a Gounaris-Sakurai vector-meson resonance model, faithfully describing the $\pi^+\pi^-$ data [71].[18] The DHMZ approaches for accounting for correlations between channels and for the *BABAR*-KLOE tension in the $\pi\pi$ channel have also been adopted for the result of the WP [5].

The procedure described above is used to derive the experimental values of $R_{e^+e^- \to \text{hadrons}}(s)$, together with its covariance matrix, up to 1.8 GeV. It is completed with contributions from inclusive experimental measurements in the range between 3.7 and 5 GeV, combined using the same methodology as above, with contributions from narrow resonances, as well as from perturbative QCD for the continuum [3]. An uncertainty accounting for possible duality violation effects is also included. A direct integration of the resulting $R_{e^+e^- \to \text{hadrons}}(s)$ on the full mass range, together with a linear propagation of the uncertainties with their correlations, yields the moment integrals with the kernels of interest here, together with their covariance matrix.

---

[18]More recently, tests based on treed Gaussian processes have confirmed the good quality of the R-ratio modeling in our spline-based approach [98].

APPENDIX B: DETERMINATION OF LATTICE OBSERVABLES AND OF THEIR CORRELATIONS

The lattice QCD results used in this work were obtained in Ref. [6]. In order to calculate the two-point function $C(t)$ defined in Eq. (1), we need to evaluate the current correlator $\langle J_\mu(x) J_{\bar\mu}(\bar x)\rangle$ in our lattice regularization. We define the regularized current correlator as the second differential of the partition function with respect to an external photon field

$$\langle J_\mu(x) J_{\bar\mu}(\bar x)\rangle \equiv \left. \frac{\delta^2 \log Z}{\delta A^{\text{ext}}_{\mu,x} \delta A^{\text{ext}}_{\bar\mu,\bar x}}\right|_{A^{\text{ext}}=0}. \tag{B1}$$

Thus, we use the conserved current at source and sink. The correlator can be evaluated as

$$\langle J_\mu(x) J_{\bar\mu}(\bar x)\rangle / e^2 = \left\langle \sum_f C^{\text{conn}}_{\mu,x,\bar\mu,\bar x}(m_f, eq_f) + C^{\text{disc}}_{\mu,x,\bar\mu,\bar x} \right\rangle, \tag{B2}$$

plus a contact term that gives no contribution to the observables of interest here. $C^{\text{conn}}_{\mu,x,\bar\mu,\bar x}$ and $C^{\text{disc}}_{\mu,x,\bar\mu,\bar x}$ arise respectively from the connected and disconnected Wick contractions of the quark fields and are given by

$$C^{\text{conn}}_{\mu,x,\bar\mu,\bar x}(m_f, eq_f) = -\frac{1}{4}\text{Tr}(M_f^{-1} D_\mu[iP_x V_U e^{ieq_f A}] \\ \cdot M_f^{-1} D_{\bar\mu}[iP_{\bar x} V_U e^{ieq_f A}]), \tag{B3}$$

$$C^{\text{disc}}_{\mu,x,\bar\mu,\bar x} = \sum_{f,\bar f} I_{\mu,x}(m_f, eq_f) I_{\bar\mu,\bar x}(m_{\bar f}, eq_{\bar f}), \tag{B4}$$

where Tr is the spin-color trace, P_x is a projection operator that sets all sites other than x to zero, M_f^{-1} is the propagator for a quark of mass m_f on a gauge background $V_U e^{ieq_f A}$, and

$$I_{\mu,x}(m_f, eq_f) \equiv \frac{1}{4}\text{Tr}(M_f^{-1} D_\mu[iP_x V_U e^{ieq_f A}]). \tag{B5}$$

These calculations are performed gauge ensemble by gauge ensemble. The SU(3) configurations are generated at the isospin-symmetric point, where $m_u = m_d$ and $e = 0$. Leading-order QED and strong-isospin-breaking (SIB) corrections are then added as perturbations.

The leading QED corrections are obtained as follows. For each SU(3) gauge configuration, a stochastic U(1) photon field is generated with the noncompact, QED_L action. The U(1) fields are exponentiated into link variables that we multiply with the corresponding SU(3) links. The covariant Dirac operator is then inverted on each of the corresponding gauge configurations and the resulting quark propagators are appropriately contracted. The first

and second derivatives of the contractions, with respect to the unit of electric charge e, are obtained by computing the contractions for $+e$ and $-e$ and by taking the appropriate finite differences. The expressions for the first and second derivatives of the fermion determinant, in terms of quark propagators, are computed analytically. For the second derivative, around 2000 photon fields are used.

These first and second derivatives are appropriately combined, with the $e = 0$ contractions as well, to obtain all $O(e^2)$ contractions relevant for the current correlator. The desired correlators are then obtained by averaging the various contractions over the SU(3) and U(1) gauge configurations of each ensemble.

At leading order, the fermion determinant does not contribute SIB corrections. The valence contributions are obtained by computing the first derivative of the contractions, with respect to light-quark mass, and then multiplying it by the u and d quark mass difference. As above, the desired correlators are then obtained by averaging the various contractions over the SU(3) and U(1) gauge configurations of each ensemble.

In order to obtain physically relevant values, we perform global fits to the lattice spacing, quark mass, and electric charge dependence of $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. These fits take the form

$$Y = A + BX_l + CX_s + DX_{\delta m} + EX_{vv} + FX_{vs} + GX_{ss}, \tag{B6}$$

where the first three terms on the right-hand side of the equation provide the isospin-symmetric contribution to Y and the other four give the isospin-breaking effects, in a scheme specified via Eqs. (B7)–(B14). The X_l and X_s variables describe the deviation from the physical light and strange mass

$$X_l = \frac{M^2_{\pi_0}}{M^2_\Omega} - \left[\frac{M^2_{\pi_0}}{M^2_\Omega}\right]_{\text{ph}}, \tag{B7}$$

$$X_s = \frac{M^2_{K_\chi}}{M^2_\Omega} - \left[\frac{M^2_{K_\chi}}{M^2_\Omega}\right]_{\text{ph}}, \tag{B8}$$

where "ph" indicates the physical value. The remaining X variables measure the distance to the isospin-symmetric limit

$$X_{\delta m} = \frac{\Delta M^2_K}{M^2_\Omega}, \tag{B9}$$

$$X_{vv} = e^2_v, \tag{B10}$$

$$X_{vs} = e_v e_s, \tag{B11}$$

$$X_{ss} = e^2_s, \tag{B12}$$

where e_v and e_s are the valence and sea electric charges, respectively. The meson masses are defined as

$$M_{K_\chi}^2 \equiv \frac{1}{2}(M_{K_0}^2 + M_{K_+}^2 - M_{\pi_+}^2), \qquad (B13)$$

$$\Delta M_K^2 \equiv M_{K_0}^2 - M_{K_+}^2. \qquad (B14)$$

The coefficients $A, B, \ldots$ in Eq. (B6) are specific to the observable Y. They can depend on the lattice spacing and also on the X variables defined above; in particular, we use

$$A = A_0 + A_2[a^2(\alpha_s(1/a))^n] + A_4[a^2(\alpha_s(1/a))^n]^2$$
$$+ A_6[a^2(\alpha_s(1/a))^n]^3, \qquad (B15)$$

$$B = B_0 + B_2 a^2, \qquad (B16)$$

$$C = C_0 + C_2 a^2, \qquad (B17)$$

$$D = D_0 + D_2 a^2 + D_4 a^4 + D_l X_l + D_s X_s, \qquad (B18)$$

$$E = E_0 + E_2 a^2 + E_4 a^4 + E_l X_l + E_s X_s, \qquad (B19)$$

$$F = F_0 + F_2 a^2, \qquad (B20)$$

$$G = G_0 + G_2 a^2. \qquad (B21)$$

For any ensemble, a is given by the ratio of the Ω mass, in lattice units, to its experimental value. For the A coefficient of observables which give a large contribution to the final result, we include powers n of the strong coupling $\alpha_s(1/a)$ in the lattice-spacing dependence [99]. We take both the commonly used $n = 0$ (corresponding to the usual polynomial expansion in a^2) and $n = 3$, as suggested in Ref. [100]. For the strong coupling, we use its four-flavor $\overline{\text{MS}}$ value at scale $1/a$. We determine this value from the world average value of $\alpha_s(M_Z)$ [101] by running down from M_Z to $1/a$ in five-loop perturbation theory [102], taking into account four-loop threshold corrections [103] at the b-quark mass that are given in Ref. [101].

The choice of n is included as a source of systematic uncertainty in the analysis described below. In addition, we vary the polynomial order of the coefficients as an additional source of uncertainty. As required by the data, we take the A coefficient to be linear, quadratic, or cubic in $a^2 \alpha_s(1/a)^n$. For all other coefficients, the lattice-spacing dependence is assumed to be a function of a^2 only. In the D and E coefficients, up to quadratic dependencies are used; in all other cases only a linear one is needed. Depending on the fit qualities, some of these parameters will be set to zero.

We perform simulations at six different values of the lattice spacing, at a range of quark masses scattered around the isospin-symmetric point. We exclude zero or more of the coarsest lattice spacings from the analysis, and the resulting variation in the global fit provides another source of systematic uncertainty.

1. Estimating the covariance matrix

The results of this lattice calculation are subject to both statistical uncertainties arising from the Monte Carlo sampling of the QCD gauge fields and systematic uncertainties arising from choices made during the analysis. We estimate the variance and covariance arising from both types of uncertainty by using a combination of statistical resampling and histogramming of systematic variations [6,104,105].

The statistical covariance matrix is computed through jackknife resampling of the two-point, electromagnetic current-current correlator $C(t)$. The gauge field configurations are computed through a Markov chain algorithm that induces an autocorrelation between subsequent configurations. We suppress this effect by introducing a blocking procedure, where the individual configurations are grouped together into N_J blocks and then the average of $C(t)$ on each block is used in the resampling procedure instead of the value on each configuration. To simplify the analysis, we take N_J to be the same on all ensembles. Specifically, we take $N_J = 48$ which gives typical blocks of 100 or more configurations, larger than the autocorrelation time of the topological charge (which is ~ 20 on the finest ensembles).

Systematic uncertainties arise from a number of choices that must be made throughout the analysis. For example, the choice of appropriate fit ranges, the experimental value of the Ω-baryon mass used to set the scale, or the particular fit form used to perform the global fit. We call the global fit resulting from a particular set of these choices an analysis. In order to quantify the systematic uncertainties, we need to estimate the variance of the final value across all analyses arising from reasonable choices. To do this we follow the strategy used in Ref. [6].

As the full value of $a_\mu^{\text{LO-HVP}}$ and the window observable $a_{\mu,\text{win}}^{\text{LO-HVP}}$ are evaluated with similar techniques on the same set of gauge configurations, a number of choices in the analyses for $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ are similar and hence introduce correlations. Therefore, the procedure described in Ref. [6] had to be adapted to take into account these correlations. $\delta(\Delta\alpha_{\text{had}}^{(5)})$ is assumed to have negligible correlations with $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ because it is mostly determined by contributions from shorter distances.

The systematic uncertainties affecting any given analysis can be arranged into two groups: those that induce correlations and those that induce no correlation. The systematics that induce no correlation are further subdivided into those that enter with a flat weight and those that are weighted with the Akaike information criterion,

$$\text{AIC} \propto \exp\left[-\frac{1}{2}(\chi^2 + 2n_{\text{par}} - n_{\text{data}})\right], \qquad (B22)$$

where the χ^2, the number of fit parameters n_{par}, and the number of data points n_{data} describe the global fit corresponding to the given analysis. These relative AIC weights are employed between analyses differing in the choice of fit functions or in the lattice spacings included. For other systematics, including all systematics that induce correlations, a flat weight is employed. The choices that are assumed to introduce correlations are the fit ranges in the extraction of masses from correlation functions, the experimental input for the Ω mass, the taste improvement time ranges, and the choice of t_c in the bounding method for the extraction of $a_\mu^{\text{LO-HVP}}$. Details of this procedure and explanation of the individual contributions to the systematic uncertainties can be found in Ref. [6]. In the following, we denote with $A(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}})$ the result for some observable A, either $a_\mu^{\text{LO-HVP}}$ or $a_{\mu,\text{win}}^{\text{LO-HVP}}$, with a specific set of systematics. Here, ϕ is a label for a set of specific choices that will induce correlations between $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$, and the labels ψ^{flat} and ψ^{AIC} refer to choices that are made independently and are weighted either by a flat weight or by an AIC weight. In this notation, the histogram for a single observable A can be expressed as

$$H(A) = \sum_{\phi, \psi^{\text{flat}}, \psi^{\text{AIC}}} w(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}})$$
$$\times \mathcal{N}_1(A, \bar{A}(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}}), \sigma_{A(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}})}), \quad \text{(B23)}$$

where $\mathcal{N}_1(x, \mu, \sigma)$ is the probability density function of the normal distribution with mean μ and standard deviation σ. $\bar{A}(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}})$ refers to the mean value of the observable A in the analysis with the systematic variation indicated in parentheses. In the same sense, $\sigma_{A(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}})}$ corresponds to the statistical uncertainty on this analysis determined from the jackknife samples. The weight function

$$w(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}}) = \frac{\text{AIC}(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}})}{\sum_{\psi^{\text{AIC}}} \text{AIC}(\phi, \psi^{\text{flat}}, \psi^{\text{AIC}})} \quad \text{(B24)}$$

ensures a correct relative weight between the analyses. Here, $\text{AIC}(\cdots)$ denotes the AIC weight of the indicated analysis.

This procedure can be generalized to the study of the correlations between $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ by introducing a two-dimensional histogram,

$$H(A, B) = \sum_{\substack{\phi, \psi_A^{\text{flat}}, \psi_A^{\text{AIC}}, \\ \psi_B^{\text{flat}}, \psi_B^{\text{AIC}},}} w_A(\phi, \psi_A^{\text{flat}}, \psi_A^{\text{AIC}}) w_B(\phi, \psi_B^{\text{flat}}, \psi_B^{\text{AIC}})$$
$$\times \mathcal{N}_2(A, B, \bar{A}(\phi, \psi_A^{\text{flat}}, \psi_A^{\text{AIC}}),$$
$$\bar{B}(\phi, \psi_B^{\text{flat}}, \psi_B^{\text{AIC}}),$$
$$C_{\text{stat}}^{(1)}(\phi, \psi_A^{\text{flat}}, \psi_A^{\text{AIC}}, \psi_B^{\text{flat}}, \psi_B^{\text{AIC}})). \quad \text{(B25)}$$

Here, $C_{\text{stat}}^{(1)}$ refers to the statistical covariance matrix determined from the jackknife samples for each individual combination of analyses and $\mathcal{N}_2$ is the bivariate normal distribution with this covariance matrix. A and B refer to either $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ or individual flavor and contraction contributions to them.

We treated the systematic error introduced by the taste correction as a special case—both $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ are affected by this correction in a correlated way. For $a_\mu^{\text{LO-HVP}}$, four time ranges for taste improvements to the correlation function are used, starting at the Euclidean times 0.4, 0.7, 1.0, and 1.3 fm. But since the window observable $a_{\mu,\text{win}}^{\text{LO-HVP}}$ is not affected by the taste improvement in the last range, for $a_{\mu,\text{win}}^{\text{LO-HVP}}$ the two last ranges give the same results. Therefore, using all four ranges for both observables would bias the window observable. To avoid this, but still retain a measure of correlation, we considered 12 systematic combinations where, for each matrix, we fixed the range of the taste correction to $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. The starting times for the taste correction ranges in all 12 cases are shown in Table III. Then, we averaged the 12 covariance matrices.

As there are a large number of choices with many of these inducing a correlation, explicitly evaluating Eq. (B25) as a function of $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ is unfeasible. Instead, we prepared three one-dimensional histograms for the distributions of $X = a_\mu^{\text{LO-HVP}}$, $Y = a_{\mu,\text{win}}^{\text{LO-HVP}}$, and $Z = a_\mu^{\text{LO-HVP}} + a_{\mu,\text{win}}^{\text{LO-HVP}}$. From these three histograms, we determined the median and the interquantile distances d_X^σ, d_Y^σ, and d_Z^σ. Here, σ indicates the number of standard deviations the interquantile distances would correspond to,

TABLE III. Starting time for the taste corrections in the 12 systematic combinations used to estimate the correlation introduced by the taste correction. The first column contains the number of the systematic combination, the second column contains the starting time for the taste correction in $a_\mu^{\text{LO-HVP}}$, and the last column contains the starting time for the taste correction for $a_{\mu,\text{win}}^{\text{LO-HVP}}$.

No. syst. comb.	$t_{a_\mu^{\text{LO-HVP}}}^{\text{start}}/\text{fm}$	$t_{a_{\mu,\text{win}}^{\text{LO-HVP}}}^{\text{start}}/\text{fm}$
1	0.4	0.4
2	0.4	0.4
3	0.4	0.4
4	0.7	0.4
5	0.7	0.7
6	0.7	0.7
7	1.0	0.7
8	1.0	0.7
9	1.0	1.0
10	1.3	1.0
11	1.3	1.0
12	1.3	1.0

TABLE IV. Upper and lower bounds of the histograms used for determining the lattice covariance matrices in different flavor channels.

Channel	Lower bound $a_{\mu,\mathrm{chan}}^{\mathrm{LO\text{-}HVP}}$	Upper bound $a_{\mu,\mathrm{chan}}^{\mathrm{LO\text{-}HVP}}$	Lower bound $a_{\mu,\mathrm{chan,win}}^{\mathrm{LO\text{-}HVP}}$	Upper bound $a_{\mu,\mathrm{chan,win}}^{\mathrm{LO\text{-}HVP}}$
Light	600	680	190	230
Strange	51	55	25	29
Disconnected	−28.5	−8.5	−1.5	−0.9

in the case of normal distributions. Using these results, we construct the covariance matrices as

$$C = \begin{bmatrix} \frac{(d_X^\sigma)^2}{\sigma^2} & \frac{(d_Z^\sigma)^2-(d_X^\sigma)^2-(d_Y^\sigma)^2}{2\sigma^2} \\ \frac{(d_Z^\sigma)^2-(d_X^\sigma)^2-(d_Y^\sigma)^2}{2\sigma^2} & \frac{(d_Y^\sigma)^2}{\sigma^2} \end{bmatrix}. \qquad (B26)$$

To avoid any bias from the details of the histogram binning, we considered a large number of 2000 bins per histogram. The upper and lower bounds of the histogram were chosen such that they do not influence the results. These values can be found in Table IV. The resulting histograms for X and Y are plotted as black curves on the upper and right sides of the heat maps of Fig. 8. For illustrative purposes, we also calculated two-dimensional histograms with a smaller number of 1000 bins in each direction. These two-dimensional histograms are also shown Fig. 8. The bins were filled by sampling five random points for each summand in Eq. (B25). Note that the random sampling is only included in the visualization and has no impact on the end result. It is noteworthy that, in the case of the light-quark contribution to the window quantity, a double-peak structure appears. This structure is due to the variation of the parameter $n = 0, 3$ in Eq. (B15). The effect of the non-Gaussianity of the distribution introduced by this structure is taken into account by looking at both the σ and 2σ quantiles.

In order to separate the systematic uncertainties and correlations from the statistical ones, we repeat, in analogy to the procedure in Ref. [6], the full procedure while replacing C with λC in Eq. (B25). We choose $\lambda = 2$. Then, we solve the equations

$$C = C_{\mathrm{stat}} + C_{\mathrm{syst}}, \qquad (B27)$$

$$C_\lambda = \lambda C_{\mathrm{stat}} + C_{\mathrm{syst}}, \qquad (B28)$$

for C_{stat} and C_{syst}. Here, C_{stat} and C_{syst} are the statistical and systematic covariance matrices of $a_\mu^{\mathrm{LO\text{-}HVP}}$ and $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$, extrapolated to $a = 0$ and interpolated to the physical mass point.

2. Systematic and statistical uncertainties on the covariance matrix

The specific value taken by the estimated covariance matrix can have an important effect on the results described

in the main paper, in particular on those for χ^2, for the corresponding p-values, as well as those resulting from the inverse problem. As such, we found it important to quantify the uncertainties in this estimate. We identified two major sources of systematic uncertainty in our covariance estimate: non-normal distributions and correlation of continuum extrapolations. Each of these sources is described below, along with our approach to estimating the associated uncertainty.

Our jackknife estimate of the statistical covariance matrix for each analysis has itself a statistical uncertainty. We estimate this uncertainty by applying an approximate second-order resampling technique. We observe that the jackknife bias in $a_\mu^{\mathrm{LO\text{-}HVP}}$ and $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$ is small. Hence, we can approximate the true bootstrap distribution by bootstrapping the reconstructed samples $X_i = N_J \bar{X} - (N_J - 1)\bar{X}_i$, where $\bar{X}$ is the ensemble mean and $\bar{X}_i$ is the ith jackknife. We construct $N_B = 1000$ bootstraps in this way, and then on each bootstrap we compute the jackknife statistical covariance and perform the histogramming technique described above to obtain the full covariance matrix. In this way we compute the bootstrapped distribution of our final covariance matrix.

To obtain a covariance from the histogram, we make the assumption that the histogram is normally distributed. Under this assumption, the 1σ quantiles described above will fall one standard deviation below and above the median. Any deviation away from the normal distribution will hence introduce a systematic uncertainty. One example of such a non-normal distribution is the double-peak structure observed in Fig. 8(a). This double peak arises from the two-point systematic related to the power n of the strong coupling in the description of the lattice-spacing dependence of $a_{\mu,l,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$ [i.e., the connected light (ud) quark contribution to the window observable] given in Eq. (B15). It is an artifact of how we sample the errors and not intrinsic to the underlying distribution, but it still introduces an uncertainty in how the covariance is defined. To quantify this uncertainty, we also consider the covariance matrix obtained by considering the 2σ quantiles to be 2 standard deviations below and above the median. If the distribution were normal, the covariance matrix obtained in each of these ways would be identical. We take the difference between these two determinations of the covariance matrix as a two-point systematic.

In producing the histograms we allow the form of the continuum fit to vary independently for $a_\mu^{\mathrm{LO\text{-}HVP}}$ and $a_{\mu,\mathrm{win}}^{\mathrm{LO\text{-}HVP}}$,

(a) light contribution

(b) strange contribution
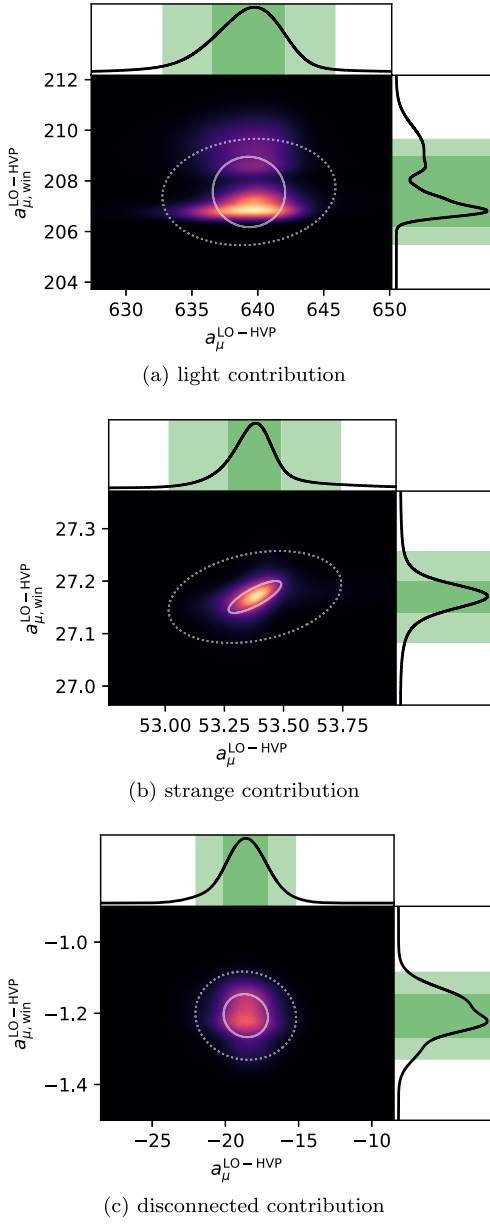

(c) disconnected contribution

FIG. 8. Histograms for the estimation of the total, statistical, and systematic lattice covariance matrices for the connected *ud* (a), the connected *s* (b), and disconnected (c) contributions to $a_\mu^{\text{LO-HVP}}$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$. These are obtained via the marginalization of Eq. (B25) to these individual contributions, as described in the text. Each of the three plots has a heat map in its center, illustrating the two-dimensional covariance distribution of the total combined uncertainty. As usual, colors from white to blue represent higher to lower probability densities. These densities are calculated using the random sampling discussed in the text. The solid and dashed ellipses in these maps indicate the 1σ and 2σ errors determined from the corresponding quantiles. In the upper and right panels around the heat maps, histograms of the marginalized distributions are shown. These are calculated exactly without random sampling. The two green bands show the 1σ and 2σ uncertainties determined by the corresponding quantiles of the distribution evaluated symmetrically around the median.

as there is no *a priori* reason that the fit forms should be the same. Hence, by construction we obtain no correlation between systematic uncertainties arising from the continuum limit of each quantity. However, since $a_\mu^{\text{LO-HVP}}$ can be considered to be the sum of $W = a_{\mu,\text{win}}^{\text{LO-HVP}}$ and its complement C, it is reasonable to surmise that there may be some correlation between their continuum limits. To estimate this potential correlation, we consider taking the continuum limits of W and C independently and adding them together to obtain $T = W + C = a_\mu$. In this scenario, let W have variance $(dW)_c^2$ associated with the continuum limit and similar for C. Then the variance of T is

$$(dT)^2 = (dT)_o^2 + (dW)_c^2 + (dC)_c^2, \qquad (\text{B29})$$

where $(dT)_o^2$ is the variance of T due to uncertainty sources other than the continuum extrapolation. Under these assumptions, the covariance matrix between T and W is

$$C_{TW} = C_{TW}^o + \begin{bmatrix} (dW)_c^2 + (dC)_c^2 & (dW)_c^2 \\ (dW)_c^2 & (dW)_c^2 \end{bmatrix}, \qquad (\text{B30})$$

where C_{TW}^o describes the covariance not associated with the continuum extrapolation. In practice, since the continuum limit of the combined $T = W + C$ is taken instead, the C component may interfere and partially spoil the correlation. As such we expect the true covariance to lie somewhere between this case and the case where this contribution is neglected. We take as an additional two-point systematic covariance matrices including either 0% or 100% of this estimated continuum limit contribution.

The connected contribution from valence charm quarks was computed in Ref. [106], and the data were not available to perform the covariance analysis described here. To address this, we assume that the correlation between $a_{\mu,c}^{\text{LO-HVP}}$ and $a_{\mu,c,\text{win}}^{\text{LO-HVP}}$ lies somewhere between 0% and 100% and treat the difference between these two extremes as an additional uncertainty. The charm contribution is a very small part of the total variance so this uncertainty is small. It is closely related to the uncertainty from the correlation of continuum extrapolations and as such is treated as part of the same two-point systematic.

Finally, the finite-size effects computed in Ref. [6] have a large contribution from the continuum limit of dedicated large-volume lattice studies, so once again the correlation of these errors is unknown. We conservatively take this correlation to lie somewhere between 0% and 100% and include this in the same two-point systematic as the other continuum limit correlations.

Taking into account both two-point systematics, we obtain four correlation matrices, each with an associated statistical uncertainty,

$$C_0 \equiv C_{\text{lat}}^{1\sigma,0\%} = \begin{bmatrix} 30.13(4.88) & -0.05(0.03) \\ -0.05(0.03) & 1.95(0.47) \end{bmatrix}, \qquad (\text{B31})$$

$$C_1 \equiv C_{\text{lat}}^{2\sigma,0\%} = \begin{bmatrix} 34.04(16.80) & 0.32(0.05) \\ 0.32(0.05) & 1.12(0.07) \end{bmatrix}, \quad \text{(B32)}$$

$$C_2 \equiv C_{\text{lat}}^{1\sigma,100\%} = \begin{bmatrix} 30.13(4.88) & 1.56(0.03) \\ 1.56(0.03) & 1.95(0.47) \end{bmatrix}, \quad \text{(B33)}$$

$$C_3 \equiv C_{\text{lat}}^{2\sigma,100\%} = \begin{bmatrix} 34.04(16.80) & 1.94(0.05) \\ 1.94(0.05) & 1.12(0.07) \end{bmatrix}. \quad \text{(B34)}$$

By performing the fits described in the main text with each bootstrap sample for each of the four representative covariance matrices, we obtain the uncertainty in the rescaling factors and observed χ^2 values arising from our uncertainty in the lattice covariance matrix.

It is important to note that the statistical covariance matrices $C_{\text{stat}}^{(1)}$ used in the histogramming procedure are determined from jackknifes over $N_J = 48$ binned "measurements." Therefore, the statistical uncertainty on these covariance matrices is determined by N_J, rather than by the number of independent samples used in the determination of the central values. This number is smaller than the total number of independent samples available in the dataset, even accounting for autocorrelations. This means that with the available data, a different sampling procedure (in particular, a larger number of bins) could produce an estimate of the statistical covariance with smaller statistical errors. However, this conservative approach to binning ensures that the systematic errors arising from autocorrelations within the data are negligible compared to these statistical errors. In addition, this choice provides consistency with the analysis of Ref. [6]. As the results of Sec. V show, the statistical error this conservative binning produces is still sufficiently small that it does not quantitatively affect the conclusions drawn in the present work.

APPENDIX C: WEIGHTED AVERAGE PROCEDURE IN PRESENCE OF UNCERTAINTIES ON UNCERTAINTIES AND CORRELATIONS

The minimum of the χ^2 function in Eq. (19) with respect to γ is obtained for

$$\gamma = \frac{\sum_{j,k} [(C_{\text{lat}}^{\tilde{\gamma}} + C_{\text{R}}^{\tilde{\gamma}})^{-1}]_{kj} \cdot \tilde{\gamma}_j}{\sum_{l,m} [(C_{\text{lat}}^{\tilde{\gamma}} + C_{\text{R}}^{\tilde{\gamma}})^{-1}]_{lm}}, \quad \text{(C1)}$$

which represents a weighted mean of the input $\tilde{\gamma}_j$ values. However, it is known that this approach can yield biased results (see e.g., Ref. [75]).

According to the Gauss-Markov theorem, this χ^2 minimization is equivalent to the minimization of the variance of a weighted average, with the sum of weights constrained to unity (see e.g., Refs. [107–109]). It is for this reason that this method is also called the "best linear unbiased estimate" (see e.g., Ref. [107] and references therein).

However, the word "unbiased" in the name above needs to be interpreted with care, as it implicitly involves several assumptions about the statistical model being employed. Indeed, in the presence of strong correlations among the inputs, the weights in Eq. (C1) can be negative or larger than unity, yielding an average value that can be even outside the range of the input values, which may be problematic. Chapter 7 of Ref. [108] points out that the measurements are likely to lie on the same side of the true value in the presence of strong positive correlations. However, it is also indicated that this interpretation is subject to the implicit assumptions about the knowledge of the covariance matrix, as discussed in more detail below.

Reference [75] points out that a bias in this procedure is caused by the input covariance matrix because of "the linearization on which the usual error propagation relies." While the χ^2 definition of Eq. (19) treats all the uncertainties as absolute (i.e., corresponding to additive effects), at least some of them should be treated as relative[19] (i.e., corresponding to multiplicative effects) [76].

Furthermore, another problem of this approach is that Eq. (19) involves an ill-behaved inversion of the covariance matrix. Indeed, a small change in the covariance matrix can have an important impact on its inverse, hence on the weights and the resulting average. Without special techniques, the inverse cannot be computed, nor can the χ^2 minimization be performed [76]. While modern numerical algorithms allow one to address the technical aspect of this matter, some more fundamental problems related to the evaluation of the covariance matrix still persist (see below).

In Ref. [75] it is discussed that a possible solution to these problems consists of introducing, in the χ^2 definition, one nuisance parameter for each correlated systematic uncertainty, together with the corresponding constraint terms. Doing so, for a normalization uncertainty, a scaling factor is applied to both data and uncorrelated uncertainties. An equivalent approach introduces the corresponding nuisance parameter as a scaling factor on the theoretical prediction [76,110]. An alternative method for the treatment of relative uncertainties consists of scaling them to the fitted value minimizing the χ^2, rederiving the corresponding covariance matrix, and iterating (see Refs. [75,111] and references therein).[20] However, we note that the rescaling of

---

[19]That is the case e.g., for the luminosity uncertainty of the experimental measurements.

[20]However, for more general applications, such a procedure faces some important challenges when the rescaling is applied for the statistical uncertainties of unfolded measurements. Indeed, in such a case the migrations of events induce nontrivial correlations between the various bins, which imply that the statistical uncertainty in a given bin also depends on (i.e., scales as a function of) the expected numbers of events in other bins. Furthermore, the regularization procedures typically used in the unfolding also induce nonlinear effects in the uncertainty propagation, making the uncertainty rescaling of such measurements even more challenging.

the uncertainties has little relevance in the current use cases, when there is little difference between the $\tilde{\gamma}_j$ values to be combined or when one of them is more precisely determined than the others and dominates the average (see Sec. V B). It is also worth recalling that, if all the nuisance parameters correspond to (are treated as) absolute systematic uncertainties, each of them acting as a coherent additive shift of the theoretical prediction, the approach using nuisance parameters is equivalent to the one of the "standard" χ^2 with correlations from Eq. (19) (see Refs. [112–117]).

The approach using nuisance parameters in the fit often leads to reduced systematic uncertainties, as they are constrained by the χ^2 minimization. However, this reduction assumes and strongly relies on a perfect knowledge of the phase-space dependence and correlations of the systematic uncertainties, information explicitly used in the χ^2 definition. The same assumptions and features are implicit for the standard χ^2 with correlations in Eqs. (12), (19), (E8), and (E10), in particular for minimizing the variance of the average, as discussed above. Indeed, the need for precisely known covariance matrices, mentioned in Refs. [14,118], becomes explicit in the equations above.

Building upon remarks made in the context of ATLAS jet performance and cross section studies [119–121] (see also the earlier discussion in Ref. [122]), these strong assumptions were also questioned for the e^+e^- annihilation data used for the evaluation of $a_\mu^{\text{LO-HVP}}$, since there are clear indications that the amplitudes of systematic uncertainties and their correlations are generally impacted by uncertainties themselves [3,5,123].[21] Indeed, the size of the systematic uncertainties of the various measurements and the correlations of a systematic uncertainty component impacting several measurements, as well as the correlations between the systematic uncertainties impacting a given measurement are never really measured, but rather *estimated*. For example, the tensions between the input measurements observed in several channels (e.g., between *BABAR* and KLOE in the $\pi^+\pi^-$ channel) are a direct indication of underestimated uncertainties for the measurements and motivated the inclusion of an extra systematic uncertainty for the resulting dispersive integrals [3] (see also Appendix A). Such two-point systematic uncertainty, although necessary when using the currently available inputs, has unknown correlations between different phase-space regions. Therefore, it comes at the price of bringing some fundamental limitations for the treatment of correlations in Eqs. (12), (19), (E8), and (E10). The recent CMD-3 measurement in the $\pi^+\pi^-$ channel [15], which

shows some significant tension with the previous precise measurements of the same channel, emphasizes even further the existence and the need for a careful treatment of the uncertainties on uncertainties. Furthermore, such limitations in the statistical treatment also originate from the uncertainties on uncertainties and on correlations for the lattice QCD calculations (see Appendix B). Therefore, the statistical procedures employing such inputs (or other quantities derived using them) should avoid overestimating the precision with which the uncertainties and their correlations are known, a conservative uncertainty treatment being preferable. This is especially important since the uncertainties on the uncertainties and on their correlations are not available in the current publications of hadronic spectra.

In order to address the limitations related to the uncertainties on the systematic uncertainties and on their correlations, we use a procedure which consists of performing a weighted average, with the weights of the input $\tilde{\gamma}_j$ values being proportional to the inverse of their squared total uncertainties. This is equivalent to using only the diagonal elements of the covariance matrices (i.e., setting the other elements of the matrices to zero) in Eq. (C1) when deriving the weights. The weights used in this procedure are in the range between 0 and 1 and are on the larger end for more precise contributions. This part of the procedure matches the corresponding one in Refs. [126,127]. At the same time, the full information of the input uncertainties, with their correlations, is propagated to the result of the weighted average (as e.g., in Ref. [127]) and is used in the χ^2 when assessing the level of agreement between the inputs and the average value. A similar fitting procedure, using only the diagonal uncertainties in the χ^2, followed by a propagation of the full set of uncertainties and their correlations, has been successfully used in Ref. [3] for combining the experimental measurements in the $e^+e^- \to \pi^+\pi^-$ channel.[22]

APPENDIX D: STABILITY OF THE RESULTS OF SEC. V B WITH RESPECT TO THE AVERAGING PROCEDURE

In this appendix we discuss the use of an alternative averaging procedure to the one employed in Sec. V B, to obtain the results for changes to the experimental R-ratio that could explain the differences between the lattice and data-driven results for $a_\mu^{\text{LO-HVP}}$, $a_{\mu,\text{win}}^{\text{LO-HVP}}$, and $\delta(\Delta\alpha_{\text{had}}^{(5)})$. In that section, we perform the average defined in Eq. (C1). That is, when we derive the averaging weights, we replace the full covariance matrix

---

[21]More recently, a similar remark has been independently made in Ref. [124]. The remarks made in the context of the muon $g-2$ studies are also relevant when using the data of hadronic spectra for other applications, e.g., for the extraction of the strong coupling employing the Adler function [125].

[22]In Ref. [122] one can find some further discussions on other unbiased combination procedures with realistic uncertainty estimates, taking into account the full information of the uncertainties and correlations of the inputs.

$(C_{\text{lat}}^{\tilde{\gamma}} + C_{\text{R}}^{\tilde{\gamma}})$, obtained by a linear propagation of lattice and experimental R-ratio uncertainties and correlations, by its diagonal form with no correlation. Here we consider what the effect of including these correlations is on our results.

Generically, the use of the full covariance matrices for determining the averaging weights in Eq. (C1) changes these weights significantly, some becoming negative and others larger than unity (see Sec. IV B). However, here, their deviations from zero or from unity are observed to be smaller than 0.06. Furthermore, in such cases either the $\tilde{\gamma}_j$ values from Eq. (18) do not differ much[23] or one of them is more precise and dominates the average. Thus, the values of the average $\gamma = \gamma_1$ are similar for the two averaging methods, within 12% in the most extreme case, and in most cases within a few percent or less.

Even if this alternative approach minimizes the χ^2 in Eq. (19) (and the uncertainty of the rescaling percentage δ_1, which is propagated from the covariance matrices of the lattice QCD and dispersive results), it only reduces it by 15% (9%) or less, compared to the values obtained with weights proportional to the inverse of the $\tilde{\gamma}_j$ uncertainties squared, which are used in the nominal approach to produce Table II. Therefore, the p-values obtained with the two approaches are also similar, leading to similar conclusions about the level of compatibility between the lattice and dispersive moment integrals for all the normalization shift scenarios considered in this paper.

The statistical variance and quantiles of the rescaling percentage δ_1, obtained from the bootstrap variations of the lattice covariance matrices, are most often enhanced (by up to 540%) in the alternative averaging method compared to the one used to obtain the results given in the main text, while in some rare cases they are reduced (by up to 87%). At the same time, the changes of δ_1 under systematic variations of the lattice covariance matrix are also generally broader for the alternative averaging method.

Concerning the statistical variance and quantiles of the *uncertainty* on δ_1 (propagated from the covariance matrices of the lattice QCD and dispersive results), they are most often enhanced (by up to 133%) in the alternative averaging method compared to the nominal one, while in some rare cases they are reduced (by up to 99%). In all cases, as for the nominal method, the uncertainty of the rescaling percentage δ_1 remains precisely determined when using the alternative averaging method.

---

[23]The $\tilde{\gamma}_j$ values used for a given averaging differ by typically only 12% or less, except for cases when the normalization fit is done in restricted $\sqrt{s}$ regions, below 0.63 GeV or above 1.8 or 3.0 GeV. In that case, the shape differences among the kernels of the various moment integrals have a more important impact for Eq. (18) and the differences between the $\tilde{\gamma}_j$ values can be as large as 30%–80%.

On the other hand, the variance and quantiles of the χ^2 values, due to the same bootstrap fluctuations as above, are generally reduced when using the alternative averaging method (by up to 44%), while in some rare scenarios they are slightly enhanced (by less than 3%). The range covered by the χ^2 values under systematic variations of the lattice covariance matrix is generally somewhat narrower for the alternative averaging method.

Some of the features described above, for the comparison between the two averaging methods, are illustrated by comparisons between the plots in the first two rows of Figs. 4–7. These features are consistent with the fact that while both procedures are sensitive to the uncertainty on the uncertainty in the determination of the weights, only the alternative approach, which uses the full covariance matrices for determining the averaging weights, relies on correlations for computing the weights to minimize the χ^2 and the uncertainty of the rescaling percentage δ_1.

APPENDIX E: GOING BEYOND THE RESCALING IN A SINGLE $\sqrt{s}$ INTERVAL

To go beyond the simple scenarios and interpretations of Appendix C, more window observables are required from the lattice, with the corresponding ones from the experimental R-ratio, split up into $\sqrt{s}$ intervals. These can be complemented by other quantities related to the HVP, such as the hadronic vacuum polarization function $\hat{\Pi}(Q^2)$ at various Q^2 accessible to the lattice. Again, all of these quantities must be obtained with a full description of the relevant covariances.

1. General formalism for comparing and possibly combining lattice and data-driven results

Consider a function $M_{\text{R}}(s, \alpha_i)$ that depends on s and on parameters α_i. It is chosen so as to provide a good description of the experimental R-ratio, or a rebinned version of it, once the α_i are appropriately adjusted. Now, consider a modified model $M_{\text{lat}}(s, \alpha_i, \beta_j)$ that depends on the additional parameters β_j. This new model represents a guess at how the experimental R-ratio could be modified to reproduce the lattice observables of interest, once integrated over s with the appropriate weights. We then define the ratio of these two models as

$$F_{\text{lat}\div\text{R}}(s, \beta_j) \equiv M_{\text{lat}}(s, \alpha_i, \beta_j)/M_{\text{R}}(s, \alpha_i), \qquad \text{(E1)}$$

which is chosen not to depend on the α_i parameters.[24] For instance, these functions can be physics-motivated models, spline-based parametrizations, orthogonal polynomials, histograms of the bins used for the experimental R-ratio,

---

[24]This factorization is used here just for an intuitive simplification of the explanation, while the methods discussed below can also be applied to more general $M_{\text{lat}}(s, \alpha_i, \beta_j)$ models.

or with some different binning, etc. For either the R-ratio or lattice model functions, we define the integrals of interest as

$$a_{jb}^{M_{R/lat}} \equiv \int_{I_b} ds \tilde{K}_j(s) M_{R/lat}(s), \quad (E2)$$

where the α_i and β_j model parameters are kept implicit. As above, I_b can be either some restricted $\sqrt{s}$ interval or it can cover the full range from threshold to ∞. In the latter case we drop the index b for the integral label, which also keeps the notations consistent with e.g., Eq. (15). Note that if $M_{R/lat}$ is a histogram, then the integral in Eq. (E2) becomes a sum over bins.

For the lattice constraints of the model integrals one can write the χ^2 function as

$$\chi_{lat}^2 = \sum_{j,k} [a_j^{lat} - a_j^{M_{lat}}][C_{lat}^{-1}]_{jk}[a_k^{lat} - a_k^{M_{lat}}], \quad (E3)$$

while for the constraints from the experimental R-ratio, one can use approaches based on different χ^2 definitions, such as

$$\chi_{R,1}^2 = \sum_{j,k} [a_j^R - a_j^{M_R}][C_R^{-1}]_{jk}[a_k^R - a_k^{M_R}], \quad (E4)$$

$$\chi_{R,2}^2 = \sum_{(jb),(kc)} [a_{jb}^R - a_{jb}^{M_R}][C_R^{-1}]_{(jb)(kc)}[a_{kc}^R - a_{kc}^{M_R}], \quad (E5)$$

$$\chi_{R,3}^2 = \sum_{b,c} [R_b - M_{R,b}][C_{R,bin}^{-1}]_{bc}[R_c - M_{R,c}], \quad (E6)$$

where, as above, we indicate whether the covariance matrix of the a_j or a_{jb} moment integrals correspond to those of the lattice (data-driven) approach via the subscript "lat" ("R"). In addition, we call $C_{R,bin}$ the covariance matrix that corresponds to (a possibly rebinned version of) the experimental R-ratio. Indeed, Eq. (E4) concerns dispersive integrals on the full s range, Eq. (E5) includes further information on the shape of the R-ratio through the integrals computed on various I_b intervals, while Eq. (E6) uses directly the experimental R-ratio (without any further convolution), either within its original bins or after merging the bins within larger intervals. Based on these χ^2 functions, one can consider several approaches for comparing and/or combining the R-ratio and lattice results.

A first approach can consist of constraining the parameters α_i of $M_R(s, \alpha_i)$ by minimizing one of the χ_R^2 functions in Eqs. (E4) and (E6), followed by a determination of the β_j parameters of $M_{lat}(s, \alpha_i, \beta_j)$ through a minimization of χ_{lat}^2 from Eq. (E3). In this two-step approach, the α_i parameters are held fixed when the second fit is performed and therefore they are not constrained by lattice results. Still, their uncertainties and correlations are propagated from the

first step to the second, e.g., by replacing C_{lat} in Eq. (E3) with the appropriate covariance matrix.

A second approach can consist of simultaneously constraining the α_i and β_j parameters of the model, by minimizing $\chi_{lat}^2 + \chi_R^2$, with either of the definitions in Eqs. (E4) and (E6) for the contribution from the experimental R-ratio. We note that this minimum is generally lower than the $\chi_{lat}^2 + \chi_R^2$ value obtained for the models fitted in the two-step approach discussed above. While the minimization of the χ^2 defined in Eq. (12), according to Eqs. (13) and (14), represents a combination of the lattice and R-ratio results in the space of the a_j integrals, without any extra constraints, the approach of fitting the α_i and β_j parameters by minimizing $\chi_{lat}^2 + \chi_R^2$ represents a similar combination in the R-ratio space, under the constraint of the $M_{lat}(s, \alpha_i, \beta_j)$ and $M_R(s, \alpha_i)$ models, respectively.

In a third approach, one can first minimize χ_{lat}^2 from Eq. (E3) with respect to the free parameters of the model $M_{lat}(s, \alpha_i, \beta_j)$, obtaining $\hat{\alpha}_i$ and $\hat{\beta}_j$. One then considers a modified version of that lattice model which accounts for possible differences with the R-ratio. Using the notation of Eq. (E1), we define this model as $\hat{M}_{lat}(s, \hat{\alpha}_i) \equiv M_{lat}(s, \hat{\alpha}_i, \hat{\beta}_j)/F_{lat \div R}(s, \hat{\beta}_j)$. This model is subsequently binned via averaging within chosen $\sqrt{s}$ intervals. We call the resulting histogram $\hat{M}_{lat}^{bin}(s, \hat{\alpha}_i)$ and its covariance matrix $\hat{C}_{lat,bin}$. In addition, we consider a model $M_R(s, \alpha_i)$ for the R-ratio and its version $\tilde{M}_R(s, \alpha_i)$, possibly rebinned to match the binning of $\hat{M}_{lat}^{bin}(s, \hat{\alpha}_i)$. The parameters α_i are to be fitted simultaneously to the experimental R-ratio and $\hat{M}_{lat}^{bin}(s, \hat{\alpha}_i)$ by minimizing

$$\chi^2 = \sum_{b,c} [\hat{M}_{lat,b}^{bin} - \tilde{M}_{R,b}][\hat{C}_{lat,bin}^{-1}]_{bc}[\hat{M}_{lat,c}^{bin} - \tilde{M}_{R,c}]$$
$$+ \sum_{b,c} [R_b - M_{R,b}][C_{R,bin}^{-1}]_{bc}[R_c - M_{R,c}], \quad (E7)$$

where the first sum runs over the lattice bins and the second runs over the ones of the experimental R-ratio. This effectively combines the lattice and experimental R-ratio results.

In this third approach, consider the case in which M_{lat} is taken to be a binned histogram with the number of bins equal to the number of (linearly independent) lattice moment integrals, the count of each bin being a free parameter to be determined. Then the minimization of χ_{lat}^2 of Eq. (E3) is equivalent to an "unregularized" unfolding toward the space of the R-ratio (in the sense that no other regularization besides the binning choice itself is being applied). Indeed, this is equivalent to solving a system of constraints through a matrix inversion approach (see e.g., Chap. 11 of Ref. [108]). However, the parametrization of $M_{lat}(s, \alpha_i, \beta_j)$ can be used to inject an effective regularization into the inverse problem. The unregularized approach avoids biases related to the

matrix inversion. However, it may induce large variances and strong (anti)correlations between the determined quantities (i.e., fitted R-ratio values in various bins) and hence a large hierarchy among the eigenvalues of the corresponding covariance matrix. In that case, any further use of such results would require a very precise determination of the covariance matrix which, in turn, would necessitate very precise knowledge of the input covariances. Alternatively, it is generally preferable to use regularized unfolding methods, aiming for a trade-off between moderate statistical variances and systematic biases induced through the regularization.[25]

It is to be noted that the three approaches described above, with their possible variants, imply different hypotheses about the validity of the fitted models and the way in which their parameters are constrained. The resulting $\chi^2/d.o.f.$ values represent tests of these hypotheses, providing information about the compatibility between these models and the input data, as well as about the compatibility between the lattice and R-ratio inputs themselves. In the presence of tensions, one can consider enhancing the uncertainties of the combination result, following e.g., the approaches from Refs. [3,71–73], summarized in Appendix A. This can be achieved by using the information from some (partial/local) χ^2 and/or adding some extra uncertainties to account for systematic deviations between the inputs.

Moreover, while Eqs. (E3) and (E6) are written using the full covariance matrices, the discussion of Appendix C concerning the treatment of the uncertainties on the uncertainties and on the correlations is relevant here too and has to be implemented accordingly. For instance, one can perform the χ^2 fits for determining the model parameters using only the diagonal elements of the covariance matrices, while the full information on the uncertainties and their correlations is propagated to evaluate the uncertainties of the model parameters, their correlations, and the fit quality.

2. The case of rescaling factors in either a single or multiple $\sqrt{s}$ intervals

To make contact with what was done in the main body of the paper, we consider here the rescaling in a single $\sqrt{s}$ interval that was introduced in Sec. IV B and employed in Sec. V B. It can be retrieved as a special case of the general approaches discussed above. Indeed, one can consider a "model" $M_R(s, \alpha_i)$ that consists of a histogram with the same binning as the experimental R-ratio itself. The α_i are parameters to be determined and correspond to the bin

counts. Employing the first, two-step approach described below Eq. (E6), with χ_R^2 given by $\chi_{R,3}^2$ from that equation, the first minimization step yields the bin counts with $\chi_R^2 = 0$. Moreover it propagates the covariance matrix of the experimental R-ratio to the histogram of $M_R(s, \alpha_i)$, i.e., to the parameters α_i. The model is completed with the function $F_{\text{lat}\div R}(s, \gamma)$, which is equal to a rescaling factor γ in a subset of the $\sqrt{s}$ intervals, while it is equal to unity for the complementary subset of bins. This yields a model $M_{\text{lat}}(s, \alpha_i, \gamma)$ that implements the prescription from Eq. (17). In particular, this model ensures that the ratios of integrals $a_{jb}^{M_R}/a_{kb}^{M_R}$ and $a_{jb}^{M_{\text{lat}}}/a_{kb}^{M_{\text{lat}}}$ are equal to a_{jb}^R/a_{kb}^R. In turn, this guarantees that these ratios are independent of the parameter γ, for any j and k integrals and any interval I_b. In other words, these ratios of integrals are driven by the corresponding kernels convoluted with (i.e., averaged over) the shape of the experimental R-ratio, while the changes of normalization through the factor γ impact coherently all the $a_{jb}^{M_{\text{lat}}}$ integrals on the rescaled $\sqrt{s}$ interval. Equation (E3), with the model $M_{\text{lat}}(s, \alpha_i, \gamma)$ and C_{lat} completed with a propagation of the covariance matrix of the parameters α_i, yields Eq. (19) [after a change of variable similar to Eq. (18), together with the corresponding linear uncertainty propagation].

The approach discussed above can be generalized by applying the constraint from Eq. (16) at the level of the corresponding R-ratio and lattice models, with multiple rescaling factors γ_b applied in the $\sqrt{s}$ intervals I_b. This corresponds to minimizing

$$\chi^2 = \sum_{j,k} \left[a_j^{\text{lat}} - \sum_b \gamma_b a_{jb}^R \right] [(C_{\text{lat}} + \tilde{C}_R)^{-1}]_{jk}$$
$$\times \left[a_k^{\text{lat}} - \sum_c \gamma_c a_{kc}^R \right], \tag{E8}$$

provided that the number of linearly independent moment integrals is greater or equal to the number of fitted γ_b rescaling factors. In this equation, $(C_{\text{lat}} + \tilde{C}_R)$ is the covariance matrix corresponding to $[a_j^{\text{lat}} - \sum_b \gamma_b a_{jb}^R]$, depending hence also on the γ_b parameters. As above, this implies assigning a prefit uncertainty to the $M_R(s, \alpha_i)$ model, which is then propagated via the corresponding covariance matrix.[26] In addition, if the I_b intervals correspond to the original experimental R-ratio bins and their number is equal to that of linearly independent lattice observables a_j^{lat}, then the rescaling hypothesis becomes

---

[25]Such regularization limits the number of free parameters that are to be determined, allowing us to also address the worries that are sometimes expressed concerning the ill-defined nature of the problem when using χ^2-based approaches (see e.g., Refs. [35–37,39,42,43]).

[26]These methods can also be seen as fits of normalization factors for templates of moment integrals derived through the dispersive approach, used to describe the corresponding lattice results, as well as possible. For a discussion on the template fitting methods, with the various approaches for treating the uncertainties of the fitted distributions and of the templates themselves, see e.g., Refs. [128–130] and references therein.

exact. In that case, one can have as many normalization rescaling factors as there are bins, allowing for a complete, unregularized reconstruction of the lattice R-ratio with the same granularity as that of the experimental R-ratio. Please see the discussion after Eq. (E7) for more details. It is also worth noting that, in the case when all the $\gamma_b = 1$, the χ^2 from Eq. (E8) matches the one of Eq. (14), providing a direct compatibility check between the lattice and the dispersive results.

3. Single or multiple rescaling factors for fitted moment integrals

An alternative to Eq. (E8), which includes multiple rescaling factors γ_b, is given by the model

$$
\begin{cases}
a_j^{\text{lat}} = \sum_b \gamma_b a_{jb} \\
a_{jb}^{\text{R}} = a_{jb}
\end{cases}, \qquad (E9)
$$

which is assumed to describe the inputs a_j^{lat} and a_{jb}^{R} within the uncertainties. The second line provides a trivial model for the R-ratio contributions a_{jb}^{R}, from the $\sqrt{s}$ intervals I_b, defined after Eq. (15) and the first a rescaling model for the corresponding lattice observables defined in Sec. III. Of course, the number of lattice observables a_j^{lat} must be larger than the number of free rescaling parameters. This is a special case of Eq. (E1), as given in Eqs. (E3) and (E5) without, however, the constraints of the models $M_{\text{R/lat}}$ for the lattice and R-ratio spectral functions.

Thus, to determine the parameters of this model, taking into account all correlations, one can minimize

$$
\chi^2 = \sum_{j,k} \left[a_j^{\text{lat}} - \sum_b \gamma_b a_{jb} \right] [C_{\text{lat}}^{-1}]_{jk} \left[a_k^{\text{lat}} - \sum_c \gamma_c a_{kc} \right] \\
+ \sum_{(jb),(kc)} [a_{jb}^{\text{R}} - a_{jb}][C_{\text{R}}^{-1}]_{(jb)(kc)}[a_{kc}^{\text{R}} - a_{kc}], \qquad (E10)
$$

with respect to the γ_b and a_{jb}. Again, lattice results are assumed to have no correlations with those obtained from the experimental R-ratio. The uncertainties on the parameters γ_b and a_{jb} can be obtained from the Hessian or by using pseudoexperiments. One can also consider the p-value obtained from the minimum value of the χ^2 and the system's number of degrees of freedom, which is the number of lattice window observables minus the number of rescaling parameters. This p-value indicates the consistency of the available lattice and R-ratio input with the rescaling hypothesis above.

Now, suppose that the rescaling model of Eq. (E9) is well justified theoretically and that the p-value of the fit is acceptable. Then, the resulting parameter values provide a description of the lattice and R-ratio observables, via the right-hand sides of Eq. (E9), that combines the information contained in the a_i^{lat} and a_{jb}^{R}. When the number of

observables and intervals increases, so will (anti)correlations between them, leading to large variances on the results, well known in inverse problems. Of course, if the number of rescaling parameters is the same as the number of lattice observables, then the system of equations (E9) can be solved exactly, and one obtains $a_{jb} = a_{jb}^{\text{R}}$ as well as the values of the γ_b obtained by minimizing Eq. (E8), in the same situation.

Beyond the possibility of combining lattice and R-ratio results for a_j^{lat} and a_{jb}^{R} via the model of Eq. (E9), the approaches of Sec. IV B and of Eqs. (E9) and (E10) differ in that the former is based on a model that is linear in the fitted parameters [Eq. (18)], with a linear propagation of uncertainties, while the latter refers to a nonlinear model [Eq. (E9)], but with no linear propagation of uncertainties. They are expected to be equivalent in the limit that the rescaling factors are such that the $|\gamma_b - 1|$ are small and as long as the relative uncertainties on the a_{jb}^{R} are also small. We have checked this analytically and numerically, in the case of a single rescaling factor γ (either fitted or fixed to 1) for the results presented in Sec. V. The nonlinear effects in the uncertainty propagation mentioned above were also studied and found to be numerically small. The nonlinearities in the deviation from 1 of the rescaling parameter γ are also generally small, though somewhat larger when the model does not describe the lattice and R-ratio results well.

The two approaches also answer slightly different questions. The one of Appendix E 2 corresponds to a rescaling of the dispersive integrals derived from the experimental R-ratio in different $\sqrt{s}$ subintervals I_b, while the one presented here corresponds to a rescaling of the fitted a_{jb}, which may deviate from the input a_{jb}^{R}.

Moreover, the approach described here provides an unregularized reconstruction of both the lattice observables, via the first line of Eq. (E9), and of the timelike contributions of $\sqrt{s}$ intervals I_b, as given in the second line of Eq. (E9). The resolution with which this sort of reconstruction can be performed is usually limited by (anti)correlations between the various observables that generally lead to large variances on the results.

Note that, when all the γ_b factors are set to unity, one can perform the change of variable $a_{j\bar{b}} = a_j - \sum_{b \neq \bar{b}} a_{jb}$ in Eq. (E10), for an arbitrary index $\bar{b}$. The covariance matrix $\bar{C}_{\text{R}}$, resulting from the propagation of C_{R} through this change of variable, can be decomposed into a set of uncertainties that are either uncorrelated ($\sigma_{j(b)}$) or that correlate ($s_{j(b)}^l$) the $a_{j(b)}^{\text{R}}$,[27]

$$
[\bar{C}_{\text{R}}]_{j(b),k(c)} = \sigma_{j(b)}^2 \cdot \delta_{j(b),k(c)} + \sum_l s_{j(b)}^l \cdot s_{k(c)}^l, \qquad (E11)
$$

---

[27]In quantities such as $a_{j(b)}^{\text{R}}$, the parentheses indicate that we are considering both a_j^{R} and a_{jb}^{R} with $b \neq \bar{b}$.

with the index l running over the corresponding independent uncertainty components [117]. This enables an equivalent rewriting of the χ^2 function from Eq. (E10) as

$$\chi^2 = \sum_{j,k}[a_j^{\text{lat}} - a_j][C_{\text{lat}}^{-1}]_{jk}[a_k^{\text{lat}} - a_k]$$
$$+ \min_{\beta_l}\left\{\sum_{j(b\neq\bar{b})}\frac{[a_{j(b)}^{\text{R}} + \sum_l s_{j(b)}^l \cdot \beta_l - a_{j(b)}]^2}{\sigma_{j(b)}^2} + \sum_l \beta_l^2\right\},$$
(E12)

using nuisance parameters $s_{j(b)}^l$ to account for the correlated uncertainties [112–117].[28] While the a_j parameters are constrained by both the lattice and the dispersive inputs, the $a_{j,b\neq\bar{b}}$ parameters remaining after the change of variable only enter the dispersive part of the χ^2 function. The minimization of this χ^2 function with respect to $a_{j,b\neq\bar{b}}$ allows us to exactly cancel the corresponding contribution to the χ^2, regardless of the contributions of the shifts of the nuisance parameters away from zero, which are parametrized by the β_l and induced by the a_j^{R} and a_j^{lat} inputs. Switching back from the representation in terms of nuisance parameters to the one with the covariance matrix, this time in the subspace of a_j parameters, allows us to show that the minimum of Eq. (E10) with respect to $a_{j,b\neq\bar{b}}$ reduces to Eq. (12), while their global minimum is provided by Eq. (14). This also implies that for $\gamma_b = 1$ the minimum of the χ^2 function from Eq. (E10) is stable with respect to the choice of the I_b intervals, a feature also observed numerically.

4. Parametrizing possible shape changes of the R-ratio suggested by lattice results

Once sufficient lattice input is available, one can refine the analysis in other ways. Since our results suggest that the ρ peak could be responsible for the disagreement with the data-driven approach, one could imagine focusing on that region and subdividing it into smaller $\sqrt{s}$ intervals. Then, with the methods described in Appendixes E 2 and E 3, one could reconstruct, via rescaling parameters, the modifications to the experimental R-ratio suggested by the lattice observables.

a. Using complete sets of polynomials

Because we are interested in understanding how a tension between lattice and data-driven results could suggest a necessary alteration of the normalization, slope, curvature, etc. of the R-ratio in a given $\sqrt{s}$ interval, it also

makes sense to parametrize the deviation from the experimental R-ratio as a sum of Legendre polynomials P_ℓ that form a complete basis on that interval. Thus for simplicity, we choose the model $M_R(s, \alpha_i)$ to correspond to the histogram of the experimental R-ratio, with the original binning and the α_i parameters tuned to the corresponding bin counts. We also choose the function $F_{\text{lat}\div\text{R}}$ of Eq. (E1) to have the following form[29]:

$$F_{\text{lat}\div\text{R}}(s, \beta_\ell) \equiv \begin{cases} \frac{2}{s_2 - s_1}\sum_{\ell\geq 0}\beta_\ell P_\ell(x(s)), & s \in I_1 \\ 1, & s \in I_2 \end{cases}, \quad \text{(E13)}$$

where $I_1 = [\sqrt{s_1}, \sqrt{s_2}]$ and I_2 is the complementary interval on the real axis.

Then we write

$$a_j^{M_{\text{lat}}} = \sum_{b=1,2} a_{j,b}^{M_{\text{lat}}}, \quad \text{(E14)}$$

where $a_{jb}^{M_{\text{lat}}}$, $b = 1, 2$, are defined via Eq. (E2), with the appropriate kernel $\tilde{K}_j(s)$. The parameters of the model are determined via the first approach described in Appendix E 1. The coefficients β_ℓ of the Legendre polynomials are then determined by minimizing χ_{lat}^2 of Eq. (E3). The resulting model, $M_{\text{lat}}(s, \alpha_i, \beta_\ell)$, serves as a regularization of the unfolding of the lattice R-ratio, in the sense discussed at the end of Appendix E 1, provided that the number of fitted Legendre coefficients is smaller than the number of moment integrals being considered.

Note that the rescaling hypothesis of Sec. IV B is a special case of the one considered here, with ℓ restricted to the value 0 and $\delta = \beta_0 - 1$.

b. Using physics-driven models

Up until this point, none of the considered modifications to the experimental R-ratio incorporate theoretical constraints which the spectral function is known to obey. Parametrizations that do that can be found in Refs. [3,26,131–135].

Here, for concreteness, we focus on the parametrization given in Ref. [26]. It relates the R-ratio to the π^+ electromagnetic form factor in the timelike region and it is the latter which is actually parametrized. This parametrization satisfies many requirements of analyticity and unitarity. It also includes two- and three-pion channels, as well as inelastic corrections via a conformal polynomial, with a threshold dictated by phenomenology. The authors estimate that this parametrization is valid from the two-pion

---

[28]In Eq. (E12), the parentheses around the b and $\bar{b}$ indices indicate that the sum in the second term of the equation runs over the a_j^{R} and a_{jb}^{R}.

[29]This function can be generalized to include Legendre polynomials over several complementary $\sqrt{s}$ intervals, with possible continuity and smoothness conditions at the edges of the intervals, obtained by imposing appropriate constraints on the Legendre coefficients.

threshold to $\sqrt{s} \simeq 1$ GeV. To fit it to the measured R-ratio, the latter must be corrected for final-state radiation [26].

Because we do not actually analyze results with this parametrization here, we do not describe it in detail. Its expression is given in Ref. [26]. However, we make a few comments on its parameters and how it may be used to describe a modification to the measured R-ratio compatible with lattice results.

The first set of parameters is the two-pion phase shift at $\sqrt{s_0} = 0.8$ and $\sqrt{s_1} = 1.15$ GeV. The second set of parameters describes $\rho - \omega$ mixing, in terms of the mass and width of the ω, as well as of a mixing parameter. The last set of parameters describes inelasticity above $\sqrt{s_{\rm in}} = M_\omega + M_{\pi^0}$, with typically four parameters. That constitutes a set of nine parameters that we label p_k.

In Ref. [27], the authors perform a study of the modifications to the measured R-ratio implied by a value of $a_\mu^{\rm LO\text{-}HVP}$ that differs from its data-driven prediction. However, the assumptions behind that study and the questions which it answers are quite different from those of the approach proposed here.

In Ref. [27] the authors first perform a fit of the parametrization of Ref. [26] to all $e^+e^- \to \pi^+\pi^-$ measurements available at the time, fixing its nine parameters. This allows them to predict the two-pion contribution to a number of quantities, including $a_\mu^{\rm LO\text{-}HVP}$, in the data-driven approach. Then they release some of the nine parameters, keeping the others fixed to their data-driven values, while performing the fit once more, with the constraint that the predicted value of $a_\mu^{\pi\pi}$ takes on a precise, prescribed value. This value is allowed to vary from its nominal, data-driven value to the value of $a_\mu^{\pi\pi}$ that one obtains by assuming that the full difference between a lattice result for $a_\mu^{\rm LO\text{-}HVP}$, close to that of Ref. [6], and its data-driven counterpart can be ascribed to the two-pion contribution. In particular, the approach allows the authors to determine the $\Delta\chi^2$ between a scenario in which $a_\mu^{\pi\pi}$ takes on its "lattice" versus its nominal value. In turn, the value of the $\Delta\chi^2$ indicates how compatible different values of $a_\mu^{\pi\pi}$ are with $e^+e^- \to \pi^+\pi^-$ data and known constraints on the pion electromagnetic form factor. The approach also allows the authors to present correlations between $a_\mu^{\pi\pi}$ values and predictions for the two-pion contribution to other observables, such as the pion electromagnetic radius or the running of α.

Here, in a sense that we now describe, we address the problem the other way around. As discussed throughout this article, we choose to consider the measured R-ratio, or quantities derived from it, and as many observables as possible that are computed in lattice QCD and that are related to HVP. The latter include, for instance, the contributions to $a_\mu^{\rm LO\text{-}HVP}$ in various time windows, the running of α in certain spacelike intervals, etc. Concerning the spectral function aspect of the comparison, we not only consider the $\sqrt{s}$ interval $I_1 = [2M_{\pi^\pm}, 1 \text{ GeV}]$, but also its complementary

interval I_2. As for all the comparisons between the two approaches that we have looked upon until now, knowledge of the correlations between all quantities considered must be determined as precisely as possible. Then, in the notation of Appendix E 1, we call $M_{\rm R}(s, \alpha_i)$ the model for the R-ratio. To be more specific, in I_1 one can consider $M_{\rm R}(s, \alpha_i)$ to be the model of Ref. [26], so that nine of the α_i correspond to the parameters p_k discussed above. We continue calling α_i the other parameters of the model. To describe the R-ratio in I_1, only the p_k are needed. The α_i are involved in describing the R-ratio in I_2. In that interval, $M_{\rm R}(s, \{p_k, \alpha_i\})$ can be chosen to be any of the models discussed above for quantities related to the R-ratio, e.g., the histogram of the measured spectral function, with parameters α_i corresponding to the counts in the original measurement bins.

Now we consider a model for the candidate lattice R-ratio, $M_{\rm lat}(s, \alpha_i, \beta_j)$. Instead of defining it via Eq. (E1), we take it to be the same model as $M_{\rm R}(s, \{p_k, \alpha_i\})$ in I_1, but where some of the parameters β_j correspond to possible shifts δp_l of a subset of the parameters p_k of the model of Ref. [26]. Thus, as for $M_{\rm R}(s, \{p_k, \alpha_i\})$, we rewrite the lattice R-ratio model $M_{\rm lat}(s, \{p_k, \alpha_i\}, \{\delta p_l, \beta_j\})$, where the parameters α_i are those needed to describe the measured R-ratio in I_2, via $M_{\rm R}$, and the β_j, those that are used to express the differences between the measured and lattice spectral functions in that same interval. For instance, the β_j could correspond to rescaling factors of the measured R-ratio in subintervals of I_2, discussed in Appendix E, or Legendre polynomial modifications, as discussed in Appendix E 4.

The next step consists of following the second approach described in Appendix E 1, in which the parameters p_k, δp_l, α_i, and β_j are simultaneously determined by minimizing $\chi_{\rm lat}^2 + \chi_{\rm R}^2$, where $\chi_{\rm lat}^2$ is given by Eq. (E3) and $\chi_{\rm R}^2$ is one of the χ^2 given in Eqs. (E4) and (E5). In this physics-motivated approach, both the candidate lattice and the measured R-ratios are described by a model that incorporates as many physical constraints as desired in interval I_1. Moreover, the possible differences between the data-driven and lattice HVP are expressed in terms of parameters which have a physical meaning. For instance, suppose that the δp_l only correspond to modifications in the two-pion phase shift at $\sqrt{s_0} = 0.8$ and $\sqrt{s_1} = 1.15$ GeV. If the minimization of $\chi_{\rm lat}^2 + \chi_{\rm R}^2$ gives a good χ^2 value and at least one of the δp_l is not consistent with zero, then we can argue that, in the interval I_1, the lattice and data-driven approaches differ in their prediction for the phase of the pion form factor. In addition, if one has reasons to believe that $M_{\rm lat}$ correctly describes the differences between the lattice and data-driven approaches and one knows which of the two approaches correctly describes HVP, then the corresponding model can be viewed as a combination of the data-driven and lattice results and can be used to make improved predictions for quantities that involve HVP.

We leave the comparison of lattice and data-driven results via physics-motivated models, as described here, for future work.

APPENDIX F: CONSIDERING MORE OBSERVABLES IN THE DATA-DRIVEN APPROACH

In Ref. [14] it was stated that several window integrals, with possibly narrower $[t_{min}, t_{max}]$ ranges compared to the ones studied up to now, could be employed to perform a more detailed comparison of the dispersive and lattice approaches. While such a larger set of window moments generically enhances the amount of available information, this increase will eventually be limited by the (anti) correlations among the windows, for both the dispersive and lattice determinations.

In the following studies we use a "blinding" approach, in the sense of only communicating the information on the dispersive uncertainties and their correlations for the moment integrals, but not the corresponding nominal values. This is motivated by the fact that these moments have not yet been determined via lattice QCD and we wish to preserve the possibility of doing so without any risk of bias. This is different to the approach followed in Ref. [14] that quotes nominal values for moment integrals derived through the dispersive approach (obtained with different data combination methodologies from those used here), even when these integrals have not yet been computed via lattice QCD. We believe that going forward, all

groups working on this problem should adopt our more conservative approach. Employing blinding approaches is very important when working at the level of precision involved in these studies. This is especially the case in the current context of the existing tensions between the dispersive approaches and the lattice calculations, as well as among the various experimental results themselves.

We consider first the seven window moments discussed in Ref. [14], then a set of nine window moments where the $[1.6, \infty[$ fm window is split into three, with extra thresholds at 2.6 and 4 fm, and finally a set of 19 moments including ten extra $\Delta\alpha_{had}^{(5)}(q^2)$ results, with spacelike q^2 varying from -10 to -1 GeV^2. Table V shows the correlation matrix of these 19 moment integrals, computed using the dispersive approach. It represents a first way of quantifying the amount of independent information available in these moments. Indeed, one can notice that moment integrals with similar kernel shapes [e.g., $\Delta\alpha_{had}^{(5)}(q^2)$ moments computed for similar q^2 values] are strongly correlated, while other moments have weaker correlations. This reflects the fact that differing kernel shapes supply a more effective way of exploiting the information provided by the timelike spectra.

A way of further quantifying the amount of independent information in the moment integrals is to evaluate the eigenvalues of the corresponding covariance matrix, since strong correlations typically imply the presence of small eigenvalues. Tables VI–VIII show the dispersive uncertainties for the three sets of moments, as well as the

TABLE V. Correlation matrix of ten $\Delta\alpha_{had}^{(5)}(q^2)$ and nine $a_{\mu,win}^{LO\text{-}HVP}$ dispersive moment integrals computed for various q^2 values and $[t_{min}, t_{max}]$ intervals with $\Delta = 0.15$ fm, respectively.

Moment integral	Correlation coefficients																		
$\Delta\alpha_{had}^{(5)}(-10$ GeV$^2)$	1																		
$\Delta\alpha_{had}^{(5)}(-9$ GeV$^2)$	0.999	1																	
$\Delta\alpha_{had}^{(5)}(-8$ GeV$^2)$	0.999	0.999	1																
$\Delta\alpha_{had}^{(5)}(-7$ GeV$^2)$	0.996	0.998	0.999	1															
$\Delta\alpha_{had}^{(5)}(-6$ GeV$^2)$	0.993	0.995	0.998	0.999	1														
$\Delta\alpha_{had}^{(5)}(-5$ GeV$^2)$	0.986	0.990	0.994	0.997	0.999	1													
$\Delta\alpha_{had}^{(5)}(-4$ GeV$^2)$	0.976	0.981	0.986	0.991	0.995	0.999	1												
$\Delta\alpha_{had}^{(5)}(-3$ GeV$^2)$	0.960	0.966	0.973	0.980	0.986	0.993	0.998	1											
$\Delta\alpha_{had}^{(5)}(-2$ GeV$^2)$	0.931	0.939	0.948	0.957	0.967	0.977	0.987	0.996	1										
$\Delta\alpha_{had}^{(5)}(-1$ GeV$^2)$	0.874	0.885	0.896	0.909	0.923	0.938	0.955	0.973	0.990	1									
$a_{\mu,win}^{LO-HVP}$ [0, 0.1] fm	0.806	0.791	0.774	0.753	0.728	0.698	0.660	0.611	0.543	0.442	1								
$a_{\mu,win}^{LO-HVP}$ [0.1, 0.4] fm	0.959	0.955	0.949	0.942	0.931	0.916	0.895	0.864	0.813	0.723	0.864	1							
$a_{\mu,win}^{LO-HVP}$ [0.4, 0.7] fm	0.876	0.887	0.899	0.912	0.926	0.940	0.954	0.966	0.972	0.958	0.428	0.786	1						
$a_{\mu,win}^{LO-HVP}$ [0.7, 1] fm	0.711	0.726	0.743	0.762	0.784	0.809	0.838	0.873	0.91	0.961	0.206	0.509	0.893	1					
$a_{\mu,win}^{LO-HVP}$ [1, 1.3] fm	0.604	0.619	0.636	0.656	0.678	0.705	0.738	0.778	0.831	0.901	0.123	0.365	0.775	0.973	1				
$a_{\mu,win}^{LO-HVP}$ [1.3, 1.6] fm	0.553	0.568	0.584	0.604	0.626	0.653	0.686	0.728	0.783	0.861	0.093	0.305	0.710	0.941	0.993	1			
$a_{\mu,win}^{LO-HVP}$ [1.6, 2.6] fm	0.508	0.522	0.537	0.556	0.577	0.604	0.636	0.677	0.733	0.814	0.074	0.260	0.647	0.891	0.963	0.987	1		
$a_{\mu,win}^{LO-HVP}$ [2.6, 4] fm	0.419	0.431	0.445	0.461	0.479	0.502	0.530	0.567	0.617	0.694	0.052	0.197	0.523	0.753	0.840	0.885	0.944	1	
$a_{\mu,win}^{LO-HVP}$ [4, ∞[fm	0.312	0.321	0.332	0.344	0.358	0.375	0.397	0.426	0.466	0.528	0.034	0.137	0.381	0.565	0.646	0.698	0.787	0.942	1

TABLE VI. Total uncertainty of seven $a_{\mu,\text{win}}^{\text{LO-HVP}}$ dispersive moment integrals computed for various $[t_{\min}, t_{\max}]$ intervals with $\Delta = 0.15$ fm. The eigenvalues of the corresponding covariance, correlation, and normalized covariance matrices are also shown.

Moment integral		Eigenvalues		
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ $[t_{\min}, t_{\max}]$ (fm)	Total uncertainty	Covariance	Correlation	Normalized covariance
[0, 0.1]	8.18×10^{-12}	3.12×10^{-20}	4.85	2.04×10^{-4}
[0.1, 0.4]	3.86×10^{-11}	3.52×10^{-21}	1.76	8.39×10^{-5}
[0.4, 0.7]	6.43×10^{-11}	6.38×10^{-22}	3.32×10^{-1}	1.45×10^{-5}
[0.7, 1]	8.00×10^{-11}	1.53×10^{-22}	4.88×10^{-2}	2.06×10^{-6}
[1, 1.3]	7.90×10^{-11}	8.77×10^{-24}	4.27×10^{-3}	1.87×10^{-7}
[1.3, 1.6]	6.32×10^{-11}	4.64×10^{-25}	4.47×10^{-4}	1.88×10^{-8}
$[1.6, \infty[$	1.15×10^{-10}	7.89×10^{-26}	1.51×10^{-5}	6.34×10^{-10}

TABLE VII. Total uncertainty of nine $a_{\mu,\text{win}}^{\text{LO-HVP}}$ dispersive moment integrals computed for various $[t_{\min}, t_{\max}]$ intervals with $\Delta = 0.15$ fm. The eigenvalues of the corresponding covariance, correlation, and normalized covariance matrices are also shown.

Moment integral		Eigenvalues		
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ $[t_{\min}, t_{\max}]$ (fm)	Total uncertainty	Covariance	Correlation	Normalized covariance
[0, 0.1]	8.18×10^{-12}	2.76×10^{-20}	6.07	2.49×10^{-4}
[0.1, 0.4]	3.86×10^{-11}	3.17×10^{-21}	1.99	9.28×10^{-5}
[0.4, 0.7]	6.43×10^{-11}	5.11×10^{-22}	6.86×10^{-1}	2.72×10^{-5}
[0.7, 1]	8.00×10^{-11}	1.33×10^{-22}	2.29×10^{-1}	9.71×10^{-6}
[1, 1.3]	7.90×10^{-11}	1.52×10^{-23}	2.12×10^{-2}	8.76×10^{-7}
[1.3, 1.6]	6.32×10^{-11}	1.20×10^{-24}	2.36×10^{-3}	9.89×10^{-8}
[1.6, 2.6]	9.32×10^{-11}	2.91×10^{-25}	3.90×10^{-4}	1.59×10^{-8}
[2.6, 4]	2.01×10^{-11}	2.15×10^{-26}	3.43×10^{-5}	1.41×10^{-9}
$[4, \infty[$	2.64×10^{-12}	1.78×10^{-27}	5.83×10^{-7}	2.50×10^{-11}

TABLE VIII. Total uncertainty of ten $\Delta\alpha_{\text{had}}^{(5)}(q^2)$ and nine $a_{\mu,\text{win}}^{\text{LO-HVP}}$ dispersive moment integrals computed for various q^2 values and $[t_{\min}, t_{\max}]$ intervals with $\Delta = 0.15$ fm, respectively. The eigenvalues of the corresponding covariance, correlation, and normalized covariance matrices are also shown.

Moment integral	Total uncertainty	Eigenvalues		
		Covariance	Correlation	Normalized covariance
$\Delta\alpha_{\text{had}}^{(5)}(-10 \text{ GeV}^2)$	4.80×10^{-5}	1.48×10^{-8}	14.74	5.21×10^{-4}
$\Delta\alpha_{\text{had}}^{(5)}(-9 \text{ GeV}^2)$	4.64×10^{-5}	2.17×10^{-10}	3.26	1.34×10^{-4}
$\Delta\alpha_{\text{had}}^{(5)}(-8 \text{ GeV}^2)$	4.48×10^{-5}	4.09×10^{-12}	7.36×10^{-1}	2.92×10^{-5}
$\Delta\alpha_{\text{had}}^{(5)}(-7 \text{ GeV}^2)$	4.29×10^{-5}	1.94×10^{-14}	2.40×10^{-1}	1.02×10^{-5}
$\Delta\alpha_{\text{had}}^{(5)}(-6 \text{ GeV}^2)$	4.09×10^{-5}	2.22×10^{-16}	2.16×10^{-2}	8.90×10^{-7}
$\Delta\alpha_{\text{had}}^{(5)}(-5 \text{ GeV}^2)$	3.85×10^{-5}	3.40×10^{-18}	2.45×10^{-3}	1.02×10^{-7}
$\Delta\alpha_{\text{had}}^{(5)}(-4 \text{ GeV}^2)$	3.57×10^{-5}	1.99×10^{-20}	4.17×10^{-4}	1.68×10^{-8}
$\Delta\alpha_{\text{had}}^{(5)}(-3 \text{ GeV}^2)$	3.23×10^{-5}	9.27×10^{-23}	3.92×10^{-5}	1.59×10^{-9}
$\Delta\alpha_{\text{had}}^{(5)}(-2 \text{ GeV}^2)$	2.78×10^{-5}	1.40×10^{-23}	2.33×10^{-6}	8.94×10^{-11}
$\Delta\alpha_{\text{had}}^{(5)}(-1 \text{ GeV}^2)$	2.07×10^{-5}	7.03×10^{-25}	1.15×10^{-7}	4.78×10^{-12}
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ [0, 0.1] fm	8.18×10^{-12}	1.47×10^{-25}	3.46×10^{-10}	1.22×10^{-14}
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ [0.1, 0.4] fm	3.86×10^{-11}	7.96×10^{-28}	3.76×10^{-13}	1.19×10^{-17}

(Table continued)

TABLE VIII. (Continued)

| | | Eigenvalues | | |
Moment integral	Total uncertainty	Covariance	Correlation	Normalized covariance
$a_{\mu,\text{win}}^{\text{LO–HVP}}$ [0.4, 0.7] fm	6.43×10^{-11}	1.66×10^{-28}	2.28×10^{-13}	7.40×10^{-18}
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ [0.7, 1] fm	8.00×10^{-11}	4.50×10^{-30}	9.74×10^{-14}	3.10×10^{-18}
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ [1, 1.3] fm	7.90×10^{-11}	5.25×10^{-31}	3.07×10^{-14}	9.72×10^{-19}
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ [1.3, 1.6] fm	6.32×10^{-11}	8.57×10^{-32}	1.40×10^{-14}	4.38×10^{-19}
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ [1.6, 2.6] fm	9.32×10^{-11}	-4.56×10^{-27}	-2.24×10^{-14}	-7.06×10^{-19}
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ [2.6, 4] fm	2.01×10^{-11}	-2.01×10^{-23}	-3.79×10^{-14}	-1.19×10^{-18}
$a_{\mu,\text{win}}^{\text{LO-HVP}}$ [4, ∞[fm	2.64×10^{-12}	-1.12×10^{-22}	-1.22×10^{-13}	-3.83×10^{-18}

eigenvalues of the corresponding covariance matrices, which typically show a fast drop over several orders of magnitude. It is worth noting that the uncertainties of the $\Delta\alpha_{\text{had}}^{(5)}(q^2)$ and $a_{\mu,\text{win}}^{\text{LO-HVP}}$ integrals typically differ by several orders of magnitude, just because of the natural normalization of these moments, which is also reflected in the range covered by the corresponding eigenvalues of the covariance matrix. This effect, also present to some extent when considering separate series of $a_{\mu,\text{win}}^{\text{LO-HVP}}$ and $\Delta\alpha_{\text{had}}^{(5)}(q^2)$ integrals, interferes with the intent of quantifying the amount of independent information in the moments.

To take this into account, we also consider the eigenvalues of the correlation matrices of the moment integrals, as well as of the covariance matrices normalized by the nominal values of the moments [i.e., $[C_R]_{ij}/(a_i^R \cdot a_j^R)$]. These are also displayed in Tables VI–VIII.[30,31] For the set of seven $a_{\mu,\text{win}}^{\text{LO-HVP}}$ dispersive moment integrals in Table VI, these eigenvalues cover 7 orders of magnitude, which reflects the strong correlations among some of these moments. For the three sets of moments, while the largest eigenvalues are enhanced by a factor of 2–3, as more moments are added, the smallest eigenvalues drop by more and more orders of magnitude. This indicates that the amount of independent information (i.e., of independent degrees of freedom) is less than the number of added extra moments. Indeed, when comparing the ratios of the various eigenvalues with the largest one, a reduction of the eigenvalues by 2 or 4 orders of magnitude for Table VI allows us to accommodate typically one extra moment in Table VII. At the same time, when including the extra $\Delta\alpha_{\text{had}}^{(5)}(q^2)$ moments in Table VIII, the potential gain in extra degrees of freedom seems rather

marginal. This is related to the fact that the $\Delta\alpha_{\text{had}}^{(5)}(q^2)$ moments considered here are rather strongly correlated among themselves and have rather strong correlations with at least one of the $a_{\mu,\text{win}}^{\text{LO-HVP}}$ moments (see Table V).

One can attempt to provide more quantitative conclusions concerning the number of independent degrees of freedom that are available in the data-driven approach for a comparison with lattice results. For this purpose, one can consider the eigenvalues of the correlation or normalized covariance matrix, rescaled by the largest eigenvalue. The number of independent degrees of freedom can then be obtained by counting the number of such ratios above some lower bound. This procedure is consistent with the idea that there is a limit to the precision with which we can obtain the uncertainty of moments of the measured R-ratio and linear combinations thereof. Such a limit approximately corresponds to that lower bound.

Here we choose this lower bound to be 10^{-6}, driven by the number of orders of magnitude covered by the eigenvalues of the correlation and normalized covariance matrices, for the moments in Table VI. With this bound we find that the seven window observables listed in Table VI contain a significant amount of linearly independent information. This number rises to eight when two more long-distance-window observables are considered in Table VII. However, when ten values of $\Delta\alpha_{\text{had}}^{(5)}(q^2)$ are added, in Table V, to the nine window observables of Table VII, no additional independent degrees of freedom appear. Thus, assuming that a lower bound of around 10^{-6} is appropriate, it is reasonable to conclude that the number of linearly independent moments which can be obtained from the experimental R-ratio is less than 10 for the types of observables considered here, which are the ones that should be computable, in the lattice approach with subpercent precision.

All these and other results presented in the paper point to the importance of determining the corresponding matrices in the lattice approach and performing studies similar to those presented here, when more moments become available in that approach.

---

[30]It is worth noting that the rotation matrices of eigenvectors from the diagonalization of the covariance, correlation, and normalized covariance matrices are different.

[31]The presence of some small negative eigenvalues in Table VIII is related to numerical effects typically occurring in the presence of strong correlations among moments.

[1] B. Abi *et al.* (Muon *g* − 2 Collaboration), Phys. Rev. Lett. **126**, 141801 (2021).

[2] G. W. Bennett *et al.* (Muon *g* − 2 Collaboration), Phys. Rev. D **73**, 072003 (2006).

[3] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang, Eur. Phys. J. C **80**, 241 (2020).

[4] A. Keshavarzi, D. Nomura, and T. Teubner, Phys. Rev. D **101**, 014029 (2020).

[5] T. Aoyama *et al.*, Phys. Rep. **887**, 1 (2020).

[6] S. Borsanyi *et al.*, Nature (London) **593**, 51 (2021).

[7] See e.g., Laurent Lellouch, *Talk Entitled Leading Hadronic Contribution to the Muon Magnetic Moment from Lattice QCD, at the Workshop "Virtual Breakfast with g–2"* (IJClab, Paris, France, 2021).

[8] D. P. Aguillard *et al.* (Muon *g* − 2 Collaboration), Phys. Rev. Lett. **131**, 161802 (2023).

[9] T. Blum, P. A. Boyle, V. Gülpers, T. Izubuchi, L. Jin, C. Jung, A. Jüttner, C. Lehner, A. Portelli, and J. T. Tsang (RBCUKQCD Collaborations), Phys. Rev. Lett. **121**, 022003 (2018).

[10] M. Cè *et al.*, Phys. Rev. D **106**, 114502 (2022).

[11] C. Alexandrou *et al.* (Extended Twisted Mass Collaboration (ETMC)), Phys. Rev. D **107**, 074506 (2023).

[12] T. Blum *et al.* (RBC, UKQCD Collaborations), Phys. Rev. D **108**, 054507 (2023).

[13] A. Bazavov *et al.* (Fermilab Lattice, HPQCD, MILC Collaborations), Phys. Rev. D **107**, 114514 (2023).

[14] G. Colangelo, A. X. El-Khadra, M. Hoferichter, A. Keshavarzi, C. Lehner, P. Stoffer, and T. Teubner, Phys. Lett. B **833**, 137313 (2022).

[15] F. V. Ignatov *et al.* (CMD-3 Collaboration), arXiv:2302.08834.

[16] R. R. Akhmetshin *et al.* (CMD-2 Collaboration), Phys. Lett. B **648**, 28 (2007).

[17] A. Crivellin, M. Hoferichter, C. A. Manzari, and M. Montull, Phys. Rev. Lett. **125**, 091801 (2020).

[18] A. Keshavarzi, W. J. Marciano, M. Passera, and A. Sirlin, Phys. Rev. D **102**, 033002 (2020).

[19] E. de Rafael, Phys. Rev. D **102**, 056025 (2020).

[20] B. Malaescu and M. Schott, Eur. Phys. J. C **81**, 46 (2021).

[21] S. Borsanyi *et al.* (Budapest-Marseille-Wuppertal Collaboration), Phys. Rev. Lett. **121**, 022002 (2018).

[22] M. Cè, A. Gérardin, G. von Hippel, H. B. Meyer, K. Miura, K. Ottnad, A. Risch, T. San José, J. Wilhelm, and H. Wittig, J. High Energy Phys. 08 (2022) 220.

[23] D. Bernecker and H. B. Meyer, Eur. Phys. J. A **47**, 148 (2011).

[24] M. Hansen, A. Lupo, and N. Tantalo, Phys. Rev. D **99**, 094508 (2019).

[25] C. Alexandrou *et al.* (Extended Twisted Mass Collaboration (ETMC)), Phys. Rev. Lett. **130**, 241901 (2023).

[26] G. Colangelo, M. Hoferichter, and P. Stoffer, J. High Energy Phys. 02 (2019) 006.

[27] G. Colangelo, M. Hoferichter, and P. Stoffer, Phys. Lett. B **814**, 136073 (2021).

[28] D. Boito, M. Golterman, K. Maltman, and S. Peris, Phys. Rev. D **107**, 034512 (2023).

[29] B. B. Brandt, A. Francis, H. B. Meyer, and D. Robaina, Phys. Rev. D **92**, 094510 (2015).

[30] M. T. Hansen, H. B. Meyer, and D. Robaina, Phys. Rev. D **96**, 094513 (2017).

[31] J. Bulava and M. T. Hansen, Phys. Rev. D **100**, 034521 (2019).

[32] C. Bonanno, F. D'Angelo, M. D'Elia, L. Maio, and M. Naviglio, Phys. Rev. D **108**, 074515 (2023).

[33] R. Frezzotti, N. Tantalo, G. Gagliardi, F. Sanfilippo, S. Simula, and V. Lubicz, Phys. Rev. D **108**, 074510 (2023).

[34] G. Bailas, S. Hashimoto, and T. Ishikawa, Prog. Theor. Exp. Phys. **2020**, 043B07 (2020).

[35] Y. Nakahara, M. Asakawa, and T. Hatsuda, Phys. Rev. D **60**, 091503 (1999).

[36] Y. Nakahara, M. Asakawa, and T. Hatsuda, Nucl. Phys. B, Proc. Suppl. **83**, 191 (2000).

[37] M. Asakawa, T. Hatsuda, and Y. Nakahara, Prog. Part. Nucl. Phys. **46**, 459 (2001).

[38] A. Rothkopf, T. Hatsuda, and S. Sasaki, Phys. Rev. Lett. **108**, 162001 (2012).

[39] A. Rothkopf, J. Comput. Phys. **238**, 106 (2013).

[40] M. Asakawa, arXiv:2001.10205.

[41] A. Rothkopf, Data **5**, 85 (2020).

[42] Y. Burnier and A. Rothkopf, Phys. Rev. Lett. **111**, 182003 (2013).

[43] A. Rothkopf, Phys. Rev. D **95**, 056016 (2017).

[44] A. K. Cyrol, J. M. Pawlowski, A. Rothkopf, and N. Wink, SciPost Phys. **5**, 065 (2018).

[45] H.-N. Li, H. Umeeda, F. Xu, and F.-S. Yu, Phys. Lett. B **810**, 135802 (2020).

[46] H.-n. Li and H. Umeeda, Phys. Rev. D **102**, 094003 (2020).

[47] H.-n. Li and H. Umeeda, Phys. Rev. D **102**, 114014 (2020).

[48] A.-S. Xiong, T. Wei, and F.-S. Yu, arXiv:2211.13753.

[49] P. Gubler and M. Oka, Prog. Theor. Phys. **124**, 995 (2010).

[50] M. Bruno and M. T. Hansen, J. High Energy Phys. 06 (2021) 043.

[51] T. Bergamaschi, W. I. Jay, and P. R. Oare, Phys. Rev. D **108**, 074516 (2023).

[52] J. Horak, J. M. Pawlowski, J. Rodríguez-Quintero, J. Turnwald, J. M. Urban, N. Wink, and S. Zafeiropoulos, Phys. Rev. D **105**, 036014 (2022).

[53] S. Offler, G. Aarts, C. Allton, J. Glesaaen, B. Jäger, S. Kim, M. P. Lombardo, S. M. Ryan, and J.-I. Skullerud, Proc. Sci., LATTICE2019 (**2019**) 076.

[54] S. Offler, G. Aarts, C. Allton, B. Jäger, S. Kim, M.-P. Lombardo, B. Page, S. M. Ryan, J.-I. Skullerud, and T. Spriggs, Proc. Sci., LATTICE2021 (**2022**) 509.

[55] L. Kades, J. M. Pawlowski, A. Rothkopf, M. Scherzer, J. M. Urban, S. J. Wetzel, N. Wink, and F. P. G. Ziegler, Phys. Rev. D **102**, 096001 (2020).

[56] N. Laanait, J. Romero, J. Yin, M. T. Young, S. Treichler, V. Starchenko, A. Borisevich, A. Sergeev, and M. Matheson, arXiv:1909.11150.

[57] S. Y. Chen, H. T. Ding, F. Y. Liu, G. Papp, and C. B. Yang, arXiv:2110.13521.

[58] L. Wang, S. Shi, and K. Zhou, Phys. Rev. D **106**, L051502 (2022).

[59] L. Wang, S. Shi, and K. Zhou, in *35th Conference on Neural Information Processing Systems* (2021); arXiv:2112.06206.

[60] S. Shi, L. Wang, and K. Zhou, Comput. Phys. Commun. **282**, 108547 (2023).

[61] T. Lechien and D. Dudal, SciPost Phys. **13**, 097 (2022).

[62] D. Boyda *et al.*, in *Snowmass 2021* (2022); arXiv:2202.05838.

[63] M. Buzzicotti, A. De Santis, and N. Tantalo, Eur. Phys. J. C **84**, 32 (2024).

[64] A. J. Chambers, R. Horsley, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, A. Schiller, K. Somfleth, R. D. Young, and J. M. Zanotti, Phys. Rev. Lett. **118**, 242001 (2017).

[65] J. Karpie, K. Orginos, A. Rothkopf, and S. Zafeiropoulos, J. High Energy Phys. 04 (2019) 057.

[66] A. Candido, L. Del Debbio, T. Giani, and G. Petrillo, Proc. Sci., LATTICE2022 (**2023**) 098.

[67] D. Dudal, O. Oliveira, and P. J. Silva, Phys. Rev. D **89**, 014010 (2014).

[68] R.-A. Tripolt, P. Gubler, M. Ulybyshev, and L. Von Smekal, Comput. Phys. Commun. **237**, 129 (2019).

[69] A. Rothkopf, Front. Phys. **10**, 1028995 (2022).

[70] A. Rothkopf, EPJ Web Conf. **274**, 01004 (2022).

[71] M. Davier, A. Hoecker, B. Malaescu, C. Z. Yuan, and Z. Zhang, Eur. Phys. J. C **66**, 1 (2010).

[72] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang, Eur. Phys. J. C **71**, 1515 (2011); **72**, 1874(E) (2012).

[73] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang, Eur. Phys. J. C **77**, 827 (2017).

[74] P. Masjuan, A. Miranda, and P. Roig, Phys. Lett. B **850**, 138492 (2024).

[75] G. D'Agostini, Nucl. Instrum. Methods Phys. Res., Sect. A **346**, 306 (1994).

[76] V. Blobel, eConf **C030908**, MOET002 (2003), https://inspirehep.net/literature/637567.

[77] C. Lehner and A. S. Meyer, Phys. Rev. D **101**, 074515 (2020).

[78] C. Aubin, T. Blum, M. Golterman, and S. Peris, Phys. Rev. D **106**, 054503 (2022).

[79] G. Wang, T. Draper, K.-F. Liu, and Y.-B. Yang (chiQCD Collaboration), Phys. Rev. D **107**, 034513 (2023).

[80] G. Benton, D. Boito, M. Golterman, A. Keshavarzi, K. Maltman, and S. Peris, Phys. Rev. Lett. **131**, 251803 (2023).

[81] G. Benton, D. Boito, M. Golterman, A. Keshavarzi, K. Maltman, and S. Peris, Phys. Rev. D **109**, 036010 (2024).

[82] M. Cè, T. Harris, H. B. Meyer, A. Toniato, and C. Török, J. High Energy Phys. 12 (2021) 215.

[83] E.-H. Chao, H. B. Meyer, and J. Parrino, Phys. Rev. D **107**, 054505 (2023).

[84] F. Ambrosino *et al.* (KLOE Collaboration), Phys. Lett. B **670**, 285 (2009).

[85] F. Ambrosino *et al.* (KLOE Collaboration), Phys. Lett. B **700**, 102 (2011).

[86] D. Babusci *et al.* (KLOE Collaboration), Phys. Lett. B **720**, 336 (2013).

[87] J. P. Lees *et al.* (*BABAR* Collaboration), Phys. Rev. D **86**, 032013 (2012).

[88] A. Keshavarzi, D. Nomura, and T. Teubner, Phys. Rev. D **97**, 114025 (2018).

[89] S. Kuberski, M. Cè, G. von Hippel, H. B. Meyer, K. Ottnad, A. Risch, and H. Wittig, arXiv:2401.11895.

[90] G. Colangelo, EPJ Web Conf. **258**, 01004 (2022).

[91] M. Davier, A. Hoecker, A. M. Lutz, B. Malaescu, and Z. Zhang, arXiv:2312.02053.

[92] J. P. Lees *et al.* (*BABAR* Collaboration), Phys. Rev. D **108**, L111103 (2023).

[93] www.gauss-centre.eu.

[94] www.genci.fr/en.

[95] V. M. Aul'chenko *et al.*, JETP Lett. **84**, 413 (2006).

[96] M. N. Achasov *et al.*, J. Exp. Theor. Phys. **103**, 380 (2006).

[97] B. Aubert *et al.* (*BABAR* Collaboration), Phys. Rev. Lett. **103**, 231801 (2009).

[98] A. Fowlie and Q. Li, Eur. Phys. J. C **83**, 943 (2023).

[99] N. Husung, P. Marquard, and R. Sommer, Eur. Phys. J. C **80**, 200 (2020).

[100] https://indico.cern.ch/event/956699/contributions/4108682/attachments/2145724/3616679/SymanzikCompressed.pdf (2020).

[101] P. A. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2020**, 083C01 (2020).

[102] F. Herzog, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, J. High Energy Phys. 02 (2017) 090.

[103] Y. Schroder and M. Steinhauser, J. High Energy Phys. 01 (2006) 051.

[104] S. Dürr *et al.* (Budapest-Marseille-Wuppertal Collaboration), Science **322**, 1224 (2008).

[105] S. Borsanyi *et al.*, Science **347**, 1452 (2015).

[106] S. Borsanyi *et al.* (Budapest-Marseille-Wuppertal Collaboration), Phys. Rev. Lett. **121**, 022002 (2018).

[107] L. Lyons, D. Gibaut, and P. Clifford, Nucl. Instrum. Methods Phys. Res., Sect. A **270**, 110 (1988).

[108] G. Cowan, *Statistical Data Analysis* (Oxford University Press, New York, 1998).

[109] K. Nakamura *et al.* (Particle Data Group), J. Phys. G **37**, 075021 (2010).

[110] C. Pascaud and F. Zomer, LAL-95-05, 1995.

[111] L. Lyons, A. J. Martin, and D. H. Saxon, Phys. Rev. D **41**, 982 (1990).

[112] L. Demortier, Equivalence of the best-fit and covariance-matrix methods for comparing binned data with a model in the presence of correlated systematic uncertainties, CDF Note No. 8661, 1999.

[113] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Phys. Rev. D **65**, 014012 (2001).

[114] M. Botje, J. Phys. G **28**, 779 (2002).

[115] R. S. Thorne, J. Phys. G **28**, 2705 (2002).

[116] G. L. Fogli, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, Phys. Rev. D **66**, 053010 (2002).

[117] B. List, Decomposition of a covariance matrix into uncorrelated and correlated errors, https://indico.desy.de/event/3009/contributions/64704/ (2010), Alliance Workshop on Unfolding and Data Correction, DESY.

[118] T. DeGrand, Phys. Rev. D **106**, 014504 (2022).

[119] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **75**, 17 (2015).

[120] M. Aaboud *et al.* (ATLAS Collaboration), J. High Energy Phys. 09 (2017) 020.

[121] M. Aaboud *et al.* (ATLAS Collaboration), J. High Energy Phys. 05 (2018) 195.

[122] B. Malaescu and P. Starovoitov, Eur. Phys. J. C **72**, 2041 (2012).

[123] B. Malaescu, Treatment of uncertainties and correlations in combinations of e^+e^- annihilation data, https://indico.him.uni-mainz.de/event/11/session/1/contribution/42/material/slides/0.pdf (2018), Muon $g - 2$ Theory Initiative workshop Mainz.

[124] G. Cowan, EPJ Web Conf. **258,** 09002 (2022).

[125] M. Davier, D. Díaz-Calderón, B. Malaescu, A. Pich, A. Rodríguez-Sánchez, and Z. Zhang, J. High Energy Phys. 04 (2023) 067.

[126] M. Schmelling, Phys. Scr. **51,** 676 (1995).

[127] S. Aoki *et al.*, Eur. Phys. J. C **74,** 2890 (2014).

[128] D. Britzger, Eur. Phys. J. C **82,** 731 (2022).

[129] K. Cranmer and A. Held, EPJ Web Conf. **251,** 03067 (2021).

[130] G. Aad *et al.* (ATLAS Collaboration), J. High Energy Phys. 06 (2023) 019.

[131] C. Hanhart, S. Holz, B. Kubis, A. Kupść, A. Wirzba, and C. W. Xiao, Eur. Phys. J. C **77,** 98 (2017); **78,** 450(E) (2018).

[132] J. F. De Troconiz and F. J. Yndurain, Phys. Rev. D **65,** 093001 (2002).

[133] J. F. de Troconiz and F. J. Yndurain, Phys. Rev. D **71,** 073008 (2005).

[134] R. Garcia-Martin, R. Kaminski, J. R. Pelaez, J. Ruiz de Elvira, and F. J. Yndurain, Phys. Rev. D **83,** 074004 (2011).

[135] B. Ananthanarayan, I. Caprini, and D. Das, Phys. Rev. D **98,** 114015 (2018).