Indirect new physics effects on σ_{had} confront the $(g-2)_\mu$ window discrepancies and the CMD-3 result

Luc Darmé,[1,*] Giovanni Grilli di Cortona,[2,3,†] and Enrico Nardi[2,4,‡]

[1]*Université de Lyon, Université Claude Bernard Lyon 1, CNRS/IN2P3, Institut de Physique des 2 Infinis de Lyon, UMR 5822, F-69622, Villeurbanne, France*
[2]*Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati, C.P. 13, 00044 Frascati, Italy*
[3]*Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Piazzale A. Moro 2, I-00185 Rome, Italy*
[4]*Laboratory of High Energy and Computational Physics, HEPC-NICPB, Rävala 10, 10143 Tallinn, Estonia*

(Received 29 September 2023; accepted 16 November 2023; published 30 November 2023)

Recent lattice determinations of the hadronic-vacuum polarization contribution to the muon anomalous magnetic moment a_μ^{HVP} have confirmed the discrepancy with the data-driven dispersive method. In the meanwhile the CMD-3 Collaboration has reported a result for the $e^+e^- \to \pi^+\pi^-$ cross section considerably larger than previous experimental results (and close to the lattice determinations) exacerbating the discordance between different e^+e^- datasets. We explore to what extent these disagreements can be accounted for by some new physics effect altering selectively the individual experimental determinations of $\sigma(e^+e^- \to \text{hadrons})$. We find that specific effects of GeV-scale new particles are able to shift upwards the KLOE and *BABAR* results in the low and intermediate energy windows, while leaving the CMD-3 energy scan unaffected. Although these new physics effects cannot fully explain all the discrepancies among the different $\sigma(e^+e^- \to \text{hadrons})$ datasets, they succeed in mitigating the overall tension between data-driven and lattice estimates of a_μ^{HVP}. Remarkably, the additional loop corrections involving the new particles concur to solve the residual discrepancy with the experimental value of $(g-2)_\mu$.

DOI: 10.1103/PhysRevD.108.095056

I. INTRODUCTION

The theoretical uncertainty in the Standard Model (SM) prediction for the anomalous magnetic moment of the muon a_μ is currently dominated by the error associated with the hadronic-vacuum polarization (HVP) contribution a_μ^{HVP}. The recommended value for the leading order HVP correction [1] is based on $e^+e^- \to$ hadrons cross section data, and reads,

$$a_\mu^{\mathrm{LO,HVP}}\Big|_{\mathrm{data-driven}} = 693.1(4.0) \times 10^{-10}. \tag{1}$$

This yields the SM prediction [2–21],

$$a_\mu^{\mathrm{SM}} = 11659181.0(4.3) \times 10^{-10}. \tag{2}$$

Comparing this result with the current experimental world average obtained by combining the previous BNL [22] and FNAL [23] results with the new FNAL determination [24],

$$a_\mu^{\mathrm{exp}} = 116592059(22) \times 10^{-11}, \tag{3}$$

leads to a tension at the level of 5σ.

On the other hand, a_μ^{HVP} can be also computed from first principles by means of QCD lattice techniques. The most precise lattice result, obtained by the BMW Collaboration [25], is

$$a_\mu^{\mathrm{LO,HVP}}\Big|_{\mathrm{BMW}} = 707.5(5.5) \times 10^{-10}, \tag{4}$$

which is in tension at 2.1σ with the data-driven determination in Eq. (1) and, most noticeably, would reduce the difference with the experimental value in Eq. (3) to 1.5σ.

Clearly, a confirmation of the correctness of the BMW result would have a major impact on assessing the need for new physics (NP) to account for the measured value of a_μ. Unfortunately, lattice-QCD results are generally affected by large systematic and statistical uncertainties mostly related to the infinite volume and continuum limits, which have so far prevented high-accuracy determinations of the HVP.

*l.darme@ip2i.in2p3.fr
†grillidc@lnf.infn.it
‡enrico.nardi@lnf.infn.it

However, specific parameter-space regions exist in which the previous uncertainties are under better control, and within these regions more precise results can be obtained. The so-called short, intermediate and long Euclidean time-distance windows (respectively labeled with SD, W, and LD) were first defined by the RBC/UKQCD Collaboration in Ref. [26]. Weight functions are introduced in the HVP integrals, which allow to select only certain regions in parameter space, and in particular regions that are less affected by the sources of uncertainty. In recent years, precise lattice-QCD determination of partial HVP integrals became available. The determination of the HVP contribution of the intermediate window a_{μ}^{HVP} provided by the BMW Collaboration [25], by the CLS/Mainz group [27], by the Extended Twisted-Mass Collaboration (ETMC) [28], by Lehner and Meyer [29], by Aubin *et al.* [30] (which updates their previous result [31]), by the χQCD Collaboration [32] and, very recently, by the Fermilab Lattice, HPQCD, MILC [33], and RBC/UKQCD lattice Collaborations [34], are in overall good agreement, giving strong support to the reliability of lattice evaluations. At the same time they shed some motivated suspicion on the result of the dispersive method which, in the same window [35], is several σ below the average of the lattice results.[1] In contrast, in the short-distance window no substantial discrepancy has been encountered [28] (see also [37]). The deviation between the lattice results and the data-driven determination may thus be interpreted as an effect of NP that is localized in the intermediate and possibly large-distance windows.

In Ref. [38] it was pointed out that, besides direct NP loop contribution to $(g-2)_{\mu}$, additional NP effects acting indirectly on the way σ_{had} is extracted from the experimental data were needed in order to reconcile the various discrepancies known at that time,[2] and a class of NP scenarios in which specific types of indirect effects can affect the various experiments in different ways was put forth.

This is an important feature in view of the fact that besides the 4.2σ discrepancy between the data-driven SM prediction for a_{μ} and the experimental result and the discrepancy of comparable significance with the lattice results for a_{μ}^{HVP}, significant disagreements between different experimental determinations of σ_{had} are also present. In particular, the long-standing $\sim 3\sigma$ discrepancy between KLOE and *BABAR*, which provided the two most accurate determinations of σ_{had}, is now overshadowed by the new CMD-3 measurement of the $e^+e^- \to \pi^+\pi^-$ cross section [45].[3] The result of this measurement yields a value of $\sigma_{\pi^+\pi^-}$ well-above previous results. In particular, in the energy range $\sqrt{s} \in [0.6, 0.88]$ GeV the CMD-3 contribution to $a_{\mu}^{\mathrm{LO,HVP}}$ is more than 5σ above the contribution estimated from KLOE data [45,47].

The aim of this work is to study in details whether the class of new physics scenarios introduced in Ref. [38] can:

(1) Solve the discrepancy between the data-driven and lattice evaluation of the HVP contribution a_{μ}^{HVP} in the intermediate window.
(2) Solve the discrepancy between the full experimental measurement a_{μ}^{exp} and its data-driven counterpart.
(3) Improve the consistency between the different datasets used for the data-driven estimate of the HVP contribution $a_{\mu}^{\mathrm{LO,HVP}}$.

We will thus study the impact of indirect NP effects on the intermediate energy window a_{μ}^{HVP} for the different datasets that we label as KLOE08 [48], KLOE10 [49] KLOE12 [50], BESIII [51], *BABAR* [52,53] (for which we perform for the first time a full study of the published results), and CMD-3 (for which we estimate the a_{μ}^{HVP} contribution using the data in the ancillary files of Ref. [45]).

Finally, we also include in the final fit the most recent SND 2020 result [54], the results from CMD-2 [55–57] and the older SND 2006 measurement [58].

This work extends the study of Ref. [38] in several important ways. First, we include a simulation of the *BABAR* analyses in the MadGraph5_aMC@NLO platform [59] that uses the muon method for the determination of σ_{had} [52,53].[4] Secondly, we estimate the contribution to a_{μ}^{HVP} in the intermediate energy window from the CMD-3 measurement reported in Ref. [45], and we compare this new piece of information with the corresponding results from other experiments and from the lattice. Third, we

[1]The tension between the dispersive method [35] and the individual results of different lattice collaborations is around 4σ [25,27,28,34]. Reference [28] quotes a 4.5σ tension for the combined BMW, CLS/Mainz, and ETMC results neglecting correlations. Reference [36] quotes a 3.8σ tension for the combined BMW, CLS/Mainz, ETMC, and RBC/UKQCD assuming 100% correlation.

[2]The need to resort to indirect effects on the measurement of σ_{had} to account for the a_{μ} discrepancies is also warranted by the fact that, as was argued in Ref. [39] (see also Ref. [40]), the possibility of solving the lattice/data-driven discrepancy relying only on direct NP contributions to the $e^+e^- \to$ hadrons process is excluded by a number of experimental constraints, as for example the global electroweak fits, which would enter in serious tension with observations, because of modifications of the hadronic contribution to the running of the fine-structure constant [41–44].

[3]It should be remarked that a recent study of higher-order radiative processes in $e^+e^- \to \mu^+\mu^-\gamma$ and $e^+e^- \to \pi^+\pi^-\gamma$ events performed by the *BABAR* Collaboration at NNLO [46], pointed out that the different treatment of these events adopted by different experiments that determine $\sigma_{\mathrm{had}}(s)$ thorough the radiative method may explain the discrepancy between KLOE and *BABAR*. However, the discrepancy with CMD-3 that is using the scanning method cannot be explained in a similar way.

[4]The *BABAR* analysis requires a visible photon in the detector acceptance and two reconstructed muons tracks. The corresponding NP process is thus $e^+e^- \to \gamma V$, along with the semivisible V decay. More details about the reconstruction procedure used for *BABAR* are given in Appendix A.

refine the simulation of the experimental efficiencies of the NP signal.

All in all, we find that within our scenario a good consistency between $(g-2)_\mu$ values inferred from lattice, data-driven, and experimental results can be recovered, although the consistency between the different $\sigma_{\rm had}$ datasets can only be marginally improved. In particular, we find that the dominant repercussions from NP processes on the determination of $\sigma_{\rm had}$ are confined to the low- and intermediate-energy windows while the high-energy window remains largely unaffected, which is in nice agreement with lattice indications. A final remark is in order regarding the recent CMD-3 measurement of $\sigma_{\pi^+\pi^-}$. Since this measurement is performed with the energy-scanning method in the energy range $\sqrt{s}\in[0.6, 0.88]$ GeV [45], that is at c.m. energies well below the V resonance, it is clear that their result cannot be affected by the NP construction of Ref. [38]. Hence, in our scenario the good agreement between CMD-3 and the lattice results, and the marked disagreement with other experiments performed with the radiative method at $\sqrt{s}\sim M_V$ or at $\sqrt{s}\gg M_V$, has a natural explanation.

In Sec. II we briefly summarize the experimental and lattice status and introduce the time windows' kernels. In Sec. III we discuss the various indirect effect that GeV-scale NP can have on the determination of $\sigma_{\rm had}$ with the dispersive approach. In Sec. IV we introduce a phenomenological NP scenario wherein a viable solution to all the a_μ window discrepancies can be provided. Finally, in Sec. V we draw our conclusions.

II. TIME WINDOWS' KERNELS AND THE DATA-DRIVEN APPROACH

The full HVP contribution to a_μ can be decomposed as the sum of three terms corresponding to the three windows SD, W, and LD [26],

$$a_\mu^{\rm HVP} \equiv a_{\rm SD}^{\rm HVP} + a_{\rm W}^{\rm HVP} + a_{\rm LD}^{\rm HVP}, \tag{5}$$

that on the QCD lattice correspond to different Euclidean time windows. Each term is obtained by modifying the integration kernel using predefined smooth step functions in order to exponentially suppress contributions from other regions. In the time-momentum representation, the HVP contribution to the muon anomalous magnetic moment is

$$a_\mu^{\rm HVP} = \left(\frac{\alpha}{\pi}\right)^2 \int_0^\infty dt \tilde{K}(t) G(t), \tag{6}$$

where $\tilde{K}(t)$ is a kernel function (given in Appendix B of Ref. [60]) and $G(t)$ is given by the correlator of two electromagnetic currents. The windows in Euclidean time are defined by means of an additional weight function:

$$\begin{aligned}
\Theta_{\rm SD}(t) &= 1-\Theta(t,t_0,\Delta),\\
\Theta_{\rm W}(t) &= \Theta(t,t_0,\Delta)-\Theta(t,t_1,\Delta),\\
\Theta_{\rm LD}(t) &= \Theta(t,t_1,\Delta),\\
\Theta(t,t',\Delta) &= \frac{1}{2}\left(1+\tanh\frac{t-t'}{\Delta}\right),
\end{aligned} \tag{7}$$

with parameters $t_0 = 0.4$ fm, $t_1 = 1.0$ fm, and $\Delta = 0.15$ fm. In order to compare with the data-driven approach, the HVP contributions can be written as

$$a_i^{\rm HVP} = \frac{1}{2\pi^3}\int_{E_{\rm th}} E\hat{K}\left(\frac{E}{m_\mu}\right)\sigma_{\rm had}(E)\hat{\Theta}_i(E)dE, \tag{8}$$

where E is the e^+e^- c.m. energy, $E_{\rm th}$ the threshold energy, $\hat{K}(x) = \int_0^1 dy \frac{(1-y)y^2}{y^2+(1-y)x^2}$ is the kernel function, m_μ the muon mass, $\sigma_{\rm had} = \sigma(e^+e^-\to \text{hadrons})$, the index $i = \text{LD, W, SD}$ refers to a specific window, and [28,35]

$$\hat{\Theta}_i(E) = \frac{\int_0^\infty dt t^2 e^{-Et} K(m_\mu t)\Theta_i(t)}{\int_0^\infty dt t^2 e^{-Et} K(m_\mu t)}, \tag{9}$$

where the kernel function is defined as

TABLE I. Results for the short-distance, intermediate- and long-distance windows contributions and for the total $a_\mu^{\rm HVP}$ contribution to a_μ. The results for the data-driven approach [35] are given in the first line, and QCD lattice results [25,27,28,34] in the following lines. The lattice average from Ref. [36] assumes 100% correlation. All numbers are in units of 10^{-10}.

	$a_{\rm SD}^{\rm HVP}$	$a_{\rm W}^{\rm HVP}$	$a_{\rm LD}^{\rm HVP}$	$a_\mu^{\rm HVP}$
Data-driven [35]	68.4(5)	229.4(1.4)	395.1(2.4)	693.0(3.9)
BMW [25]	⋯	236.7(1.4)	⋯	707.5(5.5)
Mainz/CLS [27]	⋯	237.30(1.46)	⋯	⋯
ETMC [28]	69.27(34)	236.3(1.3)	⋯	⋯
RBC/UKQCD [34]	⋯	235.56(82)	⋯	⋯
Lattice average [36]	⋯	236.16(1.09)	⋯	⋯

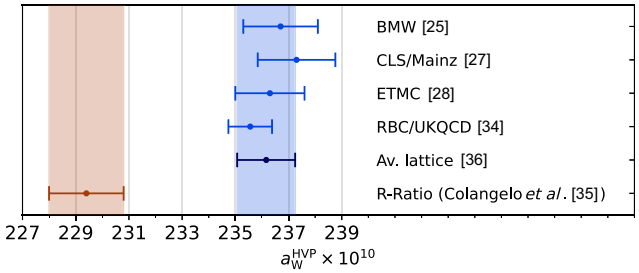

FIG. 1. Comparison of the results for the intermediate window contribution to $a_\mu^{\rm HVP}$ from the data-driven approach [35] (region in red) and from four lattice computations (BMW [25], CLS/Mainz [27] and ETMC [28], RBC/UKQCD [34] in blue). The black point and the blue region correspond to the average given in Ref. [36] of the four lattice results assuming conservatively a 100% correlation.

$$K(z)=2\int_0^1 dy(1-y)\left[1-j_0^2\left(\frac{zy}{2\sqrt{1-y}}\right)\right],$$

with $j_0(x)=\sin(x)/x$.

In Table I, we collect the results for the windows contributions to the HVP taken from Refs. [25,27,28,34,35]. We see that while there are no indications of sizeable discrepancies in the short distance window, in the intermediate window the disagreement between the lattice and the data-driven results is remarkable. Figure 1 summarizes the situation for the intermediate window. The data-driven result of [35] corresponds to the red data point and shaded region, the four lattice computations [25,27,28,34] correspond to the black data points, and their average [36] obtained by assuming conservatively a 100% correlation correspond to the blue data point and shaded region. It is clear that the 2.1σ discrepancy between the data-driven and the BMW lattice evaluations of the total $a_\mu^{\rm HVP}$ is exacerbated in the intermediate window. Most importantly, the BMW result for $a_{\rm W}^{\rm HVP}$ is confirmed independently by the recent results of the Mainz/CLS [27], ETMC [28], RBC/UKQCD [34], and Fermilab, HPQCD, and MILC [33] lattice Collaborations, strengthening the confidence in the reliability of the lattice approach.

III. GeV-SCALE NEW PHYSICS INDIRECT EFFECTS ON THE DATA-DRIVEN DETERMINATION

The data-driven method relies on the assumption that all the processes that concur to determine $\sigma_{\rm had}$ are SM processes. In presence of GeV-scale new physics, and in particular in the case the relevant couplings are sizeable enough to affect the muon $(g-2)_\mu$, this hypothesis can fail at multiple levels.

To give an example, when the hadronic cross section is directly inferred from the detected number of hadronic events, a relation similar to the following one is generally used:

$$\frac{d\sigma_{\rm had}}{ds'}=\frac{N_{\rm had}-N_{\rm bkd}}{\epsilon(s')\mathcal{L}(s')},\tag{10}$$

where $N_{\rm had}$ is the measured number of hadronic final states produced in e^+e^- annihilation with final-state invariant mass $\sqrt{s'}$, $N_{\rm bkd}$ is the estimated number of background events, $\epsilon(s')$ is the detection efficiency, and $\mathcal{L}(s')$ the luminosity for final states with invariant mass $\sqrt{s'}$, corresponding to the c.m. energy of the collision (which, in experiments that exploit initial state-radiated photons to scan over s', is different from the electron/positron c.m. beam energy $\sqrt{s}$). A crucial observation is that each one of these quantities can be affected by the presence of NP. In particular $\mathcal{L}(s')$ is generally determined by comparing the measurements of some other process (e.g., Bhabha scattering) with the SM theoretical expectation.

A. Luminosity estimate from Bhabha scattering

In order to extract the hadronic cross-section $\sigma_{\rm had}$ at the subpercent level, the experimental luminosity must be precisely measured.

In the two earlier analysis of the KLOE Collaboration [48,49] (referred to as KLOE08 and KLOE10), the luminosity was inferred by comparing measurements of the Bhabha cross section $e^+e^-\to e^+e^-$ at large angles with the SM predictions from high precision Bhabha event generators [61–63]. As discussed extensively in [38], a new physics contribution to Bhabha scattering able to affect the determination of $\sigma_{\rm had}$ at the required level via an incorrect determination of $\mathcal{L}(s')$, should be of the order of $\sim\mathcal{O}(1\ \text{nb})$. This can be obtained for instance by resonantly producing a new boson of mass close to the KLOE8/KLOE10 c.m. energies. In the following, we will assume a phenomenological setup in which a dark photon (DP) V with a mass around the GeV, decays semivisibly yielding the required excess of e^+e^- events, together with some missing/energy momentum associated with additional invisible decay products.

A remark is in order regarding the present treatment of the BESIII dataset with respect to our previous analysis [38]. The most recent measurement of BESIII [51] relies on Bhabha scattering to calibrate the luminosity instead of the ratio with muon final states (both methods are used in the paper, but the Bhabha approach is eventually preferred due to the smaller experimental errors). Thus, the overall shift that we obtain for BESIII in the present analysis is sizeably smaller than in our previous analysis. However, this does not impact strongly the overall fit due to the relatively large error bars of the BESIII [51] measurement compared to the KLOE and *BABAR* results.

B. The $\sigma(\mu\mu\gamma)$ method

More recent measurements, including KLOE12 [50] and *BABAR* [53] estimate the luminosity directly from the number of di-muon final states which can be collected along with a $\pi\pi\gamma$ dataset. While this approach allows for the approximate cancellation of many systematic uncertainties, it relies much more critically on the SM-only hypothesis due the much smaller SM $\mu\mu\gamma$ cross section compared to Bhabha. NP effects do not require a tuning of the masses of the new particles involved to be relevant for affecting analyses that adopt this strategy. Relevant effects can be quite generically expected for any GeV-scale new boson with a coupling to muons sufficiently large to explain the $(g-2)_\mu$ anomaly via new loop contributions. Since $\sigma_{\rm had}$ is eventually obtained by multiplying the ratio of events $N_{\pi\pi\gamma}/N_{\mu\mu\gamma}$ by the theoretical SM muon production cross section $\sigma^{\rm SM}_{\mu\mu}$, any excess of NP-related $\mu\mu X$ events must be subtracted from the dataset in order to obtain the correct value of $\sigma_{\rm had}$. If, under the assumption of SM μ-production only, this is not done, the inferred value of $\sigma_{\rm had}$ will be lower, given that the NP-related $\mu^+\mu^-$ events contribute to the normalization factor of the hadronic events. If instead the NP origin of some $\mu^+\mu^-$ events is accounted for, the net effect is to increase the inferred value of $\sigma_{\rm had}$.

In a general DP model, decays of the hypothetical boson, besides additional $\mu^+\mu^-$ events, may also lead to additional hadronic final states. However, the fact that at the ρ/ω peak the SM-rate for the hadronic processes are larger than that for $\mu^+\mu^-$ production by an order of magnitude, implies that the NP effects in the hadronic channel are much less important than in the muon channel. To clarify further this point, let us consider two-pion events of NP origin that, because of the missing energy/momentum associated with the V invisible decay products, are reconstructed at invariant masses laying within the ρ region. These events are produced via excitation of the V resonance at $\sqrt{s}\sim 1$ GeV and as a result, their rate is not increased by the hadronic resonances that require $\sqrt{s'}\sim m_\rho$ to contribute to enhance the two-pion channel.[5]

If we assume universality for the V couplings (for example proportionality to the electromagnetic charge), than the number of muonic and hadronic final states of NP origin that are reconstructed with invariant mass within the ρ region will be roughly comparable. As a result, this effect will be much more significant for the μ channel than for the π channel. The dominant effect of V-quark interactions is in fact that of reducing the branching ratio for decays yielding dimuons in the final state. A naive estimate of this effect by accounting for hadronic events induced by off shell V^* mixing with the ρ leads to a reduction of the NP shift by roughly one half. However, an additional complication is that if NP contributes to the hadronic channel, then the inferred $\sigma_{\rm had}=\sigma^{\rm QED}_{\rm had}+\sigma^{\rm NP}_{\rm had}$ cannot be directly related to the photon HVP, since one should first extract the pure QED contribution. While keeping in mind these caveats, in Sec. IV we will simply adopt a phenomenological model in which it is assumed that the V coupling to the pions are sufficiently suppressed with respect to the couplings to the charged leptons, leaving for future work a detailed estimate of hadronic effects in the representative case in which V-quark couplings are fixed by some universality condition.

C. Background subtraction

The precise subtraction of background events is a key issue in most experimental analyses, since only the true hadronic final states must be retained, while spurious events must be identified and rejected. For instance, an important background considered by KLOE for the analysis using the $\mu\mu\gamma$ events is the $\pi^-\pi^+\pi^0$ final state, as well as $\mu\mu\gamma$ events that can also be misidentified as $\pi^-\pi^+\gamma$. This issue may be particularly important for the KLOE analysis, due to the fact that the particle identification between muons and pions relied on the so-called computed track mass m_{tr}. The latter is defined in terms of the momenta p_+ and p_- of the reconstructed positively and negatively charged tracks, based on the energy-conservation relation

$$\left(\sqrt{s}-\sqrt{|\vec{p}_+|^2+m_{tr}^2}-\sqrt{|\vec{p}_-|^2+m_{tr}^2}\right)^2-|\vec{p}_-+\vec{p}_+|^2=0, \tag{11}$$

which assumes a SM process containing a real photon with $E_\gamma=|\vec{p}_\gamma|=-|\vec{p}_-+\vec{p}_+|$. For NP events with additional missing energy, as for example the four-body process $e^+e^-\to V\to\mu^+\mu^-\chi_1\chi_1$ where V is a hypothetical DP produced on shell whose decay products also contain the invisible particle χ_1, energy conservation would imply the replacement $|\vec{p}_-+\vec{p}_+|^2\to(E_\chi+E_{\chi'})^2$. Experimentally, muons and pions from these events would thus tend to yield track mass solutions with values somewhat larger than for the SM process. In fact, a full simulation shows that the track mass distribution for our $\mu\mu\chi_1\chi_1$ final states are roughly flat, with a lower threshold at the muon mass. As the KLOE Collaboration eventually rely on a fit on Monte Carlo (MC)-based distributions to distinguish the $\mu\mu\gamma$, $\pi\pi\gamma$, and $\pi^+\pi^-\pi^0$ sample, the effect of injecting a NP signal which does not directly match any of these distributions cannot be easily estimated, in the lack of access to the simulation tools used by the collaboration.

[5]In fact, the attempt to fit the anomaly via the direct interference of NP contributions to the $e^+e^-\to$ hadrons process carried out in Ref. [39] (see also Refs. [40,64]) requires (using our notations) $|\epsilon_e(\epsilon_u-\epsilon_d)|\sim 2\times 10^{-2}$; that is, couplings that are much larger than those required to render effective our indirect effects.

D. Efficiencies

All analyses rely to a certain extent to a tag-and-probe approach to derive their efficiencies from the data. This two step process works as follows:

(i) The selection cuts are applied on the dataset, requiring only one μ/π track. When available, particle identification requirements are made more stringent on this track. This forms the data control sample.
(ii) A kinematic fit is performed on the track along with the reconstructed photon to determine the most likely localization for the opposite charge μ/π, with the tagging and reconstruction efficiencies obtained by comparing with the reconstructed event.

Critically, the differences between the MC and the data on this sample are eventually used to apply a mass-dependent data/MC correction. If NP events are included in the data control sample, the efficiency estimate will be biased. Additionally, since the NP does not necessarily treat $\pi\pi$ and $\mu\mu$ final states on equal footing, there is no reason to expect that the efficiencies will cancel in the ratios between $\pi\pi$ and $\mu\mu$ events, as is broadly expected in the SM. In the KLOE12 analysis, these efficiencies have been found to agree with the MC simulation within a few per mil, leaving little room for a significant NP effect. On the other hand, corrections at a few percents are used in the *BABAR* analysis. A complete study of the potential NP effects in the determination of the efficiencies should be undertaken directly by the experimental collaborations.

IV. EXPLICIT MODEL AND RESULTS

An example of how the NP contributions to Bhabha and $e^+e^- \to \mu^+\mu^-$ events needed to account for the various a_μ-related discrepancies (while evading all other experimental constraints) was provided in Ref. [38], that adopted the inelastic dark matter model of Refs. [65–67] in which a dark Abelian gauge group $U(1)_D$, kinetically mixed with $U(1)_{\rm QED}$, is spontaneously broken by the vacuum expectation value of a dark Higgs S. For simplicity here we assume a simple phenomenological setup along the lines of the model of Ref. [68] in which a new $J^P = 1^-$ vector V couples to a current J^μ_V that is a linear combination of the SM fermion currents. In addition, we assume that V has also an off-diagonal coupling to two Majorana dark fermions $\chi_{1,2}$,

$$\mathcal{L} \supset -eV_\mu J^\mu_V - g_D V_\mu \bar{\chi}_2 \gamma^\mu \chi_1, \quad J^\mu_V = \sum_{i=u,d,\ell,\nu} \epsilon_i \bar{f}_i \gamma^\mu f_i, \tag{12}$$

where e is the usual QED coupling while g_D is the V coupling to the dark sector fermions. For the parameters appearing in J^V_μ we assume $\epsilon_i \ll 1$ and, in particular, a relative suppression of the neutrino with respect to the charged leptons couplings, sufficient to evade the constraints from neutrino trident production [69] as well as other neutrino-related constraints [70], that is $\epsilon_\nu < \epsilon_\mu \simeq \epsilon_e \equiv \varepsilon$ (where the parameter ε should not be confused with ϵ that refers instead to the efficiency). We also assume a certain suppression of the V couplings to the light quarks ($\epsilon_{u,d} \lesssim \varepsilon$) to justify the approximation discussed in the previous section of neglecting NP contributions to $e^+e^- \to$ hadrons. Finally, the two Majorana dark fermions are characterised by a certain mass splitting Δm_χ and, in particular, the lightest one may also play the role of a dark matter particle. This model provides all the conditions required to shift the luminosity estimates based on measurements of Bhabha events and to generate additional dimuon events. Since the most worrisome discrepancy among the various datasets is represented by the low values of $\sigma_{\rm had}$ reported by the three KLOE analyses, we fix the V mass close to the KLOE c.m. energy. More precisely, since for KLOE08/12 $\sqrt{s} = 1.020$ GeV while for KLOE10 $\sqrt{s} = 1.0$ GeV, we fix $M_V = 1.001$ GeV so that the KLOE10 measurement is also affected by NP events.

In order to avoid bounds from light resonances searches, the DP main decay channel must consist of multibody final states, including a certain amount of missing energy. With $m_{\chi_1} + m_{\chi_2} < M_V$, V decays proceed mainly via the chain $V \to \chi_1\chi_2 \to \chi_1\chi_1 e^+e^-(\mu^+\mu^-)$, with $\mathrm{BR}(V \to \chi_1\chi_2) \sim 100\%$ and $\mathrm{BR}(V \to e^+e^-, \mu^+\mu^-) \propto \varepsilon^2$. In particular, to ensure that the e^+e^- and $\mu^+\mu^-$ events from V decays will carry away sufficient energy to populate the datasets after the experimental cuts, we choose $m_{\chi_2} \sim 0.97 M_V \gg m_{\chi_1} \sim 3$ MeV (the dark matter mass does not play a critical role as long as $m_{\chi_1} \lesssim \mathcal{O}(10)$ MeV).

This is possible by assuming that a new boson produced resonantly around the KLOE c.m. energy decays promptly yielding e^+e^- and $\mu^+\mu^-$ pairs in the final state. This can give rise to three different effects:

(1) The additional e^+e^- events will affect the KLOE luminosity determination based on measurements of the Bhabha cross section, and in turn the inferred value of $\sigma_{\rm had}$.
(2) The additional $\mu^+\mu^-$ events will affect the determination of $\sigma_{\rm had}$ via the (luminosity-independent) measurement of the ratio of $\pi^+\pi^-\gamma$ versus $\mu^+\mu^-\gamma$ events.
(3) Loops involving the new boson would give a direct contribution to the predicted value of a_μ.

All these effects were discussed in detail in Ref. [38], where it was concluded that a new gauge boson V of mass close to the mass of the ϕ meson $M_V \sim M_\phi \simeq 1.020$ GeV was able to release the tensions between the KLOE and *BABAR* results for $\sigma_{\rm had}$, the data-driven and lattice determinations of $a_\mu^{\rm HVP}$, and the measured values of a_μ with the theoretical prediction, without conflicting with other phenomenological

constraints. Yet, in Ref. [38] a complete agreement among all the datasets could not be reached, and this was essentially due to the fact that one of the three KLOE measurements of $\sigma_{\rm had}$ (commonly referred as KLOE10 [49]) was performed at a c.m. energy 20 MeV below the V resonance, thus remaining unaffected by the NP.

While we adopt the same theoretical model of Ref. [38], here we shift downwards by a few MeV the location of the V resonance. With respect to the analysis in Ref. [38] such a small shift leaves the NP effects on the *BABAR* dataset essentially unmodified, because it operated at a c.m. energy much larger than M_V ($\sqrt{s} = 10.6$ GeV). For KLOE08 and KLOE12, that collected data at the ϕ resonance ($\sqrt{s} = 1.02$ GeV), the NP effects are somewhat reduced but, due to radiative return on the nearby V resonance, are still significant. Most importantly, now the KLOE10 luminosity measurement get also affected by additional Bhabha events of NP origin, resolving the tension observed in Ref. [38] between the KLOE10 and KLOE08/KLOE12 determinations of $\sigma_{\rm had}$.

A. Window anomaly and GeV-scale new physics

Once the masses of the NP particles have been fixed, we compute the absolute shift for KLOE08, KLOE10, KLOE12, and *BABAR* as the function of the coupling ε. Table II shows the shift due to DP related events for the different experiments in the energy region $\sqrt{s'} \in [0.6, 0.9]$ GeV (in which all the experiments have provided the relevant information) and for $\varepsilon = 0.0125$. It can be seen that the indirect NP contribution in the short-distance window is negligible with respect to the ones in the intermediate and long distance windows. This is because the corresponding weight function strongly suppresses the SD NP contribution.

Figure 2 shows the evolution of the shift in the contributions to $a_\mu^{\rm HVP}$ from the intermediate window, resulting from our NP model, as function of the parameter ε. The dark region at $\varepsilon \geq 0.027$ is excluded by the limit

TABLE II. Indirect new physics contribution to $a_\mu^{\rm HVP}$ for the $\pi\pi$ channel in the short-, intermediate-, and long-distance windows for KLOE08 [48], KLOE10 [49], KLOE12 [50], and *BABAR* [52,53], in units of $10^{-10}(\varepsilon/0.0125)^2$ and in the range $\sqrt{s'} = [0.6 - 0.9]$ GeV. The last column shows the total $\pi\pi$ channel SM contribution in units of 10^{-10}, computed in the relevant $\sqrt{s'}$ range from the data publicly available in the literature. The theoretical errors are not shown (see the discussion in the main text).

	$a_{\rm SD}^{\rm HVP,NP}$	$a_{\rm W}^{\rm HVP,NP}$	$a_{\rm LD}^{\rm HVP,NP}$	$a_{\rm total}^{\rm HVP,SM}$
KLOE08	0.03	0.31	0.69	368.7
KLOE10	0.65	6.66	15.07	366.0
KLOE12	0.06	0.63	1.43	366.6
BABAR	0.59	6.67	15.68	376.7
CMD-3	⋯	⋯	⋯	383.7

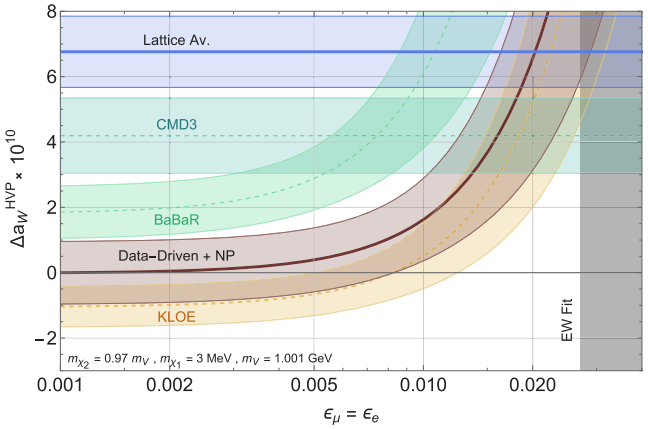

FIG. 2. Theoretical estimate for the shift in a_μ in the intermediate window compared to the data-driven result for KLOE (orange region), *BABAR* (green region), CMD-3 (aquamarine region) and for the full data-driven combination (dark red region). All the bands show the 1σ regions. $\Delta a_{\rm W}^{\rm HVP} = a_{\rm W}^{\rm HVP}(\varepsilon) - a_{\rm W}^{\rm HVP}(0)$ in the vertical axis is the NP contribution with respect to the SM data-driven estimate. The lattice result from Ref. [36] is shown in blue. For reference we also depict in dark gray the excluded region from LEP for the case of a kinetically mixed dark photon model.

from electroweak precision observables derived in Refs. [71,72] for the case in which V is a kinetically mixed dark photon. Note that in any case the direct-loop contribution to a_μ significantly overshoots the experimental value. We did not include in our estimates possible effects from background subtraction and corrections to the efficiency, as they cannot be estimated in a reliable way. For this reasons we believe that the NP effects are likely underestimated.

Several comments are in order. In first place, the shifts due to NP are estimated only for the energy range $\sqrt{s'} = [0.6, 0.9]$ GeV, which corresponds to about half of the HVP contribution in the intermediate window.[6] For $\sqrt{s'}$ above the GeV no NP effects are expected, since e^+e^- ($\mu^+\mu^-$) events with such an invariant mass cannot originate from the V resonance with $M_V \sim 1$ GeV. However, the range $\sqrt{s'} < 0.6$ GeV will also contribute to the overall shift. In order to account for (1) the missing data relative to the $\sqrt{s'} < 0.6$ GeV range, (2) the potential NP effects on background subtractions and on the efficiency calibration procedure, and (3) the effect of experimental smearing described in

[6]We do not include the results on $e^+e^- \to KK$ provided by the SND [73] and CMD-2 [74] Collaborations. Since $e^+e^- \to KK$ is dominated by the ϕ resonance peak, similarly to KLOE08 also these results will likely be shifted due to their calibration of the luminosity via Bhabha scattering events. However, we can expect that the overall effects of these additional measurements will be small, due to both the large experimental errors and the small contribution from KK final states to the total HVP.

Appendix A, we include a 50% theoretical error on the overall size of the shifts generated by the indirect NP effects.

To estimate the window contribution $a_{\rm W}^{\rm HVP}$ from the CMD-3 data in the energy range $\sqrt{s} \in [0.6, 0.9]$ GeV we have used the values of the pion form factor $|F_\pi|^2$ given in the ancillary files of Ref. [45] to compute the two-pion cross section, from which the $a_{\rm W}^{\rm HVP}$ contribution is derived. We obtain

$$a_{\rm W}^{\rm HVP,CMD\text{-}3}\Big|_{\sqrt{s}/\mathrm{GeV} \in [0.6,0.9]} = 114.5(1.2) \times 10^{-10}. \quad (13)$$

Since in our scenario the evaluation of $a_\mu^{\rm HVP}$ from CMD-3 data does not receive NP contributions ($a_{\rm W}^{\rm HVP,NP} = 0$) the value in Eq. (13) remains constant across all values of ε, see Fig. 2.

An important improvement of the present study with respect to Ref. [38] is that the residual internal discrepancy within the KLOE experimental datasets that was formerly observed in correspondence with the largest values of ε is now resolved. This is due to the fact that in Ref. [38] the V mass was fixed at the value $M_V \simeq M_\phi$, so that the NP was affecting KLOE08 and KLOE12, but not the KLOE10 data, that were taken at a c.m. energy 20 MeV below the ϕ resonance. Once ISR effects are included, the slightly lower value $M_V \simeq M_\phi - 17$ MeV adopted in this paper is enough to mitigate this issue, since now all the three KLOE datasets include ε-dependent NP contributions.

Finally, our estimate of the $a_{\rm W}^{\rm HVP}$ for CMD-3 in their full range $0.327 \le \sqrt{s}/\mathrm{GeV} \le 1.199$ is

$$a_{\rm W}^{\rm HVP,CMD\text{-}3}\Big|_{\sqrt{s} \le 1.2\ \mathrm{GeV}} = 139.4(1.6) \times 10^{-10}. \quad (14)$$

B. Internal discrepancies of the $\sigma_{\rm had}$ datasets

The measurements of $\sigma_{\rm had}$ performed with the energy scanning method is not affected by the NP when the data points are taken at $\sqrt{s} < M_V$. This is the case for the data points used by the CMD-3 Collaboration to compare their results with the ones of the other experiments, which fall in the interval $\sqrt{s} \in [0.60, 0.88]$ GeV [45]. Note that the CMD-3 result is a couple of σ above *BABAR*, several σ above the combined KLOE result, and quite close to the lattice estimate, in nice agreement with what is expected in our scenario. However, for the same reason also the CMD-2 [55–57] and SND [54,58] results are not modified. This implies that the NP scenarios discussed above cannot mitigate the discrepancy between CMD-3 and CMD-2/SND datasets.

Altogether, nine experimental results are included in our fit. While the $\sqrt{s}$ regions probed by each analysis differ, they all overlap in the $\sqrt{s} \in [0.60, 0.88]$ interval. As a first measure of the (dis)agreement between the different determinations of $\sigma_{\rm had}$, we estimate the p-value of $a_\mu^{\rm HVP}$ for the corresponding datasets, within the overlapping c.m. energy range.[7] We find that in the SM-only hypothesis, the datasets present a discrepancy at the $\sim 4.5\sigma$ level. Note that this p-value is not linked directly to the window anomaly, and simply represents a test of the self-consistency of the different datasets under the SM-only hypothesis.

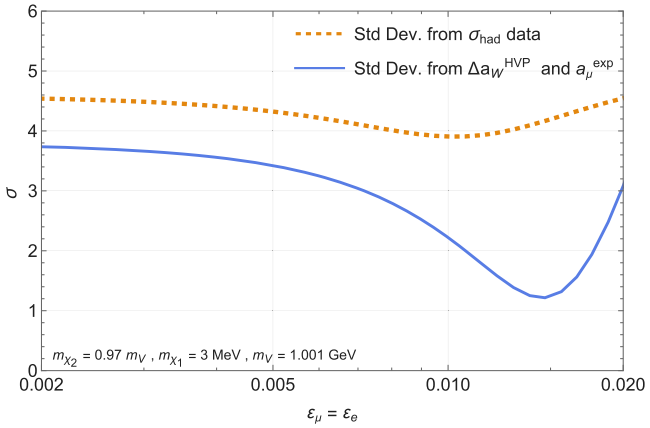

FIG. 3. Standard deviation of the global fits to the internal consistency of the data-driven dataset, and to the window anomaly as a function of ε for $m_{\chi_1} = 3$ MeV, $m_{\chi_2} = 0.97 m_V$, $m_V = 1.001$ GeV, and $\alpha_D = 0.5$.

We have also estimated the likelihood including the consistency in the intermediate window of the averaged data-driven result $a_{\rm W}^{\rm HVP}|_{\rm data}$ with the lattice estimate $a_{\rm W}^{\rm HVP}|_{\rm lat}$, as well as the consistency between the FNAL experimental result $a_\mu^{\rm exp}$ and the prediction based on the BMW lattice result. For this latter case there is substantial agreement with the SM-only case, however, our NP scenario also implies a direct (loop-induced) contribution to a_μ, that should be removed from $a_\mu^{\rm exp}$ to obtain the SM-only prediction based on the BMW's HVP result.

We show in Fig. 3 the result of both global fits, expressing for convenience the p-value in term of standard deviations for our benchmark parameter point. The window discrepancy corresponds to the blue curve, and we see that for values of the coupling above $\varepsilon \sim 0.012$ it decreases well below 2σ. In contrast, the inner tension within the experimental datasets, that correspond to the orange dashed curve, is only very mildly mitigated regardless the value of ε. Finally, in Fig. 4 we illustrate the impact of the direct and indirect NP effects on the overall scenario of the a_μ related discrepancies. We see that a remarkable improvement in the agreement among the various theoretical and experimental determinations can be obtained for $\varepsilon \sim O(1\%)$.

The fact that our mechanism can solve the discrepancies between lattice and data-driven results both for the total and intermediate windows HVP contributions without invoking modifications of $\sigma_{\rm had}$ above 1 GeV, is consistent with the

[7]Note that the KLOE08 and KLOE12 measurements are partially correlated [50].

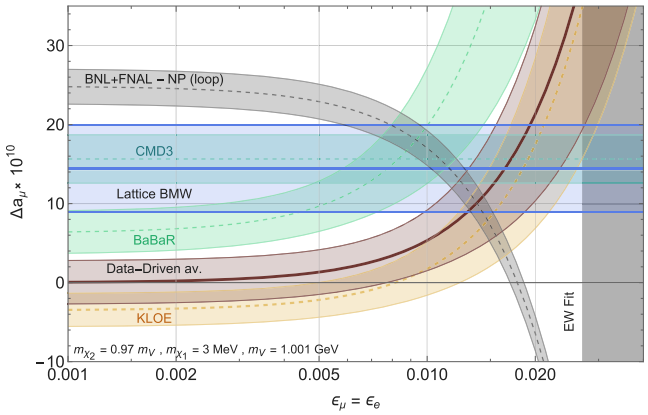

FIG. 4. Shifts in the prediction of the data-driven results from KLOE (orange band), *BABAR* (green band), CMD-3 (aquamarine band) and from the average of the e^+e^- data (brown band), as a function of ε for $m_{\chi_1} = 3$ MeV, $m_{\chi_2} = 0.97 m_V$, $m_V = 1.001$ GeV, and $\alpha_D = 0.5$. The gray band corresponds to the experimental world average Eq. (3) with the direct NP contribution to a_μ from NP loops subtracted out. For reference we also depict in dark gray the excluded region from LEP for the case of a kinetically mixed dark photon.

literature on this topic [75] in that it relies on a local change within the $\sqrt{s} \in [0.60, 0.9]$ region, and not just on an uniform shift in the low-energy $\pi\pi$ region. This possibility has been also confirmed by the recent dedicated analysis presented in Ref. [76], where it is shown that the results of the dispersive method can be reconciled with those of the lattice approach by modifying the hadronic experimental spectrum in the ρ region. On the other hand, our model does not succeed in reconciling all the discrepancies between the various $\sigma_{\rm had}$ datasets and, as we have seen, the improvement with respect to the SM-only hypothesis remains marginal for all values of ε. This issue has two facets. The first one is that the NP shift is larger for *BABAR* (and KLOE10) than for KLOE08/12. However, without NP corrections *BABAR* results are already above KLOE. This situation can be improved by adding another dark photon sitting at a somehow larger mass, corresponding to the KLOE08/12 c.m. energy. We have estimated that with this setup the overall tension in the $\sigma_{\rm had}$ dataset can be reduced to around 2.6σ, see Appendix B. The fact that the overall agreement cannot be further improved is because of the additional large discrepancies that are present between the results of experiments that rely on the energy-scanning method, and in particular CMD-3 vs CMD-2/SND. Indeed these experiments would be affected in similar ways in any NP scenario, we thus believe that this second facet points towards the possibility of experimental problems.

V. CONCLUSIONS

In this work we have studied the possibility of exploiting NP processes to solve the discrepancy between the dispersive and lattice estimates of $a_\mu^{\rm HVP}$ in the so-called intermediate window, that has persisted throughout different lattice determinations, and has by now reached a worrisome level of significance. The key NP ingredient is a GeV-scale new boson which is produced in e^+e^- collisions, and that decays promptly into e^+e^-, $\mu^+\mu^-+$ missing energy. The amount of missing energy in the decay final state must be sufficient to avoid current constraints, and is provided by light dark sector fermions produced in the decay chain. This NP scenario can affect the prediction for a_μ directly via new loop contributions, but most interestingly it can indirectly affect the determination from experimental data of $\sigma_{\rm had}$, which is used to derive the theoretical prediction for a_μ by means of the dispersive approach. We have argued that our simple phenomenological model can sizeably reduce the 'window' tension, to a significance at the 1σ level. We believe that this result can pave the way to build UV complete models able to explain the $(g-2)_\mu$ anomalies. Additionally, with respect to Ref. [38], the present work improves the technical aspects of the estimation of the indirect effects of GeV-scale NP on the data-driven $a_\mu^{\rm HVP}$ in several important directions. In particular, (i) we made a quantitative estimate of the NP effects on $a_\mu^{\rm HVP}$ for the *BABAR* dataset, (ii) we included NLO effects which modified strongly the NP shifts on the KLOE dataset, and (iii) we performed a global fit of the data-driven data including all main existing analysis. We also made a quantitative estimate of the intermediate window contribution to $a_\mu^{\rm HVP}$ inferred from the recently published CMD-3 data [45] which, in agreement with what could have been expected in our scenario, does not show any particular tension with the lattice result.

From this study a certain number of conclusions can be drawn. First, indirect effects of new GeV-scale particles on the measurements used to estimate $a_\mu^{\rm HVP}$ via the dispersive method can have important consequences, and can be significantly more ubiquitous than what was anticipated in Ref. [38]. In particular, NP effects can bias the determination of $\sigma_{\rm had}$ via the efficiency calibration and background removal processes used by the experimental collaborations. Second, in our scenario data points collected at c.m. energies below 1 GeV are not affected by NP effects. This nicely explains the qualitative agreement of the recent CMD-3 result [45] with the lattice, as well as the significant discordance with the results of other experiments operating at $\sqrt{s} \gtrsim M_V$, that have measured $\sigma_{\rm had}$ in the same energy range by exploiting the radiative method. On the other hand, for the same reason also the CMD-2 [55–57] and SND [54,58] results are not modified by this NP, and thus the discrepancy with the CMD-3 result endures. Third, for experiments running at c.m. energies at or above 1 GeV, NP effects in the determination of $a_\mu^{\rm HVP}$ are negligible for the short-distance window, while they are sizeable in the intermediate (and likely also in the

long-distance) window. Fourth, the key ingredient to evade existing limits from searches of light, weakly coupled new particles, is the presence of a significant amount of missing energy associated with their visible decay products. At the same time, as long as e^+e^- ($\mu^+\mu^-$) pairs are present in the (multibody) final state, due to the not-too-tight cuts typically set by the experimental collaborations in measuring $\sigma_{\rm had}$, the missing energy requirement does not preclude significant NP effects on the determination of $a_\mu^{\rm HVP}$.

Note added. After the first submission of this paper to the arXiv in December 2022 several new results related to the $(g-2)_\mu$ anomalies, and in particular to the photon HVP appeared. In January 2023 the results of the Fermilab Lattice, HPQCD, and MILC [33] and RBC/UKQCD RBC/UKQCD [34] appeared, giving further support to the reliability of the lattice estimates of the HVP. Most importantly, in February 2023 the CMD-3 Collaboration announced their new result on the $e^+e^-\to\pi^+\pi^-$ cross section [45], which is in dramatic tension with the previous KLOE (and, to a lesser extent, also *BABAR*) result, while it is in broad agreement with the lattice. This nicely fits within our scenario and supports its plausibility. The original article has thus been expanded to account for these important results. Additionally, towards the completion of this revised version other new results appeared which also strengthen the conclusions of this work, and that have been integrated in the text as well. First, a new experimental measurement of the muon anomalous magnetic moment by the FNAL Muon $(g-2)$ experiment [24] confirmed the first measurement by the same collaboration, yielding the present world average in Eq. (3). Second, as was pointed out in footnote 3, a study of NNLO effects by the *BABAR* Collaboration [46] remarked that it may have a significant effect for some of the KLOE analyses, and this might reduce the tensions between the $\sigma_{\rm had}$ datasets that our model cannot explain satisfactorily. Finally, as was mentioned in Sec. IV B, the new study addressing the discrepancies between the lattice QCD and data-driven results presented in Ref. [76] has shown that the two approaches can be brought into agreement by modifying the ρ peak in the experimental spectrum. This is precisely what the indirect effects generated by our model do.

ACKNOWLEDGMENTS

We thank Bogdan Malaescu for providing us with information on the *BABAR* measurement of $\sigma_{\pi\pi\gamma_{\rm ISR}}$. This work has received support from the INFN "Iniziativa Specifica" Theoretical Astroparticle Physics (TAsP-LNF) and from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 101028626. The work of G. G. d. C. was supported by the Frascati National Laboratories (LNF) through a Cabibbo Fellowship, call 2019. The work of E. N. was supported by the Estonian Research Council Grant No. PRG1884.

APPENDIX A: EXPERIMENTAL CUTS AND SMEARING

We present the main ingredients of the recasting of the $e^+e^-\to e^+e^-(\gamma^{\rm ISR})$ and $e^+e^-\to\mu^+\mu^-(\gamma^{\rm ISR})$ experimental analysis for our NP events.

We have simulated all NP events in the MadGraph5_aMC@NLO framework, then smeared the momenta of the final states particles to reflect the experimental precision. For KLOE, we use [50]

$$\sigma_{p_T} = 0.4\% \times p_T,$$
$$\sigma_E = 5.7\% \times \sqrt{E(\text{GeV})}, \tag{A1}$$

with σ_{p_T} corresponding to the typical precision on charged momenta tracks and σ_E to the precision on the photon energy reconstruction (the polar angular precision is around 1° and is further included to obtain the energy of charged tracks).[8] For *BABAR*, we use only the energy smearing [77],

$$\sigma_E = (2.3\% \times E^{3/4}) \oplus (1.35\% \times E)(E \text{ in GeV}). \tag{A2}$$

The smearing effects typically broaden the $\sqrt{s'}$ range where NP effects are relevant, and increase the selection rates for NP events, since they tend to "hide" the associated missing energy. For all the experimental processes listed below, we have applied the selection cuts in two steps: first a "broad" selection with weaker cuts is applied at the MC truth level, then an exact selection with the experimental cuts is applied after momenta smearing for the charged and photon tracks. We have applied the same procedure on both NP and SM events, and then we have used the ratio of efficiencies to derive the final shifts on $a_\mu^{\rm HVP}$ and $a_{\rm W}^{\rm HVP}$.

Below we list for convenience the selection cuts for each analysis as reported by the experimental collaborations.

(a) *KLOE08 and KLOE10*. Both analysis [48,49] relied on Bhabha scattering to calibrate the luminosity. The experimental cuts are given by [78]; $|\cos\theta_{e^\pm}| < 0.57$, $E_{e^\pm} \in [0.3, 0.8]$ GeV, $|\vec{p}_{e^\pm}| \geq 0.4$ GeV and a cut is applied on the polar angle acollinearity of the e^+ and e^- charged tracks: $\zeta \equiv |\theta_{e^+} + \theta_{e^-} - 180°| < 9°$.

(b) *KLOE12*. The KLOE12 [50] analysis used kinematic cuts on the two muons polar angles $|\cos\theta_\mu| < 0.64$, momenta $\mu \geq 160$ MeV or $|p_\mu^z| \geq 90$ MeV, and a cut on the polar angle of the missing photon (as reconstructed from the observed muons momenta) $|\cos\theta_\gamma| > \cos(15°)$. Finally, the reconstructed track mass of the $\mu^+\mu^-$ system, as defined in Eq. (11) (see

[8] We have checked that with this smearing parameter we could reproduce with good accuracy the SM muon-track mass distribution given in [50].

e.g., [49]), must satisfy $m_{\rm tr} \in [80, 115]$ MeV. This last cut is by far the most stringent one because $m_{\rm tr}$ is very sensitive to the presence of missing energy. As an example, varying the smearing of σ_{p_T} by a factor of two leads to a variation of the NP cut efficiency by around 40%–50%, depending on the parameter point.

Note that for all the three KLOE analyses, the final cross section is obtained by including soft initial-state radiation according to the analytic expressions given in Ref. [79].

(c) *BABAR*. In the last analysis of the *BABAR* Collaboration [80] the following selection cuts were applied; polar angles of charged tracks in the laboratory frame are required to be in the range $\theta_{\mu^\pm} \in [0.35, 2.4]$ rad, and $\theta_\gamma \in [0.45, 2.45]$ rad for photon tracks. The photon energy in the c.m. frame must satisfy $E^*_\gamma > 3$ Gev, and the charged track momenta $|\vec{p}_{\mu^\pm}| > 1$ GeV. Additionally, a preselection cut requiring that the ISR photon lies within 0.3 rad of the missing momentum of the charged tracks in the laboratory frame was also applied.

APPENDIX B: NEW PHYSICS VERSUS EXPERIMENTAL DATASETS TENSIONS

The main limitation of the NP model we used so far is the fact that large changes in the luminosity determination cannot be obtain for KLOE08, KLOE10 and KLOE12 simultaneously.

Thus it is interesting to explore if this limitation can be circumvented by extending our model with on additional dark photon, V_2. This attempt should be understood as a practical way to assess the best results achievable with this class of models. In Fig. 5 we plot the experimental shifts for the nine experimental results used in this work, choosing $m_{V1} = 1.001$ GeV, $\varepsilon_1 = 0.012$ and $m_{V2} = 1.0175$ GeV, $\varepsilon_2 = 0.008$ for the first and second dark photons, respectively. The error bars reflect both the experimental error as well as an educated guess of the theoretical errors associated with additional NP effects not included in the analysis, that could for example impact the experimental calibration of the efficiencies and the background subtraction procedure. Even with this *ad hoc* setup, the tension in the dataset remains at the 2.6σ level. This suggests that it is unlikely that all the discrepancies among the different datasets (as, for example, between the CMD-3 and CMD-2 results) could be accounted for by some NP.

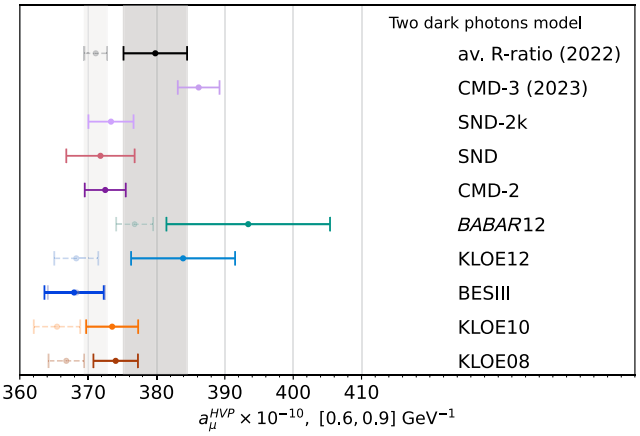

FIG. 5. Shifts in the prediction for the data-driven results for all the experimental datasets included in our analysis. The shifts are given in terms of $a_\mu^{\rm HVP}$ values in the $\sqrt{s} \in [0.60, 0.9]$ region, both for SM-only hypothesis (dashed lines) and for the best-fit model with two dark photons, with $m_{\chi_1} = 3$ MeV, $m_{\chi_2} = 0.97 m_V$, $m_{V1} = 1.001$ GeV and $m_{V2} = 1.0175$ GeV, $\alpha_D = 0.5$ and $\varepsilon_1 = 0.012$, $\varepsilon_2 = 0.008$ (solid lines). The gray band depicts the average of the e^+e^- data.

[1] T. Aoyama *et al.*, Phys. Rep. **887**, 1 (2020).
[2] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, Phys. Rev. Lett. **109**, 111808 (2012).
[3] T. Aoyama, T. Kinoshita, and M. Nio, Atoms **7**, 28 (2019).
[4] A. Czarnecki, W. J. Marciano, and A. Vainshtein, Phys. Rev. D **67**, 073006 (2003); **73**, 119901(E) (2006).
[5] C. Gnendiger, D. Stöckinger, and H. Stöckinger-Kim, Phys. Rev. D **88**, 053005 (2013).
[6] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang, Eur. Phys. J. C **77**, 827 (2017).
[7] A. Keshavarzi, D. Nomura, and T. Teubner, Phys. Rev. D **97**, 114025 (2018).
[8] G. Colangelo, M. Hoferichter, and P. Stoffer, J. High Energy Phys. 02 (2019) 006.
[9] M. Hoferichter, B.-L. Hoid, and B. Kubis, J. High Energy Phys. 08 (2019) 137.
[10] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang, Eur. Phys. J. C **80**, 241 (2020); **80**, 410(E) (2020).
[11] A. Keshavarzi, D. Nomura, and T. Teubner, Phys. Rev. D **101**, 014029 (2020).
[12] A. Kurz, T. Liu, P. Marquard, and M. Steinhauser, Phys. Lett. B **734**, 144 (2014).
[13] K. Melnikov and A. Vainshtein, Phys. Rev. D **70**, 113006 (2004).
[14] P. Masjuan and P. Sanchez-Puertas, Phys. Rev. D **95**, 054026 (2017).
[15] G. Colangelo, M. Hoferichter, M. Procura, and P. Stoffer, J. High Energy Phys. 04 (2017) 161.

[16] M. Hoferichter, B.-L. Hoid, B. Kubis, S. Leupold, and S. P. Schneider, J. High Energy Phys. 10 (2018) 141.
[17] A. Gérardin, H. B. Meyer, and A. Nyffeler, Phys. Rev. D **100**, 034520 (2019).
[18] J. Bijnens, N. Hermansson-Truedsson, and A. Rodríguez-Sánchez, Phys. Lett. B **798**, 134994 (2019).
[19] G. Colangelo, F. Hagelstein, M. Hoferichter, L. Laub, and P. Stoffer, J. High Energy Phys. 03 (2020) 101.
[20] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, and C. Lehner, Phys. Rev. Lett. **124**, 132002 (2020).
[21] G. Colangelo, M. Hoferichter, A. Nyffeler, M. Passera, and P. Stoffer, Phys. Lett. B **735**, 90 (2014).
[22] G. Bennett *et al.* (Muon $g-2$ Collaboration), Phys. Rev. D **73**, 072003 (2006).
[23] B. Abi *et al.* (Muon $g-2$ Collaboration), Phys. Rev. Lett. **126**, 141801 (2021).
[24] D. P. Aguillard *et al.* (Muon $g-2$ Collaboration), Phys. Rev. Lett. **131**, 161802 (2023).
[25] S. Borsanyi *et al.*, Nature (London) **593**, 51 (2021).
[26] T. Blum, P. A. Boyle, V. Gülpers, T. Izubuchi, L. Jin, C. Jung, A. Jüttner, C. Lehner, A. Portelli, and J. T. Tsang (RBC/UKQCD Collaborations), Phys. Rev. Lett. **121**, 022003 (2018).
[27] M. Cè *et al.* (ETMC Collaboration), Phys. Rev. D **106**, 114502 (2022).
[28] C. Alexandrou *et al.* (Extended Twisted Mass Collaboration), Phys. Rev. D **107**, 074506 (2023).
[29] C. Lehner and A. S. Meyer, Phys. Rev. D **101**, 074515 (2020).
[30] C. Aubin, T. Blum, M. Golterman, and S. Peris, Phys. Rev. D **106**, 054503 (2022).
[31] C. Aubin, T. Blum, C. Tu, M. Golterman, C. Jung, and S. Peris, Phys. Rev. D **101**, 014503 (2020).
[32] G. Wang, T. Draper, K.-F. Liu, and Y.-B. Yang (χQCD Collaboration), Phys. Rev. D **107**, 034513 (2023).
[33] A. Bazavov *et al.* (Fermilab Lattice, HPQCD, MILC Collaborations), Phys. Rev. D **107**, 114514 (2023).
[34] T. Blum *et al.*, Phys. Rev. D **108**, 054507 (2023).
[35] G. Colangelo, A. X. El-Khadra, M. Hoferichter, A. Keshavarzi, C. Lehner, P. Stoffer, and T. Teubner, Phys. Lett. B **833**, 137313 (2022).
[36] H. Wittig, Progress on $(g-2)_\mu$ from lattice QCD, in *Talk given at the 57th Rencontres de Moriond, Electroweak Interactions & Unified Theories, La Thuile, Italy* (2023).
[37] C. T. H. Davies *et al.* (Fermilab Lattice, MILC, HPQCD Collaborations), Phys. Rev. D **106**, 074509 (2022).
[38] L. Darmé, G. Grilli di Cortona, and E. Nardi, J. High Energy Phys. 06 (2022) 122.
[39] L. Di Luzio, A. Masiero, P. Paradisi, and M. Passera, Phys. Lett. B **829**, 137037 (2022).
[40] A. Crivellin and M. Hoferichter, Phys. Rev. D **108**, 013005 (2023).
[41] M. Passera, W. J. Marciano, and A. Sirlin, Phys. Rev. D **78**, 013009 (2008).
[42] A. Crivellin, M. Hoferichter, C. A. Manzari, and M. Montull, Phys. Rev. Lett. **125**, 091801 (2020).
[43] A. Keshavarzi, W. J. Marciano, M. Passera, and A. Sirlin, Phys. Rev. D **102**, 033002 (2020).
[44] B. Malaescu and M. Schott, Eur. Phys. J. C **81**, 46 (2021).
[45] F. V. Ignatov *et al.* (CMD-3 Collaboration), arXiv:2302.08834.
[46] J. P. Lees *et al.* (*BABAR* Collaboration), arXiv:2308.05233.
[47] A. Anastasi *et al.* (KLOE-2 Collaboration), J. High Energy Phys. 03 (2018) 173.
[48] F. Ambrosino *et al.* (KLOE Collaboration), Phys. Lett. B **670**, 285 (2009).
[49] F. Ambrosino *et al.* (KLOE Collaboration), Phys. Lett. B **700**, 102 (2011).
[50] D. Babusci *et al.* (KLOE Collaboration), Phys. Lett. B **720**, 336 (2013).
[51] M. Ablikim *et al.* (BESIII Collaboration), Phys. Lett. B **753**, 629 (2016); **812**, 135982(E) (2021).
[52] B. Aubert *et al.* (*BABAR* Collaboration), Phys. Rev. Lett. **103**, 231801 (2009).
[53] J. P. Lees *et al.* (*BABAR* Collaboration), Phys. Rev. D **86**, 032013 (2012).
[54] M. N. Achasov *et al.* (SND Collaboration), J. High Energy Phys. 01 (2021) 113.
[55] R. R. Akhmetshin *et al.* (CMD-2 Collaboration), Phys. Lett. B **578**, 285 (2004).
[56] V. M. Aul'chenko *et al.*, JETP Lett. **84**, 413 (2006).
[57] R. R. Akhmetshin *et al.* (CMD-2 Collaboration), Phys. Lett. B **648**, 28 (2007).
[58] M. N. Achasov *et al.*, J. Exp. Theor. Phys. **103**, 380 (2006).
[59] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, J. High Energy Phys. 07 (2014) 079.
[60] M. Della Morte, A. Francis, V. Gülpers, G. Herdoíza, G. von Hippel, H. Horch, B. Jäger, H. B. Meyer, A. Nyffeler, and H. Wittig, J. High Energy Phys. 10 (2017) 020.
[61] S. Jadach *et al.*, in *CERN Workshop on LEP2 Physics (Followed by 2nd Meeting, 15–16 June 1995 and 3rd Meeting 1995)* (1996), 10.5170/CERN-1996-001-V-2.229.
[62] A. B. Arbuzov, G. V. Fedotovich, F. V. Ignatov, E. A. Kuraev, and A. L. Sibidanov, Eur. Phys. J. C **46**, 689 (2006).
[63] G. Balossini, C. M. Carloni Calame, G. Montagna, O. Nicrosini, and F. Piccinini, Nucl. Phys. B, Proc. Suppl. **162**, 59 (2006).
[64] N. M. Coyle and C. E. M. Wagner, arXiv:2305.02354.
[65] E. Izaguirre, G. Krnjaic, and B. Shuve, Phys. Rev. D **93**, 063523 (2016).
[66] L. Darmé, S. Rao, and L. Roszkowski, J. High Energy Phys. 03 (2018) 084.
[67] A. Berlin and F. Kling, Phys. Rev. D **99**, 015021 (2019).
[68] J. L. Feng, B. Fornal, I. Galon, S. Gardner, J. Smolinsky, T. M. P. Tait, and P. Tanedo, Phys. Rev. D **95**, 035017 (2017).
[69] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, Phys. Rev. Lett. **113**, 091801 (2014).
[70] M. Bauer, P. Foldenauer, and J. Jaeckel, J. High Energy Phys. 07 (2018) 094.
[71] A. Hook, E. Izaguirre, and J. G. Wacker, Adv. High Energy Phys. **2011**, 859762 (2011).
[72] D. Curtin, R. Essig, S. Gori, and J. Shelton, J. High Energy Phys. 02 (2015) 157.
[73] M. N. Achasov *et al.*, Phys. Rev. D **63**, 072002 (2001).
[74] R. R. Akhmetshin *et al.* (CMD-2 Collaboration), Phys. Lett. B **669**, 217 (2008).

[75] G. Colangelo, M. Hoferichter, and P. Stoffer, Phys. Lett. B **814**, 136073 (2021).
[76] M. Davier, Z. Fodor, A. Gerardin, L. Lellouch, B. Malaescu, F. M. Stokes, K. K. Szabo, B. C. Toth, L. Varnhorst, and Z. Zhang, arXiv:2308.04221.
[77] A. M. Ruland, J. Phys. Conf. Ser. **160**, 012004 (2009).
[78] F. Ambrosino *et al.* (KLOE Collaboration), Eur. Phys. J. C **47**, 589 (2006).
[79] O. Nicrosini and L. Trentadue, Phys. Lett. B **196**, 551 (1987).
[80] J. P. Lees *et al.* (*BABAR* Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **726**, 203 (2013).