A Leptonic ALP Portal to the Dark Sector

Giovanni Armando,1,2 Paolo Panci,1,2 Joachim Weiss,3,4 and Robert Ziegler5,6
1 Dipartimento di Fisica E. Fermi, Università di Pisa, Largo B. Pontecorvo 3, I-56127 Pisa, Italy
2 INFN, Sezione di Pisa, Largo Bruno Pontecorvo 3, I-56127 Pisa, Italy
3 Max Planck Institute for Physics, Föhringer Ring 6, 80805 München, Germany
4 Technische Universität München, Physik-Department, James-Franck-Strasse 1, 85748 Garching, Germany
5 Institut für Theoretische Teilchenphysik, Karlsruhe Institute of Technology, Karlsruhe, Germany
6 Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, 79104 Freiburg, Germany

(Dated: October 10, 2023)

We discuss the leptonic ALP portal as a simple scenario that connects observed discrepancies in anomalous magnetic moments to the Dark Matter relic abundance. In this framework an axion-like particle in the multi-MeV range couples to SM leptons and a DM fermion, with mass above the ALP mass but below a GeV. The ALP contributes to $(g − 2)_{\mu}$ $(g − 2)_{\nu}$ dominantly through 2-loop Barr-Zee diagrams, while the DM abundance is generated by $p$-wave annihilation to ALP pairs. Constraints from beam-dump experiments, colliders and CMB probes are very stringent, and restrict the viable parameter space to a rather narrow region that will be tested in the near future.

I. INTRODUCTION

The origin of Dark Matter (DM) is arguably the most pressing problem of contemporary particle physics. Within generic theories beyond the Standard Model (BSM), the best motivated DM candidates are those that naturally arise in scenarios addressing other problems and shortcomings of the Standard Model (SM), for example the Strong CP Problem (as the QCD axion), or the hierarchy problem (as the neutralino). Another interesting class of models aims to generate the DM relic abundance within BSM scenarios that address experimental anomalies, i.e. explain observed deviations from SM predictions. Here we focus on the longstanding discrepancy in anomalous magnetic moments of leptons $(g − 2)_{\ell}$, which have been addressed in a variety of BSM scenarios, see [1] for a review.

The possible connection of DM and $(g − 2)_{\ell}$ has often been considered in the context of heavy new particles with masses $O(100\,\text{GeV})$, see e.g. Ref. [2]. In this article instead we analyze the scenario where only light new particles in the multi-MeV range are present: a pseudo-scalar particle coupling only to leptons and a new SM singlet fermion that accounts for DM. The lightness of the fermion is ensured by the chiral symmetry, while the scalar is light because it arises as a pseudo-Goldstone boson of an approximate Peccei-Quinn (PQ) symmetry, usually referred to as an axion-like particle (ALP). This ALP acts as a mediator between the SM leptons and the dark fermion, and gives rise to the observed DM relic abundance via thermal freeze-out.

Similar scenarios have been discussed in previous works [3–28], often in the context of Higgs-ALP mixing, i.e. the ALP that inherits all Higgs couplings to fermions, suppressed by a mixing angle. Here instead we only consider ALP couplings to leptons, all taken as independent parameters, and focus on a relatively low value of the ALP decay constant $f_a$. Our region of parameter space actually resembles the “visible QCD” axion proposed in Ref. [29, 30], which is a QCD axion with decay constant in the GeV range and couplings only to first generation fermions. Here instead the ALP couples to all three leptons, while couplings to quarks are completely absent. This renders this model phenomenologically viable, in contrast to the visible QCD axion of Ref. [29], which is largely (if not completely)1 excluded by NA62 searches [32] for $K \to \pi\alpha\alpha$, with all ALPs promptly decaying to electrons.

A very similar model to the one considered here has been employed in Ref. [33] to simultaneously explain $(g − 2)_{\mu}$ and an excess of electron events observed in the XENON1T experiment [34], induced by the scattering of an asymmetric DM fermion. This anomaly has now been refuted by additional data [35], along with new bounds recently derived for ALP-electron couplings from Kaon decays [36]. Here we take a somewhat smaller ALP-electron coupling in order to satisfy the latter constraints, and consider thermal freeze-out instead of an asymmetric DM scenario to generate the DM relic abundance.

The particle phenomenology in our scenario is controlled by the ALP and its couplings to leptons, since ALP decays to the dark sector are kinematically closed. As $m_a < m_{\mu}$, only ALP couplings to electrons and photons are relevant, the latter induced by loops of SM fermions. This allows to populate also regions of parameter space at lower ALP masses $(m_a \lesssim 30\,\text{MeV})$ deemed to be excluded in Ref. [37], which analyzed generic light ALP explanations of $(g − 2)_{\mu}$. Nevertheless we share the conclusion of these authors that it is challenging to build UV-complete models of these scenarios, because the PQ breaking scale is close to the GeV range, indicating the presence of other low-energy states neglected in our effective approach. However, as discussed in great details

1 The theoretical prediction might be subject to additional suppression factors, depending on different assumptions on the leading ChPT operator, see Ref. [31]
in Ref. [38], it is actually possible to construct viable UV completions for the “visible QCD” axion scenario, which is very similar to the effective model, except that the explicit ALP mass term is replaced by the QCD axion mass of similar numerical size. We expect that along the lines of Ref. [38] one can devise a viable UV extension of our scenario, although likely rather baroque. Here we refrain from presenting such a model, and focus on the analysis of collider and DM phenomenology in the effective framework, which only has a handful of parameters.

This paper is organized as follows. In Sec. II we define the basic setup of the model. The resulting particle phenomenology is analyzed in Sec. III, which includes axion decay rates, lepton anomalous magnetic moments, constraints from beam dump experiments and meson decays and a discussion of the necessary flavor alignment to satisfy LFV constraints. Sec. IV is devoted to the DM phenomenology, where we discuss the DM relic abundances, constraints from direct and indirect detection and bounds on DM self-interactions. We summarize our conclusions in Sec. V. Further details on direct detection are given in Appendix A.

II. SETUP

We consider a simplified model with a pseudo-Nambu-Goldstone (pNGB) boson \( a \), which only couples to the three SM leptons and a Dirac fermion \( \chi \) that will account for DM. The interaction Lagrangian is given by

\[
\mathcal{L} = -ig_\psi \bar{\psi} \gamma_5 \psi \gamma_5 a, \tag{1}
\]

with \( \psi = e, \mu, \tau, \chi \) and we take all couplings real. This Lagrangian is formally renormalizable, although it is an effective theory since we neglected the radial mode associated with the pNGB. Upon \( a \)-dependent fermion redefinitions (or using the fermion equations of motion with anomaly terms), we can equivalently describe the relevant interaction terms as

\[
\mathcal{L} = \frac{\partial_\mu a}{2f_a} \bar{\psi} \gamma_\mu \gamma_5 \psi + c_\gamma \frac{\alpha}{8\pi} \frac{a}{f_a} \epsilon^{\mu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}, \tag{2}
\]

with \( c^{0123} = -1 \). The couplings in Eqs. (1) and (2) are related by

\[
C_\psi \equiv \frac{c_\psi}{f_a} = \frac{g_\psi}{m_\psi}, \quad C_\gamma \equiv \frac{c_\gamma}{f_a} = \frac{g_e}{m_e} + \frac{g_\mu}{m_\mu} + \frac{g_\tau}{m_\tau}, \tag{3}
\]

for \( \psi = e, \mu, \tau, \chi \). In total there are 6 parameters, but we will fix \( g_\mu \) and \( g_\tau \) by reproducing the central value of \((g - 2)_{\ell}\), and \( g_\chi \) to reproduce the DM relic abundance, as discussed in the next sections. This leaves as free parameters \( m_\chi \) and \( g_e \) controlling the particle phenomenology, while \( m_a \) is only relevant for DM phenomenology. Besides experimental constraints, this parameter space is subject to the bounds from perturbative unitarity, which put upper bounds on the ALP couplings

\[
y_a = m_\psi c_\psi / f_a = m_\psi C_\psi < \sqrt{8\pi/3} \approx 2.9 \quad [39].
\]

Note that in contrast to Refs. [29, 38] we do not consider the possibility of a UV contribution to the effective ALP couplings to photons. Such contributions must come from charged fermions chiral under PQ, which can acquire a mass only around the PQ breaking scale, which is much below the electroweak scale in this scenario and thus likely in contrast with experimental constraints.

In the following we discuss the present particle physics constraints on the parameter space, afterwards we discuss the DM phenomenology.

III. PARTICLE PHENOMENOLOGY

A. ALP Decays

We will be interested in ALP masses below \( O(100) \text{ MeV} \), so that the ALP can only decay into electrons and photons\(^2\). The corresponding decay rates read

\[
\Gamma(a \to e^+ e^-) = \frac{m_a g^2_e}{8\pi} \sqrt{1 - \frac{4m_e^2}{m_a^2}},
\]

\[
\Gamma(a \to \gamma\gamma) = \frac{\alpha^2 m_a^2}{64\pi^3} |C_\gamma|^2, \quad \tag{4}
\]

where the effective photon coupling receives contributions from all fermions [40]

\[
C_\gamma = \sum_{\ell=e,\mu,\tau} \frac{g_\ell}{m_\ell} m_a^2 f^2 \left( \frac{4m_e^2}{m_a^2} \right). \tag{5}
\]

Here, the contribution from SM leptons is defined in terms of the loop function

\[
f(x) = \begin{cases} \arcsin \sqrt{x} & x \geq 1 \\ \frac{\pi}{2} + \frac{1}{2} \ln \frac{1 + \sqrt{x - 1}}{1 - \sqrt{x - 1}} & x < 1 \end{cases}, \tag{6}
\]

with the limit

\[
x f^2(x) = \begin{cases} \left(1 + \frac{1}{2x}\right)^2 & x \gg 1 \\ \frac{5}{4} \left(\pi + i \ln \frac{4}{x}\right)^2 & x \ll 1 \end{cases}. \tag{7}
\]

B. Lepton Anomalous Magnetic Moments

We determine the ALP couplings to heavy leptons \( g_\ell \) and \( g_\tau \) in order to reproduce the experimental values for the lepton anomalous magnetic moments \( \alpha_\ell = (g - 2)/2 \) for \( \ell = e, \mu \), which both deviate from the SM prediction. In the muon sector, the comparison of the (2021) experimental average [41] with the SM prediction of the Muon

\(^2\) As we will discuss in the next section, the ALP cannot decay to DM particles since we need \( m_\chi > m_a \), so that the relic abundance dominantly arises from p-wave annihilation \( \overline{\chi} \chi \to aa \).
g-2 Theory Initiative [42] has pointed to an intriguing 4.2σ discrepancy

\[ \Delta a_\mu^{\text{disp}}(2021) = a_\mu^{\text{EXP}} - a_\mu^{\text{SM}} = 251(59) \times 10^{-11}. \quad (8) \]

However, recent lattice results by the BMW collaboration for the hadronic vacuum polarization (HVP) contribution are in conflict with dispersive approaches based on low-energy \(e^+e^- \to \) hadrons data, and rather suggest the value (using the 2021 experimental result)

\[ \Delta a_\mu^{\text{lat}}(2021) = 107(70) \times 10^{-11}, \quad (9) \]

decreasing the discrepancy to 1.5σ. Other recent lattice calculations are consistent with the BMW result [43–47], and also a new measurement of the cross section \(e^+e^- \to \pi^+\pi^-\) between 0.32 and 1.2 GeV indicates an increase of the HVP contribution in the same direction as lattice data [48].

Given the unclear present situation, here we choose to follow an approach that mediates between the dispersive and the lattice methods, restricting the use of lattice data to the region least prone to systematic uncertainties (the so-called “window” observable [44]) and using low-energy data otherwise [49]. With the recent 2023 update of the experiment world average [50] this approach gives the value [51]

\[ \Delta a_\mu^{\text{wind}(2023)} = 181(47) \times 10^{-11}, \quad (10) \]

which corresponds to a 3.8σ deviation. For a summary of the current status of the \((g-2)_\mu\) SM prediction see Refs. [52, 53].

In the electron sector instead hadronic contributions are largely irrelevant for the SM prediction, and its uncertainty is mainly driven by the input for the fine-structure constant \(\alpha\) (see e.g. Ref. [54] for details). Unfortunately, there are two conflicting experimental determinations of \(\alpha\), obtained from spectroscopy of either Cs [55] or Rb-atoms [56]. Using the latest (2022) experimental value for \(a_\mu^{\text{EXP}}\) obtained by the Harvard group presented in Ref. [57], we obtain for the discrepancies for the Rb method

\[ \Delta a_\ell^{\text{Rb}} = 34(16) \times 10^{-14}, \quad (11) \]

corresponding to a 2.1σ deviation. Instead the Cs-method gives a lower SM prediction, resulting in

\[ \Delta a_\ell^{\text{Cs}} = -102(26) \times 10^{-14}, \quad (12) \]

which corresponds to a 3.9σ deviation from the SM prediction. Note that the significance has increased with respect to the older experimental value from 2008, giving deviations of 1.6σ (Rb) and 2.4σ (Cs). Until the conflicting experimental determinations of \(\alpha\) have been clarified, in the following we are simply taking the latter (Rb) result in Eq. (11) at face value.

The Lagrangian in Eq. (1) gives a contribution to the anomalous magnetic moment \(\Delta a_\ell\) of the lepton \(\ell = e, \mu, \tau\) at 1-loop [39, 40, 58], corresponding to the diagram on the left in Fig. 1

\[ \Delta a_\ell^{\text{1-loop}} = -\frac{g_\ell^2}{16\pi^2} h_1\left(\frac{m_\ell^2}{m_f^2}\right), \quad (13) \]

where \(h_1(u)\) is a positive-definite loop function given by

\[ h_1(u) = \int_0^1 dy \frac{2y^3}{u - uy + y^2}. \quad (14) \]

Also important are 2-loop Barr-Zee diagrams shown on the right in Fig. 1, which give the following contributions to \(\Delta a_\ell\) [33]

\[ \Delta a_\ell^{\text{2-loop}} = \frac{\alpha m_f}{8\pi^3 m_f} N_f^2 Q_f^2 g_\ell g_f F\left(\frac{m_\ell^2}{m_f^2}, \frac{m_\ell^2}{m_f^2}\right), \quad (15) \]

where \(f\) is a fermion with mass \(m_f\), color multiplicity \(N_f\), electromagnetic charge \(Q_f\) and ALP coupling \(g_f\) in Eq. (1), while \(F(u, v)\) is the loop function

\[ F(u, v) = \int_0^1 dx \int_0^1 dy \int_0^1 dz \frac{ux}{u^2 + uvx + vz^2} \left(\log\left(1 + \frac{uvx}{z}\right) - \log\left(1 + \frac{u}{z}\right) + \log\left(1 + \frac{v}{z}\right)\right), \quad (16) \]

with the shorthand \(\tau = 1 - x\), and similar for \(y, z\).

In the limit when the external lepton mass \(m_\ell\) is small compared to the ALP mass \(m_a\), i.e. \(u \gg 1\), we obtain \(F(u, v) \to -\int_0^1 dy \log(vz) / (1 - vz)\) and thus recover the result in Eq. (10) of Ref. [58] (and Eq. (58) in Ref. [54]). In the limit of large fermion masses propagating in the loop, \(v \ll 1, v \ll u\), one can treat the loop as a point-like interaction of the ALP with two photons. In this case, our full result should reproduce the leading logarithm obtained from a 1-loop calculation within an effective theory, where the heavy fermion has been integrated out. In the limit \(v \ll u \ll 1\) we obtain \(F(u, v) \to 2 - \log v\), while \(v \ll u \ll 1\) gives \(F(u, v) \to 3 - \log v/u\), which indeed matches the logarithmic dependence in Eq. (37) of Ref. [40], upon identifying the renormalization scale \(\mu\) with the heavy fermion mass \(m_f\).

This discussion also makes clear that in the case \(v \ll 1 \ll u\) the two-loop function is unsuppressed, in contrast to the one-loop function in Eq. (13), which in this limit becomes \(h_1(u) \to (-11/3 + 2\log u)/u\). Therefore the two-loop contribution in Eq. (15) can potentially dominate over the one-loop contribution in Eq. (13), whenever \(m_\ell \ll m_\ell \ll m_f\), even when \(g_f \sim g_\ell\).

Finally, the total ALP contribution \(\Delta a_\ell = \Delta a_\ell^{\text{1-loop}} + \Delta a_\ell^{\text{2-loop}}\) is compared to the difference of the experimental value and the SM expectation in Eq. (10) and (11). We then fix the value of \(c_\mu\) and \(c_\tau\) such\(^3\) to reproduce the

\(^3\)There are always two solutions for \(c_\mu\), but one of them involves a large cancellation between 1-loop and 2-loop contributions. We choose the solution that does not involve such a tuning, so that both central values for \(\Delta a_\ell\) are reproduced dominantly by the 2-loop ALP contribution.
central values, if possible at all for given values of \( m_a \) and \( g_e \). The resulting constraints in the \((m_a, g_e)\) parameter space are shown in Fig. 2 as excluded regions in light blue.

C. Beam Dump and Collider Constraints

Stringent constraints on light particles arise from electron beam-dump experiments that have searched for \( e^+e^- \) decays\(^4\) of short-lived particles produced from an electron beam stopped in an absorbing target. Relevant for our scenario are only a handful of experiments. Important bounds on the parameter space are provided by the NA64 collaboration, which originally searched for a massive vector particle [59], and has re-analyzed their results for the case of a pseudo-scalar in Ref. [60].

This analysis supersedes previous recasts in Refs. [29, 33] (that were based on a simple coupling rescaling following Ref. [61]), and gives slightly weaker bounds on ALP couplings to electrons. Also beam dump experiments carried out at SLAC (E141) [62], KEK [63] and Orsay [64] provide relevant constraints on the parameter space, besides limits from the SLAC beam dump E137 [65], which were however shown explicitly only for the lower range of couplings in the original article. For the present scenario instead only the upper range of couplings excluded by E137 is relevant, which we simply take from Fig. 2 of Ref. [36].

Finally also colliders have put important bounds on the parameter space. While searches carried out at KLOE [66] do not provide competitive constraints, the BaBar collaboration has searched for Dark Photons coupled to electrons, and the resulting limits presented in Ref. [67] can be easily recast for ALPs. All relevant constraints from beam dumps and colliders are shown as gray regions in Fig. 2.

It should be noticed that we have taken into account only experiments that look for ALPs produced off electrons. It is however clear that the ALP also couples to photons at 1-loop, and thus constraints from photo-production should also be taken into account. Nevertheless, we expect such constraints to be mild, as the loop contributions to photo-production are suppressed by the relevant energy scale, which is typically large. This is in contrast to models with effective ALP couplings to photons from loops of heavy fermions, which are severely constrained (see e.g. Ref. [40]).

D. Constraints from Pion Decays

Important constraints on leptonic ALP couplings also arise from meson decays with ALPs radiated off final-state leptons. The strongest bounds on electron couplings can be obtained from \( \pi^+ \to e^+\nu_a, a \to e^+e^- \) searches at the SINDRUM experiment [70], as recently discussed in Ref. [36].

This analysis rules out ALP couplings to electrons larger than \( 10^{-4} \) for ALP masses below \( \sim 50 \text{ MeV} \) and \( \text{BR}(a \to e^+e^-) \approx 1 \), and provides the most relevant upper bound on \( g_e \) in the present scenario, see Fig. 2, where the parameter region excluded by the SINDRUM search is shown in gray.

E. Flavor Alignment

A major shortcoming of this scenario is the large amount of required flavor alignment, in order to satisfy stringent constraints from lepton flavor violation. Specifically, a possible LFV coupling of the ALP to muons and electrons, \( \mathcal{L} \supset C_{\mu e} \partial \bar{\mu} \gamma^\mu \gamma^5 e + \text{h.c.} \), gives rise to LFV constraints. It is however clear that the ALP also couples to photons at 1-loop, and thus constraints from photo-production should also be taken into account. Nevertheless, we expect such constraints to be mild, as the loop contributions to photo-production are suppressed by the relevant energy scale, which is typically large. This is in contrast to models with effective ALP couplings to photons from loops of heavy fermions, which are severely constrained (see e.g. Ref. [40]). For example, in the relevant mass range of \((10 - 100) \text{ MeV} \) strong constraints on effective ALP couplings to photons arise from LEP [68], but we have checked that the resulting constraints on lepton couplings are weaker than constraints from perturbative unitarity.

4 We actually assume that the relevant experiments are equally sensitive to ALPs decaying to electrons or photons.
FIG. 2. Parameter space in the \((m_a, g_e)\) plane, where ALP couplings to muons and tau leptons are adjusted to reproduce the central values of \(\Delta a_{\mu}\) and \(\Delta a_{\tau}\). Gray regions show exclusion limits from beam dump and collider experiments, blue regions denote constraints from anomalous magnetic moments, and red regions are excluded by perturbativity constraints on \(g_e\). The four benchmark models defined in Table I are indicated as black diamonds.

We summarize the particle physics phenomenology in Fig. 2, which shows the relevant experimental constraints in the \((m_a, g_e)\) plane, as discussed in this section. It is clear that \((g - 2)_{\ell}\) alone would allow two separated strips of available parameter space, but the upper bound is entirely excluded by SINDRUM searches for \(\pi^+ \rightarrow e^+\nu a, a \rightarrow e^+e^-\) in the low mass regime, and by BaBar searches in the high mass range. This leaves a rather narrow strip with couplings \(10^{-6} \lesssim g_e \lesssim 10^{-5}\) and \(10\text{MeV} \lesssim m_a\), smaller masses being excluded by KEK and E137 searches. In the available parameter space we select four benchmark models (BMs), which we take as representatives for the entire region. They are defined in Table I and will be used to discuss the Dark Matter phenomenology in the next section.

### Table I. Definition of benchmark models.

<table>
<thead>
<tr>
<th>Model</th>
<th>(m_a) [MeV]</th>
<th>(g_e/10^{-7})</th>
<th>(g_\mu/10^{-4})</th>
<th>(g_\tau)</th>
<th>(\text{BR}(\gamma\gamma)[%])</th>
<th>(\tau_{\chi}[\text{ps}])</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>20</td>
<td>0.20</td>
<td>0.71</td>
<td>1.8</td>
<td>99</td>
<td>2.8</td>
</tr>
<tr>
<td>b</td>
<td>15</td>
<td>0.70</td>
<td>3.6</td>
<td>0.50</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>c</td>
<td>95</td>
<td>0.20</td>
<td>0.59</td>
<td>2.5</td>
<td>100</td>
<td>0.01</td>
</tr>
<tr>
<td>d</td>
<td>95</td>
<td>1.0</td>
<td>3.9</td>
<td>0.51</td>
<td>84</td>
<td>0.28</td>
</tr>
</tbody>
</table>

We now delve into the connection of \((g - 2)_{\ell}\) and the DM relic abundance in this scenario. Specifically, among the possible ALP couplings to SM leptons allowed by the constraints of Fig. 2, we focus on the four benchmark models as representative examples defined in Table I. The remaining free parameters of the model are the DM mass \(m_\chi\) and the DM-ALP coupling \(g_\chi\), as defined in Eq. (1). We determine them by requiring that the observed DM abundance is attained after thermal freeze-out of DM fermions from the SM plasma. Subsequently, we assess whether the model is compatible with the current constraints derived from various DM searches.

### IV. DARK MATTER PHENOMENOLOGY

For DM in the mass range \(1\text{MeV} \lesssim m_\chi \lesssim 10\text{GeV}\), the relevant DM annihilations in the early universe occur through three primary channels: i) tree-level \(\bar{\chi}\chi \rightarrow \ell^+\ell^-\) leptonic channels; ii) tree-level \(\bar{\chi}\chi \rightarrow aa\) ALP channel; and iii) the loop-induced \(\bar{\chi}\chi \rightarrow \gamma\gamma\) di-photon channel, see Fig. 3. The corresponding velocity averaged cross...
sections read:
\[
\langle \sigma v \rangle_{\ell \ell} = \frac{g^2 a^2}{2\pi} \frac{m_X^2}{(4m_X^2 - m_a^2)^2} \sqrt{1 - \frac{m_t^2}{m_X^2}},
\]
\[
\langle \sigma v \rangle_{\gamma\gamma} = \frac{g^2 a^2}{4\pi^3} \frac{m_X^4}{(4m_X^2 - m_a^2)^2} |\tilde{C}_{\gamma\gamma}|^2,
\]
\[
\langle \sigma v \rangle_{aa} = \frac{6}{x} \frac{g^4}{2\pi^3} \frac{m_X^2 (m_a^2 - m_X^2)^2}{(2m_X^2 - m_a^2)^4} \sqrt{1 - \frac{m_s^2}{m_X^2}},
\]
where \( x = m_X/T \) and \( T \) is the temperature of the thermal bath. The first two cross-sections are s-wave, while the third one is p-wave. In Eq. (20), \( C_{\gamma\gamma}^{\text{eff}} \) is analogous to the effective photon coupling \( C_{\gamma\gamma}^{\text{eff}} \) for the ALP decay defined in Eq. (5), but with the replacement \( m_s^2 \rightarrow s \simeq 4m_X^2 \). Following Ref. [73], we solve the Boltzmann equation for thermal freeze-out including both s- and p-wave contributions to the total velocity averaged cross section. Imposing that \( \Omega_{\text{DM}} h^2 \simeq 0.12 \) [74] then determines a line in the \((m_X, g_X)\) plane for a given benchmark model.

### B. Indirect Detection

Indirect detection (ID) searches for DM aim to find hints for DM by observing anomalies or distinctive features in cosmic ray fluxes measured on earth. These searches typically focus on regions where the DM density is expected to be substantial, such as the centers of galaxies or galaxy clusters. Alternatively, ID searches also target regions with relatively low astrophysical backgrounds, for example dwarf spheroidal galaxies. Moreover, the Cosmic Microwave Background (CMB) serves as a powerful tool for indirect detection: DM annihilations release energy into the surrounding plasma, causing ionization and heating of the medium [75, 76]. This injection of energy influences the evolution of the universe during recombination, leaving imprints on the anisotropies of the CMB and resulting in distortions of the CMB black body spectrum [77, 78].

We focus on two robust indirect detection probes: i) CMB anisotropies, which are highly effective in constraining sub-GeV DM annihilations into electrons and di-photons, relevant for our scenario. For a given annihilation channel we take these bounds from Ref. [78]; and ii) a compilation of X-rays and \( \gamma \)-ray data from several experiments (see e.g. Refs. [79, 80]). In particular, in the DM mass range spanning from hundreds of MeV to tens of GeV, the bounds coming from DM searches towards dwarf spheroidal galaxies [81, 82] surpass the CMB constraints only if the annihilation channel into heavy leptons (\( \mu, \tau \)) is accessible. It is important to note that above 180 MeV the constraints from the XMM-Newton satellite, provided in Ref. [80], are the strongest available. Nevertheless, these constraints are subject to considerable uncertainties, as shown in Fig. 9 of [80], and therefore, we opt not to include them in our analysis.

In the aforementioned probes the DM is highly non-relativistic. Consequently, the p-wave DM annihilations into ALPs are always negligible when compared to the s-wave annihilations into pairs of leptons and di-photons. Hence, to compute the limits from indirect detection we collect the relevant bounds and weigh them by the ratios of the respective s-wave channels. Specifically, we first calculate the total velocity averaged s-wave cross section \( \langle \sigma v \rangle_{\text{tot}}^{\text{th}} \) and the ratios \( R_i = \langle \sigma v \rangle_{ii} / \langle \sigma v \rangle_{\text{tot}}^{\text{th}} \) with \( i = \gamma, e, \mu, \tau \). Then by denoting \( \langle \sigma v \rangle_{\gamma\gamma}^{\text{CMB}} \) and \( \langle \sigma v \rangle_{\gamma\gamma}^{\text{Fermi}} \) as the best experimental limits in a given channel, we infer a limit in the \((m_X, g_X)\) plane by demanding that
\[
\langle \sigma v \rangle_{\text{tot}}^{\text{th}} < \sum_{i=\gamma, e, \mu} R_i \langle \sigma v \rangle_{\gamma\gamma}^{\text{CMB}} + R_\tau \langle \sigma v \rangle_{\gamma\gamma}^{\text{Fermi}}.
\]

### C. Direct Detection

One of the notable advantages of the present scenario is that constraints from direct detection (DD) searches for DM are easily satisfied. Model-independent processes originating from either single or double ALP exchange between the DM fermion and SM leptons give rise to loop-induced scatterings of DM off nuclei, which are highly suppressed and therefore remain entirely undetectable within current and upcoming DD experiments.

The only plausible process that could result in detectable collisions with SM quarks is induced by the model-dependent trilinear vertex \( A_{aah} a^2 h \), which couples the ALP to the SM Higgs. While not included in our setup defined in Eq. (1), such a coupling is likely generated within UV-complete models. As discussed in the introduction, a possible UV completion of our effective framework is provided in Ref. [38], where an ALP (actually a proper QCD axion) with mass in the MeV range and sizable couplings to SM fermions emerges. In this particular setup, the dimensionful coupling \( A_{aah} \) is of the order of hundreds of MeV. According to the results in Ref. [83], the vertex \( A_{aah} a^2 h \) then leads to a sizable spin-independent DM-nucleon cross section, which is already excluded by the latest results from the LUX-ZEPLIN (LZ) experiment [84], if the DM mass exceeds roughly tens of GeV (see Appendix A for more details). For this reason we restrict the parameter space to \( m_X \leq 10 \text{ GeV} \), although it might be possible to construct UV completions, where the trilinear coupling \( A_{aah} \) is much smaller.

### D. Other Complementary Bounds

In addition to these purely observational constraints, we also consider two other complementary bounds on the ALP-DM coupling from: i) perturbative unitarity which gives \( g_X < \sqrt{8\pi/3} \) [39]; and ii) DM self-interactions which gives \( g_X \lesssim 0.21 (m_X/\text{MeV})^{3/4} \) [27]. This limit arises from the fact that collision processes like \( \chi\chi \rightarrow \chi\chi \),
\( \bar{\chi} \chi \rightarrow \bar{\chi} \chi \) and \( \chi \chi \rightarrow \chi \chi \) can in principle transport heat from the hotter outer region to the colder inner region of the DM halo, leading to a thermalization of the latter if the energy transfer cross section per unit mass \( \sigma_T / m_\chi \) is larger than roughly 1 \( \text{cm}^2/\text{g} \)\cite{85, 86}.

**E. Summary**

In Fig 4, we present a summary of the results in the \((m_\chi, g_\chi)\) plane. The left column displays benchmark models \( a \) and \( c \), while the right column shows benchmarks \( b \) and \( d \). In all panels, the \( s \) - and \( p \)-wave contributions to the relic density are depicted as thin red and blue lines, respectively. The thick black line corresponds to the portion of the parameter space where the observed DM relic abundance is thermally produced. The blue shaded regions represent the bounds derived from DM indirect detection. Additionally, we use dark gray shading to indicate areas of parameter space that are excluded by the requirement of perturbative unitarity. Regions excluded by limits obtained from DM self-interactions are highlighted with dark magenta shading.

In the mass range \( m_e < m_\chi < m_a \), both \( p \)-wave annihilations into ALPs and \( s \)-wave annihilations into heavy leptons are kinematically closed. Since in all benchmarks \( C_\tau \) is the largest coupling among the leptonic couplings \( C_\ell \), it dominates the effective photon coupling in Eq. (20), \( |C_\tau^{\text{eff}}|^2 \simeq C_\tau = g_\chi \sigma T / m_\tau \). Hence, the ratio of the thermally averaged cross sections in Eq. (19) and Eq. (20) is given by \( \langle \sigma v \rangle_{\ell \ell} / \langle \sigma v \rangle_{\gamma \gamma} \simeq 2 \pi^2 / a_{\text{em}}^2 g_\chi^2 / g_e^2 m_e^2 / m_\chi^2 \). Using the reference values for the four benchmark models in Table I, it is evident that for BMs \( a \) and \( c \), the dominant contribution to the relic density arises from loop-induced annihilations into photon pairs. Conversely, for benchmarks \( b \) and \( d \), tree-level annihilations into pairs of electrons are the primary channel (if the axion is not too heavy, \( m_a \lesssim 60 \text{MeV} \), otherwise again annihilation into photons dominates). In this specific mass range and for all four benchmark models, the CMB limits probe \( s \)-wave cross section into either di-photons or electron pairs that are considerably smaller than the thermal cross-section. Thus, for \( m_\chi < m_a \), the CMB limits completely rule out the freeze-out predictions depicted in Fig. 4.

In the mass range \( m_a < m_\chi < m_\tau \), the relic abundance is dominantly generated by \( p \)-wave annihilation into ALPs. This holds for all models and is particularly important, because for sub-GeV thermal DM with \( p \)-wave annihilations indirect detection constraints are less restrictive. Specifically, for BMs \( a \) and \( c \) the ratio between the \( p \)-wave cross section at freeze-out \( \langle \sigma v \rangle_{aa} / x_F \simeq 1 / x_F g_e^4 / (128 \pi m_a^2) \) with \( x_F \sim 25 \) and the \( s \)-wave cross section into di-photons is \( \langle \sigma v \rangle_{aa} / \langle \sigma v \rangle_{\gamma \gamma} \simeq \pi^2 / (2 x_F a_{\text{em}}^2) g_e^2 / g_\chi^2 m_e^2 / m_\chi^2 \). For the reference values in Table I, it is clear that the cross section into di-photons is approximately one order of magnitude smaller than the \( p \)-wave cross-section at freeze-out for DM for \( m_\chi \sim 1 \text{GeV} \). It is worth noticing that in this specific mass range, \( \langle \sigma v \rangle_{\gamma \gamma} \) remains independent of \( m_\chi \). For DM masses in the GeV range the CMB limits are particularly stringent. They rule out cross sections that are around 30 times smaller than the thermal value, and this exclusion becomes more pronounced as the DM mass decreases. Consequently, the freeze-out predictions for BMs \( a \) and \( c \), depicted by the solid black lines in the left column of Fig. 4, are robustly excluded. For BMs \( b \) and \( d \), when \( m_\chi < m_\mu \) the primary \( s \)-wave channel is still annihilation into electron pairs, while for \( m_\chi > m_\mu \) the muon channel becomes increasingly relevant. This transition results in an enhanced \( s \)-wave contribution to the relic density, as depicted by the blue lines in the right column of Fig. 4. In this specific mass range \( \langle \sigma v \rangle_{aa} / x_F \simeq 1 / (4 x_F) g_e^2 / g_\chi^2 \). Hence, for the reference values in Table I, the \( s \)-wave channel into electrons is significantly smaller than the thermal value, at least by a factor of \( 10^4 \). Conversely, the \( s \)-wave channel into muons is also notably smaller, but by a factor of at least 50 when compared to the thermal value. The indirect detection limits for leptonic channels are not yet sensitive enough to probe cross sections of such small size. Consequently, the thermal freeze-out predictions for the benchmark models in the right column of Fig. 4 remain viable and are not excluded by these limits.

Finally, for \( m_\chi > m_\tau \), the \( s \)-wave annihilation into tau leptons is open. Since \( g_\tau \) in all benchmarks is close to unity, this process emerges as the dominant mechanism for generating the DM abundance in the multi-GeV mass range and beyond. In this specific mass range the indirect detection limits from DM searches that are directed towards dwarf spheroidal galaxies are extraordinarily stringent. These limits are so severe that they completely rule out the possibility of thermal DM with \( s \)-wave annihilations into tau lepton pairs up to DM masses of hundreds of GeV.
V. SUMMARY AND CONCLUSIONS

We have discussed a simple scenario to connect the observed discrepancies in anomalous magnetic moments with the Dark Matter relic abundance. In this framework an axion-like particle in the multi-MeV range couples solely to SM leptons and a DM fermion. In the lepton sector, the couplings to muons and tau leptons are fixed by explaining the discrepancies in \((g-2)_\mu\), leaving as free parameters only the ALP mass and its electron coupling. Imposing also present constraints from beam-dumps and collider experiments, the viable parameter space consists of a narrow strip with \(10^{-6} \lesssim g_e \lesssim 10^{-5}\) (and \(10\text{ MeV} \lesssim m_a\)), see Fig. 2. We have identified four benchmark models that represent this parameter space, and contribute to \((g-2)_\mu\) and \((g-2)_e\) dominantly through 2-loop Barr-Zee diagrams.

The remaining parameter space in the dark sector, i.e. the mass of the DM fermion and its coupling to the ALP is determined by requiring that the observed DM relic abundance is reproduced through thermal freeze-out. This determines a line in this 2d parameter region...
for a given benchmark model, which is subject to constraints from direct and indirect detection, see Fig. 4. We find that CMB constraints completely rule out s-wave annihilation channels into electrons, muons or photons, while s-wave annihilation into tau leptons is excluded by direct detection constraints. This leaves as the only viable possibility p-wave annihilation into ALPs, which essentially fixes the DM mass range \( m_a < m_\chi < m_\tau \), along with the DM-ALP coupling of order \( \sim 10^{-2} \). Still, even in this region ALP couplings to photons have to be sufficiently suppressed in order to satisfy CMB constraints, which disfavors large ALP-tau couplings, and in turn small electron couplings, so only the upper part of the phenomenologically allowed strip in the parameter space is compatible with DM phenomenology. This leaves only a relatively narrow region that will be tested by current and future experiments, such as XMM-Newton searching for X-rays generated via inverse Compton processes, the next generation of CMB probes and possibly dedicated searches at beam dump experiments.

ACKNOWLEDGMENTS

The work of GA and PP is supported by the PRIN grant 202289JEW4. JW is part of the International Max Planck Research School (IMPRS) on “Elementary Particle Physics”. We thank Giorgio Arcadi, Jeff Dror, Francesco d’Eramo, Diego Redigolo and Stefan Vogl for useful discussions.

Appendix A: Details of Direct Detection Constraints

In terms of the model parameters of Ref. [83], the trilinear vertex \( a h h \) has an effective coupling \( A_{aah} = 2m_a^2 \theta^2 / v \) in the limit \( m_A \gg m_a \). A rough estimate of the DM-nucleon spin-independent cross section induced by this interaction, expressed in terms of \( A_{aah} \), is

\[
\sigma_{SI} \simeq 5.2 \times 10^{-43} \text{ cm}^2 \times \left( \frac{A_{aah}}{1 \text{ GeV}} \right)^2 \left( \frac{100 \text{ MeV}}{\langle q_{\text{ref}} \rangle} \right)^4 \left( \frac{g_\chi}{0.5} \right)^4 \left( \frac{m_{\chi}}{30 \text{ GeV}} \right)^2 ,
\]

where \( q_{\text{ref}} = 2\mu_{Xe} v \simeq (1 / 100) \text{ MeV} \simeq m_a \) is the typical momentum exchanged in DM collisions off xenon nuclei and we have used the \( \langle q \rangle \) matrix element \( \langle N | \sum_q m_q \bar{q} q | N \rangle = 105.9 \text{ MeV} \) from the FLAG average of the lattice computations in the case of \( N_f = 2+1+1 \) [87]. An estimate of the trilinear coupling can be obtained from Eqs. (43-44) of Ref. [38] which yields \( A_{aah} \simeq 0.28 \) GeV. With this input we can calculate the limits in the \( (m_{\chi}, g_\chi) \) plane by comparing Eq. (A1) with the LZ bound of the DM spin-independent cross section [84]. This limit is roughly the same for all benchmark models and we choose to illustrate it specifically for model b in Fig. 5. As it is apparent, the LZ limit rules out the possibility of thermal DM for \( m_{\chi} \) larger than tens of GeV.


CMD-3 Collaboration, F. V. Ignatov et al., “Measurement of the $e^+e^-\to\pi^+\pi^-$ cross section from threshold to 1.2 GeV with the CMD-3 detector,” arXiv:2302.08834 [hep-ex].


