HAL
open science

Axion-like particles as mediators for dark matter: beyond freeze-out

A. Bharucha, F. Brümmer, N. Desai, S. Mutzel

To cite this version:

A. Bharucha, F. Brümmer, N. Desai, S. Mutzel. Axion-like particles as mediators for dark matter: beyond freeze-out. Journal of High Energy Physics, 2023, 02, pp.141. 10.1007/JHEP02(2023)141 . hal-03790726

HAL Id: hal-03790726

https://hal.science/hal-03790726v1

Submitted on 15 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prepared for submission to JHEP

Axion-like particles as mediators for dark matter: beyond freeze-out

A. Bharucha[a] **F. Brümmer,**[b] **N. Desai**[c] **and S. Mutzel**[a]

[a] *Aix Marseille Univ, Université de Toulon, CNRS, CPT, IPhU, Marseille, France*
[b] *LUPM, UMR5299, Université de Montpellier, 34095 Montpellier, France*
[c] *Department of Theoretical Physics, Tata Institute of Fundamental Research Mumbai, India*

E-mail: aoife.bharucha@cpt.univ-mrs.fr, felix.bruemmer@umontpellier.fr, desai@theory.tifr.res.in, sophie.mutzel@cpt.univ-mrs.fr

Abstract: We consider an axion-like particle (ALP) coupled to Standard Model (SM) fermions as a mediator between the SM and a fermionic dark matter (DM) particle. We explore the case where the ALP-SM and/or the ALP-DM couplings are too small to allow for DM generation via standard freeze-out. DM is therefore thermally decoupled from the visible sector and must be generated through either freeze-in or decoupled freeze-out (DFO). In the DFO regime, we present an improved approach to obtain the relic density by solving a set of three stiff coupled Boltzmann equations, one of which describes the energy transfer from the SM to the dark sector. Having determined the region of parameter space where the correct relic density is obtained, we revisit experimental constraints from electron beam dump experiments, rare B and K decays, exotic Higgs decays at the LHC, astrophysics, dark matter searches and cosmology. In particular, for our specific ALP scenario we (re)calculate and improve beam dump, flavour and supernova constraints. Throughout our calculation we implement state-of-the-art chiral perturbation theory results for the ALP partial decay width to hadrons. We find that while the DFO region, which predicts extremely small ALP-fermion couplings, can probably only be constrained by cosmological observables, the freeze-in region covers a wide area of parameter space that may be accessible to other more direct probes. Some of this parameter space is already excluded, but a significant part should be accessible to future collider experiments.

arXiv:2209.03932v1 [hep-ph] 8 Sep 2022

Contents

1 Introduction **1**

2 Model **3**
- 2.1 Particle content, interactions and relevant processes 3
- 2.2 General Boltzmann equations 5
- 2.3 Freeze-in 7
  - 2.3.1 Freeze-in from SM particles 7
  - 2.3.2 Freeze-in from ALPs 8
  - 2.3.3 Sequential freeze-in 9
- 2.4 Decoupled freeze-out (DFO) 10
- 2.5 Finite-temperature effects 12

3 Constraints **13**
- 3.1 Collider constraints 13
  - 3.1.1 SLAC E137 electron beam dump experiment 13
  - 3.1.2 Exotic Higgs decays at the LHC 16
  - 3.1.3 Production in decays of mesons 17
- 3.2 Astrophysical constraints 18
  - 3.2.1 Supernova SN1987A 18
  - 3.2.2 Horizontal branch stars 19
- 3.3 Cosmological constraints 21
- 3.4 Constraints on the DM 21
  - 3.4.1 DM relic density constraint 21
  - 3.4.2 (No) constraints from dark matter phenomenology 22

4 Results **23**
- 4.1 Collider and astrophysical constraints 24
- 4.2 Phase diagram 26
- 4.3 Freeze-in region 28
- 4.4 DFO region and cosmological constraints 31

5 Conclusion **35**

A Effective ALP-DM Lagrangian from a two Higgs doublet model **38**

B Collision term for the energy transfer Boltzmann equation **41**

C Cross sections and input for Boltzmann equations **42**
- C.1 Hidden sector process 44
- C.2 Connector processes 44
- C.3 SM-ALP processes 44

D Decay width of the ALP **45**

E Axion production cross sections in SLAC E137 experiment **47**
E.1 Primakoff production 47
E.2 Bremsstrahlung 48
E.3 Positron annihilation 48

F Details on the ALP's detection probability **50**

G Details on constraints from exotic Higgs decays **51**

H Loop diagrams relevant for constraint calculations **52**

---

1 Introduction

The last two decades have seen a surge in the theoretical exploration of dark matter (DM) models, and brought a wealth of new data to constrain them. The possibility that dark matter directly couples to Standard Model (SM) states via renormalizable interactions is by now severely constrained. For example, LHC data and direct detection experiments rule out vast parts of the parameter space for standard electroweak-scale WIMPs produced via thermal freeze-out [1, 2]. This has led to a paradigm shift towards various other possibilities for dark matter candidates, with masses above 1 TeV or below 10 GeV, and with other production mechanisms. In particular, models where the DM-SM interactions are mediated by additional particles have attracted much attention.

Among such models are those with a fermionic dark matter candidate and a pseudoscalar mediator a, interacting with the SM via a Lagrangian of the form

$$\mathcal{L} = a \sum_f C_f \frac{m_f}{f_a} \bar{f} i\gamma_5 f + a\, C_\chi \frac{m_\chi}{f_a} \bar{\chi} i\gamma_5 \chi + \dots \tag{1.1}$$

Here f stands for any SM fermion, and χ is the DM particle. One possible motivation for this type of interaction is that it allows the efficient suppression of direct detection cross sections while still allowing for large enough DM-SM interactions to create the observed relic abundance via thermal freeze-out. However, freeze-out is of course by no means the only mechanism for dark matter production. In fact, one of the oldest dark matter candidates is itself a pseudoscalar state which is extremely light and extremely weakly coupled, namely a non-thermally produced QCD axion. At intermediate values of the couplings, dark matter could be produced by the freeze-in mechanism: Assuming that its abundance is initially zero after reheating, dark matter will be gradually produced by scattering processes in the thermal plasma without ever reaching thermal equilibrium, until these processes decouple and the DM abundance remains constant [3, 4].

In this paper we study a dark matter model with interactions of the type of eq. (1.1) but for a range of couplings which are too small to give the correct relic density through

freeze-out. Our motivation is unrelated to direct detection constraints (which are irrelevant for such small couplings) but are rather guided by top-down reasoning: If there exists a light pseudoscalar a, it might well be an axion-like particle (ALP), by which we mean the pseudo-Goldstone boson of an approximate $\mathrm{U}(1)_{\mathrm{PQ}}$ global symmetry which is spontaneously broken at a high scale f_a; this provides a reason for it to be light. It is natural to have a emerge from an extended Higgs sector in some UV completion of the SM, which could explain the flavour-preserving couplings to SM states of eq. (1.1). In this case, besides these effectively renormalizable couplings, one also expects dimension-5 couplings of the form

$$\mathcal{L} = a \sum_f C_f \frac{y_f}{\sqrt{2} f_a} h \, \bar{f} i \gamma_5 f + \dots \tag{1.2}$$

where h is the SM Higgs boson, which will turn out to be important for the dark matter abundance in parts of the parameter space.

If there are additional fermions χ charged under $\mathrm{U}(1)_{\mathrm{PQ}}$, they would also couple to the ALP. If there are no additional fermions charged under both $\mathrm{U}(1)_{\mathrm{PQ}}$ and under the Standard Model, there exists a field basis in which no dimension-5 couplings $aF\widetilde{F}$ between the ALP and the SM gauge bosons are induced above the electroweak scale. The only relevant couplings between the SM and the ALP are then those of eqs. (1.1) and (1.2). Moreover, since there is no evidence for new physics close to the electroweak scale, the scale f_a should be large, $f_a \gg$ TeV. Hence the ALP couplings to fermions may be too small to allow for dark matter production via freeze-out, and other production mechanisms should be studied.

It is tempting to try to identify a in this scenario with a (DFSZ-like) QCD axion itself. However, if a QCD axion were massive enough to be short-lived on cosmological scales (thus mediating the SM-DM interactions, rather than being part of the dark matter sector itself), it would be subject to multiple astrophysical and laboratory constraints, see e.g. [5] for an overview. While it may be possible to circumvent these constraints by extensive model-building, in this paper we will not attempt to do so but rather assume that the a mass is mainly due to some explicitly $\mathrm{U}(1)_{\mathrm{PQ}}$-breaking effect other than the anomaly. The usual QCD axion relation between m_a and f_a therefore no longer holds, and a does not contribute to solving the strong CP problem.

To summarise our findings, we will show that the observed dark matter abundance can be produced in this model by several distinct mechanisms, depending on the values of the ALP-SM and ALP-DM couplings as well as on the ALP and DM masses. It can be produced via ALP-mediated freeze-in from the scattering of SM particles, via freeze-in from the scattering of ALPs, via freeze-out of ALP-DM interactions which kept the DM in equilibrium with the ALP, or finally via standard freeze-out. We will numerically study a number of example scenarios with regard to the most relevant cosmological, astrophysical and collider constraints.

2 Model

2.1 Particle content, interactions and relevant processes

Our model contains a single Dirac fermion χ and a pseudoscalar ALP a. The Lagrangian is

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \frac{1}{2}\partial_\mu a \partial^\mu a + \bar{\chi}(i\partial\!\!\!/ - m_\chi)\chi + i\sum_f \frac{C_f}{f_a}\left(m_f + \frac{y_f}{\sqrt{2}}h\right) a\bar{f}\gamma_5 f$$
$$- \frac{1}{2}m_a^2 a^2 + i\frac{C_\chi}{f_a}\, m_\chi\, a\bar{\chi}\gamma_5\chi\,, \tag{2.1}$$

where f is any Standard Model fermion with mass m_f and Yukawa coupling y_f and h is the Higgs boson. This Lagrangian is valid below the electroweak symmetry breaking scale; see Appendix A for some details on a possible UV completion. We have not included any coupling of the ALP to gauge bosons via $a\ \mathrm{tr}\, F\widetilde{F}$ terms. Including them would be perfectly possible but would lead to a proliferation of parameters, so we restrict our analysis to more minimal models where there are no heavy fermions charged under both the SM gauge group and $\mathrm{U}(1)_{\rm PQ}$. Nevertheless, effective $a\gamma\gamma$ and agg vertices are induced at one loop by (finite) SM fermion triangle graphs. We will often refer to the hidden sector (HS), by which we mean the ensemble of ALPs and DM particles. For future convenience, we define the effective axion couplings by

$$g_{aff} = \frac{C_f}{f_a}\,, \qquad g_{a\chi\chi} = \frac{C_\chi}{f_a}\,, \tag{2.2}$$

where we will later refer to $g_{a\chi\chi}$ as the hidden sector coupling, and to g_{aff} as the connector coupling. In this model, dark matter is stabilized by a global U(1) χ-number symmetry which also ensures that it can only be produced in $\chi\bar{\chi}$ pairs. A variant with a Majorana dark matter candidate stabilized by a $\mathbb{Z}_2$ symmetry could also be viable.

From the above Lagrangian we can deduce the various means by which:

- the dark matter can interact with Standard Model particles, e.g. through interactions with fermions mediated by the pseudoscalar, $f\bar{f} \leftrightarrow \chi\bar{\chi}$ or $f\bar{f} \leftrightarrow h\chi\bar{\chi}$, or via a loop diagram with gauge bosons, $\gamma\gamma \leftrightarrow \chi\bar{\chi}$;

- the ALP can interact with the SM, e.g. via $q\bar{q} \leftrightarrow ga$, $qg \leftrightarrow qa$, or the (inverse) decay $f\bar{f} \leftrightarrow a$, or $f\bar{f} \leftrightarrow ha$;

- the ALPs and the dark matter can interact, i.e. $aa \leftrightarrow \chi\bar{\chi}$ or $\chi a \leftrightarrow \chi a$.

These different interactions, where those responsible for DM and ALP generation are summarised in table 1, give rise to a rich phenomenology. Different DM generation mechanisms are at play in different regions of the parameter space. The study of these mechanisms was pioneered in a different model by [6], and more recently refined in [7].

As detailed in the next sections, the mechanisms at play in our model are the following:

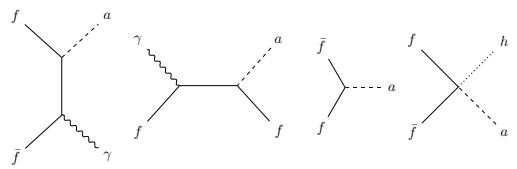
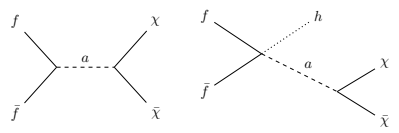
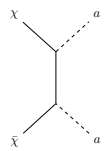
Interaction	Processes	Scaling
SM $\leftrightarrow$ ALPs		g_{aff}^2
SM $\leftrightarrow \chi$		$(g_{aff} \cdot g_{a\chi\chi})^2$
ALPs $\leftrightarrow \chi$		$g_{a\chi\chi}^4$

Table 1: Different sector interaction processes and their scaling with the hidden sector and connector couplings. Note that the diagrams $f\bar{f} \leftrightarrow \gamma a$ and $f\gamma \leftrightarrow fa$ shown are equivalent to those for $qq \leftrightarrow ga$, $qg \leftrightarrow qa$, $\bar{f}\gamma \leftrightarrow \bar{f}a$ and $\bar{q}g \leftrightarrow \bar{q}a$ and therefore the latter are not shown explicitly.

- For sufficiently small $g_{a\chi\chi}$ and g_{aff}, and assuming that the HS relic density is negligible at reheating, dark matter is produced by freeze-in. Depending on the couplings, the dominant process may be direct freeze-in $f\bar{f} \to \chi\bar{\chi}$ mediated by an off-shell ALP, or ALP production from SM states followed by $aa \to \chi\bar{\chi}$ (where the ALP may or may not be in equilibrium with the SM). These processes are infrared dominated, since they are induced by (effectively) renormalizable operators. For large reheating temperatures, the dominant process for direct freeze-in will be $2 \to 3$ scattering $f\bar{f} \to h\chi\bar{\chi}$, which is suppressed by phase space but ultraviolet-dominated, i.e. sensitive to the reheating temperature.

- For intermediate g_{aff}, ALPs will be produced abundantly from the SM thermal bath but will not reach thermal equilibrium with the SM particles. However, if $g_{a\chi\chi}$ is sufficiently large, the hidden sector will form a separate thermal bath at a lower temperature, in which DM is produced by freeze-out. We study this mechanism of freeze-out from a thermally decoupled dark sector, or to put it simply, decoupled freeze-out (DFO), in detail.

- For larger g_{aff} and $g_{a\chi\chi}$, the entire dark sector will thermalize with the SM, and DM can be produced via standard freeze-out. Since we are interested in the small-coupling regime, and since the standard freeze-out phase is already well studied in the literature, we will not analyse it in detail.

2.2 General Boltzmann equations

In order to study the various DM generation mechanisms we start from the covariant form of the Boltzmann equation for the Friedmann-Lemaître-Robertson-Walker metric, where the momenta of the particles have been integrated over. The SM particles are in equilibrium with the photon bath at temperature T, allowing us to replace the distributions of SM particles by the equilibrium distributions. We neglect quantum statistical factors, assuming Maxwell-Boltzmann statistics:

$$f_{\rm eq}(p,T) = e^{-E/T}, \tag{2.3}$$

such that the particles' number density, energy density and pressure equilibrium distributions are given by

$$\begin{aligned}
n_{\rm eq}(T) =& \frac{g}{(2\pi)^3}\int d^3p\, f_{\rm eq}(p,T) = \frac{g}{2\pi^2}\, m^2\, T K_2\left(m/T\right)\,,\\
\rho_{\rm eq}(T) =& \frac{g}{(2\pi)^3}\int d^3p\, f_{\rm eq}(p,T) E = \frac{g}{2\pi^2}\, m^2\, T\left(m K_1\left(m/T\right) + 3T K_2\left(m/T\right)\right)\,,\\
P_{\rm eq}(T) =& \frac{g}{(2\pi)^3}\int d^3p\; f_{\rm eq}(p,T)\frac{|p|^2}{3E} = \frac{g}{2\pi^2} m^2 T^2 K_2\left(m/T\right) = T n_{\rm eq}(T)\,,
\end{aligned} \tag{2.4}$$

respectively, for a given particle of mass m and degrees of freedom g, where $K_1(x)$ and $K_2(x)$ are the first and second order Bessel functions.

More generally, the hidden sector particles may be in kinetic equilibrium at a temperature T' which is different from T:

$$f(p,T') = \frac{n(T')}{n_{\rm eq}(T')} e^{-E/T'}\,. \tag{2.5}$$

Kinetic equilibrium is not reached in the freeze-in phase, where the scattering among hidden sector particles is not efficient (see [8] for a detailed analysis of freeze-in in a related model).

The Boltzmann equations governing the evolution of the a and χ number densities are given by

$$\begin{aligned}
\frac{{\rm d}n_\chi}{{\rm d}t} + 3Hn_\chi =& \sum_f \left\langle \sigma_{\chi\bar\chi\to f\bar f} v\right\rangle \left(\left(n_\chi^{\rm eq}\right)^2 - n_\chi^2\right) + \left\langle \sigma_{aa\to\chi\bar\chi} v\right\rangle n_a^2 - \left\langle \sigma_{\chi\bar\chi\to aa} v\right\rangle n_\chi^2\\
&+ \sum_{i,j,k}\left(\left\langle \sigma_{\chi\bar\chi\to ijk} v\right\rangle \left(\left(n_\chi^{\rm eq}\right)^2 - n_\chi^2\right) + \left\langle \sigma_{\chi\bar\chi i\to jk} v^2\right\rangle n_i^{\rm eq}\left(\left(n_\chi^{\rm eq}\right)^2 - n_\chi^2\right)\right)\,,\\
\frac{{\rm d}n_a}{{\rm d}t} + 3Hn_a =& \sum_{i,j,k}\left\langle \sigma_{ia\to jk} v\right\rangle \left(n_a^{\rm eq} n_i^{\rm eq} - n_a n_i^{\rm eq}\right) + \left\langle \Gamma_a\right\rangle \left(n_a^{\rm eq} - n_a\right)\\
&- \left\langle \sigma_{aa\to\chi\bar\chi} v\right\rangle n_a^2 + \left\langle \sigma_{\chi\bar\chi\to aa} v\right\rangle n_\chi^2\,.
\end{aligned} \tag{2.6}$$

Here i,j,k are SM particles which are involved in the relevant processes, as listed in table 1. For example, if all fermion couplings C_f are of the same order and at temperatures where all SM particles are relativistic, the dominant contribution for ALP production comes from $i = g,\ j = \bar k = t$. The cross sections for the processes entering eq. (2.6) are given in

appendix C. Most of the processes appearing in eq. (2.6) are $2 \to 2$ processes, and enter via the typical thermally averaged cross section $\langle \sigma v \rangle$, defined by[1]

$$\langle \sigma_{12\to 34} v \rangle = \frac{C}{2\,T K_2(m_1/T)\,K_2(m_2/T)} \int_{s_{\rm min}}^{\infty} \sigma(s) \frac{F(m_1,m_2,s)^2}{m_1^2 m_2^2 \sqrt{s}}\, K_1(\sqrt{s}/T)\, {\rm d}s \ , \tag{2.7}$$

where $\sigma(s)$ is the cross section for the process $12 \to 34$ as a function of the squared centre-of-mass energy s. Here C is an additional factor of $1/2$ for the case of identical particles in the initial state ($C = 1$ for non-identical initial state particles). We have further used the abbreviation

$$F(m_1,m_2,s) = \frac{\sqrt{(s-(m_1+m_2)^2)(s-(m_1-m_2)^2)}}{2} \tag{2.8}$$

and the lower limit of the integral is given by $s_{\rm min} = \max\left((m_1+m_2)^2, (m_3+m_4)^2\right)$. Note that, where appropriate, we have made use of detailed balance,

$$\langle \sigma_{ij\to kl} v \rangle\, n_i^{\rm eq} n_j^{\rm eq} = \langle \sigma_{kl\to ij} v \rangle\, n_k^{\rm eq} n_l^{\rm eq} \ . \tag{2.9}$$

For the (inverse) decays which contribute, we define the thermally averaged decay rate to be

$$\langle \Gamma_a \rangle = \Gamma_a \frac{K_1(m_a/T)}{K_2(m_a/T)} \ . \tag{2.10}$$

where Γ_a is the decay rate of the pseudoscalar a. We provide details of the ALP decay rate in the various decay channels in appendix D. For the calculation of the DM relic density we restrict ourselves to (inverse) decay into fermions in the perturbative regime, see eq. (D.1).[2]

The Hubble rate in eq. (2.6) is given in terms of the total energy density of the universe $\rho = \rho_{\rm SM} + \rho_{\rm HS}$, via

$$H = \left(\frac{8}{3}\pi G \rho\right)^{1/2} \ , \tag{2.11}$$

where G is the gravitational constant. The energy density of the visible sector is

$$\rho_{\rm SM} = g_{*\rho,{\rm SM}}(T) \frac{\pi^2}{30} T^4 \ , \tag{2.12}$$

with $g_{*\rho,{\rm SM}}(T)$ the SM effective degrees of freedom in energy. The hidden sector energy density differs according to the production regime as we will explain in more detail in section 2.4. For the SM effective degrees of freedom in energy and in entropy, denoted by $g_{*\rho}$ and g_{*s}, respectively, we adopt the recent results from an improved analysis, covering a wide range of temperatures [9]. This analysis is based on state-of-the-art results of perturbative and

[1] We have not explicitly stated the temperature dependence of the thermally averaged cross sections in eq. (2.6) as this depends on the production regime considered. For example, in the DFO mechanism the hidden sector interactions take place at the common temperature of the hidden sector which is different from that of the SM, see below in section 2.4.

[2] For energies in the range $\sim 0.5 - 1.2$ GeV various hadronic decay channels open up. However, when we present results on the DM relic density, we are either in the perturbative regime or apply a hard cut on the decay width below $T \leq 600$ MeV (see also the discussion in section 4.4 and in appendix D).

non-perturbative calculations of thermodynamic quantities in the SM. In particular, for an improved description of the QCD phase transition they adopt results from the "Budapest-Wuppertal"-collaboration [10] in $2+1+1$ flavour lattice QCD. We use the fit functions provided in this paper in our analysis.

Note that by solving this set of differential equations we only calculate the abundance of χ particles and later set $n_{\rm DM} = n_\chi + n_{\bar{\chi}} = 2n_\chi$. In addition, we point out that later in this section we will find that it is convenient to make a change of variables from t to $z = m_\chi/T$, and the differential equations will be written as a function of z instead of t.

2.3 Freeze-in

The freeze-in mechanism can generate the observed DM abundance for the case that both g_{aff} and $g_{a\chi\chi}$ are sufficiently small, assuming that the initial abundance of DM particles and ALPs is zero or negligible. Both a and χ will be produced from interactions amongst SM particles in the thermal bath, but for small g_{aff} and $g_{a\chi\chi}$ the DM will not thermalize, neither with the photons nor with the ALPs. The production stops once the temperature of the particles involved drops below the DM mass scale, since the abundances become Boltzmann-suppressed and the production rate becomes negligible compared to the Hubble expansion rate. The co-moving dark matter number density is effectively frozen from that point on, while the ALPs will eventually decay back to Standard Model states.

The interplay of the different sectors allows for three different freeze-in regimes, namely freeze-in of the DM directly from SM particles, freeze-in from the ALPs and sequential freeze-in [7, 8]. All three regimes are characterised by a value of $g_{a\chi\chi}$ which is too small for the hidden sector particles to establish thermal equilibrium among each other. Depending on the range in g_{aff}, the ALPs may or may not be in equilibrium with the SM, as described in the following.

2.3.1 Freeze-in from SM particles

In this regime, the DM is produced directly from the SM fermions via an s-channel ALP (see table 1). In order to generate the correct relic density via this mechanism, one requires $g_{a\chi\chi}$ to be small, such that $\langle\sigma_{aa\to\chi\bar{\chi}}v\rangle \propto g_{a\chi\chi}^4$ is small enough for $aa \to \chi\bar{\chi}$ to be negligible compared to $f\bar{f} \to (h)\chi\bar{\chi}$. In this regime we find that ALPs are in thermal equilibrium with the SM, i.e. $n_a = n_a^{\rm eq}(T)$.

In order to calculate the relic abundance we require the appropriate Boltzmann equation for this region, which we obtain by adapting eq. (2.6) accordingly. We first define the dark matter yield $Y = Y_\chi$ as usual by

$$Y = \frac{n_\chi}{s} \tag{2.13}$$

where $s = s_{\rm SM} + s_{\rm HS}$ is the total entropy density of the universe. The entropy density in the freeze-in regime is given by

$$s = \frac{2\pi^2}{45} g_{*s}(T) T^3 \ , \tag{2.14}$$

with the effective degrees of freedom in entropy given by $g_{*s}(T) = g_{*s,\rm SM}(T) + g_{*s,\rm a}(T)$ and we can neglect the entropy density of the DM particles.

The ALP abundance plays no role in this freeze-in scenario, so we neglect all terms in eq. (2.6) involving n_a. Since the DM is never in equilibrium, $Y \ll Y_{\rm eq}$, we can neglect the back-reaction as usual. Concerning the contribution to the DM yield from $2 \to 2$ scattering, it is convenient to introduce the reaction density for the interaction $ij \to kl$ [6]

$$\gamma_{ij\to kl} = \langle \sigma_{ij\to kl}\, v\rangle n_i^{\rm eq} n_j^{\rm eq} \tag{2.15}$$

in terms of which the Boltzmann equation eq. (2.6) for the yield from freeze-in becomes

$$\frac{{\rm d}Y}{{\rm d}T} = -\sum_f \frac{\gamma_{\chi\bar\chi\to f\bar f}}{3Hs^2}\frac{{\rm d}s}{{\rm d}T} + (2 \leftrightarrow 3 \text{ terms})\,. \tag{2.16}$$

Here we have replaced the z dependence by a dependence on T. The corresponding infrared-dominated contribution to the yield today at $T \approx 0$ is therefore obtained by integrating the reaction density,

$$Y_{0,\rm IR} = -\sum_f \int_0^\infty \frac{\gamma_{\chi\bar\chi\to f\bar f}}{3Hs^2}\frac{{\rm d}s}{{\rm d}T}\,{\rm d}T\,. \tag{2.17}$$

An important point concerns the relative magnitude of the contributions from $2 \to 2$ interactions and $2 \to 3$ interactions. In a hypothetical model with only renormalizable couplings between the ALP and the SM fermions, as defined by eq. (1.1), freeze-in would proceed only via $f\bar f \to \chi\bar\chi$ scattering, which is infrared dominated. However, as we stated previously, in realistic models where the ALP-fermion couplings are generated from the ALP mixing with the Standard Model Higgs, one also obtains the dimension-5 couplings of eq. (1.2). These couplings induce $2 \to 3$ interactions which are phase-space suppressed, but nevertheless cannot be neglected in general,[3] because they are ultraviolet dominated and their contributions therefore scale as the reheating temperature $T_{\rm RH}$. For instance, if $T_{\rm RH}$ is large enough to neglect all masses of the involved particles, then the dark matter yield from such a process can be estimated as [11]

$$Y_{0,\rm UV} \approx \frac{135}{(2\pi)^9}\frac{1}{1.66\, g_*^{3/2}} g_{aff}^2 g_{a\chi\chi}^2\, y_f^2\, m_\chi^2\, T_{\rm RH}\, M_P\,. \tag{2.18}$$

A thorough analysis of ALP production from UV freeze-in, allowing for all possible dimension-5 couplings between the ALP and the SM to be present, was conducted in [12], and models similar to ours have been studied in both their UV and IR freeze-in phases in [13]. We will eventually find that, for reheating temperatures $T_{\rm RH} \gtrsim 200$ GeV, UV freeze-in via $2 \to 3$ scattering gives significant contributions to the dark matter yield; see appendix C for details of the computation.

2.3.2 Freeze-in from ALPs

In this regime, the ALPs are in thermal equilibrium with the SM, and produce the DM via the t-channel process $aa \to \chi\bar\chi$. Obtaining the observed relic density via this mechanism requires a larger $g_{a\chi\chi}$ compared to the case in section 2.3.1, and a correspondingly smaller g_{aff} to avoid overproduction.

[3] Note on the other hand that the $f\bar f h \to \chi\bar\chi$ interactions included in eq. (2.6) can safely be neglected.

Since the ALPs are in thermal equilibrium with the SM particles, this regime is conceptually very similar to the previous one. We just need to replace the reaction density $\gamma_{\chi\bar{\chi}\to f\bar{f}}$ in eq. (2.17) by the appropriate $\gamma_{\chi\bar{\chi}\to aa}$:

$$Y_0 = -\int_0^\infty \frac{\gamma_{\chi\bar{\chi}\to aa}}{3Hs^2}\frac{\mathrm{d}s}{\mathrm{d}T}\,\mathrm{d}T. \tag{2.19}$$

In this regime, there are no direct sizeable UV-dominated contributions to the DM yield. However, UV-dominated $2\to 2$ processes such as $f\bar{f}\to ah$ will contribute to the ALP abundance and will therefore reduce the value of g_{aff} at which the ALPs reach equilibrium with the SM. We determine this boundary by numerically solving the Boltzmann equation for the ALPs and checking whether $n_a \gtrsim n_a^{\mathrm{eq}}$. For IR-dominated contributions, the yield is essentially insensitive to the upper integration limit (provided it is chosen above the freeze-out temperature of the top quark). Since the UV-dominated contributions depend on the reheating temperature, the appropriate integration limit is at $z_{RH} = m_\chi/T_{RH}$. As always we assume that the universe has reheated into SM particles only, i.e. $n_a(T_{RH}) \approx 0$. Note that $\mathcal{O}(1)$ corrections to the UV contributions can be expected, due to the uncertainty associated with the details of the reheating mechanism.

2.3.3 Sequential freeze-in

In the sequential freeze-in regime, dark matter is again produced from ALPs via $aa\to\chi\bar{\chi}$. The difference with the previous regime is that the ALPs are not thermalized. Nevertheless, the ALP abundance is sufficient to obtain the correct DM relic density.

Calculating the DM relic abundance is non-trivial, since the ALPs are not even in kinetic equilibrium, and therefore the unintegrated Boltzmann equations governing the individual momentum modes should be employed for a precise quantitative analysis. This was done in a study dedicated to the freeze-in regime in ref. [8] for a related model. In contrast, in ref. [7] it is proposed that one may perform a simplified analysis employing the integrated Boltzmann equations

$$\begin{aligned}\frac{\mathrm{d}n_a}{\mathrm{d}t} + 3Hn_a &= \sum_{i,j,k}\gamma^{\mathrm{eq}}_{ia\to jk} + \langle\Gamma_a\rangle\, n_a^{\mathrm{eq}}(T)\,,\\ \frac{\mathrm{d}n_\chi}{\mathrm{d}t} + 3Hn_\chi &= \gamma_{\chi\bar{\chi}\to aa}\,,\end{aligned} \tag{2.20}$$

where $\gamma_{aa\to\chi\bar{\chi}} = \langle\sigma_{aa\to\chi\bar{\chi}}v\rangle(T)\,n_a^2$, and n_a is determined from the first equation. The second term on the RHS of the first equation describes the production of the ALP via inverse decays. We have neglected the term $-\gamma_{\chi\bar{\chi}\to aa}$ in the first equation since $n_a^{\mathrm{eq}}(T) \gg n_a$. The temperature of the ALPs is taken to be the temperature T of the SM bath.

This gives a rough estimate of the parameter values leading to the observed relic density; however we stress that this method is relying on several rather crude approximations. Specifically, we assume that the axions are in kinetic equilibrium in order to be able to define a temperature, and we then equate this temperature with that of the photons. It should be possible to obtain a more precise result following the method of ref. [8], but this is beyond the scope of our work.

2.4 Decoupled freeze-out (DFO)

For a sufficiently large hidden sector coupling $g_{a\chi\chi}$, the interactions between the HS particles will be strong enough for them to thermalize. If the ALP-SM couplings g_{aff} are sufficiently small, then the hidden sector will nevertheless be thermally decoupled from the SM thermal bath, at a temperature $T' \ll T$. This regime is somewhat more complicated to analyse than the freeze-in regime, due to an increased number of interactions playing a role, and due to the non-trivial interplay between the hidden sector number densities and the temperature T'. We will therefore proceed to explain some of the technical aspects in more detail.

As in the case of freeze-in, we assume the initial number density of DM particles and of ALPs to be negligibly small. The DM and ALPs will then freeze in until the abundances are sufficient to allow the particles to thermalize. This happens once the reaction rate for $aa \leftrightarrow \chi\bar{\chi}$ scattering exceeds the expansion rate of the universe,

$$\Gamma_{aa\to\chi\bar{\chi}} \equiv \langle \sigma_{aa\to\chi\bar{\chi}}\, v\rangle n_a^{\rm eq}(T') > H\,. \tag{2.21}$$

While $aa \leftrightarrow \chi\bar{\chi}$ is the dominant process for DM and ALP annihilation and production, the DM and ALP number densities will approximately track the equilibrium values $n_\chi^{\rm eq}(T')$ and $n_a^{\rm eq}(T')$, respectively. After T' falls below the mass of the χ and a, we allow for the possibility that n_χ and n_a deviate from the equilibrium distributions. The number densities n_χ and n_a and the hidden-sector temperature T' are then given by the solution of a system of three coupled differential equations, as explained in the following.

The temperature T' of the hidden sector is calculated from the hidden-sector energy density ρ' via the equation of state. To determine ρ', we start from the Boltzmann equation for the phase-space distribution $f(p,t)$ of either of the hidden sector particle species:

$$(\partial_t - H\,p\,\partial_p)\,f(p,t) = \frac{1}{E(p)}C[f(p,t)]\,. \tag{2.22}$$

Here we have used that, by isotropy and homogeneity, $f(p,t)$ can only depend on the modulus p of its 3-momentum and on time. $C[f]$ is the collision operator and E is the energy. With the energy density given by eq. (2.4), we can integrate eq. (2.22), writing

$$\frac{\partial\rho'}{\partial t} + 3H\,\left(\rho' + P'\right) = \int \frac{d^3p}{(2\pi)^3}C[f(p,t)] \tag{2.23}$$

where we have integrated by parts and used $P' = \frac{1}{3}\langle p\frac{\partial E}{\partial p}\rangle$. The integrated collision operator for a 1 2 $\to$ 3 4 process is

$$\int \frac{d^3p}{(2\pi)^3}C[f] = g_1 g_2 \int \frac{d^3p_1}{(2\pi)^3}\frac{d^3p_2}{(2\pi)^3} f_1(p_1) f_2(p_2) v_{\rm Møl}\,\mathcal{E}(\vec{p}_1,\vec{p}_2) \tag{2.24}$$

where $v_{\rm Møl}$ is the usual Møller velocity and the energy transfer rate $\mathcal{E}$ is, in terms of the matrix element $\mathcal{M}$ and the transferred energy ΔE_{tr},

$$\mathcal{E}(\vec{p}_1,\vec{p}_2) = \frac{1}{2E_1 2E_2 v_{\rm Møl}}\int \prod_{i=3,4}\frac{d^3p_i}{(2\pi)^3}\frac{1}{2E_i}|\mathcal{M}|^2(2\pi)^4\delta^{(4)}(p_1+p_2-p_3-p_4)\Delta E_{tr}\,. \tag{2.25}$$

Solving eq. (2.23) enables us to obtain ρ' as a function of the temperature of the visible sector T. In ref. [6], eq. (2.23) was applied to a model with fermionic dark matter coupled to both the SM photon and a hidden-sector photon. In that case, the hidden sector is populated predominantly via $f\bar{f} \to \chi\bar{\chi}$ processes, and the mass degeneracies of the particles involved allow to analytically simplify the integrated collision operator along the lines of [14], leaving a single integral to be evaluated numerically. In our model, the $2 \to 2$ processes $f\bar{f} \to a\gamma$, $\gamma f \to af$, $q\bar{q} \to ag$, $gq \to aq$ and hermitian conjugates must also be taken into account, as well as (inverse) decay processes $f\bar{f} \to a$. To treat the $2 \to 2$ processes, some work is needed to cast the integrated collision operator into a form amenable to numerical integration. This is detailed and justified in Appendix B.

Given the hidden-sector energy density from eq. (2.23), we obtain T' from the equation of state of the hidden sector. The total HS energy density and pressure are given by the sum of the ALP and DM contributions,

$$\rho' + P' = \rho_a + \rho_\chi + P_a + P_\chi \,. \tag{2.26}$$

Initially the ALPs and the DM will be ultra-relativistic, $P' = \rho'/3$, and the universe will be radiation-dominated with most of its energy density in the visible sector, $\rho \propto T^4$. Changing variables using $\frac{\partial}{\partial t} \approx -HT\frac{\partial}{\partial T}$, eq. (2.23) becomes

$$\frac{\partial \rho'}{\partial t} + 4H\,\rho' = -H\left(T\frac{\partial \rho'}{\partial T} - 4\rho'\right) = -HT\rho\frac{\partial}{\partial T}\left(\frac{\rho'}{\rho}\right) = \int \frac{d^3p}{(2\pi)^3} C[f(p,t)]\,. \tag{2.27}$$

As long as the hidden sector particles are relativistic and interactions are rapid, solving this equation for ρ' will provide us with the temperature of the hidden sector via eqs. (2.4): T' is obtained by iteratively solving

$$\frac{\rho'}{\rho} = \frac{\rho_a^{\rm eq}(T') + \rho_\chi^{\rm eq}(T')}{\frac{\pi^2}{30} g_{\rm eff,SM}(T) T^4} \,. \tag{2.28}$$

As T' decreases, the hidden-sector particles will eventually become non-relativistic, and their interactions will freeze out. In this regime, T' must be determined together with the HS number densities n_χ and n_a. Kinetic equilibrium is maintained (due to efficient $\chi a \to \chi a$ scattering), hence

$$\rho_\chi = \frac{\rho_\chi^{\rm eq}(T')}{n_\chi^{\rm eq}(T')}\, n_\chi \,, \qquad P_\chi = \frac{P_\chi^{\rm eq}(T')}{n_\chi^{\rm eq}(T')}\, n_\chi = T' n_\chi \,, \tag{2.29}$$

and similarly for the ALP. The hidden-sector equation of state is then given by

$$\rho' + P' = \frac{\rho_\chi^{\rm eq}(T')}{n_\chi^{\rm eq}(T')}\, n_\chi + \frac{\rho_a^{\rm eq}(T')}{n_a^{\rm eq}(T')}\, n_a + T'\left(n_\chi + n_a\right) \,. \tag{2.30}$$

The quantities n_χ, n_a and T' are finally obtained as the solutions of a system of three coupled differential equations, namely eq. (2.23) written in the form

$$z\frac{{\rm d}\rho'}{{\rm d}T'}\frac{{\rm d}T'}{{\rm d}z} = -3(\rho' + P') + \frac{1}{H}\int \frac{d^3p}{(2\pi)^3} C[f(p,t)] \tag{2.31}$$

in conjunction with the equation of state eq. (2.30), and the Boltzmann equations

$$\begin{aligned} Hz\frac{\mathrm{d}n_\chi}{\mathrm{d}z} + 3Hn_\chi = &\sum_f \left\langle \sigma_{\chi\bar{\chi}\to f\bar{f}} v \right\rangle (T) n_\chi^{\mathrm{eq}}(T)^2 + \sum_{i,j,k} \left\langle \sigma_{\chi\bar{\chi} i\to jk} v^2 \right\rangle n_i^{\mathrm{eq}} \left(n_\chi^{\mathrm{eq}}\right)^2 \\ &+ \left\langle \sigma_{aa\to\chi\bar{\chi}} v \right\rangle (T') n_a^2 - \left\langle \sigma_{\chi\bar{\chi}\to aa} v \right\rangle (T') n_\chi^2 \end{aligned} \tag{2.32}$$

$$\begin{aligned} Hz\frac{\mathrm{d}n_a}{\mathrm{d}z} + 3Hn_a = &\left\langle \Gamma_a \right\rangle n_a^{\mathrm{eq}}(T) + \sum_{i,j,k} \left\langle \sigma_{ia\to jk} v \right\rangle (T) n_a^{\mathrm{eq}}(T) n_i^{\mathrm{eq}}(T) \\ &- \left\langle \sigma_{aa\to\chi\bar{\chi}} v \right\rangle (T') n_a^2 + \left\langle \sigma_{\chi\bar{\chi}\to aa} v \right\rangle (T') n_\chi^2 \end{aligned} \tag{2.33}$$

with the initial condition for T' provided by the solution of eq. (2.28). Here i, j, k are SM particles involved in the a-number changing processes, see table 1.

2.5 Finite-temperature effects

We incorporate finite-temperature corrections in our analysis in light of the observation in ref. [8] that these have a large impact on the relic density in the case of sequential freeze-in. Such corrections arise due to the fact that perturbation theory breaks down in the presence of an additional scale, the temperature of the plasma. Calculations therefore require the resummation of diagrams, which can be implemented in a simple way, following ref. [8], on adopting the hard thermal-loop (HTL) approximation [15], where only loops involving soft momenta $\sim gT \lesssim T$ are resummed. Within this approximation, it turns out that the masses of fermions and gauge bosons appearing in the dispersion relations can simply be replaced by temperature dependent quantities. For the gauge bosons this amounts to adding a mass to the propagators which is the thermal Debye mass (neglecting differences between the transverse and longitudinal modes), which for the case of gluons is given by [16]

$$m_g^2(T) = \frac{4\pi\alpha_s(2\pi T)T^2}{3}\left(n_c + \frac{n_f(T)}{2}\right) , \tag{2.34}$$

with n_f the number of active flavours in the plasma, and for photons [16]

$$m_\gamma^2(T) = \frac{4\pi\alpha_{\mathrm{em}} T^2 n_{\mathrm{ch}}(T)}{3} , \tag{2.35}$$

with $n_{\mathrm{ch}}(T)$ the number of electromagnetically charged particles in the plasma.

For the case of fermions we also add a thermal mass correction to the propagators, neglecting the difference between particle and hole states [17]. For the leptons this takes the form [18]

$$m_l^2(T) = \frac{4\pi\alpha_{\mathrm{em}}(2\pi T)T^2 q_f^2}{8} \tag{2.36}$$

and for quarks

$$m_q^2(T) = \frac{4\pi T^2(\alpha_s(2\pi T)/6 + \alpha_{\mathrm{em}}(2\pi T)q_f^2)}{8} . \tag{2.37}$$

In order to capture the effect of thermal corrections on interaction vertices we renormalize the coupling constants at the scale of the first Matsubara mode, $\omega = 2\pi T$, making use of the renormalization group equations in vacuum [17]. Note that we perform this running for the electromagnetic, strong couplings and Yukawa couplings, although for the latter we found that this resulted in a negligible effect on the results.

3 Constraints

In this section we review the most important existing and future experimental constraints on our model. ALPs can significantly impact astrophysical, cosmological and collider processes, and the various constraints on ALPs have widely been discussed in the literature (see [19] for a review). In section 4 we will explore the regions of parameter space of our model where the correct relic density is obtained via different production mechanisms. We will therefore focus on those collider, astrophysical and cosmological constraints which are relevant in these regions.

While the discussion has been general so far, to discuss the constraints and for the subsequent numerical analysis we concentrate on the simplest case of strictly flavor-universal axion-fermion couplings, i.e. universal values of C_f and g_{aff} in eqs. (2.1) and (2.2).

3.1 Collider constraints

3.1.1 SLAC E137 electron beam dump experiment

Electron and proton beam-dump experiments are among the accelerator-based experiments where the ALPs in our model could be produced. Within the parameter region of interest a strong constraint comes from the E137 beam dump experiment at SLAC [20]. Here ALPs could be produced via Primakoff, bremsstrahlung or positron annihilation (see figure 1) and if these decay within the detector such events might be observed.

Calculating the number of events In the SLAC E137 electron beam dump experiment no events were detected [20]. This puts an experimental upper bound on the event rates $\mathcal{N}_s$

$$\mathcal{N}_s < \mathcal{N}_s^{\text{up}} = 2.996 \text{ for } \mathcal{N}_{s,\text{det}} = 0 \text{ at } 95\% \text{ CL}\,. \tag{3.1}$$

Following [19–21] we calculate the theoretically expected event rate to obtain a bound on the ALP's coupling and mass. The theoretical number of events can be obtained by an integral over all possible production cross sections folded with the probability to detect the ALP and the differential track-length distribution of the shower particles in the beam, multiplied by the total number of electrons dumped. To obtain the theoretical number of events in a specific decay channel we additionally multiply by the corresponding branching ratio, yielding

$$\mathcal{N}_{\text{theory}}(a \to XX) = \mathcal{N}_{e,\text{ inc}} \sum_{i,Z} P_Z \int \mathrm{d}E\; T_i(E)\, p(E)\, \sigma_Z^{i\to ak}(E)\; \mathcal{B}(a \to XX)\,. \tag{3.2}$$

$\mathcal{N}_{e,\text{ inc}}$ is the total number of electrons dumped (~ 30 Coulomb), P_Z is the number density of atoms per cm^3 of atomic number Z and $\sigma_Z^{i\to ak}$ are the cross sections for the different production processes with i the initial particle. T_i is the corresponding track-length distribution of the initial incoming particles, $i \in \{\gamma, e^+, e^-\}$. $p(E)$ is the probability for the ALP to decay visibly. The differential track-length distributions for the secondary shower particles can be estimated in the Weizsäcker-Williams approximation which allows one to relate the incident electrons in the initial state to the distributions of photons, electrons and positrons which

are produced in electromagnetic shower cascades. Intuitively, the differential track-length $T(E)$ is the length a particle with given energy E will travel. One then directly considers the interactions of these particles with the target [22]. We extract the track-length distributions for photons, electrons and positrons in the beam dump (figure 14 of reference [20]) which have been calculated using the SLAC EGS program [23]. This spares us from finding an analytic expression for the Weizsäcker-Williams approximation. Through interactions of the shower particles with the target material, the ALPs in our model are predominantly produced by the Primakoff process ($Z\gamma \to Za$), by bremsstrahlung ($Ze^{\pm} \to Ze^{\pm}a$) and by (non-)resonant positron annihilation ($e^+e^- \to \gamma a$ and $e^+e^- \to a$). The Feynman diagrams for these processes are shown in figure 1. We can sum over the production channels since all other final state particles produced in the process are absorbed in the shielding. The ALP subsequently decays and its decay products can be detected in the electromagnetic calorimeter. Therefore, in order to obtain the theoretically expected number of events, we include the branching fraction for the decay channel $a \to \gamma\gamma$ (loop induced) and $a \to e^+e^-$, which opens up for $m_a > 2m_e$. The various decay channels of the ALP in our model are summarised in appendix D.

Following ref. [21] we include the probability $p(E)$ for the ALP to be detected. Taking the experimental layout into account, we can envisage at least five scenarios where the ALP could have decayed invisibly (see appendix F): Either none or only one of the decay particles reaches the detector or both of them reach the detector but the opening angle is too small and the two final state particles are indistinguishable. For instance, if the ALP is highly boosted and its lifetime is long, it will certainly decay behind the detector. We will thus have to take the ALP decay probability into account to obtain a final result for the theoretically expected number of events. Details about the calculation of the probability for the ALP to decay visibly can be found in appendix F. We assume that most of the ALPs are produced in the cooling water [19].

Details about the ALP production processes As mentioned above, the relevant production processes for our model in electron beam dump experiments are the Primakoff process, bremsstrahlung and (non-)resonant positron annihilation (see figure 1). In the Primakoff process, depicted in figure 1a), a secondary photon emits an ALP in the vicinity of the nucleus. Since the ALP in our model does not couple to photons at tree level, the process is induced at one-loop level, with a loop function depending on the momentum transfer. Details about the cross section and the loop calculation relevant for this process can be found in appendices E.1 and H, respectively. In bremsstrahlung (see figure 1b), the electron is scattered off the mass shell of the nucleus and returns to the mass shell by emission of an ALP. Lastly, the ALP can be produced by (non-)resonant positron annihilation, see figures 1c) and 1d). The cross sections for bremsstrahlung and positron annihilation can be found in appendices E.2 and E.3. It is interesting to look at the interplay of the different contributions as a function of the ALP mass. From eqs. (E.1) and (E.10) we see that for models where the ALP couples to both electrons and photons at tree level the Primakoff

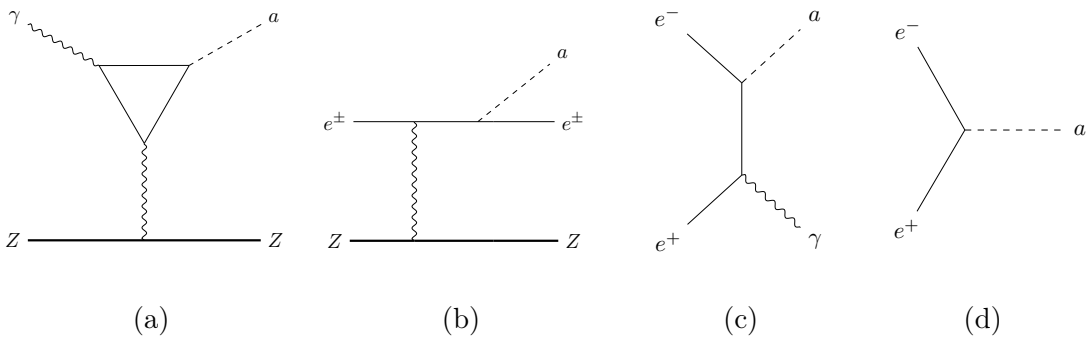


Figure 1: Feynman diagrams for the four different production mechanisms of ALPs in our model in the E137 beam dump experiment: (a) Primakoff production (b) bremsstrahlung (both off a target nucleus of atomic number Z) (c) non-resonant e^+e^- annihilation (d) resonant e^+e^- annihilation.

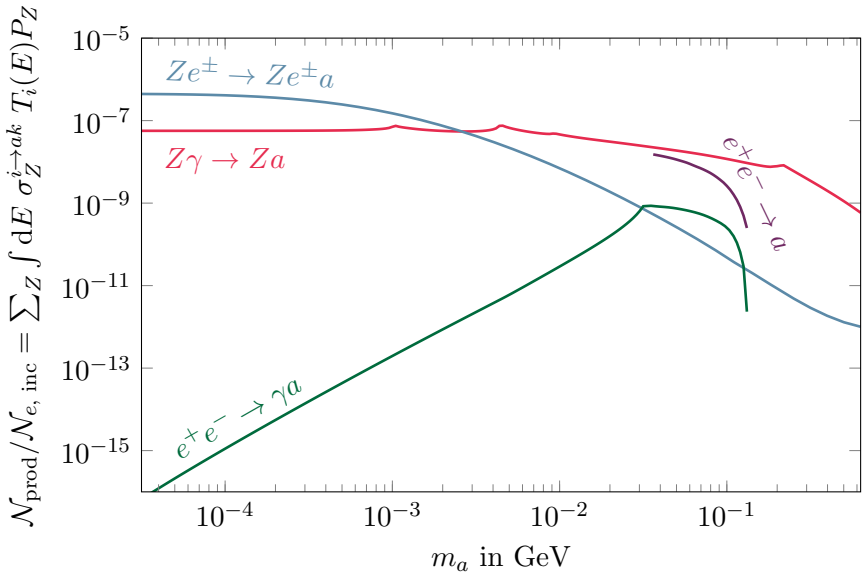


Figure 2: ALP yield per incident electron $\mathcal{N}_{\rm prod}/\mathcal{N}_{e,\,\rm inc} = \sum_Z \int {\rm d}E\ \sigma_Z^{i\to ak}\ T_i(E) P_Z$ for the various production processes of the ALP in our model in electron beam dump experiments, namely Primakoff production ($Z\gamma \to Za$), bremsstrahlung ($Ze^\pm \to Ze^\pm a$), non-resonant ($e^+e^- \to \gamma a$) and resonant ($e^+e^- \to a$) positron annihilation.

and the bremsstrahlung processes scale approximately as

$$\sigma_Z^{\gamma\to a} \equiv \sigma(Z\gamma \to Za) \sim \alpha_{\rm em} g_{a\gamma\gamma}^2 \quad \text{and} \tag{3.3}$$

$$\sigma_Z^{e\to ae} \equiv \sigma(Ze^\pm \to Ze^\pm a) \sim \alpha_{\rm em}^2 g_{aff}^2 \frac{m_e^2}{m_a^2} \ , \tag{3.4}$$

for $m_e \ll m_a \ll E_i$, with E_i the energy of the initial particle. This means that bremsstrahlung is suppressed by a factor $\sigma_Z^{e\to ae}/\sigma_Z^{\gamma\to a} \sim \alpha_{\rm em} m_e^2/m_a^2$ (see also ref. [24]). Interestingly, in our model, due to the loop suppression in the Primakoff process, see eq. (H.5), the cross sections scale as

$$\sigma_Z^{e\to ae}/\sigma_Z^{\gamma\to a} \sim m_e^2/(m_a^2 \alpha_{\rm em}) \tag{3.5}$$

and the Primakoff process becomes dominant only for larger ALP masses. This can be observed in figure 2 where we plot the ALP yield per incident electron from the various production processes depicted in figure 1,

$$\mathcal{N}_{\rm prod}/\mathcal{N}_{e,\ \rm inc} = \sum_Z \int {\rm d}E\ \sigma_Z^{i\to ak}\ T_i(E)\ P_Z \ , \tag{3.6}$$

as a function of the ALP mass m_a, for an ALP-electron coupling $g_{aff} = 1$ GeV^{-1} and ignoring the detection probability of the ALP. The range of accessible positron energies limits the range of producible ALPs to those with masses $m_{a,\rm min} \sim 35$ MeV for the resonant case and $m_{a,\rm max} \sim 135$ MeV for production by both resonant and non-resonant positron annihilation. For small ALP masses non-resonant positron annihilation scales approximately as $\sigma^{e\to a\gamma} \sim m_a^2$. While in the original analysis [20], ALP production by non-resonant positron annihilation was included, in later analyses positron annihilation is often neglected altogether since the positrons arise as secondary particles in the beam. Yet, the number of secondary positrons can be large enough for positron annihilation to become important for large ALP masses. To be precise, it turns out that it indeed dominates over the often considered bremsstrahlung for ALP masses around $50 - 100$ MeV and clearly cannot be neglected in models where the Primakoff process is highly suppressed, see also ref. [25] for a related discussion in the case of dark photons. In our case, however, the ALP coupling to photons arises from fermion loops and the suppression of photoproduction is such that positron annihilation is non-negligible, but not large enough for it to become dominant.

3.1.2 Exotic Higgs decays at the LHC

Production of the ALP in Higgs decays would be mainly via $h \to aa$ (for $m_a \leq m_h/2$) or $h \to Za$ (for $m_a < m_h - m_Z \simeq 85$ GeV). While it is possible to write extra effective operators for the haa as well as the hZa coupling, we assume for this work that only new terms are those in the Lagrangian in eq. (2.1). Therefore both these Higgs decay channels are mediated by a fermion loop and dominated by the top-quark contribution.

Higgs exotic decay searches at the LHC currently include $h \to Z(a \to gg)$ [26] and $h \to aa$ with (1) both $a \to b\bar{b}$ [27], (2) both $a \to \ell^+\ell^-$ [28], and (3) $aa \to b\bar{b}\mu^+\mu-$ [29]. A detailed study of these reveals (see appendix G for further details on the sensitivity of each of these) that current searches are not sensitive to coupling regimes that would correspond to $g_{aff} \lesssim 1/$GeV. We therefore can safely ignore limits from these searches.

3.1.3 Production in decays of mesons

For intermediate mass ranges ($2m_e \lesssim m_a \lesssim 5$ GeV), the best chance for detecting an ALP could be via decays of heavy mesons. Dedicated experiments provide the possibility to obtain rather precise measurements of branching fractions and thereby constrain couplings to very small values. We consider constraints on the emission of an ALP in rare B and K flavour-changing neutral current processes, proceeding at the quark level via $b \to sa$ or $s \to da$. Depending on the decay mode and the lifetime of the ALP, such decays could be constrained by $B \to K^{(*)}\ell^+\ell^-$ and $K \to \pi\ell^+\ell^-$ or $B \to K^{(*)}\nu\bar{\nu}$ and $K \to \pi\nu\bar{\nu}$. For ALP masses smaller than $2m_e$, the only decay possible is to photons, and the lifetime of the ALP is long enough that it decays outside the detector volume of current experiments. In this case, it would show up as an invisible decay mode of the said meson with the accompanying products identifiable.

The main experimental constraints that we consider in this work are therefore:

- $B^+ \to K^+X(\to \mu^+\mu^-)$ for long-lived scalar X (LHCb [30]), the 95% C.L. upper limits on the branching ratio are given as a function of the lifetime of X in the range 0.1 to 1000 ps.
- $B^0 \to K^{*0}X(\to \mu^+\mu^-)$ where X is a scalar particle with mass in the range 214 to 4350 MeV (LHCb [31]), the 95% C.L. upper limits on the branching ratio are given as a function of the mass and lifetime of X. The limit is of the order 10^{-9} over the majority of this range.
- $B^0 \to K^{(*)}X(\to \mu^+\mu^-)$ at fixed target experiments, limits can be extracted from CHARM results as described in ref. [32].
- $K^+ \to \pi^+\nu\bar{\nu}$ from NA62 [33], where 90% C.L. upper limits are given for the $K^+ \to \pi^+X$ branching ratio, where X is a long-lived scalar or pseudoscalar particle decaying outside the detector, for lifetimes longer than 100 ps.
- $K^\pm \to \pi^\pm e^+e^-$, where NA48/2 [34] provides a 90% C.L. upper limit (here we assume the lifetime should be less than 10 ns).
- $K^\pm \to \pi^\pm X(\to \mu^+\mu^-)$ for long-lived X (NA48/2 [35]) , the 90% C.L. upper limits on the branching ratio are given as a function of the lifetime of X in the range 100 ps to 100 ns.

We currently do not apply the $K \to \pi\gamma\gamma$ (which only dominates near 100 MeV) and $B_s \to \mu^+\mu^-$ (which only dominates for non-minimal flavour violation and for $m_a \sim 3-10$ GeV). Further the $B \to K\nu\bar{\nu}$ from BaBar [36] is not included as the relevant parameter space is covered by the $K^+ \to \pi^+\nu\bar{\nu}$ search.

In order to obtain the constraints on our parameter space, we first calculate the $B \to K^{(*)}a$ and $K \to \pi a$ branching ratios. For the former, the expression can be found in refs. [37] and [38]. Form factors for the $B \to K$ transition are taken from ref. [39] and for $B \to K^*$ from ref. [40]. For the Kaon decay and $K \to \pi a$, we follow ref. [41], taking into account the octet enhancement in non-leptonic Kaon decays.

For each point in parameter space, we then verify the ALP lifetime, which allows us to impose the appropriate limit, and then for the charged lepton channels multiply by the branching ratio for the necessary ALP decay, as given in appendix D. For the $B \to K^{(*)}X(\to \mu^+\mu^-)$, $K^+ \to \pi^+ X$ (where X is long lived) and $K^\pm \to \pi^\pm X(\to \mu^+\mu^-)$ channels, the experimental constraints are directly given for the contribution of a new particle X, as a function of the mass and lifetime of X, such that we do not require the SM branching ratio. These limits can easily be translated from the $m_a - \tau_a$ to the (m_a, g_{aff})-plane. For $K^+ \to \pi^+ e^+ e^-$ however we need to include the SM contribution, for which we adopt the chiral perturbation theory result from ref. [42], taking into account the associated theoretical uncertainty. The constraints from the CHARM experiment [43, 44] are extracted in ref. [32], by first calculating the B meson production spectra and then convoluting it with the branching ratio for $B \to K^{(*)}a$. In order to be detected, the ALP produced must then decay to muons within the detector volume. Making use of the probability of this occurring, a limit on the ALP coupling to fermions g_{aff} was obtained for a given mass m_a, for further details about this procedure see ref. [32].[4]

3.2 Astrophysical constraints

3.2.1 Supernova SN1987A

The SN1987A neutrino observations in 1987 in the Large Magellanic cloud provide us with another strong constraint. The neutrino flux coming from the core collapse, which lasted a few seconds, was measured. The presence of weakly-coupled particles in supernovae would provide an additional cooling mechanism. By comparing to data from SN1987A, bounds on the ALP coupling can therefore be derived. Here, following ref. [45, 46] we will adopt the "Raffelt criterion" to obtain these bounds, i.e. we demand that the luminosity emitted due to the ALP is less than that due to the neutrinos, $L_\nu = 3 \cdot 10^{52}$ ergs/s [47]. The cooling time should also be in agreement with the data. The high temperatures and density provided by SN1987A creates a conducive environment for the production of weakly-interacting particles such as the ALP. As the core temperature is around 30 MeV, and taking the Boltzmann tail into consideration, we could imagine that ALPs of masses up to $\mathcal{O}(100\text{ MeV})$ could be generated. As our ALPs only interact with fermions at tree level, the dominant production mechanism of the ALPs is via bremsstrahlung. In order to calculate the production rate, we first need to relate the quark Lagrangian to the nuclear Lagrangian:

$$\mathcal{L} \supset -i \sum_N \frac{m_N C_N}{f_a} a \bar{N} \gamma^\mu \gamma_5 N \ , \tag{3.7}$$

where we sum over the proton and neutron, $N = p, n$ with masses m_N and coupling to the ALP C_N respectively. Fortunately, the relation between the ALP-quark couplings and the ALP-nucleon couplings are well known, and the state of the art results can be found in ref. [48]. We then obtain $C_p = C_n = 0.43\, C_f$. Corrections to the diagrammatic calculation of the nuclear-scattering bremsstrahlung cross section are obtained using the results for

[4] We are very grateful to the authors of ref. [32] for providing us with the bounds they obtained in the $m_a - g_{aff}$ plane via private communication.

the spin-flip current at N^{3}LO in chiral perturbation theory. Three multiplicative factors reproducing these corrections were provided in ref. [46], and are included in our analysis:

- The factor $\gamma_f = 1/(1 + \frac{n_B \sigma_{np\pi}}{2\omega})$ acts as a cut-off preventing scattering at arbitrarily low energies, where n_B is the baryon number density, ω is the energy, and $\sigma_{np\pi} \simeq 15$ is the nucleon-nucleon scattering cross section when the pion is massless.
- The factor γ_p accounts for the change in phase space due to the non-zero pion mass, and is obtained from s defined in Eq. (49) of ref. [49], which we multiply by a factor $1/(1 - e^{-x})$ to account for the detailed-balance condition, following ref. [46].
- Finally γ_h contains the ratio of the mean free path of the weakly-coupled particle calculated at higher order in χPT to that at Born level, and is given by r_{Y_e} given in Eq. (5) of ref. [50].

Putting these factors together with the Born-level expression, the decay rate is given by:

$$\Gamma_a = \sum_N \frac{Y_N^2 \, C_N^2}{8 f_a^2} \, \frac{n_B^2 \, \sigma_{np\pi}}{\omega} \, \gamma_f \gamma_p \gamma_h \sqrt{1 - \frac{m_a^2}{\omega^2}} \, , \tag{3.8}$$

where Y_N is the mass fraction of the nucleon N (we adopt $Y_p = 0.3$ [45]). Making use of this decay width, on integrating over the volume V and the ALP phase-space we obtain the luminosity using [46]

$$L_a = \int_0^{R_\nu} dV \int \frac{d^3 k_a}{(2\pi)^3} \, \omega \, e^{-\omega/T} \, \Gamma_a \exp\left(- \int_0^{R_{\rm far}} dr \left(\Gamma_a + \Gamma_{all}^{\rm abs} \right) \right) \tag{3.9}$$

where k_a is the ALP momentum. The integration limits involve two radii: R_ν is the neutrinosphere radius, $\simeq$ 40 km, beyond which most neutrinos free stream until arriving at Earth, and the far radius is given by $R_{\rm far}$ $\simeq$100 km. Further the contribution to the absorptive width of the decay of the ALP to leptons is included explicitly via

$$\Gamma_{a\ell\ell}^{\rm abs} = \frac{m_a}{8\pi} \sum_\ell \Theta(m_a - 2m_\ell) \, \frac{m_\ell^2}{f_a^2} \sqrt{1 - \frac{4m_\ell^2}{m_a^2}} \, . \tag{3.10}$$

In order to calculate the luminosity, we further require the temperature and density profile of the proto-neutron star. These profiles are subject to large uncertainties, which have been estimated in the past by comparing the different profiles proposed in the literature. We adopt the profiles introduced in ref. [45], where a detailed study and comparison with other models for the hidden photon case is available. The comparison of the luminosity in the QCD axion and ALP case can be found in ref. [46].

3.2.2 Horizontal branch stars

The conditions inside horizontal branch (HB) stars in globular clusters would allow a sizeable production of ALPs inside the stellar core. HB stars lie in the horizontal branch in the Hertzsprung-Russell diagram, i.e. they contain little metal and are in the state of a

stable helium burning core and a hydrogen burning shell. Red Giants (RG) evolve to HB stars after their core becomes so hot and dense that the helium ignites. The interaction of ALPs with particles in the core of HB stars could alter the stellar evolution of these stars and in order to avoid conflicts with observations, constraints must be placed on the ALP parameter space. The constraints coming from HB stars are two-fold: If ALPs interact very weakly, they mostly escape freely, draining energy from the star. This puts an upper bound on the coupling. On the other hand, for larger couplings, they will be trapped inside the source and radiate energy which provides a lower bound.

Energy Loss Argument The first scenario we envisage is a very weakly interacting ALP, which would stream out freely of the hot core and accelerate the cooling of the star. A branch in the Hertzsprung-Russell diagram corresponds to a certain stage in the stellar evolution, depending on the type of nuclear fuel being burnt. Stars inside globular clusters travel along the different branches of the Hertzsprung-Russell diagram during their stellar evolution. Hence, the number of stars inside a certain branch is proportional to the time the star "lives" in a certain branch. The R-parameter which is the ratio of the number of HB stars over RG stars,

$$R = \frac{N_{\rm HB}}{N_{\rm RG}} = \frac{\tau_{\rm HB}}{\tau_{\rm RG}} \,, \tag{3.11}$$

is a suitable quantity to assess the impact of new particles on HB stars inside globular clusters: An additional energy loss mechanism by ALPs would manifest itself in a contraction and heating of the core, which would mainly have an effect on the nuclear fuel consumption (helium), hence reducing the star's Helium burning lifetime and therefore the number of HB stars. The Helium burning lifetime is related to the energy emitted per unit time and mass averaged over a typical HB core, $\langle\varepsilon\rangle$, via its luminosity. One usually assumes that the energy emission caused by additional particles should not exceed the energy emission from Helium burning, $\langle\varepsilon_a\rangle \lesssim \langle\varepsilon_{3\alpha}\rangle \approx 100$ erg g^{-1} s^{-1} [51]. The dominant production mechanism for ALPs with a coupling to electrons in HB stars is the Compton process. Here, we consider three possible ALP production mechanisms, the Compton process, bremsstrahlung and the loop induced Primakoff process. For the calculation of the energy emitted by ALPs we use eq. (12) in ref. [52] for the Compton process, expression (56c) given in ref. [53] for the bremsstrahlung process, and eq. (A2) in ref. [54] for the Primakoff process where we replace the axion-photon coupling by expression H.5, making the approximation of zero momentum transfer.

ALP opacity While in most models ALPs have very weak couplings, one could on the other hand imagine that the new particle has such a strong coupling that it is trapped inside the stellar core. In this scenario the ALP will scatter and decay inside the star, transporting heat between different regions and thereby contributing to radiative energy transfer. Again, observations indicate that models of stellar structure work well and that a new source of radiative energy transfer should therefore not exceed that due to photons, parametrized by the so-called Rosseland mean opacity. The photon opacity in HB stars is typically $\kappa_\gamma \sim 0.5$ cm^2/g. We include four sources for the ALP's effective opacity, namely (loop

induced) decay into photons and the inverse Primakoff process (using eqs. (A.2) and (A.6) from ref. [54], replacing $g_{a\gamma\gamma}$ by expression H.5), the Compton process and bremsstrahlung (using eq. (3.8) in ref. [51]).

3.3 Cosmological constraints

If a substantial number of ALPs are produced in the very early universe, they can affect the successful predictions for big bang nucleosynthesis (see for example [54–56]). Whether or not this is a concern depends on the ALP abundance, mass, and lifetime.

If the ALPs and the DM are still relativistic during the time of BBN (occuring at a photon temperature between $T \sim 1$ MeV and $T \sim 10$ keV), these additional relativistic degrees of freedom will contribute to the energy density of the universe, cf. eq. (2.12), which increases the speed at which it expands (see eq. (2.11)). A faster expansion causes an earlier freeze-out of neutrons, manifesting itself in a change of the neutron-to-proton ratio and changing the abundances of Helium-4 and Deuterium. The number of additional relativistic degrees of freedom around the time of BBN, $\Delta N_{\rm eff}^{\rm BBN}$, is thus constrained by measurements of these light element abundances [57]. In most of the scenarios we envisage, however, the hidden sector particles are frozen out before the QCD phase transition, and the hidden sector temperature is generally lower than the photon temperature.

Heavier ALPs with masses larger than $\sim 1 - 10$ MeV can still be constrained if they are abundant and long-lived. Out-of-equilibrium decays of ALPs after neutrino decoupling will heat up the plasma and therefore decrease the effective number of neutrinos $\Delta N_{\rm eff}^{\rm CMB}$ observed in the CMB. Since neutrino decoupling takes place just before BBN, comparably strong constraints on the ALP lifetime can also be inferred from $\Delta N_{\rm eff}^{\rm BBN}$ [55].

The combined constraints from $\Delta N_{\rm eff}$ rule out axion-like particles with either masses below $\mathcal{O}(10$ MeV$)$ or with lifetimes above $\mathcal{O}(0.01$ s$)$ when assuming that the ALPs couple predominantly to photons [55, 56]. However, this is not the case in our model.

Finally, for heavier ALPs, electromagnetic showers produced in ALP decays during or after BBN can destroy the newly created nuclei and thus directly alter the light element abundances, see [58, 59] for recent studies. For ALP masses above the GeV scale, hadronic showers give rise to additional constraints, excluding abundant hadronically decaying particles with lifetimes down to about $\mathcal{O}(0.1$ s$)$. This is because cascade hadrons can scatter off background protons, which once again increases the neutron-to-proton ratio [60]; see [61] for a recent numerical analysis.

3.4 Constraints on the DM

3.4.1 DM relic density constraint

Our model should explain the DM relic density measured today and thus has to satisfy the relic density constraint. The DM relic density today measured by Planck is [62]

$$\Omega_{\rm DM} h^2 = \frac{\rho_{\rm DM}}{\rho_{\rm crit}/h^2} = 0.120(1) \; . \tag{3.12}$$

Throughout the following sections, when calculating the DM relic densities for the different DM production mechanisms, we solve the Boltzmann equations until the comoving χ number density $Y_\chi = n_\chi/s$ stays constant. We then solve for the couplings which lead to the

observed DM relic density today, where we assume that χ and its antiparticle $\bar{\chi}$ make up all the DM. Once the number changing interactions stop, the DM number density will become redshifted due to the ongoing expansion of the universe. Using entropy conservation in a comoving volume, we relate the number densities after freeze-out (of either the DM, see section 2.4 or the bath particles annihilating into DM, see section 2.3) and the measured density parameter of DM today by

$$\Omega_\chi h^2 \equiv \frac{\rho_\chi}{\rho_{\rm crit}/h^2} = \frac{m_\chi n_{\chi,1}}{\rho_{\rm crit}/h^2}\frac{s_0}{s_1} = \frac{m_\chi Y_\chi s_0}{\rho_{\rm crit}/h^2}\,, \tag{3.13}$$

with $\rho_{\rm crit}/h^2 = 1.053672(24) \times 10^{-5}$ GeV cm^{-3} and $s_0 = 2891.2$ cm^{-3} [63] and where we have used the subscript "0" for quantities today and the subscript "1" for quantities after freeze-out, when we stop the simulation. Note that for our model $\Omega_\chi = \Omega_{\rm DM}/2$ (see section 2).

3.4.2 (No) constraints from dark matter phenomenology

We have verified explicitly that, as one would expect, the couplings between the DM particle and the SM are too small to lead to an observable signal at current and near future direct or indirect detection experiments, in both the freeze-in and the DFO region. In particular, the present limits from AMS02 on the DM annihilation cross section into SM particles [64, 65] would need to improve by several orders of magnitude to become relevant for our model. The same is true for present limits on dark matter-nucleon interactions in effective field theory for direct detection experiments [66].

Since our DM candidate has self-interactions via the axion couplings, it is interesting to ask whether it could resolve any of the small-scale structure problems plaguing the standard ΛCDM scenario. Along with possible hints for self-interacting DM come a variety of constraints from merging clusters (among them the Bullet cluster) or halo shapes, see ref. [67] for a summary. Most of the possible hints suggest a self-interaction cross section per DM mass of the order of $\sigma/m_\chi \sim 0.1 - 1$ cm^2/g. Interestingly, existing constraints turn out to be weaker or roughly of the same order as the cross sections needed to alleviate the tensions. Since the hidden sector particles are in equilibrium in the DFO region, the interaction strength might become large enough to lead to an observable signal in our model.

To quantify the self interaction strength in our model, we consider the momentum transfer cross section as proposed by [68].[5]

$$\sigma_{\rm T} = \int {\rm d}\Omega (1 - |\cos\theta|)\frac{{\rm d}\sigma}{{\rm d}\Omega}\,. \tag{3.14}$$

We find that, in the weakly coupled regime and at small velocities where the Born approximation is valid, the cross sections are too small to lead to sizeable interactions. In the non-relativistic strongly coupled regime for very light mediators, i.e. for

$$\frac{g_{a\chi\chi}^2 m_\chi^3}{4\pi m_a} > 1\,,$$

[5] Alternatively, the viscosity cross section $\sigma_{\rm V} = \int {\rm d}\Omega \sin^2\theta {\rm d}\sigma/{\rm d}\Omega$ weighting forward and backward scattering equally is sometimes used.

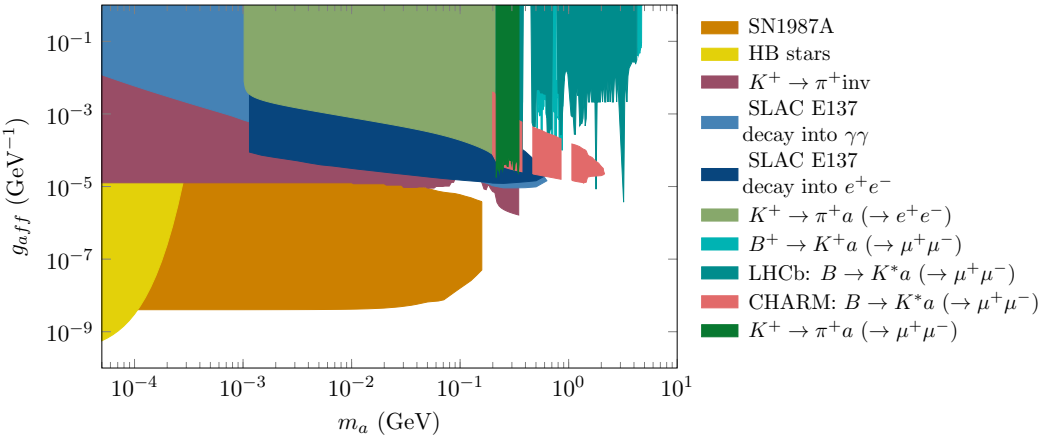


Figure 3: Full constraint plot on the ALP-fermion coupling g_{aff} as a function of the ALP mass m_a for an ALP which only couples to SM fermions at tree level. We show collider and astrophysical constraints discussed in section 3.

the cross section has been claimed to be Sommerfeld-enhanced [69–71]. However, as we will see in section 4.4, the largest possible DM-ALP coupling in the DFO region is of the order of $g_{a\chi\chi} \sim 2.5 \cdot 10^{-2}$ GeV^{-1}, such that if we keep the ratio $m_\chi/m_a = 10$ fixed, this condition is never met. Additionally, as shown recently in ref. [72] (see also ref. [73]), the potential arising from a pseudoscalar exchange does not lead to any significant Sommerfeld enhancement at small velocities. Hence, the DM particles in our model do not self-interact sufficiently to explain the discrepancies in structure formation observed at small scales.

4 Results

In this section we will first present the results of the relevant constraints on our model, providing bounds on the ALP-fermion coupling, g_{aff}, as a function of the ALP mass, m_a. As introduced in section 3, we focus on constraints from the electron beam dump experiment E137 at SLAC, from the supernova SN1987A, from horizontal branch stars and flavour constraints from heavy meson decays. Armed with the tools discussed in section 2 we subsequently study the generation of DM in the early universe, calculating the dark matter relic density generated for a large range of ALP-SM and ALP-DM couplings for a specific set of masses. This requires solving the Boltzmann equations in various different production regimes. We focus on the regions of parameter space needed to reproduce the observed DM abundance in the freeze-in and in the DFO regions, i.e. in DM genesis scenarios where the DM and the SM remain out of equilibrium, for a large range of ALP and DM masses. Finally, we compare these regions to the constraints on the ALP to study the phenomenological implications of our model. Throughout this section, when calculating the DM relic density, we fix the ratio $m_\chi/m_a = 10$.

4.1 Collider and astrophysical constraints

Figure 3 summarises the experimental bounds on ALPs discussed in section 3 we obtained for the model considered in this work, see eq. (2.1), where the ALP only couples to SM fermions at tree level.

We have considered a number of collider constraints coming from SLAC E137, NA62, NA48, LHCb and CHARM. For the electron-beam dump at SLAC we solve $\mathcal{N}_{\text{theory}}(m_a, g_{aff}) = \mathcal{N}_s^{\text{up}} = 2.996$, with $\mathcal{N}_{\text{theory}}$ given by eq. (3.2), for different values of m_a. The ALP can be produced by the Primakoff process, by bremsstrahlung or by electron-positron annihilation. It can decay into two photons or, for $m_a > 2m_e$, into an electron-positron pair. For even larger ALP masses, various other decay channels open up, cf. figure 9. The original analysis [20] considers separately 1) Primakoff production with decay into either two photons or into e^+e^-, 2) bremsstrahlung production with decay into e^+e^- and 3) non-resonant positron annihilation with subsequent decay into e^+e^-. In the present work, we provide a complete analysis without making assumptions on dominant production and/or decay channels. In order to provide a combined bound, we sum over all possible production processes, since the only observable final-state particles are the decay products of the ALP (additional final-state particles from the production process are absorbed in the beam dump absorber). In figure 3 we depict in light and dark blue the constraints from a decay of the ALP into two photons and an electron-positron pair, respectively. As explained in more detail in section 3, a too large ALP-fermion coupling would lead to an overproduction of ALPs, whereas ALPs with too small couplings are long-lived and would escape the detector. The upper bound thus stems from the fact that the detection probability of ALPs decaying too early becomes exponentially suppressed, leading to the typical nose-like shape in both SLAC beam dump constraints. Comparing our results to figure 19 of ref. [74], we find that the combination of several production processes excludes a slightly larger region of parameter space. This stems from the dominant Primakoff production process for larger ALP masses, cf. figure 2. Since the loop interaction is kinematically suppressed, the bounds on the ALP-fermion coupling g_{aff} for an ALP decaying into two photons are shifted towards larger couplings compared to the constraints on a tree-level ALP-photon coupling, see for instance figure 2 in ref. [19].

Flavour constraints also have an important impact on the parameter space considered. The calculation of the bounds from rare B and K decays shown in figure 3 was discussed in section 3. Rare B decays provide access to the parameter space for masses up to $m_a \sim m_B - m_K^{(*)}$. The LHCb bounds from $B^+ \to K^+X(\to \mu^+\mu^-)$ are shown in cyan, and from $B^0 \to K^{*0}X(\to \mu^+\mu^-)$ in dark cyan, for a long-lived scalar X, as indicated. We further show the CHARM bound coming from the $B^0 \to K^{(*)}X(\to \mu^+\mu^-)$ analysis in pink. Together, these form a powerful probe of g_{aff}, excluding almost completely couplings down to $\mathcal{O}(10^{-5})$ GeV^{-1}, as pointed out in ref. [32]. Coming to rare K decays, we show the NA48/2 limits on $K^\pm \to \pi^\pm e^+e^-$ in light green and from $K^\pm \to \pi^\pm X(\to \mu^+\mu^-)$ for long-lived X in dark green as well as the NA62 limit from $K^+ \to \pi^+\nu\bar{\nu}$ in purple. These three bounds are highly powerful and complementary probes of the parameter space, providing complete coverage down to $g_{aff} \sim 10^{-5}$ for $m_a \lesssim m_K - m_\pi$. Note that the

uneven nature of certain constraints from B and K meson decays is a direct artefact of the experimental limits, and the gaps in the constraints are due to the fact that in regions where m_a corresponds to the mass of certain mesons, a reliable limit cannot be obtained.

The astrophysical constraints from SN1987A and from HB stars in globular clusters are depicted in orange and yellow respectively. The most stringent constraint relevant for the freeze-in region is from SN1987A. This restricts the coupling g_{aff} to be less than $\sim 4 \cdot 10^{-9}$ GeV^{-1} for ALP masses up to $\mathcal{O}(100)$ MeV. The only constraint which probes lower couplings is that from HB stars, but this is only relevant for very small values of m_a. As described in section 3.2.1, the bound is obtained, following ref. [45, 46], by demanding that the luminosity emitted due to the ALP is less than that due to the neutrinos, $L_\nu = 3{\cdot}10^{52}$ ergs (i.e. the "Raffelt condition"). The upper line in the constrained region comes from the fact that if the coupling is too large the ALPs would not escape from the Supernova, and the lower line comes from the fact that if the coupling is too low, fewer ALPs would be produced. Note that our bound is obtained by following the calculation in ref. [45, 46], which involves several updated nuclear physics calculations, as well as including the energy dependence of the optical depth such that the energy loss at large couplings is correctly accounted for, and the bound therefore differs by an order of magnitude from previous work. We have verified that multiplying or dividing this luminosity by a factor two would result in a change in the limit on the coupling of less than 50%.

The constraint coming from HB stars excludes a wide region of parameter space for small ALP masses (1 eV $< m_a <$ 10 keV) and small couplings $g_{aff} \lesssim 3 \times 10^{-9}$ GeV^{-1} (yellow), by requiring $\langle \epsilon_a(m_a, g_{aff}) \rangle \lesssim 100$ erg g^{-1} s^{-1}. Not surprisingly, since the dominant production mechanism in HB stars for ALPs with a coupling to fermions is the Compton process and since the Primakoff process in our model is loop-suppressed, we essentially obtain the same bound as in figure 4 in ref. [51], which the authors calculated considering the Compton process only. It was pointed out in ref. [51] that radiation is not the dominant mechanism for heat transfer inside HB stars and that indeed convection is more efficient. Let us stress that this analysis is a crude estimate of the bound for our model and therefore most likely subject to large uncertainties. Making dedicated simulations is beyond the scope of this paper, in particular since we are mainly interested in ALPs with masses ranging in the MeV-GeV range for which the number densities are Boltzmann-suppressed in HB stars. For more recent analyses of constraints on the ALP-photon coupling from HB stars see for instance refs. [75–77]. A global analysis from HB stars, red giants and white dwarfs suggests a non-zero ALP-SM coupling (*HB-hint*) [78]. Other strong astrophysical constraints on an ALP-electron coupling come from the delay of Helium ignition in red giants, see for example [79].

It is worthwhile mentioning that we have not made any assumptions about the branching ratios or preferred decay modes for the calculation of the bounds on our model. Since the ALP only couples to fermions at tree level, the branching ratios for the various production and decay modes follow naturally.

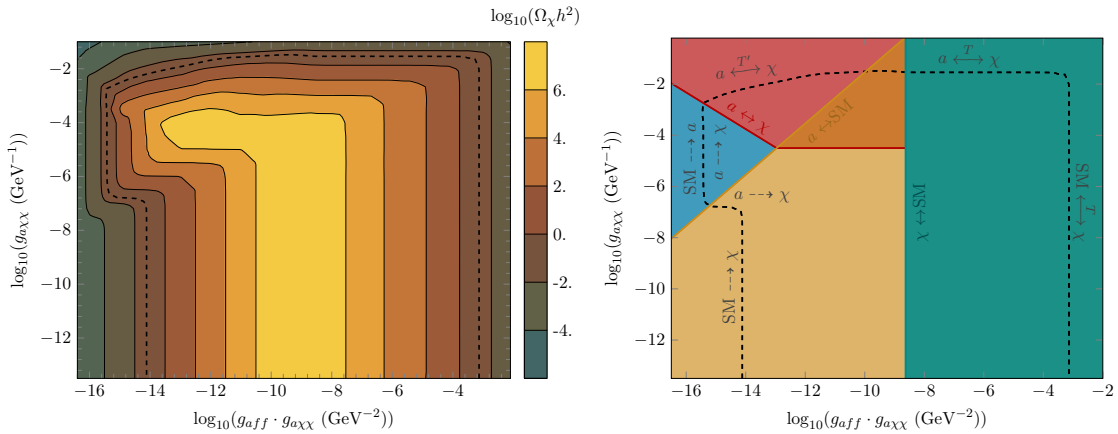


Figure 4: Hidden sector coupling $g_{a\chi\chi}$ as a function of the product of the hidden sector and the ALP-fermion coupling, $g_{a\chi\chi} \cdot g_{aff}$, for $m_\chi = 10$ GeV and $m_a = 1$ GeV. The dashed line corresponds to the combination of couplings which give the correct DM relic density $\Omega_{\rm DM}h^2 = 0.12$ as measured by the Planck telescope. *Left:* Contour plot of the DM relic density as a function of $g_{a\chi\chi} \cdot g_{aff}$ and g_{aff}, giving the so-called "Mesa" phase diagram: For our choice of masses six different DM genesis scenarios are possible, namely freeze-in from the SM, freeze-in from ALPs, sequential freeze-in, decoupled freeze-out (DFO), freeze-out from the hidden sector and freeze-out from the SM. *Right:* Different phases of the diagram, corresponding to equilibrium relations between the three sectors. The equilibrium relations essentially determine the production mechanism.

4.2 Phase diagram

As described in detail in section 2, the interplay between the three different sectors in our model, namely the SM particles, the ALP and the DM, gives rise to various DM production mechanisms. Solving the Boltzmann equations in a large range of hidden sector and connector couplings and plotting the final abundance as a function of these couplings gives what in [6] (and later in [7] for the massive mediator case) was dubbed a "Mesa" phase diagram. For our model, this phase diagram is shown in the left panel of figure 4 for an ALP with mass $m_a = 1$ GeV and the DM with mass $m_\chi = 10$ GeV. The dotted line indicates the combination of couplings $g_{a\chi\chi}$ and g_{aff} such that the final abundance corresponds to the observed DM density.[6] To aid the understanding of what follows, we also plot in the right panel of figure 4 the different phases of the diagram, corresponding to equilibrium relations between the three sectors, as these essentially determine the production mechanism. Let us have a look at the various regimes in more detail:[7]

[6]Note that we plot the abundance as a function of $g_{a\chi\chi}$ and the product of $g_{a\chi\chi} \cdot g_{aff}$ as these are the proportionality factors for the competing DM production processes, cf. table 1.

[7]This is extensively discussed in ref. [7], and we will therefore only briefly comment on the different shapes of the regions.

SM $\dashrightarrow \chi$
$(a \overset{T}{\longleftrightarrow} SM)$

Freeze-in from SM. Starting from the bottom-left (small $g_{a\chi\chi}$ and $g_{a\chi\chi} \cdot g_{aff}$) of the phase diagram in figure 4, the dominant process is $f\bar{f} \to \chi\bar{\chi}$ and, for reheating temperatures above $T_{RH} \gtrsim 200$ GeV, ultraviolet-dominated $2 \to 3$ interactions. DM is produced from scattering of SM fermions in the thermal bath without ever reaching equilibrium. From equation (2.17) we realise that the final relic density is proportional to $(g_{a\chi\chi} \cdot g_{aff})^2$ (cf. table 1). The size of each term individually is a free parameter, however, g_{aff} is large enough to ensure a-SM equilibrium. Hence, we expect a vertical line in the $(g_{a\chi\chi} \cdot g_{aff}, g_{a\chi\chi})$ phase diagram.

$a \dashrightarrow \chi$
$(a \overset{T}{\longleftrightarrow} SM)$

Freeze-in from ALPs. Going upwards in the phase diagram, i.e. increasing the value of the hidden sector coupling, while still keeping g_{aff} large enough for efficient ALP-SM scattering, collisions of thermal ALPs become more likely than collisions of thermal SM particles. Following the same arguments as above, Y_χ scales like $g^4_{a\chi\chi}$ and is therefore independent of g_{aff}. Hence, we expect a horizontal line in the $(g_{a\chi\chi} \cdot g_{aff}, g_{a\chi\chi})$ phase diagram, appearing as a plateau in figure 4.

SM $\dashrightarrow a$
$a \dashrightarrow \chi$
$(a \cdots \chi)$

Sequential freeze-in. The relic abundance is obtained by a chain of sequential reactions $\sum_{i,j,k} ij \to ak$ (with $\{i, j, k\}$ SM particles) followed by $aa \to \chi\bar{\chi}$. Hence, $Y_\chi \propto \langle \sigma_{aa\to\chi\bar{\chi}} v \rangle \, n_a^2$, i.e. $Y_\chi \propto (g_{a\chi\chi})^4$. Similarly, $n_a \propto \sum_{i,j,k} \langle \sigma_{ia\to jk} v \rangle \, n_a^{\rm eq}(T) n_i^{\rm eq}(T)$, i.e. $n_a \propto (g_{aff})^2$. Finally, $Y_\chi \propto (g_{aff} \cdot g_{a\chi\chi})^4$ and we expect a vertical line, but shifted to the left because the ALPs and the DM are now both out of equilibrium. Note that our calculation in this region is subject to uncertainties (see section 2.3.3).

$a \overset{T'}{\longleftrightarrow} \chi$
(SM $\dashrightarrow a$)
(SM $\dashrightarrow \chi$)

DFO. As explained in section 2.4, the ALP-SM and the SM-DM processes enter the final relic density only weakly via T' (see also section 4.4 for an extended discussion). The dependence on g_{aff} is mild, hence the almost horizontal line in the phase diagram. For increasing g_{aff}, more hidden sector particles are produced and the strength among the hidden sector particles has to slightly increase to ensure a later freeze-out of $\chi-$ particles.

$a \overset{T}{\longleftrightarrow} \chi$
$(a \overset{T}{\longleftrightarrow} SM)$

Thermal freeze-out from ALPs. Going further to the right in the phase diagram, i.e. further increasing g_{aff}, the ALPs and the SM and then also the SM and DM will equilibrate and all three sectors will share a common temperature. In the right half of the phase diagram DM will therefore be produced by freeze-out. However, g_{aff} is still small enough for $aa \leftrightarrow \chi\bar{\chi}$ to be the more efficient process. The behaviour is as in DFO except that the weak dependence on g_{aff} disappears completely, since the hidden and the visible sector share the same temperature.

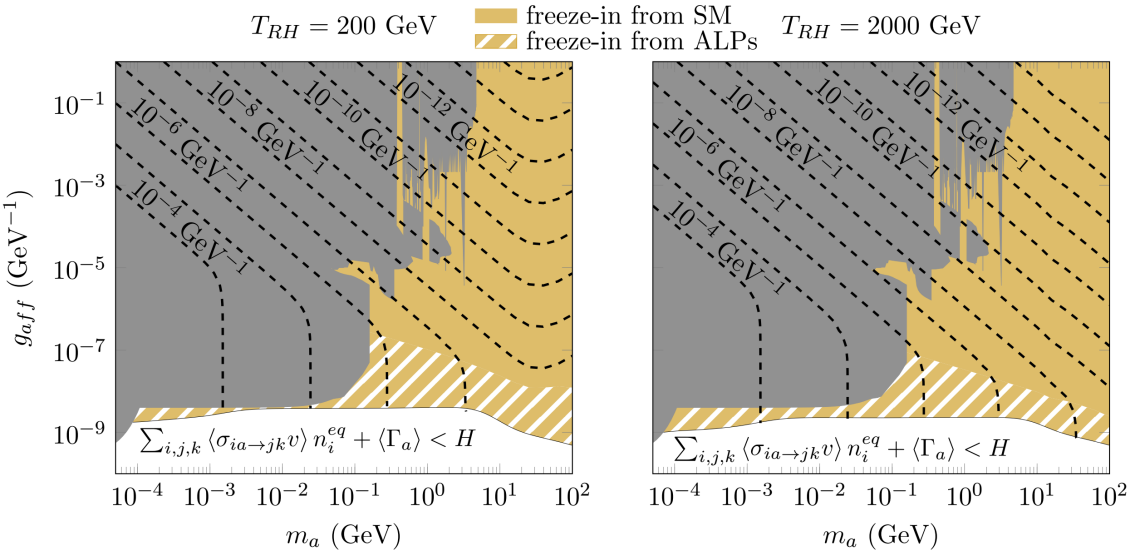


Figure 5: ALP-fermion coupling g_{aff} as a function of m_a for lines of constant hidden-sector coupling $g_{a\chi\chi}$ (plotted in dashed black) which reproduce the observed DM relic density via freeze-in from SM and freeze-in from ALPs as indicated for reheating temperatures $T_{RH} = 200$ GeV (left) and $T_{RH} = 2000$ GeV (right). We fixed the ratio $m_\chi/m_a = 10$. The lower line indicates the value of g_{aff} for which the ALPs and the SM reach equilbrium. In grey, we have included the relevant constraints on our ALP model on the connector coupling g_{aff} in this parameter region (cf. figure 3).

SM $\overset{T}{\longleftrightarrow} \chi$
$(a \overset{T}{\longleftrightarrow} SM)$

Thermal freeze-out from SM. This is the usual thermal freeze-out from $2 \to 2$ DM-SM scattering which in our model is set by $f\bar{f} \leftrightarrow \chi\bar{\chi}$ scattering and hence by the product $(g_{a\chi\chi} \cdot g_{aff})^2$. As in freeze-in from SM we expect a vertical line, except that the final relic abundance depends now inversely on $(g_{a\chi\chi} \cdot g_{aff})^2$.

4.3 Freeze-in region

Having discussed collider and astrophysical constraints on the mediator in our model in the previous section, we now investigate the phenomenological implications of a feebly interacting DM particle. The freeze-in region is defined by the couplings of the DM to both the ALPs and the visible sector being so tiny that the DM particles never reach thermal equilibrium. Their initial abundance being zero or negligible, the DM particles are gradually produced by out-of-equilibrium scattering of particles in the thermal bath. The ALPs, on the other hand, are in thermal equilibrium with the SM in the conventional freeze-in scenario. (We do not consider DM production via sequential freeze-in as its description is subject to theoretical uncertainties, as described in section 2.3.3.) Hence the reaction

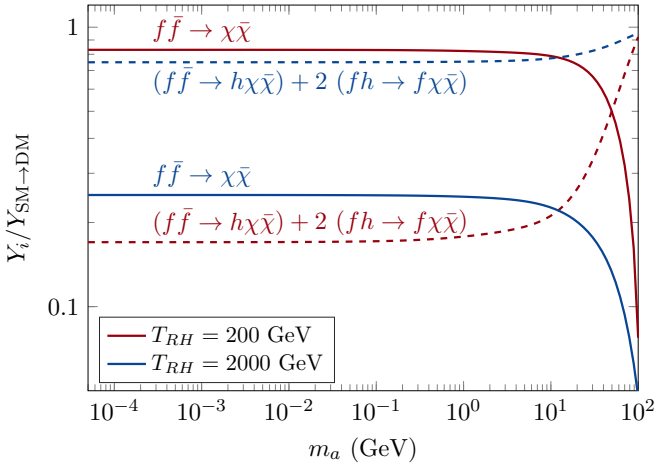


Figure 6: Relative contributions of infrared-dominated interaction $f\bar{f} \to \chi\bar{\chi}$ (solid) and ultraviolet-dominated $2 \to 3$ interactions $(f\bar{f} \to h\chi\bar{\chi})+2(fh \to f\chi\bar{\chi})$ (dashed) to the final χ-abundance as a function of the ALP mass in the freeze-in from SM regime for reheating temperatures $T_{RH} = 200$ GeV (red) and $T_{RH} = 2000$ GeV (blue). We fixed the ratio $m_\chi/m_a = 10$.

rates satisfy

$$\sum_{i,j,k} \langle \sigma_{ia\to jk} v \rangle \, n_i^{\rm eq}(m_\chi) + \langle \Gamma_a \rangle \; > H(m_\chi) \qquad \text{(equilibrium of ALP-SM)}\,,$$

$$\left.\begin{array}{l} \displaystyle\sum_{i,j,k} \left\langle \sigma_{\chi\bar{\chi}[k]\to ij} v^{[2]} \right\rangle [n_k^{\rm eq}(m_\chi)] \; n_\chi^{\rm eq}(m_\chi) \\ \qquad \langle \sigma_{\chi\bar{\chi}\to aa} v \rangle \, n_\chi^{\rm eq}(m_\chi) \end{array}\right\} < H(m_\chi) \qquad \text{(DM out of equilibrium)}\,, \quad (4.1)$$

where, in the second line, the symbols in square brackets pertain to $2 \to 3$ interactions. Depending on the relative size of the couplings g_{aff} and $g_{a\chi\chi}$, DM generation by either $ij \to \chi\bar{\chi}[k]$ (where i, j, k are SM particles) or $aa \to \chi\bar{\chi}$ scattering will be more efficient. This results in two distinct regimes which we label *freeze-in from SM particles* (cf. section 2.3.1) and *freeze-in from ALPs* (cf. section 2.3.2). In these regimes, the DM relic abundance is given by eqs. (2.17) and (2.19), respectively. Since the relic density is proportional to the couplings entering the matrix element squared of the dominant process, it suffices to factorize the couplings and solve for those couplings satisfying the relic density constraint in eq. (3.12).

As discussed in section 2.3.1, ultraviolet-dominated contributions to the DM relic density from a-mediated $2 \to 3$ scattering processes $f\bar{f} \to h\chi\bar{\chi}$, $fh \to f\chi\bar{\chi}$ and $\bar{f}h \to \bar{f}\chi\bar{\chi}$ become important for reheating temperatures above a few hundred GeV. These contributions introduce a dependence on the reheating temperature since the final relic abundance scales with T_{RH}, see eq. (2.18). Here we consider two representative scenarios: For $T_{RH} = 200$ GeV, UV-dominated processes are practically negligible; the DM abundance will therefore be independent of the exact value of T_{RH} if it is chosen even lower (but above

the temperature of top quark freeze-out). By contrast, for $T_{RH} \sim 2000$ GeV, the result is fully dominated by the UV-dominated processes. This behaviour can be observed in figure 6 where we compare the relative contribution of the IR- and UV-dominated processes to the final DM abundance as a function of the ALP mass in the freeze-in from SM scenario, i.e. where $aa \to \chi\bar{\chi}$ scattering is negligible. We keep the ratio $m_\chi/m_a = 10$ fixed. Indeed, for $T_{RH} = 200$ GeV, the DM abundance is set by $f\bar{f} \to \chi\bar{\chi}$ scattering, except for large ALP masses where the $2 \to 3$ interactions always dominate. By contrast, for reheating temperatures of the order of a few TeV, it is the ultraviolet contributions that set the final DM abundance.

In figure 5 we depict the results for these two scenarios in a large range of ALP masses, together with the constraints derived in the previous section. We plot the value for the ALP-SM coupling g_{aff} as a function of the ALP mass m_a for lines of constant hidden sector coupling $g_{a\chi\chi}$ (dashed black lines). Inside the yellow region DM production from SM fermion scattering is more efficient (freeze-in from SM particles), and inside the yellow hatched region production from ALP scattering (freeze-in from ALPs). These freeze-in regimes span a vast region of parameter space in the (m_a, g_{aff})-plane. We cut the plot at the upper theoretical bound $C_f \lesssim 4\pi$ which results in $g_{aff} = C_f/f_a \lesssim 1$ GeV^{-1} since $f_a \gg v$ with v the electroweak scale, cf. appendix A. Below the lower boundary the χ-particles are produced via sequential freeze-in or decoupled freeze-out (DFO) because the ALPs and the visible sector cease to share the same temperature.

The point where the ALPs equilibrate with the SM plasma is affected by $\mathcal{O}(1)$ finite-temperature corrections and by UV-dominated processes entering the equilibrium conditions eq. (4.1), notably $f\bar{f} \to ah$. These can enhance ALP production at earlier times. As a consequence, the simple conditions in eq. (4.1) are only approximately true and in fact, equilibration can be attained for some $T \gtrsim m_\chi$. To obtain the lower boundary in figure 5, we solve the Boltzmann equation for the ALP number density numerically and determine the couplings for which $n_a(T) = n_a^{\rm eq}(T)$ for some $T_{RH} \gtrsim T \gtrsim m_\chi$.

In general, the ALP-SM interaction strength g_{aff} must be comparably large if thermal equilibrium is to be reached, and is therefore likely in reach of upcoming experiments. For $g_{aff} \gtrsim 10^{-6} - 10^{-5}$ GeV^{-1} and m_a below a few GeV, the parameter space is ruled out by constraints from rare B and K decays and the electron beam dump experiment at SLAC. For masses below ~ 0.2 GeV, the remaining parameter space is covered by the SN constraint, see also figure 3. However, we remark that the relic density constraint inhibits large hidden sector couplings: $g_{a\chi\chi}$ has to be as small as $\sim 10^{-15}$ GeV^{-1} for the largest connector couplings and for large ALP masses as in this parameter region DM production is controlled by $f\bar{f} \to \chi\bar{\chi} + (2 \to 3$ terms) and therefore by the product of couplings $g_{aff} \cdot g_{a\chi\chi}$ (see table 1).

When the DM is dominantly produced by freeze-in from ALPs, we expect the contours in constant $g_{a\chi\chi}$ in the (m_a, g_{aff})-plane to be vertical because the relic density is independent of the connector coupling and proportional to $(g_{a\chi\chi} m_\chi)^4$. In the freeze-in from SM scenario the relic density is both proportional to $m_\chi^2 \propto m_a^2$ and to g_{aff}^2. For $T_{RH} = 200$ GeV and DM masses above the mass of the top quark, this dependence is weakened by the Boltzmann suppression of the particles predominantly producing the DM via $t\bar{t} \to \chi\bar{\chi}$, leading to

an upwards bending of lines of constant hidden sector coupling. Here, interactions have to become stronger; between reheating and freeze-out of the top there is barely enough time to produce the heavy DM particle. For the ultraviolet-dominated case, the lines of constant $g_{a\chi\chi}$ get shifted towards smaller values of g_{aff} since more DM is produced at earlier times. Not surprisingly, the freeze-in from the mediator scenario is independent of the two reheating temperatures since $aa \to \chi\bar{\chi}$ is infrared dominated and the ALPs belong to the SM bath. We would like to add that, depending on the temperature, finite-temperature corrections can influence the DM production from fermion scattering by up to $\mathcal{O}(100\%)$, resulting in a change in the final relic density of $\sim 10\%$ for $T_{RH} = 200$ GeV and $\sim 5\%$ for $T_{RH} = 2000$ GeV.

Note that, in principle, ALPs with masses $m_a \lesssim$ MeV which are still relativistic during big bang nucleosynthesis are severely constrained by cosmology, see section 3. We do not include these bounds in our analysis since the relevant parameter space is already excluded by collider experiments or astrophysics. For larger masses, lifetimes above $\tau \sim 0.01$ s can be excluded [55]; yet in the freeze-in regime the ALP-fermion coupling strength is large such that lifetimes are < 0.01 s.

In summary, we find that for feebly interacting dark matter, which has tiny couplings to the mediator, the correct relic density can be obtained in a large region of unexcluded parameter space, likely within the reach of future experiments.

4.4 DFO region and cosmological constraints

As the hidden sector coupling $g_{a\chi\chi}$ increases, the interactions among the ALPs and the DM become frequent enough to bring them in thermal equilibrium, however at a temperature T' which is distinct from that of the photons, T. These now constitute a dark sector which is thermally decoupled from the visible sector. Energy is gradually transferred to the hidden sector by rare scatterings of SM particles into both the ALPs and the DM. The DM relic density is set by $aa \leftrightarrow \chi\bar{\chi}$ interactions and the mechanism resembles ordinary freeze-out but occurs at a different temperature, i.e. by *decoupled freeze-out (DFO)*. To obtain the relic density in the DFO regime we proceed as outlined in section 2.4 and section 3.4.1. The calculation is more involved than for freeze-in, since now a stiff system of coupled differential equations (2.31), (2.32) and (2.33) has to be solved, which we have accomplished using the stiff solver package dvode [80]. We track the evolution of $Y_\chi \equiv n_\chi/s$ until the DM particles freeze out and Y stays constant (more precisely, we require the relative change in Y_χ to be smaller than $\varepsilon = 5 \cdot 10^{-4}$). As in the case of freeze-in, here too, UV contributions to SM-DM and SM-ALP interactions can introduce a dependence on the reheating temperature. However, these only enter indirectly via the energy transfer from the SM. We have checked this explicitly and find that for reheating temperatures below $\mathcal{O}$(TeV) they do not significantly influence the final DM relic abundance. To be more specific, we find that a variation of $T_{RH} = 200$ GeV and $T_{RH} = 2000$ GeV leads to a change of $\sim 22\%$ in the final DM relic abundance and a change $\sim 7\%$ in the required value of $g_{a\chi\chi}$. Hence, the DFO scenario retains very little sensitivity to the ultraviolet for reheating temperatures below a few TeV, and in the following we can neglect these contributions.

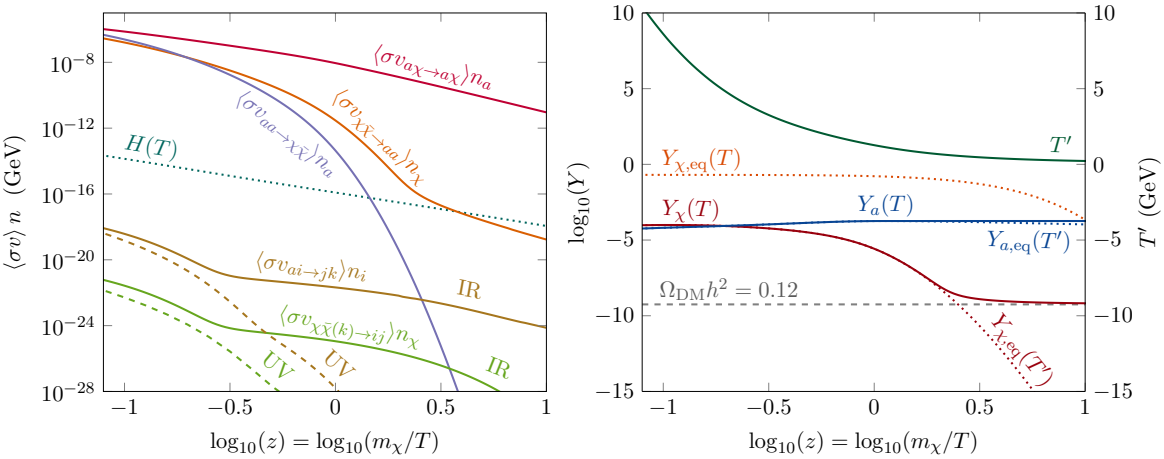


Figure 7: Illustrative example of the evolution of the number densities and the reaction rates in the DFO scenario for $m_\chi = 10$ GeV and $m_a = 1$ GeV and for a set of couplings which give the measured dark matter relic abundance: $g_{aff} = 9.44 \cdot 10^{-12}$ GeV^{-1} and $g_{a\chi\chi} = 1.06 \cdot 10^{-2}$ GeV^{-1}. *Left:* Evolution of the different terms competing on the right hand side of the Boltzmann equations for n_χ and n_a as a function of $\log_{10}(z) = \log_{10}(m_\chi/T)$. We also plot the evolution of the reaction rate for elastic scattering among the hidden sector particles, $a\chi \to a\chi$. *Right:* The evolution of the comoving number density of χ (red) and a (blue) as a function of $\log_{10}(z) = \log_{10}(m_\chi/T)$. The equilibrium distributions $Y_{\chi,\mathrm{eq}}(T')$ (dotted red) and $Y_{a,\mathrm{eq}}(T')$ (dotted blue) and $Y_{\chi,\mathrm{eq}}(T)$ (dotted orange) are also shown for reference. We plot along the evolution of the hidden sector temperature T' (green).

However, we do add finite-temperature corrections to both ALP and DM production from the SM, which result in $\mathcal{O}(1)$ corrections, mainly originating from thermal mass effects.

In the left panel of figure 7 we present an illustrative example of the evolution of the various reaction rates at play in the DFO regime, together with the Hubble rate, for $m_\chi = 10$ GeV and $m_a = 1$ GeV and for a set of couplings which give the measured dark matter relic abundance ($g_{aff} = 9.44 \cdot 10^{-12}$ GeV^{-1} and $g_{a\chi\chi} = 1.06 \cdot 10^{-2}$ GeV^{-1}). For reactions faster than the Hubble rate, i.e. lines above the turquoise dotted Hubble line, the particles involved achieve equilibrium. Once the reaction rates drop below the Hubble rate, the particles become so dilute that they no longer collide sufficiently often and the process shuts off. We also distinguish between the infrared- (solid) and ultraviolet- (dashed) dominated SM-ALP and SM-DM interactions. A few observations can be made: Firstly, $aa \leftrightarrow \chi\bar\chi$ interactions are the dominant DM number changing interactions throughout the evolution. The mechanism which sets the relic abundance is therefore conceptually very similar to the simpler freeze-out scenario, however governed by T' rather than T. Secondly, neither the SM-ALP nor the SM-DM reaction rates ever exceed the Hubble rate, and the hidden sector is indeed decoupled. Nevertheless, the SM slowly transfers energy to the hidden sector both via SM-ALP and SM-DM interactions, with the dominant contribution coming from SM-ALP interactions. Lastly, elastic scattering $a\chi(\bar\chi) \leftrightarrow a\chi(\bar\chi)$ stays active until after

freeze-out of the χ-particles, as assumed in section 2.4. For the same parameters, in the right panel of figure 7 the evolution of the comoving number densities of the DM and the ALP are shown, as well as the hidden sector temperature T'. We also plot the equilibrium distributions $Y_{\chi,\mathrm{eq}}(T')$ and $Y_{a,\mathrm{eq}}(T')$ as dotted lines, and, for comparison, $Y_{\chi,\mathrm{eq}}(T)$. As in the simpler thermal freeze-out scenario, Y_a and Y_χ follow the equilibrium distributions, $Y_{a,\mathrm{eq}}(T')$ and $Y_{\chi,\mathrm{eq}}(T')$, respectively. For rapid interactions the right hand side of the Boltzmann equation vanishes and Y_χ stays constant.[8] The ALPs and the DM separate from the equilibrium distributions at the same time, around $T' \simeq m_\chi/10$, since no other particles are around to maintain frequent number changing interactions. Since $T' < T$ the inverse SM freeze-out temperature can be as small as $z \sim 2-3$.

In figure 8 we show the region of parameter space where the correct relic density is obtained for a large range of masses and couplings, again keeping the ratio $m_\chi/m_a = 10$ fixed, together with the constraints derived in the previous section. We have checked that the observations made in the previous paragraph hold for the whole set of parameters depicted in figure 8. The region in the (m_a, g_{aff})-plane is shaped by a number of conditions:

- **Relic density** For connector couplings, g_{aff}, which are too small, the hidden sector does not become sufficiently populated. Although $g_{a\chi\chi}$ might be large enough to establish equilibrium between the hidden sector particles, $n_\chi^{\mathrm{eq}}(T')$ can never reach the amount of DM density observed today. This is indicated by the lower boundary.

- $\boldsymbol{a} \leftrightarrow$ **SM** On the other hand, if the connector coupling g_{aff} is too large, the interactions between the hidden sector and the SM become strong enough to establish thermal equilibrium. Depending on the hidden sector coupling, the DM is then either produced by freeze-in (see section 4.3) or thermal freeze-out. We remark that the numerical solution close to the transition between the freeze-in and the DFO regime is challenging and we chose this upper boundary conservatively.

- $\boldsymbol{g}_{a\chi\chi}$ The DM-mediator interaction is (cf. eq. (2.1))

$$C_\chi \frac{m_\chi}{f_a} a\bar{\chi}\gamma_5\chi \equiv g_{a\chi\chi} m_\chi a\bar{\chi}\gamma_5\chi\,. \tag{4.2}$$

  Our effective theory is valid only below the scale f_a. Thus, the reheating temperature T_{RH} should be below this scale to safely ignore UV contributions. On the other hand, T_{RH} has to be higher than m_χ. We consequently need a hierarchy $f_a \gtrsim T_{RH} \gtrsim m_\chi$, i.e. small $g_{a\chi\chi} = C_\chi/f_a$, and this is the reason why the DM (the ALP for a fixed mass ratio) should not be too heavy.

- **QCD** Finally, we employ a perturbative description of the strong interactions, only convergent at high enough energies. The DM abundance should therefore be set by interactions happening at temperatures before the QCD phase transition. In practice, we set the upper boundary labelled "QCD" by requiring that the χ-particles freeze out at temperatures above the threshold $T_{\mathrm{pert}} = 600$ MeV.

[8] Due to the relation between T and T', affected by the rate of energy transfer to the hidden sector, Y_χ and Y_a ostensibly increase in the plot; this is because the plot shows the evolution of the comoving number densities as a function of the SM temperature T.

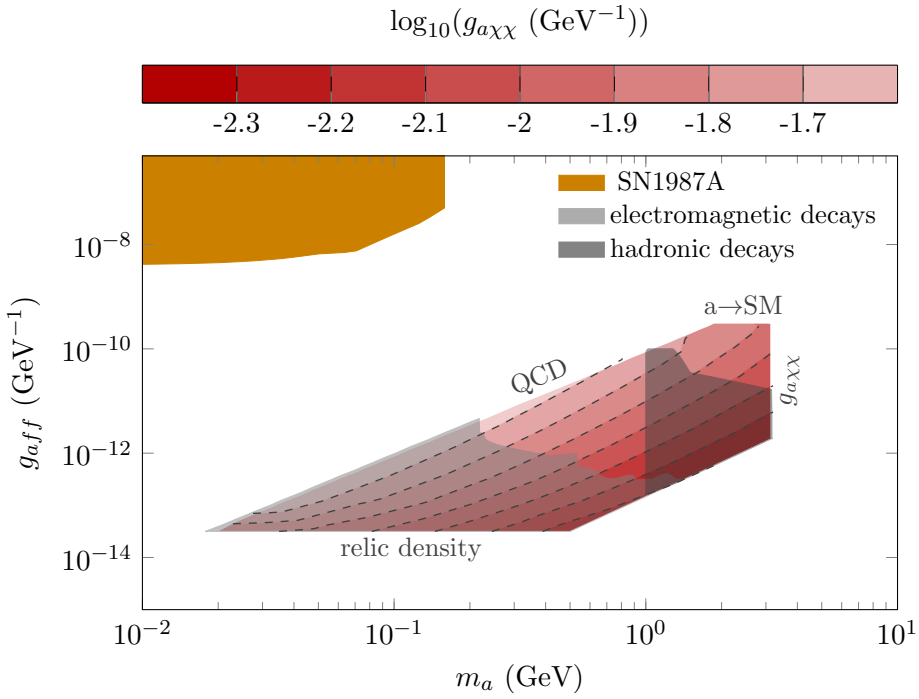


Figure 8: Contour plot of the hidden-sector couplings $g_{a\chi\chi}$ in the (m_a, g_{aff})-plane which give the observed DM relic density via decoupled freeze-out (DFO). We fixed the ratio $m_\chi/m_a = 10$. We have included the relevant constraints on our ALP model on the connector coupling g_{aff} in this parameter region (cf. figure 3). We additionally plot the region excluded by cosmological constraints of additional particles decaying electromagnetically (light grey) and decaying hadronically (dark grey). For explanations about the boundaries of the DFO region and the constraints from electromagnetic and hadronic decays see the main text.

The shape of the contours of constant $g_{a\chi\chi}$ can be understood as follows. By increasing the strength of the ALP-fermion coupling, more energy will be transferred to the hidden sector and heat up the hidden sector particles. This will increase the amount of DM particles, $n_\chi^{\rm eq}(T')$. In order to satisfy the relic density constraint (3.12), the interaction strength among the hidden sector particles has to increase to keep $\chi\bar{\chi} \to aa$ active for longer, such that the χ-particles can continue to annihilate. The DM-ALP interaction is proportional to the product $(g_{a\chi\chi} \cdot m_\chi)^4$. For increasing m_a, $g_{a\chi\chi}$ has to decrease to compensate the effect of increasing m_χ. We find that in this scenario g_{aff} enters only via SM $\to a$ processes because the energy transfer and DM production from the connector process $f\bar{f} \to \chi\bar{\chi}$ is always subdominant. The dependence on g_{aff} is therefore very mild. Furthermore, we observe that $g_{a\chi\chi}$ has to be large compared to the freeze-in scenario to maintain equilibrium among the hidden sector particles.

It turns out that it is extremely challenging to test the DFO region with collider searches for ALPs: in this scenario the visible sector and the mediator barely talk to each other, such

that the sensitivity of experiments probing the relevant mass range, in particular the SN bound and the bounds from rare B and K decays, would have to improve by several orders of magnitude. However, the tiny coupling between the SM particles and the ALP makes the ALP relatively long lived, and since the ALPs are abundantly produced along with the DM particles (cf. figure 7) their decay can have important implications for the cosmological history during and after big bang nucleosynthesis, as briefly discussed in section 3. Of course, the imprint ALPs will leave on cosmological observables strongly depends on its decay products and its lifetime. In the range of ALP masses we consider, the dominant decay channel varies, see figure 9. For ALPs with masses $m_a \lesssim 2m_\mu$, i.e. ALPs which dominantly decay into photons and e^+e^-, the constraints are very similar. Here, we apply the bounds from ref. [58] on very long-lived ALPs in the sub-GeV mass range. Generically speaking, they exclude sufficiently abundant particles decaying electromagnetically with lifetimes $\tau_a \sim 10^3 - 10^5$ s. The bound labelled "electromagnetic decays" in figure 8 was obtained by applying the bounds on the ALP's lifetime from figures 4 and 5 in ref. [58], interpolating between the mass of the decaying particle and the dominant decay channel of the ALP. For masses in the range $2m_\mu \lesssim m_a \lesssim 1$ GeV, the ALP dominantly decays into muons. In principle, the applicable constraints are the electromagnetic ones here – see also the discussion in ref. [54] – and the lifetime is short enough for them not to matter. For hadronic decays the bounds become more severe, cf. section 3, and lifetimes above $\tau \sim 0.1$ s can be excluded. In ref. [61] bounds are provided on hadronically decaying particles with masses in the GeV-TeV range. The smallest mass for which results are available is 30 GeV. We apply the corresponding bound to our model, extrapolating from the given shapes that the bounds will remain approximately constant for lower masses. We have checked that, taking into account the branching ratios of the various decay channels, the energy injection from ALPs is sufficient in the DFO region for these constraints to apply. However, as outlined in section 3, photo-dissociation is clearly not the only possible scenario. For instance, ALPs decaying into photons can re-equilibrate with the SM, a scenario which was studied in depth in ref. [55], excluding much shorter lifetimes. However, as studied in ref. [59], since the temperature of our hidden sector T' is in general well below T, we expect these bounds to be alleviated in our case. In order to obtain precise cosmological bounds a detailed analysis of the evolution of the ALP number density and its imprint on cosmological observables would hence be necessary, yet beyond the scope of this paper.

We conclude this section by remarking that, as mentioned above, the upper left boundary labelled "QCD" arises due to a theoretical limitation. It might be interesting to describe the processes involved at low energies using non-perturbative techniques in order to explore the remaining parameter space. In doing so, we should be able to close the gap between the "QCD" line and ALP-SM equilibrium condition in figure 5.

5 Conclusion

To conclude, we study a fermionic DM model with a pseudoscalar (ALP) mediator coupled to SM and DM fermions at tree level. In this model, different production mechanisms can set the relic abundance, depending on which interactions are most important. In fact,

we find that, similar to the case of a fermionic DM model with a massive dark photon mediator studied in ref. [7], six different production mechanisms are possible for the set of ALP and DM masses we have considered, namely freeze-in from SM particles, freeze-in from mediators, sequential freeze-in, decoupled freeze-out (DFO), freeze-out from mediators and freeze-out from the SM. We numerically solve the general set of coupled Boltzmann equations for a range of ALP-fermion and ALP-DM couplings to obtain the phase diagram for a fixed ALP and DM mass ($m_a = 1$ GeV, $m_\chi = 10$ GeV). We focus on two cases where the hidden sector, comprised of the DM and ALP, does not thermalise with the visible sector: freeze-in (from the SM or from ALPs) and DFO.

In the freeze-in and DFO regimes, we determine the relic density by solving the relevant Boltzmann equations. This is most challenging in the DFO region, as discussed in sec. 2.4, where a set of three coupled differential equations must be solved (the Boltzmann equations for the ALP and DM number densities and the Boltzmann equation governing the energy transfer from the SM to the dark sector), and a variety of processes contribute to ALP and dark matter production. We derive for the first time the energy transfer Boltzmann equation for the case where there is a single hidden sector particle in the final state, and present a new strategy for determining the hidden sector temperature from the evolution of the energy density. Throughout our work we take into account thermal effects by implementing temperature-dependent corrections to the SM masses and couplings, as these can alter the results at the $\mathcal{O}(1)$ level, in particular in the case of the production of ALPs. The dominant contribution stems from the increased masses of the in medium SM particles.

In a model where the ALP-fermion couplings arise from an extended Higgs sector in the UV, both infrared-dominated and ultraviolet-dominated processes contribute to freeze-in, and either of them may dominate depending on the hierarchy of the reheating temperature and the electroweak scale. In particular, we find that UV-dominated $2 \to 3$ interactions play a role for the production via freeze-in from the SM for reheating temperatures $T_{RH} \gtrsim 200$ GeV. We therefore choose to study one typical infrared-dominated ($T_{RH} = 200$ GeV) and one ultraviolet-dominated scenario ($T_{RH} = 2$ TeV). In the DFO region all ultraviolet-dominated processes enter only at the subleading level via the energy transfer to the hidden sector, such that the final DM abundance is largely insensitive to the ultraviolet contributions for reheating temperatures below a few TeV, and these are therefore neglected.

We also pay particular attention to calculating the experimental bounds on the relevant parameter space, including collider, astrophysics, cosmology and DM constraints. Some of the most powerful constraints come from collider experiments, particularly beam dump constraints coming from SLAC E137 and flavour constraints notably from NA62 and LHCb, together constraining masses up to the B meson mass and the g_{aff} coupling down to $\mathcal{O}(10^{-6}\ \text{GeV}^{-1})$. For the beam dump experiment we improve previous analyses by including all dominant ALP production processes, namely the Primakoff process, bremsstrahlung and (non-)resonant positron annihilation. In particular, we show that positron annihilation can become the dominant process for larger ALP masses in models where the ALP coupling to photons is highly suppressed and in this case clearly cannot be neglected. For the scenario considered, where the only tree-level couplings of the ALP are to fermions, we

added the different production scenarios since they cannot be distinguished by experiment. Our analysis excludes a large part of the parameter space in the (m_a, g_{aff})-plane and puts upper and lower bounds on the coupling. For the flavour constraints, we implement the state-of-the-art bounds for the most constraining channels from NA62, NA48/2, LHCb and CHARM, which together rule out a large section of parameter space, down to $g_{aff} \sim \mathcal{O}(10^{-5}$ GeV$^{-1})$ for masses up to $m_a \sim m_{K^+} - m_{\pi^+}$, and in the range $g_{aff} \sim \mathcal{O}(10^{-4}$ GeV$^{-1})$ to $\mathcal{O}(10^{-2}$ GeV$^{-1})$ up to $m_a \sim m_B - m_{K^{(*)}}$.

Out of all the constraints considered, astrophysics provides the most stringent bounds on our parameter space. In particular SN1987A and HB stars strongly probe the parameter space for $m_a \lesssim \mathcal{O}(0.1$ GeV$)$, extending down to $g_{aff} \sim \mathcal{O}(10^{-9}$ GeV$^{-1})$. For the former constraint, we adapt the state-of-the-art analysis [45, 46] to the scenario we consider, including N^{3}LO corrections in chiral perturbation theory. While we do consider DM direct detection, indirect detection and self-interactions, due to the pseudoscalar nature of the mediator, as expected, none of these provide relevant constraints.

The tiny coupling between the SM and the mediator which is required to reproduce the observed DM relic density in the DFO region is responsible for making the new sector so difficult to probe. It therefore seems that current and near future experiments cannot access this region of parameter space, yet cosmology can provide a powerful tool to constrain secluded dark sectors. Here we carry out an estimate of the consequences that the presence of ALPs would have on cosmology for the DFO region: Applying the constraints from refs. [58] and [61], we exclude very long lived ($\tau_a \gtrsim 10^3 - 10^5$ s) electromagnetically decaying ALPs and short lived ($\tau_a \gtrsim 0.1$ s) hadronically decaying ALPs as their decay products would photodissociate light elements.

In sections 4.3 and 4.4 we present the regions of parameter space in the (m_a, g_{aff})-plane where the correct relic density is obtained, with contour lines showing the required value of $g_{a\chi\chi}$. For freeze-in, as seen in figure 5, we find a large unexcluded region, extending to $g_{aff} \sim 10^{-9}$ GeV^{-1} for 10^{-4} GeV $< m_a < 10^2$ GeV. For larger values of g_{aff} and m_a, smaller values of $g_{a\chi\chi}$ are required. Probing this region, for the largest values of $g_{a\chi\chi}$, would require the flavour constraints to improve substantially, or low-mass resonance searches at ATLAS or CMS. For the DFO regime, as seen in figure 8, the correct density is obtained for relatively low values of $g_{aff} \sim 10^{-13} - 10^{-10}$ GeV^{-1} and large values of $g_{a\chi\chi}$ ($\sim 10^{-2}$ GeV^{-1}). Again it appears that cosmology offers the only means of probing such small values of g_{aff}. Moreover, it turns out the DFO region is disfavoured by the standard BBN scenario.

There are several possible improvements to our analysis that could be envisaged. As discussed in Sec. 2.3.3, a thorough study of the sequential freeze-in region requires solving the unintegrated Boltzmann equations. It would also be interesting to compare our results in the DFO region to those obtained using the unintegrated Boltzmann equations. Furthermore, it might be worthwhile using non-perturbative methods to investigate the DFO mechanism in the region above the QCD boundary. In this work we focus on a Dirac fermion dark matter particle, but, as previously mentioned, a Majorana fermion is an equally well-motivated possibility which could merit a detailed study, although one should expect qualitatively similar results. Finally, the potential sensitivity of future experiments

is of utmost interest, and in this view it should be worthwhile to assess the sensitivity of planned future flavour experiments and upgrades, beam dump experiments and cosmology measurements.

Acknowledgments

The authors thank Rupert Coy, Sacha Davidson, Fatih Ertas, Rouven Essig, Thomas Hambye, Felix Kahlhöfer, Sam McDermott, Vivian Poulin, Christopher Smith, Michel Tytgat and Wei Xue for useful discussions and correspondence. Further we thank Babette Döbrich for providing the CHARM constraint in numeric form, and Florian Domingo for sharing with us his code for the decay width of pseudoscalars in the NMSSM. The project leading to this publication has received funding from Excellence Initiative of Aix-Marseille University – A*MIDEX, a French "Investissements d'Avenir" programme (AMX-19-IET-008 - IPhU and ANR-11-IDEX-0001-02), and from the OCEVU Labex (ANR-11-LABX-0060).

A Effective ALP-DM Lagrangian from a two Higgs doublet model

The effective Lagrangian eq. (2.1) can be UV completed by a renormalizable model containing additional Higgs doublets and singlets, with the flavour-diagonal couplings of the ALP being a consequence of the Higgs coupling structure. Here we present a simple example for pedagogical purposes.

Our starting point is the Lagrangian of a type-I two Higgs doublet model with an extra scalar singlet,

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\text{kin}} &- \left(y^d_{ij}\bar{Q}_{Li}\Phi_2 D_{Rj} + y^e_{ij}\bar{L}_{Li}\Phi_2 E_R + y^u_{ij}\bar{Q}_{Li}\widetilde{\Phi}_2 U_{Rj} + \text{ h.c.}\right) - \tilde{m}_1^2\,|\Phi_1|^2 \\ &- \tilde{m}_2^2\,|\Phi_2|^2 + M^2\,|\phi|^2 - \frac{\tilde{\lambda}_1}{2}|\Phi_1|^4 - \frac{\tilde{\lambda}_2}{2}|\Phi_2|^4 - \frac{\lambda_\phi}{2}|\phi|^4 - \tilde{\lambda}_3\,|\Phi_1|^2\,|\Phi_2|^2 \\ &- \tilde{\lambda}_4\,(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) - \kappa_1\,|\Phi_1|^2\,|\phi|^2 - \kappa_2\,|\Phi_2|^2|\phi|^2 - \tilde{\lambda}_{12}\left(\Phi_1^\dagger\Phi_2\phi^2 + \text{ h.c.}\right)\,. \end{aligned} \tag{A.1}$$

Here $\mathcal{L}_{\text{kin}}$ contains all the gauge-kinetic terms for the SM gauge and fermion fields as well as for the scalars Φ_1, Φ_2 and ϕ. Φ_1 and Φ_2 are two Higgs doublets with the gauge quantum numbers of the SM Higgs, and ϕ is a complex scalar which is neutral under the SM. This is the most general renormalizable Lagrangian allowed by a $\text{U}(1)_{\text{PQ}}$ global symmetry under which Φ_2 and the SM fermions are neutral, while Φ_1 and ϕ are charged. Note that this symmetry is not a PQ symmetry in the stricter sense of being anomalous with respect to QCD, and therefore unsuitable for solving the strong CP problem.

For $M^2 > 0$, ϕ will take a VEV $\langle\phi\rangle = f_a/\sqrt{2}$, which can be taken real and positive. We define real scalar fields σ and a by

$$\phi(x) = \frac{1}{\sqrt{2}}\,(f_a + \sigma(x))\,e^{i\frac{a(x)}{f_a}}\,. \tag{A.2}$$

We anticipate that Φ_1 and Φ_2 will also take vacuum expectation values of the order $v \ll f_a$. This parameterization is useful to extract leading-order effects in v/f_a; to leading order we

have $f_a = \sqrt{\frac{2}{\lambda_\phi}} M$. The fields σ and a are approximate mass eigenstates, up to mixing with the Φ-like fields with mixing angles of order v/f_a.

At low energies, σ (whose VEV is zero to leading order) can be integrated out. The effective $\Phi_{1,2}$ mass parameters and quartic couplings at low energies will be modified according to suitable matching conditions from replacing $|\phi|^2 \to f_a^2/2$ in the above Lagrangian and from four-point interactions with σ exchange, such that $\tilde{m}_i^2 \to m_i^2$ and $\tilde{\lambda}_k \to \lambda_k$. Up to higher-dimensional operators, the effective Lagrangian is now that of a type-I two-Higgs doublet model with an additional field $a(x)$,

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\rm kin} &- \Big(y_{ij}^d \bar{Q}_{Li} \Phi_2 D_{Rj} + y_{ij}^e \bar{L}_{Li} \Phi_2 E_R + y_{ij}^u \bar{Q}_{Li} \widetilde{\Phi}_2 U_{Rj} + \text{ h.c.} \Big) - m_1^2 \, |\Phi_1|^2 \\ &- m_2^2 \, |\Phi_2|^2 - \frac{\lambda_1}{2} |\Phi_1|^4 - \frac{\lambda_2}{2} |\Phi_2|^4 - \lambda_3 \, |\Phi_1|^2 \, |\Phi_2|^2 \\ &- \lambda_4 \, (\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_1) - \frac{\lambda_{12} \, f_a^2}{2} \left(\Phi_1^\dagger \Phi_2 \, e^{2ia/f_a} + \text{ h.c.} \right) . \end{aligned} \tag{A.3}$$

With the a-dependent field redefinition $\Phi_1(x) \to e^{2ia(x)/f_a} \Phi_1(x)$, a disappears from the potential. However, the Φ_1 kinetic term is not invariant,

$$(D_\mu \Phi_1)^\dagger \, D^\mu \Phi_1 \to (D_\mu \Phi_1)^\dagger \, D^\mu \Phi_1 - 2i \frac{\partial_\mu a}{f_a} \, \Phi_1^\dagger \overleftrightarrow{D}^\mu \Phi_1 + \mathcal{O}\left(\frac{a^2}{f_a^2} \right) .$$

The "Higgs basis" in the space of Φ_1 and Φ_2 is defined by rotating $(\Phi_1, \Phi_2) \to (H, \Phi)$ by an angle β such that the Higgs VEV is contained in one doublet H only, $\tan\beta = \langle \Phi_2 \rangle / \langle \Phi_1 \rangle$. In general, β is independent of the mixing angle α which parameterizes the mixing between the mass eigenstates. However, in the decoupling limit, these two angles are aligned and the fields H and Φ contain the mass eigenstates as

$$H(x) = \begin{pmatrix} G^+(x) \\ \frac{v + h(x) + iG^0(x)}{\sqrt{2}} \end{pmatrix} , \qquad \Phi(x) = \begin{pmatrix} H^+(x) \\ \frac{H^0(x) + iA^0(x)}{\sqrt{2}} \end{pmatrix} . \tag{A.4}$$

We assume that the Φ-like states are also heavy, so they can be integrated out as well. (This step can in principle be interchanged with integrating out σ or performed simultaneously, depending on the mass hierarchies.) The effective Lagrangian becomes

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\rm kin} &- \Big(Y_{ij}^d \bar{Q}_{Li} H D_{Rj} + Y_{ij}^e \bar{L}_{Li} H E_{Rj} + Y_{ij}^u \bar{Q}_{Li} \widetilde{H} U_{Rj} + \text{ h.c.} \Big) \\ &+ m^2 \, |H|^2 - \frac{\lambda}{2} |H|^4 - 2i \frac{\partial_\mu a}{f_a} \, \cos^2\beta \, H^\dagger \overleftrightarrow{D}^\mu H \end{aligned} \tag{A.5}$$

where $Y^{u,d,e} = y^{u,d,e} \sin\beta$. The $(\partial a) \, H^\dagger \overleftrightarrow{D} H$ term would cause a to mix with the Z boson after electroweak symmetry breaking, which is avoided by another field redefinition:

$$H(x) \to H(x) e^{-2i \cos^2\beta \, a(x)/f_a} . \tag{A.6}$$

The Yukawa terms will shift accordingly, and one obtains for the final axion couplings in this field basis at leading order in $1/f_a$

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\rm kin} &+ m^2 \, |H|^2 - \frac{\lambda}{2} |H|^4 - \Big(Y_{ij}^d \, \bar{Q}_{Li} H D_{Rj} + Y_{ij}^e \, \bar{L}_{Li} H E_{Rj} + Y_{ij}^u \, \bar{Q}_{Li} \widetilde{H} U_{Rj} + \text{ h.c.} \Big) \\ &- 2\cos^2\beta \frac{a}{f_a} \Big(i \, Y_{ij}^d \, \bar{Q}_{Li} H D_{Rj} + i \, Y_{ij}^e \, \bar{L}_{Li} H E_{Rj} + i \, Y_{ij}^u \, \bar{Q}_{Li} \widetilde{H} U_{Rj} + \text{ h.c.} \Big) . \end{aligned} \tag{A.7}$$

The second line corresponds to the ALP-fermion couplings in eq. (2.1) with $C_f = \cos^2\beta$, after switching to four-spinor notation.

Other models where an ALP is obtained by more general extensions of the Higgs sector will induce different axion-fermion couplings. For example, choosing a type-II two-Higgs doublet model (as in the standard DFSZ model for QCD axions [81]) would yield different couplings for up-type and down-type quarks and leptons, unless $\tan\beta = 1$.

It is possible to transform the dimension-5 couplings in eq. (A.7) into $\partial_\mu a\, j^\mu_{\rm PQ}$ terms, with $aF\widetilde{F}$ couplings appearing due to the anomaly. These are usually eliminated again after electroweak symmetry breaking, using the equations of motion to transform the $\partial_\mu a\, j^\mu_{\rm PQ}$ terms into the Yukawa-like couplings of eq. (2.1), which involves precisely the same anomaly term. We emphasize that additional $aF\widetilde{F}$ terms are never induced at high scales, unless there are heavy fermions with charges under both PQ and the Standard Model (such as in the KSVZ model for a QCD axion).

To obtain our dark matter model, we now add two more features. The first is a mass term for a which explicitly breaks the PQ symmetry; we assume $m_a \ll f_a$ and write

$$\Delta\mathcal{L}_{\rm mass} = \frac{m_a^2}{4}\left(\phi - \phi^*\right)^2 \,. \tag{A.8}$$

The second is the dark matter candidate χ, for which there are several possibilities. In this paper we focus on a Dirac dark matter candidate $\chi = \chi_L + \chi_R$ with PQ charges allowing for a coupling

$$\Delta\mathcal{L}_\chi = -\left(y_\chi\, \phi \overline{\chi}_L \chi_R + \text{h.c.}\right) \tag{A.9}$$

which, after PQ breaking, gives rise to a Dirac mass $m_\chi = y_\chi f_a/\sqrt{2}$ and an axionic coupling:

$$\Delta\mathcal{L}_{\rm eff} = -m_\chi\, \overline{\chi}\chi - i\,\frac{m_\chi}{f_a}\, a\, \overline{\chi}\gamma^5\chi \,. \tag{A.10}$$

In this simple model, the reduced ALP-fermion couplings g_{aff} tend to be of the same order of magnitude as the reduced ALP-DM couplings $g_{a\chi\chi}$, unless $\cos\beta \ll 1$. As we show in the main text, hierarchically different couplings can lead to interesting effects. They are well-motivated in UV models with multiple axions and a clockwork-like structure for the couplings, leading to exponentially different effective decay constants for different sectors of the theory [82]. One might speculate, for instance, that the DM sector and the SM are localized on different lattice sites in theory space, giving rise to a large hierarchy $g_{aff} \ll g_{a\chi\chi}$.

Moreover, this model predicts $C_\chi = 1$ and therefore $g_{a\chi\chi} = \frac{1}{f_a}$. The DFO phase as discussed in sections 2.4 and 4.4 will therefore not be realized in this simple scenario, since DFO requires significant connector couplings $g_{a\chi\chi} \sim (100\text{ GeV})^{-1}$. A mass scale $f_a \sim 100$ GeV is of course ruled out by observation. We emphasize nevertheless that there is no obstacle in principle to raising C_χ, although doing so with the particle content of the present minimal model would require fine-tuning (by adding an additional explicitly PQ breaking term, namely a bare Dirac mass term for χ which could partially cancel the Dirac mass from spontaneous PQ breaking).

We finally remark that another possibility for the dark matter candidate would be a single Weyl dark matter candidate $\chi = \chi_L$, coupled to ϕ according to

$$\Delta\mathcal{L}_\chi = -\frac{1}{2}\left(y_\chi\, \phi \chi_L^T C \chi_L + \text{h.c.}\right) . \tag{A.11}$$

This coupling translates into a Majorana mass after PQ breaking. A third possibility would be a model with two Weyl fermions and a see-saw-like structure where both a Dirac mass and a Majorana mass are allowed, with the dark matter phenomenology depending on the hierarchy between the two.

B Collision term for the energy transfer Boltzmann equation

For a process $1\,2 \to 3\,4$, neglecting the back-reaction, the integrated collision operator is given in eq. (2.24) in terms of the energy transfer rate $\mathcal{E}(\vec{p}_1, \vec{p}_2)$ for an initial state with momenta $\vec{p}_1$ and $\vec{p}_2$, and of the Møller velocity

$$v_{\text{Møl}} = \frac{\sqrt{(s-(m_1+m_2)^2)(s-(m_1-m_2)^2)}}{2E_1E_2} \equiv \frac{F}{E_1E_2}. \tag{B.1}$$

Assuming the momentum distributions to be Maxwell-Boltzmann-like, the collision term becomes

$$\begin{aligned}\int \frac{d^3p}{(2\pi)^3} C[f] &= g_1 g_2 \int \frac{d^3p_1}{(2\pi)^3}\frac{d^3p_2}{(2\pi)^3} e^{-(E_1+E_2)/T} v_{\text{Møl}}\, \mathcal{E}(\vec{p}_1, \vec{p}_2) \\ &= \frac{g_1 g_2}{32\pi^4}\int dE_+ dE_- ds\, e^{-E_+/T} F\, \mathcal{E}(\vec{p}_1, \vec{p}_2)\end{aligned} \tag{B.2}$$

where, following ref. [14], in the second line we have replaced $E_+ = E_1 + E_2$, $E_- = E_1 - E_2$ and $s = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2$. The energy transfer rate $\mathcal{E}(\vec{p}_1, \vec{p}_2)$ can be written as in eq. (2.25). For the transferred energy ΔE_{tr}, there are two cases which we need to distinguish: For $f\bar{f} \to \chi\bar{\chi}$, $\Delta E_{tr} = E_+$, and for $f\bar{f} \to a\gamma$, $\gamma f \to af$, $q\bar{q} \to ag$, $gq \to aq$ (and the corresponding hermitian conjugate processes) we have $\Delta E_{tr} = E_3$. It turns out that in the former case, as studied in ref. [6], the identical particles in the initial state and ΔE_{tr} being independent of the final state result in several simplifications in $\mathcal{E}(\vec{p}_1, \vec{p}_2)$, absent for the latter.

In order to demonstrate the difference between these two cases let us now perform the integration in p_4 (where in the following discussion we will use the notation $p_i = |\vec{p}_i|$):

$$\mathcal{E}(\vec{p}_1, \vec{p}_2) = \frac{1}{4F}\int \frac{d^3p_3}{(2\pi)^3}\frac{1}{2E_3 2E_4}|i\mathcal{M}|^2 (2\pi)\delta(E_1+E_2-E_3-E_4)\Delta E_{tr}$$

where $E_4 = (m_4^2 + |\vec{p}_1 + \vec{p}_2 - \vec{p}_3|^2)^{1/2}$. The next steps depend on the value of ΔE_{tr}. Clearly if $\Delta E_{\text{tr}} = E_+$, it does not affect the integration over p_3 and Ω, resulting in the simple relation $\mathcal{E}(\vec{p}_1, \vec{p}_2) = \sigma(\vec{p}_1, \vec{p}_2)E_+$. However, when $\Delta E_{\text{tr}} = E_3$ integrating in p_3 and Ω is not so straightforward. For the latter case, let us consider the argument of the delta function which we will call $g(p_3) \equiv E_1 + E_2 - E_3 - E_4$:

$$g(p_3) = E_+ - E_3 - (m_4^2 + E_+^2 - s + p_3^2 - 2p_1p_3\, c_{13} - 2p_2p_3\, c_{23})^{1/2}$$

where $c_{ij} = \cos\theta_{ij}$, and we have defined θ_{ij} to be the angle between the momenta of particles i and j. Since we perform the integration in c_{13}, we need to replace θ_{23} by $\theta_{13} + \theta_{12}$, where $c_{12} = (E_+^2 - s - p_1^2 - p_2^2)/(2p_1p_2)$. In order to obtain the solution p_3^0 of the equation $g(p_3) = 0$ analytically, and to speed up the computation, we are obliged to make the assumption $E_3 = p_3$. This assumption is easily justifiable for the dominantly contributing case $f = t$ (and results in an effect of at most $\sim 1.4\%$), and we have checked that $q\bar{q} \to ga$ and $f\bar{f} \to \gamma a$ dominate over $f\gamma \to fa$ and $qg \to qa$. We then find

$$\mathcal{E}(\vec{p}_1, \vec{p}_2) = \frac{1}{8\pi F} \int p_3^2 dp_3 dc_{13} \frac{1}{4E_4} |i\mathcal{M}|^2 \frac{\delta(p_3 - p_3^0)}{|g'(p_3)|_{p_3 \to p_3^0}} \tag{B.3}$$

where we have replaced $\Delta E_{\rm tr}$ by E_3. The result for the integrated collision operator is then obtained by inserting eq. (B.3) in eq. (B.2) and performing the integration numerically, setting $E_3 = p_3$.

For the case of $f\bar{f} \to \chi\bar{\chi}$, and $\mathcal{E}(\vec{p}_1, \vec{p}_2) = \sigma(\vec{p}_1, \vec{p}_2)E_+$, the integrated collision operator can easily be simplified to [6]:

$$\int \frac{d^3p_3}{(2\pi)^3} C[f_3] = \frac{g_1 g_2}{32\pi^4} \int ds\, \sigma(s)\, (s - 4m^2)\, s\, T K_2\left(\frac{\sqrt{s}}{T}\right) \tag{B.4}$$

where we have adopted the limits of integration [6, 14]: $s > (m_1 + m_s)^2$, $E_+ > \sqrt{s}$ and $-\frac{2F}{s}\sqrt{E_+^2 - s} < E_- + E_+ \frac{m_1^2 - m_2^2}{s} < \frac{2F}{s}\sqrt{E_+^2 - s}$.

The case of inverse decays, $1\,2 \to X$, is not identical, and deserves to be studied as it has not previously been required, starting with

$$\begin{aligned}\mathcal{E}(\vec{p}_1, \vec{p}_2) =& \frac{1}{2E_1 2E_2 v_{Mol}} \int \frac{d^3p_X}{(2\pi)^3} \frac{1}{2E_X} |i\mathcal{M}|^2 (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_X) \Delta E_{tr} \\ =& \frac{1}{2E_1 2E_2 v_{Mol}} \frac{1}{2E_X} |i\mathcal{M}|^2 (2\pi) \delta(E_1 + E_2 - E_X) E_+ \,.\end{aligned} \tag{B.5}$$

We can then insert this expression in eq. (B.2), and find for $m_1 = m_2 = m$

$$\begin{aligned}\int \frac{d^3p_3}{(2\pi)^3} C[f_3] =& \frac{g_1 g_2}{32\pi^3} \int dE_+\, e^{-E_+/T}\, |i\mathcal{M}|^2\, \frac{2F\, E_+}{m_X^2} \sqrt{E_+^2 - m_X^2}\, \Theta(m_X^2 - 4m^2) \\ =& T \frac{g_1 g_2}{32\pi^3} m_X \sqrt{m_X^2 - 4m^2}\, |i\mathcal{M}|^2\, K_2\left(\frac{m_X}{T}\right) \Theta(m_X^2 - 4m^2)\end{aligned} \tag{B.6}$$

which in terms of the decay width of particle X can be expressed as

$$\int \frac{d^3p_3}{(2\pi)^3} C[f_3] = T\, \frac{g_1 g_2}{2\pi^2}\, \Gamma_X\, m_X^3\, K_2\left(\frac{m_X}{T}\right) \Theta(m_X^2 - 4m^2)\,.$$

C Cross sections and input for Boltzmann equations

In table 2 we collect the different annihilation processes for the DM and the ALPs which govern the Boltzmann equations in eq. (2.6). In this section we provide expressions for the cross sections for the required $2 \to 2$ processes and the collision term for the relevant $2 \to 3$ processes.

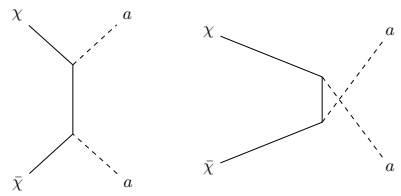
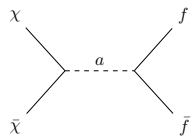
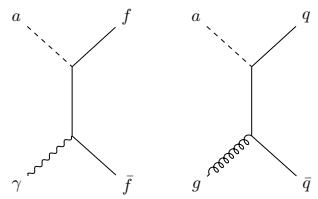
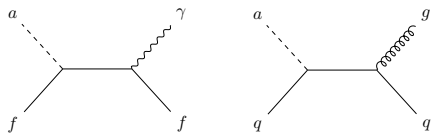
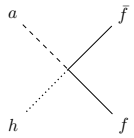
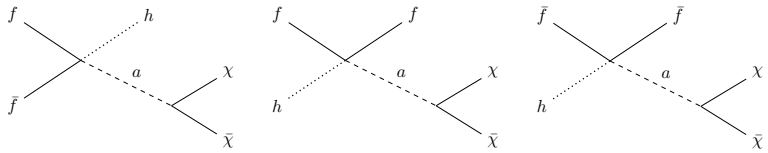
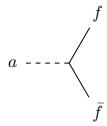
Process	Feynman-diagram
$\chi\bar{\chi} \to aa$ eq. (C.1)	
$\chi\bar{\chi} \to f\bar{f}$ eq. (C.2)	
$a\gamma \to f\bar{f}$, $ag \to q\bar{q}$ eqs. (C.4), (C.5)	
$af \to \gamma f$, $aq \to gq$ eqs. (C.6), (C.7)	
$ah \to f\bar{f}$ eq. (C.8)	
$f\bar{f} \to h\chi\bar{\chi}$, $fh \to f\chi\bar{\chi}$, $\bar{f}h \to \bar{f}\chi\bar{\chi}$ eq. (C.3)	
$a \to f\bar{f}$ eq. (D.1)	

Table 2: Feynman-diagrams of the annihilation of the DM via the hidden sector and connector processes, as well as the co-annihilation of ALPs at tree level.

C.1 Hidden sector process

For the annihilation of the DM particles to ALPs, i.e. annihilation in the hidden sector (for which the Feynman diagrams at tree level are shown in table 2), the total unpolarized cross section in the centre of mass frame is given by

$$\sigma(\chi\bar{\chi}\to aa) = \frac{(g_{a\chi\chi}m_\chi)^4}{16\pi s}\left(\frac{6m_a^4-4m_a^2s+s^2}{(s-2m_a^2)\sqrt{(s-4m_a^2)(s-4m_\chi^2)}}\ln\frac{s-2m_a^2+\sqrt{(s-4m_a^2)(s-4m_\chi^2)}}{s-2m_a^2-\sqrt{(s-4m_a^2)(s-4m_\chi^2)}}\right.$$
$$\left.-2-\frac{m_a^4}{m_a^4-4m_a^2m_\chi^2+m_\chi^2 s}\right) \tag{C.1}$$

C.2 Connector processes

The Feynman diagram of the $\chi\bar{\chi}\to f\bar{f}$ connector process is depicted in table 2. For the total cross section in the centre of mass frame we find at tree level

$$\sigma(\chi\bar{\chi}\to f\bar{f}) = \frac{(g_{a\chi\chi}m_\chi)^2(g_{aff}m_f)^2 n_f^c s\sqrt{s-4m_f^2}}{16\pi\left(m_a^2-s\right)^2\sqrt{s-4m_\chi^2}}\,, \tag{C.2}$$

with n_f^c the number of colour degrees of freedom of the fermion. For reheating temperatures above a few hundred GeV contributions from the ultraviolet-dominant $2\to 3$ processes $f\bar{f}\to h\chi\bar{\chi}$, $fh\to f\chi\bar{\chi}$ and $\bar{f}h\to\bar{f}\chi\bar{\chi}$ become important. Following ref. [11], but calculating the matrix element explicitly and including the final state masses one finds (the fact that not all involved particles are scalars does not change the integral because we neglect Fermi/Bose factors)

$$\dot{n}_\chi+3Hn_\chi = \frac{(g_{a\chi\chi}m_\chi)^2(g_{aff}m_f)^2}{v^2}\frac{n_f^c\,T}{(4\pi)^7}\times$$
$$\int_{s_{\min}}^{\infty}\mathrm{d}s\int_0^1\mathrm{d}x_2\int_{x_2}^1\mathrm{d}x_1\,s^{3/2}\,K_1\left(\frac{\sqrt{s}}{T}\right)\frac{4p_a^2 s}{(p_a^2-m_a^2)^2} \tag{C.3}$$

with $p_a^2=(x_1-x_2)s+m_h^2$ and $s_{\min}=4m_f^2(T)$ for $f\bar{f}\to h\chi\bar{\chi}$ and $p_a^2=(x_1-x_2)s+m_f^2$ and $s_{\min}=(m_f(T)+m_h)^2$ for $fh\to f\chi\bar{\chi}$, $\bar{f}h\to\bar{f}\chi\bar{\chi}$.

C.3 SM-ALP processes

For the co-annihilation of ALPs to SM particles, the Feynman diagrams at tree level are shown in table 2. Here we collect the results for the total unpolarized cross section in the

centre of mass frame:

$$\sigma(a\gamma \to f\bar{f}) = \frac{(g_{aff}m_f)^2\ \alpha_{\rm em}\ q_f^2\ n_f^c}{m_a^2+m_\gamma^2-s}\sqrt{\frac{\lambda(s,m_f^2,m_f^2)}{\lambda(s,m_a^2,m_\gamma^2)}}$$
$$\left[\frac{m_a^4-4m_a^2m_f^2+\left(m_\gamma^2-s\right)^2}{\sqrt{\lambda(s,m_a^2,m_\gamma^2)(s(s-4m_f^2)}}\ln\left(\frac{\sqrt{s}(m_a^2+m_\gamma^2-s)+\sqrt{\lambda(s,m_a^2,m_\gamma^2)}\sqrt{s-4m_f^2}}{\sqrt{s}(m_a^2+m_\gamma^2-s)-\sqrt{\lambda(s,m_a^2,m_\gamma^2)}\sqrt{s-4m_f^2}}\right)\right.$$
$$\left.-\frac{m_a^2\left(m_\gamma^2+2m_f^2\right)\left(m_a^2+m_\gamma^2-s\right)}{(m_a^2-m_\gamma^2)^2m_f^2+m_a^2m_\gamma^2s-2(m_a^2+m_\gamma^2)m_f^2s+m_f^2s^2}\right], \tag{C.4}$$

$$\sigma(ag \to q\bar{q}) = \frac{1}{6}\ \sigma(a\gamma \to f\bar{f})(\alpha_{\rm em}\to\alpha_s,\ q_f^2\to 1,\ m_\gamma\to m_g)\ , \tag{C.5}$$

$$\sigma(af \to \gamma f) = \frac{(g_{aff}m_f)^2\ \alpha_{\rm em}\ q_f^2}{4s}\sqrt{\frac{\lambda(s,m_f^2,m_\gamma^2)}{\lambda(s,m_f^2,m_a^2)}}\int_{-1}^{1}\mathrm{d}\cos(\theta)\frac{1}{\left(s-m_f^2\right)^2\left(t-m_f^2\right)^2}$$
$$\Big(m_a^2\left(m_\gamma^2\left(2m_f^2(s+t)-2m_f^4-s^2-t^2\right)-2\left(s+t-2m_f^2\right)\left(s\,t-m_f^4\right)\right)$$
$$+\left(s-m_f^2\right)\left(t-m_f^2\right)\left((s+t-2m_f^2)^2+2m_a^4\right)\Big)\ , \tag{C.6}$$

$$\sigma(aq \to gq) = \frac{4}{3}\ \sigma(af \to \gamma f)(\alpha_{\rm em}\to\alpha_s,\ q_f^2\to 1,\ m_\gamma\to m_g)\ , \tag{C.7}$$

with n_f^c the number of colour degrees of freedom of the fermion and q_f the fermion's electric charge. Note that expressions (C.6) and (C.7) are equivalent to the ones for $\bar{f}a \to \bar{f}\gamma$ and $\bar{q}a \to \bar{q}g$.

For reheating temperatures above a few hundred GeV contributions from the ultraviolet-dominant process $ah \to f\bar{f}$ become important. In this case we find for the cross section in the centre of mass frame

$$\sigma(ah \to f\bar{f}) = \frac{(g_{aff}m_f)^2 n_f^c}{8\pi v^2}\sqrt{\frac{\lambda(s,m_f^2,mf^2)}{\lambda(s,m_a^2,m_h^2)}}\ . \tag{C.8}$$

D Decay width of the ALP

The total decay width of the ALP plays an important role at various points in our work, particularly for the collider and cosmological constraints. For example, for the collider constraints it is important because the ALP lifetime determines whether it can be triggered on, or if it decays within the detector or outside (leading to invisible decay). At tree level, the ALP decays into leptons or quarks, however at loop level it may decay into photons or gluons. Our calculations for the leading partial decay widths are summarised below. In the perturbative regime the decay width of an ALP decaying to fermions is

$$\Gamma(a \to f\bar{f}) = \frac{(g_{aff}m_f)^2 n_f^c m_a}{8\pi}\sqrt{1-\frac{4m_f^2}{m_a^2}}\ , \tag{D.1}$$

with n_f^c the number of colour degrees of freedom of the fermion. This expression is used for leptons and for decays into heavy quarks $c\bar{c}$ and $b\bar{b}$, provided m_a is above the threshold

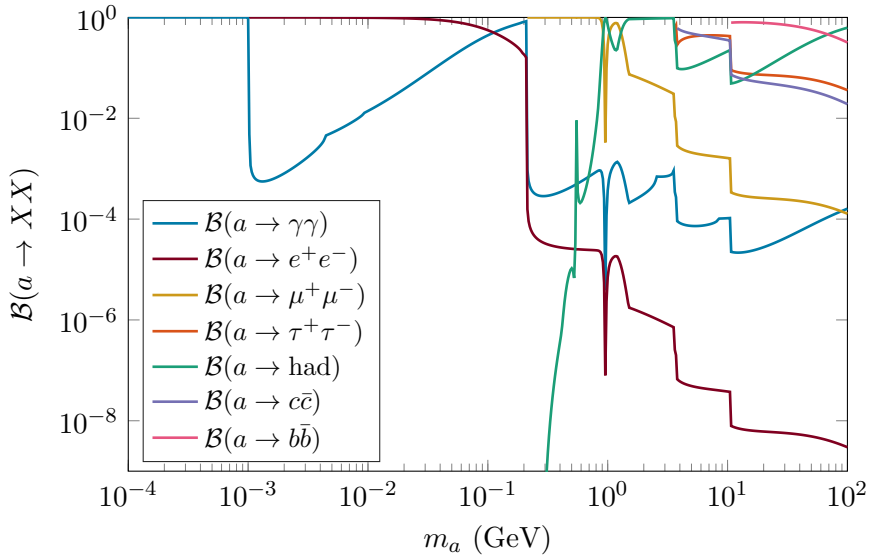


Figure 9: Branching fractions of the ALP into photons, leptons and hadrons as indicated, as a function of the ALP mass.

for the decay into a pair of D or B mesons respectively. For the contribution of the light quarks, we take into account the decays of ALPs to hadrons via

- If the mass of the ALP is less than 1.2 GeV, the decay width is calculated using code obtained by private communication, created for ref. [83]. This code was written for the decay of an NMSSM CP-odd Higgs, but with an appropriate choice of parameters ($P_{11} = \sqrt{2} v g_{aff}$, $\tan\beta = 1$ and decoupled neutralinos, charginos and heavy Higgs bosons) this particle can be identified with our ALP.

- Above 1.5 GeV, the decay width into light quarks is given by [74]

$$\Gamma(a \to \text{hadrons}) = g_{aff}^2 \frac{9 m_a^3 \alpha_s^2}{32\pi} \left(1 + \frac{83\alpha_s}{4\pi}\right), \tag{D.2}$$

  as in this regime the strong coupling is taken to be perturbative.

- Between 1.2 and 1.5 GeV we interpolate between the above determinations of the decay widths, in order to obtain a smooth result for the total decay width of the ALP, important particularly for the cosmological constraints.

This concludes the discussion of the tree-level decays of the ALP. In order to take into account experimental constraints on ALP decays to photons, we further require the decay width into two photons via a fermion loop. This is given by

$$\Gamma(a \to \gamma\gamma)_{\text{loop}} = \frac{\alpha^2 |T(m_a^2)|^2}{4\pi^3 m_a}, \tag{D.3}$$

where the triangle loop function is

$$T(s) = \sum_f g_{aff}\, m_f^2\, n_f^c\, q_f^2 \arcsin^2\left(\sqrt{s}/(2m_f)\right) \ , \tag{D.4}$$

where q_f is the electric charge of the fermion. In figure 9, we provide the branching ratios for all the above mentioned decays.

E Axion production cross sections in SLAC E137 experiment

E.1 Primakoff production

The differential cross section for the production of ALPs via the Primakoff mechanism (see figure 1a) is given by eq. (A1) in ref. [20],

$$\frac{\mathrm{d}\sigma_Z^{\gamma\to a}}{\mathrm{d}\Omega} = 8\alpha_{\mathrm{em}} \frac{\Gamma(a\to\gamma\gamma)}{m_a^3} \mid F(t) \mid^2 \frac{\theta^2}{\left(\theta^2 + \frac{1}{4}\left(\frac{m_a}{E_\gamma}\right)^4\right)^2} \ , \tag{E.1}$$

where $\Gamma(a \to \gamma\gamma)$ is the loop induced ALP decay width to two photons (see app. H for details), $t = (p_a - p_\gamma)^2$ is the momentum exchange with p_γ the four-momentum of the incoming electron and p_a the four-momentum of the outgoing ALP. Θ is the scattering polar angle in the laboratory frame. Z and $F(t)$ are the target atomic number and target form factor, respectively. For $t \leq 3\times 10^{-6}$ GeV^2 we use the atomic form factor, given by eqs. (A5) – (A6) in [20],

$$|F_{\mathrm{atomic}}(t)|^2 = Z^2 \left(\frac{a^2|t|}{a^2|t|+1}\right) + Z\left(\frac{a'^2|t|}{a'^2|t|+1}\right)^2 \tag{E.2}$$

with the parameters

$$a = \begin{cases} 122.8/m_e & \text{for hydrogen} \\ 111\ Z^{-1/3}/m_e & \text{for oxygen} \end{cases} \tag{E.3}$$

$$a' = \begin{cases} 282.4/m_e & \text{for hydrogen} \\ 773\ Z^{-2/3}/m_e & \text{for oxygen .} \end{cases} \tag{E.4}$$

For $t > 3\times 10^{-6}$ GeV^2 the elastic scattering form factor of the nuclei is used. The dipole form factor for the proton and the elastic form factor for oxygen are given by [84]

$$F_{\mathrm{di},H}(t) = \frac{1}{(1+\frac{|t|}{q_0^2})^2} \tag{E.5}$$

$$F_{\mathrm{el},O}(t) = Z\left(1 - \frac{a_0^2\, t}{8}\right) e^{-a_0^2\ t/4} \ , \tag{E.6}$$

with $q_0^2 = 0.71$ GeV^2 and $a_0 = 8.97$ GeV^{-1}. To obtain an expression for the momentum exchange t we need to find the relation between the energy of the incoming photon and the energy of the outgoing axion. Following [21], the momentum exchange can be expressed as

$$t = (p_\gamma - p_a)^2 = -\frac{m_a^4}{4\,E_a^2} - E_a^2\,\theta^2 \ . \tag{E.7}$$

under the assumption that $\theta \ll 1$, $m_a \ll E_a, m_N$ and $p_t \approx 0$. However, since both the photon track-length distribution and our expression for the Primakoff cross section are given in terms of the photon energy, we have to find an expression for the transverse momentum squared t in terms of E_γ. To first order approximation we have $E_\gamma \approx E_a$. As pointed out in [21], the term proportional to m_a^4 is negligible since we are considering small axion masses. Expanding to second order and inverting gives

$$E_a \approx E_\gamma - \frac{(E_\gamma \theta)^2}{2m_N} \tag{E.8}$$

and consequently

$$t = (p_\gamma - p_a)^2 = -2m_N(E_\gamma - E_a) \approx (E_\gamma \theta)^2 \,. \tag{E.9}$$

For our largest $m_a = 1$ GeV we get a relative error of 0.1% for the photon energy which is negligible considering the fact that we extract the track-length distribution from a graph.

E.2 Bremsstrahlung

The differential cross section for axion production by bremsstrahlung (see figure 1 b) is given by

$$\begin{aligned}\frac{\mathrm{d}^2\sigma_Z^{e\to ae}}{\mathrm{d}\Omega_a \mathrm{d}E_a} =& \frac{\alpha_{\mathrm{em}}^2 g_{aff}^2 m_e^2}{4\pi^2}\frac{E_e}{U^2}\left(x^3 - \frac{2m_a^2 x^2(1-x)}{U}\right.\\ &\left. + \frac{2m_a^2}{U^2}[m_a^2 x(1-x)^2 + m_e^2 x^3(1-x)]\right)\chi\,,\end{aligned} \tag{E.10}$$

with $U = E_e^2\theta_a^2 x + m_e^2 x + m_a^2(1-x)/x$ and $x = E_a/E_e$. χ is the integrated form factor obtained from integrating $\chi = \int_{t_{\min}} \mathrm{d}t |F(t)|^2 (t - t_{\min})/t^2$ [85],

$$\begin{aligned}\chi &= \chi_{\text{elastic}} + \chi_{\text{inelastic}}\\ &= Z^2\left[\ln\left(\frac{a^2 m_e^2(1+l)^2}{a^2 t_{\min}+1}\right) - 1\right] + Z\left[\ln\left(\frac{a'^2 m_e^2(1+l)^2}{a'^2 t_{\min}+1}\right) - 1\right]\,,\end{aligned} \tag{E.11}$$

where $t_{\min} = (U/(2E_e(1-x)))^2$ is the minimal momentum transfer squared and $l = E_e^2\theta_a^2/m_e^2$ and a and a' are to a reasonable approximation given by eqs. (E.3). m_e is the mass of the electron, E_e is the energy of the incoming electron and positron and E_a the energy of the ALP in the laboratory frame.

E.3 Positron annihilation

Non-resonant case The non-resonant cross section for positron annihilation in the centre of mass frame is given by (see figure 1c)

$$\frac{\mathrm{d}\sigma^{e^+\to a\gamma}}{\mathrm{d}\cos(\theta^*)} = \frac{g_{aff}^2 m_f^2 \alpha_{\mathrm{em}}}{2s}\left(\frac{m_a^2}{s - m_a^2}\right)\frac{1}{1-\beta^2\cos(\theta^*)^2}, \tag{E.12}$$

where θ^* is the scattering polar angle of the ALP in the centre of mass frame and $\beta = \sqrt{1 - 4m_e^2/s}$ is the speed of the electron target [20]. $s = 2E_+ m_e + 2m_e^2$ is the centre of

mass energy squared. The maximal opening angle for the ALP to be detected is given in the laboratory frame (see appendix F). We therefore have to relate the ALP's scattering angle in the centre of mass frame, θ^*, with $\theta^{\rm lab}_{\rm max}$, the angle in the laboratory frame. This can be achieved by expressing the Mandelstam variable t which is Lorentz invariant in the laboratory frame and in the centre of mass frame. We find

$$\cos(\theta^{\rm lab}) = \frac{E_+^{\rm lab} E_a^{\rm lab} - E_+^* E_a^* + \cos(\theta^*)\sqrt{E_a^{*2} - m_a^2}\sqrt{E_+^{*2} - m_e^2}}{\sqrt{E_+^{\rm lab2} - m_e^2}\sqrt{E_a^{\rm lab2} - m_a^2}}, \tag{E.13}$$

where the asterix denotes quantities in the centre of mass frame. The ALP's energy in the the laboratory frame is

$$E_a^{\rm lab} = \frac{s + m_a^2 + (s - m_a^2)\beta\cos(\theta^*)}{4m_e}. \tag{E.14}$$

In the centre of mass frame the energies of the ALP and of the positron are

$$E_a^* = \frac{4E_+^{*2} + m_a^2}{4E_+^*} \qquad E_+^* = \frac{1}{2}\sqrt{s} = \frac{1}{2}\sqrt{2(m_e^2 + m_e E_+^{\rm lab})}. \tag{E.15}$$

We have now all the ingredients to express the differential cross section in terms of the scattering angle and the positron energy in the laboratory frame over which we integrate. To deal with the infrared divergence in the cross section for $s \to m_a^2$, i.e. small photon energies, we apply a cut on the centre of mass energy around $\sqrt{s} = 3$ MeV as proposed in ref. [20].

Resonant production The resonant positron annihilation production cross section (see figure 1d) can be obtained from the Breit-Wigner formula and is in the centre of mass frame given by [86]

$$\sigma^{e^+\to a}(E_a^*) = \frac{\pi^2}{2E_a^{*2}} \frac{\Gamma(a \to e^+e^-)\Gamma(a \to XX)}{(E_a^* - m_a)^2 + \Gamma_a^2/4}\ , \tag{E.16}$$

with Γ_a the total decay width of the ALP and

$$\Gamma(a \to XX) = \begin{cases} \Gamma_{\rm loop}(a \to \gamma\gamma) & \text{for the ALP decaying into two photons} \\ \Gamma(a \to e^+e^-) & \text{for the ALP decaying into } e^+e^-. \end{cases} \tag{E.17}$$

Since the Breit-Wigner formula contains the decay width of the ALP which is strongly peaked compared to the step size in the positron energy the cross section is approximated by a delta function, i.e.

$$\sigma^{e^+\to a}(E_a^*) = \sigma_{\rm tot}^{e^+\to a}\delta(E_a^* - m_a) \tag{E.18}$$

where

$$\sigma_{\rm tot}^{e^+\to a}(a \to XX) = \int_{\infty}^{\infty} {\rm d}E_a^* \sigma^{e^+\to a}(E_a^*) = \frac{\pi^2}{2m_a^2}\Gamma(a \to e^+e^-)\mathcal{B}(a \to XX) \tag{E.19}$$

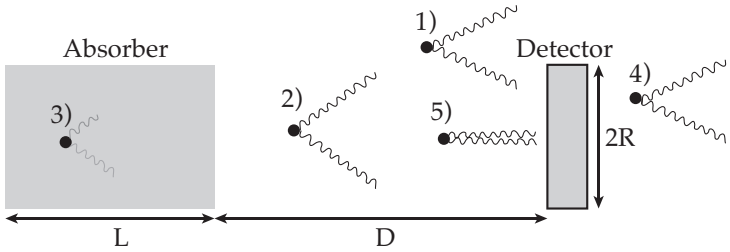


Figure 10: Different scenarios for the ALPs to decay invisibly (here schematically shown for photons).

and hence independent of the positron energy. The number of events in a specific decay channel can therefore be approximated by

$$\begin{aligned}\mathcal{N}(a \to XX) &= \mathcal{N}_{e,\,\mathrm{inc}}\, \sigma_{\mathrm{tot}}^{e^+\to a}(a \to XX) \int \mathrm{d}E_a^*\, T(E_+)\, p(E_+)\, \frac{\mathrm{d}E_+}{\mathrm{d}E_a^*}\, \delta(E_a^* - m_a) \sum_Z P_Z \\ &= \mathcal{N}_{e,\,\mathrm{inc}}\, \sigma_{\mathrm{tot}}^{e^+\to a}(a \to XX)\, T(E_+)\, p(E_+)\, \frac{m_a}{m_e} \sum_Z P_Z\,, \end{aligned} \tag{E.20}$$

where $T_+(E)$ is the track-length distribution of positrons. The energy of the positron and of the ALP are in the laboratory frame given by

$$E_+^{\mathrm{lab}} = \frac{m_a^2 - m_e^2}{2m_e}\,, \qquad E_a^{\mathrm{lab}} = E_+^{\mathrm{lab}} + m_e \tag{E.21}$$

respectively.

F Details on the ALP's detection probability

In this appendix we will give details about the probability for the ALP in the SLAC electron beam dump experiment to decay invisibly, i.e. we consider the cases where the ALP could have been produced but not been detected. In these cases the non-observation of axions in the experiment does not allow us to set bounds on the ALP's coupling and mass. To determine whether the ALP decays inside or outside the decay volume we need its decay length. The decay length of the ALP in the laboratory frame is

$$l_a = \beta\gamma_a\tau_a = \frac{\beta\gamma_a}{\Gamma_a} \approx \frac{E_a}{m_a\Gamma_a} \tag{F.1}$$

where we assume $E_a \gg m_a$ [21]. In the following discussion we consider the case where the axion decays into two photons.[9] Taking the experimental layout into account, the ALP cannot be detected if (the scenarios are summarised in figure 10):

[9] In principle, the same arguments can be applied for the case where the ALP decays to an electron-positron pair. Some approximations may not be reasonable as the decay products are massive (concerning boost, typical opening angle). To derive the bounds, we have used the same expressions and obtain very good agreement with the results presented in [20] (based on a Monte-Carlo simulation) and [74].

- the opening angle of the ALP is too big, $p(\theta_a > \theta_{\rm max}) = 0$. The detector which is approximated as cylindrical with radius $R = 1.5$ m can detect photons with a maximal separation angle of $\sin(\theta_{\rm max}) = R/(D+L) \approx \theta_{\rm max}$ (see definition in figure 10), which for the dimensions of E137 corresponds to $\theta_{\rm E137,\ max} = 0.22$ deg.

- the ALP decays too early and/or the opening angle between the photons is too large, so that both photons would miss the detector. The opening angle between the two photons is determined by the boost of the ALP. The minimal opening angle for high ALP boosts is given by $\theta_{\gamma,\ \rm min} \approx 2/\gamma_a$ [19, 21]. The distribution of the number of photons $\mathrm{d}N/\mathrm{d}\theta_\gamma$ is peaked at this minimal $\theta_{\gamma,\ \rm min}$ and the typical separation between the two photons arriving at the detector is therefore given by

$$d_{\gamma\gamma} = \sin\left[\theta_{\gamma,\ \rm min}(D+L-l_a)\right] \approx \theta_{\gamma,\ \rm min}(D+L-l_a) = \frac{2(D+L-l_a)}{\gamma_a}\,, \tag{F.2}$$

  where $D+L-l_a$ is the distance from the detector at which the ALP decays. $d_{\gamma\gamma}$ should not exceed the dimension of the detector, $p(d_{\gamma\gamma} > 2R) = p(D+L-l_a > RE_a/m_a) = 0$.

- the ALP decays inside the absorber and the photons get absorbed or the ALP decays behind the detector. The probability that the ALP decays inside the decay volume is given by

$$p(l_a) = p(l_a > D) - p(l_a > D+L) = e^{-D/l_a} - e^{-(D+L)/l_a}\,. \tag{F.3}$$

  Indeed, we find that this significantly reduces the number of detectable particles.

- the ALP decays in front of the detector but the separation between the two photons is too small. The typical separation between the photons has to be larger than the minimal resolution of the detector, $R_{\rm E137,\ min} \approx 3$ mm, which would lead to an indistinguishable signal, $p(d_{\gamma\gamma} < R_{\rm min}) = 0$.

To sum up, the probability to detect both photons is given by [21]

$$p(l_a) = \begin{cases} e^{-D/l_a} - e^{-(D+L)/l_a} & \text{if } \ E_a < 2m_a \cdot L/R_{\rm min}, \\ & E_a > (L+D)\cdot m_a/\left(R + 1/\Gamma_a\right) \ \text{and } \theta < \theta_{\rm max} \\ 0 & \text{otherwise}\,. \end{cases} \tag{F.4}$$

G Details on constraints from exotic Higgs decays

To understand the sensitivity of the LHC exotic Higgs decay searches, it is first important to calculate the decay table for the axion. A detailed calculation is available in the appendix and the branching fractions are plotted in figure 9 for a range of ALP masses from 10 MeV to 100 GeV. What is relevant for our discussion here is that the branching fraction into two photons remains consistently smaller than 10^{-3} as soon as the ALP mass is large enough to allow $a \to \mu^+\mu^-$. Next, as expected, once each new massive fermion mode turns on, it quickly dominates as the width in that channel is proportional to m_f^2. We also assume that as soon as the mass threshold is above 1.5 GeV ($\simeq 2m_\rho$), decays into hadrons via

loop-mediated $a \to gg$ open up. Therefore, once hadronic decays become allowed, any branching into muons becomes again vanishingly small. Our calculation (assuming $a \to gg$ open only above 1 GeV) gives $\mathcal{B}(\mu^+\mu^-) < 10^{-3}$ as soon as $\tau^+\tau^-$ and $c\bar{c}$ modes open (i.e. $m_a \gtrsim 1.8$ GeV). Since all current exotic Higgs decays look for $m_a \gtrsim 10$ GeV only, we can safely assume that none of the muonic searches will be able to see the signatures of our model. We therefore need only think about the $h \to aa \to 4b$ channel. The limits on $h \to aa \to 4b$ are only able to exclude branching ratios less than one in the small mass range between 18-22.5 GeV. However, this would still require a branching fraction of 0.75 of the SM Higgs into this one channel alone. Given that in our model, these channels are dependent on 1-loop contributions only the decay width of the Higgs into these modes cannot be of comparable size to the usual tree-level 2-body SM decays. We therefore conclude that none of these searches unfortunately have any sensitivity to our model.

The current limit [26] in $h \to Z(a \to gg)$ is in principle sensitive to our model in the very narrow range 0.5 GeV $< m_a \lesssim 2.7$ GeV. However, this is precisely the mass range where theoretical calculations are wildly unpredictable due to hadronic contributions and we choose not to apply these.

Lastly, the $h \to Z\gamma$ searches can be used in certain phase space regions where the two photons from a-decay are collimated, i.e. $h \to Z(a \to \gamma\gamma)$. The $h \to ZZ^* \to 4\ell$ measurements can also be used to place a limit on new physics contributions as the off-shell Z decaying into two leptons can also be interpreted as a new particle that decays into two leptons. A study of the collimated di-gamma decay was done in [87]. However, interpreted in our model, it corresponds only to coupling values $C_{aff} \gtrsim 10^2/\text{GeV}$. Since this lies wildly outside the self-consistent EFT regime, we assume that the $h \to Z\gamma$ measurements are currently not sensitive enough to provide a useful constraint. Similarly, the 4ℓ decay mode of the Higgs currently carries the signal strength accuracy $\mu = 1.44 \pm 0.4$ [88] which translates to an upper limit on the decay width $\Gamma(h \to Za) \times \mathcal{B}(a \to 2\ell) \lesssim 0.44 \times \Gamma(h \to ZZ^* \to 4\ell)$. Again, due to the loop suppression in our decay, we end up with exclusions on $C_{aff} \gg 1$ only and there is no usable limit from this measurement.

H Loop diagrams relevant for constraint calculations

The ALP in our model can only decay into photons via a fermion loop. Applying the constraints from the effective axion-photon coupling, $g_{a\gamma\gamma}$, to the axion-fermion coupling, g_{aff}, demands more theoretical groundwork. The main task is to calculate the Primakoff cross section for this coupling. The resulting cross section can then be inserted into the expression for the ALP yield for the beam dump constraint and the expressions for the total energy outflow and the ALP opacity for the HB stars constraints. The approach we adopt is to replace the expression for the axion-photon coupling in the "decay width" in equations (E.1) by an expression for the axion-photon coupling induced by a fermion loop. This coupling will now depend on the momentum transfer squared, $t = (p_\gamma - p_a)^2$. The Feynman diagram of the Primakoff process induced by a fermion loop is shown in figure 12 on the left. It should be noticed that the decay width in equation (E.1) has to be regarded

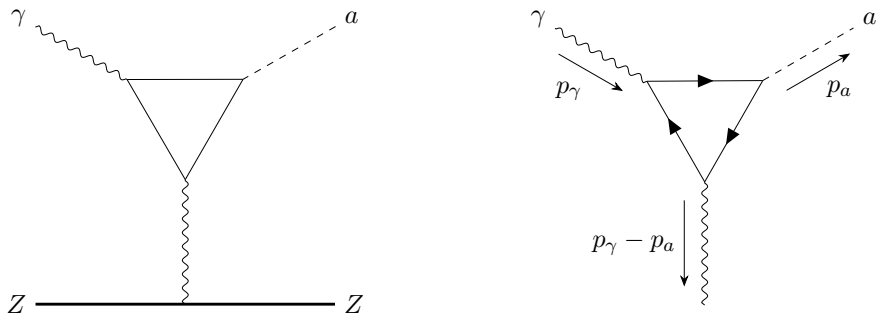


Figure 12: *Left:* Feynman diagram of the Primakoff production process via a fermion loop. *Right:* Feynman diagram of the amplitude we need to calculate to replace $g_{a\gamma\gamma}$ by $g_{a\gamma\gamma}^{\text{loop}}$ in the Primakoff production process.

as a definition of the axion-photon coupling rather than an actual physical decay width,

$$\begin{aligned}\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} &= 8\alpha_{\text{em}}\frac{\Gamma}{m_a^3}\mid F(t)\mid^2 \frac{\theta^2}{\left(\theta^2+\frac{1}{4}\left(\frac{m_a}{E_\gamma}\right)^4\right)^2} \\ &= \alpha_{\text{em}}\frac{g_{a\gamma\gamma}^2}{8\pi}\mid F(t)\mid^2 \frac{\theta^2}{\left(\theta^2+\frac{1}{4}\left(\frac{m_a}{E_\gamma}\right)^4\right)^2}\,. \end{aligned} \tag{H.1}$$

By comparing the amplitude of the loop diagram and the tree-level effective diagram we will deduce what $g_{a\gamma\gamma}$ should be replaced by. The Feynman diagram of the amplitude we have to consider is depicted in figure 12 on the right. We proceed as in ref. [89] and find that the amplitude of the triangle graph is given by

$$\begin{aligned}\mathcal{T}^{\alpha\beta} &= \int \frac{d^4k}{(2\pi)^4}(-1)\,\mathrm{Tr}\left[\frac{i}{\not{k}+\not{p}_\gamma-m_f}\gamma^\alpha\frac{i}{\not{k}-m_f}\gamma^\beta\frac{i}{\not{k}+\not{p}_\gamma-\not{p}_a-m_f}\gamma_5\right] \\ &= -i\frac{1}{2\pi^2}\varepsilon^{\alpha\beta\rho\sigma}p_{a,\rho}p_{\gamma,\sigma}m_f C_0(p_a^2,p_\gamma^2,(p_a-p_\gamma)^2,m_f^2,m_f^2,m_f^2)\,, \end{aligned} \tag{H.2}$$

with the three-point scalar function C_0 defined as in ref. [90]. The matrix element squared of the diagram is given by (with $t=(p_a-p_\gamma)^2$ in eq. (E.1))

$$|\mathcal{M}|^2_{\text{loop}} = \frac{\alpha_{\text{em}}^2(m_a^2-t)^2\left[\sum_f g_{aff}m_f^2q_f^2n_f^c|C_0(m_a^2,0,t,m_f^2,m_f^2,m_f^2)|\right]^2}{2\pi^2}\,. \tag{H.3}$$

By calculating the same diagram but with an effective ALP-photon coupling $g_{a\gamma\gamma}$ at tree level, i.e.

$$\begin{aligned}|\mathcal{M}|^2 &= g_{a\gamma\gamma}^2\epsilon_{\mu\nu\rho\sigma}\epsilon_{\alpha\beta\kappa\lambda}g^{\alpha\mu}\,g^{\beta\nu}\,p_\gamma^\rho\,p_\gamma^\kappa\,(p_\gamma-p_a)^\sigma\,(p_\gamma-p_a)^\lambda \\ &= 2\,g_{a\gamma\gamma}^2\left((p_a\cdot p_\gamma)^2-(p_a\cdot p_a)(p_\gamma\cdot p_\gamma)\right) = \frac{g_{a\gamma\gamma}^2(m_a^2-t)^2}{2} \end{aligned} \tag{H.4}$$

we obtain the replacement

$$g_{a\gamma\gamma}^2 \to \frac{\alpha_{\rm em}^2 \left[\sum_f g_{aff} m_f^2 q_f^2 n_f^c |C_0(m_a^2, 0, t, m_f^2, m_f^2, m_f^2)|\right]^2}{\pi^2} \tag{H.5}$$

in the expression for the Primakoff production in the E137 SLAC experiment, eq. (E.1), and in the expression for the Primakoff production in horizontal branch stars. For the constraints from horizontal branch stars we make the approximation of zero momentum transfer, i.e. $t \approx 0$.

References

[1] XENON Collaboration collaboration, *Constraining the spin-dependent wimp-nucleon cross sections with xenon1t*, *Phys. Rev. Lett.* **122** (2019) 141301.

[2] A. Boveia and C. Doglioni, *Dark matter searches at colliders*, *Annual Review of Nuclear and Particle Science* **68** (2018) 429.

[3] L.J. Hall, K. Jedamzik, J. March-Russell and S.M. West, *Freeze-in production of fimp dark matter*, *Journal of High Energy Physics* **2010** (2010) 80.

[4] N. Bernal, M. Heikinheimo, T. Tenkanen, K. Tuominen and V. Vaskonen, *The dawn of fimp dark matter: A review of models and constraints*, *International Journal of Modern Physics A* **32** (2017) 1730023.

[5] L. Di Luzio, M. Giannotti, E. Nardi and L. Visinelli, *The landscape of QCD axion models*, *Phys. Rept.* **870** (2020) 1 [2003.01100].

[6] X. Chu, T. Hambye and M.H. Tytgat, *The four basic ways of creating dark matter through a portal*, *Journal of Cosmology and Astroparticle Physics* **2012** (2012) 034.

[7] T. Hambye, M.H.G. Tytgat, J. Vandecasteele and L. Vanderheyden, *Dark matter from dark photons: a taxonomy of dark matter production*, *Phys. Rev. D* **100** (2019) 095018 [1908.09864].

[8] G. Bélanger, C. Delaunay, A. Pukhov and B. Zaldivar, *Dark matter abundance from the sequential freeze-in mechanism*, *Phys. Rev. D* **102** (2020) 035017 [2005.06294].

[9] K. Saikawa and S. Shirai, *Primordial gravitational waves, precisely: The role of thermodynamics in the Standard Model*, *JCAP* **05** (2018) 035 [1803.01038].

[10] S. Borsanyi et al., *Calculation of the axion mass based on high-temperature lattice quantum chromodynamics*, *Nature* **539** (2016) 69 [1606.07494].

[11] F. Elahi, C. Kolda and J. Unwin, *UltraViolet Freeze-in*, *JHEP* **03** (2015) 048 [1410.6157].

[12] A. Salvio, A. Strumia and W. Xue, *Thermal axion production*, *JCAP* **01** (2014) 011 [1310.6982].

[13] A. Biswas, S. Ganguly and S. Roy, *Fermionic dark matter via UV and IR freeze-in and its possible X-ray signature*, *JCAP* **03** (2020) 043 [1907.07973].

[14] P. Gondolo and G. Gelmini, *Cosmic abundances of stable particles: Improved analysis*, *Nuclear Physics B* **360** (1991) 145 .

[15] X. Chu, T. Hambye and M.H.G. Tytgat, *The Four Basic Ways of Creating Dark Matter Through a Portal*, *JCAP* **05** (2012) 034 [1112.0493].

[16] T. Bhattacharya, V. Cirigliano, S. Cohen, R. Gupta, H.-W. Lin and B. Yoon, *Axial, Scalar and Tensor Charges of the Nucleon from 2+1+1-flavor Lattice QCD*, *Phys. Rev. D* **94** (2016) 054508 [1606.07049].

[17] J. Bijnens, G. Colangelo and P. Talavera, *The Vector and scalar form-factors of the pion to two loops*, *JHEP* **05** (1998) 014 [hep-ph/9805389].

[18] H.A. Weldon, *Effective fermion masses of order* gT *in high-temperature gauge theories with exact chiral invariance*, *Phys. Rev. D* **26** (1982) 2789.

[19] M.J. Dolan, T. Ferber, C. Hearty, F. Kahlhoefer and K. Schmidt-Hoberg, *Revised constraints and belle ii sensitivity for visible and invisible axion-like particles*, *Journal of High Energy Physics* **2017** (2017) 94.

[20] J.D. Bjorken, S. Ecklund, W.R. Nelson, A. Abashian, C. Church, B. Lu et al., *Search for neutral metastable penetrating particles produced in the slac beam dump*, *Phys. Rev. D* **38** (1988) 3375.

[21] B. Döbrich, J. Jaeckel, F. Kahlhoefer, A. Ringwald and K. Schmidt-Hoberg, *Alptraum: Alp production in proton beam dump experiments*, *Journal of High Energy Physics* **2016** (2016) 18.

[22] V. Budnev, I. Ginzburg, G. Meledin and V. Serbo, *The two-photon particle production mechanism. physical problems. applications. equivalent photon approximation*, *Physics Reports* **15** (1975) 181.

[23] W.R. Nelson, H. Hirayama and D. Rogers, *Slac report 265*, .

[24] L. Darmé, F. Giacchino, E. Nardi and M. Raggi, *Invisible decays of axion-like particles: constraints and prospects*, *Journal of High Energy Physics* **2021** (2021) .

[25] L. Marsicano, M. Battaglieri, M. Bondí, C.D.R. Carvajal, A. Celentano, M. De Napoli et al., *Dark photon production through positron annihilation in beam-dump experiments*, *Phys. Rev. D* **98** (2018) 015031.

[26] ATLAS collaboration, *Search for Higgs Boson Decays into a Z Boson and a Light Hadronically Decaying Resonance Using 13 TeV pp Collision Data from the ATLAS Detector*, *Phys. Rev. Lett.* **125** (2020) 221802 [2004.01678].

[27] ATLAS collaboration, *Search for Higgs boson decays into two new low-mass spin-0 particles in the 4b channel with the ATLAS detector using pp collisions at* $\sqrt{s} = 13$ *TeV*, *Phys. Rev. D* **102** (2020) 112006 [2005.12236].

[28] ATLAS collaboration, *Search for Higgs bosons decaying into new spin-0 or spin-1 particles in four-lepton final states with the ATLAS detector with 139* fb^{-1} *of pp collision data at* $\sqrt{s} = 13$ *TeV*, 2110.13673.

[29] ATLAS collaboration, *Search for Higgs boson decays into a pair of light bosons in the* $bb\mu\mu$ *final state in pp collision at* $\sqrt{s}$ *=13 TeV with the ATLAS detector*, *Phys. Lett. B* **790** (2019) 1 [1807.00539].

[30] LHCb collaboration, *Search for long-lived scalar particles in* $B^+ \to K^+\chi(\mu^+\mu^-)$ *decays*, *Phys. Rev. D* **95** (2017) 071101 [1612.07818].

[31] LHCb collaboration, *Search for hidden-sector bosons in* $B^0 \to K^{*0}\mu^+\mu^-$ *decays*, *Phys. Rev. Lett.* **115** (2015) 161802 [1508.04094].

[32] B. Döbrich, F. Ertas, F. Kahlhoefer and T. Spadaro, *Model-independent bounds on light pseudoscalars from rare B-meson decays*, *Phys. Lett. B* **790** (2019) 537 [1810.11336].

[33] NA62 collaboration, *Measurement of the very rare $K^+ \to \pi^+ \nu\overline{\nu}$ decay*, *JHEP* **06** (2021) 093 [2103.15389].

[34] NA48/2 collaboration, *Precise measurement of the K+- —> pi+-e+e- decay*, *Phys. Lett. B* **677** (2009) 246 [0903.3130].

[35] NA48/2 collaboration, *Searches for lepton number violation and resonances in $K^\pm \to \pi\mu\mu$ decays*, *Phys. Lett. B* **769** (2017) 67 [1612.04723].

[36] BaBar collaboration, *Search for $B \to K^{(*)}\nu\overline{\nu}$ and invisible quarkonium decays*, *Phys. Rev. D* **87** (2013) 112005 [1303.7465].

[37] M.B. Gavela, R. Houtz, P. Quilez, R. Del Rey and O. Sumensari, *Flavor constraints on electroweak ALP couplings*, *Eur. Phys. J. C* **79** (2019) 369 [1901.02031].

[38] J. Martin Camalich, M. Pospelov, P.N.H. Vuong, R. Ziegler and J. Zupan, *Quark Flavor Phenomenology of the QCD Axion*, *Phys. Rev. D* **102** (2020) 015023 [2002.04623].

[39] A. Bharucha, T. Feldmann and M. Wick, *Theoretical and Phenomenological Constraints on Form Factors for Radiative and Semi-Leptonic B-Meson Decays*, *JHEP* **09** (2010) 090 [1004.3249].

[40] A. Bharucha, D.M. Straub and R. Zwicky, *$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules*, *JHEP* **08** (2016) 098 [1503.05534].

[41] D.S.M. Alves and N. Weiner, *A viable QCD axion in the MeV mass range*, *JHEP* **07** (2018) 092 [1710.03764].

[42] S. Friot, D. Greynat and E. De Rafael, *Rare kaon decays revisited*, *Phys. Lett. B* **595** (2004) 301 [hep-ph/0404136].

[43] CHARM collaboration, *Search for Axion Like Particle Production in 400-GeV Proton - Copper Interactions*, *Phys. Lett. B* **157** (1985) 458.

[44] CHARM collaboration, *A Search for Decays of Heavy Neutrinos*, *Phys. Lett. B* **128** (1983) 361.

[45] J.H. Chang, R. Essig and S.D. McDermott, *Revisiting Supernova 1987A Constraints on Dark Photons*, *JHEP* **01** (2017) 107 [1611.03864].

[46] J.H. Chang, R. Essig and S.D. McDermott, *Supernova 1987A Constraints on Sub-GeV Dark Sectors, Millicharged Particles, the QCD Axion, and an Axion-like Particle*, *JHEP* **09** (2018) 051 [1803.00993].

[47] G. Raffelt, *Stars as Laboratories for Fundamental Physics: The Astrophysics of Neutrinos, Axions, and Other Weakly Interacting Particles*, Theoretical Astrophysics, University of Chicago Press (1996).

[48] G. Grilli di Cortona, E. Hardy, J. Pardo Vega and G. Villadoro, *The QCD axion, precisely*, *JHEP* **01** (2016) 034 [1511.02867].

[49] S. Hannestad and G. Raffelt, *Supernova neutrino opacity from nucleon-nucleon Bremsstrahlung and related processes*, *Astrophys. J.* **507** (1998) 339 [astro-ph/9711132].

[50] A. Bartl, R. Bollig, H.-T. Janka and A. Schwenk, *Impact of Nucleon-Nucleon Bremsstrahlung Rates Beyond One-Pion Exchange*, *Phys. Rev. D* **94** (2016) 083009 [1608.05037].

[51] G. G. Raffelt and G. Starkman, *Stellar energy transfer by kev-mass scalars*, *Physical review D: Particles and fields* **40** (1989) 942.

[52] M. Pospelov, A. Ritz and M. Voloshin, *Bosonic super-wimps as kev-scale dark matter*, *Physical Review D* **78** (2008) .

[53] G.G. Raffelt, *Astrophysical axion bounds diminished by screening effects*, *Phys. Rev. D* **33** (1986) 897.

[54] D. Cadamuro and J. Redondo, *Cosmological bounds on pseudo nambu-goldstone bosons*, *Journal of Cosmology and Astroparticle Physics* **2012** (2012) 032.

[55] M. Millea, L. Knox and B. Fields, *New Bounds for Axions and Axion-Like Particles with keV-GeV Masses*, *Phys. Rev. D* **92** (2015) 023010 [1501.04097].

[56] P.F. Depta, M. Hufnagel and K. Schmidt-Hoberg, *Robust cosmological constraints on axion-like particles*, *JCAP* **05** (2020) 009 [2002.08370].

[57] B.D. Fields, K.A. Olive, T.-H. Yeh and C. Young, *Big-Bang Nucleosynthesis after Planck*, *JCAP* **03** (2020) 010 [1912.01132].

[58] M. Kawasaki, K. Kohri, T. Moroi, K. Murai and H. Murayama, *Big-bang nucleosynthesis with sub-GeV massive decaying particles*, *JCAP* **12** (2020) 048 [2006.14803].

[59] P.F. Depta, M. Hufnagel and K. Schmidt-Hoberg, *Updated BBN constraints on electromagnetic decays of MeV-scale particles*, *JCAP* **04** (2021) 011 [2011.06519].

[60] M.H. Reno and D. Seckel, *Primordial Nucleosynthesis: The Effects of Injecting Hadrons*, *Phys. Rev. D* **37** (1988) 3441.

[61] M. Kawasaki, K. Kohri, T. Moroi and Y. Takaesu, *Revisiting Big-Bang Nucleosynthesis Constraints on Long-Lived Decaying Particles*, *Phys. Rev. D* **97** (2018) 023502 [1709.01211].

[62] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, .

[63] M.T.P.D. Group), *Review of particle physics*, *Physical Review D* **98** (2018) .

[64] L. Bergstrom, T. Bringmann, I. Cholis, D. Hooper and C. Weniger, *New Limits on Dark Matter Annihilation from AMS Cosmic Ray Positron Data*, *Phys. Rev. Lett.* **111** (2013) 171101 [1306.3983].

[65] A. Cuoco, J. Heisig, M. Korsmeier and M. Krämer, *Constraining heavy dark matter with cosmic-ray antiprotons*, *JCAP* **04** (2018) 004 [1711.05274].

[66] G.B. Gelmini, V. Takhistov and S.J. Witte, *Casting a Wide Signal Net with Future Direct Dark Matter Detection Experiments*, *JCAP* **07** (2018) 009 [1804.01638].

[67] S. Tulin and H.-B. Yu, *Dark matter self-interactions and small scale structure*, *Physics Reports* **730** (2018) 1–57.

[68] F. Kahlhoefer, K. Schmidt-Hoberg, M.T. Frandsen and S. Sarkar, *Colliding clusters and dark matter self-interactions*, *Monthly Notices of the Royal Astronomical Society* **437** (2013) 2865–2881.

[69] P.F. Bedaque, M.I. Buchoff and R.K. Mishra, *Sommerfeld enhancement from goldstone pseudo-scalar exchange*, *Journal of High Energy Physics* **2009** (2009) 046–046.

[70] Z.-P. Liu, Y.-L. Wu and Y.-F. Zhou, *Sommerfeld enhancements with vector, scalar, and pseudoscalar force carriers*, *Physical Review D* **88** (2013) .

[71] M.J. Dolan, F. Kahlhoefer, C. McCabe and K. Schmidt-Hoberg, *A taste of dark matter: flavour constraints on pseudoscalar mediators*, *Journal of High Energy Physics* **2015** (2015) 171.

[72] P. Agrawal, A. Parikh and M. Reece, *Systematizing the effective theory of self-interacting dark matter*, *Journal of High Energy Physics* **2020** (2020) .

[73] F. Kahlhoefer, K. Schmidt-Hoberg and S. Wild, *Dark matter self-interactions from a general spin-0 mediator*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 003–003.

[74] M. Bauer, M. Neubert and A. Thamm, *Collider Probes of Axion-Like Particles*, *JHEP* **12** (2017) 044.

[75] P. Carenza, O. Straniero, B. Döbrich, M. Giannotti, G. Lucente and A. Mirizzi, *Constraints on the coupling with photons of heavy axion-like-particles from globular clusters*, *Physics Letters B* **809** (2020) 135709.

[76] L. Di Luzio, M. Fedele, M. Giannotti, F. Mescia and E. Nardi, *Solar axions cannot explain the xenon1t excess*, *Phys. Rev. Lett.* **125** (2020) 131804.

[77] A. Ayala, I. Domínguez, M. Giannotti, A. Mirizzi and O. Straniero, *Revisiting the bound on axion-photon coupling from globular clusters*, *Phys. Rev. Lett.* **113** (2014) 191302.

[78] M. Giannotti, I.G. Irastorza, J. Redondo, A. Ringwald and K. Saikawa, *Stellar recipes for axion hunters*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 010.

[79] Straniero, O., Pallanca, C., Dalessandro, E., Domínguez, I., Ferraro, F. R., Giannotti, M. et al., *The rgb tip of galactic globular clusters and the revision of the axion-electron coupling bound*, *A&A* **644** (2020) A166.

[80] P.N. Brown, G.D. Byrne and A.C. Hindmarsh, *Vode: A variable-coefficient ode solver*, *SIAM Journal on Scientific and Statistical Computing* **10** (1989) 1038 [https://doi.org/10.1137/0910062].

[81] M. Dine, W. Fischler and M. Srednicki, *A Simple Solution to the Strong CP Problem with a Harmless Axion*, *Phys. Lett. B* **104** (1981) 199.

[82] D.E. Kaplan and R. Rattazzi, *Large field excursions and approximate discrete symmetries from a clockwork axion*, *Phys. Rev. D* **93** (2016) 085007 [1511.01827].

[83] F. Domingo, *Decays of a nmssm cp-odd higgs in the low-mass region*, *Journal of High Energy Physics* **2017** (2017) 1.

[84] R. Hofstadter, *Nuclear and nucleon scattering of high-energy electrons*, *Annual review of nuclear science* **7** (1957) 231.

[85] Y.S. Tsai, *Axion bremsstrahlung by an electron beam*, *Phys. Rev. D* **34** (1986) 1326.

[86] B. Lu, *Search for Axion-Like Particles in the SLAC Beam Dump.*, Ph.D. thesis, VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY., Sept., 1987.

[87] G. Brooijmans et al., *Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report*, in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 2, 2020 [2002.12220].

[88] ATLAS collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment*, *Eur. Phys. J. C* **76** (2016) 6 [1507.04548].

[89] J. Quevillon and C. Smith, *Axions are blind to anomalies*, *Eur. Phys. J. C* **79** (2019) 822 [1903.12559].

[90] V. Shtabovenko, R. Mertig and F. Orellana, *New developments in feyncalc 9.0*, *Computer Physics Communications* **207** (2016) 432–444.