Dark Matter’s secret liaisons: 
phenomenology of a dark $U(1)$ sector with bound states

Marco Cirelli $^a$, Paolo Panci $^{b,c}$, Kalliopi Petraki $^{a,d}$,
Filippo Sala $^a$, Marco Taoso $^e$

$^a$ Laboratoire de Physique Théorique et Hautes Energies (LPTHE),
UMR 7589 CNRS & UPMC,
4 Place Jussieu, F-75252, Paris, France

$^b$ CERN Theoretical Physics Department, CERN,
Case C01600, CH-1211 Genève, Switzerland

$^c$ Institut d’Astrophysique de Paris,
UMR 7095 CNRS, Université Pierre et Marie Curie,
98 bis Boulevard Arago, Paris 75014, France

$^d$ Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands

$^e$ Instituto de Física Teórica (IFT) UAM/CSIC,
calle Nicolás Cabrera 13-15, 28049 Cantoblanco, Madrid, Spain

Abstract

Dark matter (DM) charged under a dark $U(1)$ force appears in many extensions of the Standard Model, and has been invoked to explain anomalies in cosmic-ray data, as well as a self-interacting DM candidate. In this paper, we perform a comprehensive phenomenological analysis of such a model, assuming that the DM abundance arises from the thermal freeze-out of the dark interactions. We include, for the first time, bound-state effects both in the DM production and in the indirect detection signals, and quantify their importance for Fermi, AMS-02, and CMB experiments. We find that DM in the mass range 1 GeV to 100 TeV, annihilating into dark photons of MeV to GeV mass, is in conflict with observations. Instead, DM annihilation into heavier dark photons is viable. We point out that the late decays of multi-GeV dark photons can produce significant entropy and thus dilute the DM density. This can lower considerably the dark coupling needed to obtain the DM abundance, and in turn relax the existing constraints.
1 Introduction

The physics of the dark sector, comprising most notably the Dark Matter (DM) particle responsible for incontrovertible astrophysical and cosmological evidence, is still largely, and aptly, obscure (see e.g. [1, 2] for reviews). Not only the mass of the DM particle is undetermined, with viable possibilities ranging from tiny fractions of one eV to hundreds of TeV, but perhaps most importantly, its interaction properties are so far unknown (beyond, of course, the coupling with gravity which is at the origin of the evidence mentioned above). Historically, much of the attention has concentrated on WIMP scenarios, in which the DM particle – often assumed, for simplicity, to consist of just one species – is charged under the well-known weak interactions of the Standard Model (SM). Under this logic, the matter content of the theory is enlarged but the gauge force sector is left unchanged. The relative predictivity of these scenarios, and their appealing embedding within broader theory constructions such as supersymmetry, have made them paradigmatic for the field in the past decades.

On the other hand, it is interesting to entertain the possibility that the DM may be charged under a new dark force, carried by a new dark mediator. We will consider, specifically, DM coupled to a dark Abelian gauge interaction $U(1)_D$, and allow for the dark gauge boson, which we shall call the dark photon, to be massive. Such a force couples DM with itself, making it potentially collisional. It also generically allows DM to couple to the Standard Model (SM), via the renormalisable kinetic mixing of the $U(1)_D$ with the hypercharge [3, 4]; this gives rise to direct and indirect detection signals, and has important implications for cosmology.
Models of this sort have been investigated in many occasions in the past, based both on theoretical [5–12] and phenomenological motivations. The latter includes, for instance, the long-standing possibility that sizeable DM self-interactions can solve some worrying inconsistencies between the predictions of collisionless cold DM and the observed large-scale structure of the Universe [13], that appear, for example, in the sizes and numbers of the galactic satellite haloes, and in the DM density profiles in the center of galaxies [14–20]. The same class of models was also invoked in the wake of the discovery of a positron excess in cosmic rays by the PAMELA satellite [21–28]; the specific implementations featured a DM particle that annihilates into mediators with a relatively small mass, below the threshold for $p\bar{p}$ production. This feature allowed to explain the leptophilic nature of the annihilation and, at the same time, to provide the very sizeable enhancement of the annihilation cross-section – the so-called Sommerfeld enhancement [28–35] – that was needed to fit the data. Later on, models in this class were constructed to explain the FERMI Galactic Center GeV excess [36–42]. In these constructions, the existence of dark mediators allows to obtain smoother gamma-ray spectra and to avoid other constraints [43–51].

As seen from the above, the suitability of the $U(1)_D$ model to address certain problems relies often on the smallness of the force mediator mass. More recently, it has been realised in the context of DM phenomenology that long-range interactions – interactions mediated by light or massless force carriers – imply the existence and formation of DM bound states. This applies both to dark force scenarios, as well as to WIMPs, provided that the force carriers (the dark mediator or the SM Weak bosons) are much lighter than the DM particles. After the pioneering work of refs. [23, 52, 53], the issue has been investigated more systematically in a series of recent studies [54–71]. Dark matter with no particle-antiparticle asymmetry, coupled to long-range interactions can form unstable particle-antiparticle bound states. Their formation and decay has an important impact on DM phenomenology: it provides an additional annihilation channel, which affects the relic abundance [57], as well as the radiative signals looked for in indirect detection strategies [62–64]. The modifications can be significant, and need to be taken into account.

The goal of this paper is to perform, for the first time, a comprehensive and self-consistent analysis of the $U(1)_D$ model that will include the impact of bound states. We compute precisely the DM relic density, incorporating the effect of bound-state formation (BSF) and decay, which reduces the predicted DM coupling [54]. We analyze the constraints on the parameter space of the model that come from beam dump experiments, supernova cooling, DM direct detection and Big Bang Nucleosynthesis. Then we consider the most promising DM indirect detection probes: the gamma-ray signals from the Milky Way galactic halo and from dwarf galaxies, the antiproton measurements by the AMS-02 experiment and the anisotropies of the Cosmic Microwave Background (CMB). Recently, two related studies with some overlap with ours have appeared in the literature [62, 72]; our results broadly agree with theirs, but our analysis has a wider scope, uses some different ingredients and points out new effects.\footnote{The main differences with [62] are the following. i) We take into account the effect of BSF on the DM relic density, hence on the DM coupling to the dark photon. ii) While they essentially focus their indirect detection analysis on $\gamma$-rays from the Galactic Center (GC), we employ a wider array of probes. iii) We avoid relying on the GC as it is a challenging region (where actually an excess is currently discussed) and therefore it is less suitable for deriving constraints. iv) We include the Inverse Compton Scattering component in the computation of the $\gamma$-ray spectra: this component significantly increases the flux at small energies, say $10^{-2}$ times the DM mass, and it cannot therefore be neglected especially if one wants to derive constraints on DM as heavy as 10 to 100 TeV using the FERMI data that extend to hundreds of GeVs at most. We will come back on these points in detail in the following sections.}

\footnote{The very recent study in [72] differs from ours in that it does not consider our main topic, i.e. the formation of bound states, while it focuses instead on DM self-interactions. Moreover, we differ in some of the probes that we consider to derive limits: we do not rely on AMS-02 positron measurements, while they do not consider the AMS-02}

For more work related to the indirect
detection of this kind of models, see also [64, 73–75].

The rest of this paper is organized as follows. In section 2, we specify the details of the particle physics model, and we recall the main features associated with the formation of bound states. We also set our notation, and define the parameter space on which we focus. In section 3, we expound on the cosmology of the model. We detail the computation of the DM relic abundance, describe the cosmological evolution of the DM and dark radiation after freeze-out, consider constraints arising from BBN, and quantify how late decays of dark photons might dilute the DM abundance. In section 4, we present the constraints we obtain from indirect detection, based on gamma-ray measurements by Fermi and anti-proton searches by AMS-02, and the constraints we obtain from CMB, based on measurements by Planck. Finally, in section 5, we summarize and we present our conclusions.

2 Setting the stage

We consider DM in the form of Dirac fermions $X$ coupled to a dark force $U(1)_D$ carried by a dark photon $V_D$, that mixes kinetically with the hypercharge $U(1)_Y$. The Lagrangian of the dark sector reads

$$\mathcal{L} = \bar{X}(i\gamma^\mu D_\mu - M_{DM})X - \frac{1}{4} F_{\mu\nu}^D F_{\mu\nu}^D - \frac{1}{2} m_{V_D}^2 V_\mu V_\mu^\dagger - \frac{\epsilon}{2c_w} F_{\mu\nu}^D F_{\mu\nu}^Y,$$  \hspace{1cm} (1)

where the covariant derivative for $X$ is $D_\mu = \partial_\mu + ig_D V_\mu$, and $F_{\mu\nu}^D = \partial_\mu V_\nu - \partial_\nu V_\mu$. The mass $m_{V_D}$ of the dark photon field may arise either via the St"uckelberg mechanism or a dark Higgs mechanism. In section 3, we shall indicate the differences in the cosmological evolution implied by the two mechanisms, whenever they arise.

Hence the parameters of the system are: $M_{DM}$ the DM particle mass, $m_{V_D}$ the dark photon mass, $\alpha_D = g_D^2/4\pi$ the dark fine structure constant and $\epsilon$ the kinetic mixing parameter. We will restrict to the case in which $M_{DM} > m_{V_D}$, so that DM annihilations into two dark photons are always kinematically possible. As described in section 3.2, we will determine $\alpha_D$, with respect to $M_{DM}$ and $m_{V_D}$, by requiring that the observed DM density arises from the thermal freeze-out of the $X$ fermions in the hidden sector.

Dark matter annihilation proceeds either directly into a pair of dark photons (fig. 1 left)

$$X + \bar{X} \rightarrow 2V_D,$$  \hspace{1cm} (2)

or, in a two-step process, through the radiative formation of particle-antiparticle bound states,

$$X + \bar{X} \rightarrow B_s(X\bar{X}) + V_D,$$  \hspace{1cm} (3)

and the subsequent decay of the latter (fig. 1 middle and right)

$$B_{s\uparrow\downarrow}(X\bar{X}) \rightarrow 2V_D,$$  \hspace{1cm} (4a)

$$B_{s\uparrow\uparrow}(X\bar{X}) \rightarrow 3V_D.$$

(4b)

In the interaction (3), the subscript $s$ denotes the spin of the bound state that forms. Spin-singlet (para-) states, denoted by $\uparrow\downarrow$, form 25% of the time; spin-triplet (ortho-) states, denoted by $\uparrow\uparrow$, antiprotons and the Fermi galactic halo measurements. Finally, the parameter space of their analysis is somewhat different from ours. We consider a larger range of dark photon masses (while they focus on light mediators), and point out important implications from the decay of the cosmological abundance of heavier dark photons.
form 75% of the time.\textsuperscript{3} The spin of the bound state determines its dominant decay mode, namely into 2 $V_D$ or 3 $V_D$, as seen in (4). The decay rates of the bound states read $\Gamma_{\uparrow\downarrow} = \alpha_D^2 M_{DM}/2$ and $\Gamma_{\uparrow\uparrow} = [4(\pi^2 - 9)/(9\pi)] \alpha_D^4 M_{DM}/2$, so that within the parameter space of interest the decay lengths range from $\sim 10^{-4}$ pc to tenths of orders of magnitude below. Since this is rather prompt in astrophysical scales, the rate of radiative signals produced depends solely on the annihilation and BSF cross-sections. We discuss these cross-sections in section 2.1.

The dark photons $V_D$ produced in the processes (2) – (4) decay into SM particles via the kinetic mixing $\epsilon$, with the decay rate and branching ratios that we discuss in detail in section 2.2. While the dark photons produced in (2) and (4) carry energy of the order of the DM mass, the dark photon produced in the BSF process (3) carries away only the binding energy and the kinetic energy of the $X - \bar{X}$ relative motion, $\omega \simeq (M_{DM}/4)(\alpha_D^2 + v_{rel}^2) \ll M_{DM}$, where $v_{rel}$ is the relative velocity of the interacting pair in units of the speed of light $c$. Because of the larger astrophysical backgrounds at low energies, the constraints we derive in section 4 emanate solely from the decay products of the high-energy dark photons emitted in (2) and (4).

\subsection{Annihilation and bound-state formation cross-sections}

Bound states may form if the following conditions are met.

(i) The interaction carried by $V_D$ is effectively long-range, i.e. the inverse mass of the carrier is larger than the Bohr radius of the bound state

$$\frac{\alpha_D M_{DM}}{2 m_{V_D}} \gtrsim 0.84.$$  \hfill (5a)

\textsuperscript{3} In the non-relativistic regime, and to lowest order in the coupling $\alpha_D$, the spin of each of the interacting particles remains unchanged by the capture process into a bound state. The BSF cross-section is the same for any initial spin configuration, with the final spin configuration mandated to be the identical to the initial one. For an unpolared ensemble of unbound particles, the probability of capture into the singlet or triplet spin states thus depends solely on their multiplicities.
Bound states exist for \( m_{VD} \lesssim \alpha_D M_{DM}/2 \) and the Sommerfeld effect is significant. However, bound states can form radiatively only below the solid blue line, where the binding energy is sufficient for a massive dark photon to be emitted. Below the red dotted lines, the average momentum transfer is larger than the mediator mass, \( (M_{DM}/2) v_{\text{rel}} \gtrsim m_{VD} \), for typical velocities in Dwarf galaxies, the Milky Way and during chemical decoupling of DM from the dark photon bath in the early universe. (We use \( v_{\text{rel}} = \sqrt{2} v_0 \).) In this regime, the Coulomb approximation for the annihilation and BSF cross-sections is satisfactory.

This condition ensures the existence of at least one bound state, but also denotes the regime where the Sommerfeld effect is important.

(ii) The binding energy of the newly formed state is large enough to provide the mass of the emitted \( V_D \),

\[
\frac{\alpha_D^2 M_{DM}}{4 m_{V_D}} \gtrsim 1. \tag{5b}
\]

Here we neglect the kinetic energy of the interacting particles, since \( v_{\text{rel}} \lesssim \alpha_D \) wherever BSF is significant. The condition (5b) is always more stringent than (5a).

In fig. 2, we present the parameter space we focus on, and depict the conditions (5a) and (5b).

For our purposes, we shall consider only capture into the ground state, with principal and angular quantum numbers \( \left\{ nlm \right\} = \left\{ 100 \right\} \). Contributions from capture into higher-energy levels may give mild enhancements to the DM radiative signals, but are mostly subdominant either with respect to capture to the ground state, or with respect to annihilation, as explained in the
discussion that follows. Moreover, the radiative capture to $n > 1$ bound states is kinematically possible for a more limited range of $m_{\nu_D}$ than that defined by the condition (5b), by roughly a factor of $n^2$. In the following, any BSF cross-section presented in equations or graphs corresponds to capture to the ground state, unless otherwise stated.

The (spin-averaged) annihilation and BSF cross-sections can be expressed as follows [69]

$$\sigma_{\text{ann}} v_{\text{rel}} = \sigma_0 S_{\text{ann}}, \quad (6a)$$
$$\sigma_{\text{BSF}} v_{\text{rel}} = \sigma_0 S_{\text{BSF}} \times \frac{pss^{1/2}(3 - pss)/2}{\sqrt{\frac{M_{DM}(\alpha_{D}^2 + v_{\text{rel}}^2)}}{2}}, \quad (6b)$$
where $\sigma_0 \equiv \pi \alpha^2_D / M_{DM}^2$ is the perturbative value of the annihilation cross-section times relative velocity, and

$$pss \equiv 1 - 16m_{\nu_D}^2 / \left[ M_{DM}(\alpha_D^2 + v_{\text{rel}}^2)^2 \right]$$

is the phase-space suppression due to the emission of a massive dark photon in the capture process. Note that the factor $3 - pss$ accounts for the contribution from the longitudinal dark photon polarisation. For a massless dark photon, $pss^{1/2}(3 - pss)/2 = 1$.

In the Coulomb limit of a massless dark photon, $S_{\text{ann}}$ and $S_{\text{BSF}}$ depend only on the ratio $\alpha_D / v_{\text{rel}}$, and can be computed analytically (see e.g. [55]),

$$S_{\text{ann}}^C = \frac{2\pi \alpha_D / v_{\text{rel}}}{1 - e^{-2\pi \alpha_D / v_{\text{rel}}}}, \quad (8a)$$
$$S_{\text{BSF}}^C = \frac{2\pi \alpha_D / v_{\text{rel}}}{1 - e^{-2\pi \alpha_D / v_{\text{rel}}}} \frac{(\alpha_D / v_{\text{rel}})^4}{\left[ 1 + (\alpha_D / v_{\text{rel}})^2 \right]^2} \frac{2^9}{3} e^{-4(\alpha_D / v_{\text{rel}})} \arccot(\alpha_D / v_{\text{rel}}). \quad (8b)$$

For $v_{\text{rel}} > \alpha_D$, BSF is very suppressed. However, in the regime where the Sommerfeld effect is important, $v_{\text{rel}} \lesssim \alpha_D$, both the annihilation and BSF cross-sections exhibit the same velocity dependence, $\sigma v_{\text{rel}} \propto 1 / v_{\text{rel}}$, with BSF being the dominant inelastic process, $\sigma_{\text{BSF}} / \sigma_{\text{ann}} \simeq 3.13$. The Coulomb limit of eqs. (8) is a satisfactory approximation provided that the average momentum transfer exceeds the mediator mass [69],

$$(M_{DM}/2) v_{\text{rel}} \gtrsim m_{\nu_D}. \quad (9)$$

Note that this condition may hold, even if the phase-space suppression $pss$ in the case of BSF implies that the latter is suppressed or entirely disallowed.

For $m_{\nu_D} > 0$, $S_{\text{ann}}$ and $S_{\text{BSF}}$ cannot be computed analytically. A detailed computation of the radiative BSF cross-sections for DM interacting via a Yukawa potential, and comparison with annihilation has been recently performed in ref. [69] (see also [62]). Here, we summarise some features that are important for the DM phenomenology.

**Resonances.** Both the annihilation and the BSF cross-sections exhibit resonances, which emanate from the scattering (initial) state wavefunction, and occur at discrete values of the ratio of the Bohr momentum to the mediator mass, $\alpha_D M_{DM}/(2m_{\nu_D})$. These values correspond to the thresholds (maximum $m_{\nu_D}$ or minimum $\alpha_D M_{DM}$) for the existence of bound-state levels. Since for a Yukawa potential, the bound-state energy levels depend both on the principal and the angular quantum numbers, $n$ and $\ell$, so do the thresholds for their existence. Thus, each $\ell$-mode of the scattering state wavefunction exhibits a different spectrum of resonances that corresponds to the thresholds for the existence of bound-state levels of the same $\ell$.

---

4 In this regime, the capture into $n = 2, \ell = 1$ bound states, if kinematically allowed, is also somewhat faster than annihilation [69]. However, it is subdominant with respect to the capture to the ground state, and we ignore it in our analysis.
Angular momentum selection rules imply that (to leading order in $\alpha_D$ and $v_{rel}$) the $\ell$-modes of the scattering state wavefunction that participate in the two processes of interest, are different. The DM annihilation into two vector bosons (2) receives its dominant contribution from the $\ell = 0$ mode. On the other hand, the radiative formation of the ground state (3), receives its dominant
contribution from the $\ell = 1$ mode; this orbital angular momentum of the initial state provides for the spin of the vector boson emitted in the capture process into the ground state (which has zero orbital angular momentum). Thus, the DM radiative signals include both $\ell = 0$ and $\ell = 1$ resonances. The resonance structure of annihilation and BSF can be seen in fig. 3.

**Velocity dependence.** Besides the resonances, the $\ell$ modes of the scattering state wavefunction that participate in a process determine the velocity dependence of the corresponding cross-section away from the Coulomb regime. At sufficiently low velocities, $\sigma v \propto v^{2\ell}_{\text{rel}}$, therefore, the annihilation cross-section saturates at a velocity-independent value, while the cross-section for capture into the ground state becomes $v^{2\ell}_{\text{rel}}$ suppressed. This behaviour holds both on- and off-resonance. The departure from the Coulombic behaviour occurs at $v_{\text{rel}} \lesssim 2m_{\nu_{\text{DM}}}/M_{\text{DM}}$. There is an important difference though, between resonant and non-resonant points. For non-resonant $\alpha_{\nu_{\text{DM}}}(2m_{\nu_{\text{DM}}})$ values near or on-resonance, the $\sigma v \propto v^{2\ell}_{\text{rel}}$ behaviour is established at $v_{\text{rel}} \lesssim 2m_{\nu_{\text{DM}}}/M_{\text{DM}}$, and the annihilation and BSF cross-sections remain always below their Coulomb values at the same speed. In contrast, for $\alpha_{\nu_{\text{DM}}}(2m_{\nu_{\text{DM}}})$ values near or on-resonance, the cross-sections grow above their Coulomb value at $v_{\text{rel}} \lesssim 2m_{\nu_{\text{DM}}}/M_{\text{DM}}$, and acquire the $\sigma v \propto v^{2\ell}_{\text{rel}}$ behaviour at much lower velocities. The above imply that BSF is comparable to, or stronger than annihilation within a range of velocities that depends on $\alpha_{\nu_{\text{DM}}}(2m_{\nu_{\text{DM}}})$. Away from resonances, this speed range is essentially limited within the Coulomb regime and by the onset of the Sommerfeld effect ($2m_{\nu_{\text{DM}}}/M_{\text{DM}} \lesssim v_{\text{rel}} \lesssim \alpha_{\nu_{\text{DM}}}$); however, it is extended to significantly lower velocities, on or near resonances \[69\].

For our purposes, BSF is significant in Dwarf galaxies, in the Milky Way and during the chemical decoupling of DM in the early universe, within large portions of the parameter space where the radiative capture is kinematically possible. In fig. 2, we circumscribe the parameter space where the Coulomb regime is attained, for the corresponding typical DM speeds (see below); in these regimes, the analytical formulae (8) may be used (although the phase-space factor appearing in eq. (6b) should still be properly taken into account). The transition from the Coulomb regime to the resonant regime, with increasing $m_{\nu_{\text{DM}}}$ or decreasing $M_{\text{DM}}$, can be seen in fig. 3. For the range of dark photon masses considered, BSF is unimportant during CMB, when the DM average speed is extremely low (c.f. section 4.3).

Before moving on, let us specify here for reference the typical speed distributions that we adopt. For the Milky Way and for dwarf galaxies, we consider a Maxwellian distribution with a cutoff at $v = v_{\text{esc}}$:

$$f(v) = N(v_0, v_{\text{esc}}) \Theta(v_{\text{esc}} - v) e^{-v^2/v_0^2}$$

The normalization factor $N(v_0, v_{\text{esc}})$ is chosen such that $\int d^3v f(v) = 1$. For the Milky Way, we take $v_0 = 220$ km/s and $v_{\text{esc}} = 533$ km/s, which are appropriate values for the DM phase-space distribution in our galaxy (see e.g. \[76\]). For dwarfs, we take $v_0 = 7$ km/s and $v_{\text{esc}} = 12$ km/s, which are typical values inferred from stellar tracers (see e.g. \[77\]). For the CMB epoch, see the dedicated discussion in section 4.3.

\[5\] Our discussion suggests that away from the Coulomb regime, the capture into the $n = 2, \ell = 1$ bound states may become the dominant inelastic process. This is because the angular momentum selection rules imply that the $\ell = 0$ and $\ell = 2$ modes of the scattering state wavefunction participate in this process. At low velocities, due to the $v^{2\ell}_{\text{rel}}$ scaling described above, the contribution from the $\ell = 0$ mode renders the capture into $n = 2, \ell = 1$ states more significant than the capture into the ground state. However, the contribution from this mode alone is subdominant to the annihilation. We conclude that the formation of $n = 2, \ell = 1$ states can be neglected, either with respect to the capture to the ground state (Coulomb limit, c.f. footnote 4), or with respect to annihilation (except perhaps very close to $\ell = 2$ resonances, which however cover very limited parameter space).
2.2 Dark photon decay

The kinetic mixing with hypercharge of eq. (1) implies that the dark photon mixes with both the SM photon and $Z$. This induces the following couplings between $V_D$ and the SM fermions $f$

$$\mathcal{L} \supset g_f V_D^\mu (f \gamma_\mu f), \quad g_f = e e \left( \frac{1}{1 - \delta^2} + \frac{Y}{2 \delta^2} \right) + O(\epsilon^2), \quad (10)$$

where $\delta = m_{V_D}/m_Z$ and $Y(\ell_L, \ell_R, u_L, \ldots) = -1/2, -1, 1/6, \ldots$. Equation (10) makes it manifest that the $V_D$ couplings to SM particles are $\epsilon$-suppressed, and that they are proportional to their electric charge for $m_{V_D} \ll m_Z$, and to their hypercharge for $m_{V_D} \gg m_Z$ (i.e. when $U(1)_Y$ is unbroken). Of course, the $\epsilon$ expansion of eq. (10) is not valid for $\delta \sim 1$. In that limit, the $V_D$ couplings depend on the $f$ quantum numbers as the $Z$ boson ones. We refer the reader to [78, 79] for the full $g_f$ expressions.

We compute the widths of $V_D$ at tree level\(^6\) as

$$\Gamma(V_D \to f \bar{f}) = \frac{N_f}{24 \pi} m_{V_D} \sqrt{1 - \frac{4 m_f^2}{m_{V_D}^2}} \left[ g_{fL}^2 + g_{fR}^2 - \frac{m_f^2}{m_{V_D}^2} (g_{fL}^2 + g_{fR}^2 - 6 g_{fL} g_{fR}) \right]$$

where $N_f = 1$ for $f = \ell, \nu$ and $N_f = 3$ otherwise. We do not report the widths into $WW$ and $Zh$, that even for $m_{V_D} > 2 m_W, m_t + m_Z$ contribute less than $\sim 10\%$ to the total $V_D$ width. In the interval $\Lambda_{QCD} < m_{V_D} < 5$ GeV we cannot treat the light quarks as free particles, because QCD is strongly coupled and decays should be described in terms of QCD resonances. For example, in \textsc{Pythia}, which is the tool we use to produce all our spectra (c.f. section 4), the decay $\tau^- \to \nu_\tau \pi^-$ is treated by an effective $\nu_\tau \tau$ vertex (see \textit{e.g.} [80]). We choose 5 GeV as the upper extreme of the interval dominated by hadronic decays, because \textsc{Pythia} has not been optimised for energies below that value, and we set the lower extreme to the reference value of 350 MeV. As customary (see \textit{e.g.} [79, 81] for recent works), we determine the hadronic decay width from measurements of $e^+ e^- \to$ hadrons. This process is dominated by $\gamma^*$ exchange, and therefore provides a good handle to describe the hadronic decays of our spin-1 $V_D$ initial state. The related $V_D$ decay width then reads

$$\Gamma(V_D \to \text{hadrons}) = R(s = m_{V_D}^2) \Gamma(V_D \to \mu^+ \mu^-), \quad (12)$$

where we extract $R(s) = \sigma(e^+ e^- \to \text{hadrons})/\sigma(e^+ e^- \to \mu^+ \mu^-)$ from [79], with $\sqrt{s}$ center-of-mass energy of the $e^+ e^-$ collision.

To our knowledge, no tool has been developed that would allow to obtain, from the hadrons produced for $\Lambda_{QCD} < m_{V_D} < 5$ GeV, the energy spectra of stable SM particles that we need in deriving our bounds (namely $\gamma$-rays, $e^\pm$, $\tilde{p}$). However, the description can be simplified by observing that $\pi^+ \pi^-$ pairs largely dominate the final states from the decays of the resonances produced in $V_D \to \text{hadrons}$ (see \textit{e.g.} [81–83]). Since $\text{BR}(\pi^\pm \to \mu^\pm \nu_\mu) > 99.9\%$, the final states from those hadrons are then dominated by $\mu^+ \mu^-$ and $\nu_\mu \bar{\nu}_\mu$ pairs, up to a very small fractions of other SM particles, like photons from $\pi_0$ decays. Therefore we employ the following approximation: we assume that half of the hadronic width consists in the $\mu^+ \mu^-$ final state, and the other half in the $\nu_\mu \bar{\nu}_\mu$ one. The resulting energy distribution of final muons is then less broad, and peaked at somehow larger energies, than the one that would arise from following properly all the steps of the various hadronic cascades (see \textit{e.g.} ref [84], which discusses the analogous case of multistep cascades of a hidden sector). We believe our approximation to be nonetheless sufficient for our

\(^6\)Loop corrections do not affect substantially this result, as can be seen comparing our values with the ones given in ref. [79], where such corrections have been taken into account.
Figure 4: Branching ratios of the dark photon into SM fermions.

purposes. On one hand, the interval of $m_{V_D}$ in which these subtleties are relevant is only a portion of the much larger range which we consider. On the other hand, even within this interval, i) the CMB constraints are not affected, because they are not sensitive to the number of steps of a cascade [84]; ii) our constraints may be taken as possibly slightly aggressive for $M_{DM} \lesssim 100$ GeV, since enhancing the lower energy $\gamma$’s from $\mu$ showers, at the expense of higher energy $\gamma$’s, would typically bring the resulting spectrum below the Fermi sensitivity; iii) the treatment chosen affects mildly a region that is anyway excluded by CMB.

To summarise, we compute the $V_D$ branching ratios to SM particles at tree level outside the ‘hadronic decays’ interval $350$ MeV $< m_{V_D} < 5$ GeV, where instead we extract the light hadrons width from measurements of $e^+e^- \rightarrow$ hadrons at colliders. We then assume this width to consist in half $\mu^+\mu^-$ and half $\nu_\mu\bar{\nu}_\mu$ pairs. The resulting BRs, that we use in the rest of this paper, are displayed in fig. 4.

2.3 Constraints on the kinetic mixing

2.3.1 Supernovae and beam dump experiments

We summarize here the constraints on the kinetic mixing $\epsilon$ that are independent of the existence of any dark sector state other than $V_D$, focusing on the region $\epsilon \lesssim 10^{-3}$. We refer the reader to [85, 86] for recent reviews.

Supernovae (SN). Light dark photons with a mixing parameter in the approximate range $\epsilon \simeq 10^{-10} - 10^{-6}$ are constrained by the neutrino spectrum observed [87–89] after the explosion of supernova SN1987A, see [90–94] for recent works. According to the standard argument [95], the cooling of the SN core becomes more efficient if enough SM photons from the explosion oscillate into $V_D$ (which requires a minimum value of $\epsilon$), and if enough $V_D$ escape the supernova without further interacting nor decaying (which implies a maximum value of $\epsilon$). Supernovae observations can thus exclude a finite band of $\epsilon$ values. The most updated analyses have been carried out in ref. [93, 94], where the plasma effects of finite temperature and density have been included for
the first time. The results of the two analyses appear to be in reasonable agreement, and for definiteness we show the fiducial exclusion from [93] as a red shaded region in fig. 5. To give a sense of possible theoretical errors, we also show with a thicker red line the contours of the region that the authors of [93] name ‘robustly excluded’, and that they find to be disallowed irrespectively of the details of their modelling. Finally, we show in thin red the contour of the region that is excluded by the analysis of ref. [91], from the non-observation of a photon excess at the SMM Gamma-Ray Spectrometer [96], that would be caused by the dark photons escaping the inner part of the SN. Since the latter bounds do not include potentially important effects [93, 94] we do not shade the related area, in order to be conservative.

**Beam dump experiments.** They consist of a high-intensity beam of particles sent on a fixed target, and a detector placed somewhat far from the target. They can constrain dark photons in the MeV to GeV range, for \( \epsilon \sim 10^{-8} - 10^{-2} \). When an electron beam hits the target, dark photons can be produced via bremsstrahlung, pass a shield that screens the SM background, and be detected via the leptons they decay into. The events expected from the dark photons have been computed in refs. [97, 98], and the limits derived from the E137 [99], Orsay [100], E141 [101] and E774 [102] data are shown as a gray shaded area in fig. 5. When a proton beam hits the target, dark photons can be produced both directly from the proton-target scattering, and from the decays of hadrons produced in the same scattering, as originally envisioned in ref. [103]. Detectors can then collect leptons from the decays of those dark photons that have travelled long enough. Neutrino experiments fit into this category [104]. Among them, the reinterpretation of LSND [105–107], CHARM [108, 109] and \( \nu \)-CAL I (at the U70 accelerator) [110, 111] data allows to exclude the gray shaded areas in fig. 5.

### 2.3.2 Direct Detection

Direct detection aims at revealing the tiny nuclear recoils produced by DM particles scattering off target nuclei in underground experiments. In the non-relativistic limit, the differential cross section can be computed from the interaction Lagrangian in eq. (10). For \( \delta \to 0 \), it reads [112, 113]

\[
\frac{d\sigma}{dE_R}(v, E_R) = \frac{8\pi \alpha_{em}\alpha D^2 m_T}{(2m_T E_R + m_{\nu_D}^2)^2} \frac{1}{v^2} Z_T^2 F_{\text{Helm}}^2(2m_T E_R),
\]

(13)

where \( m_T \) and \( Z_T e \) are the mass and the electric charge of the target nucleus \( T \) respectively, and \( \alpha_{em} = e^2/4\pi \) is the fine structure constant. Here, \( F_{\text{Helm}} \) is the Helm form factor [114, 115] related to the charge density of the nucleus.

As is evident, eq. (13) exhibits two regimes: i) for \( m_{\nu_D} \gg 2m_T E_R \), the interaction is contact-type, and for a fixed value of \( M_{DM} \) the differential cross section scales like \( \epsilon^2/m_{\nu_D}^2 \); ii) for \( m_{\nu_D} \ll 2m_T E_R \), the interaction manifests as long-range, and the cross section is independent of \( m_{\nu_D} \). For a given \( M_{DM} \), the differential cross section is simply proportional to \( \epsilon^2 \). Considering typical target nuclei (\( m_T \simeq 100 \text{ GeV} \)) and recoil energies (\( E_R \simeq 1 \text{ keV} \)) in direct searches, the transition regime occurs for \( m_{\nu_D} \simeq 15 \text{ MeV} \). Recent analysis of the phenomenology in direct detection of DM models with long-range DM-nucleus interactions can be found e.g. in [116–119].

Having at our disposal the scattering cross section, the rate of nuclear recoils, expressed in cpd

---

\(^7\) This is well justified, since as one can see in the right-panel of Fig. 5, the dark photon mass probed by direct detection experiments is much smaller than \( m_Z \) for the small values of \( \epsilon \) we are interested in.
Figure 5: **Constraints on the kinetic mixing.** Left: excluded regions in the plane \((m_{V_D}, \epsilon)\), taking into account several beam dump experiments (gray shaded areas), supernovæ (red shaded areas), direct detection (green shaded areas, for different indicative values of the DM mass) and BBN arguments (blue shaded area). See the text for more details. Right: the same excluded regions illustrated in the \((M_{DM}, m_{V_D})\) plane, for three selected values of the mixing parameter \(\epsilon\).

(counts per day) per kilogram per keV, is then

\[
\frac{dR_T}{dE_R} = \frac{\xi_T}{m_T M_{DM}} \rho_\odot \int_{v_{\text{min}}(E_R)}^{v_{\text{esc}}} d^3v f_{E}(\vec{v}) \frac{d\sigma}{dE_R}(v, E_R),
\]

where \(\xi_T\) is the mass fraction of different nuclides used in direct detection, \(\rho_\odot = 0.3 \text{ GeV/cm}^3\) is the local DM energy density and \(f_{E}(\vec{v})\) is the DM speed distribution in the Earth’s frame. In the velocity integral, \(v_{\text{min}}(E_R)\) is the minimal speed for which DM particles can provide a given recoil energy \(E_R\). \(v_{\text{esc}}\) refers of course to the Milky Way’s escape speed.

Since the cross section in eq. (13) is independent of the nucleus spin, we derive bounds on the relevant parameter space of the model considering the most sensitive detectors for spin-independent DM-nucleus scattering. More specifically, for \(M_{DM} \gtrsim 3 \text{ GeV}\) we use the latest LUX WS2014-16 run [120] to set limits, while for \(M_{DM} \lesssim 3 \text{ GeV}\) we derive bounds from the Cresst-II experiment [121]. We follow the procedure of [122] described in Appendix B.2 (Lux) and B.3 (Cresst-II), where the reader can find more details.

Figure 5 shows, in green shaded areas, the excluded regions of the parameter space. In the left panel they are reported in the \((m_{V_D}, \epsilon)\) plane, with the different lines (dotted, dashed-dotted, dashed) corresponding to \(M_{DM} = 1, 10^2, 10^5 \text{ GeV}\) respectively. In particular, for \(M_{DM} = 1 \text{ GeV}\) the exclusion is given by the Cresst-II detector, while for \(M_{DM} \gtrsim 3 \text{ GeV}\) is dominated by Lux. For light dark photons, the constraints become independent of \(m_{V_D}\). As pointed out above, this is because the interaction is entering the long-range regime. For Lux this transition occurs for \(m_{V_D} \approx 20 \text{ MeV}\), while for Cresst-II around 1 MeV, since the detector is composed by lighter target nuclei (oxygen and calcium) and employs a lower energy threshold (around 300 eV) compared to xenon-based experiments, such as Lux.
In the right panel of the same figure, we project the exclusion regions in the \((M_{DM}, m_{VD})\) plane for three fixed values of \(\epsilon\). The different lines (dashed, dotted, solid) correspond to \(\epsilon = 10^{-6}, 10^{-8}, 10^{-10}\). For \(\epsilon \lesssim 10^{-7}\) the bounds come solely from LUX, since the CRESST-II detector is not able to probe this region of the parameter space, mainly due to its lower exposure with respect to LUX. For larger \(\epsilon\), CRESST-II becomes relevant for light DM \((M_{DM} \lesssim 3 \text{ GeV})\).

### 2.3.3 Summary

We have illustrated the constraints of this section in fig. 5, adding for completeness the region disfavoured by BBN and discussed in section 3.3. We see in the left panel that, for \(m_{VD} \gtrsim 1\) GeV, a wide region is still unexplored by current measurements, and an upper limit on \(\epsilon\) is provided only by direct detection experiments (in the \(\epsilon\) region we are considering). Lower dark photon masses are more constrained, and only narrow values of \(\epsilon\) are allowed. We notice however that, for \(\epsilon \sim 10^{-(10-12)}\), a more careful study of BBN constraints (and perhaps of SN bounds) would be needed to make any conclusive statement about the low \(m_{VD}\) region (see the discussion in section 3.3). Finally, the same constraints are shown in the right-hand plot of fig. 5 on the \((M_{DM}, m_{VD})\) plane, for three fixed values of \(\epsilon\).

### 3 Cosmology

#### 3.1 Brief history of the dark sector

We shall assume that some unspecified high-energy interactions brought the dark sector plasma in thermal equilibrium with the SM particles at some high temperature, and that these interactions decoupled early, leaving the dark sector at the same temperature as the SM plasma, \(\tilde{T} = \tilde{T}_{SM} = \tilde{T}_{D}\), when all of the SM degrees of freedom where kinetically coupled to the photons.\(^8\) We also assume that at that common temperature \(\tilde{T}\), DM and the dark photon provided the only relativistic degrees of freedom in the dark sector. Beyond this point, the temperature of the dark photons \(T_{D}\) is in general different from the temperature of ordinary photons \(T_{SM}\). However, \(T_{D}\) (and the DM temperature \(T_X\), whenever different) can be computed with respect to \(T_{SM}\) by following the cosmology of the two sectors. This is the goal of this subsection. Determining \(T_{D}\) and \(T_X\) is important for the computation of the DM relic density, the estimation of dark photon cosmological abundance and the associated BBN constraints, and for the late-time DM annihilation and the resulting CMB constraints. We summarise some important events that determine the late-time cosmology in table 1.

**Dark photons.** The dark photons are the last particles in the dark sector to become non-relativistic. So long as the dark sector contains relativistic species, \(T_D\) can be easily computed with respect to \(T_{SM}\) by invoking the conservation of the comoving entropy in each sector separately. As is standard, we find

\[ r = \frac{T_D}{T_{SM}} = \left[ \frac{g_{SM}(T_{SM})}{g_D(T_D)} \right]^{1/3} \left( \frac{\tilde{g}_D}{\tilde{g}_{SM}} \right)^{1/3}, \tag{15} \]

\(^8\) The two sectors also exchange energy via elastic scatterings of DM with the SM charged fermions, via the kinetic mixing of \(U(1)_D\) with the hypercharge. However, these scatterings fail to equilibrate the two sectors if \(\epsilon^2 \alpha_D \lesssim 10^{-20}(M_{DM}/m_e)^3\), where we assume that DM is the lightest species charged under \(U(1)_D\) (see e.g. [57]). This condition is typically weaker than the constraints imposed by direct detection and other experiments (c.f. section 2.3).
The dark plasma is in thermal equilibrium with the SM plasma.

The annihilation processes $XX \leftrightarrow 2V_D$ freeze-out; the DM density departs from its equilibrium value.

The energy transfer via $X V_D \leftrightarrow X V_D$ becomes inefficient. DM kinetically decouples from the dark photons.

The DM temperature transitions from $T_X = T_D \propto 1/a$ to $T_X \propto 1/a^2$ scaling. ($a$: scale factor)

The dark photons decay.

The dark photons would dominate the energy density of the universe, had they not decayed.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T \gtrsim \bar{T} = \bar{T}<em>D = \bar{T}</em>\text{SM}$</td>
<td>The dark plasma is in thermal equilibrium with the SM plasma.</td>
</tr>
<tr>
<td>$T^{\text{f.o.}}<em>D \approx M</em>{\text{DM}}/30$</td>
<td>The annihilation processes $XX \leftrightarrow 2V_D$ freeze-out; the DM density departs from its equilibrium value.</td>
</tr>
<tr>
<td>$T^{\text{k.d.}}_D$</td>
<td>The energy transfer via $X V_D \leftrightarrow X V_D$ becomes inefficient. DM kinetically decouples from the dark photons.</td>
</tr>
<tr>
<td>$T^{\text{trans}}_D$</td>
<td>The DM temperature transitions from $T_X = T_D \propto 1/a$ to $T_X \propto 1/a^2$ scaling. ($a$: scale factor)</td>
</tr>
<tr>
<td>$T^{\text{decay}}_D$</td>
<td>The dark photons decay.</td>
</tr>
<tr>
<td>$T^{\text{dom}}_D$</td>
<td>The dark photons would dominate the energy density of the universe, had they not decayed.</td>
</tr>
</tbody>
</table>

Table 1: **Important cosmological mileposts in the evolution of the dark sector.** The subscripts $D$ and $X$ denote the temperatures of the dark photons and DM, respectively.

where $g_{\text{SM}}$ and $g_D$ stand for the relativistic degrees of freedom in the SM sector and the dark sector respectively, and $\bar{g}$ refers to their values at the common temperature $\bar{T}$. After the dark photons become non-relativistic, at $T_D \approx \frac{m_{V_D}}{3}$, their momentum redshifts as $p \propto 1/a$, where $a$ is the scale factor of the universe, and their distribution resembles a thermal one with temperature that now scales as $T_D \propto p^2 \propto 1/a^2$. Since no interactions that can change the dark photon number are in equilibrium, the dark photons develop a non-zero chemical potential. Both their comoving number and entropy are conserved.\(^9\)

**Dark matter.** Early on, DM is in kinetic and chemical equilibrium with the dark photons, via elastic and inelastic scatterings, $X V_D \leftrightarrow X V_D$ and $X X \leftrightarrow V_D V_D$. The DM density follows (approximately) an equilibrium distribution, with temperature $T_X = T_D$. At temperature $T^{\text{f.o.}}_D \approx M_{\text{DM}}/30$, the inelastic scatterings fail to keep DM in chemical equilibrium with the dark photons, and the DM density freeze-out. (However, the full chemical decoupling of $X$ and $V_D$ may occur at a somewhat later time, c.f. section 3.2.) After freeze-out, DM remains in kinetic equilibrium with the dark photons, via elastic scatterings. It continues to have an equilibrium density with $T_X = T_D$, albeit it develops a non-zero chemical potential. This remains true, up until the time of kinetic decoupling of DM with the dark photons, at a temperature $T^{\text{k.d.}}_D$ that we shall estimate below. After the kinetic decoupling, the momentum of the DM particles redshifts as $p \propto 1/a$. Their population continues to resemble an equilibrium distribution, with a chemical potential that evolves to reflect the conservation of the particle number, and a temperature that scales as $T_X \propto p^2 \propto 1/a^2$ (and departs from the temperature of the dark photons, if the latter remain relativistic at that time).

Putting the above considerations together, we conclude that the DM temperature is

$$T_X \approx \begin{cases} r(T_{\text{SM}}) \ T_{\text{SM}}, & \text{for } T_D > T^{\text{trans}}_D, \\ \frac{(1+z)^2}{(1+z^{\text{trans}})^2} \ T^{\text{trans}}_D, & \text{for } T_D < T^{\text{trans}}_D \end{cases}$$  \quad (16)

---

\(^9\) We recall that for a non-relativistic species of mass $m$, with non-zero chemical potential $\mu$, at temperature $T$, the number and entropy densities are $n = g[mT/(2\pi)^3]^{1/2} \exp[\mu - m]/T$ and $s = n(m - \mu)/T$. The scaling $T \propto 1/a^2$ and $(\mu - m) \propto T$ ensures the conservation of the comoving number and entropy: $n, s \propto 1/a^3$.  


where the temperature $T_D^{\text{trans}}$ marks the transition from the $T_X \propto 1/a$ to the $T_X \propto 1/a^2$ scaling, that occurs either when DM decouples kinetically from the dark photons, or when the dark photons become themselves non-relativistic, whichever happens first,

$$T_D^{\text{trans}} \equiv \max[T_D^{\text{k,d}}, m_{V_D}/3].$$ (17)

In eq. (16), $z$ denotes the redshift. We determine $z_{\text{trans}}$ from the conservation of entropy in the SM sector, $g_{\text{SM}} T_{\text{SM}}^{3}(1 + z)^{-3} = g_0 T_0^3$, where the index “0” refers to today’s values, as typical. Using also eq. (15), we find

$$1 + z_{\text{trans}} = (T_D^{\text{trans}}/T_0) \times (g_D^{\text{trans}}/g_{\text{SM}})^{1/3}.$$ (18)

We shall use eqs. (16) and (18) in section 4.3, to obtain constraints from the DM annihilation around CMB. We determine $T_D^{\text{trans}}$ next.

**Kinetic decoupling and transition temperature.** The DM and the dark photons are kept in kinetic equilibrium via the elastic scatterings $X V_D \leftrightarrow X V_D$. Assuming that the dark photons are relativistic, this process is described by the Thomson cross-section $\sigma_T = 8\pi \alpha_D^2/(3M_{\text{DM}}^2)$. The energy transfer rate is $d\rho/dt \sim (\delta E_1) n_x n_{V_D} \sigma_T$, where $n_x$ is the density of the DM particles, $n_{V_D} = [(3/\pi^2) g_{V_D} T_D^3]$ is the number density of the dark photons, and $\delta E_1 \approx T_D^2/(2M_{\text{DM}})$ is the average energy transfer per collision (with $\delta p_1 \sim T_D$ being the average momentum transfer). The kinetic decoupling occurs when $\rho_{X,\text{kin}}(d\rho/dt) < H$, where $\rho_{X,\text{kin}} \sim (3/2) T_D n_x$ is the DM kinetic energy density, and $H$ is the Hubble parameter. Putting everything together, we find that the kinetic decoupling occurs at $T_D^{\text{k,d}} \approx 2(M_{\text{DM}}/M_{\text{Pl}})^{1/2}/(\alpha_D r_{\text{k,d}})$. Our calculation breaks down if the estimated temperature is $T_D^{\text{k,d}} \lesssim m_{V_D}/3$. Since we are interested in $T_D^{\text{trans}}$ rather than $T_D^{\text{k,d}}$, we need not repeat the calculation for non-relativistic dark photons. We conclude that

$$T_D^{\text{trans}} \approx \max \left[ \frac{2M_{\text{DM}}^{3/2}}{r_{\text{trans}} \alpha_D M_{\text{Pl}}^{1/2}}, \frac{m_{V_D}}{3} \right],$$ (19)

where we estimate $r_{\text{trans}} \sim 0.6 - 1$ using eq. (15). (For a more precise treatment of the kinetic equilibrium and decoupling, see ref. [123].)

**Dark photon mass generation.** In the discussion above, we have not been concerned about the generation of the dark photon mass. If $m_{V_D}$ is generated via the St"{u}ckelberg mechanism, then the dark photon is massive throughout the cosmological history. However, if the dark photon obtains its mass via a dark Higgs mechanism, then it becomes massive only after a $U(1)_D$ - breaking phase transition takes place. We estimate the temperature of the latter to be $T_D^{\text{b,t}} \sim v_D = m_{V_D}/\sqrt{8\pi \alpha_D}$, where $v_D$ is the vacuum expectation value of dark Higgs. We see that for the entire range of $\alpha_D$ considered, $T_D^{\text{b,t}} \gtrsim m_{V_D}/3$, i.e. the dark photons are at least quasi-relativistic after the phase transition, and our considerations are left unaffected.

### 3.2 Relic abundance

We determine the dark fine structure constant $\alpha_D$ by requiring that the observed DM abundance is attained after the thermal freeze-out of DM from the dark photon plasma. For large $m_{V_D}$, the interaction between DM and dark photons is contact-type, and $\sigma_{\text{ann}} v_{\text{rel}} \simeq \sigma_0$ is velocity
Figure 6:  
Left: The dark fine structure constant $\alpha_D$ is determined by the thermal freeze-out of DM in the dark sector. The dotted gray line takes into account annihilation only, while the dashed gray line incorporates both annihilation and bound-state formation. The overprediction of $\alpha_D$ if BSF is neglected is (10 – 55)% for DM masses (3 – 100) TeV. Both calculations have been performed in the Coulomb limit, which is a good approximation during freeze-out, except for the phase-space suppression due to a dark photon emission in the case of BSF. Thus, for non-zero dark photon masses, we interpolate between the two calculations near the threshold where BSF becomes kinematically possible, as shown by the coloured solid lines. 
Right: Implications of BSF during freeze-out, for the DM indirect detection signals today. We show the total cross-section (annihilation plus BSF), averaged over Milky Way velocities. The dotted lines correspond to $\alpha_D$ determined by considering only annihilation during freeze-out. The solid lines correspond to $\alpha_D$ determined via the full calculation that includes both annihilation and BSF. The effect is two-fold: A lower $\alpha_D$ implies lower overall rates for the expected signals today, and shifts the resonances to larger values of the DM mass.

independent. As is standard, the predicted value for $\sigma_0$ is then essentially independent of the DM mass,\(^\text{10}\) thus setting $\alpha_D \propto M_{DM}$ (as it trivially follows from the definition of $\sigma_0$ in sec. 2.1). However, in the parameter space where the condition (5a) is satisfied, the annihilation processes are Sommerfeld enhanced. This affects significantly the DM relic density for $M_{DM} \gtrsim$ TeV [32]. More recently, it was shown that, in the same mass range, the formation and decay of particle-antiparticle bound states depletes significantly the DM abundance [54], thereby reducing the predicted $\alpha_D$ further.

We follow the analysis of ref. [54], to determine the dark fine structure constant $\alpha_D$. We perform our calculations in the Coulomb regime, using eqs. (8) to determine the annihilation and BSF cross-sections, and we verify a posteriori that this is a satisfactory approximation (c.f. appendix A). We compute $\alpha_D$ in two different cases: first, considering the DM direct annihilation only, and then considering both direct annihilation and bound-state formation and decay. The latter requires solving a set of Boltzmann equations that capture the interplay of bound-state formation, ionisation and decay processes [54]. However, for a massive dark photon, BSF is not always

\(^{10}\)A step-like variation at $M_{DM} \lesssim 10$ GeV arises due to the decoupling of the QCD degrees of freedom during freeze out (see e.g. [124]).
kinematically allowed. To account for this kinematic cutoff and the phase-space suppression, in our calculations of the various DM interaction rates throughout our work, we adopt the former computation of $\alpha_D$ when $p_{s\text{SS}}^{1/2}(3-p_{s\text{SS}})/2 < 0.5$, and the latter when $p_{s\text{SS}}^{1/2}(3-p_{s\text{SS}})/2 > 0.5$, where $p_{s\text{SS}}$ is defined in eq. (7) [see also eq. (6b)]. Figure 6 (left panel) shows $\alpha_D$ as a function of the DM mass.\(^{11}\)

Note that $\alpha_D$ may be lower than estimated here, if in the early universe the dark sector was at a significantly lower temperature than the SM plasma. This would imply overall more relaxed direct and indirect detection constraints, as well as weaker DM self-interactions. However, a significantly colder dark sector requires either a large number of yet unknown particles with sizeable couplings to the SM, whose cosmological decoupling increased the entropy of the SM plasma after its decoupling from the dark sector, or appropriate initial conditions set by inflation and absence of any interactions that would equilibrate the two sectors. We do not explore this possibility further in this paper.

**Phenomenological implications of the BSF effect on the DM density.** Neglecting BSF in the determination of the DM relic density overpredicts $\alpha_D$ by about $(15 - 55)\%$ in the DM mass range $(5 - 100)$ TeV. This overestimates the direct detection constraints on the DM-nucleon cross-section, and consequently on $\epsilon^2$, by the same amount. The effect on the DM annihilation and BSF cross-sections relevant for indirect searches and CMB constraints, is greater. In the Coulomb regime, which pertains to much of the parameter space of the model for indirect searches in the Milky Way and the Dwarfs (c.f. fig. 2), $\sigma_{\text{ann, BSF}} \propto \alpha_D^3$. Then, the overestimation of the expected indirect detection signals ranges from 50% up to a factor of $\sim 4$, in the same DM mass interval. A somewhat milder but still significant effect is expected away from the Coulomb regime, at non-resonant points, where $\sigma_{\text{ann, BSF}} \propto \alpha_D^2$. The discrepancy is also stark at resonant points, which appear at discrete values of $\alpha_D M_{\text{DM}}/(2m_{V_D})$, as described in section 2.1; the overestimation of $\alpha_D$ implies that the resonances would be expected to occur at lower $M_{\text{DM}}$ values. These effects are depicted in fig. 6 (right panel).

### 3.3 BBN constraints

The cosmological abundance of dark photons freezes-out essentially at the time of DM freeze-out, when the interaction that can alter their number become inefficient. After the dark photons become non-relativistic, they may dominate the energy density of the universe, and cause cosmological problems. In particular, they may inject a large amount of entropy if they decay around or after BBN, and/or shift the time of matter-radiation equality. Here, we estimate the constraints implied by these considerations.

Assuming they were relativistic at the time of DM freeze-out ($m_{V_D}/3 \lesssim T_{\text{f.o.}}^{D,0} \approx M_{\text{DM}}/30$), the ratio of the dark photon number density to SM entropy density was

$$f_{V_D} = \frac{n_{V_D}}{s_{\text{SM}}} \approx \frac{45\zeta(3)}{2\pi^4} \frac{g_{V_D} r_{\text{f.o.}}^3}{g_{\text{SM}}^{f.o.}} = \frac{45\zeta(3)}{2\pi^4} \frac{g_D^f}{g_{\text{SM}}}, \quad (20)$$

where $r_{\text{f.o.}}$ is the dark-to-ordinary temperature ratio at the time of DM freeze-out, computed according to eq. (15), and we used $g_{V_D}^{f.o.} = g_{V_D}$. From eq. (20), we estimate $f_{V_D} \sim 0.02$. In the

\(^{11}\) While in fig. 6 we present our results only for $M_{\text{DM}} \gtrsim 100$ GeV, we carry out our computations for the entire mass range considered in this work, $1$ GeV $\lesssim M_{\text{DM}} \lesssim 100$ TeV. We note that we improve with respect to the computation of ref. [54], by taking into account the variation of the dark-to-ordinary plasma temperature ratio, due to the decoupling of the SM degrees of freedom, according to eq. (15). This is particularly important for low DM masses.
Figure 7: **Regions disfavoured by BBN.** Outside the shaded area, the dark photons either decay safely before BBN ($\tau_{V_D} \lesssim 0.03\,\text{s}$ above border to the right), or their cosmological abundance is not significant (above border to the left). The disfavoured area is a rough estimate only (see text for discussion). For large $m_{V_D}$ and small $\epsilon$, the decay of the dark photons injects significant entropy in the universe and dilutes DM. The solid orange lines indicate the estimated dilution factor. In this regime, $\alpha_D$ should be significantly lower than estimated, in order for the dark fermions to account for the observed DM abundance. This would relax the CMB and indirect detection constraints.

The absence of decay, the dark photons are set to exceed a fraction $f$ of the energy density of the universe when $m_{V_D} f_{V_D} s_{\text{SM}} \gtrsim f (\pi^2/30) g_{\text{SM}} T_{\text{SM}}^4$, or

$$T_{\text{SM}}^{\text{dom}} \approx (4m_{V_D} f_{V_D})/(3f). \quad (21)$$

An early time of matter domination induced by dark photons, is not in conflict with late-time cosmology, provided that radiation domination is re-established before BBN. We shall thus require that the dark photons either decay before BBN, or that their energy density does not exceed a critical fraction $f$ of the energy density of the universe prior to their decay,

$$\tau_{V_D} \lesssim \max[\tau_{\text{BBN}}, H(T_{\text{SM}}^{\text{dom}})^{-1}] \quad (22)$$

where $\tau_{V_D}$ is the dark photon lifetime computed according to section 2.2, and $H(T_{\text{SM}}^{\text{dom}})$ is the Hubble parameter at the temperature $T_{\text{SM}}^{\text{dom}}$ estimated from eq. (21).

The parameter space disfavoured by the condition (22) is illustrated in fig. 7. We have taken $f \sim 0.5$ and $\tau_{\text{BBN}} \approx 0.03\,\text{s}$. This choice of $\tau_{\text{BBN}}$ is motivated by the known BBN constraints on dark photons; our disfavoured region encompasses all the parameter space that was excluded by a proper BBN analysis in ref. [125], under minimal assumptions about the dark photon cosmological density. In the our case, the dark photon primordial density can be significantly larger. Nevertheless, the effect of dark photon decay on BBN is intricate, and depends not only on the dark photon energy density and decay rate, but also on their mass, which determines their decay channels, and possibly on other details [125]. Our constraints are thus a rough estimate only; it is possible that some parts of the parameter space close to the border of the disfavoured region are in fact allowed. The derivation of more precise BBN constraints is beyond the scope of our work.
3.4 Entropy production from dark photon decay

If the dark photons decay before BBN, but after they constitute a sizeable fraction of the energy density of the universe, i.e. if \( H(T_{\text{dom}}^{-1})^{-1} \lesssim \tau_{VD} \lesssim \tau_{\text{BBN}} \), their decay generates significant entropy, which dilutes any decoupled relic, including DM. We may estimate the dilution factor using the simultaneous decay approximation, and invoking the conservation of energy before and after the decay (see e.g. ref. [126]),

\[
\rho_{\text{before}} = \left(\frac{\pi^2}{30}\right) g_{\text{SM}} T_{\text{SM}}^4 + m_{VD} f_{VD} \left(2\frac{\pi^2}{45}\right) g_{\text{SM}} T_{\text{SM}}^3, \tag{23a}
\]

\[
\rho_{\text{after}} = \left(\frac{\pi^2}{30}\right) g_{\text{SM}} T_{\text{SM}}^4, \tag{23b}
\]

\[
\rho_{\text{before}} = \rho_{\text{after}} = \frac{3M_{\text{Pl}}^2}{(8\pi\tau_{VD}^2)}, \tag{23c}
\]

where \( f_{VD} \) has been estimated in eq. (20). The entropy densities are

\[
s_{\text{before}} = \left(2\frac{\pi^2}{45}\right) g_{\text{SM}} \left(1 + g_{VD} f_{\text{SM}} r_{f.o.}/g_{\text{SM}}\right) T_{\text{SM}}^3, \tag{24a}
\]

\[
s_{\text{after}} = \left(2\frac{\pi^2}{45}\right) g_{\text{SM}} T_{\text{SM}}^3, \tag{24b}
\]

where in eq. (24a) we have accounted both for the SM sector and the dark photon entropy densities (see footnote 9). We determine the temperatures \( T_{\text{SM}}^{\text{before}} \) and \( T_{\text{SM}}^{\text{after}} \) from eqs. (23), and estimate the dilution factor to be

\[
\frac{s_{\text{after}}}{s_{\text{before}}} \approx 1 + 1.72 f_{VD} g_{\text{SM}}^{1/4} m_{VD}^{1/2} \sqrt{\tau_{VD}} M_{\text{Pl}} \left(1 + 3.6 f_{VD}\right). \tag{25}
\]

Figure 7 shows that the dilution can be very significant for \( m_{VD} \gtrsim 1 \text{ GeV} \) and small \( \epsilon \). In this event, the X fermions can account for the observed DM density only if they freeze-out with a significantly larger abundance. This requires a smaller \( \alpha_D \) than that estimated in section 3.2 in the absence of any excessive entropy production, which in turn implies overall smaller annihilation and BSF cross-sections at late times, thereby relaxing the CMB and indirect detection constraints. We leave a more detailed study of this possibility for future work.

4 Constraints from indirect DM searches and the CMB

In this section we move to the indirect detection probes of the model outlined above. In the commonly assumed sense, indirect detection strategies aim at revealing excesses and features in cosmic ray fluxes collected at Earth, which could be ascribed to DM annihilations. The searches focus in particular on regions where DM is most dense or where the astrophysical backgrounds are particularly reduced. Here we concentrate in particular on three indirect detection probes: gamma ray searches with the Fermi satellite in the Milky Way galactic halo and in dwarf spheroidal galaxies (section 4.1) and antiproton searches with the Ams-02 experiment (section 4.2).

In a broader sense, the CMB is also a powerful indirect probe, since it is sensitive to DM annihilations into SM particles at the time of recombination. Such annihilations inject energy in the plasma, causing ionisation and heating of the medium, production of low energy photons, etc. A particular important effect is ionisation: the increased amount of free electrons affects the CMB...
anisotropies, and makes them potentially inconsistent with observations (see refs. [127–131] for the first DM studies using WMAP data).

The compilation of the constraints that we derive in this section is presented in fig. 11.

As mentioned in the introduction, DM bound states will annihilate into 2 or 3 dark mediators $V_D$, depending on the spin state. In turn, $V_D$ decays into pairs of SM particles such as $e^+e^-$, $\mu^+\mu^-$, light quarks and heavy quarks, according to the BRs discussed in detail in section 2.2. Such SM states will then shower and hadronize, when applicable, producing fluxes of stable particles ($\gamma$-rays, $e^{\pm}$ and $\bar{p}$) which can be compared meaningfully with the observations in gamma-rays and antiprotons. We take these spectra from the detailed work in [84], where they are provided for a variety of channels and for a large range of masses. Their results are based on previous PPPC4DMID and PYTHIA computations. A few technical comments are however in order. i) While the provided spectra are computed for the case of scalar mediators, they can also be adopted for the case of vector mediators in which we are interested, as discussed in particular in [51]. ii) The spectra for the decay of spin-singlet states in 2 $V_D$ are given explicitly, while those for spin-triplet states in 3 $V_D$ have to be approximately derived. As discussed in detail in [84], the direct decay into 2n SM states is roughly equivalent to the decay via an n-step cascade, at least for $n = 1, 2$. We therefore determine the spectra for our 3-step cascade by averaging the $n = 1$ and $n = 2$ spectra, and verify that, for our purposes, this is effectively the same of taking the $n = 2$ spectra. We further check, in specific cases, that our averaged spectra are in very good agreement with those obtained by the analytical form given in ref. [84] (eq. (A4) therein). iii) The spectra from [84] are parametrized in terms of the quantity $2m_{SM}/m_{V_D}$ (with $m_{SM}$ the mass of the outgoing SM particles, if non vanishing), within a limited range: when needed (e.g. for large $m_{V_D}$) we will linearly extrapolate the spectra and check the consistency of the procedure. iv) The spectra for light quarks are not provided in [84]: we use the ones for gluon pairs, given their close similarity. Such spectra are given for $m_{V_D} \geq 10$ GeV so we extrapolate down to 5 GeV, below which we adopt the refined treatment discussed in section 2.2.

For the case of the CMB, the impact on the medium depends on the fraction $f$ of the injected power (the products of DM annihilation) that is actually deposited in the environment and thus matters for reionisation. This ‘efficiency factor’ $f$ is, in general, a function of the DM annihilation channel, the injection energy, and the injection redshift $z$. Reference [132] demonstrated however that it is possible to account for the redshift dependence with a constant effective function $f_{\text{eff}}$, and allowed to compute such a function for a wide set of injection energies (i.e. DM masses) and arbitrary annihilation channels. We will use this quantity in the following.

### 4.1 Gamma-ray searches

**Dwarf spheroidal galaxies (dSphs).** These objects are among the best targets to look for DM signals with gamma-ray observations. So far, no evidence for an excess of photons over the background has been found. This allows to set stringent constraints on the DM annihilation cross-section. For this purpose, we follow closely the analysis performed by the FERMI collaboration, which is based on the observation of 15 dwarf galaxies in the energy range 500 MeV-500 GeV [133].

The statistical analysis is performed as follow. For each target, the likelihood $L_i(\mu, J_i|D_i)$ depends on the gamma-ray data, $D_i$, the parameters of the model $\mu$ (i.e. the particular point in the $(M_{DM}, m_{V_D})$ parameter space under consideration) and the so-called $J$-factors $J_i$. The latter quantities correspond to the integral of the square of the DM density profile in the window of the observation. These parameters, specific for each

---

13The numerical results are also linked from the PPPC4DMID website.
Fig. 8: **Gamma-ray constraints** from dwarf galaxies (left) and from the Galactic Halo (right) in the plane $M_{\text{DM}}/m_{V_D}$. If neglecting the phenomenon of bound state formation, only the darker shaded region is probed. The inclusion of bound states extends the area.

dwarf galaxy, are needed to determine the DM fluxes. For each dwarf galaxy, FERMI provide the likelihood in different energy bins as a function of the integrated signal flux. Thus, for a given point $\mu$ in the parameter space and choice of $J_i$, we compute the DM flux in each energy bin and determine the corresponding likelihood. Then, we simply multiply these likelihoods for all the energy bins in the analysis. This gives the likelihood $\mathcal{L}_i(\mu, J_i|D_i)$. The total likelihood $\mathcal{L}_i(\mu, J_i|D_i)$ is obtained multiplying by an additional contribution:

$$\mathcal{L}_i(\mu, J_i|D_i) = \mathcal{L}_i(\mu, J_i|D_i) \times \frac{1}{\ln(10) J_{\text{obs},i} \sqrt{2\pi \sigma_i}} e^{-\frac{(\log_{10}(J_i) - \log_{10}(J_{\text{obs},i}))^2}{2\sigma_i^2}}$$

This method allows to take into account the uncertainty on the determination of the $J$-factors. We take the measured values $J_{\text{obs},i}$ and uncertainties $\sigma_i$ from [133]. Finally, we combine the likelihoods of all the 15 dwarf galaxies and treat the $J$-factors $J_i$ as nuisance parameters. We can then determine the regions of the parameter space excluded at 95% CL performing a test-statistic, comparing the likelihood with and without the DM signal.

We recall that the annihilation and BSF cross-sections are computed by averaging over the speed distribution of DM inside the dwarf galaxies as described in sec. 2.1.

Our results are shown in fig. 8 left. The presence of resonances of the annihilation cross-section are clearly visible. At low DM masses ($M_{\text{DM}} \lesssim 10$ GeV), only values of $m_{V_D} \gtrsim 3$ GeV are excluded. This is because lighter dark photons decay mostly/only into leptonic channels, producing less pronounced gamma-ray spectra than those from hadronic channels. At larger DM masses this effect is compensated by a larger annihilation cross-section and by the fact that the photon spectrum shifts at higher energies. The consequence of the formation of bound states can be isolated comparing the lighter and darker regions of fig. 8, left. The inclusion of this effect rules out an additional region of the parameter space ($2 \lesssim M_{\text{DM}} \lesssim 5$ TeV).
Milky Way galactic halo (GH). We derive another set of constraints from FERMI observations of the diffuse gamma-ray emission from our Galaxy. The strategy that we adopt is based on the analysis in [134]. We consider two mid-latitude regions of interest (RoI 12 and 24 in [134]) defined by $|\ell| < 80^\circ$ and $5^\circ < b < 15^\circ$ ($-15^\circ < b < -5^\circ$) for the region 12 (24), with $\ell$ and $b$ the galactic longitude and latitude. We model the astrophysical emission inside these RoIs as a superposition of different templates. We include i) the emission from the interactions of the cosmic-rays with the interstellar radiation field and the gas, ii) a template for point sources, iii) a template for the so-called ‘FERMI bubbles’ and iv) the isotropic gamma-ray background. For the DM signal we include both the primary gamma-rays produced by the decays of the dark photon and the Inverse Compton secondary emission (for this purpose we use the tool in [135, 136]). We derive constraints on the model comparing the signal and the background fluxes in the RoIs with FERMI observations. We exclude the points with $\Delta \chi^2 > 9$ with respect to the background-only hypothesis. We refer to [134] for further details. As explained in [134], we limit our analysis to energies above $\sim 1.5$ GeV, since only for these energies our background model provides a reliable description of the diffuse emission. We average the annihilation and BSF cross sections considering the Maxwellian speed distribution described in sec. 2.1.

The results are shown in fig. 8 right. Some features are common to the constraints from dwarf galaxies, for instance those associated to the resonances of the annihilation cross-section. Light DM candidates are not excluded mainly for two reasons: the cross-section (at fixed value of $m_{\nu_D}$) decreases at smaller masses and we are focusing on gamma-ray energies above 1.5 GeV. Including the formation of bound states in the analysis allow to rule out an additional region of the parameter space at $M_{DM} \sim 1.5 - 10$ TeV.

4.2 Antiproton searches

The antiproton component in cosmic rays has been recognized since decades as a potentially important channel for DM searches. The recent release [137] of high-precision data by the AMS-02 experiment concerning the antiproton flux and antiproton to proton ratio has on one side strengthened the relevance of antiprotons and on the other side pointed to a need for a better determination of the astrophysical background (called ‘secondary antiprotons’). At the moment, such background can explain the data, within the rather large uncertainties, without the need for an additional exotic component. Hence, it remains meaningful to derive constraints based on antiprotons with the best tools currently at disposal.

We follow closely the analysis strategy of [144], which in turn builds on [145]. We just recall here the most important features. AMS-02 has measured the flux in an energy range that extends from 1 to 450 GeV. The $\bar{p}/p$ ratio, which we will use for consistency with [144], shows a very good agreement with the predicted astrophysical background up to about 50-100 GeV and then remains rather flat in energy. The data can be reasonably well fit with secondaries assuming a MAX galactic propagation scheme and adapting the normalization and the impact of solar modulation within the uncertainties (we refer to [144] for any further detail). For the DM contribution, we choose for definiteness an Einasto profile for the distribution in the MW galactic halo and we include all the relevant propagation phenomena as discussed in [145]. By requiring that the DM contribution does not worsens the secondary-only fit by more than $\Delta \chi^2 = 9$, we can derive conservative DM bounds.

The constraints that we obtain are reported in fig. 9. With respect to the other probes that we consider, they cover smaller areas, necessarily located above $m_{\nu_D} \gtrsim m_p$ where antiproton

\footnote{For recent studies on the subject, in some cases advocating a different conclusion, see [138–143].}
production from $V_D$ decays is kinematically allowed. More precisely, the excluded region consists of two separate areas. The low mass one, for $M_{DM} \lesssim 200$ GeV, corresponds to the case in which the DM $\bar{p}$ flux falls in the high precision part of the AMS-02 data: even a small DM component is constrained by the small experimental error bars. The high mass one ($300$ GeV $\lesssim M_{DM} \lesssim 3$ TeV) corresponds instead to the case in which the large enhancements of the resonance peaks in the annihilation cross-section imply a large DM flux and therefore an overshooting of the measurements.

We notice that the constraints in the low mass area rule out certain models which have been put forward to explain the GC GeV excess, featuring a DM with a mass around 30-50 GeV and dark photons heavier than 1 GeV (see e.g. [45, 47]). While we caution that this is a complex region, where significant astrophysical uncertainties come into play [144, 146] and that would therefore require a dedicated study, we find it worthwhile to stress for the first time that the AMS-02 antiproton data have the power to test this portion. Naturally, models for the GC GeV excess where $m_{V_D} \lesssim 1$ GeV remain unconstrained.

The impact of adding the formation of bound states is limited for antiprotons: only a small additional region around $M_{DM} \simeq 7$ TeV, $m_{V_D} \simeq 7$ GeV is excluded. This is not surprising: antiproton constraints necessarily appear only for $m_{V_D} \gtrsim m_p$ and bound state formation only occurs to the right of the solid blue line, thus restricting the applicable area only to the upper-right portion of the parameter space. On the other hand, for DM masses larger than $\sim 10$ TeV the constraints quickly lose power because the maximal energy of the AMS-02 data and because the annihilation cross section decreases.
4.3 CMB constraints

**Bound on the annihilation cross section.** We base our DM constraints on the recent detailed analysis of ref. [132] (itself based on the inputs of [147]), and on the PLANCK measurements of CMB anisotropies [148]. We take $f_{\text{eff}}$ from [132], and use it in conjunction with the bound of ref. [148], to derive constraints on our parameter space. We shall require that

$$\sum_{\text{SM}=e^e,\mu\mu,...} \langle \sigma_{\text{tot}} v_{\text{rel}} \rangle \text{BR}_{\nu\nu \rightarrow \text{SM}} f_{\text{eff}}(M_{\text{DM}}, \text{SM}) < 8.2 \times 10^{-28} \text{ cm}^3 \text{s} \left( \frac{M_{\text{DM}}}{\text{GeV}} \right),$$

(26)

where we remark that CMB bounds are insensitive to the number of steps of the cascade [84], in contrast to the indirect detection bounds of sections 4.1 and 4.2.

In eq. (26) $\langle \sigma_{\text{tot}} v_{\text{rel}} \rangle$ accounts for all processes via which DM annihilates. It depends, in general, on the DM speed distribution, and should be evaluated at the redshift that affects the ionisation of the medium most strongly. The effect on the ionisation has been shown to peak at $z \sim 600$ and extend down to $z \sim \text{few hundred}$ [149]. We need therefore to determine the DM speed distribution at those redshifts.

**Dark matter velocity around CMB.** The DM temperature at redshifts relevant to CMB can be found from eqs. (16), (18) and (19). It indicates a Maxwellian DM speed distribution with central value $v_0 = \sqrt{2T_X/M_{\text{DM}}}$, where

$$v_0 \approx 10^{-8} \left( \frac{1 + z}{600} \right) \sqrt{\min \left[ \left( \frac{\text{MeV}}{m_{\nu_D}} \right) \left( \frac{\text{GeV}}{M_{\text{DM}}} \right), \left( \frac{\alpha_D}{10^{-6}} \right) \left( \frac{\text{GeV}}{M_{\text{DM}}} \right)^{5/2} \right]}.$$  

(27)

Note that we have assumed a sharp transition between the $T_X \propto 1/a$ and $T_X \propto 1/a^2$ scalings, which is more than sufficient for our purposes. As always, for the relative velocity of the DM particles, we use $v_{\text{rel,0}} = \sqrt{2}v_0$.

**Annihilation cross section.** We now turn to the evaluation of $\langle \sigma_{\text{tot}} v_{\text{rel}} \rangle$ entering the constraint (26).

- For the range of DM velocities indicated by eq. (27) and the range of DM and dark photon masses we consider in this work, only the direct DM annihilation contributes, while BSF is extremely suppressed due to its $\sigma_{\text{BSF}} v_{\text{rel}} \propto v_{\text{rel}}^2$ scaling at low velocities, discussed in section 2.1. Hence, $\langle \sigma_{\text{tot}} v_{\text{rel}} \rangle \simeq \langle \sigma_{\text{ann}} v_{\text{rel}} \rangle$.

- Because $v_0$ is very low, the resonances of $S_{\text{ann}}$ are extremely dense, and their numerical calculation is impractical. We use instead a standard analytical approximation obtained by replacing the Yukawa with the Hulthén potential,\(^{15}\) $V_H = -\alpha_D m_* e^{-m_* r}/(1 - e^{-m_* r})$, where $m_* \sim O(m_{\nu_D})$. Then, the Sommerfeld enhancement factor is (see e.g. [150, 151])

$$S_H = \frac{2\pi \alpha_D}{v_{\text{rel}}} \frac{\sinh(\pi M_{\text{DM}} v_{\text{rel}}/m_*)}{\cosh(\pi M_{\text{DM}} v_{\text{rel}}/m_*) - \cosh \left( \pi \sqrt{M_{\text{DM}}^2 v_{\text{rel}}^2/m_*^2 - 4M_{\text{DM}} \alpha_D/m_*^2} \right)}. \quad (28)$$

While $S_H$ is a good approximation off-resonance, it does not reproduce the precise position of the resonances of the Yukawa potential [150], beyond the first one, whose position we

\(^{15}\) For the other indirect detection analysis in this section, we instead calculate $S_{\text{ann}}$ fully numerically, using the Yukawa potential.
can define. Indeed, we choose $m_\ast = 1.68 m_{V_D}$, such that the first resonances of the Yukawa and the Hulthén potentials coincide.\footnote{This is slightly different from the more common choice found in the literature, $m_\ast/m_{V_D} = \pi^2/6 \simeq 1.64$ [150].} Despite the discrepancy in the location of the higher resonances, the Hulthén potential reproduces correctly the fact that they become denser at increasing values of $\alpha_D M_{DM}/m_{V_D}$, and is overall sufficient for our purposes.

It is well-known that at very low velocities, the growth of the resonances of the Hulthén (or the Yukawa) potential is unphysical, a fact that is manifested by the apparent violation of the unitarity limit on the inelastic cross-section. We use the prescription of ref. [152] to ensure that the resonant growth is curtailed below the unitarity limit. For this purpose, we replace the Sommerfeld enhancement factor $S_H$ with $S_{reg}^H$, where

$$S_{reg}^H = \frac{S_H}{(1 + S_H \alpha_D^2 v_{rel}/16)^2}. \quad (29)$$

\(\diamond\) In the parameter space of interest, $\sigma_{ann} v_{rel}$ is well within the saturated regime, \textit{i.e.} it is independent of $v_{rel}$. It is then consistent to use the constraint (26), which applies to a velocity-independent $\sigma v_{rel}$. Moreover, the precise value of the speed around $z \sim 600$ is unimportant, and there is no need to average over the DM speed distribution.

**Results.** The portion of our parameter space excluded by CMB is shown as a shaded green area in fig. 10. Values of $m_{V_D} \lesssim 1$ GeV are ruled out for any DM mass, with the exception of a ‘gap’ at $M_{DM} \sim (10 - 200)$ GeV, that extends down to $m_{V_D} \sim 200$ MeV. For larger values of $m_{V_D}$, and for $M_{DM} \gtrsim 1$ TeV, the excluded regions follow the resonances of the annihilation cross-section. Our CMB bounds are in good agreement with those derived in ref. [72] for the same model, and extend to larger values of $m_{V_D}$ and $M_{DM}$. Moreover, they broadly agree with those derived in [84, 132]. We complement them by considering the dark photon mass as an explicit parameter and combining the different annihilation channels as dictated by the actual branching ratios of $V_D$. For $m_{V_D} \gtrsim 1$ GeV and $M_{DM} \gtrsim 1$ TeV, our inclusion of BSF in the DM relic density computation is important, because it alters the predicted $\alpha_D$ with respect to a calculation that includes annihilation only (see section 3.2). This, in turn, affects the position of the resonances (which is, however, difficult to visualize because they are very dense), but mildly so on the overall size of the excluded area.

5 Conclusions

Dark $U(1)$ sectors are ubiquitous in beyond-the-Standard-Model physics, ranging from string-theory constructions to supersymmetric extensions, and models with phenomenological motivation. Such sectors often involve stable particles that may have been produced with cosmologically significant abundances in the early universe, and make up part or all of the DM today. In this work, we considered DM as a Dirac fermion coupled to a dark $U(1)_D$ force, and carried out a comprehensive study of its phenomenology.

As with all relics from the early universe, the phenomenology of DM today depends on the preceding cosmology. For the first time in the literature, the present work incorporates self-consistently, in the cosmology and the phenomenology of this model, effects that had been previously only separately considered. In particular, we accounted for the effect of bound states on the
DM relic density—thus on the estimated dark sector couplings—and computed direct and indirect detection constraints on the model parameters.

Our main points can be summarized as follows:

- We identify the regions of the parameter space of the model in which bound state formation is relevant (section 2.1) and quantify precisely its impact. In these regions, BSF increases the total effective annihilation rate and therefore allows to obtain the observed DM density with a smaller dark coupling. Neglecting the effect of BSF on the relic density can result in an error of 50% and up to a factor of $\sim 4$ in the expected signal strength, for a given DM mass.

- Despite the decrease in the predicted coupling, the expected indirect detection signals are increased, due to the BSF contribution, leading to the exclusion of additional regions of the parameter space (see figs. 8 and 9).

- We detailed the cosmological history of the dark sector, and estimated the constraints implied from not disrupting BBN and changing the time of matter-radiation equality (section 3.3).

- MeV – GeV dark-photon masses, which are relevant for the leptophilic realizations of the model, are highly constrained by an overlap of CMB and gamma-ray bounds for any DM
mass in our range (see fig. 11). This sub-GeV dark-photon mass range has also been invoked in the context of self-interacting dark matter: the combined cosmological and indirect detection considerations imply that the realization of self-interacting dark matter within a $\mathbb{U}(1)_D$ dark sector is viable either for very light (eV-scale)/massless dark photons, or provided that dark matter possesses a particle-antiparticle asymmetry.

* Dark photons heavier than 1 GeV are somewhat less constrained in the context of the present model, with the antiproton measurements and the CMB being the most powerful probes in this regime. The antiproton bounds put under strain models that explain the GC GeV excess in terms of 40 GeV DM coupled to heavy (above GeV) dark photons (although astrophysical uncertainties make a definite statement challenging).

* In the multi-GeV dark photon region, the indirect detection constraints of fig. 11 would be weakened for a very small value of the kinetic mixing $\epsilon$. Small $\epsilon$ values (between $10^{-9}$ and $10^{-12}$, depending on the $m_{V_D}$ considered), while still guaranteeing a dark photon lifetime shorter than $\tau_{\text{BBN}}$, would cause $V_D$ to decay when it dominates the energy density of the universe, thus injecting significant entropy in the SM plasma and diluting DM. The dilution factors are shown in fig. 7. The dilution of DM would imply a smaller coupling $\alpha_D$, in order to account for the observed DM density, and in turn the possibility to have DM of thermal origin heavier than $\sim 100$ TeV without violating the unitarity bound. We leave the study of the related phenomenology for future work.

Moreover, if in the early universe the dark plasma was at a significantly lower temperature than assumed in this work, $\alpha_D$ would be expected to be smaller, in the entire parameter space considered. This would relax the CMB, direct and indirect detection constraints, but would also shift the parameter space within which the DM self-interactions are sizeable.

Appendix

A Coulomb approximation in relic density calculations

To ensure the validity of our computation of the DM relic density, we consider the time of chemical decoupling, when the DM density has approached its final value, $Y_X \simeq \Omega_{\text{DM}} \rho_c / (2 M_{\text{DM}} s_0)$, where $\rho_c$ and $s_0$ are the critical energy density and the entropy entropy of the universe today. Because of the Sommerfeld effect, the annihilation and BSF processes remain sizeable after freeze-out. While freeze-out – conventionally defined as the time when the density of the DM particles departs from its equilibrium value – occurs at $T_{\text{fo}}^{\nu} \approx M_{\text{DM}}/30$, the chemical decoupling may occur significantly later.

We shall define the chemical decoupling as the time when the rate of change of the DM comoving density becomes $|d \ln Y_X / d \ln x| \lesssim 10\%$, where $x \equiv M_{\text{DM}} / T_{\text{SM}}$. Around that time, the DM density is governed by the Boltzmann equation $d \ln Y_X / d \ln x \approx -\sqrt{g} \left[ \sigma_0 / (1.2 \times 10^{-26} \text{cm}^3 / \text{s}) \right] \left[ S_{\text{eff}}(x) / x \right]$, where on the right side of the equation, we have substituted the final value of the DM density. The “effective” Sommerfeld factor $S_{\text{eff}}$ is defined as in ref. [54]; it includes the thermally averaged Sommerfeld enhancement factor $\left< S_{\text{ann}} \right>$, as well as $\left< S_{\text{BSF}} \right>$ but only after the bound-state ionisation rate has become lower than the bound-state decay rate. Using these equations, we determine the temperature of chemical decoupling $T_{\text{c}}^{\nu}$, for a given $M_{\text{DM}}$ and the associated $\alpha_D$.

The condition (9) for the Coulomb approximation is fulfilled provided that the chemical decoupling occurs at $T_{\text{c}}^{\nu} \gtrsim m_{V_D}^2 / M_{\text{DM}}$. We find that this is satisfied for the $m_{V_D}$ range below the upper dotted red line in fig. 2. Since this encompasses essentially all of the parameter where the
Sommerfeld effect is important (below dashed blue line), including the entire region where bound 
states can form, we deem the Coulomb approximation to be satisfactory for the computation of 
the relic density, in the entire region where \( S_{\text{ann}}, S_{\text{BSF}} > 1 \).

Our computation also introduces a Coulombic Sommerfeld enhancement in the parameter 
space above the dashed blue line of fig. 2, where in fact \( S_{\text{ann}} \simeq 1 \) at any speed, due to \( m_{\nu_D} \) being 
large. This raises the concern that \( \sigma_{\text{ann}} v_{\text{rel}} \) may be overestimated in this regime. However, for 
\( M_{\text{DM}} \lesssim \text{TeV} \), which encompasses most of the parameter space above the dashed blue line, the 
Sommerfeld enhancement has no significant effect on the DM density, even when its Coulomb 
limit is employed, because \( \alpha_D / v_{\text{rel}} \) remains small before chemical decoupling. Thus, including a 
Coulombic Sommerfeld enhancement in the computation does not affect the predicted \( \alpha_D \) in this 
region, and our results remain valid above the dashed blue line of fig. 2 as well.

Acknowledgements

We thank Francesc Ferrer, Andreas Goudelis, Felix Kahlhoefer, Bradley Kavanagh, Tien-Tien Yu 
and Gabrijela Zaharijas for interesting discussions. We acknowledge the hospitality of the Institut 
d‘Astrophysique de Paris (IAP) where a part of this work was done.

Funding and research infrastructure acknowledgements:

* European Research Council (ERC) under the EU Seventh Framework Programme (FP7/2007-
2013)/ERC Starting Grant (agreement n. 278234 — ‘NEWDark’ project).
* ANR (Agence Nationale de la Recherche) under contract ACHN 2015 ‘THE INTRICATE DARK’ 
project [work of K.P.].
* NWO (The Netherlands Organisation for Scientific Research), VIDI grant, ‘SELF-INTERACTING 
ASYMMETRIC DARK MATTER’ project [work of K.P.].
* ERC Advanced Grant project 267117 (‘DARK’) hosted by Université Pierre et Marie Curie – Paris 
6 [initial work of P.P.].
* Centro de Excelencia Severo Ochoa Programme SEV-2012-0249 [work of M.T.].
* MINECO through a Severo Ochoa fellowship with the Program SEV-2012-0249 [work of M.T.].
* FPA2015-65929-P and Consolider MultiDark CSD2009-00064 [work of M.T.].

References

[1] G. Bertone, D. Hooper, and J. Silk, Particle dark matter: Evidence, candidates and constraints, 
[2] J. L. Feng, Dark Matter Candidates from Particle Physics and Methods of Detection, 
023503, [hep-ph/0610133].


[71] S. P. Liew and F. Luo, Effects of QCD bound states on dark matter relic abundance,


