Supersymmetric resonant dark matter:
A thermal model for the AMS-02 positron excess

Yang Bai, Joshua Berger, and Sida Lu

Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

We construct a thermal dark matter model with annihilation mediated by a resonance to explain the positron excess observed by PAMELA, Fermi-LAT and AMS-02, while satisfying constraints from cosmic microwave background (CMB) measurements. The challenging requirement is that the resonance has twice the dark matter mass to one part in a million. We achieve this by introducing an $SU(3)_f$ dark flavor symmetry that is spontaneously broken to $SU(2)_f \times U(1)_f$. The resonance is the heaviest state in the dark matter flavor multiplet, and the required mass relation is protected by the vacuum structure and supersymmetry from radiative corrections. The pseudo-Nambu-Goldstone bosons (PNGBs) from the dark flavor symmetry breaking can be slightly lighter than one GeV and dominantly decay into two muons just from kinematics, with subsequent decay into positrons. The PNGBs are produced in resonant dark matter semiannihilation, where two dark matter particles annihilate into an anti–dark matter particle and a PNGB. The dark matter mass in our model is constrained to be below around 1.9 TeV from fitting thermal relic abundance, AMS-02 data and CMB constraints. The superpartners of Standard Model (SM) particles can cascade decay into a light PNGB along with SM particles, yielding a correlated signal of this model at colliders. One of the interesting signatures is a resonance of a SM Higgs boson plus two collimated muons, which has superb discovery potential at LHC Run 2.

I. INTRODUCTION

It is beyond doubt that the majority of matter in the Universe is composed of dark matter, yet we still don't know how to describe the particle properties, *if any*, of dark matter as we can with other particles in the Standard Model (SM). The mechanism by which the abundance of observed dark matter is generated is not known either, though thermal freeze-out has long been regarded as the simplest explanation of the dark matter relic abundance. For order-one coupling strength between dark matter particles and SM particles or other mediators, the dark matter mass is anticipated to be around the TeV scale in freeze-out models. These models furthermore generically predict additional contributions to the cosmic ray spectra of electrons/positrons, protons/antiprotons, photons and neutrinos, generated by dark matter annihilations in the present day. Among the experimental searches for such cosmic rays, known as indirect detection searches, AMS-02 has provided the most precise measurement of the electron and positron energy spectrum up to 1 TeV [1]. Their measurement of the positron fraction shows an excess above the standard background estimation up to an energy of 0.6 TeV [2,3]. This interesting excess has also been seen in earlier experiments including HEAT [4], PAMELA [5], Fermi-LAT [6].

On the one hand, the energy scale of the positron excess matches the generic mass scale of thermal dark matter models, which provides a strong hint that the positron excess may be explained by thermal dark matter annihilations. On the other hand, the preferred annihilation rate from data, $\mathcal{O}(10^{-23} \text{ cm}^3/\text{s})$, is 2 to 3 orders of magnitude higher than the required rate, $\sim 3 \times 10^{-26} \text{ cm}^3/\text{s}$, of the simple s-wave annihilation thermal dark matter models. Additional complications are required in the dark matter sector to explain why the dark matter annihilation rate is higher in the present day Milky Way halo than during the time of thermal freeze-out. One frequent approach to accommodate both annihilation rates utilizes the so-called "Sommerfeld enhancement," whereby attractive long-range interactions among the dark matter particles yield a $1/v$ enhancement to the cross section for short-range annihilation. Because the dark matter averaged velocity in Milky Way is around 10^{-3}, 2 orders of magnitude smaller than that during thermal freeze-out, the large present day annihilation rate preferred by data can be naturally explained [7,8]. For this class of models, dark matter annihilation rates during the recombination era, where
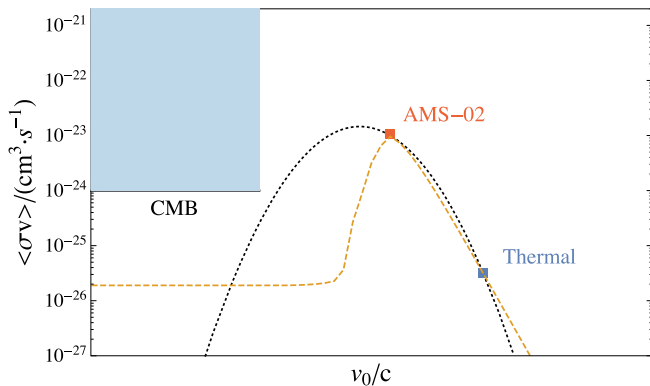
FIG. 1. A schematic plot to show the required annihilation rates for AMS-02 positron fraction data and dark matter thermal relic abundance, as well as the constraints of the CMB power spectrum from Planck. The dashed orange line is the possible behavior from resonance-mediated annihilation, while the dotted black line is a phenomenological fit to the data.

$v \ll 10^{-3}$, are further enhanced and dump energetic electrons and positrons into the plasma, which interact with the CMB photons and lead to excluded distortions of the CMB power spectrum (see Refs. [9,10] recent analysis for this class of models and Refs. [11–14] and Planck constraints [15] for general models). Up to the model-dependent absorption efficiencies of electrons and positrons energy, the constraint on the dark matter annihilation rate is $\langle \sigma v \rangle \lesssim \mathcal{O}(10^{-24} \text{ cm}^3/\text{s})$.

Before we move to discuss other possible models, we first study the schematic picture told by the experimental data in Fig. 1. Because of the relations among three characteristic velocities, $v_{\text{CMB}} < v_{\text{AMS}} < v_{\text{thermal}}$ and the large rate required for the AMS-02 data, the underlying dark matter annihilation rate $\langle \sigma v \rangle$ has to have a peak structure around v_{AMS}. This observation is based on a simplified early universe model for the dark matter relic abundance. More possibilities are allowed within some nonstandard cosmological models. From the particle physics point of view, the simplest explanation for the peak in Fig. 1 is to have the dark matter annihilation mediated by a resonance, though other possibilities exist [16]. If the resonance particle mass is very close to twice of the dark matter mass, then the dark matter may reach its peak annihilation rate around v_{AMS}, yielding a much larger rate for the dark matter annihilation in the Milky Way halo.

Several studies of so-called "Breit-Wigner Enhancement" [17–23] phenomenological models exist. Such models can simultaneously fit to both the AMS-02 data and thermal relic abundance. For some models considered in the literature (for instance Ref. [19]), a light PNGB exists with a mass below one GeV and mainly decay into two electrons and two muons, which can satisfy gamma ray experimental constraints [24–26] from Fermi-LAT [27]. Although this class of models seem to be in the right direction to provide a thermal dark matter explanation for the data, it suffers its

own problem from the theoretical point of view. The biggest issue with resonant annihilation is explaining why the resonance mass M_R is close to twice of the dark matter mass M_X at per million level: $(M_R - 2M_X)/(2M_X) \sim \mathcal{O}(v_{\text{AMS}}^2) \sim \mathcal{O}(10^{-6})$.

One of the simplest ways to explain the small mass splitting is to have the resonance be a bound state of two dark matter particles. This scenario requires a long-range force to provide the binding energy and also suffers the additional $1/v$ Sommerfeld enhancement for the annihilation rate at the CMB era. To solve this problem, one could have two interacting dark matter states, X_1 and X_2, with the mass splitting $\delta M \sim 1$ MeV and only the long-range force for the X_2 state. Then, the lighter dark matter states X_1 can annihilate via exchanging an X_2 bound state in the s-channel. To have a natural model without fine-tuning the parameter space, the gauge coupling, if mediated by $U(1)'$ force, has to have the binding energy $\mathcal{O}(\alpha'^2 M_X)$ match to the dark matter kinetic energy $\mathcal{O}(1 \text{ MeV})$, so $\alpha' \sim 10^{-3}$. However, the additional gauge interaction for the X_2 can also shift its mass at loop level, with a contribution $\mathcal{O}(\alpha' M_X) \sim \mathcal{O}(1 \text{ GeV})$, dramatically higher than the required small mass splitting $\mathcal{O}(1 \text{ MeV})$. Another option is a higher-dimensional model such as the Universal Extra Dimension (UED) model [28] (as mentioned in Ref. [17]). The second Kaluza-Klein (KK) mode has the mass close to twice the first KK-mode mass. This seems to be an interesting way to construct a UV model for resonant dark matter. However, additional quantum corrections are anticipated to generate at least loop suppressed brane-localized kinetic terms [29]. Taking those terms into account, if the couplings in the dark sector are large, then the ratio of the second KK-mode mass over the first KK-mode mass can be close to two only at 10^{-3} level and still far from phenomenologically needed 10^{-6} level. Small couplings may be possible from a low energy phenomenology perspective, but the implementation of boundary conditions could require some large couplings that would reintroduce problematically large corrections to the mass spectrum. It becomes a nontrivial model-building challenge to obtain a natural model to realize resonant dark matter annihilation.

In this paper, we explore a new and natural way to realize resonant dark matter annihilation based on the symmetry breaking vacuum structure of non-Abelian global symmetry. As worked out a long time ago, the renormalizable potential of a certain representation of $SU(N)$ symmetry has only a few discrete vacuum structures [30]. For instance, one could have $SU(3) \rightarrow SU(2) \times U(1)$ or $SU(3) \rightarrow SU(2)$, depending on the coupling relations in the potential. Those symmetry breaking patterns are fairly stable against radiative correction and higher-order dimensional operator correction. The breaking pattern $SU(3) \rightarrow SU(2) \times U(1)$ is achieved, for instance, when an $SU(3)$ octet scalar gets a vacuum expectation value (VEV). The order parameter is $\langle \Phi \rangle \propto \text{diag}(1, 1, -2)$ after a particular

choice of $SU(3)$ basis. If this order parameter spurion couples to dark matter fields in $3(\bar{3})$ representation, the ratio of the masses of the heavier dark matter state over the lighter dark matter states is therefore two. In order to have resonant annihilation, at least one of the dark sector states should be a boson. Spin one dark matter would have additional model building difficulties, so we are left with scalar dark matter at the TeV scale. Additional symmetries are needed to explain why such a scalar is light, as it has a hierarchy problem analogous to that of the SM Higgs boson. Supersymmetry (SUSY) remains a leading candidate to solve the hierarchy problem, so we construct a "supersymmetric resonant dark matter" (SRDM) model, based on the dark matter flavor symmetry breaking of $SU(3)_f/SU(2)_f \times U(1)_f$.

One interesting coincidence is that the symmetry breaking of $SU(3)_f/SU(2)_f \times U(1)_f$ also provides a PNGB supermultiplet. The PNGB states could naturally have a mass at scale dramatically below the dark matter mass scale. If their mass is below around 1 GeV, the leading decays into SM particles will likely be two muons or two electrons, just from kinematics, which are functionally the "best" dark matter annihilation channels [24–26]. The SRDM model thus solves two problems at once, providing further motivation for its structure.

The remainder of the paper is organized as follows. In Sec. I, we explicitly write down the necessary superpotential and soft terms for the dark matter states and the interactions to break the global symmetry. We then calculate the particle spectrum and decay rates in Sec. III. In Sec. IV, we calculate the annihilation cross section for the processes mediated by the resonance and the corresponding dark matter relic abundance. We show the parameter space to fit the AMS-02, Fermi-LAT gamma ray and CMB data in Sec. V and show additional signals of the SRDM model in Sec. VI. We conclude our paper in Sec. VII.

II. THE MODEL BASED ON $SU(3)_f/SU(2)_f \times U(1)_f$ SYMMETRY BREAKING

As is well known for nonsupersymmetric theories, certain renormalizable potential of some representations of global Lie group can only have some finite possible vacuum symmetries. For $SU(3)_f$ with an adjoint representation $\Phi \equiv \Phi_a t_a$ with t_a as the generators and $a = 1, …, 8$, one possible vacuum is $\langle\Phi\rangle \propto \text{diag}(1, 1, -2)$ with the unbroken symmetry $SU(2)_f \times U(1)_f$. If the unbroken symmetry is nearly exact in the low energy theory, the vacuum structure of Φ should stay fairly stable and is not easily modified by quantum correction or higher-dimensional operators. If this symmetry-breaking spurion, Φ, couples linearly to other matter fields that is fundamental under $SU(3)_f$, the $SU(2)_f$-doublet field should have a mass half of the $SU(2)_f$-singlet field, which is exactly the required condition to realize resonant dark matter or Breit-Wigner features for dark

matter annihilations. In this section, we build the SRDM model based on this simple observation in a supersymmetric theory. The reason that we choose a supersymmetric model is to extend the factor of two mass relation to particles with different spins or different parities.

In the SRDM model, we have the spontaneously breaking global dark matter flavor symmetry to be $SU(3)_f$, under which we have two superfields, X and $\overline{X}$, as 3 and $\bar{3}$ and one superfield Φ as 8. Based on the global symmetry, we have following renormalization interactions in the superpotential

$$W_{SU(3)} = -y\overline{X}_i\Phi^i_j X^j + \mu_\Phi \text{Tr}(\Phi\Phi) + \lambda_\Phi \text{Tr}(\Phi\Phi\Phi). \quad (1)$$

where $i = 1, 2, 3$ are the indices of $SU(3)_f$. For simplicity, we assume that all parameters are positive and this superpotential conserves P and CP. Minimizing the potential, one can have two degenerate supersymmetric minima for $\langle X\rangle = \langle\overline{X}\rangle = 0$ with $\langle\Phi\rangle = 0$ or $\langle\Phi\rangle = 4\mu_\Phi/(\sqrt{3}\lambda_\Phi)t^8$, up to an arbitrary $SU(3)_f$ transformation. In the symmetry breaking vacuum, $SU(3)_f$ breaks to its subgroup $SU(2)_f \times U(1)_f$ with two massless complex superfields that are doublets under $SU(2)_f$. If there are no additional interactions in the superpotential or soft potential, one has simple relations among the three dark matter states: $M_{X_3} = 2M_{X_1} = 2M_{X_2}$, which is simply from the symmetry breaking pattern of dark flavor symmetry. Note that we have neglected a few terms, such as $\overline{X}X$. The additional terms could be forbidden by a $\mathcal{Z}_2$ under which X and Φ are charged, for example, which λ_Φ breaks by a small amount. Furthermore, since this interaction would be in the superpotential, it is not radiatively generated in the absence of SUSY breaking effects.

Other than the $SU(3)_f$-conserving superpotential, we also need to introduce an explicit $SU(3)_f$ breaking superpotential for three purposes: (a) introducing interactions among the dark matter states; (b) providing masses for the pseudo-Nambu-Goldstone bosons and fermions (PNGB/F) of the coset space of $SU(3)_f/SU(2)_f \times U(1)_f$; (c) introducing interactions of the dark matter sectors with the SM sector. We introduce the following $SU(3)_f$ breaking potential with the three terms fulfilling the three goals above, respectively,

$$W_{SU(3)} \supset \frac{\lambda_X}{2}(X_1 X_1 X_3 + \overline{X}_1\overline{X}_1\overline{X}_3) + \epsilon_1\mu^2_\Phi\Phi_8$$
$$+ \epsilon_2\sum_a \Phi^a H_u H_d. \quad (2)$$

For the first term, one could also have interactions like $X_2 X_2 X_3$ or $X_1 X_2 X_3$. Since we will have the lightest stable dark matter state in X_1, those interactions are not important for dark matter phenomenology. We also note that there is a $\mathcal{Z}_3$ symmetry for the two superpotentials in Eqs. (1) and (2), under which $X_i \to \omega X_i$, $\overline{X}_i \to \omega^2\overline{X}_i$ and $\Phi \to \Phi$ with

TABLE I. The fermion mass spectrum in the SRDM model. Here, the parameter $\delta m \equiv \sqrt{3} b_\Phi \lambda_\Phi f / 8 = b_\Phi \mu_\Phi / 2$ and $M_X \equiv y f / (2\sqrt{3}) - \delta m$. We neglect some leading corrections in ϵ_1 and b_Φ to the $\tilde{\Phi}_T$ and $\tilde{\Phi}_8$ masses that are unimportant to the remaining discussion.

Fermions	$\tilde{X}_{1,2}, \tilde{\bar{X}}_{1,2}$	$\tilde{X}_R, \tilde{\bar{X}}_R$	$\tilde{\Phi}_T$	$\tilde{\Phi}_D(\tilde{\bar{\Phi}}_D)$	$\tilde{\Phi}_8$
Mass	$M_X + \delta m$	$2M_X + 2\delta m$	$\frac{3\sqrt{3}}{4} \lambda_\Phi f$	$\frac{4b_\Phi + \sqrt{3}\epsilon_1 \lambda_\Phi}{4} \mu_\Phi$	$\frac{\sqrt{3}}{4} \lambda_\Phi f$

$\omega = e^{i2\pi/3}$. This $\mathcal{Z}_3$ symmetry is sufficient to protect the dark matter states from decaying in the SRDM model. The second term in the above superpotential can provide masses for the PNGB/F. $SU(3)_f$ freedom allows us to choose a basis where this term goes as Φ^8 and Φ^3, though we neglect the small correction due to a Φ^3 term in what follows. This pushes the vacuum into the Φ^8 direction, whereas it was previously arbitrary up to an $SU(3)_f$ rotation. Here, the dimensionless parameter, $\epsilon_1 \ll 1$, will perturb the vacuum expectation value (VEV) of Φ and introduce tiny mass differences among different flavor components of $X(\overline{X})$. For the last term in Eq. (2), the coefficient $\epsilon_2 \ll 1$ is introduced to mediate interactions of the dark sector to the SM sector. This term has negligible effects on vacuum structures and spectra of the SRDM model and will be only responsible for the PNGB/F decays to SM particles.

As in the MSSM sector, various soft-mass terms could exist for the dark matter sector. One necessary soft term is needed to break the degeneracy of two vacua with $\langle \Phi \rangle = 0$ and $\langle \Phi \rangle \neq 0$. We choose this term to respect the $SU(3)_f$ global symmetry. The second soft term that we also need is to break the degeneracy of the two dark matter states X_1 and X_2 and potentially have X_1 as the lightest dark matter state. They are

$$V_{\text{soft}} \supset -b_\Phi \mu_\Phi^2 \text{Tr}(\Phi^2) + \text{H.c.} + b_{X_i} \mu_\Phi^2 (X_i X_i^\dagger + \overline{X}_i \overline{X}_i^\dagger). \quad (3)$$

The dimensionless parameters $0 < b_\Phi, b_{X_i} \ll \epsilon_1 \ll 1$ and will be determined later from our fit to the AMS-02 positron signal.

To the first order in both b_Φ and ϵ_1 and minimizing the potential, we have $\langle X \rangle = \langle \overline{X} \rangle = 0$ and

$$f = \langle \Phi_8 \rangle = \frac{4\sqrt{3} + 4\sqrt{3}b_\Phi + 3\lambda_\Phi \epsilon_1}{3\lambda_\Phi} \mu_\Phi. \quad (4)$$

It's easy to check that $V|_{\Phi=0} - V|_{\Phi=f} = (16 b_\Phi \mu_\Phi^4)/(3\lambda_\Phi^2) > 0$, the symmetry-breaking vacuum with $f \neq 0$ is the global minimum of the potential. Knowing the vacuum structure of the SRDM model, we first work out the particle spectrum and properties, followed by the dark matter annihilation rate.

III. PARTICLE MASS SPECTRA AND DECAYS

For a small soft parameters, the mass spectrum in the SRDM model is nearly supersymmetric. The tiny mass splittings among different components could be crucial for the dark matter phenomenology and also provide a natural model for the dark matter annihilation mediated by a resonance. The soft mass parameter modifies the scalar particle masses, while the fermion masses only change by the SUSY-breaking shift in f, so we first discuss the fermion masses and then come back to scalar masses.

A. Fermion mass spectrum

Under the remaining approximately global symmetry $SU(2)_f \times U(1)_f$, three components of $\tilde{X}(\tilde{\bar{X}})$ each separate into a doublet and a singlet. In the basis we choose, the doublets can be written as $\tilde{X}_D = (\tilde{X}_1, \tilde{X}_2)^T$ and $\tilde{\bar{X}}_D$. The two singlets are $\tilde{X}_3 \equiv \tilde{X}_R$ and $\tilde{\bar{X}}_3 \equiv \tilde{\bar{X}}_R$. The Weyl fermions $\tilde{X}$ and $\tilde{\bar{X}}$ can be combined to form Dirac fermions. For the fermion fields contained in the superfield Φ, there is an $SU(2)_f$ triplet field $\tilde{\Phi}_T = (\tilde{\Phi}^1, \tilde{\Phi}^2, \tilde{\Phi}^3)^T$, two doublets as $\tilde{\Phi}_D(\tilde{\bar{\Phi}}_D) = (\tilde{\Phi}^4 \pm i\tilde{\Phi}^5, \tilde{\Phi}^6 \pm i\tilde{\Phi}^7)^T$ and a singlet $\tilde{\Phi}^8$. The two Weyl fermion doublets combine to form Dirac fermions, while the triplet and singlet are Majorana fermions. In Table I, we show the mass square of various fermions in leading orders of ϵ_1 and b_Φ, where one can see that only the PNGF masses are suppressed by the small coefficients ϵ_1 and b_Φ. All other fermion masses are anticipated at the scale of f for Yukawa couplings of order of unit. It also easy to check that the singlet fermion $\tilde{X}_R$ has its mass to be twice of the doublet fermion $\tilde{X}_{1,2}$, as one anticipates.

B. Scalar mass spectrum

The scalar fields X and $\overline{X}$ separate into different parity states. The mass eigenstates X_s and X_p as

$$X_s = \frac{1}{\sqrt{2}}(X + \overline{X}^\dagger), \qquad X_p = \frac{1}{\sqrt{2}}(X - \overline{X}^\dagger). \quad (5)$$

Under the unbroken $\mathcal{Z}_3$ symmetry, we simply have $X_s \to \omega X_s$ and $X_p \to \omega X_p$. The subscripts s and p here indicate "scalar" and "pseudoscalar," respectively. There are both doublet and singlet scalars and pseudoscalars. We show the mass spectrum in Table II at the leading order in b_Φ, b_{X_i} and ϵ_1.

TABLE II. Mass spectrum of X and $\bar{X}$ scalars. Here, $X^R_{s,p} \equiv X^3_{s,p}$; $\delta m' \equiv \frac{1}{2} b_{X_2} \mu^2_\Phi / M_X$ and $\delta m'' \equiv \frac{1}{4} b_{X_3} \mu^2_\Phi / M_X$ with $b_{X_1} = 0$.

Scalars	X^1_p	X^1_s	X^R_s
	X^2_p	X^2_s	X^R_p
Mass	M_X	$M_X + 2\delta m$	$2M_X + \delta m + \delta m''$
	$M_X + \delta m'$	$M_X + 2\delta m + \delta m'$	$2M_X + 3\delta m + \delta m''$

From Tables I and II, it is easy to see that the state X^1_p is the lightest dark matter state for all positive dimensionless couplings and is the dark matter candidate in our model. For our later phenomenological studies, we will have all parameters $\delta m, \delta m', \delta m'' = \mathcal{O}(\text{MeV})$ to be around the dark matter kinetic energy. The two heaviest scalars, X^R_s and X^R_p, are just lightly heavier than twice of dark matter mass. We will show that one of them will be the relevant s-channel state to mediate resonant annihilations of dark matter particles.

The scalar Φ soft potential contains terms like $\Phi\Phi$ and $\Phi^\dagger\Phi^\dagger$, which split the real and imaginary parts of the original scalar fields in the supermultiplets. To leading order in b_Φ and ϵ_1, we have the mass spectrum shown in Table III. The light scalar fields are $\Phi^D_{s,p}$ with their masses suppressed by b_Φ and ϵ_1 at linear order.

Before we end this section, we briefly discuss the mass scales in our model. The dark matter mass, M_X will be chosen to be $\mathcal{O}(1 \text{ TeV})$. The doublet $X_{1,2}(\bar{X}_{1,2})$ fermions and scalars all have masses around M_X. The singlet $X_R(\bar{X}_R)$ fermions and scalars have masses around $2M_X$. The fermions and scalars in Φ, excluding the Goldstone doublet superfield, have masses of $\mathcal{O}(\lambda_\Phi / y M_X)$ and could be dramatically heavier than the dark matter and decouple from dark matter phenomenology. The PNGB masses will be $\mathcal{O}(500 \text{ MeV})$ to select muons as the leading decay channel from kinematics. The mass splittings, related to $\delta m, \delta m', \delta m''$, among different $X(\bar{X})$ states will be $\mathcal{O}(1 \text{ MeV})$. Altogether, the small dimensionless parameters in our model have magnitudes

$$\delta m = \frac{1}{2} b_\Phi \mu_\Phi = \frac{3}{4} b_\Phi \frac{\lambda_\Phi}{y} M_X \sim 1 \text{ MeV}$$
$$\Rightarrow b_\Phi \sim 10^{-6} \times \frac{y}{\lambda_\Phi}, \tag{6}$$

$$\delta m'(\delta m'') = \frac{1}{2} b_{X_2}(b_{X_3}) \mu^2_\Phi / M_X = \frac{9}{16} b_\Phi \frac{\lambda^2_\Phi}{y^2} M_X \sim 1 \text{ MeV}$$
$$\Rightarrow b_{X_2}(b_{X_3}) \sim 10^{-6} \times \frac{y^2}{\lambda^2_\Phi}, \tag{7}$$

TABLE III. Mass spectrum of Φ scalars.

| Scalars | Φ^T_s, Φ^T_p | $\Phi^D_{s,p} = \frac{1}{\sqrt{2}}(\Phi^D \pm \bar{\Phi}^{D\dagger})$ | Φ^8_s, Φ^8_p |
| Mass | $m_{\tilde{\Phi}^T} \mp \delta m$ | $\sqrt{m_{\tilde{\Phi}^D}(m_{\tilde{\Phi}^D} \mp 2\delta m)}$ | $m_{\tilde{\Phi}^8} \pm \delta m$ |

$$M_{\Phi^D_{s,p}} \approx m_{\tilde{\Phi}_D} \approx \frac{3\sqrt{3}}{8} \epsilon_1 \frac{\lambda^2_\Phi}{y} M_X \sim 0.5 \text{ GeV} \Rightarrow \epsilon_1 \sim 10^{-3} \times \frac{y}{\lambda^2_\Phi}, \tag{8}$$

which justifies our assumption of perturbative calculations so far.

C. Interactions and heavier particle decay widths

We have demonstrated so far that the state X^1_p is the lightest stable states in our model. In principle, the supersymmetric partner states of X^1_p and other states in the superfield of X^2 could also be stable on a cosmological time scale. Similarly, additional interactions are required to mediate the decay of the other states of the PNGB supermultiplet aside from Φ^D_s. Because of the $\mathcal{O}(\text{MeV})$ mass splitting among some states and in order for not having too much kinematic suppression factors, we introduce the following "neutrino portal" to mediate the heavier state decays. The relevant higher-dimensional operators in the superpotential are

$$W_{\text{decay}} \supset \frac{1}{\Lambda_X}[a_{11}X_1\bar{X}_1 + a_{12}X_1\bar{X}_2 + a_{22}X_2\bar{X}_2]LH_u$$
$$+ \frac{1}{\Lambda_\Phi} \Phi^D\bar{\Phi}^D LH_u, \tag{9}$$

where we have ignored the flavor index for the SM leptons. For the first term, we keep the different flavor couplings independent for the moment, for reasons that we discuss in detail below.

Based on the first operator in Eq. (9), other states in the superfields X^1 and X^2 can either directly or in the cascade way decay into X^1_p. Assuming $\delta m' > \delta m$, the allowed decay channels are $X^1_s, X^2_s, X^2_p \to \tilde{X}^{1(2)}$ and $\tilde{X}^{1(2)} \to X^1_p$. Their decay widths are calculated as

$$\Gamma[X^i_{s(p)} \to \tilde{X}^j + \nu_L] = \Gamma[X^i_{s(p)} \to \tilde{X}^j + \bar{\nu}_L]$$
$$= \frac{a^2_{ij} v^2_u [M_{X^i_{s(p)}} - m_{\tilde{X}^j}]^2}{16\pi M_X \Lambda^2_X}, \tag{10}$$

$$\Gamma(\tilde{X}^i \to X^j_p + \nu_L) = \Gamma(\tilde{X}^i \to X^j_p + \bar{\nu}_L)$$
$$= \frac{a^2_{ij} v^2_u (m_{\tilde{X}^i} - M_{X^j_p})^2}{32\pi M_X \Lambda^2_X}, \tag{11}$$

where $i, j = 1, 2$, $v_u/\sqrt{2} = \langle H^0_u \rangle$. For a reasonably low cutoff scale, all heavier dark matter states decay fast to be treated as unstable particles for both dark matter thermal and indirect signal calculations. For instance, choosing $M_X = 1 \text{ TeV}$, $\Lambda_X = 1000 \text{ TeV}$, $v_u \sim 246 \text{ GeV}$, $a_{ij} = 1$ and a mass splitting of 1 MeV, the lifetime of these states are evaluated to be $\mathcal{O}(10^{-6}\text{s})$. We also note that both decays

into neutrinos and antineutrinos can happen here because of the lepton-number violating interactions in our model.

For the real scalar of the PNGB, Φ_s^D, it's mixing with the CP-even Higgs in the MSSM can be derived from the two interaction terms, $\epsilon_2 \Phi^D H_u H_d$ and $\mu H_u H_d$, in the superpotential. The mixing angle is at the order of $\epsilon_2 v_u \mu / M_h^2$. For the mass range between the muon and kaon thresholds, $2m_\mu < M_{\Phi_s^D} < 2M_K$, the leading decay channel is

$$\Gamma(\Phi_s^D \to \mu + \bar{\mu}) = \frac{\epsilon_2^2 \mu^2 m_\mu^2}{4\pi M_h^4} M_{\Phi_s^D} \left(1 - \frac{4m_\mu^2}{M_{\Phi_s^D}^2}\right)^{3/2}$$
$$\sim 1.8 \times 10^{-16} \text{ GeV}, \qquad (12)$$

for $\mu \sim 1$ TeV, $\epsilon_2 \sim 10^{-4}$, and $M_{\Phi_s^D} \sim 0.5$ GeV. The corresponding lifetime of this PNGB is around 10^{-11} s. The superpartner states of Φ_s^D can have decays mediated by the second operator in Eq. (9) and have prompt decay widths for a not-too-high cutoff scale, Λ_Φ. The main decay of the PNGF is $\tilde{\Phi}^D \to \Phi_s^D + \nu_L(\overline{\nu_L})$ with its formula given by

$$\Gamma[\tilde{\Phi}^D \to \Phi_s^D + \nu_L(\overline{\nu_L})] = \frac{v_u^2 (m_{\tilde{\Phi}^D} - M_{\Phi_s^D})^2}{32\pi M_{\Phi_s^D} \Lambda_\Phi^2}. \qquad (13)$$

The decay of Φ_p^D is $\Phi_p^D \to \tilde{\Phi}^D + \nu_L(\overline{\nu_L})$ with its decay width calculated as

$$\Gamma[\Phi_p^D \to \tilde{\Phi}^D + \nu_L(\overline{\nu_L})] = \frac{v_u^2 (M_{\Phi_p^D} - m_{\tilde{\Phi}^D})^2}{16\pi M_{\Phi_s^D} \Lambda_\Phi^2}. \qquad (14)$$

Both of them have a lifetime of 10^{-9} s for a mass splitting of order MeV, $M_{\Phi_s^D}$ GeV and $\Lambda_\Phi \sim 1000$ TeV.

Having discussed the decays of lighter dark matter states, we now turn to the properties of the heavy dark matter state, $X^3 \equiv X^R$, which will play the role of the resonance in our model. The decays of the X^R singlet states can be related in the Breit-Wigner formalism to annihilation cross section of the X_p^1 dark matter state. From the kinematics and parity conservation, the only possible mass-on-shell state to mediate resonant annihilation of dark matter states is the X_s^R scalar, so we pay special attention to this particle. Its decay width into the two dark matter states is

$$\Gamma(X_s^R \to X_p^1 + X_p^1) \approx \frac{\lambda_X^2 M_X^2}{8\pi \sqrt{s}} v_{\text{rel}}, \qquad (15)$$

with the $v_{\text{rel}} = 2\sqrt{1 - 4M_X^2/s}$ ($s = M_{X_s^R}^2$ for an on-shell resonance) or twice of the velocity of the particle in the final state. Another decay channel, $X_s^R \to \tilde{X}^D + \tilde{X}^D$, is at higher order of v_{rel} and given by

$$\Gamma(X_s^R \to \tilde{X}^D + \tilde{X}^D) = \frac{1}{64\pi} \lambda_X^2 \sqrt{s} v_{\text{rel}}'^3, \qquad (16)$$

with $v_{\text{rel}}' = 2\sqrt{1 - 4M_{\tilde{X}_D}^2/s}$ and suppressed. From an explicit calculation, one can show that there is no cubic interaction among X_s^R and two X_s^1 scalars. Finally, there are also decays into PNGBs and PNGF's, their decay branching widths are

$$\Gamma(X_s^R \to X_s^1 + \Phi_s^D) = \Gamma(X_s^R \to X_s^2 + \Phi_s^D) = \frac{3}{512\pi} y^2 M_X, \qquad (17)$$

$$\Gamma(X_s^R \to X_p^1 + \Phi_p^D) = \Gamma(X_s^R \to X_p^2 + \Phi_p^D) = \frac{27}{512\pi} y^2 M_X, \qquad (18)$$

$$\Gamma(X_s^R \to \tilde{X}^1 + \tilde{\Phi}^D) = \Gamma(X_s^R \to \tilde{X}^2 + \tilde{\Phi}^D) = \frac{9}{256\pi} y^2 M_X. \qquad (19)$$

It is understood that only one component of the Φ^D state enters the final state because of the $SU(2)_f$ symmetry for this decaying interaction.

IV. RESONANT ANNIHILATIONS

In our model, the leading annihilation channels for dark matter at the current universe are dominated by the one mediated by an on-shell resonance. Specifically, we have the semiannihilation processes of

$$X_p^1 + X_p^1 \to X_s^{R\dagger} \to X_p^{1,2\dagger} + \Phi_p^{D\dagger}, X_s^{1,2\dagger} + \Phi_s^{D\dagger}, \tilde{\overline{X}}_{1,2} + \tilde{\overline{\Phi}}_D. \qquad (20)$$

All the other states in X^1 and X^2 will decay into X_p^1. The states, Φ_p^D and $\tilde{\Phi}^D$, decay to Φ_s^D, which decays to two muons in the SM. So, from the cosmic ray positron signal point of view, we can sum all the annihilation channels and group them together as the outgoing Breit-Wigner width.

Using the standard nonrelativistic Breit-Wigner formula [31], the dark matter annihilation cross section $X_p^1 X_p^1 \to X_s^{R\dagger} \to (X^{1,2\dagger}, \tilde{\overline{X}}_{1,2}) + (\Phi^{D\dagger}, \tilde{\overline{\Phi}}_D)$ is given by

$$\sigma_{X_p^1 X_p^1}(\sqrt{s}) = 2 \frac{(2J+1)}{(2s_1+1)(2s_2+1)} \frac{4\pi}{k k_{\text{in}}}$$
$$\times \left[\frac{\Gamma^2/4}{(\sqrt{s} - M_{X_s^R})^2 + \Gamma^2/4}\right] B_{\text{in}} B_{\text{out}}. \qquad (21)$$

Here, the $J = 0$ is the resonance spin; $s_1 = s_2 = 0$ are the dark matter spin; $k = M_X v_{\text{rel}}/2$ is the center-of-mass momentum in the initial state with v_{rel} as the relative speed of the two dark matter particles; $k_{\text{in}} \approx M_X \sqrt{1 - 4M_X^2/s} = M_X v_{\text{rel}}/2$; the center-of-mass energy $s \approx 4M_X^2 + M_X^2 v_{\text{rel}}^2$; the resonance mass is $M_{X_s^R} = 2M_X + \delta m + \delta m''$. We

include an overall factor of two in Breit-Wigner formula to account for the fact that the initial state is made up of identical particles [32]. The total width of the resonance as a function of s is

$$\Gamma(X_s^R) = \frac{1}{16\pi}\lambda_X^2 M_X v_{\text{rel}} + \frac{3}{16\pi}y^2 M_X. \qquad (22)$$

Introducing the resonance velocity, $v_R \approx \sqrt{M_{X_s^R}^2/M_X^2 - 4} \approx 2\sqrt{(\delta m + \delta m'')/M_X}$, we have the effective annihilation rate

$$
\begin{aligned}
(\sigma v_{\text{rel}})_{\text{eff}} &= \frac{1}{2}\sigma_{X_p^1 X_p^1}(\sqrt{s}) \\
&= \frac{3y^2\lambda_X^2}{4\pi M_X^2[(v_{\text{rel}}^2 - v_R^2)^2 + \frac{1}{64\pi^2}(3y^2 + \lambda_X^2 v_{\text{rel}})^2]},
\end{aligned}
\qquad (23)
$$

where the factor of 1/2 for the first equality comes from the fact that our dark matter field is a complex scalar instead of a real scalar.

In the narrow width approximation, the term inside the square bracket of Eq. (21) can be replaced by $\frac{\pi\Gamma}{2}\delta(\sqrt{s} - M_{X_s^R})$. The annihilation rate takes the form of

$$
\begin{aligned}
(\sigma v_{\text{rel}})_{\text{eff}} &= 2\frac{8\pi}{M_X^2 v_{\text{rel}}}\frac{\pi\Gamma}{2}B_{\text{in}}B_{\text{out}}\delta(\sqrt{s} - M_{X_s^R}) \\
&= 2\frac{8\pi}{M_X^2 v_{\text{rel}}}\frac{\pi\Gamma}{2}B_{\text{in}}B_{\text{out}}\frac{2}{M_X v_{\text{rel}}}\delta(v_{\text{rel}} - v_R).
\end{aligned}
\qquad (24)
$$

Substituting the widths and branching ratios into Eq. (24), we have the annihilation rate to be

$$(\sigma v_{\text{rel}})_{\text{eff}} = \frac{3\pi\lambda_X^2 y^2}{v_{\text{rel}}(3y^2 + \lambda_X^2 v_{\text{rel}})M_X^2}\delta(v_{\text{rel}} - v_R). \qquad (25)$$

A. Annihilation rates in our Galactic halo and the CMB era

Assuming the normalized and isotropic dark matter velocity distribution is

$$f(v) = \frac{4\pi}{(\pi v_0^2)^{3/2}}e^{-v^2/v_0^2}, \qquad (26)$$

such that $\int_0^\infty f(v)v^2 dv = 1$. Here, $v_0 \approx 220$ km/s for the Standard Halo model [33]. We ignore the distortions due to the escape velocity and will have the resonance velocity be close to v_0. In terms of the two dark matter relative velocity, the normalized distribution is

$$f(v_{\text{rel}}) = \frac{4\pi}{(2\pi v_0^2)^{3/2}}e^{-v_{\text{rel}}^2/2v_0^2}. \qquad (27)$$

Using the delta function approximation, the averaged dark matter annihilation rate is

$$
\begin{aligned}
\langle(\sigma v_{\text{rel}})_{\text{eff}}\rangle &\equiv \int dv_{\text{rel}} v_{\text{rel}}^2 (\sigma v_{\text{rel}})_{\text{eff}} f(v_{\text{rel}}) \\
&= \frac{3\pi\lambda_X^2 y^2 v_R}{(3y^2 + \lambda_X^2 v_R)M_X^2}\frac{4\pi}{(2\pi v_0^2)^{3/2}}e^{-v_R^2/2v_0^2},
\end{aligned}
\qquad (28)
$$

which provides a good approximation for the narrow-width case.

In Fig. 3, we show the averaged annihilation rate as a function of v_0 for fixed model parameters. For a benchmark model point with $M_X = 1.7$ TeV, $v_R = 2.75 \times 10^{-3}$, $\lambda_X = 0.027$ and $y = 1 \times 10^{-3}$, we have the prediction of $\langle\sigma v_{\text{rel}}\rangle_{\text{eff}}^{\text{AMS}} = 7.9 \times 10^{-23}$ cm³/s. For a small value of $v_0 \ll 10^{-4}$, relevant for the CMB era, the predicted annihilation rate is $\langle\sigma v_{\text{rel}}\rangle_{\text{eff}}^{\text{CMB}} = 1.2 \times 10^{-23}$ cm³/s, around a factor of seven smaller than the prediction for AMS-02. One can see that the resonant dark matter could provide a sufficiently large rate for AMS-02 and at the same time satisfy the constraints from CMB. To understand better of this good feature of resonance effect, we can set $v_{\text{rel}} = 0$ in Eq. (23) and have the averaged annihilation rate at the CMB era to be

$$\langle\sigma v_{\text{rel}}\rangle_{\text{eff}}^{\text{CMB}} \approx \frac{3y^2\lambda_X^2}{4\pi M_X^2 v_R^4}, \qquad (29)$$

and the ratio of the two annihilation rates as

$$\frac{\langle\sigma v_{\text{rel}}\rangle_{\text{eff}}^{\text{CMB}}}{\langle\sigma v_{\text{rel}}\rangle_{\text{eff}}^{\text{AMS}}} = \frac{(3y^2 + \lambda_X^2 v_R)}{4\pi^2 v_R^5}\frac{(2\pi v_0^2)^{3/2}}{4\pi}e^{v_R^2/2v_0^2}. \qquad (30)$$

For $v_R \sim v_0$, one can have a suppressed annihilation rate during the CMB era only when $y \lesssim v_R$ and $\lambda_X \lesssim v_R^{1/2}$.

B. Dark matter thermal relic abundance and kinetic decoupling

In order to determine the thermal relic abundance, we first note that resonant models are known to have the potential for early kinetic decoupling [21,23,34]. The completion of the process of thermal and kinetic freeze-out is slow, due to the increasing resonant enhancement of the semiannihilation cross section. For kinetic decoupling in our semiannihilation case, this behavior is exaggerated by the large energy of the final state dark matter, which is produced with a momentum of order $3M_X/4$. If the semi-annihilation process dominates the evolution of the dark matter energy too early or if the dark matter fails to maintain kinetic equilibrium even with itself, then the dark matter could end up going through the resonance in semiannihilation too late, leading to phenomenological issues such as an overproduction of dark matter or a large cross section around the time of recombination that is ruled

out by CMB data. In order to satisfy these phenomenological constraints, we consider parameter space for the SRDM model that has a schematic cosmological evolution as follows:

(i) Dark matter loses chemical equilibrium and ceases to follow a Boltzmann suppressed number density. The dark matter continues to deplete due to an enhanced semiannihilation cross section approaching the resonance. It maintains kinetic equilibrium dominantly via resonantly enhanced elastic scattering off neutrinos, effectively via the a_{12} term in Eq. (9). The diagram for elastic scattering is shown in Fig. 4.

(ii) As the dark matter and SM bath temperature goes through the resonance at $T \sim 1$ MeV in both the semiannihilation and neutrino scattering processes, these processes become significantly less efficient. The abundance of dark matter per comoving volume freezes out to nearly its present day value. There is a large initial energy dump into the dark matter at $x \gtrsim 1/\delta$, as kinetic equilibrium is lost and the semiannihilation process comes to dominate the kinetic evolution of the dark sector. After this energy dump, the dark sector cools slowly as it redshifts. There are two possibilities for the dark matter evolution at this point. If a new mediator is introduced that couples only to the dark matter, then the dark matter may thermalize among itself. Otherwise, the quasi-relativistic dark matter produced in residual semiannihilation comes to dominate and has only a small annihilation rate. The choice between the two possibilities has little relevance for indirect detection, CMB constraints or the relic abundance of dark matter, but may have other interesting cosmological consequences as the dark sector gains a large amount of extra energy from the residual semiannihilation events.

(iii) As the dark matter slowly cools in either way, it eventually crosses through the semiannihilation resonance pole once more, then decouples and cools rapidly compared to the SM sector, with only a small residual semiannihilation rate.

In the remainder of this subsection, we study each of these pieces in turn.

The process of chemical decoupling is governed by the Boltzmann equation for the number density, n, of X_p^1. It is convenient to define the quantity "yield", $Y \equiv n/s$ with the entropy $s = 2\pi^2 g_* T^3/45$. Then, the resulting Boltzmann equation for Y due to the resonant semiannihilation process [35–37] is

$$\frac{dY}{dx} = -\frac{1}{2}\frac{s}{xH} Y(\langle \sigma v_{\rm rel}\rangle_{{\rm eff},x_{\rm DM}} Y - \langle \sigma v_{\rm rel}\rangle_{{\rm eff},x} Y^{\rm eq}), \quad (31)$$

where for future convenience we define $x = M_X/T$ with T as the SM particle temperature and $x_{\rm DM} = M_X/T_X$ for the

dark matter temperature could be different from x after the dark matter sector kinetically decouples from the SM sector. We include a factor of $1/2$ that arises due to having identical particles in the initial state phase space integration. The number density is the summation of dark matter and antidark matter number densities. In the radiation dominated era, $H = (8\pi\rho/3M_{\rm pl}^2)^{1/2}$, $t = 1/(2H)$, $\rho(T) = g_*\pi^2 T^4/30$ and $n^{\rm eq}(T) = g(M_X T/2\pi)^{3/2} e^{-M_X/T}$, where we will choose $g_* = 10.75$ because of the freeze-out temperature at a few MeV and $g = 2$ for the dark matter degrees of freedom. $Y^{\rm eq} \equiv n^{\rm eq}/s$ is the kinetic and chemical equilibrium of Y. The final dark matter relic abundance has $\Omega_{\rm DM} = M_X n/\rho_c = M_X s_0 Y(\infty)/\rho_c$ with the critical density $\rho_c = 3H_0 M_{\rm pl}^2/8\pi = 1.0539 \times 10^{-5} h^2$ GeV cm^{-3} and $s_0 = 2889.2$ cm^{-3} as the entropy today. From the Planck collaboration, the measured dark matter energy density has $\Omega_{\rm DM} h^2 = 0.1199 \pm 0.0022$ [15].

For the dark matter thermal averaged annihilation rate and in terms of the parameter $x_{\rm DM}$, we have

$$\langle \sigma v_{\rm vel}\rangle_{{\rm eff},x_{\rm DM}} = \frac{x_{\rm DM}^{3/2}}{2\pi^{1/2}} \int_0^\infty dv_{\rm rel} v_{\rm rel}^2 e^{-v_{\rm rel}^2 x_{\rm DM}/4} (\sigma v_{\rm vel})_{\rm eff}. \quad (32)$$

To understand the behavior during the different periods of the extended freeze-out process, we use the general parametrization for the resonance annihilation (see Appendix A for details),

$$(\sigma v_{\rm rel})_{\rm eff} = \sigma_0 \frac{\delta^2 + \gamma^2}{(\delta - v_{\rm rel}^2/4)^2 + \gamma^2}, \quad (33)$$

with the resonance speed $v_{\rm R} = 2\sqrt{\delta}$ and the parameter γ related to the resonance width. During earlier times with $x \ll 1/\delta$, the resonance annihilation cross section is growing with $x_{\rm DM}$ and is given to very good approximation by

$$\langle \sigma v_{\rm rel}\rangle_{{\rm eff},x_{\rm DM}} = \sigma_0 \frac{2\sqrt{\pi} x_{\rm DM}^{3/2} \delta^{5/2}}{\gamma}. \quad (34)$$

When the resonant semiannihilation process is efficient, the dark matter follows its equilibrium distribution as usual, with $Y = Y^{\rm eq}$. The chemical equilibrium is lost when $\langle \sigma v_{\rm rel}\rangle_{\rm eff} n^{\rm eq} \lesssim H$ as in the usual s-wave or p-wave annihilation cases, and with the chemical decoupling $x_{\rm cd} \sim 20$ as usual.

After chemical decoupling, since the cross section grows as the temperature approaches the resonant temperature at $x_{\rm DM} \to 1/\delta$, the dark matter number density continues to decrease as a power law, instead of rapidly freezing out to a plateau as in the standard freeze-out calculation. At this point, $Y^{\rm eq} \ll Y$ and can be neglected, so that the differential equation in Eq. (31) can be solved easily and provides a power-law behavior for Y before freeze-out at $x_{\rm fo}$

$$Y \propto x^{-1/2}, \quad \text{for } x_{\text{cd}} < x < x_{\text{fo}}. \tag{35}$$

We have assumed at this point that kinetic equilibrium is maintained between dark matter and the SM sector. This is a nontrivial assumption and we consider this further now. At temperatures below the weak scale, there are two processes that could potentially maintain kinetic equilibrium: scattering off the relativistic Φ_D states via quartic interactions as well as via t-channel Φ_3 and Φ_8 exchange or scattering off neutrinos via the first operator in Eq. (9). The first of these possibilities necessarily is lost at $x \sim M_X/M_{\Phi_D}$ as the Φ_D become nonrelativistic and rapidly decay away. We therefore consider parameter space where the second possibility is large. It in fact receives a resonant enhancement as well, as is clear from the structure of dominant diagram in Fig. 4. We consider only the dark matter flavor-changing superpotential operators $X_1 \overline{X}_2 L H_u$ to be significant. If the dark matter flavor-conserving operators have large coefficients, then they contribute to the neutrino Majorana mass at an unacceptably large level. Kinetic equilibrium is maintained until $x \sim 1/\delta$ for $\Lambda_X/a_{12} \sim v_{\text{EW}}$, the Higgs VEV. The operator should be UV completed at the TeV scale, though we do not study such a completion further in this work. To study the process of kinetic decoupling and freeze-out, we follow Refs. [38,39] and define a normalized measure of the dark matter temperature

$$y \equiv \frac{M_X T_{\text{DM}}}{s^{2/3}} \equiv \frac{4g}{Y s^{5/3}} \int \frac{d^3 p}{(2\pi)^3} \mathbf{p}^2 f_X(p), \tag{36}$$

where f_X is the phase space distribution of dark matter. The variable y is defined such that after kinetic freeze-out, $y \to$ constant and in kinetic equilibrium $y \propto x^{-1}$ up to changes in the composition of the SM relativistic fluid. The evolution of y is governed by a higher moment of the Boltzmann equation and is given in this era by [38]

$$\frac{dy}{dx} = -\frac{1}{Hx} 2M_X c(T)(y - y_{\text{EQ}}), \tag{37}$$

with $c(T)$ defined and calculated in Appendix B. We will choose the relevant model parameters in $c(T)$ to keep dark matter kinetically coupled to SM before its chemical freeze-out. For the purposes of determining the relic abundance this is equivalent to the choice of kinetic decoupling temperature as $x_{\text{kd}} = \infty$.

At $x \gtrsim 1/\delta$, the abundance of dark matter per comoving volume freezes out. At this point, kinetic equilibrium with the SM sector is lost as well. Since the abundance of dark matter is already frozen, this will turn out to not have a large effect on the phenomenology relevant to this work, provided that dark matter is sufficiently cold by the time the SM goes through recombination, which is the case regardless of the kinetic decoupling phenomenology. We therefore only briefly describe a couple of possibilities here and

present some in more details in Appendix B. If there is insufficient interaction within the dark sector, then the dominant dark matter population will be produced with an energy very close to $3M_X/4$ as predicted from semiannihilation with an initial state at rest. As the universe expands, this initial energy redshifts. Even if there is sufficient interaction among dark matter particles, the energy dumped into the dark matter fluid due to residual semiannihilations is significant. In either case, the dark sector does eventually cool back to a point where very few particles have a velocity above the resonant velocity, at which point the semiannihilation process shuts off and the dark matter quickly redshifts to a very cold distribution. At this point, the averaged dark matter semiannihilation goes to its zero velocity level. All of this happens before recombination in the SM sector, so it does not have any bearing on the CMB bound as derived above.

In Fig. 5, we show the behavior of the yield function in terms of the temperature parameter x. We show the results from solving the full coupled differential equations of Eqs. (31), (37) and (B9) in the red curve and from the approximate approach in the blue curve. The good overlapping between those two curves justify our understanding that the kinetic decoupling in our model happens after the chemical freeze-out. Furthermore, the final freeze-out temperature is around 1 MeV and larger than the one without considering kinetic decoupling. We also note that the temperature of dark matter has a nontrivial evolution even after the chemical freeze-out. Since this is not important for the main purpose of our current paper, we show its behavior in Fig. 11 of Appendix B.

In Appendix A, we have used a general parametrization for the resonance annihilation in Eq. (33) and found a simple relation of parameters to fit the thermal relic abundance

$$\frac{\gamma}{\sigma_0 \delta^2} = 1.35 \times 10^{10} \text{ GeV}^2. \tag{38}$$

Applying that to our model, the condition to satisfy the thermal dark matter relic abundance is approximately

$$M_X^2 \frac{3y^2 + \lambda_X^2 v_R}{y^2 \lambda_X^2} \approx 2.0 \times 10^{10} \text{ GeV}^2, \tag{39}$$

Though in later sections the (y, λ_X) value we use to satisfy the dark matter thermal relic abundance is obtained by parameter space scanning, we would still point out that the result from Eq. (39) matches very well to the exact value.

V. FIT TO AMS-02 DATA AND CONSTRAINTS

In the SRDM model, there is thus a nontrivial interplay between various constraints and the fit to the cosmic positron excess. Annihilation just after recombination must
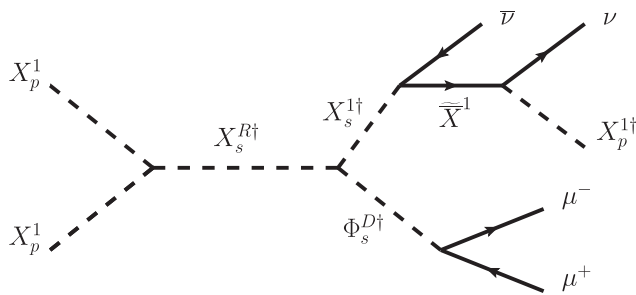
FIG. 2. A representative Feynman diagram for the semianni-hilation of dark matter mediated by a resonance. An unbroken $\mathcal{Z}_3$ is responsible for the stability of dark matter. The muon energies can be up to the dark matter mass, while the neutrino energies are $\mathcal{O}(\text{MeV})$, determined by the mass difference of different dark matter states.

be sufficiently small to not generate CMB distortions, annihilation during freeze-out must be large enough to not overproduce dark matter, and annihilation in the Milky Way today must have the correct rate to generate the observed cosmic ray spectrum, while not producing too many gamma rays. In this section, we determine model parameters that provide a good fit to the observed positron excess. We then turn to the remaining constraints on the model from cosmic gamma ray observations and CMB spectrum measurements.

A. Fitting the positron excess

The primary motivation for the models considered in this work are as explanation for the excess in the cosmic ray positron flux above a few GeV, as seen at PAMELA [5], Fermi-LAT [6], and AMS-02 [2]. Dark matter annihilations have been considered as a potential explanation for the excess in the past. Annihilations into quarks, gluons or electroweak gauge bosons can provide a good fit to the
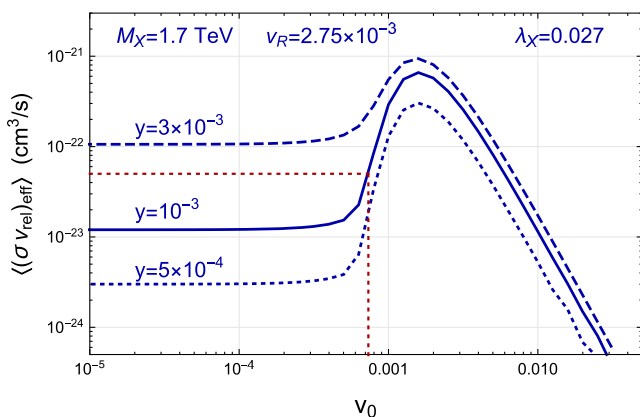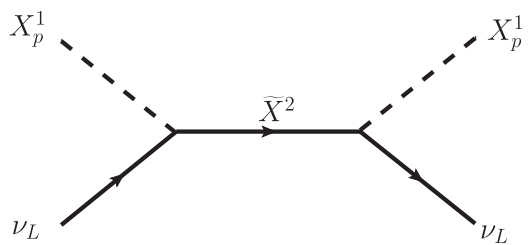

FIG. 4. Feynman diagrams for the dark matter elastically scatter off neutrinos via a resonance in the s-channel.

positron data alone, but there are additional stringent constraints from gamma ray searches at VERITAS [40], Fermi-LAT [27], MAGIC [41], and H.E.S.S. [42] that rule out these possibilities by orders of magnitude [24]. Direct annihilations into electron pairs do not provide a good fit to the data [24], while direct annihilations into μ and τ pairs provide good fits, but are also highly constrained by the same gamma ray searches. The remaining possibility considered in the literature that can have a sufficiently large cross section, while providing a reasonable fit to the data, is annihilation into two light particles that each decay to a pair of leptons, generally referred to as $\phi\phi \to 4\ell$ models [24–26] with $\ell = e, \mu$. Such models additionally have relatively weak constraints from annihilated products distorting the CMB spectrum [43], though the constraints are still significant and we revisit them below. The models considered here effectively behave like $\phi\phi \to 4\ell$ bench-mark models for sufficiently light PNGBs Φ_s^D (see the annihilation Feynman diagram in Fig. 2), up to a remapping of the parameter space that we discuss below. If the mass of the PNGB is above $2m_K$, then significant hadronic modes open up generate conflicts with the data outlined above, so we consider Φ_s^D masses below $2m_K \sim 1$ GeV and the

FIG. 3. Averaged annihilation rate as a function of the averaged dark matter velocity v_0. For the dark matter velocity in our galaxy halo with $v_0 \approx 7.3 \times 10^{-4}$, the averaged annihilation rate is 5×10^{-23} cm³/s for the benchmark point with $y = 1 \times 10^{-3}$, while it is 1.2×10^{-23} cm³/s for a small velocity at the CMB era.
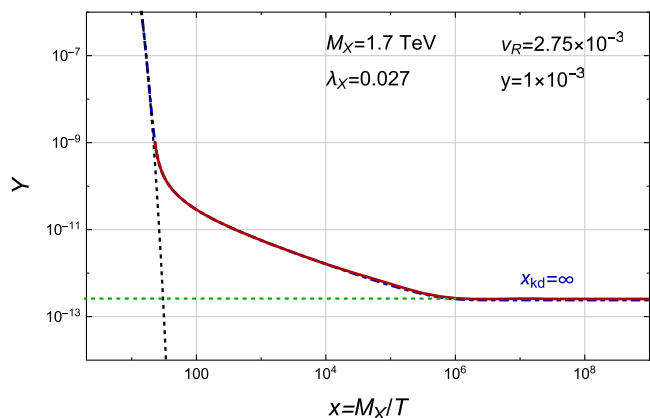

FIG. 5. The dark matter yield as a function of the temperature parameter x for a benchmark point. The red line is the result from solving the full coupled Eqs. (31), (37) and (B9), while the blue line is using $x_{\text{DM}} = x$ and only solving Eq. (31), assuming $x_{\text{kd}} = \infty$ as phenomenologically justified in the text. This benchmark point has $\Omega_X h^2 \approx 0.12$.

dominant decay into $\mu^+\mu^-$, which subsequently decay into e^+e^- to increase the positron fraction. For our benchmark point, we will choose $M_{\Phi^D_s} = 0.5$ GeV.

In order to fit the AMS-02 data, we calculate the contribution of our model to the positron and electron fluxes. Noting that the energy of a Φ^D_s particle produced in the annihilation channel studied in the model above is

$$E_{\Phi^D_s} = \frac{3M_X}{4}. \qquad (40)$$

from neglecting the tiny Φ^D_s particle mass and the small dark matter kinetic energy. This is different from the ordinary annihilation case with $E_{\Phi^D_s} = M_X$ in Refs. [24–26]. We can then successively determine the spectrum of positrons from their parent chain as for example in Ref. [44]. The spectrum of positrons produced at the time and location of DM annihilation is given by

$$\frac{dN}{dE} = \frac{4}{3M_X} \int_{\frac{4E}{3M_X}}^{1} \frac{dx_\phi}{x_\phi} \int_{x_\phi}^{1} \frac{dx_\mu}{x_\mu} \frac{dN}{dx_\mu}, \qquad (41)$$

where $x_\phi = 2E_e^{\Phi^D_s\,\mathrm{rest}}/M_{\Phi_D}$, $x_\mu = 2E_e^{\mu\,\mathrm{rest}}/m_\mu$, and dN/dx_μ is the spectrum of positrons produced in a muon decay in the muon rest frame. By $E_e^{P\,\mathrm{rest}}$, we mean the energy of the positron in the rest frame of P. This last spectrum can be calculated from the four-Fermi interaction for muon decay and is given by

$$\frac{dN}{dx_\mu} = 2x_\mu^2(3 - 2x_\mu), \qquad (42)$$

such that the resulting positron spectrum in the dark matter rest frame is given by

$$\frac{dN}{dE} = \frac{2}{27M_X}(-8x^3 + 27x^2 - 30\log x - 19), \qquad (43)$$

with $x = 4E/3M_X$ as the fraction of electron energy over its maximum energy.

The spectrum at annihilation is related to the spectrum observed at detectors at or near the Earth by propagation through the galactic medium. For this work, we calculate the spectrum using the PPPC4DMID package [45]. In their formalism, the flux of positrons at Earth is given by

$$\frac{d\Phi}{dE} = \frac{v_e}{4\pi b(E)} \frac{1}{2} \left(\frac{\rho_\odot}{M_X}\right)^2 \int_E^{3M_X/4} dE' \langle\sigma v\rangle_{\mathrm{eff}} \frac{dN}{dE'} I(E,E'), \qquad (44)$$

where $v_e \approx c$ is the electron velocity, $\rho_\odot = 0.3$ GeV/cm^3 is the local dark matter energy density. $b(E)$ is the energy loss function of electron and positron at the location of the Earth and $I(E,E')$ is the generalized halo function at the Earth.

The energy loss function $b(E,\vec{x})$ is the summation of four different terms, which characterize the energy loss of a electron or positron by Coulomb interaction and ionization, bremsstrahlung, inverse Compton scattering (ICS) and synchrotron emission, respectively. We explicitly write out the form of the piece for ICS for later convenience,

$$b_{\mathrm{ICS}} = 3c\sigma_T \int_0^\infty d\epsilon\epsilon \int_{1/4\gamma^2}^1 dq n(\epsilon) \frac{(4\gamma^2 - \Gamma_\epsilon)q - 1}{(1 + \Gamma_\epsilon q)^3}$$
$$\times \left[2q\ln q + q + 1 - 2q^2 + \frac{1}{2}\frac{(\Gamma_\epsilon q)^2}{1 + \Gamma_\epsilon q}(1 - q)\right], \qquad (45)$$

with $\sigma_T = 8\pi\alpha_{\mathrm{em}}^2/(3m_e^2)$, $\gamma = E/m_e$, $\Gamma_\epsilon = 4\epsilon\gamma/m_e$, $n(\epsilon)$ as the number density of photons with energy ϵ. For the expressions of the other three terms we refer the reader to Ref. [45] for detail. The generalized halo function $I(E,E')$ is the solution to the following differential equation [46]:

$$\nabla^2\tilde{I}(\epsilon,\epsilon',\vec{x}) + \frac{1}{\mathcal{K}_0\tau_\odot\epsilon^{\delta-2}}\frac{\partial}{\partial\epsilon}\left[\frac{b(\epsilon,\vec{x})}{b_T(\epsilon)}\tilde{I}(\epsilon,\epsilon',\vec{x})\right] = 0, \qquad (46)$$

$$I(E,E',\vec{x}) = \tilde{I}(\epsilon,\epsilon',\vec{x})\left[\frac{b_T(\epsilon)}{b(\epsilon,\vec{x})}\left(\frac{\rho(\vec{x})}{\rho_\odot}\right)^2\right]^{-1}, \qquad (47)$$

with the boundary condition

$$\begin{cases} \tilde{I}(\epsilon',\epsilon',\vec{x}) = \frac{b_T(\epsilon')}{b(\epsilon',\vec{x})}\left(\frac{\rho(\vec{x})}{\rho_\odot}\right)^2, \\ \tilde{I}(\epsilon,\epsilon',\vec{x})|_{\mathrm{galaxy\ boundary}} = 0. \end{cases} \qquad (48)$$

Here $\epsilon = E/$GeV, $\epsilon' = E'/$GeV, $\tau_\odot = 5.7 \times 10^{15}$ sec, $\rho(\vec{x})$ is the dark matter density profile and $b_T(\epsilon) = \epsilon^2$ GeV/$\tau_\odot$ is the normalizing factor. $\mathcal{K}_0$ and δ depend on propagation model, which can be found in [45]. For our benchmark, we use "EinastoB" profile, "MAX" propagation model and "MF2" galactic magnetic field model, though several other propagation models and magnetic field models are also shown for comparison.

In order to predict the positron fraction, which is the most sensitive observable to the model presented by cosmic ray experiments, we must also determine the background electron and positron fluxes. The positron fraction can be written as

$$\frac{\Phi^{e^+}}{\Phi^{e^-} + \Phi^{e^+}} = \frac{\Phi^{e^+}_{\mathrm{sig}} + \Phi^{e^+}_{\mathrm{bkg}}}{2\Phi^{e^+}_{\mathrm{sig}} + \Phi^{e^-}_{\mathrm{bkg}} + \Phi^{e^+}_{\mathrm{bkg}}}, \qquad (49)$$

where we have used the fact that the signal positron and electron fluxes are the same in our model. Here and where relevant, we use Φ to represent $d\Phi/dE$ for continuous spectra and the flux in a given bin for binned distributions. The background positron flux $\Phi^{e^+}_{\mathrm{bkg}}$ appearing in the numerator of Eq. (49) must be taken from a theoretical
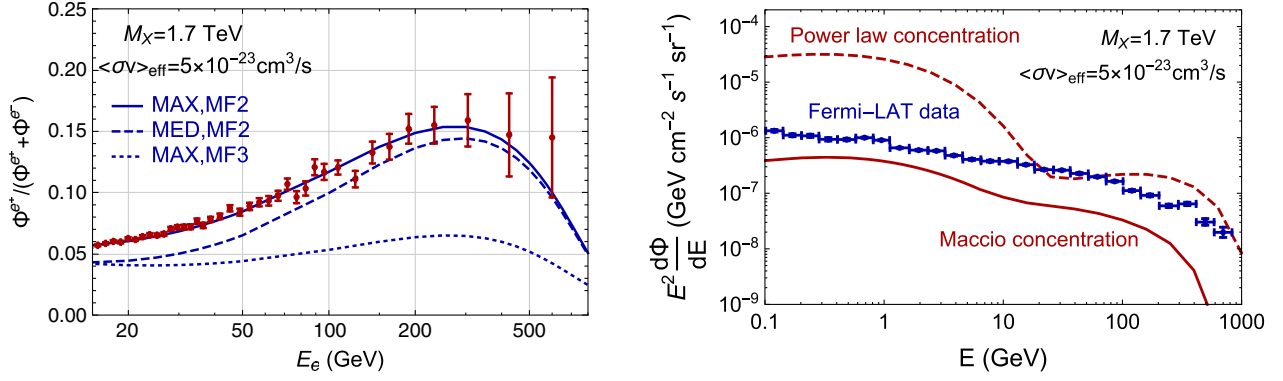
FIG. 6.   Left: Fits to the AMS-02 data with a heavy dark matter state, with dark matter profile EinastoB and propagation/magnetic field models (MAX, MF2), (MAX, MF3) and (MED, MF2). All fits assume $M_X = 1.7$ TeV, $M_{\Phi_s^D} = 0.5$ GeV and annihilation rates $\langle \sigma v \rangle_{\rm eff} = 5 \times 10^{-23}$ cm^3/s, using tools from Ref. [45]. The AMS measurement of the positron fraction is also shown for comparison. Right: Comparison of the gamma ray flux produced by our model and the constraints from Fermi-LAT measurements at $|b| > 20°$ [49].

model, but the background lepton fluxes may be taken either from a model or from data. We have applied both methods in this work. For the theoretical model method, we use the background fluxes from Refs. [47,48], given by

$$\frac{d\Phi_{\rm bkg}^{e^-}}{dE} = \frac{0.16 E^{-1.1}}{1 + 11 E^{0.9} + 3.2 E^{2.15}} + \frac{0.70 E^{0.7}}{1 + 110 E^{1.5} + 660 E^{2.9} + 580 E^{4.2}}, \quad (50)$$

and

$$\frac{d\Phi_{\rm bkg}^{e^+}}{dE} = \frac{4.5 E^{0.7}}{1 + 650 E^{2.3} + 1500 E^{4.2}}, \quad (51)$$

for E in GeV and $d\Phi/dE$ in GeV^{-1} cm^{-2} s^{-1} sr^{-1}. When applying data to determine the sum of electron and positron fluxes, we interpolate the data in Ref. [1] from AMS-02.

For a fixed dark matter mass of $M_X = 1.7$ TeV, we then determine a cross section that fits the available data. The data is largely taken from Ref. [2], though we supplement this data with a new data point presented in Ref. [3]. The data and the spectrum resulting from our model are shown in Fig. 6. For the benchmark "MAX" propagation model, the required annihilation rate is around 5×10^{-23} cm^3/s and matched by our benchmark model point in Fig. 3 or 5.

B. Gamma ray constraints

Given this fit to the AMS data, we now turn to the constraints from the flux of gamma-rays. The dominant of such constraints come from the isotropic diffusive gamma-ray background (IGRB) measured by Fermi-LAT. The contribution of dark matter annihilation to IGRB is characterized by two pieces, one from annihilations in the Milky Way and one from extragalactic annihilations. The calculation of both pieces are included in the PPPC4DMID package [46] and, in the model considered in this work, are

dominated by ICS. The gamma-ray flux from ICS within the Milky Way can be calculated by

$$\frac{d^2\Phi_{\rm IC\gamma}}{dE_\gamma d\Omega} = \frac{1}{E_\gamma^2} \frac{r_\odot}{4\pi} \frac{1}{2} \left(\frac{\rho_\odot}{M_X} \right)^2$$
$$\times \int_{m_e}^{3M_X/4} dE' \langle \sigma v \rangle_{\rm eff} \frac{dN}{dE'} I_{\rm IC}(E_\gamma, E', b, l), \quad (52)$$

where $I_{\rm IC}(E_\gamma, E', b, l)$ is the halo function of ICS, which is given by

$$I_{\rm IC}(E_\gamma, E', b, l) = 2 E_\gamma \int_{\rm l.o.s.} \frac{ds}{r_\odot} \left[\frac{\rho(\vec{x}(s, \theta))}{\rho_\odot} \right]^2$$
$$\times \int_{m_e}^{E'} dE \frac{\mathcal{P}_{\rm IC}(E_\gamma, E, \vec{x}(s, \theta))}{b(E, \vec{x}(s, \theta))}$$
$$\times I(E, E', \vec{x}(s, \theta)), \quad (53)$$

$$\mathcal{P}_{\rm IC}(E_\gamma, E, \vec{x}) = \frac{3\sigma_T}{4\gamma^2} \int_{1/4\gamma^2}^{1} dq \left[E_\gamma - \frac{E_\gamma}{4 q \gamma^2 (1 - \eta)} \right]$$
$$\times \frac{n(E_\gamma^0(q), \vec{x})}{q} \left[2q \ln q + q + 1 - 2q^2 \right.$$
$$\left. + \frac{1}{2} \frac{\eta^2}{1 - \eta} (1 - q) \right], \quad (54)$$

with

$$q = \frac{\eta}{\Gamma_E (1 - \eta)}, \qquad \Gamma_E = \frac{4 E_\gamma^0 E}{m_e^2}, \qquad \eta = \frac{E_\gamma}{E}. \quad (55)$$

Here, E_γ^0 is the initial energy of photon in the background bath, and $n(E_\gamma^0(q), \vec{x})$ is the photon distribution in the photon gas frame.

The formula of extragalactic gamma-ray flux for observer at redshift $z = 0$ is given by

$$\frac{d\Phi_{\text{EG}\gamma}}{dE_\gamma} = c \frac{1}{E_\gamma} \int_0^\infty dz' \frac{1}{H(z')(1+z')^4} j_{\text{EG}\gamma}(E_\gamma', z') e^{-\tau(E_\gamma, z')},$$

(56)

with $E_\gamma' = E_\gamma(1 + z')$. As there is no prompt gamma-ray production in the SRDM model, the local gamma-ray emissivity for annihilating dark matter $j_{\text{EG}\gamma}$ has the main contribution from inverse Compton radiation, which is expressed as

$$j_{\text{EG}\gamma}(E_\gamma', z') = \int_{m_e}^{3M_X/4} dE_e \frac{\mathcal{P}_{\text{IC}}^{\text{CMB}}(E_\gamma', E_e, z')}{b_{\text{IC}}^{\text{CMB}}(E_e, z')}$$
$$\times \int_{E_e}^{3M_X/4} dE \langle \sigma v \rangle_{\text{eff}} \frac{dN}{dE} B(z') \left[\frac{\bar{\rho}(z')}{M_X} \right]^2. \quad (57)$$

The power function $\mathcal{P}_{\text{IC}}^{\text{CMB}}$ has similar definition as Eq. (53), but has the photon distribution $n(E_\gamma^0, \vec{x})$ replaced with the CMB distribution at redshift z'. Similarly, the energy loss function $b_{\text{IC}}^{\text{CMB}}$ is just Eq. (45) with the corresponding replacement. The cosmological boost factor $B(z, M_{\min})$ characterizes the effect of dark matter clustering and is defined as

$$\frac{\langle \rho_{\text{DM}}^2(z) \rangle}{\bar{\rho}_{\text{DM}}^2(z)} \equiv B(z, M_{\min})$$
$$= 1 + \frac{\Delta_c}{3\bar{\rho}_{m,0}} \int_{M_{\min}}^\infty dM M \frac{dn}{dM}(M, z) f[c(M, z)],$$

(58)

with dn/dM is the halo mass function. The function $f(c)$ is given by

$$f(c) = \frac{c^3}{3} \left[1 - \frac{1}{(1+c)^3} \right] \left[\log(1+c) - \frac{c}{1+c} \right]^{-2}. \quad (59)$$

Within in the `PPPC4DMID` package, the halo concentration function $c(M, z)$ is set according to a choice of two models. For the "power-law" model, $c(M, z) = 6.5[H(z)/H_0]^{-2/3}(M/M_*)^{-0.1}$, with $M_* = 3.37 \times 10^{12} h^{-1} M_\odot$. For the model proposed by Macciò *et al.* [50] $c(M, z) = 3.9[H(z_c(M))/H(z)]^{2/3}$, where z_c is defined via $\sigma[M_*(z_c)]D(z_c) = 1.686$, with $\sigma(M)$ the root-mean-square over-density at the redshift $z = 0$ for a mass scale M and $D(z)$ the linear growth rate of perturbation.

Summing the galactic and extragalactic gamma-ray contribution to IGRB together, we show the total gamma-ray flux in the right panel of Fig. 6 and compare it with the high latitude Fermi-LAT measurement result taken from Ref. [49]. Our SRDM model is severely ruled out when using the "power-law" concentration parameter function, but is still within the constraint for the "Macciò" model.

C. CMB constraints

The next important constraint we consider comes from the spectrum of the CMB radiation. Dark matter annihilations just after the time of recombination can dump energy into the photon bath, leading to distortions in the CMB spectrum. At the relevant times for this process, the typical dark matter velocity is very small. The annihilation cross section is thus close to its zero velocity value.

The PLANCK collaboration has presented general bounds on annihilation of dark matter into electromagnetic particles, i.e. electrons and photons [15]. They are sensitive to the rate of energy deposition per unit volume, which can generally be written as

$$\frac{dE}{dVdt}(z) = 2\tilde{g} \rho_{\text{DM}}^2(z) \frac{f(z) \langle \sigma v_{\text{rel}} \rangle}{m_\chi}, \quad (60)$$

where $\tilde{g}$ is a degeneracy factor of $1/2$ for Majorana or real scalar dark matter and $1/4$ for Dirac or complex scalar dark matter; ρ_{DM} is the dark matter energy density at a given redshift z; $f(z)$ is an efficiency factor for absorption of electromagnetic energy that we discuss in greater detail below; and m_χ is the dark matter mass. We discuss each of these factors in greater detail as they pertain to the model above.

The degeneracy factor $\tilde{g}$ is taken to be $1/2$ for the determination of constraints by the PLANCK collaboration. This is consistent with the model we work with here, which effectively behaves like Majorana or real scalar dark matter, with only XX annihilation. Particles annihilate with identical particles, rather than antiparticles. The dark matter density as a function of z is an observable that is fixed independently. The factor of $f(z)$ is the efficiency for produced electrons and photons to dump energy into the matter-radiation bath. At the narrow relevant range of $z = 600$–1000, it has been shown [13,51,52] that f is nearly independent of z, such that a constraint can be determined in terms of a constant f_{eff}. To determine f_{eff} for the model considered here, we follow Ref. [14]. In particular, this factor can be determined from Eq. (2) therein, which we modify to the model above as

$$f_{\text{eff}}(M_X) = \frac{\int_0^{3M_X/4} dE E 2 f_{\text{eff}}^{e^+e^-}(E) dN/dE}{2M_X}, \quad (61)$$

where $f_{\text{eff}}^{e^+e^-}(E)$ is the efficiency for absorption of energy from electrons and positrons and dN/dE is the spectrum of positrons produced in dark matter annihilation given in Eq. (41). The spectrum $f_{\text{eff}}^{e^+e^-}(E)$ is taken from the calculation in Ref. [14]. Note once more that the maximum energy that an electron or positron produced in annihilation is $3M_X/4$, unlike in models where two light scalars are produced in annihilation.

The remaining factors in Eq. (60) are trivial. $\langle \sigma v_{\text{rel}} \rangle$ should just be taken to be $\langle (\sigma v_{\text{rel}})_{\text{eff}} \rangle$ in our model, while
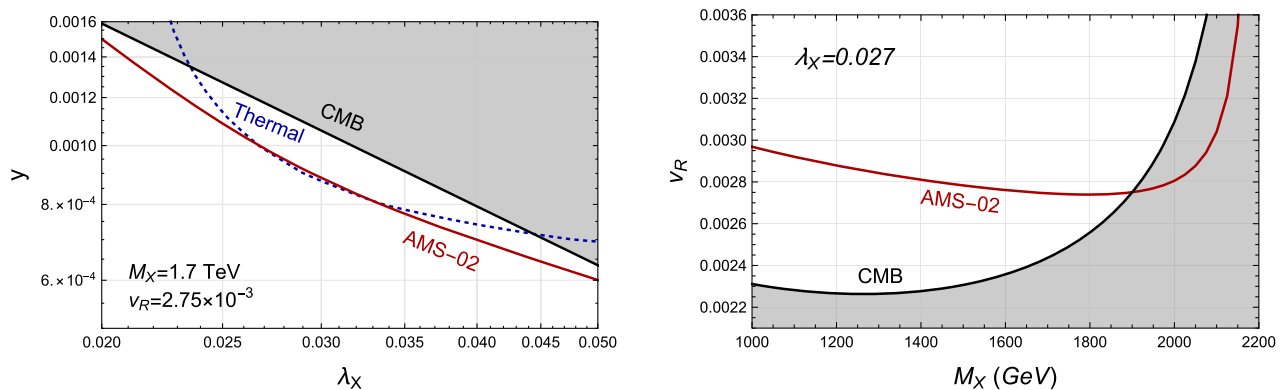
FIG. 7. Left panel: The parameter space in λ_X and y to fit the dark matter thermal abundance and the AMS-02 signal with $\langle \sigma v_{\text{rel}} \rangle^{\text{AMS}}_{\text{eff}} = 5 \times 10^{-23}$ cm³/s $\times (M_X/1.7 \text{ TeV})^2$. The constraint from CMB is approximately taken as $\langle \sigma v_{\text{rel}} \rangle^{\text{CMB}}_{\text{eff}} < 1.7 \times 10^{-23}$ cm³/s $\times (M_X/1.7 \text{ TeV})$. Right panel: after satisfying the thermal relic abundance with the approximation Eq. (39), the allowed parameter space to fit the AMS-02 rate and satisfy the CMB constraints.

$m_\chi = M_X$. The resulting bound from combining the PLANCK collaboration constraint [15] with Eq. (61) can be written as

$$\langle (\sigma v_{\text{rel}}) \rangle^{\text{CMB}}_{\text{eff}} < 1.7 \times 10^{-23} \text{ cm}^3/\text{s} \times \frac{M_X}{1.7 \text{ TeV}}. \quad (62)$$

The constraints above can also be approximately obtained by reading constraints on $VV \to 4\mu$ models with V as a light vector boson or exactly obtained by reading constraints on $\phi\phi \to 4\mu$ models. All that is required is some rescaling factors that we derive below. Before beginning, note that the spectrum of electrons in $VV \to 4\mu$ is not very different from those in $\phi\phi \to 4\mu$, leading to similar constraints and predictions for the two models. The rescaling that we derive here is physically due to two important, but somewhat superficial differences between the SRDM model and the $\phi\phi \to 4\mu$ model. The first is that the semiannihilation process in the SRDM model produces only a single scalar that decays to muons, leading to a trivial relative factor of 1/2. The second difference has less trivial repercussions: the momentum of the scalar produced in annihilation is $3M_{\text{DM}}/4$ in the SRDM as opposed to M_{DM} in the $\phi\phi \to 4\mu$ model. In order to read off constraints on SRDM model using constraints on $\phi\phi \to 4\mu$ or $VV \to 4\mu$, one needs to rescale the constrained mass $M^{\text{eff}}_{\text{DM}}$ and $\langle (\sigma v_{\text{rel}}) \rangle_{\text{eff},4\mu}$ to the physical parameters of the SRDM model M_X and $\langle (\sigma v_{\text{rel}}) \rangle_{\text{eff}}$. We derive this rescaling below.

Since the scalar decaying to muons is produced with fixed momentum in both models, the energy spectrum of positrons (and electrons) is related by a constant factor, so that we have the mapping, $M^{\text{eff}}_{\text{DM}} = 3M_X/4$. Since the number density of dark matter is given by $\rho_{\text{DM}}/M_{\text{DM}}$ and $M_{\text{DM}} = M_X = 4M^{\text{eff}}_{\text{DM}}/3$, the annihilation rate for SRDM dark matter is suppressed by a factor of $(3/4)^2 = 9/16$, in addition to the factor of 1/2 described above. Note that this factor applies to CMB constraints as well, since one explicit factor of 3/4 is obtained from the

factor of m_χ in the denominator of Eq. (60) and second factor of 3/4 enters in from writing f_{eff} as

$$f_{\text{eff}} = \frac{3}{4} \frac{\int_0^1 dx\, x f^{e^+ e^-}_{\text{eff}} (x M^{\text{eff}}_{\text{DM}}) dN/dx}{2}, \quad (63)$$

where we change variables to $x = 4E/3M_X = E/M^{\text{eff}}_{\text{DM}}$. Combining these two factors, we obtain the relation

$$\langle (\sigma v_{\text{rel}}) \rangle_{\text{eff},4\mu} = \frac{9}{32} \langle (\sigma v_{\text{rel}}) \rangle_{\text{eff}}. \quad (64)$$

This relation allows us to read off any constraint on the annihilation rate of $\phi\phi \to 4\mu$ directly.

Combining the required parameters to achieve the observed thermal relic abundance, the fit to the positron fraction data and the constraints from annihilations in the recombination epoch, we arrive at results shown in Fig. 7. To explain the AMS-02 preferred annihilation rate, the dark matter mass has to be below around 1.9 TeV. This is another interesting general feature of the resonant dark matter model, in which an upper bound on dark matter mass exists.

VI. ADDITIONAL SIGNALS

The main signal of the SRDM model is annihilation that produces muons and thereby cosmic ray positrons. In this section, we discuss other possible mechanisms for discovering dark matter. As usual, there are three possible approaches for confirming the SRDM model: direct detection, indirect detection, and collider production. We discuss these in turn.

Direct detection of the SRDM model is challenging. The light Goldstone modes do not couple directly to the dark matter state without sufficient energy to excite the heavy X^R states and can only mediate interactions with the proton at one loop, while the heavy Φ modes lead to a large mass suppression of their interaction cross section. The dominated heavy Φ mediated tree-level interaction, on the other hand, has an estimated cross section of
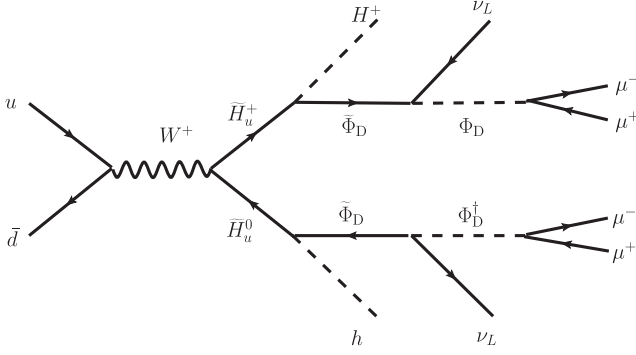
FIG. 8. A representative Feynman diagram for the collider signatures with Higgs fields and collimated (displaced) "dimuon jet" in the final state.

$$\sigma_{\Phi_T, \Phi_8} \sim \frac{1}{16\pi M_X^2} \epsilon_2^2 y^2 \frac{M_p^4}{M_h^4} \sim 10^{-58} \text{ cm}^2$$
$$\times \left(\frac{1.7 \text{ TeV}}{M_X}\right)^2 \left(\frac{\epsilon_2}{10^{-4}}\right)^2 \left(\frac{y}{10^{-3}}\right)^2, \quad (65)$$

which is too small to be accessed in the near future.

Indirect detection of positrons and gamma rays from the SRDM model has been discussed above, but as seen in Fig. 2, another SM product of dark matter annihilation is neutrinos. The neutrinos are emitted with an energy $E \sim \delta m \sim 1$ MeV. Such neutrinos are relatively low energy and are difficult to detect without an enormous flux, such as that from the Sun. To detect such neutrinos, detectors use either inverse beta decay or scattering off electrons. In either case, the cross section is determined by the four Fermi weak interaction and is thus highly suppressed. In addition, the flux is expected to be of order

$$\Phi_\nu \sim J\rho_\odot^2 r_\odot \frac{\langle \sigma v \rangle_{\text{eff}}}{M_X^2} \sim 10^{-11} \text{ cm}^{-2} \text{ s}^{-1} \times J, \quad (66)$$
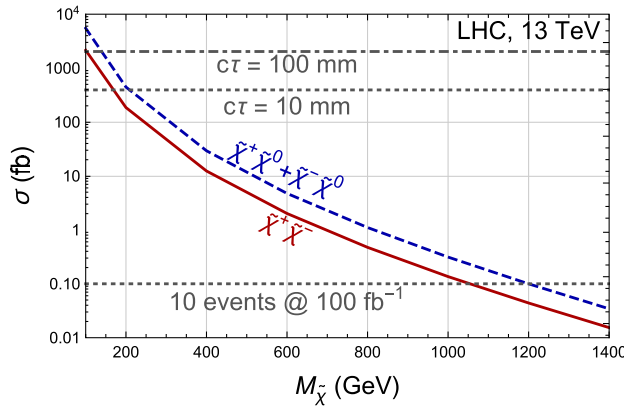
where J is a dimensionless line-of-sight integral that is of order 100–1000 for the dominant contribution from the Galactic center. For comparison, the flux of solar neutrinos near $E = 1$ MeV is few $\times 10^8$ cm^{-2} s^{-1} [53]. The predicted MeV neutrino flux in our model is far too small to be detected with current detectors.

The most promising additional means of detecting the SRDM model is by producing states at colliders. Our model is by necessity added onto a supersymmetrized SM. Since the Goldstone multiplet states are rather light, with masses $\lesssim 1$ GeV, the lightest SM superpartner states can decay to them if other R-parity violating decay modes are subdominant. If this decay is sufficiently fast, then the SRDM leads to a striking prediction of several muon jets arising from the collimated muons in the Goldstone scalar decays that end the decay chain [54–57]. The mediation to the dark sector can be achieved from production of Higgsino, charginos or sneutrinos. We therefore consider the minimal scenario in which either a Higgsino or a sneutrino is the lightest SM partner. Any spectrum in which there are light squarks or gluinos will of course be easier to detect. In Fig. 8, we show a representative Feynman diagram for the signature at the LHC. Higgsino states are dominantly pair produced via s-channel γ^*, Z, and W. The chargino states can be produced as well, assuming they are nearly degenerate with the neutralino states. The dominant production modes are $\tilde{\chi}^{\pm}\tilde{\chi}^0$ and $\tilde{\chi}^+\tilde{\chi}^-$, as the $\tilde{\chi}^0\tilde{\chi}^0$ only couples via the Z and accidental cancellation suppresses this channel by a factor of a few. Depending on the cutoff of the higher-dimensional operator $\Phi_D\bar{\Phi}_D H_u L/\Lambda_\Phi$ operator in Eq. (9) and the coefficient of the SUSY Higgs-portal operator $\epsilon_2 \Phi_D H_u H_d$, the neutralino could decay into the SM Higgs plus $\Phi_D\nu$ and the chargino could decay into charged Higgs plus $\Phi_D\nu$, or they can also decay to multiple PNGBs as $\Phi_D\Phi_D\nu$ and $\Phi_D\Phi_D\ell$. For the first possibility with the dominant decay for Higgsino as $h\tilde{\Phi}_D$ (ignoring the soft neutrino), one has two collimated
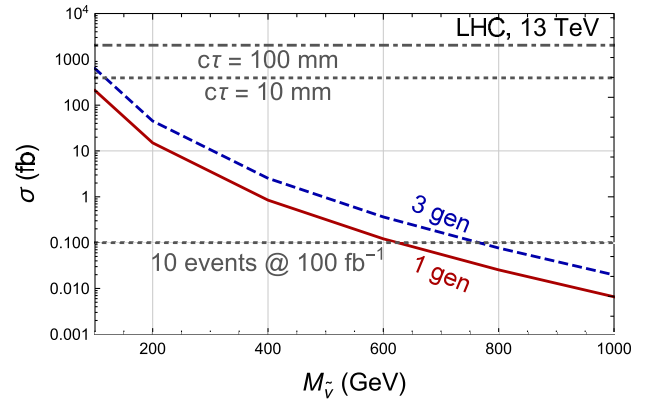

FIG. 9. Left panel: Production cross section for the dominant Higgsino states when the Higgsino is the lightest SM partner. We show the channels $\tilde{\chi}^+\tilde{\chi}^-$ (solid red) and $\tilde{\chi}^{\pm}\tilde{\chi}^0$ (dashed blue). Right panel: Production cross section for sneutrinos when the sneutrino is the lightest SM partner. We present cases where there is only one light sneutrino state (solid red) and where there are three degenerate light sneutrino states (dashed blue). Current estimated constraints from an ATLAS search for displaced lepton tracks [58] are shown for lifetimes of 100 mm (dot dashed gray) and 10 mm (dotted gray).

muons along with a Higgs that together reconstruct the Higgsino mass. The production cross section for the dominant modes is shown in the left panel of Fig. 9. Because of the striking signature properties, we anticipate a very good coverage of the model parameter space at the LHC Run 2.

The other minimal scenario, with only light sneutrinos, has a smaller cross section both due to the fact that the sneutrinos don't carry charge and that they are scalars. The left-handed sleptons corresponding to the relatively light sneutrinos may also be light, but we neglect this possibility here. The dominant decay also occurs via the operator $\Phi_D \bar{\Phi}_D H_u L$ in Eq. (9). The decay products are $\Phi_D \tilde{\Phi}_D$. The production cross section is shown in the right panel of Fig. 9.

In either of the above cases, the final state may contain four Φ_D PNGBs, each of which decays to a collimated pair of muons since the Φ_D's are produced at highly relativistic energies. Therefore, the signal of these models is multiple muon jets, along with other possible SM particles. The upper bound on Λ_Φ in the SRDM model is very mild and arises by requiring that the heavier states in the Φ_D supermultiplet decay sufficiently promptly on a cosmological. Their decays to Φ_D therefore may or may not be prompt in colliders, depending on the value of Λ_Φ. The Φ_D, on the other hand, has a lifetime of order 10^{-11}s, as seen in Eq. (12). The decays of Φ_D are therefore generally displaced by $\mathcal{O}(3\ \text{mm})$ at rest and $\mathcal{O}(30\ \text{cm})$ with a Lorentz boost of $\mathcal{O}(100)$.

CMS and ATLAS have searched for long-lived scalars having displaced decays to muons at 7 [59] and 13 TeV [58], respectively. The ATLAS search has greater sensitivity to the SRDM model at the moment. Since there is some flexibility in the lifetime of the Φ_D states, we indicate two benchmark lifetimes of $c\tau = 10$ mm and $c\tau = 100$ mm in Fig. 9. These bounds should be interpreted as rough, since the models considered in Ref. [58] do not map exactly onto the model we consider. We focus on the FRVZ [57] $4\gamma_d$ model with $m_H = 800$ GeV and $m_{\gamma_d} = 400$ MeV since this is the model that most closely matches the interesting portion of parameter space of SRDM model. We note that the LHC is just beginning to have sufficient data to probe these processes in the most difficult channels. For our SRDM model, because of the additional energetic SM particles in the final state, both the trigger choices and reduction of backgrounds in our model are much easier than the FRVZ model. A more careful collider study is needed to know the final reach of neutralino or sneutrino masses at the LHC Run 2, which we leave to future study.

VII. DISCUSSION AND CONCLUSIONS

There are some effects we have neglected up to this point in our calculations that should have a minimal effect. We have thus far neglected the $SU(3)_f$-breaking effects in the superpotential that will feed back at loop-level into the Kähler potential, which could cause small distortions of

the delicate factor of in the X mass spectrum required for the resonant enhancement of the semiannihilation cross section. The largest of these effects comes from the λ_X interaction in Eq. (2). The one-loop contribution to the wave-function renormalization of X in the Kähler potential is of order $\lambda_X^2/(16\pi^2) \sim 10^{-6}$, leading to a contribution to $\delta m/M_X \sim 10^{-6}$, comparable to the benchmark discussed above. We note this additional contribution, but it is not an issue for the phenomenological analysis above. Other contributions to the mass splitting are suppressed smaller parameters, higher scales, or more loops and are negligible.

Another loop-level effect on the spectrum affects the SM particle properties. The first term in the decay superpotential of Eq. (9) can generate a Majorana neutrino mass. The contribution is of order

$$\Delta m_\nu \sim \frac{v_u^2 M_X a_{ii}^2}{16\pi^2 \Lambda_X^2} \frac{\delta m}{M_X} \sim 0.07\ \text{eV}$$
$$\times \left(\frac{10^5\ \text{GeV}}{\Lambda_X/a_{ii}}\right)^2 \left(\frac{\delta m/M_X}{10^{-6}}\right) \left(\frac{M_X}{1.7\ \text{TeV}}\right). \quad (67)$$

Note that the generation of Majorana neutrino masses necessarily requires SUSY breaking by holomorphy, so additional suppression factor of $\delta m/M_X$ is anticipated. A lower bound is set on Λ_X, since the largest possible neutrino mass is around 0.23 eV from cosmological constraints [15] and is around 0.31 eV for Majorana effective mass determined by neutrinoless double beta decay ($0\nu\beta\beta$) [60]. Since the mass generated is in fact a Majorana mass, the constraint from $0\nu\beta\beta$ should be taken into account and we can set a constraint of $\Lambda_X/a_{ii} \gtrsim 5 \times 10^4$ GeV. We also note that the dark flavor-changing operator with an coefficient of a_{12} will not generate neutrino Majorana mass by itself. This is because the lepton-number is conserved if other coefficients a_{11}, and a_{22} are tiny.

One final effect to consider is the feedback of SM SUSY breaking into the dark sector. The dominant mediation of SUSY breaking occurs via the SUSY Higgs portal parameter $\epsilon_2 \lesssim 10^{-4}$. All such effects on the spectrum are highly suppressed as they require a double insertion of operators coupling to the SM. Loop effects from SUSY breaking could also in principle destabilize the desired vacuum alignment which is dominantly along the Φ^8 direction. Such effects should only led to small corrections that could not induce a large change to a completely different vacuum structure.

In addition to these higher-order effects on the spectrum, we have made a few assumptions in our solution to the Boltzmann equations. The semiannihilation process generates a quasirelativistic X or $\overline{X}$ in the final state. We have made the assumption that there is an interaction that allows these relativistic dark matter particles to rapidly thermalize with the nonrelativistic dark matter population. This interaction also erases any distortions caused by the nontrivial dark matter momentum dependence of the semiannihilation

cross section on the nonrelativistic dark matter population. Independent of the interpretation of the AMS-02 positron excess, the velocity-dependent semiannihilation scenario introduced in our model may have other interesting implications on the large scale structure with a mixture of cold and warm dark matter components.

Before concluding, we note that the precise predictions for the spectrum and flux of cosmic ray positrons is quite sensitive to the details of the electron and positron propagation in the galaxy, the size of the standard background of positrons and the velocity distribution of dark matter in the galactic halo. We have shown a few examples of how the spectrum could change due to the first of these. One particular effect is on the upper bound on the allowed dark matter mass to accommodate all the data fit in this work. This uncertainty could leave a larger window of viability for the SRDM model.

In summary, we have developed and studied a SRDM model where annihilations producing positrons are resonantly enhanced in the Milky Way. The model evades strong constraints from precision CMB measurements, while explaining the observed dark matter relic abundance and high energy cosmic ray positron excess. The resonance must have a mass very close to twice the dark matter mass, which is achieved by spontaneously breaking an $SU(3)_f$ flavor symmetry to $SU(2)_f \times U(1)_f$ under which dark matter is charged. The factor of two relation in the VEV-induced mass for a triplet of $SU(3)_f$ provides a novel mechanism for naturally explaining resonant dark matter. The PNGB of spontaneous flavor symmetry breaking decay dominantly to muons, leading to an implementation of a class of models where dark matter annihilates into particles that only later decay to the SM. Such models are known to be far safer from gamma ray constraints at the large cross sections required to explain the positron excess. Since the model structure requires supersymmetry for stability, weak scale superpartners of the SM lead to distinctive signals for colliders, such as decay chains with an SM Higgs and collimated muons. We have shown that searches for displaced lepton jets could have sensitivity to such models, while searches more optimized for the particular structure of this model could increase sensitivity at the LHC.

APPENDIX A: GENERAL PARAMETRIZATION

Starting from the generalized parametrization, we have the effective annihilation rate as

$$(\sigma v_{\rm rel})_{\rm eff} = \sigma_0 \frac{\delta^2 + \gamma^2}{(\delta - v_{\rm rel}^2/4)^2 + \gamma^2}. \tag{A1}$$

Here, the parameter γ is treated to be independent of $v_{\rm rel}$. If only the complex scalar X_p^1 is the dark matter, the degrees of freedom is $g = 2$. To suppress the prediction for the CMB while explain the rate for the AMS-02, we need to work in the limit of $\delta \gg \gamma$.

To satisfy the relic abundance, $\Omega h^2 = 0.112$, the following simple relation is needed:

$$\frac{\Omega h^2}{0.112} = 7.4 \times 10^{-11} \frac{\gamma}{\sigma_0 \delta^2} \Rightarrow \frac{\gamma}{\sigma_0 \delta^2} = 1.35 \times 10^{10} \, {\rm GeV}^2. \tag{A2}$$

For a narrow resonance and if the pole can be reached for the integration, the annihilation rate is

$$(\sigma v_{\rm rel})_{\rm eff} = \frac{\pi \sigma_0 \delta^{3/2}}{\gamma} \delta(v_{\rm rel} - 2\sqrt{\delta}). \tag{A3}$$

So, the prediction for AMS-02 is

$$\langle \sigma v_{\rm rel} \rangle_{\rm eff}^{\rm AMS} = \frac{4\pi \sigma_0 \delta^{5/2}}{\gamma} \frac{4\pi}{(2\pi v_0^2)^{3/2}} e^{-2\delta/v_0^2}. \tag{A4}$$

For our case, we have the effective annihilation rate as

$$(\sigma v_{\rm rel})_{\rm eff} = \frac{3y^2 \lambda_X^2}{4\pi M_X^2 [(v_{\rm rel}^2 - v_{\rm R}^2)^2 + \frac{1}{64\pi^2}(3y^2 + \lambda_X^2 v_{\rm rel})^2]}. \tag{A5}$$

In the limit of $\gamma \ll \delta$, we have the following match relations

$$\sigma_0 \delta^2 = \frac{3y^2 \lambda_X^2}{64\pi M_X^2}, \quad \gamma = \frac{1}{32\pi}(3y^2 + \lambda_X^2 v_{\rm rel}), \quad \delta = v_{\rm R}^2/4. \tag{A6}$$

To simplify our discussion, we first treat the $v_{\rm rel}$ in γ as a constant with $v_{\rm rel} \approx v_{\rm R} = 2\sqrt{\delta}$. In general, a larger value of γ (allowed by CMB constraints) can make our model easier to fit. This is because of the simple math relation $2ab \leq a^2 + b^2$. Defining the parameter $\kappa = 1.35 \times 10^{10} \, {\rm GeV}^2$ from Eq. (A2), the condition to have real solutions for y and λ_X is

$$\delta < \frac{4\pi^2 \kappa^2}{M_X^4} \gamma^2. \tag{A7}$$

For $M_X = 1.7 \, {\rm TeV}$ and $\gamma = 5 \times 10^{-8}$, one needs $\delta < 2.15 \times 10^{-6}$. Let's choose $\gamma = 5 \times 10^{-8}$ and $\delta = 1.89 \times 10^{-6}$, the corresponding model parameters are $v_{\rm R} = 2.75 \times 10^{-3}$ and $(y, \lambda_X) = (7.9 \times 10^{-4}, 3.4 \times 10^{-2})$ or $(1.0 \times 10^{-3}, 2.6 \times 10^{-2})$. For this model point, we have $\langle \sigma v_{\rm rel} \rangle_{\rm eff}^{\rm AMS} = 5 \times 10^{-23} \, {\rm cm}^3/{\rm s}$ and $\langle \sigma v_{\rm rel} \rangle_{\rm eff}^{\rm CMB} = 1.2 \times 10^{-23} \, {\rm cm}^3/{\rm s}$. The predicted cross-sections in this parametrization for annihilation in the Milky Way and during the recombination era are shown in Fig. 10.
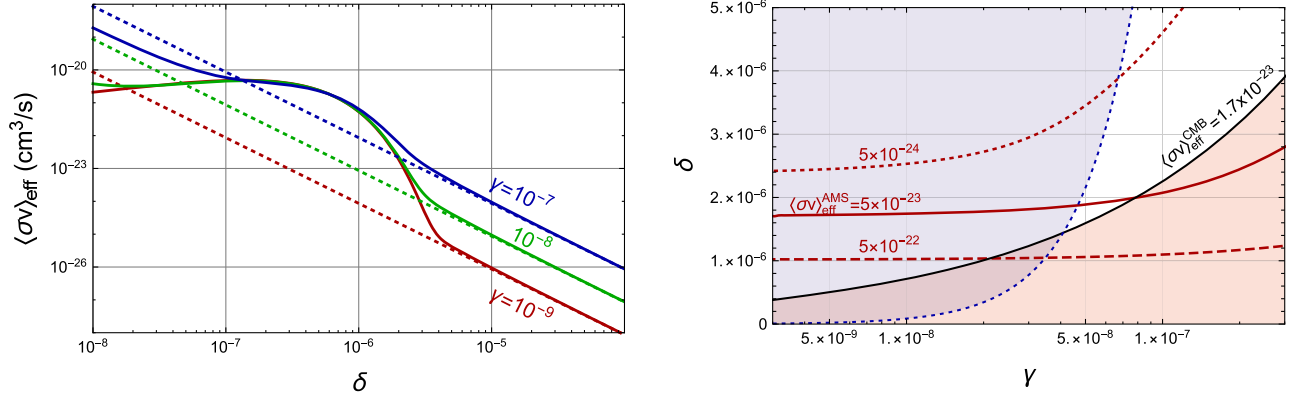
FIG. 10. Predicted cross sections (solid lines) for annihilation in the Milky Way (as would be seen by AMS-02) and during the recombination era (dotted lines), which lead to CMB constraints. The low velocity cross section is fixed by requiring that the model achieves the observed present day dark matter abundance, while M_X is fixed to 1.7 TeV in order to obtain a good fit to the AMS-02 data. The AMS-02 data fit above had a best fit cross section of $\langle \sigma v \rangle_{\rm eff}^{\rm AMS} = 5 \times 10^{-23}$ cm^3/s, while the constraint from annihilations during recombination on $M_X = 1.7$ TeV dark matter is $\langle \sigma v \rangle_{\rm eff}^{\rm CMB} < 1.7 \times 10^{-23}$ cm^3/s. The blue dotted line is the additional constraint on the SRDM model as in Eq. (A7).

Finally, we also note that there is an upper bound on the dark matter mass in our model. From Eq. (A4), we can rewrite the annihilation rate for AMS-02 as

$$\langle \sigma v_{\rm rel} \rangle_{\rm eff}^{\rm AMS} \approx \frac{\eta(\delta, v_0)}{\kappa}, \quad \text{with}$$

$$\eta(\delta, v_0) \equiv \delta^{1/2} \frac{16\pi^2}{(2\pi v_0^2)^{3/2}} e^{-2\delta/v_0^2}. \quad \text{(A8)}$$

In this general parametrization, the prediction for the annihilation rate in the CMB era is simply $\langle \sigma v_{\rm rel} \rangle_{\rm eff}^{\rm CMB} = \sigma_0 = \gamma/(\kappa \delta^2)$. Requiring an upper bound on the ratio of those two annihilation rates, we have

$$\frac{\langle \sigma v_{\rm rel} \rangle_{\rm eff}^{\rm CMB}}{\langle \sigma v_{\rm rel} \rangle_{\rm eff}^{\rm AMS}} \leq R^{\rm max} \times \left(\frac{1.7 \text{ TeV}}{M_X} \right)$$

$$\Rightarrow \gamma \leq R^{\rm max} \delta^2 \eta \times \left(\frac{1.7 \text{ TeV}}{M_X} \right), \quad \text{(A9)}$$

where numerically we have $R^{\rm max} < 0.34$ from Sec. V. Combining the two inequalities in Eqs. (A7), (A9), we arrive at the following upper bound on the dark matter mass

$$M_X \leq [2\pi \kappa R^{\rm max} \delta^{3/2} \eta(\delta, v_0)]^{1/3} \times (1.7 \text{ TeV})^{1/3} \approx 1.6 \text{ TeV},$$

$$\text{(A10)}$$

after we use $\delta = 1.89 \times 10^{-6}$ to fit the rate for AMS-02. There is another weaker bound from satisfying the narrow-width condition of $\gamma < \delta$ and Eq. (A7). We note that the above approximated upper bound on the dark matter mass becomes less precise if the narrow resonance condition $\gamma \ll \delta$ is not satisfied, in which case the actual bound should rely on numerical calculations.

APPENDIX B: DETAILS OF BOLTZMANN EQUATIONS AND KINETIC DECOUPLING

In order to study kinetic decoupling, additional moments of the full Boltzmann equation are required. It is conventional and convenient to parametrize the kinetic coupling by the variable y defined as

$$y = \frac{1}{s^{2/3} n} \int \frac{d^3 p}{(2\pi)^3} \mathbf{p}^2 f(p), \quad \text{(B1)}$$

where f is the phase space distribution of X_p^1. The variable y is chosen such it goes to a constant after kinetic decoupling and such that in kinetic equilibrium with the SM bath $y_{\rm EQ} = 3M_X^2 x / s^{2/3} (x = 1)$. The contribution to the evolution of y due to elastic scattering off of other species, namely neutrinos in the SRDM model, has been studied in Ref. [61], where it is found that

$$\frac{dy}{dx} \bigg|_{\rm elas} = -\frac{1}{Hx} 2M_X c(T)(y - y_{\rm EQ}), \quad \text{(B2)}$$

where $c(T)$ is given by

$$c(T) = \frac{1}{12(2\pi)^3 M_X^4 T} \int dk k^4 (e^{k/T} + 1)^{-1}$$

$$\times [1 - (e^{k/T} + 1)^{-1}] \sum_f |\widetilde{\mathcal{M}}|^2, \quad \text{(B3)}$$

for scattering off of massless fermions f. Here, k is the momentum of the relativistic fermion. We define the Mandelstam t averaged amplitude by

$$|\widetilde{\mathcal{M}}|^2 = \frac{1}{8k^4} \int_{-4k^2}^{0} |\mathcal{M}|^2 (-t) dt. \quad \text{(B4)}$$

For scattering off neutrinos in the SRDM model, there are two different relevant amplitudes. If the incoming state has a neutrino (antineutrino) and the final state has a neutrino (antineutrino), then the amplitude is given by

$$|\mathcal{M}|^2 = \frac{\lambda_\nu^4[(s - M_X^2)^2 + st]}{(s - M_{\tilde{X}^2}^2)^2 + \Gamma_{\tilde{X}^2}^2(s)s}. \tag{B5}$$

If the incoming state has a neutrino (antineutrino) and the final state has an antineutrino (neutrino), then the amplitude is given by

$$|\mathcal{M}|^2 = \frac{\lambda_\nu^4 M_{\tilde{X}^2}^2(-t)}{(s - M_{\tilde{X}^2}^2)^2 + \Gamma_{\tilde{X}^2}^2(s)s}. \tag{B6}$$

Here, we define $\lambda_\nu = a_{12}v_u/\sqrt{2}\Lambda_X$. After integrating over t as prescribed by Eq. (B4), summing over all four possible combinations of neutrinos, and using $s = M_X^2 + 2M_X k$, we find, to leading order in k/M_X and $\delta_2 = \delta m/M_X$,

$$\sum_f |\widetilde{\mathcal{M}}|^2 = \frac{2\lambda_\nu^4 k^2}{(k - \delta m)^2 + \Gamma_{\tilde{X}^2}^2/4}. \tag{B7}$$

This contribution tends to push the dark sector toward equilibrium with the relativistic fermion bath off of which it is scattering. There is an additional contribution in the model considered here due to the semiannihilation process. To leading order in $1/x$, the dominant contribution to the evolution of y is due to the large momentum of the final state X_1^p state, which is produced with a momentum

$$\mathbf{p}_{X,\text{out}}^2 = \frac{9M_X^2}{16} + \mathcal{O}(\mathbf{p}_{X,in}^2), \tag{B8}$$

where the additional contributions go like the very non-relativistic momentum of the incoming X particles, which is suppressed by $1/x$. By integrating the Boltzmann equation weighted by $\mathbf{p}^2$, the contribution to y of this injection is then given by

$$\left.\frac{dy}{dx}\right|_{\text{semiann}} = \frac{9M_X^2 s^{1/3}}{32Hx}\langle\sigma v_{\text{rel}}\rangle_{\text{eff},x_{\text{DM}}}Y. \tag{B9}$$

Our full solution for the abundance of dark matter is obtained by solving the coupled Eqs. (31), (B2), and (B9) under the assumption that dark matter is in kinetic equilibrium among itself at all temperatures. From this solution, we are able to derive the constraint $\Lambda_X/a_{12} \sim v_{\text{EW}}$, as well as to verify that the dark matter cools to an effectively low temperature by the time of recombination. The temperature of the dark sector as parametrized by x_{DM} is shown in
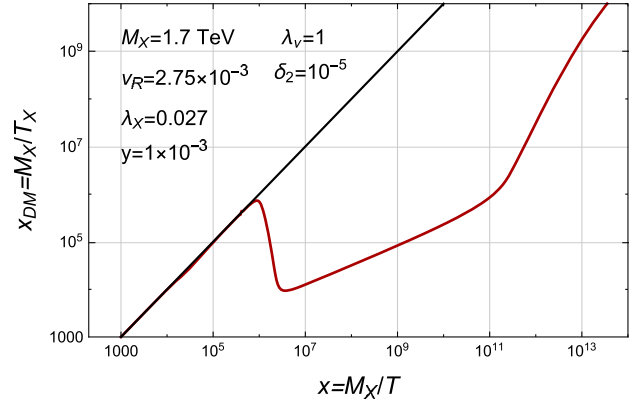

FIG. 11. The dark matter temperature parameter as a function of the SM temperature parameter.

Fig. 11. One can clearly see from Fig. 11 that the dark matter is hotter than the SM particles after chemical freeze-out era with $x \approx 10^6$. As a result, there is a delayed behavior for the resonant semiannihilation process to reach the physical pole at $x \sim 10^{11}$ before the recombination era. For $10^6 < x < 10^{11}$, x_{DM} follows a simple power-law behavior as $x_{\text{DM}} \propto x^{2/5}$. After that, the dark matter temperature continues to cool down. At the start of the recombination era with $z = 1000$ and $x \approx 6 \times 10^{12}$, the corresponding dark matter averaged speed is $v \leq 4 \times 10^{-5}$ and well below the resonance pole.

Maintaining kinetic equilibrium within the dark sector requires the introduction of a new mediator state that interacts with the dark matter. The interaction must be quite large, so the assumption of kinetic equilibrium within the dark matter sector could be difficult to achieve. Then, the dark matter produced in a small time window will begin with a large momentum very close to $3M_X/4$. We briefly verify that this does not pose a problem for our assumption that dark matter annihilates at effectively zero velocity around the time of recombination. The semiannihilation cross section for these states is small at the time they are produced, since they are far from the resonance. They thus redshift as essentially free nonrelativistic particles with kinetic energy scaling as $1/x^2$ so that the hottest dark matter states around the time of recombination have kinetic energy

$$E_{\text{kin}} < \frac{9M_X x_{\text{kd}}^2}{32x_{\text{rec}}^2} \sim 10^{-3} \text{ eV} \ll \delta m. \tag{B10}$$

Whether or not the dark matter remains in kinetic equilibrium among itself, the cross section at the time of recombination is well approximated by its zero velocity value.

[1] M. Aguilar *et al.* (AMS Collaboration), Precision Measurement of the $(e^+ + e^-)$ Flux in Primary Cosmic Rays from 0.5 GeV to 1 TeV with the Alpha Magnetic Spectrometer on the International Space Station, Phys. Rev. Lett. **113,** 221102 (2014).

[2] L. Accardo *et al.* (AMS Collaboration), High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station, Phys. Rev. Lett. **113,** 121101 (2014).

[3] S. Ting, The first five years of the alpha magnetic spectrometer on the international space station, *CERN Colloquium, Geneva,* (2016).

[4] J. J. Beatty *et al.*, New Measurement of the Cosmic-Ray Positron Fraction from 5 to 15-GeV, Phys. Rev. Lett. **93,** 241102 (2004).

[5] O. Adriani *et al.* (PAMELA Collaboration), Cosmic-Ray Positron Energy Spectrum Measured by PAMELA, Phys. Rev. Lett. **111,** 081102 (2013).

[6] M. Ackermann *et al.* (Fermi-LAT Collaboration), Measurement of Separate Cosmic-Ray Electron and Positron Spectra with the Fermi Large Area Telescope, Phys. Rev. Lett. **108,** 011103 (2012).

[7] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, A theory of dark matter, Phys. Rev. D **79,** 015014 (2009).

[8] M. Pospelov and A. Ritz, Astrophysical signatures of secluded dark matter, Phys. Lett. B **671,** 391 (2009).

[9] M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, Dark matter's secret liaisons: phenomenology of a dark U(1) sector with bound states, J. Cosmol. Astropart. Phys. 05 (2017) 036.

[10] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg, and P. Walia, Strong Constraints on Self-Interacting Dark Matter with Light Mediators, Phys. Rev. Lett. **118,** 141802 (2017).

[11] X.-L. Chen and M. Kamionkowski, Particle decays during the cosmic dark ages, Phys. Rev. D **70,** 043502 (2004).

[12] N. Padmanabhan and D. P. Finkbeiner, Detecting dark matter annihilation with CMB polarization: Signatures and experimental prospects, Phys. Rev. D **72,** 023508 (2005).

[13] D. P. Finkbeiner, S. Galli, T. Lin, and T. R. Slatyer, Searching for dark matter in the CMB: A compact parametrization of energy injection from new physics, Phys. Rev. D **85,** 043522 (2012).

[14] T. R. Slatyer, Indirect dark matter signatures in the cosmic dark ages. I. Generalizing the bound on s-wave dark matter annihilation from Planck results, Phys. Rev. D **93,** 023527 (2016).

[15] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, Astron. Astrophys. **594,** A13 (2016).

[16] J. Buch, P. Ralegankar, and V. Rentala, Late decaying 2-component dark matter scenario as an explanation of the AMS-02 positron excess, J. Cosmol. Astropart. Phys. 10 (2017) 028.

[17] M. Ibe, H. Murayama, and T. T. Yanagida, Breit-Wigner Enhancement of Dark Matter Annihilation, Phys. Rev. D **79,** 095009 (2009).

[18] W.-L. Guo and Y.-L. Wu, Enhancement of dark matter annihilation via Breit-Wigner resonance, Phys. Rev. D **79,** 055012 (2009).

[19] M. Ibe, Y. Nakayama, H. Murayama, and T. T. Yanagida, Nambu-Goldstone dark matter and cosmic ray electron and positron excess, J. High Energy Phys. 04 (2009) 087.

[20] X.-J. Bi, X.-G. He, and Q. Yuan, Parameters in a class of leptophilic models from PAMELA, ATIC and FERMI, Phys. Lett. B **678,** 168 (2009).

[21] X.-J. Bi, P.-F. Yin, and Q. Yuan, Breit-Wigner enhancement considering the dark matter kinetic decoupling, Phys. Rev. D **85,** 043526 (2012).

[22] H. An and M. Pospelov, Dark resonance, J. Cosmol. Astropart. Phys. 11 (2012) 061.

[23] M. Duch and B. Grzadkowski, Resonance enhancement of dark matter interactions: The case for early kinetic decoupling and velocity dependent resonance width, J. High Energy Phys. 09 (2017) 159.

[24] M. Boudaud *et al.*, A new look at the cosmic ray positron fraction, Astron. Astrophys. **575,** A67 (2015).

[25] A. Lopez, C. Savage, D. Spolyar, and D. Q. Adams, Fermi/LAT observations of Dwarf Galaxies highly constrain a Dark Matter Interpretation of Excess Positrons seen in AMS-02, HEAT, and PAMELA, J. Cosmol. Astropart. Phys. 03 (2016) 033.

[26] A. Scaffidi, K. Freese, J. Li, C. Savage, M. White, and A. G. Williams, Gamma rays from muons from WIMPs: Implementation of radiative muon decays for dark matter analyses, Phys. Rev. D **93,** 115024 (2016).

[27] M. Ackermann *et al.* (Fermi-LAT Collaboration), Dark matter constraints from observations of 25 Milky Way satellite galaxies with the Fermi Large Area Telescope, Phys. Rev. D **89,** 042001 (2014).

[28] T. Appelquist, H.-C. Cheng, and B. A. Dobrescu, Bounds on universal extra dimensions, Phys. Rev. D **64,** 035002 (2001).

[29] H.-C. Cheng, K. T. Matchev, and M. Schmaltz, Radiative corrections to Kaluza-Klein masses, Phys. Rev. D **66,** 036005 (2002).

[30] L.-F. Li, Group theory of the spontaneously broken gauge symmetries, Phys. Rev. D **9,** 1723 (1974).

[31] C. Patrignani *et al.* (Particle Data Group Collaboration), Review of particle physics, Chin. Phys. C **40,** 100001 (2016).

[32] P. Törmä and K. Sengstock, *Quantum Gas Experiments: Exploring Many-Body States*, Cold Atoms (World Scientific, Singapore, 2014).

[33] M. Vogelsberger, A. Helmi, V. Springel, S. D. M. White, J. Wang, C. S. Frenk, A. Jenkins, A. D. Ludlow, and J. F. Navarro, Phase-space structure in the local dark matter distribution and its signature in direct detection experiments, Mon. Not. R. Astron. Soc. **395,** 797 (2009).

[34] T. Binder, T. Bringmann, M. Gustafsson, and A. Hryczuk, Early kinetic decoupling of dark matter: When the standard way of calculating the thermal relic density fails, Phys. Rev. D **96,** 115010 (2017).

[35] K. Griest and D. Seckel, Three exceptions in the calculation of relic abundances, Phys. Rev. D **43,** 3191 (1991).

[36] F. D'Eramo and J. Thaler, Semi-annihilation of dark matter, J. High Energy Phys. 06 (2010) 109,

[37] G. Belanger, K. Kannike, A. Pukhov, and M. Raidal, Impact of semi-annihilations on dark matter phenomenology - an example of Z_N symmetric scalar dark matter, J. Cosmol. Astropart. Phys. 04 (2012) 010.

[38] T. Bringmann, Particle models and the small-scale structure of dark matter, New J. Phys. **11,** 105027 (2009).

[39] L. G. van den Aarssen, T. Bringmann, and Y. C. Goedecke, Thermal decoupling and the smallest subhalo mass in dark matter models with Sommerfeld-enhanced annihilation rates, Phys. Rev. D **85,** 123512 (2012).

[40] E. Aliu *et al.* (VERITAS Collaboration), VERITAS Deep Observations of the Dwarf Spheroidal Galaxy Segue 1, Phys. Rev. D **85,** 062001 (2012); Erratum, Phys. Rev. D**91,** 129903 (2015).

[41] J. Aleksic *et al.*, Optimized dark matter searches in deep observations of Segue 1 with MAGIC, J. Cosmol. Astropart. Phys. 02 (2014) 008.

[42] A. Abramowski *et al.* (H.E.S.S. Collaboration), Search for dark matter annihilation signatures in H.E.S.S. observations of Dwarf Spheroidal Galaxies, Phys. Rev. D **90,** 112012 (2014).

[43] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, CMB constraints on WIMP annihilation: Energy absorption during the recombination epoch, Phys. Rev. D **80,** 043526 (2009).

[44] G. Elor, N. L. Rodd, T. R. Slatyer, and W. Xue, Model-independent indirect detection constraints on hidden sector dark matter, J. Cosmol. Astropart. Phys. 06 (2016) 024.

[45] J. Buch, M. Cirelli, G. Giesen, and M. Taoso, PPPC 4 DM secondary: A poor particle physicist cookbook for secondary radiation from dark matter, J. Cosmol. Astropart. Phys. 09 (2015) 037.

[46] M. Cirelli, G. Corcella, A. Hektor, G. Hutsi, M. Kadastik, P. Panci, M. Raidal, F. Sala, and A. Strumia, PPPC 4 DM ID: A poor particle physicist cookbook for dark matter indirect detection, J. Cosmol. Astropart. Phys. 03 (2011) 051; Erratum, J. Cosmol. Astropart. Phys.10 (2012) 01.

[47] E. A. Baltz and J. Edsjo, Positron propagation and fluxes from neutralino annihilation in the halo, Phys. Rev. D **59,** 023511 (1998).

[48] M. Cirelli, R. Franceschini, and A. Strumia, Minimal Dark Matter predictions for galactic positrons, anti-protons, photons, Nucl. Phys. **B800,** 204 (2008).

[49] M. Ackermann *et al.* (Fermi-LAT Collaboration), The spectrum of isotropic diffuse gamma-ray emission between 100 MeV and 820 GeV, Astrophys. J **799,** 86 (2015).

[50] A. V. Maccio', A. A. Dutton, and F. C. v. d. Bosch, Concentration, spin and shape of dark matter haloes as a function of the cosmological model: WMAP1, WMAP3 and WMAP5 results, Mon. Not. R. Astron. Soc. **391,** 1940 (2008).

[51] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, Updated CMB constraints on dark matter annihilation cross sections, Phys. Rev. D **84,** 027302 (2011).

[52] G. Giesen, J. Lesgourgues, B. Audren, and Y. Ali-Haimoud, CMB photons shedding light on dark matter, J. Cosmol. Astropart. Phys. 12 (2012) 008.

[53] J. N. Bahcall, A. M. Serenelli, and S. Basu, New solar opacities, abundances, helioseismology, and neutrino fluxes, Astrophys. J. **621,** L85 (2005).

[54] N. Arkani-Hamed and N. Weiner, LHC signals for a superunified theory of dark matter, J. High Energy Phys. 12 (2008) 104.

[55] M. Baumgart, C. Cheung, J. T. Ruderman, L.-T. Wang, and I. Yavin, Non-Abelian dark sectors and their collider signatures, J. High Energy Phys. 04 (2009) 014.

[56] Y. Bai and Z. Han, Measuring the Dark Force at the LHC, Phys. Rev. Lett. **103,** 051801 (2009).

[57] A. Falkowski, J. T. Ruderman, T. Volansky, and J. Zupan, Discovering Higgs Decays to Lepton Jets at Hadron Colliders, Phys. Rev. Lett. **105,** 241801 (2010).

[58] ATLAS Collaboration, Search for long-lived neutral particles decaying into displaced lepton jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Report No. ATLAS-CONF-2016-042, 2016.

[59] V. Khachatryan *et al.* (CMS Collaboration), Search for long-lived particles that decay into final states containing two electrons or two muons in proton-proton collisions at $\sqrt{s} = 8$ TeV, Phys. Rev. D **91,** 052012 (2015).

[60] P. Guzowski, L. Barnes, J. Evans, G. Karagiorgi, N. McCabe, and S. Soldner-Rembold, Combined limit on the neutrino mass from neutrinoless double-β decay and constraints on sterile Majorana neutrinos, Phys. Rev. D **92,** 012002 (2015).

[61] T. Bringmann and S. Hofmann, Thermal decoupling of WIMPs from first principles, J. Cosmol. Astropart. Phys. 04 (2007) 016; Erratum, J. Cosmol. Astropart. Phys.03 (2016) 02.