Galactic positron excess from selectively enhanced dark matter annihilation

Anirban Das,[*] Basudeb Dasgupta,[†] and Anupam Ray[‡]

Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India

Precision measurements of the positron flux in the cosmic ray have revealed an unexplained bump in the spectrum around $E \simeq 300$ GeV, not clearly attributable to known astrophysical processes. We propose annihilation of dark matter of mass $m_\chi = 780$ GeV with a late-time cross section $\sigma v = 4.63 \times 10^{-24}$ cm^3 s^{-1} as a possible source. The nonmonotonic dependence of the annihilation rate on dark matter velocity, owing to a selective p-wave Sommerfeld enhancement, allows such a large signal from the Milky Way without violating corresponding constraints from cosmic microwave background and dwarf galaxy observations. We briefly explore other signatures of this scenario, and outline avenues to test it in future experiments.

I. INTRODUCTION

Observations of the positron flux in cosmic rays have seen tremendous progress and excitement in recent years. PAMELA [1], Fermi-LAT [2], MASS [3], Wizard/CAPRICE [4], AMS-01 [5], and HEAT [6] have reported the positron flux in the cosmic ray, with PAMELA and Fermi-LAT providing evidence for a rising positron flux around 100 GeV. The AMS-02 collaboration, in its latest published results with three times more statistics than previous measurements, has presented the positron flux up to energy 1 TeV, and the existence of a bump in the flux spectrum at $E \simeq 300$ GeV followed by a cutoff at $E \simeq 810$ GeV has been confirmed [7].

The positron flux spectrum shows several distinct features. The flattening at $E \simeq 8$ GeV seen in Fig. 1 is understood to be mainly caused by the diffusion of the positrons produced from the decay of the scattering products of the cosmic ray with the interstellar gas [8,9], as described in Ref. [7]. However, the shape of the spectrum in the region $E > 100$ GeV, a rise followed by a sharp drop, is not yet fully understood.

Dark Matter (DM) annihilation/decay and particle acceleration by pulsars have been shown to be able to reproduce the positron spectrum, but several concerns remain [10–16]. DM of mass (1.5–3) TeV and annihilation cross

section $(6–23) \times 10^{-24}$ cm^3 s^{-1} [11,12], such that it first annihilates into light intermediate states that subsequently decay to charged leptons, can produce a good fit to the earlier AMS data [17]. However, this large annihilation cross section into leptons is disallowed by the Fermi-LAT observations of the dwarf galaxies [18], which show no excess gamma-ray flux. Such a large DM annihilation cross
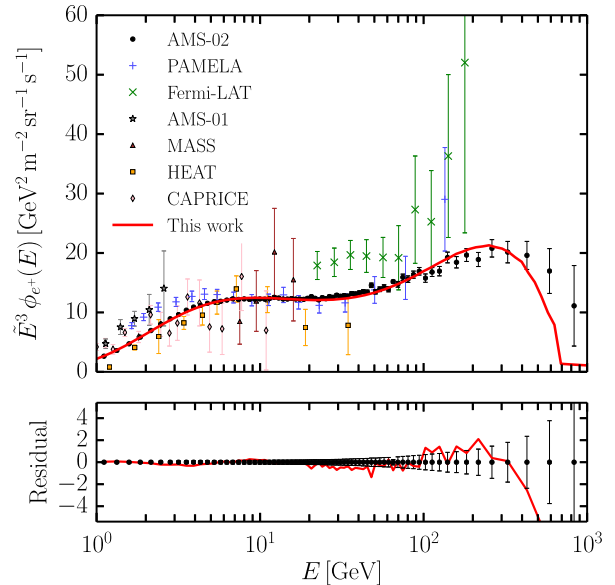

FIG. 1. AMS-02 positron flux measurements (black dots), previous data (other symbols), and our theoretical prediction (red solid curve) due to Sommerfeld-enhanced p-wave DM annihilation rate $(\sigma v)_p = 4.63 \times 10^{-24}$ cm^3 s^{-1} and DM mass $m_\chi = 780$ GeV. The red curve includes positrons from both the astrophysical background and DM annihilation. The theoretical residuals over the AMS-02 data are shown in the lower panel. The goodness of fit χ^2/d.o.f. $= 2.07$.

[*]anirbandas@theory.tifr.res.in
[†]bdasgupta@theory.tifr.res.in
[‡]anupam.ray@theory.tifr.res.in

section into visible sector particles during the recombination era is also highly constrained by cosmic microwave background (CMB) observation [19–23]. Moreover, note that this cross section is approximately 2 orders of magnitude larger than the canonical value, $(2–3) \times 10^{-26}$ cm^3 s^{-1}, needed to produce the correct relic abundance of thermal DM [24]. The possibility of a boost in the annihilation due to the substructures in the DM halo was also explored [25], but such large boosts are unlikely due to substructures alone. Alternatively, the pulsars in the Milky Way (MW) have also been used to explain the positron spectrum with a possibility that the nearby young pulsars (e.g., Geminga and B0656 + 14) are contributing the most [10–12,26]. Authors in Ref. [26] used the HAWC data to show that these nearby pulsars are able to produce high-energy electrons, and could source from the local positron flux. However more recently, Ref. [27] used the HAWC and Fermi-LAT data to show that the diffusion of the positrons from the pulsars Geminga and B0656 + 14 is less efficient in the energy range (50–500) GeV in a two-zone diffusion model of the cosmic ray, but can contribute a substantial fraction of the positron flux. Therefore, the pulsar origin of the positron flux is also debated. All things considered, the shape of the positron flux spectrum in the high-energy region still remains poorly understood.

In this work, we show that it is possible to explain the positron excess using selectively enhanced, velocity-dependent p-wave annihilation of Majorana DM particles, following our earlier work [28]. An angular momentum and spin-dependent selection rule in the Sommerfeld effect enhances the p-wave DM annihilation. The p-wave nature implies that the rate is a nonmonotonic function of DM velocity. With a suitable choice of parameters, it can be maximum in the Milky Way-like galaxies, and less in smaller dwarf galaxies. The DM annihilation rate in the MW is shown in the contour plot in Fig. 2. In this parameter space, there are a few islands, marked with yellow asterisks, that yield such large galactic annihilation signal, and satisfy the thermal relic constraint, i.e., $\langle \sigma v \rangle_{\text{relic}} \simeq (2–3) \times 10^{-26}$ cm^3 s^{-1} [24], the dwarf galaxy bound from Fermi-LAT [18], and DM annihilation constraint from CMB observation [19], all at the same time. A representative positron flux spectrum is shown in Fig. 1. In this scenario, Majorana DM particles of mass $m_\chi = 780$ GeV annihilate into light dark sector scalars that promptly decay into electrons and muons. The annihilation rate in the Milky Way is $\sigma v \simeq 4.63 \times 10^{-24}$ cm^3 s^{-1}. Our purpose, in the remaining part of the paper, is to provide some details of this result and the related technical background, and to outline other phenomenological consequences of this model.

The paper is structured as follows: In Sec. II, we describe the dark sector of the model and its connection to the Standard Model (SM), followed by Sec. III discussing the cosmological thermal history. In Sec. IV, we compute and
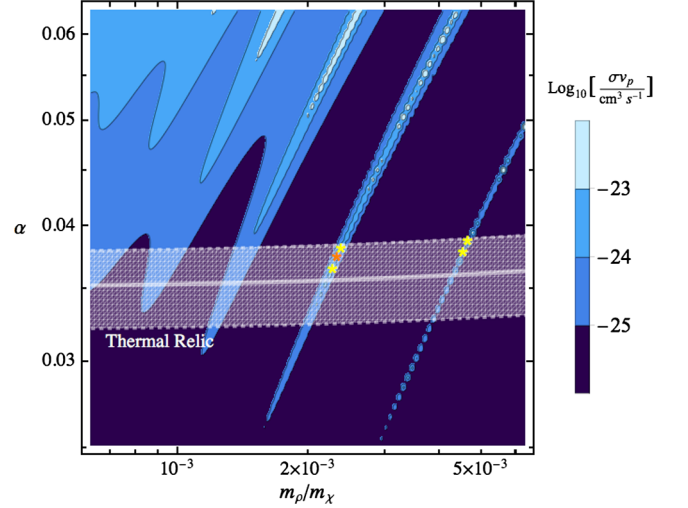

FIG. 2. A contour plot showing the DM annihilation rate in the Milky Way galaxy in the $m_\rho/m_\chi - \alpha$ plane. Overlaid is a white band of the region of parameter space that yields correct thermal relic abundance. The thick white line denotes the contour of $\sigma v_{\text{relic}} = 2.5 \times 10^{-26}$ cm^3 s^{-1}, bordered by the dashed contours of $\sigma v_{\text{relic}} = 3 \times 10^{-26}$ cm^3 s^{-1} (above) and 2×10^{-26} cm^3 s^{-1} (below). The points marked with asterisks can provide large enough annihilation cross section to explain the positron flux excess. In this work, we use the point marked with the orange asterisk.

explain the selective Sommerfeld enhancement in annihilation, and in Sec. V outline the relevant phenomenology. We summarize and conclude in Sec. VI.

II. THE MODEL

We extend the SM by a dark sector that has a complex scalar Φ and a Dirac fermion χ, with charges -2 and $+1$, respectively, under an approximate global $U(1)$ symmetry. The right-handed leptons in the Standard Model, l_R, are also charged $+2$ under this symmetry. The SM Lagrangian is therefore extended by $\mathcal{L}_{\text{BSM}} \supset \mathcal{L}_{\text{dark}} + \mathcal{L}_{\text{portal}}$, where

$$\mathcal{L}_{\text{dark}} = \partial^\mu \Phi^\dagger \partial_\mu \Phi + \mu^2 |\Phi|^2 - \lambda |\Phi|^4 + i\bar{\chi}\partial\!\!\!/\chi - M\bar{\chi}\chi$$
$$- (f\Phi\chi^T C\chi/\sqrt{2} + \text{H.c.}) \quad \text{and}$$
$$\mathcal{L}_{\text{portal}} = \sum_{l=e,\mu} \frac{c_l}{\Lambda} \Phi H \bar{l}_L l_R + \text{H.c.} \tag{1}$$

The dark complex scalar Φ and the dark fermion χ are coupled to each other through a Majorana-type interaction [29] and the five-dimensional nonrenormalizable operator, involving the SM Higgs doublet H, acts as a portal between the dark and the visible sectors. After the spontaneous breaking of the $U(1)$ and the electroweak symmetries, it would lead to decay of the dark scalars into electrons and muons.

Note that the $U(1)$ is also explicitly broken at the tree level by the electron and muon masses. It is also possible to include explicit $U(1)$-breaking terms in $\mathcal{L}_{\text{dark}}$ itself. Such terms are crucial for determining the mass of the pseudo-Goldstone mode of Φ, but we do not model build that aspect and treat the pseudo-Goldstone mass as a free parameter.

The scalar potential induces nonzero vacuum expectation value (vev) v_Φ and splits Φ into a radial mode ρ and a pseudo-Goldstone mode η: $\Phi \to (v_\Phi + \rho + i\eta)/\sqrt{2}$, spontaneously breaking the $U(1)$ symmetry. Because of the Majorana-type coupling with the scalar, χ also splits into two Majorana particles, $\chi_1 = (\chi - \chi^c)/\sqrt{2}$ and $\chi_2 = (\chi + \chi^c)/\sqrt{2}$, with masses m_χ and $m_\chi + \Delta$ where $m_\chi \equiv M - fv_\Phi$ and $\Delta \equiv 2fv_\Phi$. After dark symmetry breaking, the resultant interactions between the fermions and the scalars are

$$-\frac{f}{2}\rho(\bar{\chi}_1\chi_1 - \bar{\chi}_2\chi_2) - \frac{f}{2}\eta(\bar{\chi}_2\chi_1 + \bar{\chi}_1\chi_2). \qquad (2)$$

With these interactions, $\{\chi_1,\chi_2\}$ constitute a two-level inelastic self-interacting dark matter (SIDM). The stable χ_1 is the DM candidate. After χ_1, χ_2 fall out of chemical equilibrium from the thermal bath, χ_2 quickly decays $\chi_2 \to \chi_1\eta$, and χ_1 forms the DM abundance. While the ρ couples similar DM states, η provides an off-diagonal interaction and couples different DM states. The symmetry-broken scalar potential reads

$$V(\rho,\eta) = \frac{1}{2}m_\rho^2\rho^2 + \frac{1}{2}m_\eta^2\eta^2$$
$$+ \lambda v_\Phi(\rho^3 + \rho\eta^2) + \frac{\lambda}{4}(\rho^4 + \eta^4 + 2\rho^2\eta^2), \qquad (3)$$

where $m_\rho = \sqrt{2\lambda}v_\Phi$ and the mass of η arises from the $U(1)$-breaking terms and its interaction with other particles in the thermal bath.

The higher-dimensional part of the Lagrangian yields

$$\mathcal{L}_{\text{portal}} = \sum_{l=e,\mu} \frac{c_l}{2\Lambda}\bar{l}_L l_R(v_H\rho + iv_H\eta$$
$$+ \rho h + i\eta h + v_\Phi h + v_\Phi v_H). \qquad (4)$$

These terms allow the dark scalars ρ and η to decay into a pair of leptons. Further, the Higgs boson h develops new decay channels $h \to \rho l^+ l^-$ and $h \to \eta l^+ l^-$ and a new contribution to its decay into lepton pairs. The masses of the e and μ get a contribution from the dark symmetry breaking as well.

At low energy, the DM annihilation phenomenology is controlled by the five free parameters $\{m_\chi, m_\rho, m_\eta, \alpha, \Delta\}$ where $\alpha \equiv f^2/(4\pi)$. We are interested in $m_\chi \gtrsim 300$ GeV to explain the AMS data, and fix $\Delta = 10^{-3}m_\chi$. We vary other parameters in the ranges $10^{-4} \leq m_\rho/m_\chi \leq 10^{-1}$ and

$10^{-2} \leq \alpha \leq 10^{-1}$. We discuss more about parameter values later. The decay of the scalars is determined by c_e, c_μ, and Λ. We take the value of the SM vev as $v_H = 256$ GeV.

III. COSMIC HISTORY

After the initial production of the dark sector particles through some mechanism like reheating, they decouple from the SM sector at a temperature T_*. We assume this to occur at a scale much higher relative to other energy scales in the theory. After decoupling, the SM and the dark sectors evolve independently with separate temperatures T (SM) and T_d (dark sector) with a ratio $\xi_d \equiv T_d/T$. After the dark sector symmetry breaking at $T_d \simeq v_\Phi$, the dark sector consists of two Majorana fermions χ_1, χ_2 and two scalars ρ, η. The comoving entropy of this sector changes only during the decays of χ_2 and ρ. In general ξ_d can be written as

$$\xi_d(T) = \left(\frac{g_d(T_*)g_{\text{SM}}(T)}{g_d(T_d)g_{\text{SM}}(T_*)}\right)^{1/3}. \qquad (5)$$

Here $g_d(T)$ and $g_{\text{SM}}(T)$ are the relativistic degrees of freedom in the dark and SM sectors, respectively, at temperature T. From Fig. 3, it is clear that the ratio ξ_d is never too far away from unity. After ρ and η go out of chemical equilibrium, they decay. Note that ρ could decay into η if $m_\eta < m_\rho/2$. However, as we show later, this is not the case in the region of the parameter space we are interested in. If the dark sector decouples from the visible sector early enough then ρ, η do not affect the predictions for N_{eff} and the Hubble parameter much during big bang nucleosynthesis, as was shown in [30]. Dark sector temperature T_d rises after the DM decouples from the thermal bath. Later during the QCD phase transition at $T \sim 200$ MeV, a large amount of entropy is dumped into the SM thermal bath, heating it up. This history depends on the relative values of M and T_* (see Fig. 3 in Ref. [30]). In this paper, we only consider the case in which the DM mass is above the QCD condensation scale.

We always assume that $M < v_\Phi$, so that the symmetry breaking occurs earlier than the chemical freeze-out of DM. The ensuing phenomenology was discussed in Refs. [29–31]. Before thermal freeze-out, both χ_1 and χ_2 are present in the thermal bath. Afterwards, out-of-equilibrium χ_2 particles decay and χ_1 forms the DM abundance. During relic annihilation, the interactions in Eq. (2) provide both annihilation and coannihilation,

$$\chi_1\chi_1, \chi_2\chi_2 \to \rho\rho, \eta\eta \qquad \{\text{annihilation}\},$$
$$\chi_1\chi_2 \to \rho\eta \qquad \{\text{coannihilation}\}. \qquad (6)$$

Figure 4 shows the Feynman diagrams for the annihilation and coannihilation. Even though in this case the DM is lighter than the symmetry breaking scale v_Φ, the scalar particle can be much lighter than $\chi_{1,2}$ if the hierarchy $v_\Phi > M > \mu_\Phi$ is maintained.
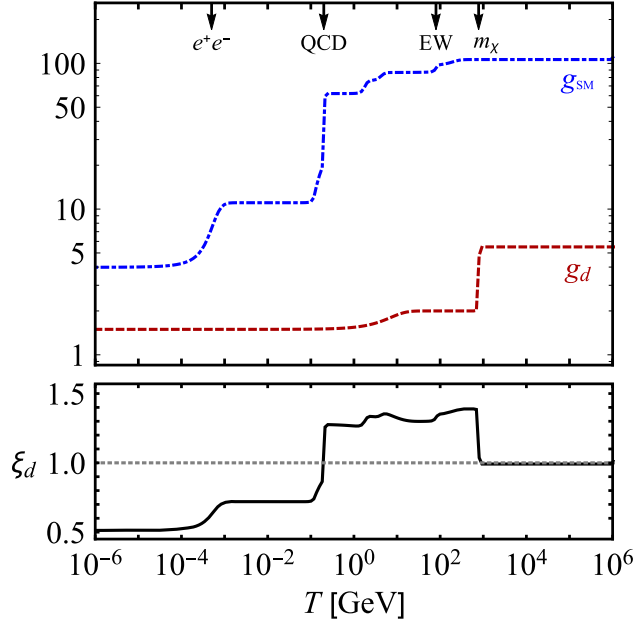
FIG. 3. Variation of the relativistic degrees of freedom in the SM g_{SM} (dot-dashed blue) and the dark sector g_d (dashed red). The lower panel shows the ratio of the dark sector temperature to the SM temperature ξ_d (solid black).
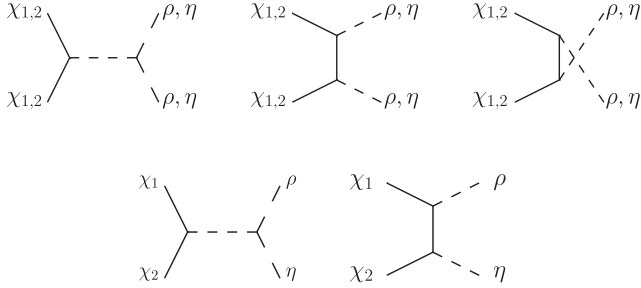

FIG. 4. The Feynman graphs for DM annihilation and coannihilation.

As χ_1, χ_2 are both Majorana particles, their annihilation into two scalars does not have any s-wave ($\ell = 0$) component, and hence is p-wave ($\ell = 1$) suppressed. But coannihilation is not subject to such suppression. Even after the DM is not in chemical equilibrium with the other particles in the dark sector, it exchanges kinetic energy with them as kinetic decoupling happens later than the chemical decoupling. In this case, the elastic scattering between χ_1 and η helps keep the DM in kinetic equilibrium with η. It was shown in Ref. [30] that a small mass split between the two DM states enhances this scattering cross section through a resonance. This delays the DM kinetic decoupling and may ameliorate the missing satellite problem [32,33].

IV. ENHANCED ANNIHILATION

When the annihilating DM particles interact with each other through a long-range potential $V(r)$, their initial plane waves are modified. The Sommerfeld effect is the change in the annihilation rate resulting from this wave function modification. It is a nonperturbative effect. It is quantified by the Sommerfeld factor S_ℓ and is written as

$$\sigma_\ell = S_\ell \sigma_{0\ell}. \qquad (7)$$

Here σ_ℓ and $\sigma_{0\ell}$ are the total and the perturbative cross sections, respectively, for the orbital angular momentum $\ell = 0, 1, 2$, and so on. The process is enhanced if $S_\ell > 1$, suppressed if $S_\ell < 1$.

This effect is most pronounced when the DM particles are nonrelativistic. In this limit, it is useful to use the formalism of nonrelativistic effective theory (NREFT). In NREFT, the wave function Ψ_ℓ of the heavy particles obeys the Schrödinger equation with the interaction potential $V(r)$. Relative to the range of $V(r)$, the actual annihilation process is short range and is usually taken to be happening at the origin. Hence the Sommerfeld factor S_ℓ is defined as the change in the amplitude of the wave function $\Psi_\ell(\mathbf{r})$ at the origin due to the potential,

$$S_\ell \equiv \left| \frac{\Psi_\ell(0)}{\Psi_\ell(0)_{\text{free}}} \right|^2. \qquad (8)$$

Here, $\Psi_\ell(\mathbf{r})_{\text{free}}$ is the wave function in the absence of any potential, i.e., the free particle plane wave. Depending on the nature of the mediator and the charges of the interacting particles, $V(r)$ can be either attractive or repulsive, causing enhancement or suppression, respectively, to the annihilation rate.

The possible two-particle states in the present case are the following.

$$\left.\begin{array}{c} |\chi_1\chi_1\rangle \\ |\chi_2\chi_2\rangle \end{array}\right\} \text{Annihilation}$$

$$\left.\begin{array}{c} |\chi_1\chi_2\rangle \\ |\chi_2\chi_1\rangle \end{array}\right\} \text{Coannihilation.} \qquad (9)$$

The η interaction leads to mixing between the two states in the annihilation space. But the annihilation and coannihilation subspaces remain decoupled as neither ρ nor η interaction can couple one to the other. As mentioned before, the annihilation subspace does not have any s-wave process, but coannihilation has both s- and p-wave contributions. Although the two states for coannihilation consist of the same particles, we write them separately because it is easier to understand the transition from one to the other due to the η exchange. We first compute the perturbative cross sections for these three processes using NREFT, and then calculate the Sommerfeld factors by solving the matrix Schrödinger equations numerically.

A. Annihilation matrices

DM annihilation typically happens at a length scale $1/m_\chi$. Therefore, typical momentum exchange in such a process is $\sim \mathcal{O}(m_\chi)$. In NREFT, the annihilation process cannot be described using tree-level graphs. However, like in any EFT, all information regarding these high-energy processes is contained the series of higher-dimensional operators,

$$\mathcal{L}_{\text{eff}} = \sum_{i,j,k,l} c_{ij,kl}(m_\chi)\bar{\chi}_i\chi_j\bar{\chi}_k\chi_l + \cdots. \quad (10)$$

Here $c_{ij,kl}(m_\chi)$ are the Wilson coefficients computed at the scale m_χ, and $\bar{\chi}_i\chi_j\bar{\chi}_k\chi_l$ are four-Fermi operators, $\bar{\chi}_1\chi_1\bar{\chi}_1\chi_1$, $\bar{\chi}_1\chi_2\bar{\chi}_1\chi_1, \bar{\chi}_2\chi_2\bar{\chi}_1\chi_1$, and so on. The $c_{ij,kl}$ coefficients can be calculated by matching a four-point amplitudes in the full theory with the corresponding four-Fermi operator in the effective Lagrangian.

To classify the effective operators according to the spin and angular momentum of the two-body states, we write the Dirac spinors of χ_i using the Pauli two-component spinors ξ_i, η_i as

$$u_i(\mathbf{p}) = \sqrt{\frac{E_i + m_i}{2E}}\begin{pmatrix} \xi_i \\ \frac{\sigma \cdot \mathbf{p}}{E_i + m_i}\xi_i \end{pmatrix},$$

$$v_i(-\mathbf{p}) = \sqrt{\frac{E_i + m_i}{2E}}\begin{pmatrix} \frac{-\sigma \cdot \mathbf{p}}{E_i + m_i}\eta_i \\ \eta_i \end{pmatrix}. \quad (11)$$

With these expressions for the spinors, we expand the operators in Eq. (10) in powers of $|\mathbf{p}|/m_\chi$ to find the effective operators for each spin and angular momentum,

$$\mathcal{L}_{\text{eff}} = \sum_{i,j,k,l}\sum_{\ell,s} c_{ij,kl}(^{2s+1}\ell_J)\mathcal{O}_{ij,kl}(^{2s+1}\ell_J), \quad (12)$$

where $\mathcal{O}_{ij,kl}(^{2s+1}\ell_J)$ are the effective operators consisting of the spinors ξ_i, η_i corresponding to spin s and angular momentum ℓ of the associated two-body states [34] (see the Appendix A for details about these operators).

Once the Wilson coefficients are known, the annihilation cross sections into the light scalars can be found by using the Cutkosky theorem: the annihilation cross section of a process $\chi_i\chi_j \to X_A X_B$ is proportional to the imaginary part of the Wilson coefficient of the operator $\chi_i\chi_j \to X_A X_B \to \chi_i\chi_j$ [35]. If $(\sigma v)_{\chi_i\chi_j \to X_A X_B} \equiv \Gamma(\chi_i\chi_j \to X_A X_B)$ is the annihilation rate, then

$$(\sigma v)_{\chi_i\chi_j \to X_A X_B} = 2\text{Im}[c_{ij,ij}(^{2s+1}\ell_J)]. \quad (13)$$

In addition to $|\mathbf{p}|/m_\chi$, we also expand the operators in powers of other dimensionless parameters, such as $m_\rho/m_\chi, m_\eta/m_\chi$, and Δ/m_χ. We computed the Wilson

coefficients for both annihilation and coannihilation using the FeynCalc package in *Mathematica*[1] and classified them according to their spin and angular momentum [36]. The procedure is described in detail in Appendix A.

To the leading order in the dimensionless parameters mentioned above, we get the following annihilation matrices.

$$\Gamma^{\text{ann}}_{\ell=1,s=1} = \frac{2\pi\alpha^2 v^2}{m_\chi^2}\begin{pmatrix} +1 & +1 \\ +1 & +1 \end{pmatrix},$$

$$\Gamma^{\text{co-ann}}_{\ell=0,s=1} = \frac{\pi\alpha^2}{16m_\chi^2}\begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix},$$

$$\Gamma^{\text{co-ann}}_{\ell=1,s=1} = \frac{\pi\alpha^2 m_\rho^2 v^2}{16m_\chi^4\Delta^2}\begin{pmatrix} +1 & +1 \\ +1 & +1 \end{pmatrix}. \quad (14)$$

The $\ell = 0$ coannihilation matrix has opposite signs for the off-diagonal entries relative to the diagonal elements. However, all elements of the $\ell = 1$ matrices have similar sign. This fact is related to the particle exchange symmetry, and is discussed in more detail when we describe our results.

B. Sommerfeld factors

The nonrelativistic wave function of the DM particles obeys the matrix Schrödinger equation,

$$[\mathcal{D}_{ab} + V_{ab}(r)](u_\ell(r))_{bc} = 0. \quad (15)$$

Here $u_\ell(r)$ is a 2×2 matrix proportional to the radial parts $R(r)(=u(r)/r)$ of the full wave function $\Psi_\ell(\mathbf{r})$, and $\mathcal{D}_{ab} = (-\frac{1}{2\mu}\frac{d^2}{dr^2} + \frac{\ell(\ell+1)}{2\mu r^2} - E)\delta_{ab}$ is the differential operator. The elements of $u_\ell(r)$ are the transition amplitudes $(u_\ell(r))_{ab} \equiv \langle b|a\rangle$ of going from state $|a\rangle$ to $|b\rangle$, with $|a\rangle, |b\rangle$ being one of the three possible two-body states in Eq. (6). As the particles are in the scattering state before annihilation, the total energy of the ath two-body state is

$$E_a \equiv \frac{1}{2}\mu_a v_{\text{rel}}^2 = 2\mu_a v^2 = \frac{k_a^2}{2\mu_a}, \quad (16)$$

where v is the velocity of individual particles in the center-of-mass frame, and $k_a = \mu_a v_{\text{rel}}$ is the momentum of the ath two-body state. The potential $V(r)$ arises from the exchange of the scalars. It is given by

$$V(r) = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \quad (17)$$

with $V_{11} = -\alpha e^{-m_\rho r}/r$, $V_{12} = V_{21} = -\alpha e^{-m_\eta r}/r$, and $V_{22} = -\alpha e^{-m_\rho r}/r + 2\Delta$ for annihilation and $V_{22} = -\alpha e^{-m_\rho r}/r$ for

---

[1]*Mathematica* notebooks are available at this https://github.com/anirbandas89/NREFT_Matching.

coannihilation. The extra term 2Δ in V_{22} for annihilation comes from the mass gap between the two two-body states $|\chi_1\chi_1\rangle$ and $|\chi_2\chi_2\rangle$. No such mass gap exists for coannihilation.

We follow Ref. [37] to compute the Sommerfeld factors for the multilevel DM. We quote the final expression for the Sommerfeld factor, $S_a(^{2s+1}\ell_J)$, that is given as

$$\left[\frac{(2\ell-1)!!}{k_a^\ell}\right]^2 \frac{(T^\dagger)_{ab}\mathcal{O}_{bc}(^{2s+1}\ell_J)T_{ca}}{\mathcal{O}_{aa}(^{2s+1}\ell_J)}. \qquad (18)$$

No summation is implied over the repeated index a in the denominator. The matrix T is defined as the complex conjugate of the amplitude of the irregular solutions $v_\ell(r)$ of Eq. (15) in the large r limit,

$$(v_\ell(r \to \infty))_{ab} = (T^\dagger)_{ab}e^{-ik_a r}. \qquad (19)$$

The detailed procedure to compute the T matrix is given in Appendix B.

We solved the Schrödinger equation using two methods: (i) directly using NDSolve in *Mathematica*, and (ii) using the variable phase method [38,39]. The direct method works well as long as the kinetic energy of the incoming particles is above the threshold for the $|\chi_2\chi_2\rangle$ state, i.e.,

$$E \geq 2\Delta. \qquad (20)$$

When the incoming particles are below threshold, the wave functions for the $|\chi_2\chi_2\rangle$ final state are not a scattering state anymore. In this case, the direct method fails as it simultaneously tries to solve for scattering and bound state solutions within a single system. Reference [39] proposed that the variable phase method can be used in such cases. Instead of solving for the full solution, this method solves for the modification in the wave function from the free particle solution because of the long-range potentials. We have verified the results from the variable phase method with direct method solutions for the above threshold parameters. We found an excellent match between the results from two methods.

To reduce the number of parameters, we fix $m_\eta = 0.7m_\rho$ for all the results shown in this paper. A different choice for m_η does not change our results qualitatively. In Figs. 5 and 6, we show the Sommerfeld factors and the annihilation rates, respectively, as functions of DM velocity v. For the figures in the next two subsections, we choose a representative set of values for the parameters: $\alpha = 0.0373, m_\rho/m_\chi = 0.0023, \Delta/m_\chi = 0.001, v = 100 \text{ km s}^{-1}$, which approximately corresponds to the point marked with an orange asterisk in Fig. 2.

Both p-wave annihilation and coannihilation show large enhancement in the small velocity limit in Fig. 5. The enhancement factor has a $1/v^3$ dependence in the intermediate velocity regime. The difference between the
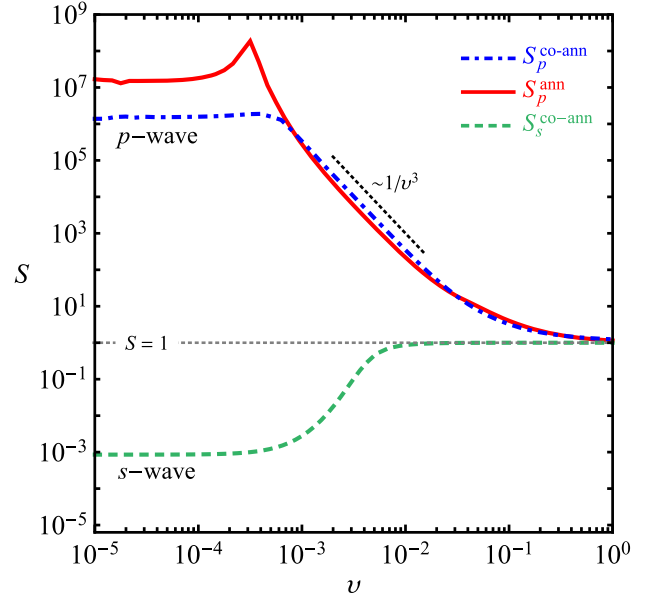

FIG. 5. Sommerfeld factors for annihilation (p wave) and coannihilation (s and p wave) as a function of DM velocity v.

annihilation and coannihilation factors is due to the mass gap between the two states. However, the s-wave Sommerfeld factor is less than 1. The p-wave annihilation and the s-wave coannihilation rates are shown in Fig. 6. The $\sim v^2$ scaling of the bare p-wave cross section yields the $\sim v^2$ and $\sim v^{-1}$ behavior of σv_p in the small and intermediate velocity regimes, respectively. We note that the
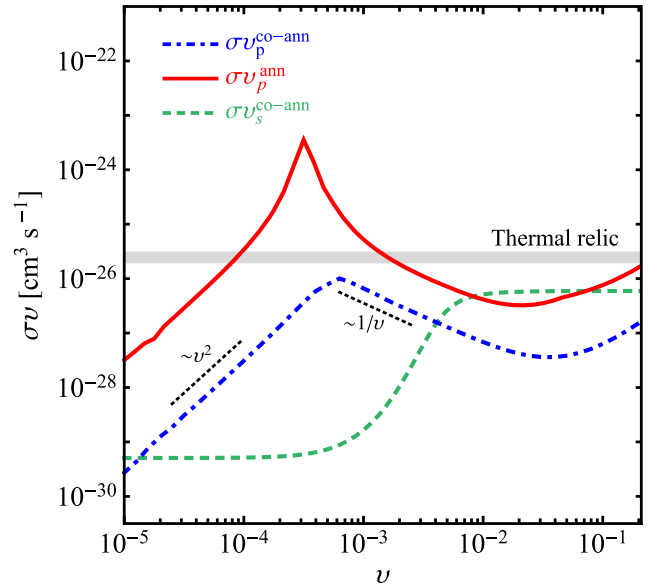

FIG. 6. Sommerfeld-enhanced p-wave annihilation rate $S_p(\sigma v)_p$ (solid red) as a function of DM velocity v. The s- and p-wave coannihilation rates are shown for comparison as the dashed green and dot-dashed blue curves. The thermal relic annihilation rate $\sigma v_{\text{relic}} = (2–3) \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$ is shown as a gray band.

relic annihilation cross section is dominated by annihilation because of a residual Sommerfeld enhancement factor $S_p \simeq 2$ even for $v_{\text{relic}} \simeq 0.2$. Therefore, the p-wave annihilation process contributes more by a factor of ~ 2 than coannihilation in the early Universe. This more carefully computed result disagrees somewhat with the conclusions made in previous papers [28,30,31]. However, we must note that at this large velocity $v \sim 0.2$, the partial wave expansion of the cross section may not be very accurate as higher partial wave contributions become important. Figure 7 shows the m_ρ dependence of the factors. In addition to a large overall enhancement for p wave, a resonance feature is also present for certain values of m_ρ. We explain the origin of these features in the rest of this section.

C. Particle exchange symmetry

We use the particle exchange symmetry following Ref. [28]. Suppose A and B are two fermions. They can form two two-body states, namely $|AB\rangle$ and $|BA\rangle$ of total angular momentum ℓ and spin s. However, these two states consist of the same set of particles and are related to each other through

$$|AB\rangle = (-1)^{\ell+s}|BA\rangle. \tag{21}$$

This factor has three components: $(-1)^\ell$ from relative angular momentum, $(-1)^{s+1}$ from their spins, and (-1) due to Wick exchange of two fermions.

We can directly apply this to the coannihilation channel in the present model implying that two states $|\chi_1\chi_2\rangle$ and

$|\chi_2\chi_1\rangle$ in the matrix Schrödinger equation are related to each other,

$$|\chi_1\chi_2\rangle = (-1)^{\ell+s}|\chi_2\chi_1\rangle. \tag{22}$$

It further implies that we can combine the two equations into a single equation with an effective potential that is a linear combination of the diagonal and off-diagonal potentials (see Appendix C for details),

$$V_{\text{eff}} = V_{11} + (-1)^{\ell+s}V_{12}. \tag{23}$$

Therefore, V_{eff} has the following forms for the two cases,

$$\begin{aligned}
V_{\text{eff}}^{\ell=0,s=1} &= -\frac{\alpha e^{-m_\rho r}}{r} + \frac{\alpha e^{-m_\eta r}}{r}, \\
V_{\text{eff}}^{\ell=1,s=1} &= -\frac{\alpha e^{-m_\rho r}}{r} - \frac{\alpha e^{-m_\eta r}}{r}.
\end{aligned} \tag{24}$$

We note that in the $\ell = 0$ case, the difference between the two potentials is acting as the effective 1D potential. When $r \ll 1/m_\rho$ and $1/m_\eta$, V_{eff} saturates at the value $(m_\rho - m_\eta)$. For larger r, it gradually decreases to 0, never becoming negative. The nature of the net potential thus becomes *repulsive* (See Fig. 8). However, for $\ell = 1$, the diagonal and off-diagonal potentials are added with the same sign, and hence yield an even stronger attractive potential. This explains the behavior of the different Sommerfeld factors in Figs. 5 and 7. The effective repulsive nature of V_{eff} leads to the suppression in $S_s^{\text{co-ann}}$.
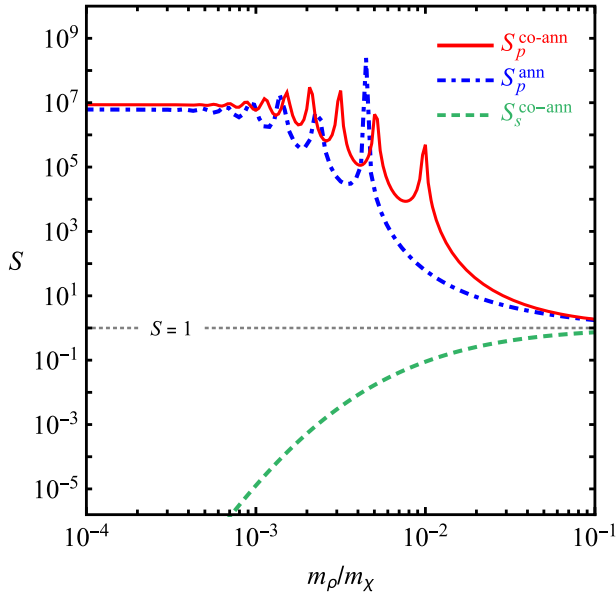

FIG. 7. Sommerfeld factors for annihilation (p wave) and coannihilation (s and p wave) as a function of the diagonal mediator mass m_ρ.
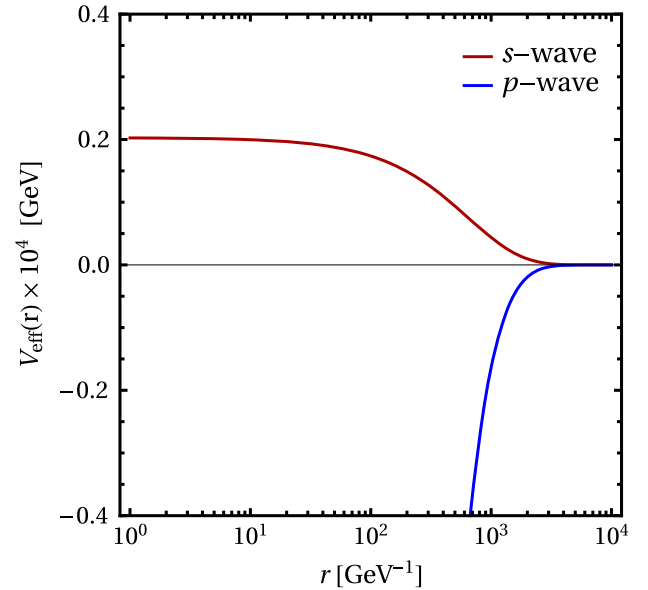

FIG. 8. The one-dimensional (1D) effective potentials for coannihilation in Eq. (24).

D. Velocity dependence and resonances

The $1/v^3$ dependence of S_p in Fig. 5 can be understood by taking the limit $m_\rho/m_\chi \ll 1$ and $\alpha/v \gg 1$. In this Coulomb limit, an analytic expression for S_ℓ for general angular momentum ℓ is given by [40]

$$S_\ell = S_0 \prod_{j=1}^{\ell} \left(1 + \frac{\alpha^2}{j^2 v^2} \right), \qquad (25)$$

where S_0 is the s-wave factor in the Coulomb approximation

$$S_0 = \frac{\pi\alpha/v}{1 - e^{-\pi\alpha/v}}. \qquad (26)$$

Therefore, for small velocity, $S_\ell \sim 1/v^{2\ell+1}$. However, for finite m_ρ, the Sommerfeld factor does not grow indefinitely for smaller velocity. Instead, it saturates to a constant value when the de Broglie wavelength of the DM particles gets much longer than the range of the potential, i. e., approximately when the condition $\mu_1 v/m_\rho \ll 1$ is met. All of these features are evident in Fig. 5.

Resonances from virtual bound states arise and cause large Sommerfeld enhancement as can be seen in Fig. 7. This occurs whenever the range of the potential matches a multiple n of the Bohr radius $1/(\alpha m_\chi)$ of the DM particles, i.e., [40]

$$\frac{\alpha m_\chi}{\kappa m_\rho} = n^2, \quad n = 1, 2, 3, \ldots. \qquad (27)$$

Here $\kappa \simeq \pi^2/6$. Note that this resonance condition is only approximately true in the present model as we have two mediators with slightly different masses. Also the resonance positions are shifted by the mass gap Δ [37].

A consequence of the nonmonotonic velocity dependence of the p-wave Sommerfeld factor is that it predicts different annihilation rates of DM in astrophysical objects of different size. In the net p-wave annihilation rate, the perturbative cross section provides a velocity scaling $\sim v^2$. Therefore, $S_p(\sigma v)_p$ increases as $\sim v^2$ for small velocity, reaches a maximum, and then falls as $\sim 1/v$ as shown in Fig. 6. The position of the maximum annihilation depends on the ratio m_ρ/m_χ. Hence this model naturally predicts large DM annihilation signal in the galaxies, but very small signal from the dwarf galaxies.

V. PHENOMENOLOGICAL CONSEQUENCES

A. Galactic positron excess

The AMS-02 and several other experiments have observed an excess in the positron spectrum from the Galaxy, as mentioned in the introduction. DM annihilation in our model can be used to explain such observation. To this end, we show the variation of the p-wave annihilation

rate $(\sigma v)_p$ in the Milky Way with m_ρ/m_χ and α in Fig. 2. The cross section is enhanced for large α and small m_ρ/m_χ ratio, with a resonance feature as discussed above. The points within the overlaid white band yield a relic annihilation cross section within $(2-3) \times 10^{-26}$ cm^3 s^{-1} to satisfy the relic abundance constraint [24].

The points marked with asterisks yield DM annihilation cross section $\sigma v > 10^{-24}$ cm^3 s^{-1}, sufficient to explain the AMS-02 positron flux excess, without running into any problems with thermal relic or dwarf galaxy constraints. The typical annihilation cross section in dwarf galaxies is $\lesssim 6 \times 10^{-28}$ cm^3 s^{-1}, which is well below the Fermi-LAT bound in the e^+e^- and $\mu^+\mu^-$ channels [18]. Because of the suppression in the s-wave channel as shown in Fig. 6, the DM annihilation rate during recombination era (when DM velocity $v \sim 10^{-8}$) is also predicted to be suppressed at $\sim 10^{-29}$ cm^3 s^{-1}, which is much smaller than the current experimental bound [21]. Note that the p-wave rate decreases as $\sim v^2$ towards small velocity and is insignificant during recombination. We show the positron flux from such a representative point in Fig. 1. We used the publicly available code PPPC4DMID to compute the positron flux spectrum in the $\chi_1\chi_1 \to VV \to 4e, 4\mu$ channel, where the scalar V represents ρ and η, from the Galaxy after diffusion [41].

An astrophysical background of positron flux was assumed, following Ref. [7],

$$F_{e^+}(E) = C_d^2 \left(\frac{E^2}{\hat{E}^2} \right) \left(\frac{\hat{E}}{E_1} \right)^{\gamma_d}, \qquad (28)$$

where $\hat{E} = E + \phi_{e^+}$ is the positron energy in the interstellar space. The dark sector and background parameters used in Fig. 1 are listed in Table I. For $m_\chi = 780$ GeV, the required mediator mass is about $m_\rho \simeq 1.5$ GeV (see Fig. 2). Therefore, ρ and η can decay to the electron and muon. Note that we are not using the background parameters quoted in Ref. [7] as this is an independent model. The fit shown in Fig. 1 corresponding to the parameter values quoted in Table I was found by a crude scan over the

TABLE I. The dark sector and background parameters [see Eq. (28)] used in Fig. 1.

Parameter	Value
m_χ	780 GeV
σv	4.63×10^{-24} cm^3 s^{-1}
B.R.(e^+e^-)	38%
B.R.$(\mu^+\mu^-)$	62%
Halo profile	Einasto
C_d	6.42×10^{-2} (GeV m^2 s sr)$^{-1}$
ϕ_{e^+}	0.869 GeV
γ_d	-3.6
E_1	7 GeV

parameter space. Although the fit of the theoretical curve to the data appears acceptable, the goodness of the fit is rather modest with $\chi^2/\text{d.o.f} = 2.07$.

Three dark sector parameters $m_\chi, \sigma v$, and $\text{BR}(e^+e^-)$ determine the shape and position of the bump in the positron flux. DM mass m_χ fixes its position as the DM particles are highly nonrelativistic today; σv and $\text{BR}(e^+e^-)$ together determine the amplitude and the spread of the peak. It might seem from Fig. 1 that increasing m_χ could give a better fit to the last three data points in the high-energy end. However, as the peak shifts towards higher energy, the fit in the intermediate energy becomes poor. Increasing σv has the effect of increasing the overall amplitude of the flux. The height of the peak is also partially controlled by $\text{BR}(e^+e^-)$; decreasing this parameter lowers the height increasing its width. We used the Einasto profile for the MW halo,

$$\text{Einasto:}\ \rho(r) = \rho_s \exp\left[-\frac{2}{\alpha}\left(\left(\frac{r}{r_s}\right)^\alpha - 1\right)\right]. \quad (29)$$

We used the default values for the parameters implemented in PPPC4DMID: $\rho_s = 0.033\ \text{GeV cm}^{-3}$, $r_s = 28.44$ kpc, fand $\alpha = 0.17$ [42]. With these parameters, the local DM density is $0.32\ \text{GeV cm}^{-3}$. In this context, we mention the effect of the morphological dependence of the J factor on the positron flux [43]. Averaging the annihilation rate over velocity decreases the effective rate, given that our parameters correspond to a resonance at some v. However, several factors mitigate this effect. The Einasto density profile that we use is expected to give a smaller morphological dependence than a cuspy profile used in Ref. [43]. Further, the AMS-02 data for the positron flux is not directional; rather it is the total flux integrated over all directions as observed by the spectrometer, thus diluting this effect. To arrive at a more quantitative conclusion, however, the calculation of Ref. [43] must be repeated for the nonmonotonic velocity dependence of the Sommerfeld factor.

The $\chi^2/\text{d.o.f}$ reported here can be easily improved. For example, fitting the low and intermediate energy data points more precisely, at the cost of fitting the higher energy data less well, improves the χ^2 because of the relatively large error bars for the high-energy data points. If more data shrink these error bars (without changing the central values), they would easily rule out such parameters. Thus we have chosen to show a benchmark point that will survive such a scenario to a greater extent. We also note the glitches in the AMS-02 data around $E \sim 100$ GeV, which worsen the fit. We believe more data in future will settle these issues.

The annihilation products ρ, η couple to the visible sector through the charged leptons e and μ with branching ratios 38% and 62%, respectively, to get an acceptable fit. One may ask if it is possible to get these branching ratios within our model. The $\Phi H \bar{l}_L l_R$ term, shown in Eq. (1), leads to decay of ρ and η generated in the DM annihilations. We compare their decay times with $\tau_c (\simeq 47$ yrs), which is the typical time required by them to escape the MW without any scattering. The decay widths are parametrized by c_e, c_μ, and Λ. For $c_e, c_\mu \gtrsim 10^{-5}$ and $\Lambda = 10^{13}$ GeV, their lifetimes $\tau_{\rho,\eta}$ are less than τ_c, as shown in Fig. 9 and 10. For our purpose, their decay can be assumed to be prompt.

In passing, we emphasize that the aim of this work is to demonstrate the possibility of explaining the positron flux excess using p-wave Sommerfeld-enhanced DM annihilation in the MW. Given the moderate goodness of fit, we do not claim that this particular DM model is an excellent candidate to explain the excess. Rather, we point out the uniqueness of this solution as it can accommodate the required large annihilation cross section without conflicting the dwarf galaxy or CMB data. We hope that this work will spawn a new direction in model-building and associated cosmic ray data analysis. Note that the selective Sommerfeld enhancement described in this paper is a general feature of any multilevel self-interacting DM models with light mediator.

B. Other BSM effects

The $\Phi H \bar{l}_L l_R$ operator also modifies e and μ masses, as well as their coupling with the SM Higgs boson h. The electron and muon masses are predicted to change by an amount $\sim c_l v_\Phi v_H/\Lambda$. The present experimental precisions
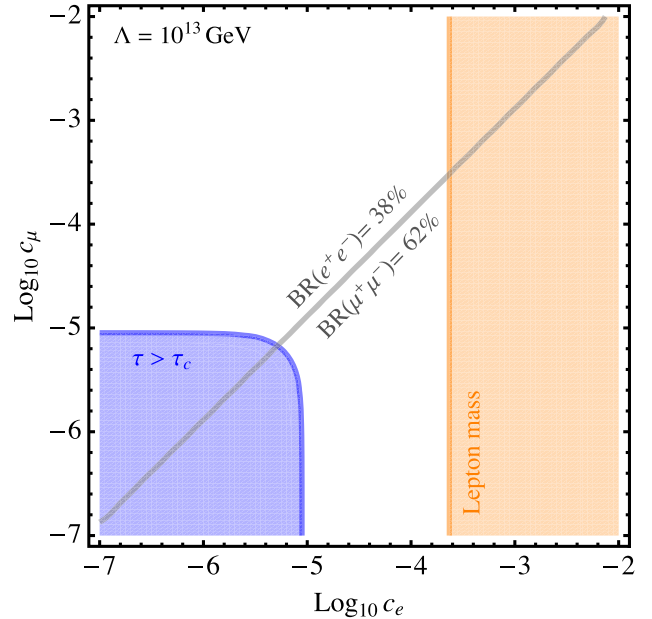

FIG. 9. The model parameter space in the $c_e - c_\mu$ plane for $\Lambda = 10^{13}$ GeV. The gray line is the contour of $\text{BR}(e^+e^-) = 38\%$ and $\text{BR}(\mu^+\mu^-) = 62\%$. In the blue-shaded region, ρ and η decay times are larger than τ_c. See the text for definition. The orange-shaded region is excluded by lepton mass measurement.
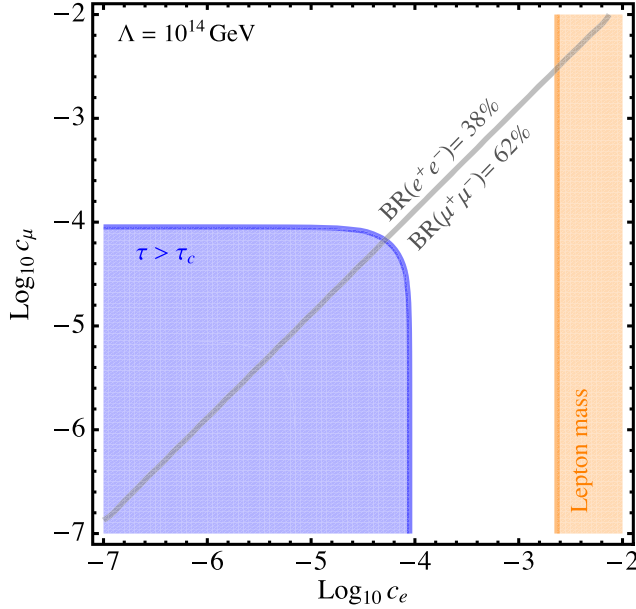
FIG. 10. Same as Fig. 9 but with $\Lambda = 10^{14}$ GeV. The position of the gray line does not change as the branching ratio does not have any Λ dependence.
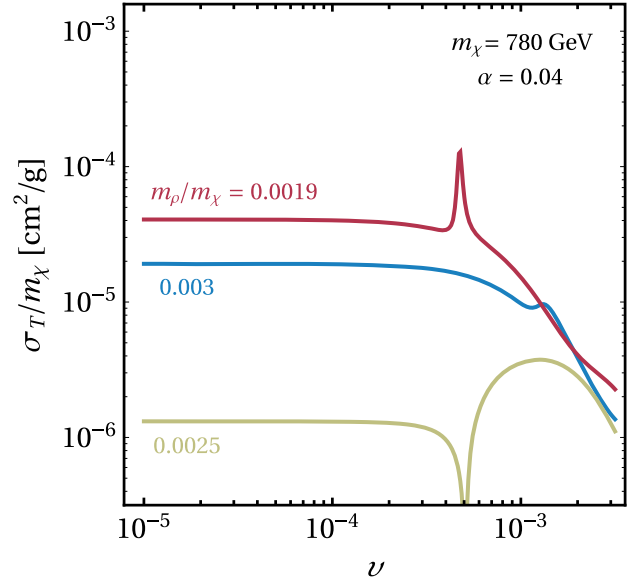

FIG. 11. The elastic transfer cross section σ_T for $\chi_1\chi_1 \rightarrow \chi_1\chi_1$ as a function of DM velocity. The mass gap is taken to be small, $\Delta = 0.1$ MeV, to ensure numerical stability. Larger Δ changes the cross section at most by a factor of a few.

of their mass measurement are very high, at the level of $\sim 10^{-8}$ [44]. Specifically, electron mass measurement constrains our parameter space significantly. The exclusion regions due to this are shown as orange-shaded regions in Figs. 9 and 10. The branching ratio of Higgs decay into muons is also affected. However, the resulting change in the Higgs signal strength in the $\mu^+\mu^-$ channel is very small relative to the current measurement; hence we do not consider it any further [45,46].

In Figs. 9 and 10, we show the contour of $BR(e^+e^-) = 38\%$ and $BR(\mu^+\mu^-) = 62\%$ with a gray solid line. Everywhere on the gray lines, the model produces a good fit to data, though the blue shaded region in the lower left is disfavored because the scalars decay too slowly ($\tau > \tau_c$) and the orange shaded region in the top right is disfavored because the fractional shift to lepton masses is larger than 10^{-8}.

C. Self-scattering

The long-range attractive potential would enhance the DM self-scattering cross section as well. The large self-scattering cross section, say, $\sigma/m_\chi \simeq 1 - 10$ cm^2 g^{-1}, may affect DM halo formation and produce a core at the center [47,48]. At the same time, the bullet cluster observation puts an upper limit $\sigma/m_\chi \lesssim 0.1$ cm^2 g^{-1} on this cross section [49].

In Fig. 11 we show the elastic transfer cross section σ_T as a function of DM velocity v for three different mediator masses. The red curve corresponds to a resonant point in Fig. 2, and has $\sigma_T \simeq 4 \times 10^{-5}$ cm^2 g^{-1} at dwarf-scale velocity $v \simeq 10^{-4}$, which is not large enough to produce

cores in the DM halos. This is not surprising because indirect detection and CMB experiments put strong constraints on SIDM models with light mediators, and it is difficult to satisfy these constraints along with large self-scattering cross sections to produce the core in the DM halo [50].

However, multilevel SIDM models can generically have interesting scattering phenomenology originating from the inelasticity in the system. DM particles in the ground state can upscatter to the excited state and quickly deexcite to the ground state by emitting a light mediator particle that escapes the halo. As a result, the halo dissipates energy at a certain rate and cools [51,52]. This cooling mechanism can significantly affect the halo dynamics and profile in other parts of the parameter space where self-scattering is large [53,54].

D. Kinetic decoupling of dark matter

The scattering between DM χ_1 and η keeps the DM in kinetic equilibrium with the relativistic η particles longer than the usual. The kinetic decoupling happens roughly when the momentum transfer rate $\Gamma_{\eta\chi} = (T_d/m_\chi)n_\chi\sigma_{\eta\chi}$ becomes smaller than the Hubble expansion rate, $\Gamma_{\eta\chi} \lesssim H(T)$. A t-channel resonance due to the mass gap Δ between χ_1 and χ_2 enhances the cross section, and delays the kinetic decoupling of DM further. Such late kinetic decoupling of DM suppresses structure formation at low halo mass scale, and can explain the apparent under-abundance of satellite galaxies around the MW [33]. It was already pointed out in Ref. [30] that for $m_\chi \simeq 1$ GeV, $\Delta/m_\chi \simeq 10^{-4.5}$, and massless η (acting as dark radiation),
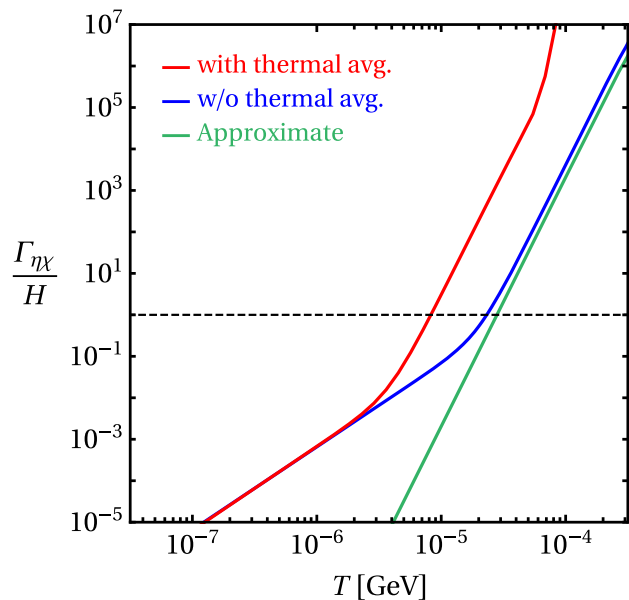
FIG. 12.　The ratio of the momentum transfer rate $\Gamma_{\eta\chi}$ between χ_1 and η to the Hubble parameter H. The chosen parameters are $m_\chi = 1$ GeV, $m_\rho = 1$ MeV, $\Delta = 1$ MeV, $\alpha = 10^{-4.5}$.

the kinetic decoupling temperature is ~0.5 keV, which is essential to solve the missing satellite problem. This temperature decreases further by a factor of a few when a more exact analytic expression for the cross section is used and the thermal distributions of χ_1 and η are taken into account, as shown in Fig. 12.

However, note that explaining the positron excess requires the DM mass to be $m_\chi \approx 800$ GeV, which would increase the decoupling temperature by several orders of magnitude from the keV scale. Also in the present scenario, the mediator mass $m_\eta \sim 1$ GeV, which ceases to be relativistic much earlier. Thus in the model we present, there is no delayed kinetic decoupling. Of course, one could consider DM with $M \approx 800$ GeV and a nearly massless η, but with much smaller Δ/m_χ to keep the decoupling temperature at keV scale. The Sommerfeld effect phenomenology would change significantly in such a scenario and requires a separate study.

VI. SUMMARY AND OUTLOOK

In this work, we have explored the possibility of explaining the bumps in the galactic positron flux, seen by AMS-02 and several other experiments, using Sommerfeld-enhanced DM annihilation without overpredicting the DM annihilation signal from dwarf galaxies.

We computed the Sommerfeld-enhanced annihilation rate in a two-level SIDM model, following our previous work in Ref. [28]. We show that even though the tree-level annihilation process is p-wave suppressed due to the Majorana nature of the DM, the nonperturbative Sommerfeld effect enhances the rate by ~6 orders of magnitude. At the same

time, the s-wave coannihilation is suppressed. This selective Sommerfeld enhancement/suppression was explained to be arising due to a particle exchange symmetry, and occurs only in multilevel DM models.

The dominating p-wave annihilation rate has a unique dependence on DM velocity that leads to a lower annihilation rate in smaller objects, like dwarf galaxies, but a large rate $\sigma v \sim 10^{-24}$ cm^3 s^{-1} in MW-sized galaxies. This can naturally explain that perhaps the positron flux excess in the cosmic ray observed by AMS-02 and several other experiments is due to the decay of the DM annihilation products into the charged leptons through a higher-dimensional operator. We investigated this possibility and found a reasonably good match between the theoretical and observed spectra.

This is the first work, to the best of our knowledge, interpreting the positron excess as a result of DM annihilation without violating the bounds from dwarf galaxy and CMB observations. One way to test this model is to search for gamma-ray lines in, e.g., H.E.S.S. and Fermi-LAT data, originating as final state radiation from the electron and muon. Bremsstrahlung radiation typically shows up as a peak in the gamma-ray spectrum around DM mass. However, in the present scenario DM annihilation products ρ/η cannot emit a photon. Only the $e^\pm/\mu^\pm$, to which the scalars decay into, can emit bremsstrahlung photon. Hence the resulting gamma-ray spectral shape is expected to be different from the usual case. A search for such gamma-ray spectrum can be done from the H.E.S.S. and Fermi-LAT data [55,56]. Nonobservation of such gamma-ray emission from dwarf galaxies could serve as an evidence for this model. We intend to do such analysis in a future work. Moreover, the high-energy charged leptons in the decay cascade can also inverse Compton scatter with photons to produce additional signature [57]. The self-scattering cross section in such a DM model turns out to be inadequate to explain the existence of the cored density profiles of the DM halos.

Even though the positron excess has been seen by both Fermi-LAT and AMS-02, there exists tension between their data in the region (20–200) GeV. The reason behind this is not known. Future independent cosmic ray experiments with improved statistics and systematics, and more precise directional flux measurements, will pin down the source of the excess.

ACKNOWLEDGMENTS

APPENDIX A: NONRELATIVISTIC FOUR-FERMI OPERATORS

In this section, we briefly review the two-spinor operators that may arise from the Yukawa interaction between a Dirac fermion ψ of mass m and a scalar ϕ or a vector A_μ,

$$\mathcal{L} \supset y\phi\bar{\psi}\psi \quad \text{or} \quad \mathcal{L} \supset gA^\mu\bar{\psi}\gamma_\mu\psi. \tag{A1}$$

As mentioned before, we use the two-component spinor notation for the fermions,

$$u(\mathbf{p}) = \sqrt{\frac{E+m}{2E}} \begin{pmatrix} \xi \\ \frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{E+m}\xi \end{pmatrix},$$

$$v(-\mathbf{p}) = \sqrt{\frac{E+m}{2E}} \begin{pmatrix} \frac{-\boldsymbol{\sigma}\cdot\mathbf{p}}{E+m}\eta \\ \eta \end{pmatrix}, \tag{A2}$$

with $E = \sqrt{\mathbf{p}^2 + m^2} \simeq m + mv^2/2$ in the nonrelativistic limit. With these definitions, all possible bispinor contractions $\bar{v}(-\mathbf{p})\Gamma u(\mathbf{p})$ reduce to

$$\bar{v}(-\mathbf{p})u(\mathbf{p}) = -\frac{1}{m}\eta^\dagger\mathbf{p}\cdot\boldsymbol{\sigma}\xi,$$

$$\bar{v}(-\mathbf{p})\gamma^0 u(\mathbf{p}) = 0,$$

$$\bar{v}(-\mathbf{p})\gamma^0\gamma^5 u(\mathbf{p}) = \left(1 - \frac{\mathbf{p}^2}{2m^2}\right)\eta^\dagger\xi,$$

$$\bar{v}(-\mathbf{p})\gamma^i u(\mathbf{p}) = \eta^\dagger\boldsymbol{\sigma}\xi - \frac{\mathbf{p}^i}{2m^2}\eta^\dagger\mathbf{p}\cdot\boldsymbol{\sigma}\xi,$$

$$\bar{v}(-\mathbf{p})\gamma^0\gamma^i u(\mathbf{p}) = -\left(1 - \frac{\mathbf{p}^2}{m^2}\right)\eta^\dagger\sigma^i\xi - \frac{\mathbf{p}^i}{2m^2}\eta^\dagger\mathbf{p}\cdot\boldsymbol{\sigma}\xi,$$

$$\bar{v}(-\mathbf{p})\gamma^i\gamma^5 u(\mathbf{p}) = \frac{i}{m}\eta^\dagger(\mathbf{p}\times\boldsymbol{\sigma})^i\xi,$$

$$\bar{v}(-\mathbf{p})\gamma^5 u(\mathbf{p}) = -\eta^\dagger\xi. \tag{A3}$$

The two bispinors $f(\mathbf{p})\eta^\dagger\xi$ and $g(\mathbf{p})\eta^\dagger\boldsymbol{\sigma}\xi$ are the scalar (spin $s = 0$) and vector (spin $s = 1$) combinations, respectively. All expressions are approximated to the order $\mathcal{O}((|\mathbf{p}|/m)^3)$. In the four-Fermi operators, two bispinors are combined together. The orbital angular momentum of such combinations is dictated by the rotational symmetry property of individual operator. A few relevant examples are listed in Table II. Here we used the notations $\mathbf{v} = \mathbf{p}/m$, $(\xi'^\dagger\eta')(\eta^\dagger\xi) = \mathbf{1}\otimes\mathbf{1}$, $(\xi'^\dagger\sigma^i\eta')\cdot(\eta^\dagger\sigma^i\xi) = \sigma^i\otimes\sigma^i$, $(\xi'^\dagger\mathbf{v}'\cdot\boldsymbol{\sigma}\eta')(\eta^\dagger\mathbf{v}\cdot\boldsymbol{\sigma}\xi) = \mathbf{v}'\cdot\boldsymbol{\sigma}\otimes\mathbf{v}\cdot\boldsymbol{\sigma}$, and $(\xi'^\dagger\mathbf{v}\cdot\boldsymbol{\sigma}\eta')\times(\eta^\dagger\mathbf{v}'\cdot\boldsymbol{\sigma}\xi) = \mathbf{v}\cdot\boldsymbol{\sigma}\otimes\mathbf{v}'\cdot\boldsymbol{\sigma}$.

TABLE II. Symmetries of the four-Fermi operators.

Operator	$^{2s+1}\ell_J$
$\mathbf{1}\otimes\mathbf{1}$	$f(^1S_0)$
$\sigma^i\otimes\sigma^i$	$f(^3S_1)$
$v^2\mathbf{1}\otimes\mathbf{1}$	$h(^1S_0)$
$\mathbf{v}'\cdot\mathbf{v}\mathbf{1}\otimes\mathbf{1}$	$g(^1P_1)$
$\mathbf{v}'\cdot\mathbf{v}\sigma^i\otimes\sigma^i$	$\frac{1}{2}(g(^3P_2) + g(^3P_1))$
$\mathbf{v}'\cdot\boldsymbol{\sigma}\otimes\mathbf{v}\cdot\boldsymbol{\sigma}$	$\frac{1}{3}(g(^3P_0) - g(^3P_2))$
$\mathbf{v}\cdot\boldsymbol{\sigma}\otimes\mathbf{v}'\cdot\boldsymbol{\sigma}$	$\frac{1}{2}(g(^3P_2) - g(^3P_1))$

To find the annihilation matrices in Eq. (14) for a particular process $|a\rangle \to X_A X_B$ where $|a\rangle = \{|\chi_1\chi_1\rangle, |\chi_2\chi_2\rangle\}$ for annihilation, and $\{|\chi_1\chi_2\rangle, |\chi_2\chi_1\rangle\}$ for coannihilation, we performed the following steps:-

(1) Write down the amplitude for $\Gamma_{ab} = |a\rangle \to X_A X_B \to |b\rangle$ using two-component spinors.
(2) Expand it in powers of $|\mathbf{p}|/m_\chi$ and other small numbers, like m_ρ/m_χ, Δ/m_χ.
(3) Collect the terms that correspond to an operator $f(^{2s+1}\ell_J)$.

APPENDIX B: SOMMERFELD FACTOR

The effect of the potential is contained in the phase shifts of the wave functions of the incoming particles, which can be found by solving the Schrödinger equation [37]. The full wave function can be expanded in the partial wave basis as

$$\Psi(\mathbf{r})_{ij} = \sum_\ell \frac{u_\ell(r)_{ia}}{r} A_{aj} P_\ell(\cos\theta). \tag{B1}$$

Here the expansion coefficients A_{aj} are understood to have an implicit partial wave ℓ index. The radial part of the wave function obeys the equation

$$\left[\left(\frac{d^2}{dr^2} + k^2 - \frac{\ell(\ell+1)}{r^2}\right)\delta_{ij} - 2\mu V(r)_{ij}\right]u_\ell(r)_{jk} = 0. \tag{B2}$$

The asymptotic form of the regular solutions of the above equation is

$$u_\ell(r)_{ij} \overset{r\to\infty}{=} N_{ij}\sin\left(k_i r - \frac{\ell\pi}{2} + s_{ij}\right), \tag{B3}$$

which we compare with the large radius asymptotic form of the total wave function $\Psi(\mathbf{r})$ in Eq. (B1) to yield

$$A_{ij} = i^\ell(2\ell+1)\frac{[M^{-1}]_{ij}}{k_i}, \tag{B4}$$

where we used $M_{ij} = N_{ij}e^{-is_{ij}}$. Here s_{ij} denote the phase shifts due to the scattering in the respective channel. In the last two equations, no summation is implied over any

repeated index. To compute the Sommerfeld factor, we need the value and the radial derivative of the wave function at the origin. For this purpose, we series expand the reduced wave function $u_\ell(r)$ as

$$u_\ell(r) = \frac{1}{(\ell+1)!} \frac{d^{\ell+1} u_\ell(r)}{dr^{\ell+1}}\bigg|_{r=0} r^{\ell+1} + \cdots \quad \text{(B5)}$$

for all four elements. To use this form of the wave function near the origin, one needs to know the phase shift matrix M. Instead we use the r-independence of the Wronskian of the linearly independent solutions of the Schrödinger equation. We compute the Wronskian both at $r \to 0$ and $r \to \infty$ limits and compare them to obtain the M matrix in terms of the amplitude of the wave functions at infinity. The irregular set of solutions to Eq. (B2), defined as $v(r)_{ij}$, has the asymptotic behavior

$$v(r)_{ij} \stackrel{r \to 0}{=} \delta_{ij} r^{-\ell}, \qquad v(r)_{ij} \stackrel{r \to \infty}{=} T_{ij}^\dagger e^{-ik_i r}. \quad \text{(B6)}$$

The Wronskian is defined as

$$W_\ell(r)_{ij} \equiv v_\ell^\dagger(r)_{ik} u_\ell'(r)_{kj} - v_\ell^{\dagger\prime}(r)_{ik} u_\ell(r)_{kj}. \quad \text{(B7)}$$

Using the asymptotic forms of the solutions, we find

$$W_\ell(r)_{ij} \stackrel{r \to 0}{=} \frac{2\ell+1}{(\ell+1)!} \frac{d^{\ell+1} u_\ell(r)}{dr^{\ell+1}}\bigg|_{r=0}, \quad \text{(B8)}$$

$$W_\ell(r)_{ij} \stackrel{r \to \infty}{=} i^\ell \sum_a k_a T_{ia} M_{aj}. \quad \text{(B9)}$$

Now equating these two quantities yields the phase shift matrix M in terms of the large-r amplitude T, which used in the full solution of Eq. (15) gives the value of the wave function at origin in terms of T,

$$\Psi_\ell(0)_{ij} = (2\ell-1)!! i^{-\ell} \frac{T_{ij}^\dagger}{k_i^\ell}. \quad \text{(B10)}$$

Using this expression for the wave function at the origin in the definition of Sommerfeld factor yields Eq. (18). The T matrix is computed by solving the Schrödinger equation numerically. A brief algorithm is provided below, following Ref. [37].

(1) Equation (B2) is solved in between $r = r_0$ and r_f. The point r_0 is chosen such that the centrifugal term dominates over the potential and Eq. (B5) is valid. The initial conditions are as follows:

$$u_\ell(r_0)_{ij} = \frac{r_0^{\ell+1}}{2\ell+1} \delta_{ij},$$
$$u_\ell'(r_0)_{ij} = \frac{(\ell+1)r_0^\ell}{2\ell+1} \delta_{ij}. \quad \text{(B11)}$$

With this choice of the boundary conditions, the Wronskian turns out to be exactly unity from Eq. (B7).

(2) At $r = r_f$, the Wronskian is written as below, using Eqs. (B6) and (B7).

$$(W_\ell)_{ij} = T_{ia}[u_\ell'(r_f)_{aj} - ik_i u_\ell(r_f)_{aj}]e^{ik_i r_f}$$
$$= \delta_{ij}. \quad \text{(B12)}$$

(3) The T matrix obtained by inverting the B matrix is defined below,

$$T = B^{-1},$$
$$B_{ij} \equiv [u_\ell'(r_f)_{ij} - ik_i u_\ell(r_f)_{ij}]e^{ik_i r_f}. \quad \text{(B13)}$$

This procedure works well when the heavier annihilation channel $|\chi_2 \chi_2\rangle$ is kinematically open. When that is not the case, the wave functions $\langle \chi_2 \chi_2 | \chi_1 \chi_1 \rangle$ or $\langle \chi_2 \chi_2 | \chi_2 \chi_2 \rangle$ are exponentially growing/decaying. The matrix inversion in Eq. (B13) with those solutions becomes difficult and gives rise to numerical instabilities. To mitigate this problem, we have followed the modified variable phase method as prescribed in Ref. [39].

APPENDIX C: EQUIVALENT ONE-LEVEL SYSTEM

The two-level Schrödinger equations in Eq. (15) are restated below for readers' convenience.

$$[\mathcal{D}_{ab} + V_{ab}(r)](u_\ell(r))_{bc} = 0. \quad \text{(C1)}$$

Note that the indices a, b, c run over the two-body states $|\chi_1 \chi_1\rangle$, $|\chi_2 \chi_2\rangle$ for annihilation, and $|\chi_1 \chi_2\rangle$, $|\chi_2 \chi_1\rangle$ for coannihilation. Let us explicitly write down the first equation (11-component) from Eq. (C1),

$$(\mathcal{D}_{11} + V_{11})u_{11} + V_{12}u_{21} = 0, \quad \text{(C2)}$$

where we have suppressed all other indices except that of two-body states to avoid clutter. Now using the particle exchange symmetry in the annihilation subspace,

$$u_{21} = \langle \chi_2 \chi_2 | \chi_1 \chi_1 \rangle \approx (-1)^{\ell+s} \langle \chi_1 \chi_1 | \chi_1 \chi_1 \rangle$$
$$= (-1)^{\ell+s} u_{11}. \quad \text{(C3)}$$

This equation relates the transition amplitude u_{21} to u_{11} through the factor $(-1)^{\ell+s}$. Using it to substitute u_{21} in Eq. (C2) gives

$$[\mathcal{D}_{11} + V_{\text{eff}}]u_{11} = 0, \quad \text{(C4)}$$

which is the equivalent one-level Schrödinger equation with the effective potential given by

$$V_{\text{eff}} = V_{11} + (-1)^{\ell+s} V_{12}. \quad \text{(C5)}$$

A similar derivation is applicable for coannihilation as well.

[1] O. Adriani *et al.* (PAMELA Collaboration), Phys. Rev. Lett. **111,** 081102 (2013).

[2] M. Ackermann *et al.* (Fermi-LAT Collaboration), Phys. Rev. Lett. **108,** 011103 (2012).

[3] C. Grimani *et al.*, Astron. Astrophys. **392,** 287 (2002).

[4] M. Boezio *et al.*, Adv. Space Res. **27,** 669 (2001).

[5] M. Aguilar *et al.* (AMS 01 Collaboration), Phys. Lett. B **646,** 145 (2007).

[6] M. A. DuVernois *et al.*, Astrophys. J. **559,** 296 (2001).

[7] M. Aguilar *et al.* (AMS Collaboration), Phys. Rev. Lett. **122,** 041102 (2019).

[8] R. Trotta, G. Jóhannesson, I. V. Moskalenko, T. A. Porter, R. R. de Austri, and A. W. Strong, Astrophys. J. **729,** 106 (2011).

[9] T. Delahaye, J. Lavalle, R. Lineros, F. Donato, and N. Fornengo, Astron. Astrophys. **524,** A51 (2010).

[10] D. Hooper, P. Blasi, and P. D. Serpico, J. Cosmol. Astropart. Phys. 01 (2009) 025.

[11] Q. Yuan, X.-J. Bi, G.-M. Chen, Y.-Q. Guo, S.-J. Lin, and X. Zhang, Astropart. Phys. **60,** 1 (2015).

[12] I. Cholis and D. Hooper, Phys. Rev. D **88,** 023013 (2013).

[13] A. Ibarra, A. S. Lamperstorfer, and J. Silk, Phys. Rev. D **89,** 063539 (2014).

[14] E. Nardi, F. Sannino, and A. Strumia, J. Cosmol. Astropart. Phys. 01 (2009) 043.

[15] Y. Farzan and M. Rajaee, J. Cosmol. Astropart. Phys. 04 (2019) 040.

[16] S. Profumo, F. Queiroz, and C. Siqueira, arXiv:1903.07638.

[17] M. Aguilar *et al.* (AMS Collaboration), Phys. Rev. Lett. **110,** 141102 (2013).

[18] M. Ackermann *et al.* (Fermi-LAT Collaboration), Phys. Rev. Lett. **115,** 231301 (2015).

[19] N. Aghanim *et al.* (Planck Collaboration), arXiv:1807.06209.

[20] H. Liu, T. R. Slatyer, and J. Zavala, Phys. Rev. D **94,** 063507 (2016).

[21] M. Kawasaki, K. Nakayama, and T. Sekiguchi, Phys. Lett. B **756,** 212 (2016).

[22] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, Phys. Rev. D **84,** 027302 (2011).

[23] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, Phys. Rev. D **80,** 043526 (2009).

[24] G. Steigman, B. Dasgupta, and J. F. Beacom, Phys. Rev. D **86,** 023506 (2012).

[25] D. Hooper, J. E. Taylor, and J. Silk, Phys. Rev. D **69,** 103509 (2004).

[26] D. Hooper, I. Cholis, T. Linden, and K. Fang, Phys. Rev. D **96,** 103013 (2017).

[27] S.-Q. Xi, R.-Y. Liu, Z.-Q. Huang, K. Fang, and X.-Y. Wang, Astrophys. J. **878,** 104 (2019).

[28] A. Das and B. Dasgupta, Phys. Rev. Lett. **118,** 251101 (2017).

[29] S. Weinberg, Phys. Rev. Lett. **110,** 241301 (2013).

[30] X. Chu and B. Dasgupta, Phys. Rev. Lett. **113,** 161301 (2014).

[31] C. Garcia-Cely, A. Ibarra, and E. Molinaro, J. Cosmol. Astropart. Phys. 11 (2013) 061.

[32] S. Garrison-Kimmel, M. Boylan-Kolchin, J. Bullock, and K. Lee, Mon. Not. R. Astron. Soc. **438,** 2578 (2014).

[33] T. Bringmann and S. Hofmann, J. Cosmol. Astropart. Phys. 04 (2007) 016; 03 (2016) E02.

[34] G. T. Bodwin, E. Braaten, and G. P. Lepage, Phys. Rev. D **51,** 1125 (1995); **55,** 5853(E) (1997).

[35] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, Reading, MA, 1995).

[36] V. Shtabovenko, R. Mertig, and F. Orellana, Comput. Phys. Commun. **207,** 432 (2016).

[37] T. R. Slatyer, J. Cosmol. Astropart. Phys. 02 (2010) 028.

[38] S. N. Ershov, J. S. Vaagen, and M. V. Zhukov, Phys. Rev. C **84,** 064308 (2011).

[39] M. Beneke, C. Hellmann, and P. Ruiz-Femenia, J. High Energy Phys. 05 (2015) 115.

[40] S. Cassel, J. Phys. G **37,** 105009 (2010).

[41] M. Cirelli, G. Corcella, A. Hektor, G. Hutsi, M. Kadastik, P. Panci, M. Raidal, F. Sala, and A. Strumia, J. Cosmol. Astropart. Phys. 03 (2011) 051; 10 (2012) E01.

[42] M. Boudaud, M. Cirelli, G. Giesen, and P. Salati, J. Cosmol. Astropart. Phys. 05 (2015) 013.

[43] K. K. Boddy, J. Kumar, and L. E. Strigari, Phys. Rev. D **98,** 063012 (2018).

[44] P. J. Mohr, D. B. Newell, and B. N. Taylor, Rev. Mod. Phys. **88,** 035009 (2016).

[45] M. Aaboud *et al.* (ATLAS Collaboration), Phys. Rev. Lett. **119,** 051802 (2017).

[46] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **122,** 021801 (2019).

[47] D. N. Spergel and P. J. Steinhardt, Phys. Rev. Lett. **84,** 3760 (2000).

[48] A. Loeb and N. Weiner, Phys. Rev. Lett. **106,** 171302 (2011).

[49] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, Astrophys. J. **648,** L109 (2006).

[50] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg, and P. Walia, Phys. Rev. Lett. **118,** 141802 (2017).

[51] K. K. Boddy, M. Kaplinghat, A. Kwa, and A. H. G. Peter, Phys. Rev. D **94,** 123017 (2016).

[52] A. Das and B. Dasgupta, Phys. Rev. D **97,** 023002 (2018).

[53] K. Todoroki and M. V. Medvedev, Mon. Not. R. Astron. Soc. **483,** 4004 (2019).

[54] M. Vogelsberger, J. Zavala, K. Schutz, and T. R. Slatyer, Mon. Not. R. Astron. Soc. **484,** 5437 (2019).

[55] H. Abdallah *et al.* (HESS Collaboration), Phys. Rev. Lett. **120,** 201101 (2018).

[56] S. Li, Z.-Q. Xia, Y.-F. Liang, K.-K. Duan, Z.-Q. Shen, X. Li, L. Feng, Q. Yuan, Y.-Z. Fan, and J. Chang, Phys. Rev. D **99,** 123519 (2019).

[57] A. V. Belikov and D. Hooper, Phys. Rev. D **81,** 043505 (2010).