PHYSICAL REVIEW D **83,** 083511 (2011)

Metastable dark matter mechanisms for INTEGRAL 511 keV γ rays and DAMA/CoGeNT events

James M. Cline,* Andrew R. Frey,† and Fang Chen‡

Physics Department, McGill University, 3600 University Street, Montréal, Québec, Canada H3A 2T8

(Received 19 August 2010; published 14 April 2011)

We explore dark matter mechanisms that can simultaneously explain the galactic 511 keV gamma rays observed by INTEGRAL/SPI, the DAMA/LIBRA annual modulation, and the excess of low-recoil dark matter candidates observed by CoGeNT. It requires three nearly degenerate states of dark matter in the 4–7 GeV mass range, with splittings, respectively, of order MeV and a few keV. The top two states have the small mass gap and transitions between them, either exothermic or endothermic, and can account for direct detections. Decays from one of the top states to the ground state produce low-energy positrons in the Galaxy whose associated 511 keV gamma rays are seen by INTEGRAL. This decay can happen spontaneously, if the excited state is metastable (longer lived than the age of the Universe), or it can be triggered by inelastic scattering of the metastable states into the shorter-lived ones. We focus on a simple model where the dark matter is a triplet of an $SU(2)$ hidden sector gauge symmetry, broken at the scale of a few GeV, giving masses of order $\lesssim 1$ GeV to the dark gauge bosons, which mix kinetically with the standard model hypercharge. The purely decaying scenario can give the observed angular dependence of the 511 keV signal with no positron diffusion, while the inelastic scattering mechanism requires transport of the positrons over distances ~ 1 kpc before annihilating. We note that an x-ray line of several keV in energy, due to single-photon decays involving the top dark matter states, could provide an additional component to the diffuse x-ray background. The model is testable by proposed low-energy fixed-target experiments.

DOI: 10.1103/PhysRevD.83.083511 PACS numbers: 98.80.Cq, 12.60.Cn, 95.35.+d, 98.70.Rz

I. INTRODUCTION

Annihilation of positrons near the galactic center gives rise to a narrow 511 keV gamma-ray line that was first observed in 1972 [1], and which has been confirmed by numerous experiments since then, most recently by the SPI spectrometer aboard the INTEGRAL satellite [2]. The signal has two distinct components, one associated with the central region of the Galaxy (bulge) and another with the disk. There is as yet no consensus as to a conventional astrophysical origin for these gamma rays [3–5], which evidently originate from excess positrons annihilating nearly at rest. The apparent axial symmetry of the bulge component is a point in favor for proposals of models of dark matter (DM) that decays or annihilates into low-energy positrons, since DM should be distributed symmetrically near the galactic center.[1] However early proposals of this sort were driven toward DM candidates that were nearly as light as the electron itself [7], since the injection energy of the positrons can be no greater than a few MeV [8] (see however [9]). Models of MeV-scale dark matter that couples to e^+e^- are highly constrained by low-energy collider data, and are not (in our opinion) theoretically attractive. If dark matter is the source of 511 keV gamma rays, one will need to verify its properties by direct detection or other complementary means to make the explanation of the 511 keV signal convincing. For a recent and thorough review of the 511 keV signal and an evaluation of possible sources including DM, see [10].

In this work we propose and revisit scenarios in which a long-lived excited state of DM with mass at the 10 GeV scale can scatter into a nearby unstable state, whose mass differs by only a few keV. The unstable state decays into the ground state with the emission of a low-energy e^+e^- pair. The decay can be relatively fast, but the energy is only released after the slow process of inelastic DM-DM collisions occurs. (See however the purely decaying variant described below.) There are two qualitatively different ways to realize this, depending upon whether the metastable state is the middle one, requiring endothermic scattering, or the top one, leading to exothermic. The mass spectra and sequence of transitions are sketched in Fig. 1. The endothermic version, in the context of 500–1000 GeV DM, was first presented in [11] to try to simultaneously explain INTEGRAL, PAMELA (Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics) [12], and ATIC (Advanced Thin Ionization Calorimeter) [13] excess electron observations. It was subsequently discussed with applications to direct DM detection in [14], for ~ 100 keV values of the small mass splitting.

*fangchen@physics.mcgill.ca
†jcline@physics.mcgill.ca
‡frey@physics.mcgill.ca

[1]In addition, the east-west asymmetry in the disk component claimed by [5] is not confirmed by the more recent analysis of [6], using more accumulated data from INTEGRAL.

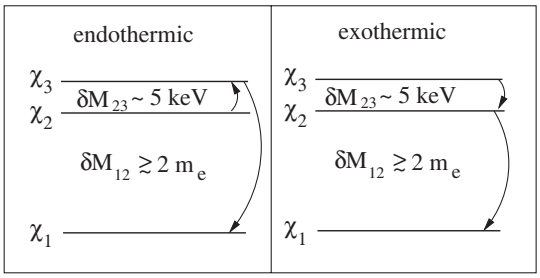

FIG. 1. Spectrum of states for metastable dark matter models. Curves with arrows indicate the sequence of transitions for the endothermic case, $\chi_2 \to \chi_3 \to \chi_1$ (left) and the exothermic one, $\chi_3 \to \chi_2 \to \chi_1$ (right).

The viability of the scenario for INTEGRAL was further explored in [15], but only in the heavy DM regime.[2]

It would clearly be interesting if the DM mechanism for the INTEGRAL observations was somehow tied to direct detection of the DM [14,18]. Our exothermic mechanism is partly motivated by Ref. [19], which proposed a model involving only the states χ_2 and χ_3 (in our notation), as a means of explaining two indications of direct detection of dark matter, namely, the long-standing DAMA/LIBRA annual modulation [20], and the more recent observation of excess low-recoil events by the CoGeNT (Coherent Germanium Neutrino Technology) experiment [21]. Reference [19] showed that DM with a mass of $\cong 4$ GeV and mass splitting of a few keV could be consistent with these observations, using the exothermic nuclear scattering $\chi_3 N \to \chi_2 N'$. Their observation is that the shape of recoil spectrum is sensitive to modulations of the local DM velocity when the scattering is exothermic, and this can explain the DAMA observations. Additionally the overall rate for the same parameters is correct for explaining the excess CoGeNT events. The idea of Ref. [19] is related to the inelastic dark matter proposal [22], which however assumed the scatterings to be endothermic rather than exothermic. (See [23] for another discussion of exothermic scatterings.)

There have been several proposals for DM in the 5–10 GeV mass range to explain the DAMA and CoGeNT observations [24]. Most recently, Ref. [25] showed that elastic DM interactions could simultaneously explain the DAMA/LIBRA and CoGeNT observations if the dark matter mass is near 7 GeV and the cross section on nucleons around 2×10^{-40} cm^2. We will argue that our endothermic scenario is close enough to being elastic, if the small mass splitting δM_{23} is of order a few keV, so that the same analysis applies. Even though such a small splitting has little effect on DM-nucleus scattering, it is important for DM-DM scattering in the Galaxy, in the present case where the DM is lighter than the nuclei in the direct detectors. Getting the observed rate of galactic positrons limits the maximum mass splitting in this case, to values somewhat lower than those that would strongly affect the direct detection rates.

An interesting variant of the above mechanisms is to assume that the unstable excited state is so long lived that it still has a relic population in the Galaxy, and so does not need to be produced by DM collisions.[3] This version has more freedom, in that the rate of producing positrons (via decays into $\chi_1 e^+ e^-$) does not depend upon the small mass splitting δM_{23}, whereas the rate of inelastic scattering $\chi_3 \chi_3 \leftrightarrow \chi_2 \chi_2$ is rather sensitive to δM_{23}.

Our proposals fit nicely within the framework of dark matter with a non-Abelian gauge symmetry in a hidden sector, as suggested by [18], since such DM automatically consists of multiple states, and small mass splittings are radiatively generated if the gauge symmetry is spontaneously broken. The simplest example that contains three DM states is a hidden $SU(2)$ gauge sector, where the DM is in the triplet representation. After the hidden $SU(2)$ breaks, two colors of the dark gauge boson must acquire small kinetic mixing $\epsilon_i \sim 10^{-3}$–10^{-6} with the standard model (SM) hypercharge Y,

$$\mathcal{L}_{\rm mix} = \sum_i \epsilon_i B_i^{\mu\nu} Y_{\mu\nu}, \tag{1}$$

while the remaining one must have negligible mixing to keep the long-lived state stable against decays to $\chi_1 e^+ e^-$. The mixing B_i's couple weakly to charged standard model particles, and mediate the scatterings with nucleons or decays into $e^+ e^-$.[4] We find that the hidden gauge symmetry should break at the 10 GeV scale (resulting in gauge boson masses of order several hundred MeV) to give the right cross sections for DM scattering in the galactic center and in detectors. An attractive feature of these hidden sector gauge boson masses and couplings is that they are in the right range to be directly probed by new proposed fixed-target experiments [27].

We will present our main results first, in Sec. II. The remaining parts of the paper supply the many details leading to these results. Section III specifies the hidden sector $SU(2)$ particle physics models we consider. The gauge coupling α_g of this $SU(2)$ is calculated in Sec. IV by the requirement of getting the right thermal relic density for the DM. There we also work out the crucial relative abundances of the excited states. Section V describes how the rate and angular distribution of 511 keV gamma rays are computed. Here we also summarize what is believed about

[2]The original idea of excited dark matter assumed that only the ground state was significantly populated, so that excitation through the $\gtrsim$ MeV mass gap must occur in galactic inelastic collisions [16]. However more detailed computations showed that the collision rate is not high enough with such a large energy barrier to overcome [11,15,17].

[3]Earlier work on decaying DM as the source of 511 keV gamma rays can be found in [17,26].

[4]Except in Eq. (11), we will rescale the ϵ_i to be the mixing parameter of B_i with the photon for notational convenience.

the location of gaseous media in the galactic bulge where positron annihilation is supposed to take place, in response to criticisms of DM interpretations of the INTEGRAL observations in Ref. [4]. In Sec. VI we explain how the gauge kinetic mixing parameter ϵ_1 is constrained to get the desired rates for DAMA. Various astrophysical constraints are addressed in Sec. VII. Our predictions for the masses and couplings of the gauge boson B_1 that mediates the DAMA and CoGeNT reactions, relevant for direct laboratory searches, are presented in Sec. VIII. We conclude in IX. The appendices give further details about the kinetic equilibrium of the DM with the SM, and the cross section for DM annihilation.

II. MAIN RESULTS

In this section we summarize our main results. The details leading up to them will be given in subsequent sections. The relevant parameter space is the average DM mass M_χ, the mass splitting δM_{23}, and the masses μ_i of the hidden sector gauge bosons B_i that mediate the interactions with the standard model. We fix the larger mass splitting to be $\delta M_{12} = 1.1$ MeV so that there is sufficient phase space for the decay into $\chi_1 e^+ e^-$ while insuring that the decay products are not very relativistic, as required by constraints on the injection energy of the low-energy positrons [8]. Larger values of δM_{12} tend to suppress the positron rate, and the direct detection rate for exothermic DM, because of greater depletion of the excited state abundance, but our results are not greatly sensitive to the exact value so long as it is less than a few MeV.

Because the non-Abelian $SU(2)$ gauge interactions take the form

$$g\bar{\chi}_1 B_2 \chi_3 + \text{cyclic permutations}, \tag{2}$$

B_1 mediates the transition $\chi_3 \to \chi_2$, *et cyc*. The strength of the gauge coupling g is fixed by the requirement of getting the observed relic density of DM from thermal freeze-out,

$$\alpha_g \cong c_g \left(1 - \frac{\bar{\mu}^2}{M_\chi^2}\right)^{-1/4} \frac{M_\chi}{\text{GeV}}, \tag{3}$$

where $\alpha_g = g^2/4\pi$ and $\bar{\mu}$ is the average mass of the gauge bosons. The value of the constant c_g depends upon the number of hidden sector Higgs bosons that can be present in the final state of $\chi\chi \to HH$ annihilations; it can lie in the range $c_g \cong (1.7 - 2.5) \times 10^{-5}$ for the scenarios we consider. In the following, we assume the dark Higgs bosons are heavier than the DM, which yields the top value in this range, hence a larger rate of positron production. The value of c_g is derived in Sec. IV.

A. Exothermic dark matter

We first consider the exothermic case where χ_3 is the stable excited state. In Fig. 2 we plot contours of $\log R_{e^+}/R_{\text{obs}}$, the predicted rate of positron production at the galactic center versus the measured rate, in the M_χ-δM_{23} plane, where δM_{23} is the small mass splitting between DM states 2 and 3. The contours are superimposed upon the DAMA/LIBRA allowed region of Ref. [19]. To illustrate the dependence on the gauge boson masses, each graph has a different value of μ_2, the mass of B_2. In the exothermic case, B_2 is the color that has negligible mixing with the SM (to avoid $\chi_3 \to \chi_1$ decays), and so μ_2 does not directly affect the rates of either direct detection nor galactic positron production. However, the class of models we describe in Sec. III predicts relations between μ_2 and the other gauge boson masses,

$$\mu_1 \gtrsim \sqrt{\mu_3^2 - \mu_2^2}, \tag{4}$$

$$\mu_3 = \frac{2}{\alpha_g}\delta M_{23} + \mu_2. \tag{5}$$

The first condition (4) depends on details of how the Higgs mechanism in the hidden sector gives masses to the gauge bosons; we take the inequality to be saturated, which helps

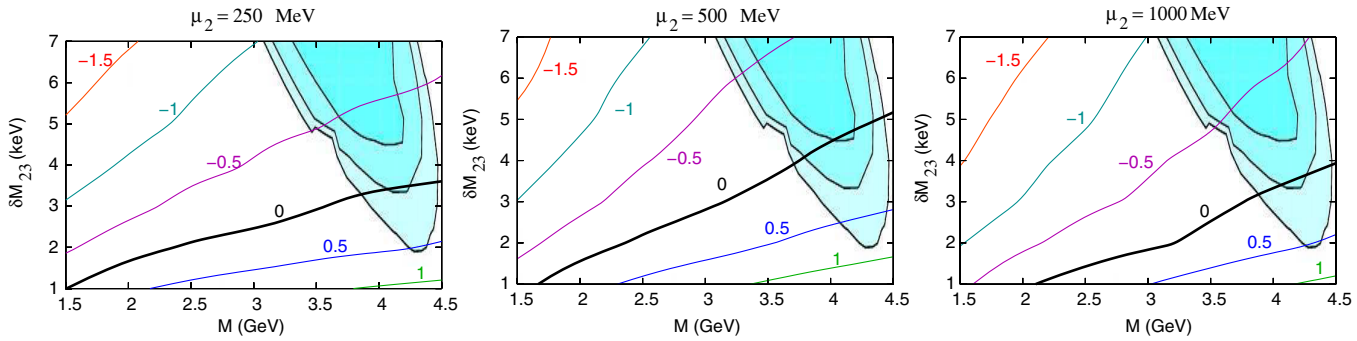

FIG. 2 (color online). Solid curves: contours of $\log R_{e^+}/R_{\text{obs}}$ (the positron production rate) for exothermic dark matter, in the plane of the average DM mass M_χ and mass splitting δM_{23}. Heavy contour labeled "0" matches the observations. Shaded regions are allowed by DAMA/LIBRA, from Ref. [19]. Each plot takes a different value of gauge boson mass μ_2, with $\mu_{1,3}$ given by Eqs. (4) and (5). DM halo parameters are given by Eq. (7).

to increase the rate of $\chi_3\chi_3 \to \chi_2\chi_2$ scatterings (since B_1 is the exchanged boson). The second condition (5) arises because the mass difference $\mu_3 - \mu_2$ determines the radiatively generated splitting $\delta M_{23} = -\frac{1}{2}\alpha_g(\mu_3 - \mu_2)$.

From Fig. 2 one observes that larger values of μ_2 help to achieve a large enough rate of positron production, up to some optimal value $\mu_2 \sim 600$ MeV, beyond which the rate starts to slowly fall with μ_2. The rise at small μ_2 occurs because increasing μ_2 inhibits $\chi_3\chi_3 \to \chi_1\chi_1$ downscatterings in the early Universe, lessening the depletion of the χ_3 state. Figure 3, left panel, illustrates this more directly, where the relic abundance of the excited state Y_3 relative to that of total DM abundance $Y_{\rm tot}$ is plotted as a function of M_χ for several values of μ_2. There is a saturation $Y_3/Y_{\rm tot} \to \sim 1/3$ as μ_2 approaches the GeV scale, for the fiducial value $\delta M_{12} = 1.1$ MeV of the large mass splitting. (The right panel of Fig. 3 indicates that this saturation would occur at higher values of μ_2 if δM_{12} is increased. The rate of $\chi_3\chi_3 \to \chi_1\chi_1$ increases with δM_{12} due to the larger phase space.) Further increase of μ_2 beyond the optimal point decreases the positron rate, because μ_1 is an increasing function of μ_2, and the rate of $\chi_3\chi_3 \to \chi_2\chi_2$ transitions goes like μ_1^{-4}.

It may seem surprising that the rate of positron production is a decreasing function of the mass splitting δM_{23}, since the phase space for $\chi_3\chi_3 \to \chi_2\chi_2$ increases with δM_{23}. However, so does the exchanged momentum that appears in the propagator of the virtual gauge boson, and this has the more important effect of suppressing the amplitude; see Eq. (23).

Figure 2 shows some overlap between the desired rate of positron production and the DAMA-allowed region for $\mu_2 \gtrsim 200$ MeV. For each point in the M_χ-δM_{23} plane, we have adjusted the value of ϵ_1 to obtain the DAMA detection rate assumed by Ref. [19]. ϵ_3 is taken to be $\lesssim 10^{-3}$; the results shown are insensitive to the exact value. Concerning ϵ_3, an intriguing prediction of our model is that each direct detection of the process $\chi_3 N \to \chi_2 N'$ must be accompanied by the subsequent production of e^+e^- through the decay $\chi_2 \to \chi_1 e^+ e^-$ (whose rate scales as ϵ_3^2), so in principle one could look for the positron in coincidence. However, the lifetime for the decay cannot be much less than 10^3 s, as we will show in Sec. VII C. Since the speed of DM in the Galaxy is of order $10^{-3}c$, this occurs too far from the experiment to detect the e^+e^- pair. In fact this lifetime is much longer than the age of the Universe for $\epsilon_3 \lesssim 10^{-8}$, leading to an alternative possibility for explaining the 511 keV signal via decays of primordial χ_2, more about which in Sec. II C.

The rate of positron production through DM excitations is sensitive to the density profile ρ of the DM halo; it scales like ρ^2 evaluated near the galactic center. We parametrize the shape using the Einasto profile

$$\rho = \rho_\odot \exp\left(-\frac{2}{\alpha}((r/r_s)^\alpha - (r_s/r_\odot)^\alpha)\right). \qquad (6)$$

A set of values that are often considered to be standard are $\alpha = 0.17$, $r_s = 20$ kpc, $\rho_\odot = 0.3$ GeV/cm^3, $r_\odot = 8.5$ kpc. These values for α and r_s are based upon pure dark matter N-body simulations that do not take into account the effects of baryons in the central region of the Galaxy [28]. We do not obtain a large enough rate of positron production using these numbers. However, there is strong evidence that the halo is much more concentrated (cuspy) near the center than these values indicate, due to the presence of the baryons, which have the effect of contracting the density [29]. Table I shows the profile parameters for six Milky Way-like galaxies from the Aquarius simulation, which have been reanalyzed to include baryonic contraction [30]. Furthermore it has been argued that the local density may be larger than the canonical value by a factor of 1.3–2 [31,32]. We find that the exothermic scenario gives acceptable overlap between the INTEGRAL and DAMA-allowed regions only if we adopt

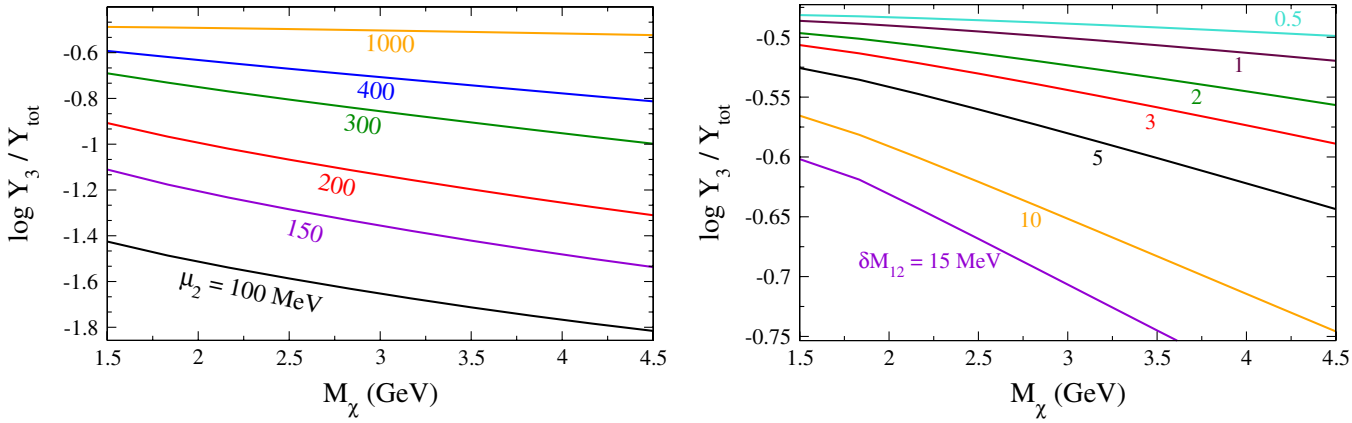

FIG. 3 (color online). Left: in the exothermic case, log of $Y_3/Y_{\rm tot}$, abundance of stable excited state χ_3 relative to the total DM abundance, as a function of M_χ, for several values of the gauge boson mass μ_2, with $\mu_{1,3}$ fixed as in Eq. (4) and (5) and $\delta M_{23} = 5$ keV. δM_{12} is fixed at 1.1 MeV. Right: same but with varying $\delta M_{12} = 0.5$–15 MeV, and fixed $\mu_2 = 1000$ MeV.

TABLE I. Characteristics of the density profiles of the haloes in the Aquarius galaxy formation simulations of Ref. [30]. Column 1 gives the name of each halo. Columns 2–3 list α, r_s, the parameters of the best fitting Einasto model, in the inner region of the Galaxy.

Galaxy	α	r_s (kpc)
Aq-A-5	0.065	5.3
Aq-B-5	0.145	15.6
Aq-C-5	0.115	10.2
Aq-D-5	0.102	14.7
Aq-E-5	0.098	11.1
Aq-F-5	0.112	15.6

a DM halo that is very cuspy and has a somewhat large density in the solar neighborhood. We take the most concentrated example in Table I,

$$\alpha = 0.065, \qquad r_s = 5.3 \text{ kpc}, \qquad \rho_\odot = 0.42 \text{ GeV/cm}^3 \tag{7}$$

to obtain Fig. 2. This could still be considered a conservative choice, since Ref. [32] argues for $\rho_\odot = 0.43(11) \times (10)$ GeV/cm^3. With these error estimates, one might reasonably consider $\rho_\odot = 0.6$ GeV/cm^3. This allows for some reduction of the cuspiness of the halo with very similar results, to $\alpha = 0.08$, $r_s = 7.5$ kpc for example. Moreover, we can still achieve reasonable consistency using the same cuspy profile while keeping $\rho_\odot = 0.42$ GeV/cm^3; see Fig. 4.

B. Endothermic dark matter

If χ_2 is the stable state, then the transitions $\chi_2 \to \chi_3$ are endothermic. The energy barrier would tend to reduce the rate of such transitions compared to the exothermic case, but there are other differences that also affect the rate. Most importantly, even though Eqs. (4) and (5) are unchanged, the roles of the gauge bosons B_2 and B_3 become interchanged relative to exothermic DM. μ_3 now controls the rate of $\chi_2\chi_2 \to \chi_1\chi_1$ downscattering in the early Universe,

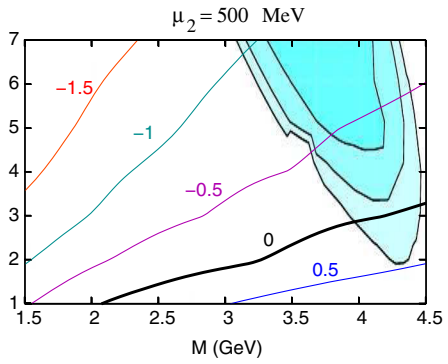

FIG. 4 (color online). As in Fig. 2, but for less cuspy Einasto profile with $\alpha = 0.08$, $r_s = 7.5$ kpc, $\rho_\odot = 0.42$ GeV/cm^3.

hence the relic density of χ_2. Because μ_3 is naturally the heaviest of the three gauge boson masses in our model (see Sec. III), this means that the endothermic scenario leads to a significantly larger rate of galactic positrons than the corresponding exothermic one. We thus adopt a less cuspy halo profile in this case, I,

$$\alpha = 0.08, \qquad r_s = 8 \text{ kpc}, \qquad \rho_\odot = 0.42 \text{ GeV/cm}^3. \tag{8}$$

Our findings for the 511 keV signal for endothermic DM are illustrated in Fig. 5, We have the freedom to choose even less cuspy profiles if desired, with some accompanying decrease in the value of δM_{23}, as shown in Fig. 6, using the more moderate parameter values $\alpha = 0.12$ and $r_s = 12$ kpc. Figure 7 shows how the relative abundance of the stable state, Y_2/Y_{tot}, depends upon the masses M_χ, μ_2 and mass splitting δM_{12}. For the examples shown, δM_{23} should be $\lesssim 4$ keV to match the direct detection rate corresponding to Fig. 8, in the allowed M_χ region that is shaded in Figs. 5 and 6.

Similarly to the exothermic case, we fix the value of ϵ_1 to get the desired direct detection rate, while ϵ_2, which controls the rate of decay $\chi_3 \to \chi_1 e^+ e^-$, is a free parameter. We assumed $\epsilon_2 = 10^{-3}$ in Fig. 5. In contrast to the exothermic case, the results are somewhat sensitive to this choice: taking much smaller values of ϵ_2 mildly suppresses the rate due to its effect on the relic abundance Y_2, as we will further discuss in Sec. IV C.

Ideally, the analysis of Ref. [25] should be redone for our slightly inelastic scattering to see how the overlap of the DAMA and CoGeNT allowed regions of Fig. 8 might be modified. (For this reason we display our results in the same range of DM masses as in Fig. 8.) We leave such an investigation to future work.

C. Decaying DM scenario

If the gauge mixing parameter ϵ_2 or ϵ_3 is sufficiently small, then the excited state χ_2 or χ_3 (depending upon whether the DM is exothermic or endothermic) which we have referred to as "unstable" can be as long lived as the Universe. Let us denote the stable and "unstable" excited states by χ_s and χ_u. Instead of being produced in $\chi_s\chi_s \to \chi_u\chi_u$ scattering, χ_u can have a significant relic density and produce e^+e^- from its slow decays. In Sec. V B we show that the correct lifetime for producing the observed rate of positrons results from taking

$$\epsilon_{2,3} \sim 10^{-11}. \tag{9}$$

The exact expression depends upon other parameters and is given by Eq. (41). In particular, the dependence upon the DM halo profile is much weaker for decays than for the inelastic scattering scenarios discussed above. We are no longer constrained to consider profiles such as (7).

It is intriguing that for reasonable choices of the halo profile, the decaying DM scenario can explain the

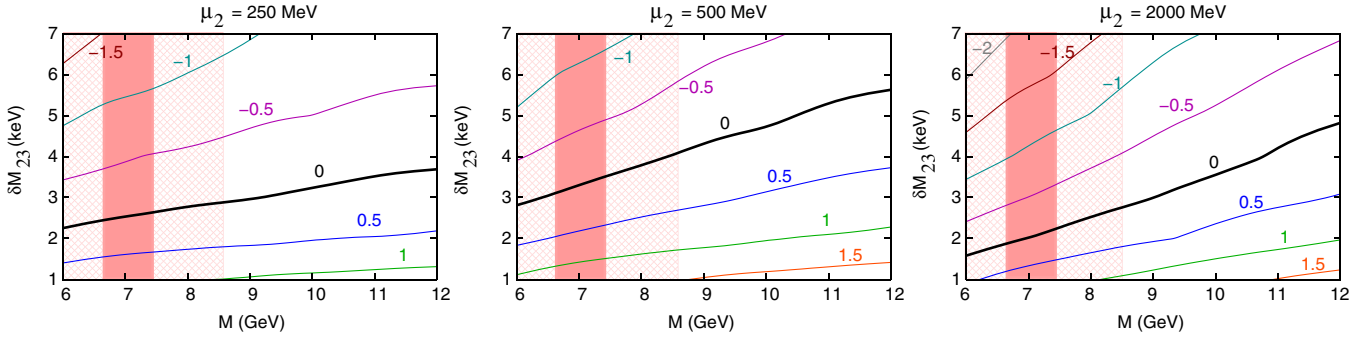

FIG. 5 (color online). Solid curves: contours of $\log R_{e^+}/R_{\rm obs}$ (the positron production rate) for endothermic dark matter, analogous to Fig. 2 for the exothermic case. Einasto halo parameters are given in (8). Columns correspond to gauge boson masses $\mu_2 = 250$, 500 and 1000 MeV, respectively. Shaded regions are 90% and 99% C.L. preferred DM masses for fitting DAMA/CoGeNT data, from Ref. [25].

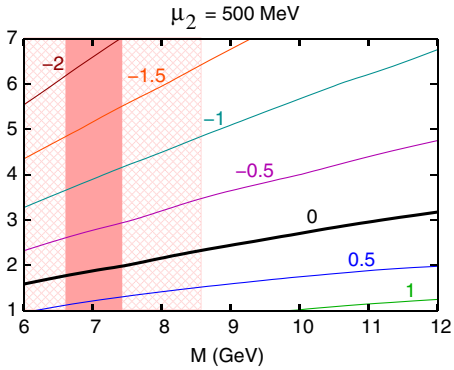

FIG. 6 (color online). As in Fig. 5, but for less cuspy Einasto profile with $\alpha = 0.12$, $r_s = 12$ kpc, $\rho_\odot = 0.42$ GeV/cm^3.

morphology of the 511 keV signal without the need to invoke propagation of the positrons before annihilation. This is in contrast to the inelastic scattering mechanisms which localize the positron production much closer to the galactic center, as we next discuss.

D. Angular profile of 511 keV signal

If positrons annihilate before propagating, the predicted intensity I_{e^+} of the 511 keV signal as a function of angle is just a reflection of the DM density profile,[5] through a line-of-sight integral, whose form depends upon whether the positrons were created through scattering or decay:

$$I_{e^+}(\hat{x}) = \int_{\rm l.o.s.} dx \begin{cases} \frac{1}{2}\langle\sigma v\rangle \frac{\rho_s^2}{M_\chi^2}, & \text{scattering} \\ \frac{\rho_u}{M_\chi \tau_s}, & \text{decay.} \end{cases} \quad (10)$$

The integral is taken along the $\hat{x}$ direction, where $\rho_{s,u}$ is the density of the stable or unstable excited DM component $\chi_{s,u}$, proportional to the total density ρ, and τ_u is the lifetime of χ_u.

For the Einasto profile (7) we considered for scatterings, ρ^2 is practically a delta function, and so the signal would appear to come from a point source. ρ^2 falls to e^{-8} of its maximum value at a distance of $r_8 = r_s(2\alpha)^{1/\alpha}$. Even for the standard profile with $\alpha = 0.17$ and $r_s = 20$ kpc, r_8 is only 35 pc, which subtends an angle of 0.2°. However the observed signal subtends at least 8° [6]; see Fig. 9. Therefore if the scattering explanation is correct, all of the observed width must be due to propagation.[6] The distance corresponding to 8° is 1.1 kpc, which may be astrophysically reasonable, depending upon the structure of the galactic magnetic field and the injection energy of the positrons.

Apart from astrophysical mechanisms of positron transport [34,35], which strongly depend upon the poorly constrained magnetic field of the inner Galaxy, our model suggests another way in which this widening could occur due to the streaming of χ_3 (χ_2 in exothermic case) before it decays. If the gauge mixing parameter for $B_{2,3}$, the hidden gauge boson mediating the $\chi_{3,2} \to \chi_1 e^+ e^-$ decay, is sufficiently small, $\epsilon_{2,3} \sim 10^{-7}$, then $\chi_{3,2}$ is so long lived that it will travel approximately 1 kpc before decaying.

For the decaying DM scenario, it is possible to fit the observed angular distribution without any smoothing from positron diffusion. Figure 9 shows several examples with Einasto parameters $\alpha = 0.11$ and $\alpha = 0.13$ that pass through all the error bars. These examples are close to the ones given in Table I, and so could be considered realistic in light of baryonic compression of the inner part of the DM halo.

[5] In the case of inelastic scattering, there is some mild r dependence of $\langle\sigma v\rangle$ due to the r dependence of the velocity dispersion, which we neglect here.

[6] Our results differ somewhat from those of Ref. [33], which assumed a less cuspy halo.

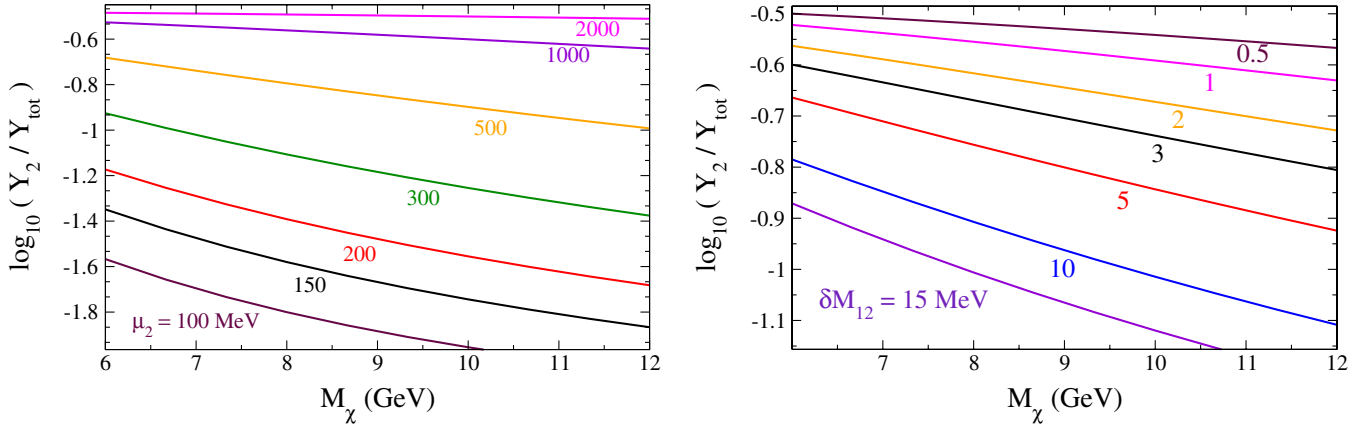

FIG. 7 (color online). Similar to Fig. 3, but for endothermic model.

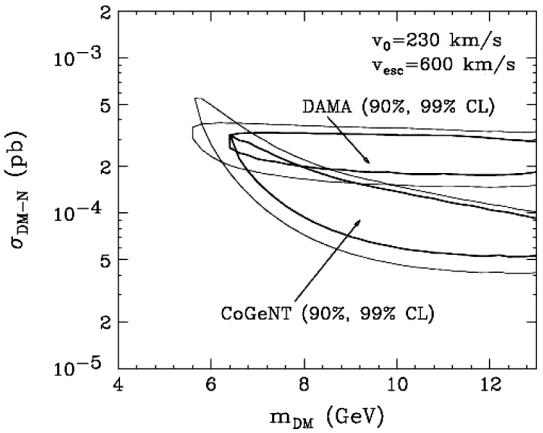

FIG. 8. Allowed regions of Ref. [25] for DM to explain DAMA/LIBRA and CoGeNT events.

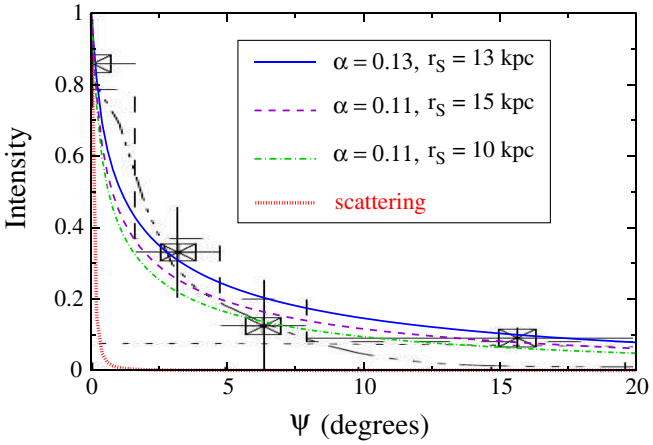

FIG. 9 (color online). Data points show angular distribution from galactic center of observed 511 keV signal, reproduced from Ref. [6]; labeled curves are predictions from DM scattering (lowest, dotted curve) and from decay (upper curves), assuming no propagation of positrons before they annihilate.

III. PARTICLE PHYSICS MODELS

The simplest example of a non-Abelian hidden sector model consistent with the observations we discuss has a dark $SU(2)$ gauge group under which the DM transforms as a triplet. The most general form of the Lagrangian that we will need is

$$\mathcal{L} = \frac{1}{2}\bar{\chi}_a(i\not{\partial}_{ab} - M_\chi \delta_{ab})\chi_b - \frac{1}{4g^2}B^a_{\mu\nu}B^{\mu\nu}_a - \sum_i \frac{1}{\Lambda_i}\Delta^{(i)}_a B^{\mu\nu}_a Y_{\mu\nu} - \frac{1}{2}y\bar{\chi}_a\Sigma_{ab}\chi_b + \mathcal{L}_{\text{Higgs}}(\Delta^{(i)}, \Sigma). \quad (11)$$

Here $\Delta^{(i)}$ and Σ are triplets and a quintuplet, respectively, of the hidden $SU(2)$. Two such triplets are needed in order to get the kinetic mixing (1) required for the direct detection signal and galactic positron production. The mixing parameters $\epsilon_i = \langle\Delta_i\rangle/\Lambda_i$ arise when the triplets acquire vacuum expectation values (VEVs). These mixing parameters lead to a coupling $e\epsilon_i \cos\theta_W$ of the electric current to B_i; in the remainder of this paper, we rescale ϵ_i to remove the Weinberg angle from this coupling. A third triplet is required to get the right pattern of mass splittings for exothermic DM. The VEV of the quintuplet Σ_{ab} gives the large $\sim$ MeV mass splitting.

We studied this class of models previously in Ref. [36]. It is convenient to take the triplet VEVs to be mutually orthogonal $\langle\Delta^{(i)}_a\rangle \equiv \delta_{ia}\Delta_a$. Reference [36] shows that it is easy to construct a potential leading to this pattern. We further restrict the traceless symmetric tensor Σ_{ab} to have VEVs only on the diagonal,

$$\langle\Sigma\rangle = \text{diag}(A - B, 2B, -A - B). \quad (12)$$

This alignment can be accomplished by suitable small interactions between Σ and the triplets. With these VEVs, the masses of the gauge bosons are given by

$$\mu_1^2 = g^2(\Delta_2^2 + \Delta_3^2 + 2(A + 3B)^2),$$
$$\mu_2^2 = g^2(\Delta_1^2 + \Delta_3^2 + 8A^2), \quad (13)$$
$$\mu_3^2 = g^2(\Delta_1^2 + \Delta_2^2 + 2(A - 3B)^2).$$

The corresponding mass shifts in the χ_a states relative to the average mass M_χ are given by

$$\delta M_1 = -\tfrac{1}{2}\alpha_g(\mu_2 + \mu_3) + y(A - B),$$
$$\delta M_2 = -\tfrac{1}{2}\alpha_g(\mu_1 + \mu_3) + y(2B), \quad (14)$$
$$\delta M_3 = -\tfrac{1}{2}\alpha_g(\mu_1 + \mu_2) + y(-A - B).$$

We have introduced the Yukawa coupling contribution in order to explain two different scales of mass splittings: $\delta M_{23} \sim$ keV, and $\Delta M_{12} \sim \delta M_{13} \sim$ MeV. This can occur if the quintuplet VEVs satisfy $A = -3B$; then the Yukawa term only contributes to the large mass splittings and not to δM_{23}. Let us assume this to be the case; we will presently show how it can come about. Then the small mass splitting comes entirely from the one-loop self-energy contribution from gauge boson exchange, $\delta M_{23} = \delta M_3 - \delta M_2 = \frac{1}{2}\alpha_g(\mu_3 - \mu_2)$. The assumed order $\delta M_3 > \delta M_2$ requires that $\mu_3 > \mu_2$, hence $\Delta_2 > \Delta_3$. The most economical choice would be to remove Δ_3 from the spectrum altogether, which is permissible if the kinetic mixing parameter ϵ_3 is allowed to vanish. In fact for the endothermic scenario, this is exactly what we want, in order to forbid $\chi_2 \to \chi_1 e^+ e^-$ decays, so that χ_2 can be stable.[7] For the exothermic case, it is opposite: we need to insure the stability of χ_3 against decays to χ_1, hence ϵ_2 must be negligible, while ϵ_3 is needed for the $\chi_2 \to \chi_1 e^+ e^-$ decays. And for both scenarios, ϵ_1 must be nonzero to enable direct detection via inelastic $\chi_{2,3}$ scattering on nucleons. The upshot is that we need all three triplets for exothermic DM (although only two of them should lead to kinetic mixing), but only two, $\Delta_{1,2}$, for endothermic.

Now we return to the question of why the quintuplet VEVs should satisfy the seemingly fine-tuned relation $A = -3B$. Interestingly, the desired VEVs can arise from the simple renormalizable potential

$$V(\Sigma) = \lambda_\Sigma(\mathrm{tr}\Sigma^2 - v^2)^2 + \mu \det\Sigma, \quad (15)$$

which has three degenerate minima at $A = \pm 3B$ and $A = 0$. (In the absence of the triplet VEVs, this would leave one of the three gauge bosons massless, breaking $SU(2) \to \mathrm{U}(1)$). We assume that it is possible to design small interactions with the triplets that align $\langle\Sigma\rangle$ along the diagonal, and which might perturb A slightly away from $-3B$. For simplicity we take $A = -3B$ in the following, so that $\delta M_{12} \cong 6yB$.

[7]See Sec. VII A, though, for further discussion of some subtleties.

Considering the smaller mass splitting, if $M_\chi \sim 5$ GeV, the gauge coupling is of order $\alpha \sim 10^{-4}$, and the difference in gauge boson masses should be of order $|\delta M_{23}|/\alpha_g \sim 100$ MeV. This is consistent with triplet VEVs at the scale of $\sim$30 GeV, since the gauge coupling is $g \sim 0.035$. To get the correct sign for the mass difference, $M_3 > M_2$ only requires that $\Delta_2 > \Delta_3$ (given our assumption $A = -3B$), so that $\mu_3 > \mu_2$.

IV. RELIC DENSITY

In Ref. [36] a first attempt was made to compute the value of the gauge coupling α_g corresponding to the observed DM density through thermal freeze-out. In this section we correct and refine that result, taking into account some subleading effects, including extra annihilation channels into dark Higgs bosons, dependence of the Born cross section on the DM velocity, and Sommerfeld enhancement [37] at the time of freeze-out. Moreover we estimate the amount of dilution of the excited states due to downscattering in the early Universe.

A. Annihilation cross section; determination of α_g

Reference [36] derived the annihilation cross section for $\chi\chi \to BB$ by separately considering $\chi_1\chi_1 \to B_jB_j$ (with $j = 2, 3$) and $\chi_1\chi_j \to B_1B_j$ (again with $j = 2, 3$), and explicitly averaging over the initial state and summing over the final state colors. In this paper, we make several improvements to the previous calculation as well as minor corrections. Details of our calculation are given in Appendix A.

First, we now include the process where $\chi\chi$ goes to two hidden sector Higgs bosons through exchange of a virtual B in the s channel. Furthermore, as pointed out in [38], velocity-dependence (including Sommerfeld enhancement) may make significant contributions to the cross section in some models. Therefore, we include corrections to order v^2 in the tree-level cross section as well as the leading contribution from Sommerfeld enhancement.

Moreover, in the present application, the dark matter is sufficiently light that its annihilation may take place in the broken phase of the theory. Therefore, we have computed the cross section for $\chi\chi \to BB$ taking into account the gauge boson masses. The main effect comes simply from the reduction in phase space, which is a factor of $(1 - \mu_i^2/M_\chi^2)^{1/2}$ in the cross section, where μ_i is the mass of the gauge boson in the final state. Annihilation to light Higgs bosons is also modified only by this factor (of course, Higgs bosons heavier than the DM are not annihilation products). For simplicity we replace μ_i by the average mass $\bar{\mu}$ of the gauge bosons and light Higgs bosons in this part of the calculation.

The full cross section for annihilation into gauge bosons, N_3 Higgs triplets, and N_5 Higgs quintuplets, including leading velocity dependence, is

$$\sigma v_{\rm rel} = \frac{\pi}{12}\frac{\alpha_g^2}{M_\chi^2}\left[\left(\frac{25}{2}+2N_3+10N_5\right)\left(1+\frac{\pi\alpha_g}{v_{\rm rel}}\right) + \left(\frac{317}{48}-\frac{5N_3}{12}-\frac{25N_5}{12}\right)v_{\rm rel}^2\right]\left(1-\frac{\bar{\mu}^2}{M_\chi^2}\right)^{1/2} \quad (16)$$

in the center-of-momentum (CM) frame. The factor $(1+\pi\alpha_g/v_{\rm rel})$ incorporates Sommerfeld enhancement neglecting the masses of the gauge bosons. This neglect is valid for large M_χ, such that the freeze-out temperature $\sim M_\chi/20$ is above the symmetry breaking scale. For smaller M_χ, this factor is roughly an upper bound on Sommerfeld enhancement (except very close to a resonance) and furthermore α_g is sufficiently small that Sommerfeld enhancement is an unimportant correction during freeze-out. Conversion to the rest frame of the cosmic fluid introduces an additional correction of order V^2 for center-of-momentum velocity $\vec{V}$.

To compute the relic density, one needs the thermal average $\langle\sigma v_{\rm rel}\rangle$. Using the Maxwell-Boltzmann distribution, we find

$$\langle\sigma v_{\rm rel}\rangle = \frac{\pi}{12}\frac{\alpha_g^2}{M_\chi^2}\left[\left(\frac{25}{2}+2N_3+10N_5\right)\times\left(1+\alpha_g\sqrt{\frac{\pi M_\chi}{T}}-\frac{1}{2\pi}\frac{T}{M_\chi}\right) + \left(\frac{317}{8}-\frac{5N_3}{2}-\frac{25N_5}{2}\right)\frac{T}{M_\chi}\right]\left(1-\frac{\bar{\mu}^2}{M_\chi^2}\right)^{1/2}. \quad (17)$$

Annihilations go out of equilibrium at a temperature given by $M_\chi/T_f = x_f \cong \ln\xi - \frac{1}{2}\ln\ln\xi$ with $\xi = 1.0\times10^{12}(M/\text{TeV})$ for triplet DM, giving $x_f = 20.4$ for $M_\chi = 5$ GeV and $x_f = 23.3$ for $M_\chi = 100$ GeV. We should equate (17) at the T_f with the cross section needed to match current constraints on the DM density. This varies mildly with M_χ as $\langle\sigma v\rangle_0 \cong (3.2\text{–}0.24\log(M/\text{GeV}))\ \text{cm}^3/\text{s}$ [39]. However, the latter expression assumes the usual particle content of the standard model at the time of freeze-out, whereas in our model there are three additional gauge bosons and extra dark Higgs bosons. This increases both the Hubble rate and redshifting between freeze-out and the present. Thereby, the extra particle content decreases $\langle\sigma v\rangle_0$ by a factor of $\sqrt{1+(6+3N_3+5N_5)/61.75}$. Details are given in Appendix A. For two Higgs triplets and one quintuplet, we find that a good approximation is given by

$$\alpha_g \cong \frac{1.7\times10^{-5}}{(1-\bar{\mu}^2/M_\chi^2)^{1/4}}\frac{(M_\chi/\text{GeV})}{\sqrt{1+7.7\alpha_g}}. \quad (18)$$

For the range of M_χ we are interested in, α_g is so small that the Sommerfeld enhancement factor $\sqrt{1+7.7\alpha_g}$ can be neglected. Further results for other light Higgs states are given in (A20).

B. Relative density of excited state

Let us denote the stable excited state by χ_s, and the unstable one as χ_u. At the freeze-out temperature, all three DM states are equally populated, but if the rate of downscattering transitions $\chi_s\chi_s \to \chi_1\chi_1$ remains larger than the Hubble rate at temperatures below the mass splitting δM_{1s}, the density n_s of the excited state gets suppressed relative to n_1 of the ground state. The rate of $\chi_s\chi_s \to \chi_u\chi_u$ transitions in the Galaxy at the present epoch scales with $(n_s/n_{\rm tot})^2$ (for $n_{\rm tot} = n_1+n_s+n_u$). We must therefore compute this ratio to accurately predict the rate of positron production. The direct detection rate for $\chi_s N \to \chi_u N'$ similarly scales like $(n_s/n_{\rm tot})$.

To compute the dilution of χ_s from downscattering, we solve the Boltzmann equation for the abundance $Y_s = n_s/s$, where s is the entropy density. Defining $z = \delta M_{1s}/T$, it can be cast in the form [40]

$$\frac{dY_s}{dz} = -\frac{\lambda}{z^2}\left(Y_s^2 - Y_1^2\left(\frac{Y_s^{\rm eq}}{Y_1^{\rm eq}}\right)^2\right), \quad (19)$$

where λ is related to the cross section $\sigma_\downarrow$ for $\chi_s\chi_s \to \chi_1\chi_1$ downscattering by

$$\lambda = \langle\sigma_\downarrow v\rangle\frac{s}{H}\bigg|_{z=1} \quad (20)$$

except (as we shall describe below) the multiplicity factors g_* and g_{*s} that appear in λ should retain their z dependence (only explicit factors of T get replaced by δM_{1s}).

We can simplify this by assuming that the abundance of the ground state does not change significantly during the depletion of χ_s, so Y_1 is just a constant. Furthermore $Y_2^{\rm eq}/Y_1^{\rm eq} \cong e^{-z}$ to a good approximation if the DM is in kinetic equilibrium with the standard model particles (we will discuss this caveat below). Defining the fraction $f = Y_s/Y_1$, (19) becomes

$$\frac{df}{dz} = -\frac{\bar{\lambda}}{z^2}(f^2 - e^{-2z}), \quad (21)$$

where $\bar{\lambda} = \lambda Y_1$. (However the e^{-2z} will be modified when we take into account kinetic decoupling effects; see the next subsection.) To explicitly compute λY_1, let us parametrize the DM ground state density as $n_1 = (g_{*s}/g_{*s,0})\xi T^3$, where the g_{*s} factors take into account the dilution of n_1 as a function of temperature due to entropy production after freeze-out. Then

$$\bar{\lambda} = \frac{g_{*s}\xi\delta M_{1s}M_p}{1.66\sqrt{g_*}g_{*s,0}}\langle\sigma_\downarrow v\rangle, \quad (22)$$

where $M_p = 1.22\times10^{19}$ GeV and $\xi = 7\times10^{-10}\ \text{GeV}/M_\chi$ to match the observed DM density.

The cross section $\sigma_\downarrow$ is straightforward to compute, since it is similar to e^-e^- scattering, with just two diagrams, exchange of a gauge boson in the t and u channels. In the low-velocity limit, we obtain

$$\langle \sigma_\downarrow v \rangle = 4\pi\alpha_g^2 \frac{M_\chi^2 v_t}{(\mu_i^2 + M_\chi^2 v_t^2)^2}, \tag{23}$$

where $v_t = \sqrt{2\delta M_{1s}/M_\chi}$ is the velocity of χ_1 at threshold (when the incoming χ_s particles are at rest) and μ_i is the mass of the exchanged gauge boson. For the exothermic DM model, where $\chi_s = \chi_3$, $i = 2$, while for the endothermic case where $\chi_s = \chi_2$, $i = 3$. Using this constant cross section is actually a conservative estimate, as it is near the maximum value of the full velocity dependent cross section given in Eq. (36) for $\delta M_{1s} \sim 1.1$ MeV and our typical values of M_χ, μ_i.

To solve the Boltzmann equation, we first tried to employ the semianalytic technique popularized in Kolb and Turner [41]. Namely, one writes $f = e^{-z} + \Delta$ and linearizes the equation in Δ for the early time behavior, giving $\Delta = z^2/(2\bar{\lambda})$, while $\Delta' = -\bar{\lambda} z^{-2}\Delta^2$ at late times. Integrating the latter equation between the z of freeze-out, z_f, and infinity gives the final abundance $\frac{Y_s}{Y_1} = \Delta_\infty = z_f/\bar{\lambda}$. The trick then is to appropriately determine the value of z_f. One does this by assuming that

$$\Delta(z_f) = \frac{z_f^2}{2\bar{\lambda}} = c e^{-z_f} \tag{24}$$

and then finding the value of c for which this procedure best reproduces the full numerical solution. However we find that this procedure is not sufficiently accurate for the regime we are interested in, where $Y_s/Y_1 \gtrsim 0.1$ rather than the exponentially small values of interest for Y_1 itself. There is no fixed value of c that accurately gives the same as the numerical result as Y_s/Y_1 ranges between 0.1 and 1. Therefore we numerically solve the Boltzmann equation in all cases.

To relate Y_s/Y_1 to the fractional abundance of the stable state to the total dark matter population at the present time, we must remember that the unstable state χ_u is also kept in equilibrium with χ_1 until a similar freeze-out temperature (which is the same in the limit that $\mu_2 = \mu_3$). Only at much later times $>10^3$ s, χ_u decays to the ground state. The total abundance of dark matter is then $Y_1 + Y_u + Y_s$. The fractional abundance of χ_s is therefore

$$\frac{Y_s}{Y_{\rm tot}} \cong \frac{Y_s/Y_1}{1 + Y_s/Y_1 + Y_u/Y_1}, \tag{25}$$

where $Y_{s,u}/Y_1$ denotes the values at freeze-out, from solving the Boltzmann equation. We compute Y_u/Y_1 in exactly the same way as Y_s/Y_1. The only difference is the exchange of $\mu_2 \leftrightarrow \mu_3$ for the gauge boson mass appearing in the propagator of the cross section (23). An astute reader may wonder whether $\chi_3\chi_3 \to \chi_2\chi_2$ scatterings change the ratio Y_u/Y_s additionally; however, the cross section (23) is greatly reduced for $\delta M_{23} \sim 1$–10 keV, and we find like [19] that this process freezes out at temperatures well above the mass splitting δM_{23}.

C. Kinetic equilibrium

The preceding discussion of the Boltzmann equation assumed that the DM is in kinetic equilibrium until the freeze-out of downscattering. If this is not the case, the relic density of χ_s will be smaller than estimated there. The reason is that the equilibrium density depends upon the kinetic temperature T_k and this redshifts with the expansion of the Universe $T_k \sim 1/a^2$, in contrast to the temperature of particles that are still coupled to the thermal bath, $T \sim 1/a$. To get some sense of the size of the effect, we can follow the analytic procedure for an approximate result, even though in the end we solve the Boltzmann equation numerically.

If T_d is the kinetic decoupling temperature, then $T_k = T^2/T_d$ for $T < T_d$. Let $z_d = \delta M_{1s}/T_d$. Then the term e^{-2z} in (19) must be replaced by $\exp(-2\,\max(z, z^2/z_d))$. Following the semianalytic approach described above, one finds that Eq. (24) is replaced by

$$\Delta(z_f) = \frac{z_f^3}{c\lambda z_d} = e^{-z_f^2/z_d}, \tag{26}$$

which can be rewritten as $z_f = (z_d \ln(c\bar{\lambda} z_d/z_f^3))^{1/2}$. As a consequence the relic abundance of χ_s is suppressed by $\sqrt{z_d}$ in this case. Thus it is preferable for kinetic decoupling to occur after the chemical freeze-out of χ_s, for maximizing its relic density.

The principal interaction for maintaining kinetic equilibrium with the SM is the electron scattering diagram shown in Fig. 10. The rate for this process is computed in Appendix B. The decoupling temperature as a function of ϵ (the kinetic mixing parameter for whichever gauge boson is exchanged) is shown in Fig. 11 for the case $M_\chi = 5$ GeV and $\mu = 100$ MeV. This can easily be generalized to other DM and gauge boson masses by noticing that the rate scales like $\alpha_g \epsilon^2/\mu^4$ and α_g is proportional to M_χ. Hence the scaling of ϵ in Fig. 11. For lower values of ϵ than shown in the figure, the relation extrapolates to a power law,

$$\frac{T}{10\ {\rm MeV}} \cong \left(\frac{\epsilon}{1.2\times 10^{-6}}\right)^{-2/3}. \tag{27}$$

In reality there are two transitions with two different mass splittings that can maintain kinetic equilibrium, since

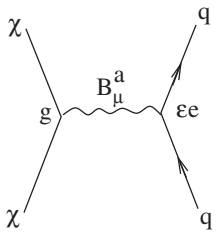

FIG. 10. Scattering of χ on charged particle q that keeps DM in kinetic equilibrium.

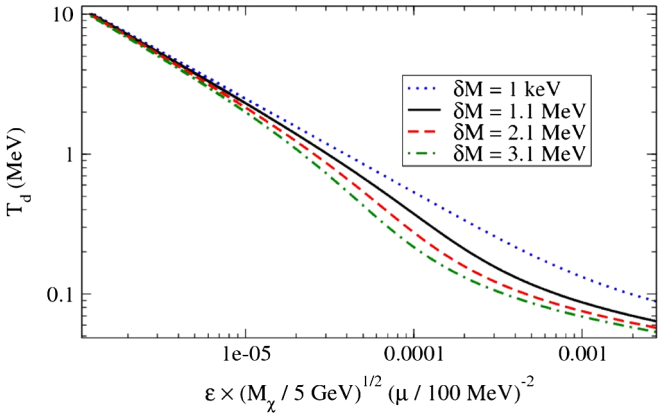

FIG. 11 (color online). Decoupling temperature for process of Fig. 10 as a function of gauge kinetic mixing parameter, for several values of the large mass splitting, and for DM mass $M_\chi = 5$ GeV and gauge boson mass $\mu = 100$ MeV.

we also have the $\chi_2 e \leftrightarrow \chi_3 e$ reaction with the small mass splitting δM_{23}. We compute the decoupling temperature for both reactions and take the smaller of the two as the true T_d. Roughly speaking, only the larger of the two ϵ's is therefore relevant for kinetic equilibrium. Figure 11 shows that there is a weak dependence upon δM with the large mass gap giving a bigger effect. There is also dependence upon the gauge boson masses.

For the exothermic model, the coupling ϵ_1 is large enough so that ϵ_2 is practically irrelevant for kinetic equilibrium. This is illustrated in Fig. 12 (left panel), which shows that contours of $\log R_{e^+}/R_{\rm obs}$ hardly change between $\epsilon_2 = 10^{-1}$ and 10^{-10}. (The example shown is for $\mu_2 = 250$ MeV; for larger μ_2 the dependence is even weaker.) For the endothermic model, ϵ_1 is smaller and so ϵ_3 can have a bigger impact. The right panel of Fig. 12 shows that δM_{23} must decrease by about 1 keV in the INTEGRAL/DAMA-allowed region to compensate the effect of making ϵ_3 arbitrarily small.

We found that for $\epsilon \sim 10^{-3}$, the reaction of Fig. 10 stays in equilibrium to such low temperatures that z_d is nearly always greater than z_f. This is the case for Figs. 2–5. One starts to see the effect of high decoupling temperatures already for $\epsilon = 10^{-4}$. This is illustrated in Fig. 12 where we show how decreasing ϵ must be compensated by a corresponding decrease in δM_{23} in order to keep the rate of positron production constant.

V. POSITRON PRODUCTION RATE AND ANGULAR PROFILE

A. Rate from inelastic collisions

The most recent determination of the observed positron annihilation rate in the bulge is 1.1×10^{43}/s [6]. This value depends upon the assumed distance between the sun and the galactic center; consistently with [6] we take $r_\odot = 8.5$ kpc [42]. For the predicted rate, we have

$$R_{e^+} = \frac{1}{2}\left(\frac{Y_s}{Y_{\rm tot}}\right)^2 \int d^3x \langle\sigma v\rangle \frac{\rho^2}{M_\chi^2}, \tag{28}$$

where σ is the cross section for $\chi_s\chi_s \to \chi_u\chi_u$ (recall that $\chi_{s,u}$ are the stable and unstable excited states). The $1/2$ is to avoid double-counting, and the abundance factor $Y_s/Y_{\rm tot}$ is given by (25). We integrate over a region of radius 1.5 kpc, corresponding to an angular diameter of approximately 10°. The observed profile, Fig. 9 suggests that the signal falls below the sensitivity of INTEGRAL near this angle.

The phase space average of σv is given by

$$\langle\sigma v\rangle = \int d^3v_1 d^3v_2 f(v_1) f(v_2) \sigma |\vec{v}_1 - \vec{v}_2|. \tag{29}$$

We take a Maxwellian velocity distribution

$$f(v) = N e^{-v^2/v_0^2} \theta(v - v_{\rm esc}) \tag{30}$$

cut off at the escape velocity

$$v_{\rm esc}^2(r) = 2v_0^2(r)[2.39 + \ln(10\ {\rm kpc}/r)] \tag{31}$$

and having velocity dispersion

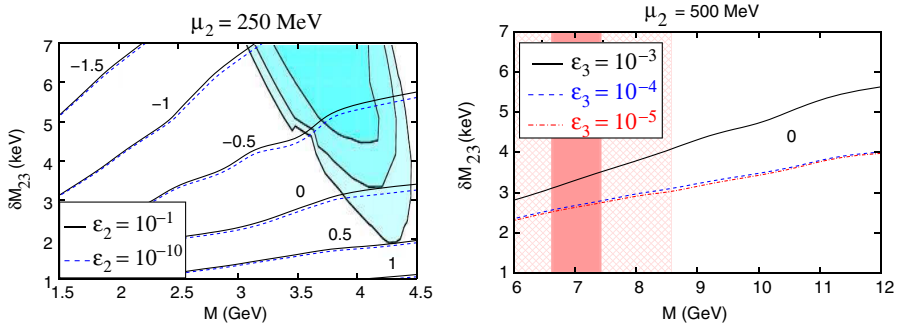

FIG. 12 (color online). To illustrate the effect of $\epsilon_{2,3}$ on kinetic decoupling and the relic density of the excited state, left: contours of $\log R_{e^+}/R_{\rm obs}$ in the exothermic model, varying ϵ_2 between 10^{-1} and 10^{-10}. μ_2 is fixed at 250 MeV and other parameters are as in Fig. 2. Right: similar plot for the endothermic model, with $\mu_2 = 500$ MeV. For clarity only the contours with $\log R_{e^+}/R_{\rm obs} = 0$ are shown. Dependence on ϵ_3 is saturated for $\epsilon_3 > 10^{-3}$ or $\epsilon_3 < 10^{-5}$.

$$v_0(r)^3 \propto r^\chi \rho(r), \tag{32}$$

with $\chi = 1.64$, and the normalization such that $v_0(r_\odot)$ is 220–230 km/s. This form of v_0 is suggested by N-body simulations that include the effects of baryonic contraction [30]. Our choice of $v_{\rm esc}$ follows Ref. [43]; see Appendix C of that paper.

In our previous work, the major challenge was to compute σ since we were concerned with DM at the TeV scale, implying gauge couplings α_g larger than the average DM velocity v. In this case a nonperturbative calculation of σ was necessary, since multiple gauge boson exchanges occur when $v < \alpha_g$, similarly to the Sommerfeld enhancement in DM annihilation [37]. However in the present situation $\alpha_g \ll v$ and a perturbative treatment suffices.

We define some kinematic variables to facilitate the presentation of the cross section:

$$v_t^2 = 2\frac{\delta M_{23}}{M_\chi}, \qquad \Delta = \frac{v^2}{v_t^2}, \tag{33}$$

where δM_{23} is the small splitting between the two excited states, v is the DM velocity in the center-of-mass frame, and v_t is the threshold velocity for $\chi_2\chi_2 \to \chi_3\chi_3$ excitations. The cross section for excitations can be expressed as

$$\sigma_\uparrow v_{\rm rel} = 4\pi\alpha_g^2\sqrt{\Delta - 1}\,\frac{v_t M_\chi^2}{\mathcal{D}^2}\left(\frac{2}{1-\eta^2} - \frac{1}{2\eta}\ln\frac{1+\eta}{1-\eta}\right), \tag{34}$$

where

$$\mathcal{D} = M_\chi^2 v_t^2(2\Delta - 1) + \mu_1^2; \qquad \eta = 2\frac{M_\chi^2 v_t^2\sqrt{\Delta(\Delta-1)}}{\mathcal{D}}. \tag{35}$$

Notice that $\sigma_\uparrow v_{\rm rel}$ vanishes at threshold, $\Delta = 1$. The related cross section for $\chi_3\chi_3 \to \chi_2\chi_2$ de-excitations is

$$\sigma_\downarrow v_{\rm rel} = 4\pi\alpha_g^2\sqrt{\Delta + 1}\,\frac{v_t M_\chi^2}{\bar{\mathcal{D}}^2}\left(\frac{2}{1-\bar\eta^2} - \frac{1}{2\bar\eta}\ln\frac{1+\bar\eta}{1-\bar\eta}\right), \tag{36}$$

where

$$\bar{\mathcal{D}} = M_\chi^2 v_t^2(2\Delta + 1) + \mu_1^2; \qquad \bar\eta = 2\frac{M_\chi^2 v_t^2\sqrt{\Delta(\Delta+1)}}{\bar{\mathcal{D}}}. \tag{37}$$

As expected, $\sigma_\downarrow v_{\rm rel}$ can be obtained from $\sigma_\uparrow v_{\rm rel}$ by changing $\delta M_{23} \to -\delta M_{23}$, which implies $v_t^2 \to -v_t^2$ and $\Delta \to -\Delta$ (notice that $v_t\sqrt{\ldots} = \sqrt{v_t^2\ldots}$). In the limit $\Delta \to 0$, and substituting $\delta M_{23} \to \delta M_{1s}$ and $\mu_1 \to \mu_i$, we recover the zero-velocity cross section for downscattering through the large mass gap, (23).

B. Rate from decaying DM

We consider the scenario where the unstable state χ_u is so long lived that it is already present in the Galaxy due to its relic density, and decays with a lifetime τ_u greater than the age of the Universe. Assuming that the 511 keV gamma rays observed by INTEGRAL come from a central region of radius r_c, the rate of positrons is

$$R_{e^+} = \frac{4\pi}{M_\chi\tau_u}\int_0^{r_c} dr r^2\rho_u(r) \equiv 4\pi\zeta\frac{\rho_\odot {\rm kpc}^3}{M_\chi\tau_u}\left(\frac{Y_u}{Y_{\rm tot}}\right), \tag{38}$$

where ρ_u is the density of χ_u and $Y_u/Y_{\rm tot}$ is the abundance of χ_u relative to the total DM population. We assume the Einasto profile to obtain the dimensionless factor

$$\zeta = \left(\frac{r_s}{\rm kpc}\right)^3 e^{(2/\alpha)(r_\odot/r_s)^\alpha}\frac{1}{\alpha}\left(\frac{\alpha}{2}\right)^{3/\alpha} \times\left[\Gamma\left(\frac{3}{\alpha}\right) - \Gamma\left(\frac{3}{\alpha}, \frac{2}{\alpha}\left(\frac{r_c}{r_s}\right)^\alpha\right)\right]. \tag{39}$$

Matching R_{e^+} to the observed rate 3.4×10^{42}/s, we find that the lifetime of χ_u relative to the age of the Universe is

$$\frac{\tau_u}{\tau_U} = 6.5\times10^3\,\zeta\left(\frac{Y_u}{Y_{\rm tot}}\right)\left(\frac{5\ {\rm GeV}}{M_\chi}\right)\left(\frac{\rho_\odot}{\bar\rho_\odot}\right), \tag{40}$$

where $\bar\rho_\odot = 0.3\ {\rm GeV/cm}^3$. The factor ζ is plotted over a wide range of Einasto parameters in Fig. 13, showing that it is between 25 and 75 for reasonable profiles. With $Y_u/Y_{\rm tot} \sim 1/3$ this gives $\tau_u \sim 10^4$ times the age of the Universe $\tau_U \cong 10^{10}$ y.

To see how small $\epsilon_{2,3}$ this corresponds to, we can rescale the bound (51) to be derived below from demanding that similar decays of the "stable" excited state must take longer than τ_U. It implies that

$$\epsilon_{2,3}^2 \cong (7\times10^{-12})^2\left(\frac{10^{-4}}{\alpha_g}\right)\left(\frac{\mu_{2,3}}{1\ {\rm GeV}}\right)^4\left(\frac{0.1\ {\rm MeV}}{\delta M_{1s-}}\right)^3\left(\frac{50}{\zeta}\right) \times\left(\frac{Y_{\rm tot}}{3Y_u}\right)\left(\frac{M_\chi}{5\ {\rm GeV}}\right)\left(\frac{\bar\rho_\odot}{\rho_\odot}\right), \tag{41}$$

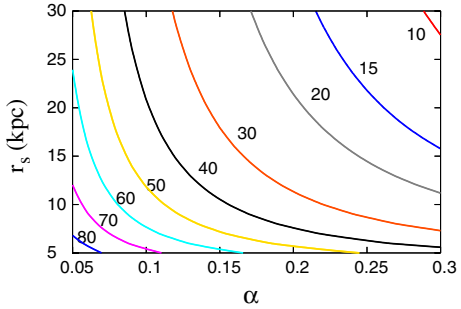

FIG. 13 (color online). Contours of ζ, defined in (39), in the plane of the Einasto halo parameters, with $r_c/r_\odot = 0.176$. ζ is related to the volume integral of the DM density in the region of the INTEGRAL 511 keV signal, Eq. (40).

where $\delta M_{1s-} = \delta M_{1s} - 2m_e$ is the energy available for the decay. It is theoretically easy to achieve the desired rate of positron creation just by adjusting $\epsilon_{2,3}$ to this small value, since there is no other constraint on $\epsilon_{2,3}$.

C. Angular distribution

In this section we elaborate on the angular profile of the 511 keV signal in the case of scatterings only, since only there is it definitely necessary to consider the effects of positron propagation. The intensity of the signal as a function of angle is found by computing the line-of-sight integral (10) where the line is oriented along the direction $\hat{x}$ specified by angles θ, ϕ relative to the galactic center. This expression assumes that the positrons decay at the same position where they were created. To model the effects of propagation before decay, we smear the angular distribution by averaging $\hat{x}$ weighted by some function $f(\cos\theta)$,

$$\bar{I}_{e^+}(\hat{x}') = \int d\Omega f(\hat{x}\cdot\hat{x}') I_{e^+}(\hat{x}). \tag{42}$$

The integral over solid angle can be combined with the integral over the line of sight and rewritten in terms of a volume integral, $dx d\Omega = d^3x/x^2$, with the origin of coordinates at the earth. Now, given that ρ^2 is strongly peaked near the galactic center, we can write $\rho^2 \sim \delta^{(3)}(\vec{x} - \vec{x}_0)$, where x_0 is the position of the galactic center. Then we find that

$$\bar{I}_{e^+}(\theta) \sim f(\cos\theta). \tag{43}$$

The intensity has the same shape as the smoothing function. As argued in Sec. II D, this is a good approximation for the DM halo profiles that we are considering for the inelastic scattering mechanism.

It is interesting to notice that even though ρ^2 looks like a delta function with respect to the measure d^3x/x^2, not so for the usual volume measure d^3x. Indeed, the function $r^2\rho^2(r)$ has a maximum near $r = \sqrt{2}r_s$ even in the limit $\alpha \to 0$. Therefore the total rate of positron production in the Galaxy gets significant contributions away from the galactic center, although these are not counted in the observations of the bulge component upon which we focus in this paper, since only near the center is the intensity high enough to be detected.

D. Regions of positron annihilation

In the above discussion we have assumed that positrons are able to migrate to the regions where positronium forms and where they can subsequently annihilate. It is known from fitting the observed γ ray spectrum that ~97% of the positrons indeed form positronium before annihilating [44,45]. This is because orthopositronium decays to three photons, and comparison of the 511 keV line flux with the continuum level is consistent with nearly all annihilations coming from positronium rather than positrons encountering free electrons. The spectral shape also shows that most annihilations take place in warm (~ 8000 K) [46] regions, which may be mostly ionized [44] or else a combination of neutral and ionized regions [45]. Note that annihilations in molecular regions are disfavored, with at most 8% of annihilations occurring in molecular regions [45].

Efforts have been made to independently map out the positions of the warm regions in the galactic bulge (GB); doing so could provide a consistency check on the above determinations, since then the morphology of the INTEGRAL detection of the galactic bulge 511 keV gamma rays should match the position of the warm regions. Reference [47] has modeled the spatial distribution of molecular gas in the GB based on CO emission data [48] for the central molecular region (CMZ) in the inner 150 pc, and borrowing an older model [49] based on H*I* observations for the "holed GB disk" region extending to radii of ~1 kpc. It has been suggested that the warm neutral or ionized regions relevant for positronium ionization coincide with these molecular gas clouds [4]. In Fig. 14, we have transposed an outline of the CMZ and holed disk regions (Fig. 4 of Ref. [47]) on the most recent INTEGRAL 511 keV intensity map [6].

However, the assumption that these regions coincide with the warm ionized or neutral + ionized regions of positronium annihilation is questionable, particularly because the positronium annihilation signal from molecular regions is strictly limited by the γ-ray spectrum. The direct measurements of the ionized component are based upon pulsar observations [50], which suggest the existence of warm H^+ regions with similar morphology to the molecular gas. But this is not considered to be a very reliable measurement of the ionized gas density in the GB due to the scarcity of pulsars in this region [51]. Thus we do not know with a high level of confidence where the warm

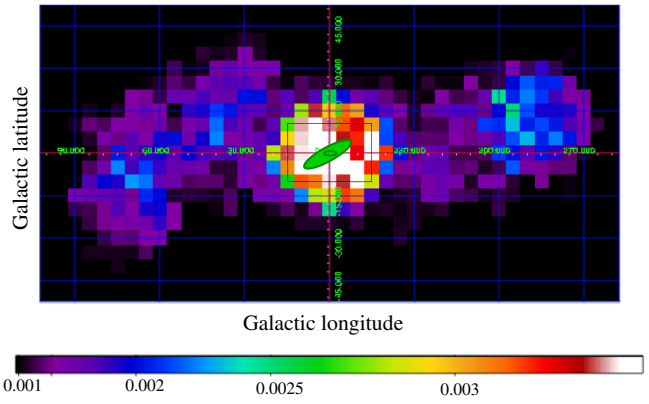

FIG. 14 (color online). Superposition of molecular gas regions of [47] (green ellipses in center) on the intensity map of the INTEGRAL 511 keV observations from Ref. [6]. Innermost ellipse is the CMZ; outer tilted ellipse is the holed galactic bulge disk.

regions of positronium annihilation are really located. It is possible that they extend beyond the molecular gas regions identified by [47], filling the inner galactic bulge. Unfortunately, the INTEGRAL/SPI instrument does not have sufficient spatial resolution to test whether the positron annihilation is localized around the molecular gas regions or fills more of the bulge (see Fig. 14).

Propagation of positrons before their annihilation is likewise uncertain. In one study, Ref. [4] assumed that positrons from the radioactive decays of supernova ejecta can be transported from the galactic disk into the GB, arguing that these positrons account for the entire observed 511 keV signal in both the GB and disk. However, as summarized in [10], their model makes many strong assumptions that are open to criticism. For one, their model of electron transport is extrapolated from the interplanetary medium and may be inappropriate for the GB, which has an unknown magnetic field configuration [35,52]. Furthermore, [4] requires a high fraction of ^{56}Co to escape from supernovae, but this escape must occur early in the explosion and is not supported at this point either by theory or observation. In any case, the rate of type Ia supernovae in the GB is uncertain by at least a factor of 2, rendering precise predictions moot. While the model of [4] is not ruled out, it seems too early to claim that it explains the 511 keV signal, particularly from the GB. Alternately, as proposed in [34], positrons from the disk can follow poloidal magnetic fields into the GB (again, assuming that ^{56}Co escapes from supernovae), but other spiral galaxies have instead an X-shaped field structure. Thus, it is highly uncertain that this model could work either.

Because of the lack of very reliable information as to the spatial distribution of the warm ionized (+ neutral) regions, an uncertainty that is acknowledged in careful studies such as [47], it is possible that positronium production occurs in the vicinity of the initially produced positrons.[8] In this case, positrons could annihilate near the decay of metastable DM particles, so that the INTEGRAL signal would be a reflection of the underlying DM distribution. On the other hand, if the positrons result from DM scattering, we have shown that they are initially produced within $1^\circ \sim 150$ pc of the galactic center (dotted red curve of Fig. 9). Then transport of the positrons to larger radii is probably necessary to be consistent with the observed extent of the 511 keV signal. Reference [35] recently showed that, under the assumption that positrons do not scatter effectively from magnetic fields in the GC, they can travel well outside of the GC before annihilating. If this is the case, the initially highly localized source from DM scattering will be widened to fill the interstellar medium. However, the nature of the magnetic fields in this region is largely unknown, and positron transport is highly sensitive to it.

More information will be needed to attach a firm interpretation to the angular distributions of the DM decays or annihilations. Observations of the 511 keV gamma rays using a future instrument with better spatial resolution would clearly be desirable for helping to settle these questions. In particular, if a new measurement revealed stronger localization of the GB component of the 511 keV emission toward the galactic center, it would favor the DM explanation over astrophysical sources. In the interim, we note that metastable DM collision and decay are still viable explanations with sufficiently cuspy halo profiles.

VI. DIRECT DETECTION RATES

In our computation of the 511 keV rate, we fixed the value of gauge kinetic mixing parameter ϵ_1 so as to match the direct detection rates determined, respectively, by Refs. [19,25] for the exothermic and endothermic DM models. Although ϵ_1 does not directly affect the rate of $\chi_2\chi_2 \leftrightarrow \chi_3\chi_3$ transitions, it does so indirectly, because of its influence (through kinetic decoupling) on the relic density of the stable excited state. Here we give details on the determination of ϵ_1 in these two cases. We note briefly that the nuclear form factor is trivial for collisions studied here, so we will ignore it.

A. Exothermic dark matter

Reference [19] determined the elastic limit of the DM-nucleon cross section needed to get the right rate of DAMA transitions:

$$\sigma_{n,\mathrm{el}} = \frac{\mu_n^2}{4\pi\Lambda^4}, \tag{44}$$

where $\mu_n = m_n M_\chi/(m_n + M_\chi)$ is the reduced mass and $\Lambda = 340$ GeV.[9] In our model, the coupling is to protons only, and the analogous quantity is given by

$$\sigma_{p,\mathrm{el}} = 16\pi\epsilon_1^2\alpha\alpha_g\frac{\mu_n^2}{\mu_1^4}. \tag{45}$$

To determine the value of ϵ_1 needed to match the observed rates, we must account for the coupling to protons only since the rate is proportional to $A(Zf_p + (A-Z)f_n)^2$ for atomic number and mass Z, A and relative strengths of couplings to protons and neutrons f_p, f_n. In Ref. [19], the couplings were assumed to be $f_p = f_n = 1$, but we have $f_p = 1$, $f_n = 0$. Moreover, we have a different local density of the excited state than that assumed by [19] because of the abundance factor Y_s/Y_{tot} [which also appears in the positron rate (28)], and because we allow the local

[8] As another example, Ref. [53] have proposed that positron annihilation in a hot but adiabatically cooling medium can fit the spectrum, allowing positron annihilation near their sources throughout the GB.

[9] Note that taking this elastic limit negates the need to average over DM speeds as in (49).

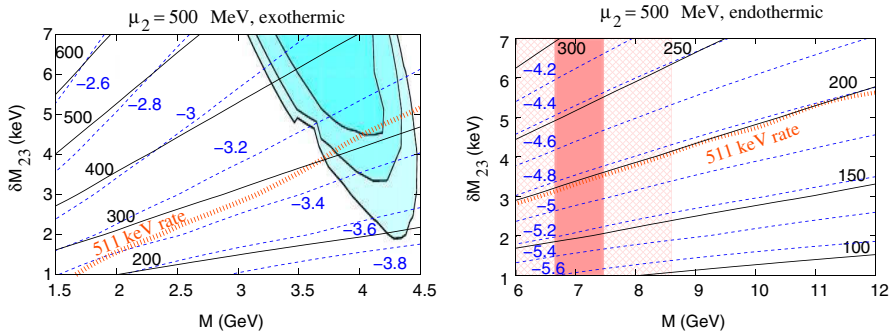

FIG. 15 (color online). Contours of $\log\epsilon_1$ (dashed) and μ_1 (solid, in MeV) for $\mu_2 = 500$ MeV, in the exothermic (left) and endothermic (right) models. Thick curve labeled "511 keV rate" is the contour where the predicted positron rate matches the INTEGRAL observation. Shaded areas are the DAMA-allowed regions.

DM density to vary with respect to the fiducial value $\rho_0 = 0.3\ \mathrm{GeV/cm^3}$. The result is

$$\epsilon_1 = \frac{\mu_1^2}{8\pi\Lambda^2}\frac{A}{Z}\left[\frac{1}{\alpha\alpha_g}\frac{Y_{\mathrm{tot}}}{Y_s}\frac{\rho_0}{\rho_\odot}\right]^{1/2}. \tag{46}$$

If $\epsilon_{2,3} \ll \epsilon_1$ so that ϵ_1 determines the kinetic decoupling temperature of the DM, then Y_s depends implicitly on ϵ_1 and (46) must be solved numerically. The factor with A/Z depends upon which nucleus we are talking about, and is given by 2.4 and 2.28, respectively, for I and Na. As [19] notes, scattering from Na nuclei is preferentially detected in our region of parameter space, so we choose the latter number.

In Fig. 15 (left panel) we plot contours of ϵ_1 corresponding to the $\mu_2 = 500$ MeV example shown in Fig. 2, to give a sense for how large ϵ_1 must be. Near $M_\chi = 4$ GeV, $\delta M_{23} = 4.5$ keV, where the INTEGRAL and DAMA rates best fit simultaneously, $\epsilon_1 \sim 10^{-3.27}$, significantly larger than the generic estimate 10^{-5} given in Ref. [19]. This is due to the A/Z correction, the fact that we need μ_1 to be heavier than the nominal 100 MeV value assumed in [19], and that Y_s/Y_{tot} can be significantly less than 1 in our model.

B. Endothermic dark matter

Reference [25] finds that DM with a mass of approximately 7 GeV and cross section on nucleons

$$\sigma_n = 2 \times 10^{-4}\ \mathrm{pb} \tag{47}$$

can simultaneously fit the DAMA/LIBRA annual modulation and the CoGeNT low-recoil events. Their allowed regions of σ_n versus M_χ are reproduced in Fig. 8. The logic for matching our cross section to theirs is similar to the exothermic case, except for the fact that endothermic scatterings are kinematically blocked if the DM velocity is below the threshold value

$$v_t = \sqrt{2\delta M_{23}/\mu_N}, \tag{48}$$

where $\mu_N = m_N M_\chi/(m_N + M_\chi)$ is the nucleus-DM reduced mass. We take this into account by doing the phase space average of σv. The phase space factor in σv that is sensitive to the threshold is $\sqrt{v^2 - v_t^2}$. For elastic scattering, this factor would be v. Therefore we match the quantity (45) that also appears in our slightly inelastic cross section to (47) using

$$\sigma_{p,\mathrm{el}} = \frac{\langle v\rangle}{\langle\sqrt{v^2 - v_t^2}\rangle}\left(\frac{A}{Z}\right)^2\frac{\rho_0}{\rho_\odot}\frac{3Y_1}{Y_s}\sigma_n, \tag{49}$$

where the averages over velocity are performed with the Maxwellian distribution function $f = Ne^{-v^2/v_0^2}$ cut off at the escape velocity v_{esc}. Since we are comparing with Ref. [25], we use their values $v_0 = 230$ km/s and $v_{\mathrm{esc}} = 600$ km/s for this part. Once again, scatterings from Na are preferentially detected, so $A/Z = 2.28$. Similarly, we take the threshold velocity for sodium in the above.

Figure 15 (right panel) shows contours of ϵ_1 for the case of $\mu_2 = 500$ MeV. In the overlap region for INTEGRAL and DAMA, $\epsilon_1 \cong 10^{-5}$. This is smaller than required in the exothermic model because the corresponding value of μ_1 is smaller, and also the cross section (47) is approximately 0.15 of that in (44). Kinetic equilibrium of the DM with the SM in the early Universe is not as efficiently maintained by $\chi_2 \leftrightarrow \chi_3$ transitions in this case. This is why the relic density of χ_2 is sensitive to the value of the other nonvanishing kinetic mixing parameter for the endothermic model, whereas it is practically insensitive in the exothermic case.

VII. ASTROPHYSICAL CONSTRAINTS

In this section we address the astrophysical and cosmological constraints on our proposal that are complementary to the 511 keV and direct detection signals, as well as to laboratory constraints from electron beams.

A. Lifetime of metastable state

We need to insure that the stable excited state χ_s is either truly stable or else sufficiently long lived. The most dangerous process is $\chi_s \to \chi_1 e^+ e^-$. At the phenomenological level, we suppress this by setting the kinetic mixing of the gauge boson that mediates this process to a sufficiently small value. The rate for this decay can be computed analytically with the approximation that $\delta M_{1s-} = \delta M_{1s} - 2m_e$ is sufficiently small for the final state particles to be nonrelativistic. Then

$$\Gamma_{\chi_s} \cong 2\alpha_g \alpha \epsilon^2 m_e^2 \delta M_{1s-}^3 \mu^{-4}. \tag{50}$$

Demanding that τ_s exceed 10^{10} y requires that

$$\epsilon < 2 \times 10^{-9} \left(\frac{10^{-4}}{\alpha_g}\right)^{1/2} \left(\frac{\mu}{1\ \text{GeV}}\right)^2 \left(\frac{0.1\ \text{MeV}}{\delta M_{1s-}}\right)^{3/2}. \tag{51}$$

In the endothermic model, we can set $\epsilon_3 = 0$ at tree level by removing the Δ_3 Higgs boson, but there seems to be no symmetry to ensure that ϵ_3 is not generated by loops if the other two mixing parameters are nonzero. Nonetheless, we are not able to find an example of a loop diagram that generates nonzero ϵ_3; any that superficially seem promising vanish because of Furry's theorem. Instead, we find a one-loop process where B_3 acquires a magnetic moment coupling to the electron, $\mu_B B^{\mu\nu} \bar{e} \sigma_{\mu\nu} e$. The decay of χ_2 proceeds by B_3 exchange in the one-loop diagrams of Fig. 16. These diagrams would cancel exactly if $\mu_2 = \mu_1$, so the magnetic moment can be estimated as

$$\mu_B \sim g \alpha \epsilon_1 \epsilon_2 \frac{(\mu_2 - \mu_1)}{4\pi \bar{\mu}^2} \ln \frac{\Lambda}{\bar{\mu}}, \tag{52}$$

where $\bar{\mu} = \frac{1}{2}(\mu_1 + \mu_2)$ and Λ denotes the hidden $SU(2)$ symmetry breaking scale, above which the kinetic mixing of $B_{1,2}$ is replaced by an interaction with the triplet Higgs fields. The squared matrix element of Fig. 16 can be estimated as $|\mathcal{M}|^2 \sim g^2 \mu_B^2 M_\chi^2 m_e^2 \delta M_{12}^2 \mu_3^{-4}$, and the decay rate in the limit of small $\delta M_{12-} \equiv \delta M_{12} - 2m_e$ is

$$\Gamma_{\chi_2 \to \chi_1 e^+ e^-} \sim \frac{\alpha_g}{32\pi} \mu_B^2 m_e^3 \delta M_{12}^2 \delta M_{12-}^2 \mu_3^{-4}. \tag{53}$$

For $\alpha_g \sim 10^{-4}$ and $\epsilon_1 \sim \epsilon_2 \sim 10^{-3}$, $\delta\mu \sim \mu_3 \sim \bar{\mu} \sim 1$ GeV, $\delta M_{12-} \sim 0.1$ MeV, $\Lambda \sim 10$ GeV, we find a lifetime of 10^{26} s, much larger than the age of the Universe. Therefore it seems technically natural to neglect the dangerous kinetic mixing term and assume the stable state is sufficiently long lived. As it turns out, a careful calculation is even more suppressed; see Appendix c.

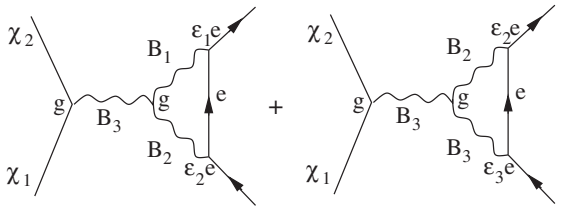

FIG. 16. Decay of metastable χ_2 state due to nonzero $\epsilon_{1,2}$.

The $\chi_3 \to \chi_2 X$ decay in the exothermic model is not problematic, since the mass splitting is much smaller and the only available decay channels are with $X = \gamma$, $X = \nu\bar{\nu}$, and $X = 3\gamma$. These have been studied previously [14,23,36]. The single-photon decay has a lifetime longer than the age of the Universe for the value of ϵ_1 required for direct detection; since it could be observed, we discuss it in more detail in Sec. VII B below.

The partial width for the $\nu\bar{\nu}$ final state is easy to estimate in analogy with (50). For this channel, there is an additional suppression in the kinetic mixing. First, the mixing of B_1 with the Z boson current has an extra factor of μ_1^2/m_Z^2 [36], and, second, the SM Z boson mixes with the B_1 current with opposite sign such that the two mixings nearly cancel at small energy-momentum transfer (see Appendix B for discussion of the same cancellation in $\chi\nu$ scattering). The ensuing bound on ϵ_1 is much weaker than that on ϵ_2; practically speaking there is no constraint.

The decay $\chi_3 \to \chi_2 + 3\gamma$ is due to the operator $\sim(\epsilon\alpha^2/90m_e^4) B_1^{\mu\nu} F_{\mu\nu} F^2$ induced by an electron loop, similar to the Euler-Heisenberg F^4 interaction in QED. (Furry's theorem forbids a term of the form $B_1 F^2$ and $B_1 F$ mixing is already taken into account by diagonalizing the kinetic terms.) The rate is suppressed by δM_{23}^{13} [14,23], leading to lifetimes that far exceed the age of the Universe for the small $\sim$5 keV splittings relevant to our exothermic model.

B. Single-photon decays of metastable state

The exothermic version of our proposal faces the challenge that the excited state can decay by emission of a single photon, via $\chi_3 \to \chi_2 \gamma$. The origin of this decay was pointed out in [36]: the non-Abelian term in the field strength $B_1^{\mu\nu}$ leads to interactions of the form $\epsilon_1 B_2^\mu B_3^\nu F_{\mu\nu}$ with the photon, from the gauge kinetic mixing operators (1). This can be put into a loop diagram which results in a transition magnetic moment χ_2-χ_3,

$$\mu_{23} \cong \frac{\epsilon_1 g^2}{128\pi^2 M_\chi} \left(\ln \frac{M_\chi}{\mu} - 1\right), \tag{54}$$

where μ is of order μ_2, μ_3. Therefore, there is a decay channel $\chi_3 \to \chi_2 \gamma$. The rate is

$$\Gamma_\gamma = \frac{\mu_{23}^2}{8\pi} (\delta M_{23})^2. \tag{55}$$

For $\epsilon_1 \cong 10^{-5}$, $M_\chi \cong 4.5$ GeV and $\delta M_{23} \cong 5$ keV, the lifetime is 4×10^{19} s which is much longer than the age of the Universe. However, this is not necessarily enough because such photons could be visible in astronomical searches.

Of the various instruments that could be sensitive to low-energy γ rays, INTEGRAL/SPI comes the closest.

Reference [54] gives limits on the intensity of gamma-ray lines that could come from such decays in the Galaxy; however INTEGRAL's sensitivity cuts out below 20 keV, making our scenario just out of reach. Interestingly limits on the diffuse gamma-ray background put a δM_{23} dependent lower limit on the partial lifetime times the mass [55] of approximately

$$\tau_\gamma M_\chi > 1 \times 10^{20} \left(\frac{3Y_3}{Y_{\rm tot}}\right)\left(\frac{\delta M_{23}}{10\ {\rm keV}}\right)^{1.2}\ {\rm GeV\,s} \qquad (56)$$

for $\delta M_{23} > 10$ keV. Again because of INTEGRAL's energy sensitivity, data is not given for lower photon energies. Nevertheless, extrapolating the bound to $\delta M_{23} = 5$ keV gives 4.3×10^{19} GeV s, which is not even 5 times less than our nominal value 1.8×10^{20} GeV s, assuming $M_\chi = 4.5$ GeV. Therefore an instrument sensitive to these lower energies might detect this low-energy photon, which is in the x-ray part of the spectrum.

In fact, observations of the galactic center by the Chandra x-ray telescope [56] may rule out this particular model. Observations are presented for a region of size 35 arcmin^2 that is 7.5 arcmin away from the GC. No evidence of an unidentified line is observed in the 1–8 keV band (Fig. 17), and the continuum seen there is modeled by thermal sources with a flux of 6×10^{-12} erg cm^{-2} s^{-1}. We can compute the expected flux by integrating over the line of sight and the solid angle ($d\Omega = d\phi d\cos\psi$) of the observed region [55],

$$F_{\rm th} = \frac{Y_3/Y_{\rm tot}}{4\pi M_\chi \tau_\gamma} \int d\Omega \int dl \rho(|\vec{l} - \vec{r}_\odot|) \equiv \frac{\rho_\odot r_s}{2M_\chi \tau_\gamma} \frac{Y_3}{Y_{\rm tot}} I \qquad (57)$$

where $I = e^{(2/\alpha)y^\alpha} \int d\cos\psi \int d\hat{l} e^{-(2/\alpha)(y^2+\hat{l}^2-2\hat{l}y\cos\psi)^{\alpha/2}}$, $y = r_\odot/r_s$ and $\hat{l} = l/r_s$. We numerically integrate over an annular region of similar area and displacement from the GC to the observed one, using the Einasto parameters (7) to find $I \cong 10^{-4}$. Using the value of $M_\chi \tau_\gamma$ determined above, this gives $F_{\rm th} \cong 0.001$ photons cm^{-2} s^{-1} (for the maximal ratio $Y_3/Y_{\rm tot}$). The corresponding energy flux for a 5 keV mass difference is 10^{-11} erg cm^{-2} s^{-1}, not quite 2 times greater than the observed continuum flux.

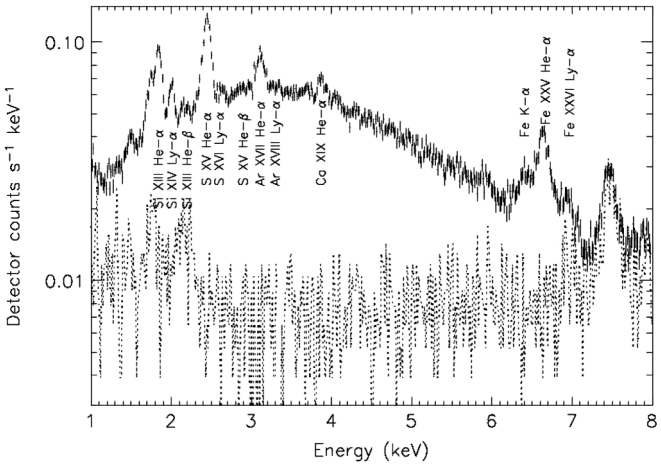

FIG. 17. Chandra spectrum from region near galactic center, where ~5 keV x-ray from $\chi_3 \to \chi_2 \gamma$ decay might manifest itself.

In the above estimate we did not take account of absorption of the decay signal, which would help to soften the discrepancy, especially if the photon energy is somewhat lower. It may also be possible to evade the problem by extending the gauge group to $SU(2) \times {\rm U}(1)$ [57] and replace the kinetic mixing of B_1 by that of the extra U(1) gauge boson; this would remove the μ_{23} transition magnetic moment. Notice that this problem does not affect the endothermic model because $\chi_3 \to \chi_1 e^+ e^-$ proceeds much more quickly, as we discuss in Sec. VII C below.

C. Lifetime of unstable state

In passing, we can also estimate the decay rate for the unstable excited state into $e^+ e^-$ using (50). It has the same form, except for the substitutions of ϵ and μ by the corresponding quantities for B_3, in the exothermic model; for the endothermic model, (50) applies as written, to the unstable excited state. Laboratory experiments constrain the appropriate ϵ_i to be $\lesssim 10^{-3}$, so the lifetime could be 10^{10} times shorter then the above estimate using $\epsilon \sim 10^{-8}$, thus on order of 1 y. This assumes the large mass splitting is only 1.1 MeV. With a 2.1 MeV splitting one gains a factor of 10^5 in the rate due to the larger phase space, giving a lifetime of several hundred seconds. It cannot be significantly smaller in our model.

D. Single-photon decays of unstable state

The decay mechanism discussed in Sec. VII B was originally conceived for the decay of the unstable state in Ref. [36]. This goes through the MeV-scale mass gap, so the photon in this case is a gamma ray. For definiteness let us consider the endothermic model, so χ_3 is the unstable state and the relevant decay is $\chi_3 \to \chi_1 \gamma$ via the μ_{13} transition magnetic moment, which is proportional to ϵ_2, in analogy to (54). The partial decay rate is the obvious generalization of (55). The branching ratio for the single-photon decay relative to that into $e^+ e^-$ is [36]

$${\rm BR}_\gamma = \frac{\alpha_g^2/\alpha}{8192\pi^2} \frac{\mu^4 (\delta M_{13})^3}{M_\chi^2 (\delta M_{13-})^3 (\delta M_{13+})^2} \ln^2 \frac{M_\chi}{e\mu}, \qquad (58)$$

where $\mu \sim \mu_1,\ \mu_3,\ \delta M_{13\pm} = \delta M_{13} \pm 2m_e$ and $e = 2.71\,828\ldots$. The resulting photon might be observed by INTEGRAL in the diffuse γ ray background. In Ref. [36], a bound was derived, which however overestimated the sensitivity of INTEGRAL to the signal. We therefore reconsider it here.

The analysis of Ref. [54] is particularly relevant for us, since they searched for line sources from the galactic

center region, having a spatial distribution similar to that of the 511 keV line. They limit the flux of such a line, for energies between 1 and 2 MeV, to less than $\sim 3 \times 10^{-5}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This is to be compared to the flux from positrons, $\sim 3.6 \times 10^{-3}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Therefore BR_γ should not be greater than about 0.01. However, using the typical values of parameters of interest for our present application, we find $\mathrm{BR}_\gamma \sim 10^{-7}$, far below the sensitivity of current searches.

E. Cosmic ray and cosmic microwave background constraints

DM annihilations can occur even after freeze-out, with the production of gamma rays or charged particles that can have an observable effect. In the class of models we consider, the DM annihilates directly into hidden sector gauge bosons, which in turn decay into any charged SM particles that are sufficiently light. Gamma rays emerge only as secondary products of these charged particles. Their contribution to the diffuse gamma-ray background can potentially give interesting constraints [58], but currently the uncertainties from details of structure formation do not allow one to draw firm conclusions. The production of antiprotons in the Galaxy gives more definite constraints, which can be quite stringent [59]. To avoid them, we need to assume that the gauge bosons which mix with the photon are lighter than $2m_p$ so that $p\bar{p}$ pairs are not produced.

Charged particles that are produced around the time of recombination reionize the plasma and change the optical depth to the surface of last scattering, a quantity that affects the Doppler peaks of the cosmic microwave background [60,61]. The effect is particularly strong for DM with mass $M_\chi \lesssim 10$ GeV, as in the exothermic proposal for DAMA. Reference [61] shows that such DM is marginally ruled out if it decays exclusively into e^+e^-, while it is marginally allowed if it decays into heavier charged particles (which decay into electrons that are less energetic than if they were primary products). The relevant bounds are reproduced in Fig. 18. In most of our examples, the gauge bosons can decay into muons and charged pions, so the branching ratio into electrons will be suppressed and the bound should be somewhere between the two cases shown in Fig. 18. Furthermore, as discussed in Appendix A, the annihilation cross section in our model is somewhat reduced compared to the usual case, also lessening the tension with the bound. Reference [60] also derives a bound from excess heating of the interstellar gas, which is more stringent than the cosmic microwave background CMB bound in this small M_χ region, but which is also more subject to uncertainty because of its dependence upon assumptions about the details of structure formation.

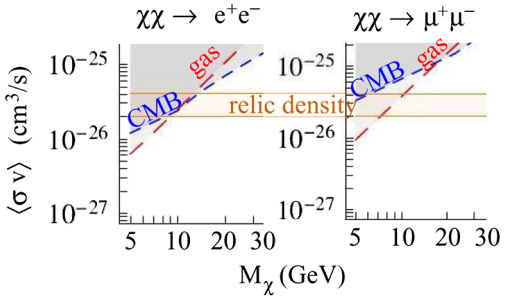

FIG. 18 (color online). Adapted from Ref. [61], showing constraints on the DM annihilation cross section versus mass from optical depth (''CMB'') and excess heating of the intergalactic medium (''gas''). Regions above the diagonal lines are excluded. ''Relic density'' region indicates the desired value of the cross section for the correct thermal abundance.

F. Long-lived gauge bosons and nucleosynthesis

It is interesting to consider possible effects of the hidden sector gauge bosons in the early Universe. Decays around the time of big bang nucleosynthesis (BBN) or later can be deleterious, although it is also possible to improve the predictions of BBN, notably for lithium [62]. In our models, B_1 is usually the lightest gauge boson, and it couples to electrons (and muons and pions) with $\epsilon_1 \sim 10^{-5}$, fixed by the rate of direct DM detection. Its decay rate is therefore of order $\alpha\epsilon_1^2\mu_1$ which for $\mu_1 \sim 500$ MeV leads to a lifetime of 10^{-10} s, which is clearly harmless.

On the other hand, if $\epsilon_2 \lesssim 10^{-10}$ in the exothermic model, where we have the constraint $\epsilon_2 \lesssim 10^{-8}$, then B_2 can have a lifetime greater than 1 s and possibly be relevant for nucleosynthesis. The question is whether its relic density is large enough to have an effect. We have computed the cross section for $B_2B_2 \to B_1B_1$ using FEYNCALC [63]. The cross section as $v \to 0$ can be expressed as

$$\sigma v = \frac{\pi\alpha_g^2}{2\mu_2^2} f(x), \tag{59}$$

where $x = 1 - \mu_2/\mu_3 = 2\delta M_{23}/\alpha_g\mu_3$ and f has a minimum value of 18 at $x = 0$ (treating (4) as an equality to eliminate μ_1). This neglects dark Higgs exchange in the s channel, but we have checked that including it makes no dramatic change unless the virtual Higgs goes on shell. For typical values we find that the standard relic abundance calculation gives a freeze-out temperature around 8 MeV for B_2, and an abundance 10^{-2} times smaller than that of the baryon asymmetry. This is too small to have any effect on BBN.

Reference [62] point out that a more likely candidate for giving interesting effects is the dark Higgs bosons. In particular, if there exist a Higgs boson that is lighter than the gauge bosons, it would decay into 4 leptons through emission of two virtual gauge bosons, with a rate suppressed by $\epsilon^2\alpha_g\alpha^2(m_h/\mu)^8$. The annihilation cross section is suppressed for similar reasons. This can more naturally give long-lived relics (on the time scale of BBN) that could solve the lithium problem.

VIII. LABORATORY SEARCHES

A. Beam dump experiments

An interesting feature of the class of models we consider is that they can be tested in proposed low-energy laboratory experiments. A beam dump on an absorbing target can produce the weakly interacting B bosons that mix with the photon. These can decay into e^+e^- or other charged particles before reaching the detector, providing a signal not present in the standard model.

In our scenario, two of the three colors of bosons should mix with the photon: B_1, with strength $\epsilon_1 \sim 10^{-3}$–10^{-5} to get the right rate of direct detection, and either B_3 or B_2, depending upon whether χ_2 or χ_3 is the stable excited state. Let us denote the corresponding mixing parameter by $\epsilon_{3,2}$. We noted in Sec. IV C (Fig. 12) that this parameter is essentially unconstrained. If $\epsilon_{3,2} \gtrsim 10^{-6}$ then the effects of $B_{3,2}$ could be discovered in laboratory searches. But since we have more definite predictions for B_1, we will focus here on its discovery potential. Moreover we have argued that there are certain advantages to having very small values of $\epsilon_{3,2}$ which could make laboratory detection of $B_{3,2}$ impossible for the present.

The authors of Ref. [64] have recently summarized the current experimental constraints in the ϵ_i-μ_i plane (where μ_i is the mass of the relevant gauge boson), and they have also proposed strategies for new experiments that can cover more of the still-allowed region in this plane. Figure 19 reproduces some of their results. On top of these we plot several examples of predictions from our endothermic and exothermic models (circles containing "n" or "x," respectively), corresponding to those shown in Figs. 2 and 5 (see also Fig. 15). Almost all of these points are contained within the contours denoting the reach of feasible new experiments suggested by Ref. [64]: the solid (blue) line denoting the high resolution, high rate trident spectrometer, and the dashed (red) one for the thin-target with double arm spectrometer. It is suggested that such experiments would be feasible at several existing laboratories, including JLab (Thomas Jefferson National Accelerator Facility), SLAC (Stanford Linear Accelerator Center), ELSA (Electron Stretcher and Accelerator), and MAMI (Mainzer Mikrotron). We see that only one of our examples (the right-most "x") would lie outside of the reach of the proposed experiments. This corresponds to the extreme case where $\mu_2 = 2$ GeV in Fig. 5. The more typical models would therefore be in the discoverable region.

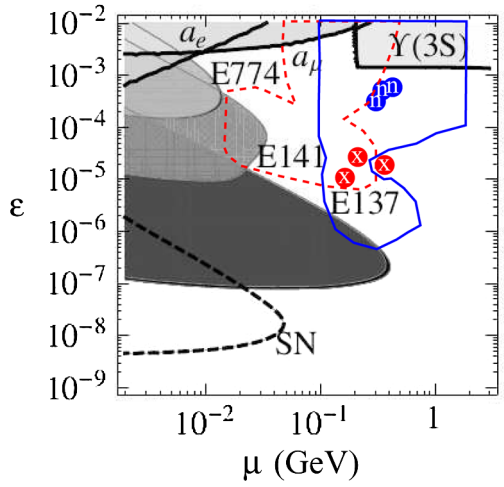

FIG. 19 (color online). Potential for discovery of light mixed gauge bosons in plane of kinetic mixing parameter ϵ and gauge boson mass μ. Shaded regions are ruled out by existing laboratory or astrophysical constraints. Unshaded enclosed regions denote the reach of experimental strategies proposed in Ref. [64]. Circles containing "n" or "x" are typical predictions of our endothermic or exothermic DM models, respectively.

B. Invisible width of Z boson

The non-Abelian gauge kinetic mixing portal (1) provides two invisible decay channels for the Z boson: $Z \to \Delta_i B_i$ since ϵ_i stands for the VEV of the Higgs triplet Δ_i over the heavy scale Λ (Eq. (11)), and $Z \to B_j B_k$ where i, j, k are a cyclic permutation of 1, 2, 3. The latter arises because the non-Abelian field strength $B_i^{\mu\nu}$ contains $g\epsilon_{ijk} B_j^\mu B_k^\nu$. Considering the first process, the partial width is

$$\Gamma_{Z\to\Delta B} = \frac{m_Z^3}{96\pi\Lambda^2} \tag{60}$$

in the approximation m_Z is much greater than the masses of the decay products. Demanding that this be less than the experimental error on the invisible Z width, 1.5 MeV [65], we find that $\Lambda > 1.3$ TeV. For $\langle\Delta\rangle \sim 10$ GeV, this leads to the bound $\epsilon \lesssim 10^{-2}$, which is less stringent than other constraints shown in Fig. 19. For the $Z \to B_j B_k$ channel, the partial width is of order $\alpha_g \epsilon^2 m_z$. This leads to a weaker bound on ϵ than does the $Z \to \Delta_i B_i$ channel.

IX. CONCLUSIONS

If the anomalous 511 keV gamma rays from the galactic center are truly distributed in an axisymmetric manner, as suggested by the INTEGRAL observations, this provides strong motivation to seriously consider DM decays or scatterings as their source, rather than localized sources such as supernovae or x-ray binaries. A new measurement with higher spatial resolution would be very desirable to help settle this question. In the meantime, it seems worthwhile to explore possible DM interpretations, especially if they can explain more than just the 511 keV signal. In the present work we have shown how a three-component DM model with two mass splittings and a hidden $SU(2)$ gauge boson might address both the 511 keV observation and indications of DM detection by DAMA/LIBRA and possibly CoGeNT.

The scenarios we have presented involve slightly inelastic nuclear scatterings, either endothermic or exothermic, in the direct detection experiments: $\chi_{2,3}N \to \chi_{3,2}N'$. The endothermic version with $M_\chi \cong 4$ GeV is under stronger pressure from astrophysical constraints from the CMB (Fig. 18) and especially from the decay $\chi_3 \to \chi_2\gamma$, not observed by Chandra (Sec. VII B). The latter could possibly be softened by some modification of the particle physics model, such as extending the gauge group to $SU(2) \times \mathrm{U}(1)$. The exothermic model also requires a more cuspy halo than the endothermic one to get the observed 511 keV rate from $\chi_3\chi_3 \to \chi_2\chi_2$ scattering, although still consistent with examples from N-body simulations that take into account compression by baryons.

We have highlighted two distinct mechanisms for getting the 511 keV signal: either inelastic $\chi_{2,3}\chi_{2,3} \to \chi_{3,2}\chi_{3,2}$ scatterings followed by $\chi_{3,2} \to \chi_1 e^+ e^-$ decays, or the decay process by itself when $\chi_{3,2}$ has a lifetime of order 10^5 times the age of the Universe. Whereas the first mechanism requires some mutual adjustments of the particle physics and DM halo parameters to get the right rate, the second is more easily arranged by just tuning the gauge kinetic mixing parameter $\epsilon_{2,3} \sim 10^{-11}$ that controls the decay rate. The decay mechanism points to the exciting possibility that the angular profile of the 511 keV signal is actually a picture of the DM halo profile in the inner part of the Galaxy, if positron diffusion is a negligible effect. The scattering mechanism on the other hand *requires* significant positron diffusion, or else propagation of the excited DM state before decay, since otherwise it predicts too narrow angular profile. It is interesting that our model can naturally explain such long-distance travel of the excited DM prior to its decay, by tuning $\epsilon_{2,3} \sim 10^{-7}$.

A very encouraging aspect of these proposals is their testability in low-energy electron beam dump experiments. The kinetic mixing parameter $\epsilon_1 \sim 10^{-4}$ and the mass $\mu_1 \lesssim 1$ GeV of the gauge boson mediating the direct detection scatterings are in prime territory for their discovery by such experiments, which could be done at existing laboratories. The models presented here are also potentially rich in consequences for cosmic rays, the diffuse x-ray or gamma-ray backgrounds, the CMB, and big bang nucleosynthesis.

ACKNOWLEDGMENTS

We thank C. Boehm, L. Bouchet, M. Cirelli, N. Dalal, M. Fairbairn, J. Feng, K. Ferriere, B. Grinstein, P. Jean, M. Kaplinghat, V. Kaspi, J. McDowell, G. Moore, N. Produit, G. Skinner, T. Slatyer, and G. Starkman for helpful correspondence or discussions. Our work is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. J. C. thanks the CERN theory group and the Perimeter Institute for their hospitality during part of this work.

APPENDIX A: ANNIHILATION AMPLITUDES AND RATES

In this appendix, we derive the invariant amplitudes (squared) for annihilation of DM particles χ to both gauge and Higgs bosons, including the lowest order corrections due to dark matter velocity, which we use to find the relation between the thermal relic density and the dark gauge coupling α_g in Sec. IV A. Including annihilation to Higgs bosons extends and corrects the results listed in [36]; in addition, we correct the final state polarization and color sums carried out in [36]. To keep the final result simple, we will first assume that symmetry breaking occurs at a lower temperature than DM freeze-out, so the gauge and Higgs bosons may be treated as massless. The effects of symmetry breaking are discussed at the end. For reference, we will consider a general gauge group and general representations for both DM and the Higgs.

We consider first DM annihilation to Higgs bosons. This process is mediated by t- and u-channel diagrams involving Yukawa couplings at each vertex, s-channel diagrams with an intermediate Higgs particle connecting a Yukawa coupling at one vertex to a scalar potential vertex, and an s-channel diagram with an intermediate gauge boson coupling to the DM and Higgs particles at either end. Since we are concerned in this paper with either parametrically small or vanishing Yukawa couplings, we assume that the s-channel diagram with an intermediate gauge boson dominates.

Consider incoming DM states χ_i in representation R and outgoing Higgs states Δ_I in representation R'; the incoming momenta are p_i and outgoing momenta q_I. The matrix element is[10]

$$\mathcal{M} = \frac{ig^2}{(p_i + p_j)^2}\bar{v}_j T^a_{ji}\gamma^\mu u_i T^a_{JI}(q_I - q_J)_\mu. \tag{A1}$$

Once summed over outgoing colors and averaged over incoming colors and spins, it is

$$|\mathcal{M}|^2 = \frac{1}{4d_R^2}\frac{g^4}{s^2}\,\mathrm{tr}[(\not{p}_i + M_\chi)(\not{q}_I - \not{q}_J)(\not{p}_j - M_\chi) \times (\not{q}_I - \not{q}_J)]T^b_{ji}T^a_{ij}T^b_{JI}T^a_{IJ}, \tag{A2}$$

where d_R is the dimension of representation R. The color sums both take the form

$$\mathrm{tr}\,{}_R T^b T^a = (d_R/d_{\mathrm{adj}})C_2(R)\delta^{ab}, \tag{A3}$$

where C_2 is the quadratic Casimir; the sum over the adjoint indices gives $\delta^{ab}\delta^{ab} = d_{\mathrm{adj}}$. For nonrelativistic dark matter at center-of-mass velocity $v = v_{\mathrm{rel}}/2$ and scattering angle θ, the amplitude becomes

[10]We use a mostly plus metric with Dirac algebra $\{\gamma^\mu, \gamma^\nu\} = -2\eta^{\mu\nu}$.

$$|\mathcal{M}|^2 = \left(\frac{d_{R'}}{d_{adj}d} C_2(R)C_2(R')\right)\frac{g^4}{2}(1 - v^2\cos^2\theta). \quad \text{(A4)}$$

Annihilation to gauge bosons (of colors a, b and momenta $q_{a,b}$) receives contributions from s-, t-, and u-channels. The amplitudes for each channel are (as in [36])

$$\begin{aligned}
\mathcal{M}_s &= \frac{g^2}{s}\bar{v}_j T^c_{ji}\gamma_\lambda u_i f^{abc}\varepsilon^\star_\mu(a)\varepsilon^\star_\nu(b)[\eta^{\mu\nu}(q_b - q_a)^\lambda - \eta^{\nu\lambda} \\
&\quad \times (q_b + p_i + p_j)^\mu + \eta^{\mu\lambda}(q_a + p_i + p_j)^\nu], \\
\mathcal{M}_t &= i\frac{g^2}{t - M_\chi^2} T^b_{jk}T^a_{ki}\bar{v}_j \not{\varepsilon}^\star(b)(\not{p}_i - \not{q}_a - M_\chi)\not{\varepsilon}^\star(a)u_i, \\
\mathcal{M}_u &= i\frac{g^2}{u - M_\chi^2} T^a_{jk}T^b_{ki}\bar{v}_j \not{\varepsilon}^\star(a)(\not{p}_i - \not{q}_b - M_\chi)\not{\varepsilon}^\star(b)u_i.
\end{aligned} \quad \text{(A5)}$$

We need to account for both direct squares and cross terms in the amplitudes. After some tedious algebra including Dirac traces, we find the following results for the (color and spin summed and averaged) square amplitude:

$$\begin{aligned}
|\mathcal{M}_s|^2 &= \frac{g^4}{d_R^2}\mathrm{tr}_R T^c T^d f^{abc}f^{abd}\left(-\frac{19}{4} + \frac{1}{8}v^2(11 - 5\cos 2\theta)\right), \\
|\mathcal{M}_{t,u}|^2 &= \frac{2g^4}{d_R^2}\mathrm{tr}_R(T^aT^bT^bT^a)(1 \pm v\cos\theta + v^2), \\
\mathcal{M}_s\bar{\mathcal{M}}_t &= \frac{ig^4}{d_R^2}f^{abc}\,\mathrm{tr}_R(T^cT^bT^a)(2 - v^2\sin^2\theta), \\
\mathcal{M}_s\bar{\mathcal{M}}_u &= -\frac{ig^4}{d_R^2}f^{abc}\mathrm{tr}_R(T^cT^aT^b)(2 - v^2\sin^2\theta), \\
\mathcal{M}_t\bar{\mathcal{M}}_u &= \frac{2g^4}{d_R^2}\mathrm{tr}_R(T^aT^bT^aT^b)v^2.
\end{aligned} \quad \text{(A6)}$$

Here, we have expanded around zero DM velocity as for annihilation to Higgs bosons.

We can evaluate the group theory factors using $f^{abc}f^{abd} = C_2(adj)\delta^{cd}$ and $T^aT^a = C_2(R)$, antisymmetry of the structure constants, and the group algebra. We find

$$\begin{aligned}
|\mathcal{M}_s|^2 &= \frac{g^4}{d_R}C_2(adj)C_2(R)\left(-\frac{19}{4} + \frac{1}{8}v^2(11 - 5\cos 2\theta)\right), \\
|\mathcal{M}_{t,u}|^2 &= \frac{2g^4}{d_R}C_2(R)^2(1 \pm v\cos\theta + v^2), \\
\mathcal{M}_s\bar{\mathcal{M}}_t &= \frac{g^4}{2d_R}C_2(adj)C_2(R)(2 - v^2\sin^2\theta), \\
\mathcal{M}_s\bar{\mathcal{M}}_u &= \frac{g^4}{2d_R}C_2(adj)C_2(R)(2 - v^2\sin^2\theta), \\
\mathcal{M}_t\bar{\mathcal{M}}_u &= \frac{2g^4}{d_R}(C_2(R)^2 + \frac{i}{2}C_2(adj)C_2(R))v^2.
\end{aligned} \quad \text{(A7)}$$

In the above discussion, we took the sum over gauge boson polarization vectors to give the metric for simplicity; this includes longitudinal and timelike polarizations as well as the transverse ones. In a non-Abelian gauge theory, the unphysical polarizations do not automatically vanish when contracted in the amplitudes, so we must correct for their inclusion. We can do this by *subtracting* the squared amplitude for ghost production. The amplitude is

$$\mathcal{M} = \frac{g^2}{s}\bar{v}_j T^c_{ji}\not{q}_a u_i f^{abc}, \quad \text{(A8)}$$

so we find

$$|\mathcal{M}|^2 = \frac{g^4}{8d_R}C_2(\mathrm{adj})C_2(R)(1 + v^2(1 - \cos^2\theta)). \quad \text{(A9)}$$

As an example of this effect, we can consider $\chi_1\chi_2 \to B_1B_2$ scattering (with fixed colors for $SU(2)$ triplet DM). Taking just the kinematical factors, the amplitude for annihilation into gauge bosons is given by $|\mathcal{M}_s + \mathcal{M}_u|^2 = 5g^4/4$, as in [36]. The ghosts subtract $g^4/8$ for a total of $|\mathcal{M}|^2 = 9g^4/8$, in agreement with the massless limit of the amplitude in the symmetry breaking phase.

We are primarily interested in $SU(2)$ triplet DM with triplet and quintuplet Higgs fields, and we now specialize to that case, assuming N_3 Higgs fields in the triplet and N_5 in the quintuplet. The total squared amplitude, with gauge and Higgs boson final states added incoherently, is

$$\begin{aligned}
|\mathcal{M}|^2 = \frac{g^4}{3}\Bigg[&\left(\frac{25}{2} + 2N_3 + 10N_5\right) + \frac{63}{2}v^2 \\
&+ \left(\frac{7}{2} - 2N_3 - 10N_5\right)v^2\cos^2\theta\Bigg].
\end{aligned} \quad \text{(A10)}$$

For identical nonrelativistic initial particles and identical massless final particles, the differential cross section is

$$\frac{d\sigma}{d\Omega} = \frac{1}{2}\frac{1}{64\pi^2 s}\frac{|\vec{q}_a|}{|\vec{p}_i|}|\mathcal{M}|^2, \quad \text{(A11)}$$

where the factor of $1/2$ is due to overcounting identical final states in the color sum. (Here we have assumed that the Higgs particles are in a real representation; if not, drop the factor of $1/2$ for the Higgs final states.)

So far we have worked at tree-level and in the CM frame, but there are small corrections to both approximations. First, the annihilation cross-section experiences Sommerfeld enhancement since the DM is nonrelativistic. Under the assumption that the gauge symmetry is not yet broken, the boost factor is

$$S(v) = \frac{\pi\alpha_g/v}{1 - e^{-\pi\alpha_g/2v}}, \qquad \alpha_g = \frac{g^2}{4\pi}. \quad \text{(A12)}$$

In the parameter space appropriate to the thermal DM freeze-out, $\pi\alpha_g/v < 1$, so we treat it as a small parameter and expand $S(v) \sim 1 + \pi\alpha_g/2v$. Next, since the CM frame is not quite the rest frame of the cosmic fluid, we must include the Lorentz transformation of the cross section. For nonrelativistic center-of-mass velocity $\vec{V}$,

this correction takes $\sigma \to \sigma(1 - V^2\sin^2\phi/2)$, where ϕ is the angle between $\vec{V}$ and $\vec{v}$.

Finally, we must average the cross section over the DM velocity distribution, which is Maxwell-Boltzmann. In terms of the $\vec{V}$ and $\vec{v}$, this average takes the form

$$\begin{aligned}\langle\sigma v_{\rm rel}\rangle &= \left(\frac{M_\chi}{2\pi T}\right)^3 \int d^3\vec{V} d^3\vec{v} (\sigma v_{\rm rel})(v) \\ &\quad \times \left(1 - \frac{1}{2}V^2\sin^2\phi\right) e^{-M_\chi(v^2+V^2)/T} \\ &= \frac{\pi}{12}\frac{\alpha_g^2}{M_\chi^2}\left[\left(\frac{25}{2} + 2N_3 + 10N_5\right)\left(1 + \alpha_g\sqrt{\frac{\pi M_\chi}{T}} - \frac{1}{2\pi}\frac{T}{M_\chi}\right)\right. \\ &\quad \left. + \left(\frac{317}{8} - \frac{5}{2}N_3 - \frac{25}{2}N_5\right)\frac{T}{M_\chi}\right], \end{aligned} \tag{A13}$$

where the relative velocity $v_{\rm rel} = 2v$. We have expanded the result to first order in the small parameters $\alpha_g\sqrt{M_\chi/T}$ and T/M_χ. In the sequel, we will treat the ratio M_χ/T as roughly independent of the DM mass and use an approximate value of $M_\chi/T \sim 20$.

We should also address the issue of symmetry breaking. We will assume that the gauge bosons are light compared to the DM and that Higgs bosons are either light compared to the DM or too heavy to be produced in DM annihilation. The annihilation amplitudes are slightly modified by the gauge boson masses (in a manner that does not respect the gauge symmetry, of course). For example, $\chi_1\chi_2 \to B_1B_2$ annihilation has

$$|\mathcal{M}|^2 = g^4\frac{\mu_1^4 + 2\mu_1^2(5\mu_2^2 - \mu_3^2) + (\mu_2^2 - \mu_3^2)^2}{8\mu_1^2\mu_2^2} \tag{A14}$$

at zero velocity and lowest order in the gauge boson masses. This deviates from the leading order massless result only in as much as the gauge boson masses differ from each other. Note that the Goldstone boson states become longitudinal gauge bosons in gauges such as the unitary gauge. The significant effect of symmetry breaking is to change the kinematical factors in the cross section. Using the average mass $\bar{\mu}$ for all the light gauge bosons and Higgs states,

$$\begin{aligned}|\vec{q}_a| &= M_\chi(1 + v^2/2) \to \sqrt{M_\chi^2(1+v^2) - \bar{\mu}^2} \\ &\cong M_\chi\left(1 - \frac{\bar{\mu}^2}{M_\chi^2}\right)^{1/2}\left(1 + \frac{M_\chi^2 v^2}{2(M_\chi^2 - \bar{\mu}^2)}\right). \end{aligned} \tag{A15}$$

For symmetry breaking at a small scale compared to the DM mass, this just renormalizes the cross section by a factor of $(1 - \bar{\mu}^2/M_\chi^2)^{1/2}$.

We can now compare the cross section to that required for the correct relic density of dark matter. Under the normal assumption that only SM particles are lighter than the dark matter, the required cross section is approximately $\langle\sigma v_{\rm rel}\rangle_0 \sim 2.84 \times 10^{-26}$ cm^3/s, with a logarithmic dependence on DM mass. We will take this central value. However, our dark matter models contain extra light degrees of freedom, which affects the required cross section in two ways. First, the Hubble parameter is greater at a given temperature, which causes freeze-out to occur earlier. Second, due to heating of photons by annihilation of these light degrees of freedom (see big bang nucleosynthesis constraints given in [36]), the Universe expands more between freeze-out and the present day, which means that the freeze-out density of DM must be higher than in minimal DM models. As a result, the desired cross section satisfies

$$\langle\sigma v_{\rm rel}\rangle = \frac{\sqrt{g_*}/g_{*s}}{(\sqrt{g_*}/g_{*s})_0}\langle\sigma v_{\rm rel}\rangle_0. \tag{A16}$$

For DM masses near $M_\chi = 5$ GeV, this is

$$\langle\sigma v_{\rm rel}\rangle = \frac{\langle\sigma v_{\rm rel}\rangle_0}{\sqrt{1 + (6 + 3N_3 + 5N_5)/61.75}}. \tag{A17}$$

Including all these corrections, we can write the desired cross section in terms of the gauge coupling as

$$\langle\sigma v_{\rm rel}\rangle = (A\alpha_g^2 + B\alpha_g^3)/M_\chi^2, \tag{A18}$$

which has the iterative solution

$$\alpha_g \cong \left(\frac{\langle\sigma v_{\rm rel}\rangle}{A + B\alpha_g}\right)^{1/2} M_\chi. \tag{A19}$$

Our $SU(2)$ models have

$$\begin{aligned}\alpha_g &\cong \frac{M_\chi}{\rm GeV}\left(1 - \frac{\bar{\mu}^2}{M_\chi^2}\right)^{-1/4} \\ &\quad \times \begin{cases} 2.5\times10^{-5} & (N_3 = 0, N_5 = 0; \quad \text{all} \quad m_h > M_\chi) \\ 2.2\times10^{-5} & (N_3 = 2, N_5 = 0) \\ 2.0\times10^{-5} & (N_3 = 3, N_5 = 0) \\ 1.7\times10^{-5} & (N_3 = 2, N_5 = 1) \\ 1.7\times10^{-5} & (N_3 = 3, N_5 = 1). \end{cases} \end{aligned} \tag{A20}$$

In the end, corrections due to the initial velocity of the DM particles contribute at the 5–15% level, while corrections from Sommerfeld enhancement contribute only 1 part in 10^4 due to their additional dependence on α_g (since the coefficient B is of the same order as A).

APPENDIX B: KINETIC COUPLING TO SM

In this appendix, we find the freeze-out temperature of the dark matter kinetic coupling to the standard model. For convenience calculating phase space factors, we consider downscattering $\chi_{2,3}f \to \chi_1 f$, labeling the χ_j momenta as p_j and the initial and final energy and momenta of the SM fermions f as $E_{i,f}$ and $q_{i,f}$ respectively (similarly for other subscripts). Here, the relevant initial DM state is χ_3 for the endothermic case and χ_2 for exothermic. In this appendix,

we consider the endothermic case, but it should be clear that these results apply equally well in both scenarios. At temperatures under 100 MeV (and, in particular near the important scale of $\delta M_{13} \sim$ MeV), only scattering from $e^{\pm}$ will be important. Furthermore, if this process occurs roughly once per Hubble time for each of the more massive DM particles, it efficiently maintains the distribution of these two DM states given by the Boltzmann factor at the SM temperature. Also, if $\chi\chi$ scattering is still in equilibrium, this reaction can insure that all the DM states maintain a thermal velocity distribution at the SM temperature. (As we have seen in the main text, $\chi\chi$ scattering typically freezes out later than this process.) We note that the same calculations apply for scattering between the two top states $\chi_3 f \to \chi_2 f$ with the appropriate replacements of ϵ_2, μ_2, and δM_{13}. As shown in Fig. 11, the cross section is only slightly smaller for keV mass splittings.

The matrix element for the scattering process shown in Fig. 10 is (taking electrons for specificity)

$$\mathcal{M} = -i\frac{ge\epsilon_2}{(p_3 - p_1)^2 - \mu_2^2}\bar{u}_1\gamma^\mu u_3 \bar{u}_f \gamma_\mu u_i. \tag{B1}$$

After the spin sum and average,

$$\begin{aligned}\frac{1}{4}\sum|\mathcal{M}|^2 = 4\frac{(ge\epsilon_2)^2}{(\mu_2^2 - t)^2}\Big[& s^2 + \frac{1}{2}t^2 + st \\ & - s(2m_e^2 + 2M_\chi^2 + 2M_\chi\delta M_{13} + \delta M_{13}^2) \\ & - \frac{1}{2}t\delta M_{13}^2 + M_\chi^4 + 2M_\chi^3\delta M_{13} \\ & + M_\chi^2(\delta M_{13}^2 + 2m_e^2) + 2M_\chi\delta M_{13}m_e^2 + m_e^4\Big].\end{aligned} \tag{B2}$$

With the replacement $m_e \to m_p$, $\delta M_{13} \to \delta M_{23}$, and $\mu_2 \to \mu_1$, this is also the result for inelastic DM scattering off protons, which is relevant to direct detection experiments.

Since we are interested in temperatures much less than the DM mass M_χ (and all other energy scales are also much less than M_χ), we can work in center-of-momentum frame up to overall errors of order $\sqrt{T/M_\chi} \ll 1$ in the cross section compared to the cosmic rest frame. To lowest order in T/M_χ, the final electron energy is

$$E_f = E_i + \delta M_{13}, \tag{B3}$$

or

$$|\vec{q}_f|^2 = \delta M_{13}^2 + 2\delta M_{13}E_i + |\vec{q}_i|^2. \tag{B4}$$

Then the Mandelstam t ranges between t_- and t_+ satisfying

$$\begin{aligned}t_\pm = & -2|\vec{q}_i|^2 - 2\delta M_{13}E_i \pm 2|\vec{q}_i| \\ & \times [\delta M_{13}^2 + 2\delta M_{13}E_i + |\vec{q}_i|^2]^{1/2}\end{aligned} \tag{B5}$$

to lowest order. Because of cancellations in (B2), we will need s to the same (second) order in small quantities:

$$\begin{aligned}s = & M_\chi^2 + 2M_\chi(\delta M_{13} + E_i) + \delta M_{13}^2 \\ & + 2\delta M_{13}E_i + m_e^2 + 2|\vec{q}_i|^2.\end{aligned} \tag{B6}$$

To lowest order in small quantities, we find

$$|\mathcal{M}|^2 = 16\frac{(ge\epsilon_2)^2M_\chi^2}{(\mu_2^2 - t)^2}\left[E_i^2 + \delta M_{13}E_i + \frac{1}{4}t\right] \tag{B7}$$

for the (spin averaged and summed) squared amplitude. Since $d\sigma/dt = (1/64\pi s)|\mathcal{M}|^2/|\vec{q}_i|^2$ and the relative velocity is dominated by the electron velocity, we find that

$$\begin{aligned}\sigma v_{\rm rel} = & \frac{(ge\epsilon_2)^2}{4\pi}\frac{1}{E_i|\vec{q}_i|}\Big[\left(E_i^2 + \delta M_{13}E_i + \frac{\mu_2^2}{4}\right) \\ & \times\left(\frac{1}{\mu_2^2 - t_+} - \frac{1}{\mu_2^2 - t_-}\right) - \frac{1}{4}\ln\left(\frac{\mu_2^2 - t_-}{\mu_2^2 - t_+}\right)\Big].\end{aligned} \tag{B8}$$

We are especially interested in whether the DM and SM can maintain kinetic equilibrium at temperatures $T \lesssim m_e$, since those temperatures are relevant for $\chi - \chi$ downscattering. Then temperatures are parametrically less than μ_2, so

$$\begin{aligned}\sigma v_{\rm rel} = & \frac{(ge\epsilon_2)^2}{2\pi}\frac{1}{\mu_2^4}\frac{1}{E_i}(2m_e^2 + |\vec{q}_i|^2 + \delta M_{13}E_i) \\ & \times(\delta M_{13}^2 + |\vec{q}_i|^2 + 2\delta M_{13}E_i)^{1/2}\end{aligned} \tag{B9}$$

times corrections of relative order t/μ_2^2 and $\sqrt{|\vec{q}_i|/M_\chi}$, where $v_{\rm rel}$ is the CM frame electron speed. At a fixed temperature T, we find the total scattering rate by integrating over the Fermi-Dirac distribution for marginally relativistic electrons (since $T \sim m_e$). Assuming a thermal origin for the dark matter abundance so $\alpha_g \propto M_\chi$, this total rate can be written in terms of a normalized rate $\hat{\Gamma}$ as

$$\langle n_e\sigma v_{\rm rel}\rangle \equiv \left(\frac{\epsilon_2^2M_\chi}{\mu_2^4}\right)\hat{\Gamma}(\delta M_{13}, T). \tag{B10}$$

In n_e, we include both spin states of electrons and positrons.

Therefore, the ratio $3H/\hat{\Gamma}$ considered as a function of T inverts to give the DM/SM decoupling temperature as a function of $\epsilon_2^2M_\chi/\mu_2^4$. This decoupling temperature is shown in Fig. 11 for several values of δM_{13}. Since we consider temperatures near the electron mass, we calculate the effective species number g_* numerically. This includes heating of photons due to $e^{\pm}$ annihilation and also the neutrino density.

We note in passing that the scattering $\chi_3 e^{\pm} \to \chi_2 e^{\pm}$ is not as effective at maintaining kinetic equilibrium between the DM and SM. First of all, for matching to direct detection experiments, we require ϵ_1 to be small, so that the criterion used above yields a large freeze-out temperature.

Second, since δM_{23} is keV scale, the cross section is also slightly smaller, leading to a larger freeze-out temperature. And last, since $\delta M_{23} \ll m_e$, each scattering event transfers only a small fraction of the DM particle's kinetic energy. Therefore, at the temperatures of interest, kinetic equilibrium requires many scatterings per Hubble time.

Since the kinetic mixing is between the B^a_μ bosons and SM hypercharge and therefore includes mixing with Z_μ, we can ask if the DM comes into equilibrium with neutrinos through the weak force. Above the electron mass, scattering from electrons will always dominate scattering from neutrinos because $m_Z \gg \mu_2$ (and because the B_μ coupling to the weak current is suppressed by $\sim \mu_2^2/m_Z^2$) [36]. However, below the electron mass, there are many more neutrinos than electrons, so neutrino scattering is potentially important. Scattering from neutrinos progresses through two Feynman diagrams similar to Fig. 10, one with a B propagator and one with a Z propagator, but there is a relative sign between the two in the kinetic mixing. Therefore, to include both diagrams properly, we should replace

$$\frac{1}{(\mu_2^2 - t)^2} \to \left[\frac{\mu_2^2/m_z^2}{\mu_2^2 - t} - \frac{1}{m_Z^2 - t}\right]^2 \sim \frac{t^2}{\mu_2^4 m_Z^4} \tag{B11}$$

in Eq. (B7) at low temperatures. Also, taking ϵ to be the B coupling to the electric current, we should replace $\epsilon \to \epsilon \tan\theta_w$. Because of the fact that the two Feynman diagrams nearly cancel, the cross section is highly suppressed. A straightforward estimate of the total scattering rate and comparison to the Hubble rate indicates that ϵ would need to be of order 10^5 for $\chi - \nu$ scattering to equilibrate at temperatures below m_e!

APPENDIX C: $\chi_2 \to \chi_1 e^+ e^-$ AT ONE-LOOP

In this appendix, we give a careful derivation of the lifetime for $\chi_2 \to \chi_1 e^+ e^-$ decay at the one-loop level when $\epsilon_3 = 0$. This is the dominant decay process we have been able to find for this case.[11]

Consider the loops in Fig. 16. As noted above the figure, the two diagrams nearly cancel due to the opposite signs of the non-Abelian 3-gauge-boson couplings; in fact, they do not cancel completely only because the two gauge bosons in the loop have different masses. The complete amplitude can be written as

$$\mathcal{M} = ig\frac{\bar{u}_1 \gamma^\mu u_2 \bar{u}_e L_\mu v_e}{q^2 - \mu_3^2}, \tag{C1}$$

where the spinors of the χ particles are labeled by their color (as their momenta $k_{2,1}$ will be), we take the outgoing momenta of the positron and electron to be $p_\pm$ respectively, and $q = k_2 - k_1 = p_+ + p_-$.

The momenta running counter clockwise around the loops are $l + \delta p$ on the electron line, $l - \bar{p}$ on the upper gauge boson line, and $l + \bar{p}$ on the lower gauge line, where $\bar{p} = q/2$ and $\delta p = (p_+ - p_-)/2$. With these conventions, the loop integrals are

$$\begin{aligned} L_\mu = ige^2\epsilon_1\epsilon_2 \int \frac{d^4 l}{(2\pi)^4}\left[\frac{\gamma^\nu(l + \delta\not{p} + m_e)\gamma^\lambda}{(l+\delta p)^2 - m_e^2}\right][\eta_{\mu\nu}(3\bar{p} - l)_\lambda \\ + \eta_{\nu\lambda}(2l)_\mu - \eta_{\mu\lambda}(l + 3\bar{p})_\nu] \\ \times\left[\frac{1}{[(l+\bar{p})^2 - \mu_2^2][(l-\bar{p})^2 - \mu_1^2]} - (\mu_1 \leftrightarrow \mu_2)\right]. \end{aligned} \tag{C2}$$

As usual, we can rewrite the denominators with Feynman parameters as

$$\begin{aligned} &\int \frac{dxdydz 2\delta(1 - x - y - z)}{[l^2 + 2l\cdot(\bar{p}(y-x) + \delta p z) + \bar{p}^2(x+y) + \delta p^2 z - \mu_1^2 x - \mu_2^2 y - m_e^2 z]^3} \\ &\quad = \int \frac{dxdydz 2\delta(1 - x - y - z)}{[l'^2 + \delta p^2 z(1-z) + \bar{p}^2(x(1-x) + y(1-y) + 2xy) - \mu_1^2 x - \mu_2^2 y - m_e^2 z]^3}, \end{aligned} \tag{C3}$$

where $l' = l + \bar{p}(y - x) + \delta p z$ and we have used $\bar{p}\cdot\delta p = 0$. From the second form, it is clear that taking $\mu_1 \leftrightarrow \mu_2$ is the same as swapping the Feynman parameters $x \leftrightarrow y$. Therefore, the only terms that survive taking the difference in (C2) must be antisymmetric in $x \leftrightarrow y$, and these must come from shifting l to l'.

In the end, we find

$$\begin{aligned} L_\mu &= -8(\bar{p}_\mu \delta\not{p} + \delta p_\mu \not{\bar{p}}) \int_0^1 dx \int_0^{1-x} dy \int \frac{d^4 l}{(2\pi)^4} \frac{(y-x)(1-x-y)}{[l^2 - \Delta(x,y)]^3} \\ &\cong \frac{-i}{12\pi^2}(\bar{p}_\mu \delta\not{p} + \delta p_\mu \not{\bar{p}})\left[\frac{1}{\mu_2^2 - \mu_1^2} + \frac{\mu_1^2 + \mu_2^2}{[\mu_2^2 - \mu_1^2]^2}\ln\left(\frac{\mu_1}{\mu_2}\right)\right], \end{aligned} \tag{C4}$$

[11] As this article went to press, we discovered other diagrams that dominate the amplitude for this decay. We estimate that the lifetime is 10^{23} using the parameters given under Eq. (53).

with $\Delta(x, y)$ given as in (C3). Note that this loop appears to generate an interaction with one extra derivative compared to a magnetic moment operator. In the approximation, we have taken the gauge boson masses to be much larger than any of the momenta in the denominator. We also have $\bar{u}_e \delta\!\!\!/ v_e = 2m_e$ and $\bar{u}_e \bar{p}\!\!\!/ v_e = 0$, so the amplitude is finally

$$\mathcal{M} \cong -\frac{8\alpha_g \alpha \epsilon_1 \epsilon_2 m}{3\mu_3^2} \bar{u}_1 \delta\!\!\!/ u_2 \bar{u}_e v_e \Lambda, \tag{C5}$$

where Λ is the function of $\mu_{1,2}$ in square brackets in (C4).

In the nonrelativistic limit, the spin-summed and averaged square amplitude is

$$|\mathcal{M}|^2 \cong \frac{1024}{9} \frac{\alpha_g^2 \alpha^2 \epsilon_1^2 \epsilon_2^2 m_e^2 M_\chi^2}{\mu_3^4} \times \Lambda^2 E_+ E_- (E_+ E_- - \vec{p}_+ \cdot \vec{p}_- - m_e^2), \tag{C6}$$

with $E_\pm$ and $\vec{p}_\pm$ the energy and 3-momentum of the $e^\pm$. Integrating over the nonrelativistic phase space, we find

$$\Gamma = \frac{16}{9\pi^2} \alpha^2 \alpha_g^2 \epsilon_1^2 \epsilon_2^2 \frac{m_e^5}{\mu_3^4} (\delta M_{12} - 2m_e)^3 (\delta M_{12} - m_e) \Lambda^2. \tag{C7}$$

Using the same estimated parameters as below Eq. (53), we find a lifetime of more than 10^{33} s.

[1] R. Diehl and M. Leising, Proc. Sci. INTEGRAL08 (2008) 001 [arXiv:0906.1503].

[2] J. Knödlseder *et al.*, Astron. Astrophys. **411**, L457 (2003); P. Jean *et al.*, Astron. Astrophys. **407**, L55 (2003); J. Knödlseder *et al.*, Astron. Astrophys. **441**, 513 (2005); L. Bouchet, E. Jourdain, J. P. Roques, A. Strong, R. Diehl, F. Lebrun, and R. Terrier, Astrophys. J. **679**, 1315 (2008).

[3] W. Wang, C. S. J. Pun, and K. S. Cheng, Astron. Astrophys. **446**, 943 (2006); M. Casse, B. Cordier, J. Paul, and S. Schanne, Astrophys. J. **602**, L17 (2004); G. Bertone, A. Kusenko, S. Palomares-Ruiz, S. Pascoli, and D. Semikoz, Phys. Lett. B **636**, 20 (2006); , J. D. Kurfess, R. L. Kinzer, and M. D. Leising, New Astron. Rev. **46**, 553 (2002); K. Ahn, E. Komatsu, and P. Hoflich, Phys. Rev. D **71**, 121301 (2005); R. M. Bandyopadhyay, J. Silk, J. E. Taylor, and T. J. Maccarone, Mon. Not. R. Astron. Soc. **392**, 1115 (2009); K. S. Cheng, D. O. Chernyshov, and V. A. Dogiel, Astrophys. J. **645**, 1138 (2006); A. Calvez and A. Kusenko, arXiv:1003.0045.

[4] J. C. Higdon, R. E. Lingenfelter, and R. E. Rothschild, Astrophys. J. **698**, 350 (2009); Phys. Rev. Lett. **103**, 031301 (2009).

[5] G. Weidenspointner *et al.*, Nature (London) **451**, 159 (2008).

[6] L. Bouchet and J.-P. Roques, E. Jourdain, Astrophys. J. **720**, 1772 (2010).

[7] C. Boehm, T. A. Ensslin, and J. Silk, J. Phys. G **30**, 279 (2004); C. Boehm and P. Fayet, Nucl. Phys. **B683**, 219 (2004); C. Boehm, D. Hooper, J. Silk, M. Casse, and J. Paul, Phys. Rev. Lett. **92**, 101301 (2004); C. Boehm, P. Fayet, and J. Silk, Phys. Rev. D **69**, 101302 (2004); D. Hooper, F. Ferrer, C. Boehm, J. Silk, J. Paul, N. W. Evans, and M. Casse, Phys. Rev. Lett. **93**, 161302 (2004); C. Boehm and Y. Ascasibar, Phys. Rev. D **70**, 115013 (2004); P. Fayet, Phys. Rev. D **70**, 023514 (2004); P. D. Serpico and G. G. Raffelt, Phys. Rev. D **70**, 043526 (2004); J. F. Beacom, N. F. Bell, and G. Bertone, Phys. Rev. Lett. **94**, 171301 (2005); K. Ahn and E. Komatsu, Phys. Rev. D **71**, 021303 (2005); **72**, 061301 (2005); Y. Rasera, R. Teyssier, P. Sizun, B. Cordier, J. Paul, M. Casse, and P. Fayet, Phys. Rev. D **73**, 103518 (2006); J. F. Gunion, D. Hooper, and B. McElrath, Phys. Rev. D **73**, 015011 (2006); P. Sizun, M. Casse, and S. Schanne, Phys. Rev. D **74**, 063514 (2006); C. Jacoby and S. Nussinov, J. High Energy Phys. 05 (2007) 017; Y. Kahn, M. Schmitt, and T. M. P. Tait, Phys. Rev. D **78**, 115002 (2008).

[8] J. F. Beacom and H. Yuksel, Phys. Rev. Lett. **97**, 071102 (2006).

[9] D. O. Chernyshov, K. S. Cheng, V. A. Dogiel, C. M. Ko, and W. H. Ip, arXiv:0912.0889.

[10] N. Prantzos *et al.*, arXiv:1009.4620.

[11] F. Chen, J. M. Cline, and A. R. Frey, Phys. Rev. D **79**, 063530 (2009).

[12] O. Adriani *et al.* (PAMELA Collaboration), Nature (London) **458**, 607 (2009).

[13] J. Chang *et al.*, Nature (London) **456**, 362 (2008).

[14] D. P. Finkbeiner, T. R. Slatyer, N. Weiner, and I. Yavin, J. Cosmol. Astropart. Phys. 09 (2009) 037.

[15] F. Chen, J. M. Cline, A. Fradette, A. R. Frey, and C. Rabideau, Phys. Rev. D **81**, 043523 (2010).

[16] D. P. Finkbeiner and N. Weiner, Phys. Rev. D **76**, 083519 (2007).

[17] M. Pospelov and A. Ritz, Phys. Lett. B **651**, 208 (2007).

[18] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, Phys. Rev. D **79**, 015014 (2009).

[19] P. W. Graham, R. Harnik, S. Rajendran, and P. Saraswat, Phys. Rev. D **82**, 063512 (2010); R. Essig, J. Kaplan, P. Schuster, and N. Toro, arXiv:1004.0691.

[20] R. Bernabei *et al.* (DAMA Collaboration), Eur. Phys. J. C **56**, 333 (2008); **67**, 39 (2010).

[21] C. E. Aalseth *et al.* (CoGeNT Collaboration), arXiv:1002.4703.

[22] D. Tucker-Smith and N. Weiner, Phys. Rev. D **64**, 043502 (2001); **72**, 063509 (2005); S. Chang, G. D. Kribs, D. Tucker-Smith, and N. Weiner, Phys. Rev. D **79**, 043513 (2009).

[23] B. Batell, M. Pospelov, and A. Ritz, Phys. Rev. D **79**, 115019 (2009).

[24] A. L. Fitzpatrick, D. Hooper, and K. M. Zurek, Phys. Rev. D **81**, 115005 (2010); S. Andreas, C. Arina, T. Hambye,

F. S. Ling, and M. H. G. Tytgat, Phys. Rev. D 82, 043522 (2010); S. Chang, J. Liu, A. Pierce, N. Weiner, and I. Yavin, arXiv:1004.0697; R. Foot, arXiv:1004.1424; K. J. Bae, H. D. Kim, and S. Shin, arXiv:1005.5131; Y. Mambrini, arXiv:1006.3318.

[25] D. Hooper, J. I. Collar, J. Hall, and D. McKinsey, Phys. Rev. D **82**, 123509 (2010).

[26] D. Hooper and L. T. Wang, Phys. Rev. D **70**, 063506 (2004); C. Picciotto and M. Pospelov, Phys. Lett. B **605**, 15 (2005).

[27] B. Batell, M. Pospelov, and A. Ritz, Phys. Rev. D **80**, 095024 (2009).

[28] J. F. Navarro *et al.*, Mon. Not. R. Astron. Soc. **402**, 21 (2010).

[29] E. Romano-Diaz, I. Shlosman, Y. Hoffman, and C. Heller, Astrophys. J. **685** L105 (2008); M. G. Abadi, J. F. Navarro, M. Fardal, A. Babul, and M. Steinmetz, arXiv:0902.2477; S. E. Pedrosa, P. B. Tissera, and C. Scannapieco, arXiv:0910.4380.

[30] P. B. Tissera, S. D. M. White, S. Pedrosa, and C. Scannapieco, Mon. Not. R. Astron. Soc. **406**, 922 (2010); arXiv:0911.2316.

[31] M. Pato, O. Agertz, G. Bertone, B. Moore, and R. Teyssier, Phys. Rev. D **82**, 023531 (2010).

[32] P. Salucci, F. Nesti, G. Gentile, and C. F. Martins, Astron. Astrophys. **523**, A83 (2010).

[33] Z. Abidin, A. Afanasev, and C. E. Carlson, arXiv:1006.5444.

[34] N. Prantzos, Astron. Astrophys. **449**, 869 (2006); New Astron. Rev. **52**, 457 (2008).

[35] P. Jean, W. Gillard, A. Marcowith, and K. Ferriere, arXiv:0909.4022.

[36] F. Chen, J. M. Cline, and A. R. Frey, Phys. Rev. D **80**, 083516 (2009).

[37] A. Sommerfeld, Ann. Phys. (Leipzig) **403**, 257 (1931); J. Hisano, S. Matsumoto, and M. M. Nojiri, Phys. Rev. Lett. **92**, 031303 (2004); J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, Phys. Rev. D **71**, 063528 (2005); M. Cirelli, A. Strumia, and M. Tamburini, Nucl. Phys. **B787**, 152 (2007); M. Lattanzi and J. I. Silk, Phys. Rev. D **79**, 083523 (2009); see also previous work in K. Belotsky, D. Fargion, M. Khlopov, and R. V. Konoplich, Phys. At. Nucl. **71**, 147 (2008) and references therein.

[38] J. L. Feng, M. Kaplinghat, and H. B. Yu, Phys. Rev. D **82**, 083525 (2010).

[39] T. Slatyer (private communication).

[40] F. D'Eramo and J. Thaler, J. High Energy Phys. 06 (2010) 109.

[41] Kolb and Turner, *The Early Universe* (Addison-Wesley, Reading, MA, 1988).

[42] F. J. Kerr and D. Lynden-Bell, Mon. Not. R. Astron. Soc. **221**, 1023 (1986).

[43] M. Cirelli and J. M. Cline, Phys. Rev. D **82**, 023503 (2010).

[44] E. Churazov, R. Sunyaev, S. Sazonov, M. Revnivtsev, and D. Varshalovich, Mon. Not. R. Astron. Soc. **357**, 1377 (2005).

[45] P. Jean *et al.*, Astron. Astrophys. **445**, 579 (2006).

[46] N. Guessoum, P. Jean, and W. Gillard, arXiv:astro-ph/0504186.

[47] K. Ferriere, W. Gillard, and P. Jean, Astron. Astrophys. **467**, 611 (2007).

[48] T. Sawada, T. Hasegawa, T. Handa, and R. J. Cohen, Mon. Not. R. Astron. Soc. **349**, 1167 (2004).

[49] H. S. Liszt and W. B. Burton, Astrophys. J. **236**, 779 (1980).

[50] J. M. Cordes and T. J. W. Lazio, arXiv:astro-ph/0207156.

[51] V. Kaspi (private communication).

[52] G. K. Skinner, arXiv:1009.2098.

[53] E. Churazov, S. Sazonov, S. Tsygankov, R. Sunyaev, and D. Varshalovich, arXiv:1010.0864.

[54] B. J. Teegarden and K. Watanabe, Astrophys. J. **646**, 965 (2006).

[55] H. Yuksel and M. D. Kistler, Phys. Rev. D **78**, 023502 (2008).

[56] M. P. Muno *et al.*, Astrophys. J. **613**, 326 (2004).

[57] M. Baumgart, C. Cheung, J. T. Ruderman, L. T. Wang, and I. Yavin, J. High Energy Phys. 04 (2009) 014.

[58] C. Arina and M. H. G. Tytgat, J. Cosmol. Astropart. Phys. 01 (2011) 011.

[59] J. Lavalle, Phys. Rev. D **82**, 081302 (2010).

[60] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, Phys. Rev. D **80**, 023505 (2009); T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, Phys. Rev. D **80**, 043526 (2009).

[61] M. Cirelli, F. Iocco, and P. Panci, J. Cosmol. Astropart. Phys. 10 (2009) 009.

[62] M. Pospelov and J. Pradler, Phys. Rev. D **82**, 103514 (2010).

[63] R. Mertig, M. Bohm, and A. Denner, Comput. Phys. Commun. **64**, 345 (1991).

[64] J. D. Bjorken, R. Essig, P. Schuster, and N. Toro, Phys. Rev. D **80**, 075018 (2009); R. Essig, P. Schuster, N. Toro, and B. Wojtsekhowski, J. High Energy Phys. 02 (2011) 009.

[65] K. Nakamura *et al.*, J. Phys. G **37**, 075021 (2010).