

Sommerfeld enhancements for thermal relic dark matter

Jonathan L. Feng, Manoj Kaplinghat, and Hai-Bo Yu

Department of Physics and Astronomy, University of California, Irvine, California 92697, USA

(Received 17 August 2010; published 26 October 2010)

The annihilation cross section of thermal relic dark matter determines both its relic density and indirect detection signals. We determine how large indirect signals may be in scenarios with Sommerfeld-enhanced annihilation, subject to the constraint that the dark matter has the correct relic density. This work refines our previous analysis through detailed treatments of resonant Sommerfeld enhancement and the effect of Sommerfeld enhancement on freeze out. Sommerfeld enhancements raise many interesting issues in the freeze out calculation, and we find that the cutoff of resonant enhancement, the equilibration of force carriers, the temperature of kinetic decoupling, and the efficiency of self-interactions for preserving thermal velocity distributions all play a role. These effects may have striking consequences; for example, for resonantly-enhanced Sommerfeld annihilation, dark matter freezes out but may then chemically *recouple*, implying highly suppressed indirect signals, in contrast to naive expectations. In the minimal scenario with standard astrophysical assumptions, and tuning all parameters to maximize the signal, we find that, for force-carrier mass $m_\phi = 250$ MeV and dark matter masses $m_X = 0.1, 0.3,$ and 1 TeV, the maximal Sommerfeld enhancement factors are $S_{\text{eff}} = 7, 30,$ and $90,$ respectively. Such boosts are too small to explain both the PAMELA and Fermi excesses. Nonminimal models may require smaller boosts, but the bounds on S_{eff} could also be more stringent, and dedicated freeze out analyses are required. For concreteness, we focus on 4μ final states, but we also discuss $4e$ and other modes, deviations from standard astrophysical assumptions and nonminimal particle physics models, and we outline the steps required to determine if such considerations may lead to a self-consistent explanation of the PAMELA or Fermi excesses.

DOI: 10.1103/PhysRevD.82.083525

PACS numbers: 95.35.+d, 95.85.Ry

I. INTRODUCTION

Dark matter may be composed of thermal relics with mass near the weak scale $m_{\text{weak}} \sim 100$ GeV–1 TeV. Such candidates are produced in the hot early Universe and then freeze out when the Universe cools and expands. Their annihilation cross section therefore determines both the relic density and the rate of annihilation today. The requirement that the candidate be much or all of the dark matter therefore constrains its annihilation rate now, with important implications for indirect searches for dark matter.

Although annihilation in the early Universe and now is determined by the same dynamics, the kinematics are vastly different: at freeze out, thermal relics have relative velocity $v_{\text{rel}} \sim 0.3,$ whereas today, $v_{\text{rel}} \sim 10^{-3}.$ As a result, the numerical values of the annihilation cross sections may differ significantly. For example, neutralino annihilation is in many cases dominated by P -wave processes [1], and so it is suppressed at low velocities. In this work, we consider Sommerfeld-enhanced cross sections, which have the opposite behavior: they are enhanced at low velocities and therefore boost present-day annihilation signals. We refine a previous study [2] and examine how large these boosts may be, subject to the constraint that the thermal relic has the correct thermal relic density to be dark matter.

Along with its general implications for future indirect searches, this study also has direct implications for the interpretation of current data. Following earlier excesses

reported by the HEAT Collaboration [3,4], the PAMELA [5], ATIC [6], and Fermi [7] Collaborations have reported excesses of cosmic positrons over an estimate of expected background [8]. Dark matter annihilation with Sommerfeld enhancement [9] of dark matter annihilation [10], generalized to massive force carriers [11–15], has been proposed as an explanation [16–19]. In its simplest form, this scenario assumes that dark matter is composed of a single particle species $X,$ which interacts with light force carriers ϕ with fine-structure constant α_X and $m_\phi \ll m_X \sim m_{\text{weak}}.$ This new interaction modifies dark matter annihilation and scattering properties. The resulting annihilation cross section multiplied by relative velocity is $(\sigma_{\text{an}} v_{\text{rel}})_0 S,$ where $(\sigma_{\text{an}} v_{\text{rel}})_0$ is its tree-level value, and S is Sommerfeld's original enhancement factor [9]

$$S^0 = \frac{\pi\alpha_X/v}{1 - e^{-\pi\alpha_X/v}} \xrightarrow{\alpha_X \gg v} \frac{\pi\alpha_X}{v}, \quad (1)$$

generalized to massive $\phi,$ where $v \equiv v_{\text{rel}}/2$ is the dark matter particle's velocity in the center-of-mass frame. In the proposed explanation, dark matter freezes out at early times when the Sommerfeld effect is negligible with $(\sigma_{\text{an}} v_{\text{rel}})_0 \approx 3 \times 10^{-26}$ cm³/s, leading to the correct relic density $\Omega_{\text{DM}} h^2 \approx 0.114.$ At present, however, when v_{rel} is much smaller, the Sommerfeld effect becomes important, and, for example, for $m_X \sim 2$ TeV and an assumed enhancement factor of $S \sim 1000,$ such annihilations are sufficient to explain the positron excesses [20,21].

In Ref. [2], we showed that in straightforward models, such large Sommerfeld enhancements cannot be self-consistently realized. The problem is simple to state: large Sommerfeld enhancement requires strong interactions, and strongly-interacting particles annihilate too efficiently in the early Universe to be all of the dark matter. More quantitatively, the required new force-carrier interaction necessarily induces an annihilation process $XX \rightarrow \phi\phi$, with cross section

$$(\sigma_{\text{an}}v_{\text{rel}})_0 \sim \frac{\pi\alpha_X^2}{m_X^2}. \quad (2)$$

Conservatively neglecting other annihilation processes, for a typical weak-scale mass m_X , the thermal relic density implies a typical weak coupling $\alpha_X \approx 0.05[m_X/2 \text{ TeV}]$, as dictated by the WIMP miracle. The resulting Sommerfeld enhancement now is $\sim \pi\alpha_X/v \approx 100[m_X/2 \text{ TeV}]$, an order of magnitude too small to explain the PAMELA and Fermi signals. Alternatively, to achieve $S \sim 1000$, one requires $\alpha_X \sim 1$, which implies $\Omega_X \sim 0.001$, 2 orders of magnitude smaller than the value assumed in deriving the requirement $S \sim 1000$.

In this work, we refine our previous analysis in several ways. In Ref. [2] we approximated the Sommerfeld enhancement S by its value at $m_\phi = 0$, given by S^0 in Eq. (1). For massive ϕ , the Sommerfeld enhancement cuts off at a value proportional to m_X/m_ϕ and also exhibits resonant structure. Here we use numerical results (and a highly accurate analytic approximation to the numerical results [22–25]) for S .

In addition, we refine our previous work to include the effect of Sommerfeld enhancement on freeze out. One might expect this effect to be negligible, because freeze out is typically thought to occur at $v \sim 0.3$, when the Sommerfeld effect is insignificant. However, annihilation continues to much later times, and Sommerfeld enhancement has an impact when the dark matter cools. Sommerfeld effects on freeze out have been considered previously in Ref. [26] for the case of Wino dark matter and in Refs. [27–30] for hidden sector dark matter. In the analyses published after the PAMELA and Fermi excesses were reported [28–30], $(\sigma_{\text{an}}v_{\text{rel}})_0$ was taken as a free parameter, and the fact that it depends on the same parameters that determine S was not used. Here we make essential use of the observation that there is an irreducible contribution to $(\sigma_{\text{an}}v_{\text{rel}})_0$ of the form of Eq. (2), and so constraints on the relic density bound the fundamental parameters that also determine S .

As we will see, the combination of these two refinements highlights a number of effects that may typically be ignored, but now require exploration. For example, the relic density is in principle affected by the cutoff of resonant Sommerfeld enhancement, the production and decay rates of force-carrier particles, the temperature of kinetic decoupling, and the efficiency of self-scattering for

preserving the dark matter’s thermal velocity distribution. Very close to resonances, we also find the intriguing possibility of chemical *recoupling*, in which dark matter freezes out and then melts back in, with annihilations becoming important again at late times. This implies that exact resonances suppress, rather than enhance, indirect signals.

We examine the quantitative impact of all of these effects, and present results for the maximal Sommerfeld enhancement of indirect signals. In the most optimistic case, that is, tuning all parameters to maximize the signal, we find that, for $m_\phi = 250 \text{ MeV}$ and $m_X = 0.1, 0.3$, and 1 TeV , the largest possible effective Sommerfeld enhancements are $S_{\text{eff}} = 7, 30$, and 90 , respectively. This refined analysis therefore strengthens our previous results: Sommerfeld enhancements may be significant, but they fall short of explaining the current PAMELA and Fermi cosmic ray excesses. We then discuss various astrophysical effects that may reduce the discrepancy between the maximal enhancements derived here and those required to explain the data, and critically examine several nonminimal models and the issues that must be addressed to determine if more complicated particle physics models may provide viable explanations.

This paper is organized as follows. In Sec. II, we describe our underlying model assumptions and our treatment of Sommerfeld enhancements with resonances. In Sec. III we analyze the effect of Sommerfeld enhancement on freeze out. In Sec. IV we assemble these pieces and present the results for the maximal Sommerfeld enhancement achievable now. In Sec. V, we compare these to the current data, and discuss nonstandard astrophysical effects and nonminimal particle physics models. We present our conclusions in Sec. VI.

II. SOMMERFELD ENHANCEMENT WITH RESONANCES

We consider a simple model with a hidden sector dark matter particle X , which couples to a light force carrier ϕ with coupling $\sqrt{4\pi}\alpha_X$. The annihilation cross section is then $(\sigma_{\text{an}}v_{\text{rel}})_0 S$, where $(\sigma_{\text{an}}v_{\text{rel}})_0$ is the tree-level cross section and S is the Sommerfeld enhancement.

To maximize the Sommerfeld enhancement, we assume that $(\sigma_{\text{an}}v_{\text{rel}})_0$ is dominated by S -wave processes, and so is unsuppressed at low velocities. We also consider only the “irreducible” annihilation channel $XX \rightarrow \phi\phi$, and take the tree-level cross section

$$(\sigma_{\text{an}}v_{\text{rel}})_0 = \frac{\pi\alpha_X^2}{m_X^2}. \quad (3)$$

This may be modified by $\mathcal{O}(1)$ prefactors, depending, for example, on whether X is a Majorana or Dirac fermion, and whether ϕ is a scalar or a gauge boson. In addition, even in simple models, $(\sigma_{\text{an}}v_{\text{rel}})_0$ will typically receive additional contributions from other annihilation channels;

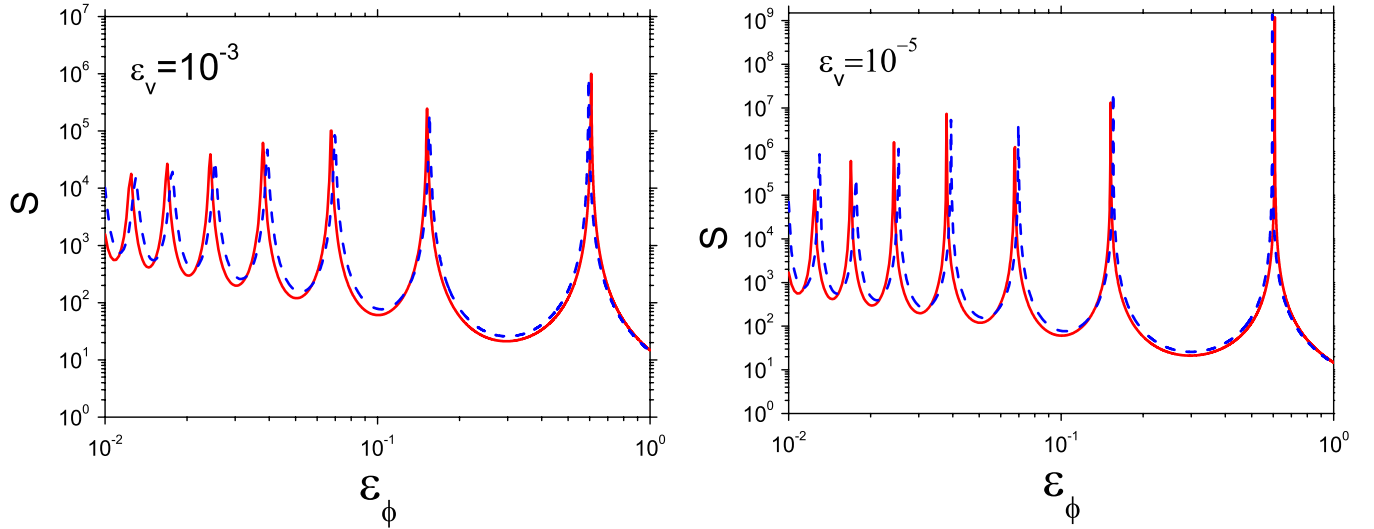


FIG. 1 (color online). The Sommerfeld enhancement factor S as a function of $\epsilon_\phi \equiv m_\phi/(\alpha_X m_X)$ for the constant values of $\epsilon_v \equiv v/\alpha_X$ indicated. The solid red curves are the analytic approximation of Eq. (6), and the dashed blue curves are numerical results.

to maximize the Sommerfeld effect on indirect search signals, we neglect these other channels here, but discuss their impact in Sec. V.

To determine the enhancement factor S , we numerically solve the differential equation

$$\frac{1}{m_X} \frac{d^2 \chi}{dr^2} + \frac{\alpha_X}{r} e^{-m_\phi r} \chi = -m_X v^2 \chi, \quad (4)$$

with the boundary conditions $\chi'(r) = im_X v \chi(r)$ and $\chi(r) = e^{im_X v r}$ when $r \rightarrow \infty$. The Sommerfeld enhancement factor is given by

$$S = \frac{|\chi(\infty)|^2}{|\chi(0)|^2}. \quad (5)$$

The Sommerfeld enhancement may also be obtained by approximating the Yukawa potential by the Hulthen potential, for which an analytic solution is possible [24]. The resulting analytic approximation to the Sommerfeld enhancement is [24,25]

$$S = \frac{\pi}{\epsilon_v} \frac{\sinh(\frac{2\pi\epsilon_v}{\pi^2\epsilon_\phi/6})}{\cosh(\frac{2\pi\epsilon_v}{\pi^2\epsilon_\phi/6}) - \cos(2\pi\sqrt{\frac{1}{\pi^2\epsilon_\phi/6} - \frac{\epsilon_v^2}{(\pi^2\epsilon_\phi/6)^2}})}, \quad (6)$$

where $\epsilon_v \equiv v/\alpha_X$ and $\epsilon_\phi \equiv m_\phi/(\alpha_X m_X)$. The analytic expression of Eq. (6) is compared to the numerical solution in Fig. 1. We see that the analytic result is an excellent approximation, typically reproducing the numerical results to within fractional differences of 10%, and accurately reproducing the resonant behavior. Given Eq. (6), we see that for $\epsilon_v \ll \epsilon_\phi$, these resonances are at

$$m_\phi \simeq \frac{6\alpha_X m_X}{\pi^2 n^2}, \quad n = 1, 2, 3, \dots \quad (7)$$

At these resonances, with m_ϕ determined by Eq. (7), the Sommerfeld enhancement for low v is

$$S \simeq \frac{\pi^2 \alpha_X m_\phi}{6m_X v^2}. \quad (8)$$

S is therefore enhanced by v^{-2} at resonances, as opposed to v^{-1} away from resonances.

Note, however, that both the numerical and analytic results become infinite on resonance. This is unphysical, a result of the fact that the quantum mechanical treatments do not include the effect of bound state decay. In fact, the finite lifetime of bound states implies that S saturates at $v \sim \alpha_X^3 (m_\phi/m_X)$ [13], which, given the condition for a strong resonance, Eq. (7) with small n , is $v \sim \alpha_X^4$. In this study, we use the analytic form for S given in Eq. (6) and include the saturation by making the substitution $v \rightarrow v + \alpha_X^4$.

III. THERMAL FREEZE OUT WITH SOMMERFELD ENHANCEMENT

A. General formalism

Given values of m_X , m_ϕ , and α_X , we must determine the relic density Ω_X . The formalism of thermal freeze out is well developed [31,32]. Here we review this formalism in sufficient generality to accommodate novel effects resulting from Sommerfeld enhancements.

The evolution of the abundance of a thermal relic X is governed by the Boltzmann equation

$$\frac{dn_X}{dt} + 3Hn_X = -\langle\sigma_{\text{an}} v_{\text{rel}}\rangle(n_X^2 - n_X^{\text{eq}2}), \quad (9)$$

where n_X is the number density of the dark matter particles, n_X^{eq} is its value in equilibrium, H is the Hubble constant,

and $\langle\sigma_{\text{an}}v_{\text{rel}}\rangle$ is the annihilation cross section multiplied by the relative velocity, averaged over the dark matter velocity distribution. Changing variables from $t \rightarrow x = m_X/T$ and $n_X \rightarrow Y = n_X/s$, where s is the entropy density, the Boltzmann equation becomes

$$\frac{dY}{dx} = -\sqrt{\frac{\pi}{45}}m_{\text{Pl}}m_X\frac{(g_{*s}/\sqrt{g_*})}{x^2}\langle\sigma_{\text{an}}v_{\text{rel}}\rangle(Y^2 - Y^{\text{eq}2}), \quad (10)$$

where $m_{\text{Pl}} = G_N^{-1/2} \simeq 1.2 \times 10^{19}$ GeV is the Planck mass, and g_{*s} and g_* are the effective relativistic degrees of freedom for entropy and energy density, respectively. In this section we will assume that the particles X annihilates to are in thermal equilibrium at the time of freeze out. We will revisit this issue in Sec. III B, where we show that this requirement leads to significant constraints if the dominant annihilation is to the dark force carriers.

To evaluate the relic density, we first evolve it to the time of chemical decoupling, when the annihilation rate $\Gamma_{\text{an}} \equiv n_X\langle\sigma_{\text{an}}v_{\text{rel}}\rangle$ is approximately the expansion rate H . As discussed in Sec. II, we assume the annihilation cross section has the form $\sigma_{\text{an}}v_{\text{rel}} = (\sigma_{\text{an}}v_{\text{rel}})_0 S$, where $(\sigma_{\text{an}}v_{\text{rel}})_0$ is the S -wave, tree-level cross section, and S is the Sommerfeld enhancement. Up to the time of freeze out, the Sommerfeld effect is insignificant, and so the standard results for S -wave annihilators apply: defining freeze out by

$$Y(x_f) = (1 + c)Y^{\text{eq}}(x_f), \quad (11)$$

where c is a constant, freeze out occurs at

$$\begin{aligned} x_f &\approx \ln\xi - \frac{1}{2}\ln(\ln\xi) \\ \xi &= 0.038c(2 + c)m_{\text{Pl}}m_X(g/\sqrt{g_*})(\sigma_{\text{an}}v_{\text{rel}})_0, \end{aligned} \quad (12)$$

where g is the number of degrees of freedom of the dark matter particle; we take $g = 2$. To match numerical results, $c \sim 1$; we choose $c = 1$ and have checked that varying c from $1/2$ to 2 has no appreciable effect. To further test this approximation we have also numerically evolved Eq. (10) from $x = 20$ to $x = 100$ (when the equilibrium abundance is effectively zero) and then used Eq. (12) to evolve into the late Universe. This numerical calculation typically results in 10% smaller relic abundance. Given the excellent agreement, we set $c = 1$ and with this choice, Eqs. (11) and (12) determine $Y(x_f)$, the abundance at freeze out.

After freeze out, Y^{eq} quickly becomes insignificant. Neglecting it, we may solve Eq. (10) to find

$$\begin{aligned} \frac{1}{Y(x_s)} &= \frac{1}{Y(x_f)} + \sqrt{\frac{\pi}{45}}m_{\text{Pl}}m_X \int_{x_f}^{x_{\text{kd}}} \frac{(g_{*s}/\sqrt{g_*})\langle\sigma_{\text{an}}v_{\text{rel}}\rangle}{x^2} dx \\ &+ \sqrt{\frac{\pi}{45}}m_{\text{Pl}}m_X \int_{x_{\text{kd}}}^{x_s} \frac{(g_{*s}/\sqrt{g_*})\langle\sigma_{\text{an}}v_{\text{rel}}\rangle}{x^2} dx, \end{aligned} \quad (13)$$

where x_{kd} is the value of x at kinetic decoupling and x_s is its value when annihilations become insignificant and we may stop the integration. We have broken the integral into two to emphasize that there are two eras: before and after the

temperature of kinetic decoupling T_{kd} . Before T_{kd} , the dark matter's velocity distribution is thermal with temperature $T_X = T$. After T_{kd} , the dark matter's velocity distribution initially remains thermal, but T_X drops as a^{-2} , while T drops as a^{-1} , where a is the scale factor, and so $T_X = T^2/T_{\text{kd}}$. Eventually, the dark matter's velocity distribution need not even be thermal. We discuss the value of T_{kd} and the issue of nonthermal velocity distributions in Secs. III C and III D, respectively.

When the dark matter distribution is thermal with temperature T_X , the thermally-averaged cross section in the nonrelativistic limit is

$$\begin{aligned} \langle\sigma_{\text{an}}v_{\text{rel}}\rangle &\equiv \int f(\vec{v}_1)f(\vec{v}_2)\sigma_{\text{an}}v_{\text{rel}}d^3\vec{v}_1d^3\vec{v}_2 \\ &= \int \sqrt{\frac{2}{\pi}}\frac{1}{v_0^3}v_{\text{rel}}^2e^{-(v_{\text{rel}}^2/2v_0^2)}\sigma_{\text{an}}v_{\text{rel}}dv_{\text{rel}}, \end{aligned} \quad (14)$$

where $f(\vec{v})$ is the dark matter's velocity distribution, $v_{\text{rel}} = |\vec{v}_1 - \vec{v}_2|$, and

$$v_0 = \sqrt{\frac{2T_X}{m_X}} \equiv \sqrt{\frac{2}{x_X}}, \quad (15)$$

the most probable velocity. For the Sommerfeld-enhanced annihilation cross section, this becomes

$$\begin{aligned} \langle\sigma_{\text{an}}v_{\text{rel}}\rangle &= \frac{x_X^{3/2}}{2\sqrt{\pi}} \int_0^\infty (\sigma_{\text{an}}v_{\text{rel}})v_{\text{rel}}^2e^{-x_Xv_{\text{rel}}^2/4}dv_{\text{rel}} \\ &= (\sigma_{\text{an}}v_{\text{rel}})_0\bar{S}(x_X), \end{aligned} \quad (16)$$

where

$$\bar{S}(x_X) = \frac{x_X^{3/2}}{2\sqrt{\pi}} \int_0^\infty S v_{\text{rel}}^2 e^{-x_X v_{\text{rel}}^2/4} dv_{\text{rel}} \quad (17)$$

is the Sommerfeld enhancement averaged over a thermal distribution with temperature $T_X = m_X/x_X$.

B. Equilibration of force carriers

In the relic density calculation, we have implicitly assumed that, when the X particles chemically decouple, the force carriers ϕ are in thermal equilibrium with the massless particles of the standard model. The force carriers are expected to interact with the standard model, as their decays to standard model particles typically provide the indirect signals. In a simple example, if ϕ is a $U(1)$ gauge boson, it may mix with the standard model photon through kinetic mixing terms $\sim \epsilon F_{\mu\nu}^{\text{EM}} F^{\phi\mu\nu}$, where $F_{\mu\nu}^{\text{EM}}$ is the standard model's electromagnetic field strength tensor. After diagonalizing the ϕ -photon system, standard model particles with charge Q have hidden charge ϵQ [33], and so the ϕ particles decay through $\phi \rightarrow f\bar{f}$, where $f = e, \mu, u, d, s, \dots$. The largest kinetic mixing parameter allowed by current particle physics constraints is $\epsilon \sim 10^{-3}$ [34,35].

The existence of ϕ interactions with the visible sector does not, however, guarantee that they are efficient enough to bring the ϕ particles in thermal equilibrium. Here we determine sufficient conditions for the kinetic mixing example to guarantee the equilibration of ϕ particles. The leading ϕ number-changing interactions between the ϕ particles and the visible sector are decays $\phi \rightarrow f\bar{f}$ and inverse decays $f\bar{f} \rightarrow \phi$. Other ϕ number-changing processes, such as $f\bar{f} \rightarrow \phi\gamma$, $q\bar{q} \rightarrow \phi g$, and $f\gamma \rightarrow f\phi$, are parametrically suppressed compared to these and subdominant. For temperatures $T \gg m_\phi$, the decay and inverse decay processes balance (since f, \bar{f} are in chemical equilibrium), and we may simply compare the decay rate to the expansion rate [36].

For $T \gg m_\phi$, the thermally-averaged decay rate for $\phi \rightarrow f\bar{f}$ in the lab frame is $\langle \Gamma_\phi \rangle \simeq (\epsilon^2/3) \sum_f (Q_f^2 N_c^f) \times \alpha_{\text{EM}} m_\phi (m_\phi/T)$, where we have averaged over three ϕ polarizations, Q_f and N_c^f are the standard model charge and number of colors for fermion f , the sum is over all fermions with mass $m_f < m_\phi/2$, and the last factor of m_ϕ/T is from time dilation. Note that the decay process is inefficient at early times and high temperatures, but as T drops, $\langle \Gamma_\phi \rangle$ increases and H decreases, and so the ϕ particles may come into thermal equilibrium with the standard model. The expansion rate is $H \simeq 1.66 g_*^{1/2} T^2/m_{\text{Pl}}$; at freeze out, $g_* \simeq 100$. Requiring $\langle \Gamma_\phi \rangle \gtrsim H$ at $T_f \simeq m_X/25$ yields $\epsilon \gtrsim 3 \times 10^{-6}$ for $m_\phi = \text{GeV}$ and $m_X = \text{TeV}$.

If the above constraint is not satisfied, then the hidden sector may be at a different temperature from the standard model, modifying the dark matter freeze out. Such scenarios have been considered in Ref. [37], and the difference in temperatures has a significant impact on the dark matter relic density. These considerations may be important for some models, but here we assume that Eq. (18) is satisfied and $m_\phi \ll T_f \simeq m_X/25$, so that the ϕ particles are in thermal contact with the standard model at freeze out and their number density is close to the thermal prediction.

We note here that the calculation above does not guarantee chemical equilibrium. For the ϕ particles to produce X through $\phi\phi \rightarrow XX$, we need the tail of the ϕ distribution to be populated thermally. To check this, we approximate the decay rate for ϕ with energies of order m_X as $(\epsilon^2/3) \sum_f (Q_f^2 N_c^f) \alpha_{\text{EM}} m_\phi (m_\phi/m_X)$. If this decay rate is larger than the expansion rate, then the inverse process will also be in equilibrium and hence thermally populate the tail of the ϕ distribution. Comparing this to the expansion rate at $T_f \simeq m_X/25$ we obtain that

$$\epsilon \gtrsim 1.5 \times 10^{-5} \left[\frac{4}{\sum_f Q_f^2 N_c^f} \right]^{1/2} \left[\frac{\text{GeV}}{m_\phi} \right] \left[\frac{m_X}{\text{TeV}} \right]^{3/2}. \quad (18)$$

In setting the bound above, we have not considered scattering interactions with standard model fermions (and the resulting changes in kinetic energy for ϕ) because these

processes are proportional to ϵ^4 and hence considerably slower than the inverse decay process.

To summarize, for the ϕ particles to be in thermal equilibrium with the standard model at freeze out requires that Eq. (18) is satisfied and $m_\phi \ll T_f \simeq m_X/25$. We will assume these conditions hold in our analysis, but we note that they do not necessarily hold. In particular, if $m_\phi \sim T_f$, the number of ϕ particles at freeze out is reduced, which reduces the X relic density, strengthening the bounds on S_{eff} we determine below.

C. Kinetic decoupling of dark matter

After freeze out, dark matter particles remain kinematically coupled to the thermal bath through elastic scattering. Kinetic decoupling occurs later, when the momentum transfer rate drops below the Hubble expansion rate. We define the kinetic decoupling temperature T_{kd} by $\Gamma_k(T_{\text{kd}}) = H(T_{\text{kd}})$, where Γ_k is the momentum transfer rate. It may be approximated as

$$\Gamma_k \sim n_r \langle \sigma_{\text{el}} v_{\text{rel}} \rangle \frac{T}{m_X}, \quad (19)$$

where n_r is the number density of the relativistic species in the thermal bath, and $\langle \sigma_{\text{el}} v_{\text{rel}} \rangle$ is the thermally-averaged cross section for elastic scattering between dark matter particles and the relativistic particles in the thermal bath.

We first consider the model-independent elastic scattering off the hidden sector thermal bath of ϕ particles [38]. For $T \lesssim m_\phi$, n_ϕ is Boltzmann suppressed, and so this process will not be able to maintain kinetic equilibrium. For $T \gtrsim m_\phi$, the thermally-averaged $X\phi \rightarrow X\phi$ cross section is

$$\langle \sigma_{\text{el}} v_{\text{rel}} \rangle \sim \frac{\alpha_X^2}{m_X^2}, \quad (20)$$

the Thomson scattering cross section. Given the ϕ number density $n_\phi \sim T^3$, the momentum transfer rate is $\Gamma_k \sim \alpha_X^2 T^4/m_X^3$, and this is equal to $H \sim T^2/m_{\text{Pl}}$ at

$$T \sim \left[\frac{m_X^3}{\alpha_X^2 m_{\text{Pl}}} \right]^{1/2} = 0.43 \text{ MeV} \left[\frac{0.021}{\alpha_X} \right] \left[\frac{m_X}{\text{TeV}} \right]^{3/2}, \quad (21)$$

where we have normalized the expression to typical parameters that give the correct relic density in the presence of Sommerfeld enhancement. Halo shape constraints require $m_\phi \gtrsim 30 \text{ MeV}$ [2], and so for all physically viable and relevant parameters,

$$T_{\text{kd}}^\phi \sim m_\phi. \quad (22)$$

This is the temperature of kinetic decoupling from the hidden thermal bath. It is quite model independent, as it assumes only the X - ϕ interactions required in all Sommerfeld enhancement scenarios, and so it provides a maximal value of T_{kd} in Sommerfeld scenarios.

Dark matter particles may also be kept in kinetic equilibrium through scattering off the visible sector's thermal bath. This is more model dependent, but as discussed in Sec. III B, if ϕ is a U(1) gauge boson, it may mix with the standard model photon. After diagonalizing the ϕ -photon system, standard model particles with charge Q have hidden charge ϵQ , inducing new energy transfer processes. The most efficient process is $Xe \rightarrow Xe$ scattering through t -channel ϕ exchange. For $T \lesssim m_e$, Boltzmann suppression of the electron number density makes the interaction inefficient. At temperatures $m_\phi \gtrsim T \gtrsim m_e$, the corresponding cross section is [39]

$$\langle \sigma_{\text{el}} v_{\text{rel}} \rangle \sim \frac{\epsilon^2 \alpha_{\text{EM}} \alpha_X T^2}{m_\phi^4}. \quad (23)$$

The momentum transfer rate $\Gamma_k \sim \epsilon^2 \alpha_{\text{EM}} \alpha_X T^6 / (m_\phi^4 m_X)$ is equal to the expansion rate at temperature

$$T \sim \left[\frac{m_\phi^4 m_X}{\epsilon^2 \alpha_{\text{EM}} \alpha_X m_{\text{Pl}}} \right]^{1/4}, \quad (24)$$

and so the resulting temperature of kinetic decoupling from the visible sector's thermal bath is

$$T_{\text{kd}}^e(\epsilon) \sim \max \left\{ m_e, 0.82 \text{ MeV} \left[\frac{10^{-3}}{\epsilon} \right]^{1/2} \left[\frac{m_\phi}{30 \text{ MeV}} \right] \times \left[\frac{0.021}{\alpha_X} \right]^{1/4} \left[\frac{m_X}{\text{TeV}} \right]^{1/4} \right\}, \quad (25)$$

where we have again normalized α_X and m_X to typical freeze out parameters, and additionally normalized m_ϕ to its smallest possible value and ϵ to its largest allowed value. When ϵ is near its maximal value, $T_{\text{kd}}^e(\epsilon) < T_{\text{kd}}^\phi$, and so interactions with the visible thermal bath delay kinetic decoupling. We denote the lowest possible kinetic decoupling temperature

$$T_{\text{kd}}^e \equiv T_{\text{kd}}^e(\epsilon = 10^{-3}), \quad (26)$$

and will explore the dependence of our results on the temperature of kinetic decoupling by varying it between its maximal value T_{kd}^ϕ and its minimal value T_{kd}^e .

By crossing symmetry, the visible sector scattering interaction also implies an annihilation process $XX \rightarrow e^+e^-$. In principle, this also enters the thermal relic density calculation. The cross section is $\langle \sigma_{\text{an}} v_{\text{rel}} \rangle \sim \epsilon^2 \alpha_{\text{EM}} \alpha_X / m_X^2$, however, and the ϵ^2 suppression makes this subdominant for all relevant cases.

D. Velocity distribution after kinetic decoupling

Before kinetic decoupling, dark matter particles have the same temperature as the thermal bath and a Maxwell-Boltzmann phase space distribution. Usually one assumes that the phase space distribution remains Maxwell-Boltzmann after kinetic decoupling. However, this is not necessarily true in scenarios with Sommerfeld-enhanced

annihilation, because slow particles annihilate with larger cross sections. This preferentially depletes the low velocity population and may distort the phase space distribution. In this case, the standard formulas used to compute the relic density of dark matter are not valid and one needs to explicitly consider the effect of annihilations using the full phase space distribution.

As discussed in Refs. [2,39–42], however, the ϕ field mediates self-scattering between dark matter particles. If the self-scattering rate is higher than the Hubble expansion rate, the momentum exchanged in these self-interactions will be sufficient to maintain thermal equilibrium. The momentum transfer of particles interacting through Yukawa potentials has been studied in Refs. [43,44]. Although the authors of these studies were interested in slow and highly charged particles moving in plasmas with screened Coulomb potentials, they approximated these potentials by Yukawa potentials, and so their results are exactly applicable in the current particle physics context. The numerical results of these studies for the momentum transfer cross section are accurately reproduced by [43,44]

$$\begin{aligned} \sigma_T &\approx \frac{4\pi}{m_\phi^2} \beta^2 \ln(1 + \beta^{-1}), & \beta < 0.1 \\ \sigma_T &\approx \frac{8\pi}{m_\phi^2} \frac{\beta^2}{1 + 1.5\beta^{1.65}}, & 0.1 \leq \beta \leq 1000 \\ \sigma_T &\approx \frac{\pi}{m_\phi^2} \left(\ln\beta + 1 - \frac{1}{2} \ln^{-1}\beta \right)^2, & \beta > 1000, \end{aligned} \quad (27)$$

where $\beta = 2\alpha_X m_\phi / (m_X v_{\text{rel}}^2)$.

Given these results, we may calculate the dark matter self-scattering rate and compare it to the Hubble expansion rate. The self-scattering rate is given by [2]

$$\Gamma_s = \int f(\vec{v}_1) f(\vec{v}_2) n_X (\sigma_T v_{\text{rel}}) \frac{v_{\text{rel}}^2}{v_0^2} d^3\vec{v}_1 d^3\vec{v}_2, \quad (28)$$

where n_X is the dark matter density. This sets the time scale to change velocities by $\mathcal{O}(1)$. To find the rate to change the kinetic energy by $\mathcal{O}(1)$, one would divide the above rate by a factor of 3. However, such details are not important for the present calculation. H and Γ_s are presented in Fig. 2 for two representative cases.

We may define the self-scattering decoupling temperature T_{nt} by $\Gamma_s(T_{\text{nt}}) = H(T_{\text{nt}})$; after T_{nt} , the dark matter velocity distribution may become nonthermal. To derive an approximate expression for T_{nt} , note that to have the right relic abundance, the comoving number density of dark matter during the self-scattering epoch cannot be smaller than the comoving number density at present. The dark matter (DM) number density n_X may therefore be taken to be

$$n_X \sim s \left(\frac{n_X}{s} \right)_0 = s \frac{\Omega_{\text{DM}} \rho_c}{m_X s_0} \simeq s \times 3.9 \times 10^{-13} \left[\frac{\text{TeV}}{m_X} \right]. \quad (29)$$

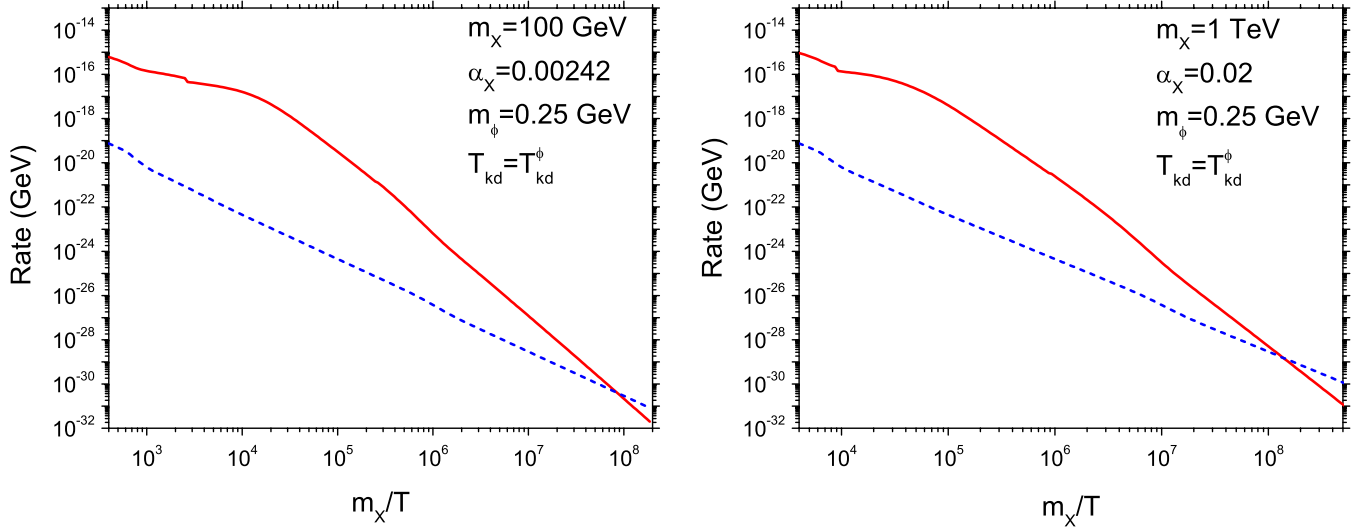


FIG. 2 (color online). The self-scattering rate Γ_s (solid red lines) and the Hubble rate H (dashed blue lines) as functions of $x = m_X/T$ for the values of m_X , m_ϕ , α_X , and T_{kd} indicated.

During the self-scattering decoupling epoch, typically $\beta > 1000$, and so $\sigma_T \sim \kappa/m_\phi^2$, where $\kappa \sim \pi \ln^2 \beta \sim 600\text{--}1000$ for the parameters of interest. With this approximate expression for σ_T , we find

$$T_{\text{nt}} \sim 20 \text{ keV} \left[\frac{m_\phi}{250 \text{ MeV}} \right] \left[\frac{m_X}{1 \text{ TeV}} \right]^{3/4} \left[\frac{T_{\text{kd}}}{250 \text{ MeV}} \right]^{1/4} \times \left[\frac{\kappa}{800} \right]^{-(1/2)}. \quad (30)$$

This agrees well with our numerical results, and we use this approximate form with $\kappa = 800$ in deriving our results below.

From both the numerical and analytical results, we see that the self-scattering rate efficiently preserves the thermal distribution to temperatures $T_{\text{nt}} \sim 10$ keV, when dark matter velocities are $v_{\text{rel}} \sim T/(T_{\text{kd}} m_X)^{1/2} \sim 10^{-5}$. This is a low velocity, and in most cases, we will find that it is sufficiently low that the impact of nonthermality after T_{nt} has a negligible impact on the thermal relic density. Note, however, that T_{nt} is not low enough to ensure a thermal distribution down to $v_{\text{rel}} \sim \alpha_X^4$, where the Sommerfeld resonances cut off, and so very close to resonances, the nonthermality will have an effect.

IV. RESULTS

A. Maximal Sommerfeld enhancements

In this section, we present results for the maximal Sommerfeld enhancement. We have defined several Sommerfeld factors, including S^0 , the original Sommerfeld factor for massless ϕ given in Eq. (1), and S , the Sommerfeld factor generalized to massive ϕ , which includes resonances and is given in Eqs. (5) and (6). Here we also define the effective Sommerfeld enhancement factor

$$S_{\text{eff}} = \frac{(\sigma_{\text{an}} v_{\text{rel}})_0 \bar{S}_{\text{now}}}{3 \times 10^{-26} \text{ cm}^3/\text{s}}, \quad (31)$$

where

$$\bar{S}_{\text{now}} \simeq \frac{x_{\text{now}}^{3/2}}{2\sqrt{\pi}N} \int_0^{v_{\text{max}}} S v_{\text{rel}}^2 e^{-x_{\text{now}} v_{\text{rel}}^2/4} dv_{\text{rel}}, \quad (32)$$

$x_{\text{now}} \equiv 2/v_0^2$, and $N = \text{erf}(z/\sqrt{2}) - (2/\pi)^{1/2} z e^{-z^2/2}$ with $z \equiv v_{\text{max}}/v_0$. S_{eff} is the experimentally relevant parameter, as it is the factor by which indirect fluxes are enhanced relative to the case without Sommerfeld enhancement. Unlike the early Universe case, in the halo we have to cut off the velocity integral at some maximum relative speed v_{max} , which is related to the escape speed from the local neighborhood. This maximum speed is a function of the angle between the dark matter particles in the lab frame. However, we have checked that setting v_{max} to be equal to the escape speed v_{esc} and integrating as in Eq. (32) makes a difference of less than 10%. The escape speed at about 10 kpc from the center of the halo is estimated to be about 500 km/s to 550 km/s [45]. We use $v_{\text{esc}} = 525$ km/s to derive our results, noting that 10% variations in v_{esc} have a much smaller effect on S_{eff} . We have assumed that the velocity dispersion tensor is isotropic with the 1D velocity dispersion given by $v_0/\sqrt{2}$. We use $v_0 = 210$ km/s consistent with determinations of the circular velocity (e.g., [45]). Inferring the dispersion from the circular velocity curve requires knowledge of the dark matter density profile. The value we use for v_0 is consistent with a Navarro-Frenk-White density profile with a scale radius of about 20 kpc. Here again, we note that there are uncertainties in the value of v_0 at the 10% level. This translates directly into an uncertainty in S_{eff} of about

10% if $S \propto 1/v$ and 20% around resonance where $S \propto 1/v^2$.

For comparison purposes, we also define

$$S_{\text{eff}}^0 = \frac{(\sigma_{\text{an}} v_{\text{rel}})_0 \bar{S}_{\text{now}}^0}{3 \times 10^{-26} \text{ cm}^3/\text{s}}, \quad (33)$$

where

$$\bar{S}_{\text{now}}^0 = \frac{x_{\text{now}}^{3/2}}{2\sqrt{\pi}N} \int_0^{v_{\text{esc}}} S^0 v_{\text{rel}}^2 e^{-x_{\text{now}} v_{\text{rel}}^2/4} dv_{\text{rel}} \quad (34)$$

is the effective Sommerfeld enhancement without resonances, but with the Sommerfeld effect on freeze out included.

To calculate the largest possible S_{eff} for a given m_X and m_ϕ , we make a number of optimistic (S_{eff} -maximizing) assumptions:

- We fix $T_{\text{kd}} = T_{\text{kd}}^e$. This delays kinetic decoupling as much as possible, keeps the dark matter as hot as possible, and so reduces the Sommerfeld effect on freeze out, maximizing S_{eff} .
- We fix x_s in Eq. (13) by stopping the dark matter evolution at T_{nt} ; that is, we neglect all annihilations that occur after the dark matter distribution becomes nonthermal. This is certainly optimistic, as the distribution will remain thermal for some time and dark matter annihilations will continue, but this again minimizes the Sommerfeld effect on freeze out, maximizing S_{eff} .
- We require $\Omega_X h^2 = 0.114$. This might appear to be too restrictive; after all, there is no requirement that the observed signals arise from a particle that makes up all of the dark matter. However, if a flux

arises from Sommerfeld-enhanced annihilation, it scales as $n_X^2 \langle \sigma_{\text{an}} v_{\text{rel}} \rangle S \sim \alpha_X^{-1}$, because its number density scales as $n_X \sim \Omega_X \sim \langle \sigma_{\text{an}} v_{\text{rel}} \rangle^{-1} \sim \alpha_X^{-2}$ and $S \sim \alpha_X$. The flux can therefore always be increased by lowering α_X until $\Omega_X h^2$ is all of the dark matter, and so choosing $\Omega_X h^2$ in fact maximizes indirect signals.

- We choose the maximal α_X that yields $\Omega_X h^2 = 0.114$. Roughly speaking, $\Omega_X h^2$ decreases as α_X increases, and so typically, there is a unique choice of α_X that yields the correct $\Omega_X h^2$. In some cases with strong resonances, however, there are three choices of α_X that give the correct Ω_X , as shown in Fig. 3. When we do a coarse scan there is no guarantee that we always pick the solution with the largest value of S_{eff} . Note, however, that the values of S_{eff} only change by about 20%–30% between the three allowed solutions, and such variations do not modify our conclusions. The values of α_X that yield the correct relic density for given values of m_X and m_ϕ are given in Fig. 4. Note that close to a resonance, one needs extreme fine-tuning to avoid efficient annihilation in the early Universe and obtain the correct relic density.

In Fig. 5, we show the maximal values of S_{eff} as a function of m_X for $m_\phi = 250$ MeV and 1 GeV. To understand the impact of resonances and the Sommerfeld effect on freeze out on these results, we also plot other Sommerfeld enhancements. S^0 , which includes neither resonances nor freeze out effects, was used in Ref. [2]. We see that it is almost always overestimates the maximal Sommerfeld enhancement. This is because, as evident in Fig. 5, the effect of Sommerfeld-enhanced annihilation on

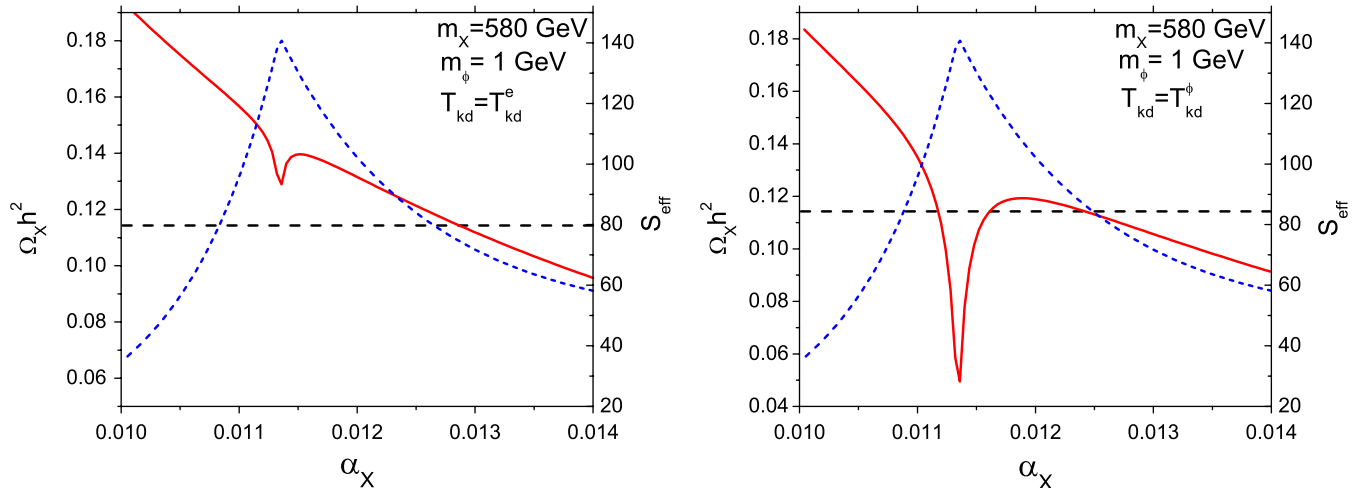


FIG. 3 (color online). The relic density $\Omega_X h^2$ (solid red lines) and S_{eff} (dotted blue lines) as a function of the fine-structure constant α_X for the fixed values of m_X , m_ϕ , and T_{kd} indicated and the observed value of $\Omega_X h^2$ (dashed black lines). In most cases, there is a unique choice of α_X that yields the correct $\Omega_X h^2 = 0.114$ (left panel). In the presence of strong resonances, however, there are cases where three different choices of α_X all yield the correct Ω_X (right panel). In these cases, S_{eff} varies by about 20%–30% between the different solutions.

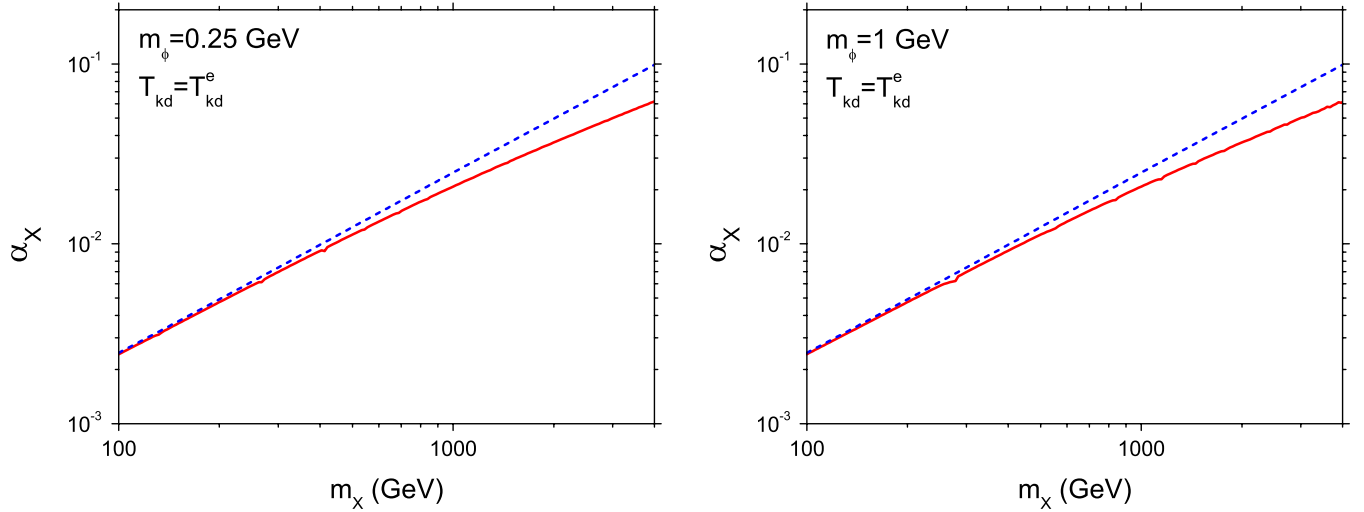


FIG. 4 (color online). The value of α_X required to achieve a relic density of $\Omega_X h^2 = 0.114$ as a function of the dark matter particle mass m_X (solid red lines) for $m_\phi = 250$ MeV (left panel) and $m_\phi = 1$ GeV (right panel). We also plot the required α_X (dotted blue lines) if Sommerfeld effects are neglected in the early Universe. The tree-level cross section (without Sommerfeld enhancement) is $(\sigma_{\text{an}} \nu_{\text{rel}})_0 = \pi \alpha_X^2 / m_X^2$. Because α_X varies over almost 2 orders of magnitude in this plot, the dips near resonance are not immediately apparent.

freeze out is a significant suppression. This may be understood as follows. If Sommerfeld enhancement reduces the thermal relic density by a factor ζ , the tree-level cross section $(\sigma_{\text{an}} \nu_{\text{rel}})_0$ must be reduced by ζ to keep Ω_X fixed. However, $S_{\text{eff}} \propto \alpha_X^3 \propto (\sigma_{\text{an}} \nu_{\text{rel}})_0^{3/2}$, and so reducing Ω_X by ζ implies a reduction in S_{eff} by a factor $\zeta^{3/2}$. For example, for $m_X = 1$ TeV, the maximal coupling is $\alpha_X \simeq 0.021$, and $\zeta \sim 1.5$, consistent with the results of Refs. [29,30]. Including the effect of freeze out here therefore reduces S^0 to S_{eff}^0 by a factor of $(1.5)^{3/2} \sim 2$, a significant reduction.

Adding resonances then produces oscillations about the S^0 and S_{eff}^0 contours, with peak positions given by Eq. (7). We see that, although resonances can significantly enhance the effective Sommerfeld factor, these enhancements are significant only for low values of n ; for $m_\phi \lesssim$ GeV, these lie at low m_X , and the effect becomes negligible for $m_X \gtrsim 1$ TeV.

In Fig. 5, we have fixed m_ϕ to typical values considered in the literature. In Fig. 6, we plot upper bounds on S_{eff} as a function of m_ϕ for fixed values of m_X . We see that the

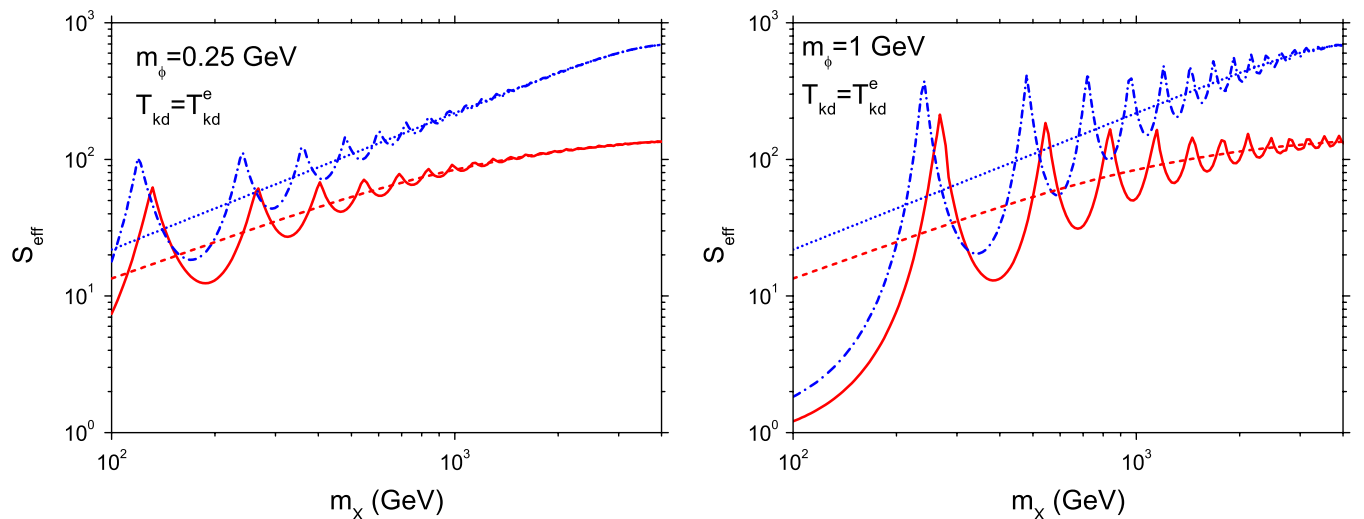


FIG. 5 (color online). The effective Sommerfeld enhancement factor S_{eff} (solid red lines) as a function of m_X for $m_\phi = 250$ MeV (left panel) and 1 GeV (right panel) and the set of S_{eff} -maximizing assumptions listed in Sec. IV A. Also shown for comparison are S^0 (dotted blue lines), the Sommerfeld factor without resonances and neglecting the Sommerfeld effect on freeze out; \tilde{S} (dot-dashed blue lines), the Sommerfeld factor with resonances but neglecting the Sommerfeld effect on freeze out; and S_{eff}^0 (dashed red lines), the Sommerfeld factor without resonances, but including the Sommerfeld effect on freeze out.

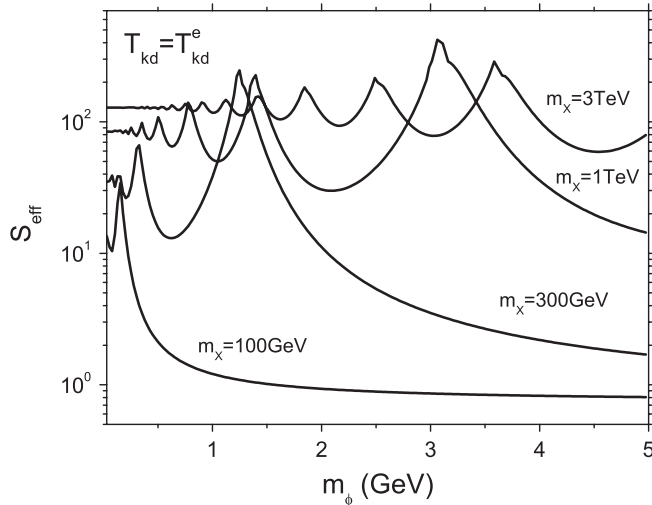


FIG. 6. The effective Sommerfeld enhancement factor S_{eff} as a function of m_ϕ for the m_X indicated and the set of S_{eff} -maximizing assumptions listed in Sec. IV A.

resonances have little impact for $m_\phi \lesssim 1$ GeV, but can produce enhancements by factors of 2 to 3 for larger m_ϕ . Large values of $m_\phi \gtrsim 1$ GeV have been considered disfavored, however, as they eliminate the kinematic suppression of antiproton fluxes, which are typically considered to be consistent with astrophysical backgrounds.

What is the effect on S_{eff} of deviating from the S_{eff} -maximizing assumptions? The effect of the choice of T_{kd} is highly sensitive to how close one is to a resonance. Far from a resonance, there is little sensitivity to T_{kd} values between T_{kd}^ϕ and T_{kd}^e . As one approaches a resonance, however, S_{eff} may vary by $\sim 10\%$ or more. In particular, we found regions close to resonances for $m_\phi \sim \text{GeV}$ where

S_{eff} changed by about 30%. This behavior is shown in Fig. 7. If one is not very close to a resonance, the choice of x_s has a small impact. In contrast, as discussed above, the assumption of $\Omega_X h^2 = 0.114$ has a large effect; for other choices, the maximal value of S_{eff} scales as $(\Omega_X h^2)^{3/2}$. Finally, as can be seen in Fig. 3, different choices of α_X also decrease $S_{\text{eff}} \propto \alpha_X^3$ by $\sim 20\%$ – 30% .

For typical parameters that are not very close to a resonance, then, we expect that assumptions different from those listed above would produce a small decrease in S_{eff} when away from resonances. Note, however, that very near a resonance, all of these choices become very important, as we now discuss.

B. Resonances and chemical recoupling

Very near a resonance, the annihilation cross section becomes sensitive to details of the dark matter's evolution to low velocities at late times. The effective Sommerfeld enhancement then becomes highly sensitive to details of T_{kd} , the velocity distribution of dark matter after T_{kd} , and the cutoff of resonant enhancements. As this is far from generic, we do not present details of these dependences, but we note that, contrary to naive expectations, S_{eff} is not maximized by sitting exactly at resonance. In fact, exact resonances lead to extremely efficient annihilation in the early Universe and minimize current indirect signals.

Resonant Sommerfeld enhancement's effect on freeze out can also be so large that it leads to the intriguing phenomenon of chemical recoupling. This is illustrated for two representative cases in Fig. 8. Without resonant Sommerfeld enhancement, dark matter freezes out and remains frozen out, but with resonant enhancement, dark matter may melt back in at late times, or chemically recouple, leading to a second era of efficient annihilation.

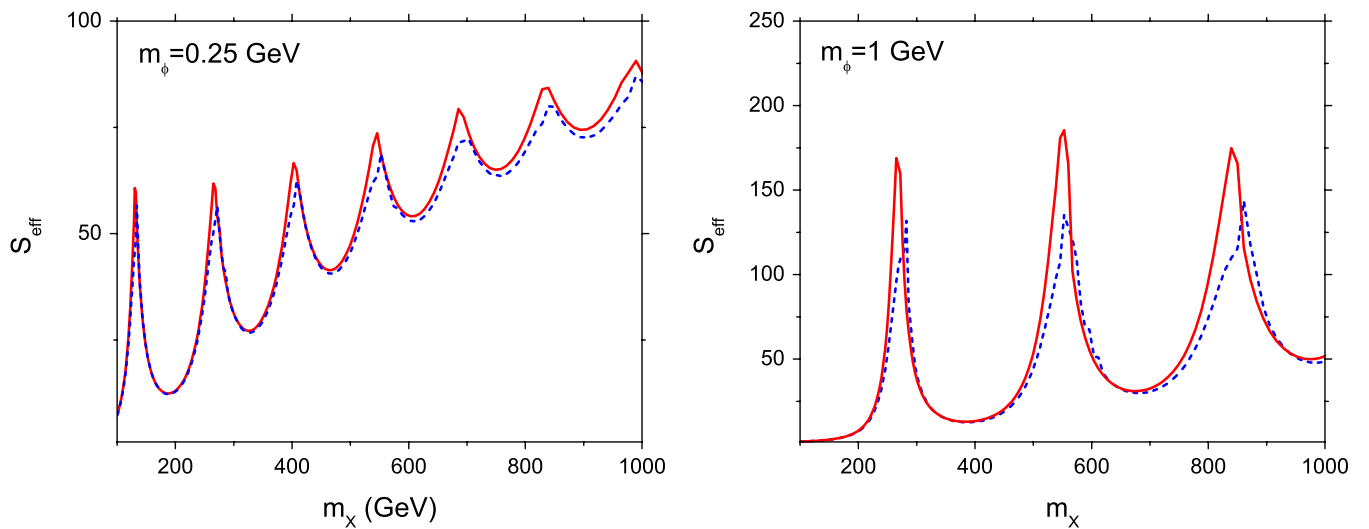


FIG. 7 (color online). The effective Sommerfeld enhancement factor S_{eff} for $T_{\text{kd}} = T_{\text{kd}}^e$ (solid red lines) and $T_{\text{kd}} = T_{\text{kd}}^\phi$ (dashed blue lines), for $m_\phi = 250$ MeV (left panel) and 1 GeV (right panel). All other S_{eff} -maximizing assumptions have been made. Near resonances, S_{eff} is sensitive to the temperature of kinetic decoupling.

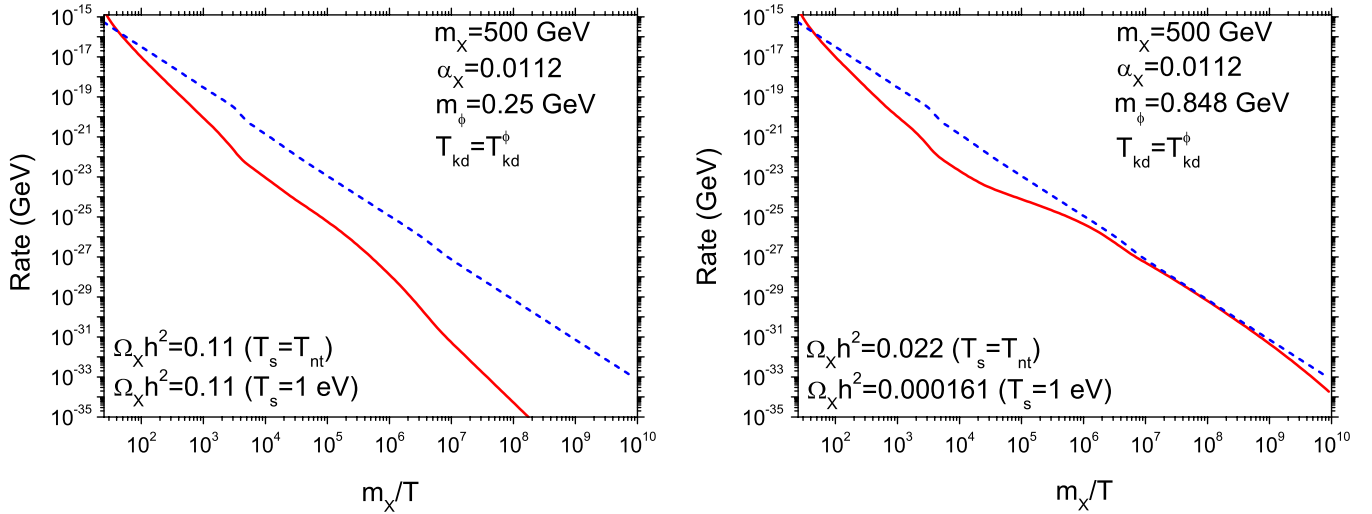


FIG. 8 (color online). The annihilation rate $\Gamma_{\text{an}} = n_X \langle \sigma_{\text{an}} v_{\text{rel}} \rangle$ (solid red lines) and the Hubble rate H (dashed blue lines) as functions of $x = m_X/T$ for the values of m_X , m_ϕ , α_X , and T_{kd} indicated. Without resonant Sommerfeld enhancement (left panel), $\Gamma_{\text{an}} < H$ after freeze out; but with resonant Sommerfeld enhancement (right panel), Γ_{an} may become comparable to H at late times, leading to chemical recoupling and an new era of annihilation. Also indicated on the plots are the resulting values of $\Omega_X h^2$ that result from assuming annihilations up to a temperature of T_{nt} when the dark matter distribution function is no longer able to maintain kinetic equilibrium, as well as a lower temperature of 1 eV.

The phenomenon of chemical recoupling in the context of Sommerfeld enhancements may be understood as follows. The Hubble parameter scales as $H \sim T^2$. The annihilation rate is $\Gamma_{\text{an}} = n_X \langle \sigma_{\text{an}} v_{\text{rel}} \rangle$. The dark matter density scales as $n_X \sim T^3$, and $\langle \sigma_{\text{an}} v_{\text{rel}} \rangle$ scales as $v_{\text{rel}}^{-1} \sim T_X^{-1/2}$ off resonance and $v_{\text{rel}}^{-2} \sim T_X^{-1}$ on resonance. Before T_{kd} , $T_X = T$, and so the annihilation rate cannot grow relative to the expansion rate. However, after T_{kd} , $T_X \propto T^2$, and so $\Gamma_{\text{an}} \sim T$. At late times, then, the annihilation rate decrease slowly enough to become comparable to or greater than the expansion rate. The dark matter then melts back in, and there is a new era of annihilation. In these cases, the relic density is very sensitive to the temperature at which we stop including the annihilation process. As our default case, we have been conservative and stopped the annihilation process when the self-interactions of dark matter are no longer able to maintain kinetic equilibrium. Annihilations will proceed beyond this temperature; however, we cannot assume a Maxwellian distribution for the dark matter particle, because the annihilations preferentially deplete the low momentum tail. As an example, we show in Fig. 8 what happens when we allow the annihilation process to proceed down to 1 eV. The relic density is essentially negligible for the case where we have chemical recoupling, whereas it is unchanged for the case away from resonance.

V. COMPARISON TO PAMELA AND FERMI

A. Maximal enhancements and best-fit parameters

In Figs. 9 and 10, we compare the maximal S_{eff} presented in Fig. 5 to the boost factors required to explain PAMELA

and Fermi data. The PAMELA and Fermi regions, derived in Ref. [20], assume an isothermal halo and a 250 MeV force carrier that decays with 100% branching ratio to $\mu^+ \mu^-$, leading to annihilations $XX \rightarrow \mu^+ \mu^- \mu^+ \mu^-$. The fits are insensitive to m_ϕ , provided $m_\phi \ll m_X$. The best fit is for $(m_X, S_{\text{eff}}) = (2.35 \text{ TeV}, 1500)$, and the

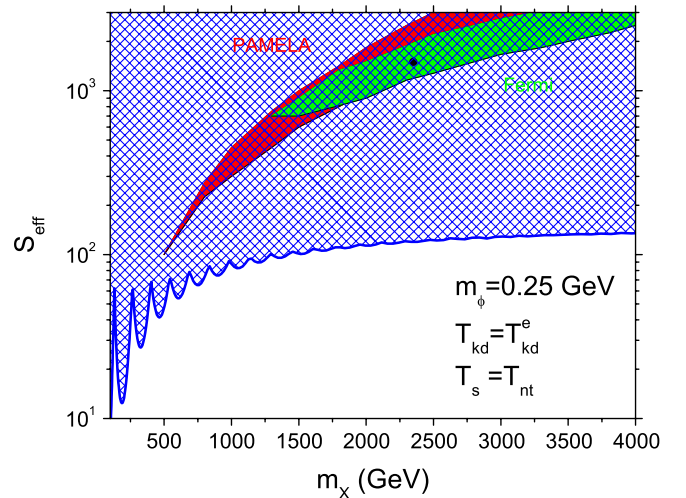


FIG. 9 (color online). The maximal effective Sommerfeld enhancement factor S_{eff} compared to data for force-carrier mass $m_\phi = 250 \text{ MeV}$. The crosshatched region is excluded by requiring consistency with the thermal relic density and adopting all of the S_{eff} -maximizing assumptions listed in Sec. IV A. The red and green shaded regions are 2σ PAMELA- and Fermi-favored regions for the 4μ channel, and the best-fit point is $(m_X, S_{\text{eff}}) = (2.35 \text{ TeV}, 1500)$ [20].

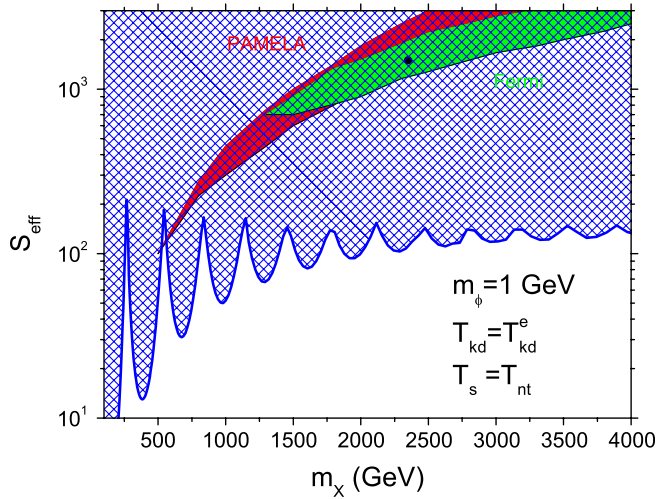


FIG. 10 (color online). The maximal effective Sommerfeld enhancement factor S_{eff} compared to data for $m_\phi = 1$ GeV. The crosshatched and shaded regions are as in Fig. 9. The PAMELA- and Fermi-favored regions are for $m_\phi = 250$ MeV, but do not change appreciably for $m_\phi = 1$ GeV [21] in the case that only the 4μ channel is included.

regions are 2σ contours relative to the best-fit parameters. Results for other halo profiles, other particle physics models, and other final states have also been considered [20,21,46]. These studies find that the 4μ final state provides a better fit than all other considered final states, but see Sec. VD for a discussion of this issue.

Despite choosing T_{kd} and all other parameters to maximize S_{eff} , we find that the possible values of S_{eff} fall short of explaining the PAMELA and Fermi excesses. For example, at $m_X = 2.35$ TeV the maximal Sommerfeld enhancement is $S_{\text{eff}} \approx 100$, a factor of 15 below the best-fit value. For lower m_X , the data may be fit with smaller S_{eff} , but the relic density bounds are also stronger; for example, for $m_\phi = 250$ MeV and $m_X = 0.1, 0.3,$ and 1 TeV, the maximal values of S_{eff} are 7, 30, and 90, respectively.

B. Astrophysical uncertainties

Our analysis so far has assumed standard astrophysics and astroparticle physics. To enable Sommerfeld enhancements to explain PAMELA and Fermi, one may consider uncertainties in these assumptions.

The annihilation signal is sensitive to the square of the dark matter density ρ_X in the local neighborhood. The best-fit regions given in Ref. [20] assume the conventional value $\rho_X = 0.3$ GeV/cm³, which has traditionally been considered to be uncertain up to a factor of 2 [47]. More recently, studies have tried to refine the calculation of the uncertainty in the local density of dark matter. One study found $\rho_X = 0.389 \pm 0.025$ GeV/cm³ [48], using an ensemble of kinematic tracers to constrain a standard galactic model with a spherical halo, stellar and gas disk

and stellar bulge. Another recent study [49] finds $\rho_X = 0.2\text{--}0.4$ GeV/cm³. A third study Taylor expanded the local rotation velocity curve to first order and used local kinematic tracers and an estimate of the baryonic contribution to the rotation curve to find $\rho_X = 0.430 \pm 0.113 \pm 0.096$ GeV/cm³ [50], where the first uncertainty is from the slope of the circular velocity $V(r)$ at the Sun's radius and the second is from uncertainty in distance of the Sun from the Galactic center. The three studies are consistent with each other, but the first one finds significantly smaller uncertainty in the value of ρ_X . If we use the results of Ref. [48], then the best-fit regions plotted in Figs. 9 and 10 shift down by a factor of $(0.4/0.3)^2$. Of course, it is not appropriate to simply scale up the signal by some ρ_X^2 required to fit the positron and electron data; in a proper analysis, one should include the appropriate likelihood for ρ_X and $\Omega_X h^2$. We have also discussed some of the uncertainties in the velocity distribution function and their effects in Sec. IVA, and, as with ρ_X and $\Omega_X h^2$, these uncertainties should be included in the full likelihood analysis.

Another effect that was pointed out recently [51] is that the positrons produced within the unseen bound subhalos out in the Milky Way dark matter halo could propagate to the local neighborhood and contribute to the PAMELA and Fermi signal. For the best-fit point, this study found that about 30% of the positrons in PAMELA could be due to annihilations in the subhalos [51]. Taking $\rho_X = 0.4$ GeV/cm³ and an enhancement factor of 30% from subhalos, the discrepancy is reduced from a factor of 15 to a factor of 4. We see that these changes are insufficient to reach the best-fit regime, even if all the parameters are simultaneously pushed in the most optimistic direction. A more troubling aspect of this calculation is however the fact that the subhalos considered were those resolved by the Via Lactea II simulation. This implies that the much larger contribution from the lower mass subhalos would over produce the signal. Appealing to subhalos to $\mathcal{O}(1)$ changes to the required S_{eff} is therefore unmotivated.

Finally, we may also appeal to small scale structure in the immediate local neighborhood to enhance the positron signals. Such a contribution is largely unconstrained by data, but present simulations, even without the deleterious effect of the disk of stars, do not predict large enhancements from such an effect [52]. A recent study [53] using the Via Lactea II simulation and including the Sommerfeld effect for annihilation in the local clumps found that the probability of finding a subhalo close enough to make an $\mathcal{O}(1)$ effect on the e^+e^- flux above about 100 GeV is about 4% and that this contribution could be about 15% if the Via Lactea II subhalo mass function is extrapolated down to lower (unresolved) masses. Two points are worthy of further note here. First, these subhalo results are going to be crucially affected by the disk. The zeroth order expectation is that the gravity of the stellar disk will not allow

such subhalos to survive or even if they do, truncate it significantly. Second, for small m_ϕ and near resonances self-interactions are important [2,41] and that could modify the internal structure of the subhalo as well as the mass function.

A more complicated set of astrophysical uncertainties is related to astrophysical backgrounds and the signals themselves. For example, the required Sommerfeld enhancement may be reduced if conventional astrophysical backgrounds have been underestimated, cosmic ray propagation models are modified, pulsars contribute significantly to the positron flux, or the PAMELA e^+ sample includes some proton contamination. Any of these effects could reduce the required dark matter signal, but they could also plausibly eliminate the need for dark matter altogether, as has been argued in numerous studies [54–59]. It is therefore typically unclear what the motivation would be for assuming that such effects account for part of the signal but not all of it, independent of a desire to leave room for a dark matter signal.

A possible exception to the above argument is the possibility of adjusting cosmic ray propagation models to match the Fermi spectrum. Such adjustments could remove the need to explain Fermi, without changing the PAMELA excess significantly [60]. As shown in Fig. 10, for $m_X \lesssim \text{TeV}$ and $m_\phi \gtrsim 1 \text{ GeV}$, the allowed Sommerfeld enhancements may be in marginal agreement with PAMELA data. This scenario is, however, problematic. A contribution to the e^\pm flux at about 100 GeV to explain the observed PAMELA positron fraction implies a modification to the e^\pm flux observed by Fermi. For example, a 500 GeV dark matter explanation for PAMELA with the correct relic density is marginally possible for $m_\phi = 1 \text{ GeV}$ according to the 4μ fits of Ref. [21] as shown in Fig. 10. This would, however, imply a drop in the e^\pm flux beyond 500 GeV, which is not seen by Fermi, a point stressed in previous works [21,46]. Analysis of such a scenario with modified propagation models would be another avenue for further study.

C. Cosmic microwave background bounds

In this section, we include bounds from the effect of residual annihilation at last scattering on the ionized electron fraction and therefore on the cosmic microwave background (CMB) power spectrum. The CMB bound may be written as [61,62]

$$S|_{\nu=0} = \frac{12}{\epsilon_\phi [1 - \cos(\sqrt{24}/\epsilon_\phi)]} < \frac{120}{f} \left(\frac{m_X}{\text{TeV}} \right), \quad (35)$$

where f is the average fraction of the energy produced in annihilation that reionizes hydrogen between redshifts of 800 and 1000. For the e^\pm final state, $f \simeq 0.7$, while for the μ^\pm final state, $f \simeq 0.25$ [62]. The left-hand side $S|_{\nu=0}$ is the saturated Sommerfeld enhancement obtained by setting

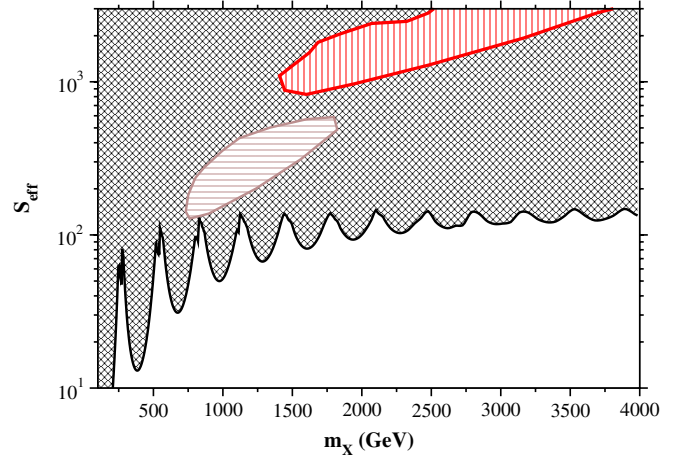


FIG. 11 (color online). The maximal effective Sommerfeld enhancement S_{eff} (black solid curve) after including bounds from CMB (see Sec. V C) compared to fits for $m_\phi = 1 \text{ GeV}$ for exclusive $4e$ and 4μ modes. The dark red (vertically hatched) region is the 99% contour for the 4μ mode and the light red (horizontally hatched) contour is for the $4e$ mode [21]. The best-fit regions for the $4e$ mode are at lower S_{eff} and m_X , but the best-fit $4e$ mode point has a significantly lower likelihood than the best-fit 4μ mode point [21] for a fixed cosmic ray propagation model.

$\nu = 0$ in Eq. (6). This is sufficient to approximate the Sommerfeld enhancement during recombination, unless one is right on top of a resonance, which is highly disfavored by this bound.

Equation (35) is not a monotonic function of α_X . To impose this bound, we consider a region in α_X that is bounded by values 10% larger and smaller than the value of α_X dictated by the relic density bound of $\Omega_X h^2 = 0.114$. We find that large resonances are no longer allowed by the CMB bound as a comparison of Figs. 10 and 11 will show.

D. Force-carrier decay channels

The best-fit regions of the (m_X, S_{eff}) plane shown in Figs. 9 and 10 assume that the force-carrier decays solely to muons, leading to the $XX \rightarrow \phi\phi \rightarrow 4\mu$ annihilation channel. The 4μ final state has been shown [20,21,46,63] to be the final state that yields the best fit of all considered so far, as it leads to smooth e^\pm distributions that can simultaneously explain both PAMELA and Fermi. At the same time, this is an inefficient mode, as significant energy is lost to neutrinos. To ameliorate the discrepancy between the required S_{eff} to explain the anomalies and the maximal S_{eff} allowed by the relic density constraint, one might consider other channels, which could potentially allow similar contributions to the e^\pm spectrum with lower S_{eff} and at lower m_X .

The existence of other decay channels is theoretically well-motivated and there are many possibilities. If the decay $\phi \rightarrow \mu^+ \mu^-$ is possible, the decay $\phi \rightarrow e^+ e^-$ is

also kinematically allowed. For $m_\phi \sim 1$ GeV, other decay modes, such as $\phi \rightarrow \pi\pi, K\bar{K}$, are also possible. As an example, if ϕ decays are controlled by its kinetic mixing with the standard model photon, then there will also be e^\pm and π^\pm channels with roughly similar branching ratios, depending on m_ϕ [64]. Alternatively, ϕ particles may decay preferentially to the heaviest available states if they are coupled through standard model Yukawa couplings, or may have other interesting dynamics. In fact, in full generality, the ϕ particles may decay not only to pairs of standard model particles, but also through final states involving hidden Higgs bosons and other hidden particles, leading, for example, to XX annihilations to 6 or 8 particle final states.

Studies so far have found that the 4μ mode is a better fit than other channels [21,63]. A precise quantitative statement requires inclusion of correlations through covariance matrices for each experiment, and these are not available. Given this fundamental uncertainty, conclusions about the μ mode and other channels are far from firm and require additional assumptions.

In Ref. [63], the $4e$ mode is shown to have a χ^2/dof that is larger by about ~ 1 compared to the 4μ mode. These values of χ^2 are computed by assuming that the data points are uncorrelated and allowing for the spatial diffusion parameter to vary. Other results obtained by fixing the diffusion parameter seem to agree qualitatively with this result [21]. In fact, marginalizing over m_X and other parameters, Ref. [21] found that the possibility of mixed modes in the ratio of $4e:4\mu = 1:1$ is excluded at 99.9% C.L.

Despite these results, we may consider what impact other modes might have, especially in light of the fact that there is a lot of freedom in the cosmic ray propagation model parameters. As an illustration, we consider the best-fit regions from Ref. [21] for both the $4e$ and 4μ modes separately. We plot these best-fit regions with the relic density and CMB upper limits on S_{eff} in Fig. 11. The best-fit point for the $4e$ mode is at lower S_{eff} and m_X and the edge of the region is in marginal agreement with the relic density constraint. Of course, as discussed above, the best-fit point for the $4e$ mode has a much lower likelihood than the best-fit point for the 4μ mode, according to Ref. [21]. For a 1:1 ratio, the best-fit point would move to X masses somewhere between 1.5 and 2 TeV. This is disfavored by the relic density requirement even if we set the local dark matter density to be $0.4 \text{ GeV}/\text{cm}^3$.

For the reasons noted above, all of these conclusions are subject to many uncertainties. The present conclusion in the literature seems to be that significant deviations from the pure 4μ channel do not provide a good fit to data. One way forward would be to redo fits of the kind discussed in literature, allowing for variations in cosmic ray propagation model parameters to see if this conclusion holds up in general. In any case, a careful statistical analysis is required before appeals to alternative channels alone may

be considered as a way to allow Sommerfeld scenarios to explain the data.

E. Bound states

For the case of attractive forces, pairs of dark matter particles may form bound states by radiating force-carrier particles if the force-carrier mass m_ϕ is below the binding energy $\alpha_X^2 m_X/4$ [18,65,66]. As in the case of positronium, when this is kinematically possible, annihilation through bound states is preferred. In the limits $m_\phi \ll \alpha_X^2 m_X/4$ and $v \gg m_\phi/m_X$, the annihilation rate enhancement is about $20\alpha_X v_\phi(3 - v_\phi^2)/(2v)$ for fermions, where v_ϕ is the velocity of the final state ϕ particle when the bound state forms [18]. This is roughly a factor of 7 larger than S^0 when $v_\phi \simeq 1$.

Unfortunately, such bound state effects do not help Sommerfeld enhancement models explain the PAMELA and Fermi data for two reasons. First, bound state formation is kinematically forbidden in large parts of parameter space that yield the best fits to data, and the large factor of 7 enhancement alluded to above is not realized for most of the parameter space plotted in Fig. 10. To see this, we first need to compute the value of α_X required to get the right relic density. Now, bound states may form in the early Universe when they are energetically viable, that is when $v < \alpha_X$, and hence affect the relic density calculation. For simplicity we neglect this effect here, and use the same α_X values from Fig. 4, which leads to a conservative bound. Then, for $m_\phi = 0.25$ GeV (1 GeV), the bound state enhancement formula quoted above is valid only for $m_X \gg 1.5$ TeV (2.5 GeV). These inequalities are not satisfied for parameters that are most promising for fitting the data ($m_\phi \geq 0.25$ GeV and $m_X \leq 4$ TeV).

Second, if bound states form, they are likely to be in conflict with bounds from the CMB. To see this, let us saturate the bound state enhancement at $v \sim m_\phi/m_X$ and estimate the enhancement at recombination as $20\alpha_X m_X/m_\phi$. This should be compared to the CMB bound of $(120/f)(m_X/\text{TeV}) \simeq 500(m_X/\text{TeV})$, where we have set $f = 0.25$ to be conservative. To satisfy the CMB bound, then, we require $\alpha_X < (m_\phi/\text{GeV})/40$. This is generically violated if α_X is fixed by the relic density. For example, for $m_\phi = 1$ GeV, the bound implies $\alpha_X < 0.025$. Requiring the correct relic density then implies $m_X < 1$ TeV. However, we need $m_X > 2.5$ TeV for bound states to form with $m_\phi = 1$ GeV. The situation is no different with $m_\phi = 0.25$ GeV. Thus our saturation approximation above (for bound state enhancement) suggests that regions of parameter spaces where bound states can form are generically in conflict with the CMB.

F. Nonminimal particle physics models

A potentially more promising direction is to construct more complicated particle physics models to achieve

Sommerfeld enhancements that are large enough to explain PAMELA and Fermi. Such models will generically include additional annihilation channels. We begin by discussing the impact of these channels in general, and then examine various strategies one might explore to achieve larger S_{eff} .

1. Additional annihilation channels

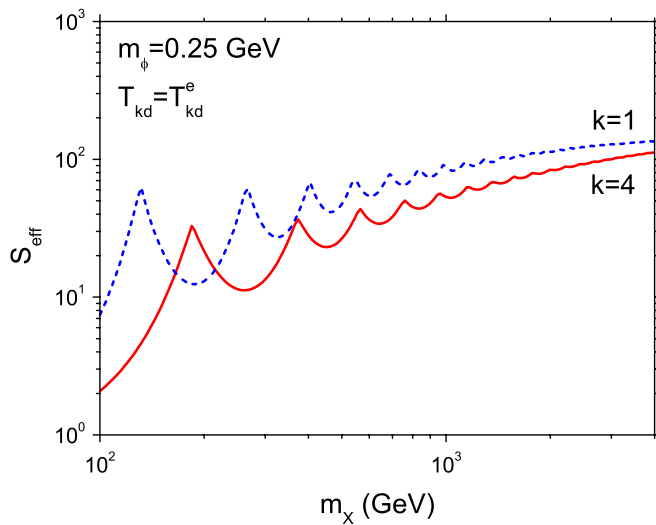
As discussed above, to maximize S_{eff} , we have considered only $XX \rightarrow \phi\phi$ annihilation, leading to the typical tree-level cross section estimate of Eq. (2). Even in the simplest models, however, one may have additional annihilation channels. For example, if ϕ is a U(1) gauge boson with mass generated by spontaneous symmetry breaking through the Higgs potential $\frac{1}{2}\lambda(|H|^2 - v^2)^2$, where H is a complex scalar, there is a physical hidden Higgs boson h with mass $m_h \sim \sqrt{2\lambda}v$, where v is related to the ϕ mass by $m_\phi = \sqrt{8\pi\alpha_X}v$. For perturbative λ , $m_h \lesssim 10m_\phi$; h may be much lighter than ϕ , but it cannot be much heavier. The hidden Higgs boson therefore cannot be decoupled either dynamically or kinematically, and there is necessarily an additional annihilation channel $XX \rightarrow \phi h$. In the nonrelativistic limit, and assuming there is no Yukawa coupling $h\bar{X}X$, the cross section is

$$\sigma(XX \rightarrow \phi h)v_{\text{rel}} = \frac{\pi\alpha_X^2}{m_X^2} \frac{|\bar{p}_\phi|}{m_X} \frac{4m_X^4 + 8m_X^2 m_\phi^2}{(4m_X^2 - m_\phi^2)^2}, \quad (36)$$

where

$$|\bar{p}_\phi| = \frac{[(4m_X^2 - (m_\phi + m_h)^2)(4m_X^2 - (m_\phi - m_h)^2)]^{1/2}}{4m_X}. \quad (37)$$

In the limit $m_\phi, m_h \ll m_X$,



$$\sigma(XX \rightarrow \phi h)v_{\text{rel}} = \frac{1}{4} \frac{\pi\alpha_X^2}{m_X^2}. \quad (38)$$

Additional annihilation channels and their impact on relic densities in Sommerfeld-enhanced models have also been considered in, for example, Ref. [27].

To roughly accommodate such modifications, we may parametrize the effect of additional annihilation channels by assuming a tree-level annihilation cross section

$$(\sigma_{\text{an}}v_{\text{rel}})_0 = k \frac{\pi\alpha_X^2}{m_X^2}, \quad (39)$$

where k is a constant. If the new structures do not significantly modify the Sommerfeld enhancement S , the desired relic density will be achieved by modifying $\alpha_X \rightarrow \alpha_X/\sqrt{k}$, moving the positions of resonances and reducing the maximal S_{eff} by the factor \sqrt{k} . In general, Sommerfeld enhancement in the early Universe is important and the reduction in maximal S_{eff} is smaller. The effects of varying k are shown in Fig. 12.

Of course, in general, new annihilation channels may have different couplings α_X , different kinetic decoupling features, and different T_{nt} , all leading to different Sommerfeld effects, resonances in different places, and many other effects that cannot be captured by a single extra parameter k . The availability of additional annihilation channels also has a complicated effect on the signal. For example, for the case of decays to Higgs bosons and assuming $h \rightarrow 4l$ through real or virtual ϕ pairs, the $h\phi$ mode leads to six or more leptons. This channel changes the energy spectrum of positrons that would be seen by PAMELA and Fermi and thus requires new fits to the data. The full analysis is therefore complicated, and requires a dedicated study incorporating all of these effects into the

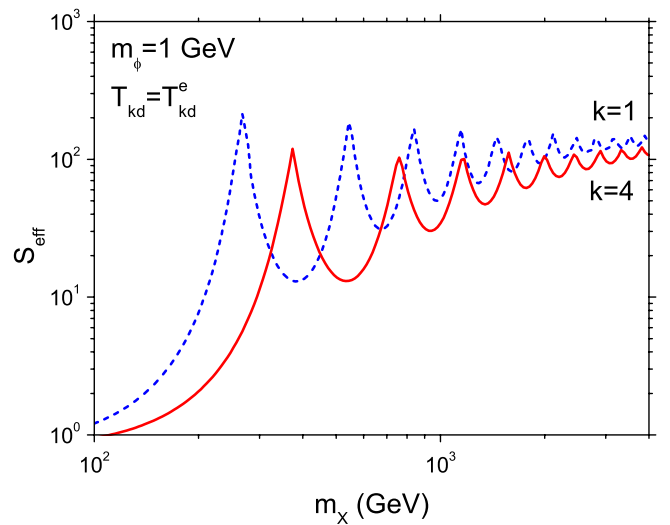


FIG. 12 (color online). The maximal effective Sommerfeld enhancement factor S_{eff} for $k = 1, 4$, where the tree-level annihilation cross section is $(\sigma_{\text{an}}v_{\text{rel}})_0 = k\pi\alpha_X^2/m_X^2$, for $m_\phi = 0.25$ GeV (left panel) and $m_\phi = 1$ GeV (right panel). The effects scale roughly as $1/\sqrt{k}$, with the deviation from $1/\sqrt{k}$ scaling due to the Sommerfeld effect in the early Universe.

freeze out analysis. The main point is that nonminimal models will typically include additional hidden sector states and additional annihilation channels, generically reducing the maximum possible S_{eff} consistent with thermal relic density constraints.

2. Multistate dark matter

Two-state dark matter models have received considerable attention recently. For example, “inelastic dark matter” has been motivated by the DAMA NaI signal [67,68] and “exciting dark matter” has been proposed to explain the INTEGRAL/SPI excess towards the Galactic center [69]. They have also been proposed as explanations of the PAMELA excess and the WMAP haze towards the Galactic center [38,70]. In these models, the stable dark matter state is accompanied by a more massive unstable state, which is separated from the stable state by a mass gap that is chosen to fit the data and is much smaller than the dark matter mass. The presence of this almost degenerate excited state significantly complicates the early Universe freeze out analysis. The Sommerfeld enhancements are also different [25]. To compute the relic density, one would need to include the additional annihilation channels and the two-state dark matter Sommerfeld enhancement in a self-consistent early Universe freeze out calculation.

Dark matter models with multiple *stable* dark matter states have also been proposed to explain the PAMELA and Fermi excesses and the DAMA/LIBRA signal [71–73]. These models provide an interesting avenue to reduce the required Sommerfeld enhancement to explain both the PAMELA and Fermi excesses [73]. This is because the heavier (\sim TeV) component contributes mainly to the Fermi excess, whereas the lighter (\sim 100 GeV) dark matter particle contributes predominantly to the PAMELA positron excess. The required Sommerfeld enhancements are about 100 for the TeV state.

Such enhancements may appear to be consistent with our relic density constraint to within a factor of 2, but the relic density calculation that we have presented here is not complete in these scenarios. In the simplest such models [73], the two states couple to the dark gauge boson with the same charge. As stated, however, these models are ruled out by the direct detection limits, because the $Xq \rightarrow Xq$ cross section due to the kinetic mixing would be too large for $\epsilon > 10^{-6}$ [17,36]. The way out of this constraint is to specify that each state is further split into a metastable and stable state with very small mass gap and that the ϕ particles only have off-diagonal couplings, so that kinetic energy larger than the mass gap is required to initiate $Xq \rightarrow X^*q$, where X^* denotes the excited state of X [17]. These mass gaps may then be chosen to be $\sim \mathcal{O}(\text{MeV})$ or $\sim \mathcal{O}(0.1 \text{ MeV})$ to explain INTEGRAL/SPI and DAMA/LIBRA signals, respectively [73]. The relic density calculation is now significantly more complicated. Additional annihilation channels, for example, annihilations between

the heavy and light states and annihilation to $h\phi$, and the presence of an almost degenerate state must be taken into account in the early Universe calculation. The Sommerfeld effect for the lighter state is generically important because the coupling to the dark gauge boson would be set by the relic density constraint for the heavy particle, and so $\alpha_X \sim 0.02$, much larger than the required α_X for thermal dark matter with mass ~ 100 GeV. The new channels, such as the $h\phi$ channel, also modify comparisons to PAMELA and Fermi as these new channels may contribute to the e^\pm flux with a different energy spectrum. These issues are beyond the scope of the present work.

Multistate dark matter theories may also have significantly suppressed k . For example, in the non-Abelian models of Ref. [74], values of k as small as 0.14 are possible, and these have recently been claimed to provide marginally consistent explanations of the cosmic ray anomalies [75]. In these studies, however, the irreducible annihilations to Higgs bosons have not been included, and the Sommerfeld effect on freeze out has been neglected. Given the low value of k , these changes to the annihilation cross section may be important. It would be interesting to see if the consistency remains after a complete analysis.

To summarize, finding a model with more degrees of freedom that requires smaller boosts to fit the PAMELA and Fermi data does not by itself imply an improvement, as the bounds on S_{eff} will generically also be stronger. Generalizations of the freeze out analysis described here are required to determine if these scenarios may self-consistently explain PAMELA and Fermi data.

3. Very heavy force carriers with $m_\phi \gg 1$ GeV

One reason the maximal S_{eff} is somewhat limited is that the resonances are not significant for $m_\phi \lesssim 1$ GeV and $m_X \gtrsim 1$ TeV. For larger m_ϕ , however, the resonances move to larger m_X , as evident from Eq. (7). Generically, for $m_\phi > 2$ GeV, ϕ decay may produce antiprotons. However, if such contributions to the antiproton spectrum may be otherwise suppressed or accommodated [76], such large m_ϕ may allow one to improve the consistency of Sommerfeld-based explanations of the PAMELA and Fermi data.

As an example, in Ref. [19], a leptophilic model was constructed where the antiproton flux is suppressed by dynamics, freeing the force-carrier mass to be much larger than \sim GeV. For $m_\phi \sim 10$ GeV and $m_X \approx 800$ GeV, the authors noted that resonant Sommerfeld enhancement could enhance signals with enhancement factors as large as $S \sim 1000$. However, the study of Ref. [19] did not include the Higgs annihilation channel, and did not include the impact of Sommerfeld enhancement on freeze out. Note also that, as discussed in Sec. III B, for very large m_ϕ near the freeze out temperature, ϕ particles may not be in complete thermal equilibrium with the standard model, modifying the freeze out analysis. Nevertheless, such

possibilities certainly merit further study. Note, though, that these models are inaccessible to searches for GeV forces.

4. Running α_X

In all of our analysis above, we have assumed that the running of the gauge coupling in the dark sector is not significant. Although this is certainly a good approximation for the minimal Abelian model we have considered, the situation is much more complicated for non-Abelian models. It has been recently pointed out that a non-Abelian hidden sector could explain both PAMELA and Fermi, while resulting in the correct relic density [77]. This work notes that the Sommerfeld effect is a soft process and the enhancement factor S is determined by the coupling at scale m_ϕ , whereas for the tree-level cross section the coupling must be evaluated at m_X . For a non-Abelian theory where the coupling constant is smaller at higher energies, this could significantly relax the tension between the relic density constraint and the requirements for explaining PAMELA and Fermi.

However, this analysis [77] did not include the Sommerfeld effect on the annihilation in the early Universe. This would be particularly important because the Sommerfeld enhancement is now much larger even in the early Universe. The model also includes a Higgs doublet H and the dark matter annihilation $XX \rightarrow HZ$ must also be included. The stability of the Higgs sector is a major concern for these models because the Higgs annihilation cross section is too small to suppress their relic density [77]. The decay lifetimes that have been computed [77] are long ($\sim 10^6$ s or longer) and could destroy the successful predictions of big bang nucleosynthesis (BBN) because of the injection of electromagnetic energy [78,79]. Given these issues, it is not clear that these models can successfully satisfy the relic density constraints, but future work addressing these points could be an interesting way forward to explain PAMELA and Fermi excesses using Sommerfeld-enhanced annihilation.

We note that the arguments regarding the decay of the hidden sector Higgs boson are quite general. The hidden Higgs boson should decay to standard model particles, otherwise it would contribute to the dark matter abundance. In doing so, the decays should not destroy the successful predictions of BBN.

VI. CONCLUSIONS

In this work, we have investigated scenarios with Sommerfeld-enhanced annihilation of thermal dark matter and determined the largest possible indirect signals, subject to the most basic of constraints, namely, that the thermal dark matter have the correct thermal relic density. In addition to their potential impact on cosmic e^\pm spectra, such scenarios may also predict observable effects in the cosmic microwave background and the diffuse extragalactic γ -ray

background, excess γ -ray, neutrino, and radio wave emission from the Galactic center, and excess γ -rays from dwarf spheroidal satellites of the Milky Way and extended regions centered on the Galactic center [21,46,61,62,80–88]. The absence of each of these potential signatures provides significant constraints. The bounds derived here are complementary to those, in that these bounds are derived by studying not the annihilation products, but the production. These results are therefore model-independent and robust, in that they assume only that the dark matter is produced by thermal freeze out, which is the central assumption motivating the consideration of Sommerfeld enhancements in the first place.

This analysis included the possibility of resonant Sommerfeld enhancement and the effect of Sommerfeld enhancement in the thermal relic density calculation. The possibility of Sommerfeld enhancement at low velocities, with annihilation cross sections proportional to v_{rel}^{-1} , or even v_{rel}^{-2} on resonance, implies that the thermal relic density is sensitive to the dark matter density's evolution long after conventional freeze out at $T_f \sim m_X/25$. As a result, the relic density depends on many aspects absent in conventional scenarios. These include the cutoff of resonant Sommerfeld annihilation, the equilibration of force-carrier particles, the temperature of kinetic decoupling T_{kd} , and the efficiency of self-scattering to thermalize the dark matter velocity distribution after T_{kd} .

These dependences may be very important. A nongeneric but intriguing example is that, on resonance, dark matter freezes out but may then later melt in, chemically recoupling for a second era of efficient annihilation. As a result, exact resonances suppress rather than enhance indirect signals, in contrast to naive expectations. More generally, the Sommerfeld enhancement of freeze out implies that smaller tree-level annihilation cross sections are required to give the correct relic density, and so reduces the effective Sommerfeld enhancement of indirect signals. We have found reduction factors of a few, with several $\sim 10\%$ – 30% variations possible, depending on the value of T_{kd} and other freeze out parameters.

For the minimal scenario analyzed here, and adopting the most optimistic assumptions to maximize indirect signals, the largest possible values of S_{eff} are 7, 30, and 90 for $m_\phi = 250$ MeV and $m_X = 0.1, 0.3,$ and 1 TeV, respectively. As shown in Figs. 9 and 10, such enhancements fall short of explaining the PAMELA and Fermi cosmic e^\pm excesses. For the best-fit point, as discussed in Sec. V, this discrepancy is over an order of magnitude and cannot be eliminated by appeals to enhanced local dark matter density or boosts from small scale structure. Bound state effects are also unable to eliminate the inconsistency, as they are typically insignificant in the parameter regions that may potentially explain the cosmic ray anomalies, and are in any case stringently constrained by the CMB. Along with the astrophysical uncertainties discussed in Sec. V and

the complementary constraints from the cosmic microwave background and other sources, these results imply that the dark matter motivations for Sommerfeld enhancements and GeV dark forces at present are, at best, strained.

We have discussed several possibly interesting directions for further study in Sec. V. Nonminimal models will generically include additional annihilation channels, such as those involving hidden Higgs bosons. These will generically strengthen the bounds on S_{eff} . We noted also that the decay of the hidden sector Higgs boson to standard model particles could be constrained by demanding consistency with the measured light element abundances produced during BBN. Additional features and constraints particular to specific models have been noted in Sec. V. We have focused on the 4μ final state here, which generically give the best fits to the data. Different annihilation channels, such as $4e$ final states, are optimized at lower S_{eff} and m_X , but also yield worse fits to the data according to present estimates in the literature. However, this conclusion may depend significantly on the cosmic ray propagation model parameters used. At present, no Sommerfeld models have been shown to explain the PAMELA and Fermi excesses consistent with the constraint of thermal freeze out, and we stress that complete and dedicated studies of freeze out are required to judge properly whether more complicated models can self-consistently explain PAMELA and Fermi data.

The maximal Sommerfeld enhancements derived here, although not sufficient to explain current PAMELA and Fermi excesses, are nevertheless significant enhancements. These will be probed as PAMELA and Fermi continue to gather data, and, in the near future, at AMS on the International Space Station. The interpretation of future data as dark matter may, of course, continue to be clouded by confusion from astrophysical uncertainties, but the diverse set of particle physics and astroparticle physics experiments that may probe weak-scale dark matter in the near future at least provides some reason for optimism.

ACKNOWLEDGMENTS

We thank Masahiro Ibe, Jason Kumar, Erich Poppitz, and Neil Weiner for useful conversations and correspondence, and Tracy Slatyer for motivating more detailed investigation of resonant Sommerfeld enhancements and comments on an early draft of this work. We thank Marc Kamionkowski for pointing out the importance of inverse decays for the force carrier in equilibrium, and the participants of the Aspen Summer 2010 workshop ‘‘From Colliders to the Dark Sector’’ for stimulating discussions. The work of J. L. F. and H. Y. was supported in part by NSF Grants No. PHY-0653656 and No. PHY-0709742. The work of M. K. was supported in part by NSF Grant No. PHY-0855462 and NASA Grant No. NNX09AD09G.

-
- [1] H. Goldberg, *Phys. Rev. Lett.* **50**, 1419 (1983); **103**, 099905(E) (2009).
 - [2] J. L. Feng, M. Kaplinghat, and H. B. Yu, *Phys. Rev. Lett.* **104**, 151301 (2010).
 - [3] S. W. Barwick *et al.* (HEAT Collaboration), *Astrophys. J.* **482**, L191 (1997).
 - [4] J. J. Beatty *et al.*, *Phys. Rev. Lett.* **93**, 241102 (2004).
 - [5] O. Adriani *et al.* (PAMELA Collaboration), *Nature (London)* **458**, 607 (2009).
 - [6] J. Chang *et al.* (ATIC Collaboration), *Nature (London)* **456**, 362 (2008).
 - [7] A. A. Abdo *et al.* (The Fermi LAT Collaboration), *Phys. Rev. Lett.* **102**, 181101 (2009).
 - [8] A. W. Strong *et al.*, contribution to the 31st ICRC, Lodz, Poland, July 2009.
 - [9] A. Sommerfeld, *Ann. Phys. (Leipzig)* **403**, 257 (1931).
 - [10] H. Baer, K. m. Cheung, and J. F. Gunion, *Phys. Rev. D* **59**, 075002 (1999).
 - [11] J. Hisano, S. Matsumoto, and M. M. Nojiri, *Phys. Rev. D* **67**, 075014 (2003).
 - [12] J. Hisano, S. Matsumoto, and M. M. Nojiri, *Phys. Rev. Lett.* **92**, 031303 (2004).
 - [13] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, *Phys. Rev. D* **71**, 063528 (2005).
 - [14] J. Hisano, S. Matsumoto, O. Saito, and M. Senami, *Phys. Rev. D* **73**, 055004 (2006).
 - [15] M. Cirelli, A. Strumia, and M. Tamburini, *Nucl. Phys.* **B787**, 152 (2007).
 - [16] M. Cirelli, M. Kadastik, M. Raidal, and A. Strumia, *Nucl. Phys.* **B813**, 1 (2009).
 - [17] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, *Phys. Rev. D* **79**, 015014 (2009).
 - [18] M. Pospelov and A. Ritz, *Phys. Lett. B* **671**, 391 (2009).
 - [19] P. J. Fox and E. Poppitz, *Phys. Rev. D* **79**, 083528 (2009).
 - [20] L. Bergstrom, J. Edsjo, and G. Zaharijas, *Phys. Rev. Lett.* **103**, 031103 (2009).
 - [21] P. Meade, M. Papucci, A. Strumia, and T. Volansky, *Nucl. Phys.* **B831**, 178 (2010).
 - [22] R. Iengo, *J. High Energy Phys.* 05 (2009) 024.
 - [23] R. Iengo, arXiv:0903.0317.
 - [24] S. Cassel, *J. Phys. G* **37**, 105009 (2010).
 - [25] T. R. Slatyer, *J. Cosmol. Astropart. Phys.* 02 (2010) 028.
 - [26] J. Hisano, S. Matsumoto, M. Nagai, O. Saito, and M. Senami, *Phys. Lett. B* **646**, 34 (2007).
 - [27] J. March-Russell, S. M. West, D. Cumberbatch, and D. Hooper, *J. High Energy Phys.* 07 (2008) 058.
 - [28] Q. Yuan, X. J. Bi, J. Liu, P. F. Yin, J. Zhang, and S. H. Zhu, *J. Cosmol. Astropart. Phys.* 12 (2009) 011.

- [29] J. B. Dent, S. Dutta, and R. J. Scherrer, *Phys. Lett. B* **687**, 275 (2010).
- [30] J. Zavala, M. Vogelsberger, and S. D. M. White, *Phys. Rev. D* **81**, 083502 (2010).
- [31] P. Gondolo and G. Gelmini, *Nucl. Phys.* **B360**, 145 (1991).
- [32] E. W. Kolb and M. S. Turner, *Front. Phys.* **69**, 1 (1990).
- [33] B. Holdom, *Phys. Lett. B* **166**, 196 (1986).
- [34] B. Batell, M. Pospelov, and A. Ritz, *Phys. Rev. D* **79**, 115008 (2009).
- [35] J. D. Bjorken, R. Essig, P. Schuster, and N. Toro, *Phys. Rev. D* **80**, 075018 (2009).
- [36] M. Pospelov, A. Ritz, and M. B. Voloshin, *Phys. Lett. B* **662**, 53 (2008).
- [37] J. L. Feng, H. Tu, and H. B. Yu, *J. Cosmol. Astropart. Phys.* **10** (2008) 043.
- [38] D. P. Finkbeiner, T. R. Slatyer, N. Weiner, and I. Yavin, *J. Cosmol. Astropart. Phys.* **09** (2009) 037.
- [39] J. L. Feng, M. Kaplinghat, H. Tu, and H. B. Yu, *J. Cosmol. Astropart. Phys.* **07** (2009) 004.
- [40] L. Ackerman, M. R. Buckley, S. M. Carroll, and M. Kamionkowski, *Phys. Rev. D* **79**, 023519 (2009).
- [41] M. R. Buckley and P. J. Fox, *Phys. Rev. D* **81**, 083522 (2010).
- [42] M. Ibe and H. B. Yu, *Phys. Lett. B* **692**, 70 (2010).
- [43] S. A. Khrapak *et al.*, *Phys. Rev. Lett.* **90**, 225002 (2003).
- [44] S. A. Khrapak *et al.*, *IEEE Trans. Plasma Sci.* **32**, 555 (2004).
- [45] X. X. Xue *et al.* (SDSS Collaboration), *Astrophys. J.* **684**, 1143 (2008).
- [46] K. N. Abazajian, P. Agrawal, Z. Chacko, and C. Kilic, *arXiv:1002.3820*.
- [47] M. Kamionkowski and A. Kinkhabwala, *Phys. Rev. D* **57**, 3256 (1998).
- [48] R. Catena and P. Ullio, *J. Cosmol. Astropart. Phys.* **08** (2010) 004.
- [49] M. Weber and W. de Boer, *arXiv:0910.4272*.
- [50] P. Salucci, F. Nesti, G. Gentile, and C. F. Martins, *arXiv:1003.3101*.
- [51] J. M. Cline, A. C. Vincent, and W. Xue, *Phys. Rev. D* **81**, 083512 (2010).
- [52] M. Vogelsberger *et al.*, *Mon. Not. R. Astron. Soc.* **395**, 797 (2009).
- [53] M. Kuhlen and D. Malyshev, *Phys. Rev. D* **79**, 123517 (2009).
- [54] D. Hooper, P. Blasi, and P. D. Serpico, *J. Cosmol. Astropart. Phys.* **01** (2009) 025.
- [55] H. Yuksel, M. D. Kistler, and T. Stanev, *Phys. Rev. Lett.* **103**, 051101 (2009).
- [56] S. Profumo, *arXiv:0812.4457*.
- [57] S. Dado and A. Dar, *Mem. Sci. Astron. Ital.* **81**, 132 (2010).
- [58] P. L. Biermann *et al.*, *Phys. Rev. Lett.* **103**, 061101 (2009).
- [59] B. Katz, K. Blum, J. Morag, and E. Waxman, *Mon. Not. R. Astron. Soc.* **405**, 1458 (2010).
- [60] D. Grasso *et al.* (FERMI-LAT Collaboration), *Astropart. Phys.* **32**, 140 (2009).
- [61] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, *Phys. Rev. D* **80**, 023505 (2009).
- [62] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, *Phys. Rev. D* **80**, 043526 (2009).
- [63] V. Barger, Y. Gao, W. Y. Keung, D. Marfatia, and G. Shaughnessy, *Phys. Lett. B* **678**, 283 (2009).
- [64] A. Falkowski, J. T. Ruderman, T. Volansky, and J. Zupan, *J. High Energy Phys.* **05** (2010) 077.
- [65] J. D. March-Russell and S. M. West, *Phys. Lett. B* **676**, 133 (2009).
- [66] W. Shepherd, T. M. P. Tait, and G. Zaharijas, *Phys. Rev. D* **79**, 055022 (2009).
- [67] D. Tucker-Smith and N. Weiner, *Phys. Rev. D* **64**, 043502 (2001).
- [68] S. Chang, G. D. Kribs, D. Tucker-Smith, and N. Weiner, *Phys. Rev. D* **79**, 043513 (2009).
- [69] D. P. Finkbeiner and N. Weiner, *Phys. Rev. D* **76**, 083519 (2007).
- [70] I. Cholis, L. Goodenough, and N. Weiner, *Phys. Rev. D* **79**, 123505 (2009).
- [71] K. M. Zurek, *Phys. Rev. D* **79**, 115002 (2009).
- [72] A. Katz and R. Sundrum, *J. High Energy Phys.* **06** (2009) 003.
- [73] I. Cholis and N. Weiner, *arXiv:0911.4954*.
- [74] F. Chen, J. M. Cline, and A. R. Frey, *Phys. Rev. D* **80**, 083516 (2009).
- [75] M. Cirelli and J. M. Cline, *Phys. Rev. D* **82**, 023503 (2010).
- [76] G. Kane, R. Lu, and S. Watson, *Phys. Lett. B* **681**, 151 (2009).
- [77] H. Zhang, C. S. Li, Q. H. Cao, and Z. Li, *Phys. Rev. D* **82**, 075003 (2010).
- [78] M. Kawasaki and T. Moroi, *Astrophys. J.* **452**, 506 (1995).
- [79] R. H. Cyburt, J. R. Ellis, B. D. Fields, and K. A. Olive, *Phys. Rev. D* **67**, 103521 (2003).
- [80] M. Kamionkowski and S. Profumo, *Phys. Rev. Lett.* **101**, 261301 (2008).
- [81] A. V. Belikov and D. Hooper, *Phys. Rev. D* **80**, 035007 (2009).
- [82] S. Profumo and T. E. Jeltema, *J. Cosmol. Astropart. Phys.* **07** (2009) 020.
- [83] G. Huetsi, A. Hektor, and M. Raidal, *Astron. Astrophys.* **505**, 999 (2009).
- [84] M. Cirelli, F. Iocco, and P. Panci, *J. Cosmol. Astropart. Phys.* **10** (2009) 009.
- [85] T. Kanzaki, M. Kawasaki, and K. Nakayama, *Prog. Theor. Phys.* **123**, 853 (2010).
- [86] R. M. Crocker, N. F. Bell, C. Balazs, and D. I. Jones, *Phys. Rev. D* **81**, 063516 (2010).
- [87] A. Calvez, W. Essey, M. Fairbairn, A. Kusenko, and M. Loewenstein, *arXiv:1003.1113*.
- [88] S. Palomares-Ruiz and J. M. Siegal-Gaskins, *J. Cosmol. Astropart. Phys.* **07** (2010) 023.