PHYSICAL REVIEW D **79,** 115019 (2009)

Direct detection of multicomponent secluded WIMPs

Brian Batell,[1] Maxim Pospelov,[1,2] and Adam Ritz[2]

[1]*Perimeter Institute for Theoretical Physics, Waterloo, ON, N2J 2W9, Canada*

[2]*Department of Physics and Astronomy, University of Victoria, Victoria, BC, V8P 1A1 Canada*

(Received 27 March 2009; published 25 June 2009)

Dark matter candidates comprising several substates separated by a small mass gap Δm, and coupled to the standard model by (sub-)GeV force carriers, can exhibit nontrivial scattering interactions in direct-detection experiments. We analyze the secluded $U(1)_S$-mediated weakly interacting massive particle (WIMP) scenario, and calculate the elastic and inelastic cross sections for multicomponent WIMP scattering off nuclei. We find that second-order elastic scattering, mediated by virtual excited states, provides strong sensitivity to the parameters of the model for a wide range of mass splittings, while for small Δm the WIMP excited states have lifetimes exceeding the age of the Universe, and generically have a fractional relative abundance above 0.1%. This generates even stronger constraints for $\Delta m \lesssim 200$ keV due to exothermic deexcitation events in detectors.

DOI: 10.1103/PhysRevD.79.115019 PACS numbers: 95.35.+d

I. INTRODUCTION

Many concordant aspects of cosmology and astrophysics present us with compelling evidence for a universe in which dark matter (DM) comprises about one-quarter of the total energy density in the current Universe. However, while we have ample evidence for its gravitational interaction, the details of its nongravitational dynamics, if any, and thus its place within particle physics, remain obscure. The importance of this question, and the strong motivation it implies for physics beyond the standard model (SM), has led to an expansive experimental program, both terrestrially and in space, aimed at detecting dark matter through nongravitational interactions, namely, annihilation, scattering, or decay [1].

Among the best motivated scenarios for particle dark matter is a generic weakly interacting massive particle (WIMP)—namely, a weak-scale massive particle, thermally populated during the early Universe, and subsequently depleted by a weak-scale annihilation rate [2]. This framework has important implications, as it suggests the natural scale for self-annihilation and, albeit in a less direct manner, gives guidance as to the likely level of scattering with normal matter. Consequently, searches for WIMPs present in the galactic halo via their scattering off nuclei in radioactively pure underground detectors have become an integral part of modern subatomic physics [3–5]. Existing searches have primarily focused on nucleon-WIMP elastic scattering, and the null results have placed significant upper limits on the elastic cross section that are now beginning to probe the natural parameter range [1].

The chances for successful direct detection depend rather sensitively on whether the WIMP is an isolated state, such as a real scalar or a Majorana fermion, or a multicomponent state, such as a complex scalar or a Dirac fermion. Since the latter allows for the existence of vector and/or tensor currents, in many scenarios Z- or γ-mediated scattering can result in large elastic cross sections. However, even a small mass splitting Δm between the WIMP components—that we will generically denote χ_1 and χ_2—which allows for coupling to these currents may significantly alter the scattering signal if Δm is in excess of the typical kinetic energy of the WIMP-nucleus system. This was first noticed a decade ago [6,7], where it was shown that an otherwise large sneutrino scattering cross section off nuclei, mediated by Z exchange, could be drastically reduced by splitting the two components of the complex scalar sneutrino by Δm, with $E_{\rm kin} < \Delta m \ll m_{\tilde{\nu}}$. This idea was later exploited in Ref. [8], where the inelastic scattering of WIMP states split by $\Delta m \sim \mathcal{O}(100)$ keV was utilized to reconcile the annual modulation of energy deposition seen by DAMA with the null results of other direct-detection experiments using lighter nuclei. It is of relevance here that the signature of inelastic endothermic WIMP scattering, which is enhanced for heavy nuclei [8], differs significantly from elastic scattering.

The possibility of *exothermic* WIMP-nucleus scattering, i.e. with significant energy release, was studied in Ref. [9], where it was shown that a splitting $\Delta m \sim \mathcal{O}(10$ MeV) between neutral χ_1 and electrically charged $\chi_2^{\pm}$ components of the WIMP sector will lead to WIMP-nucleus recombination with heavy nuclei via the formation of a bound state between the nucleus and the higher-mass negatively charged partner χ_2^-. A significant amount of energy, of $\mathcal{O}(1$–$10)$ MeV, can be released this way via e^+, γ, n, or ν emission depending on the particle physics realization of this scenario. Reference [9] also demonstrated the possibility for an $\mathcal{O}(\Delta m)^{-1}$ enhancement in the elastic scattering amplitude due to virtual excitation of the WIMP substructure.

In general terms, the consideration of multicomponent WIMP scenarios with relatively small splittings has in the past been motivated on several fronts. On one hand, due to

collider constraints, supersymmetric scenarios such as the constrained minimal supersymmetric standard model generally require an enhanced annihilation cross section in the early Universe to avoid overproducing neutralino dark matter, and viable parameter ranges usually make use of coannihilation with nearby charged states [10]. Another motivation comes from the possibility of the catalysis of nuclear reactions in the early Universe leading to the resolution of the cosmological lithium problem [11]. MeV-scale splittings also allow for the possibility of sourcing the galactic 511 keV line via the decay of excited WIMP states [12,13].

Recently, further motivation for the study of multicomponent WIMP scenarios has come from claims of a positron excess in cosmic rays above 10 GeV [14], and an enhancement in the total electron/positron flux around 800 GeV [15]. While the interpretation of these anomalies remains an active topic of debate, it is tempting to speculate on an origin related to dark matter. However, this requires a significant enhancement of the annihilation rate associated with cosmological freeze-out, by factors of 10–1000 depending on the WIMP mass. This is perhaps most naturally achieved via a new light mediator with GeV or sub-GeV mass [16,17], which allows for a Sommerfeld-type enhancement of annihilation at low velocities in the present halo. Decays into these metastable mediators then lead to a dominant leptonic branching fraction for kinematic reasons [16–18]. For the present discussion, the salient feature of the enhancement mechanism is that the required Coulomb-like potential arises most naturally if the WIMPs are multicomponent states, such as complex scalars or Dirac fermions, or have at most a small mass splitting between the components [16,17].[1]

Perhaps the simplest and most natural realization of this proposal [16,17] is via a secluded $U(1)_S$ gauge interaction that acts in the dark matter sector and couples to SM particles via kinetic mixing with SM hypercharge $U(1)_Y$ [19]. Such a coupling, $\kappa F^S_{\mu\nu} F^Y_{\mu\nu}$, provides one of the few renormalizable portals for coupling the SM to new (SM-singlet) physics [20]. The particle phenomenology of a (sub-)GeV secluded $U(1)_S$ sector has recently been addressed in a number of studies [21], with the conclusion that the current sensitivity to κ is in the range $\kappa \sim \mathcal{O}(10^{-3}$–$10^{-2})$. This does not place significant restrictions on WIMP properties, as κ does not enter the freeze-out cross section [19], but is interesting in its own right as it is close to the natural radiatively generated value [21,22].

The question of direct WIMP-nucleus scattering has not yet been addressed in detail for $U(1)_S$-mediated scenarios, despite the high level of recent interest. The cross section was calculated to first order in [19] in the limit of small Δm, where scattering is mediated by a vector current and the WIMP charge radius. This cross section is generically large, and leads to strong limits on κ [16,17], well outside both its natural range due to radiative mixing, and consequently the experimental reach [21]. While this "secluded" regime is of interest for indirect detection, it is important to realize, as discussed above, that such a scattering process can be switched off by even a small mass splitting of the multicomponent WIMPs that couple to the $U(1)_S$ vector current [16,17], in complete parallel with earlier studies [6–8]. In this case, elastic WIMP-nucleus scattering may still proceed, but now at second order with off-shell excited states. This echoes earlier calculations in [9], where an electrically charged partner of the WIMP state is virtually produced and absorbed in the elastic scattering process.

In this paper, we calculate the WIMP-nucleus cross sections for inelastic scattering at first order, and elastic scattering at second order, for the minimal $U(1)_S$-mediated model, and set constraints on combinations of parameters, such as the mass splitting Δm, the kinetic mixing parameter κ, and the mediator mass m_V. We also observe that for light mediator masses and large mixing, the interaction may become sufficiently strong that perturbation theory breaks down. Although Sommerfeld-type enhancement is not important for direct scattering with $\kappa \ll 1$ and $v \sim 10^{-3}$ in the halo, recombination through the formation of a WIMP-nucleus bound state can be kinematically accessible for a relevant range of parameters [17]. Another interesting aspect of the minimal $U(1)_S$-mediated WIMP model is the long lifetime of χ_2, the excited WIMP state(s), for small $\Delta m < 2m_e$. For lifetimes in excess of the age of the Universe, this immediately raises the possibility of collisional deexcitation within the detector. Not only are such processes allowed at first order in perturbation theory, but they will proceed at a constant rate even in the limit of very small relative velocity. This can significantly enhance the energy deposition within the detector, and we use this signal to set stringent constraints on the parameter space of the model. These limits employ an estimate of the minimal concentration of excited states surviving from the early Universe or regenerated locally in the galaxy.

This paper is organized as follows. In the next section we introduce the $U(1)_S$ model with a mass splitting of the multicomponent WIMPs, its low-energy effective Lagrangian, and the lifetime of the excited states. Section III contains the calculation of elastic and inelastic scattering cross sections for various parameter choices, and considers the effects of WIMP-nucleus binding. Section IV estimates the fractional cosmological freeze-out abundance of excited states and their regeneration in the galaxy, and sets limits on the parameters of the model from deexcitation in inelastic exothermic scattering for small Δm. We finish with some concluding remarks in Sec. V.

[1]We note in passing that pseudodegenerate χ_1 and $\chi_2^\pm$ WIMP multiplets would also boost the annihilation cross section by orders of magnitude due to the possibility of $(\chi_2^+ \chi^-)$ resonant scattering, which, in particular, enhances neutralino annihilation into gauge bosons and leptons provided the neutralino-slepton system is split by $\Delta m \sim 10$ MeV [9].

II. MULTICOMPONENT WIMPS WITH $U(1)_S$ MEDIATORS

The model introduced in Ref. [19] utilizes the kinetic mixing portal to couple the SM and (SM-singlet) DM sectors via a $U(1)_S$ mediator,

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \mathcal{L}_{U(1)_S} + \mathcal{L}_{\rm DM}, \tag{1}$$

where $\mathcal{L}_{\rm DM}$ describes the WIMP sector and its $U(1)_S$ interactions. $\mathcal{L}_{U(1)_S}$ includes the standard U(1) Lagrangian, the kinetic mixing portal, and the associated Higgs sector that gives a mass to the $U(1)_S$ gauge boson V_μ. After this symmetry breaking, the relevant part of the low-energy $U(1)_S$ Lagrangian becomes

$$\mathcal{L}_{U(1)_S} = -\frac{1}{4}V_{\mu\nu}^2 + \frac{1}{2}m_V^2 V_\mu^2 + \kappa V_\nu \partial_\mu F_{\mu\nu}, \tag{2}$$

where $F_{\mu\nu}$ is the electromagnetic field strength, and a small multiplicative shift by $\cos\theta_W$ is absorbed into κ. In this paper, for convenience, we choose the Higgs sector to have twice the $U(1)_S$ charge of the WIMPs. This allows us to introduce two types of (renormalizable) mass terms in the WIMP action, which itself can be either bosonic or fermionic:

$$\mathcal{L}_{\rm DM}^f = \bar{\psi}(iD_\mu\gamma_\mu - m_\psi)\psi + (\lambda H'\psi\psi + \text{H.c.})$$
fermionic DM, (3)

$$\mathcal{L}_{\rm DM}^b = (D_\mu\phi)^*(D_\mu\phi) - m_\phi^2\phi^*\phi + (\lambda H'\phi\phi + \text{H.c.})$$
bosonic DM. (4)

Here ψ (or, respectively, ϕ) is the WIMP field in the form of a Dirac fermion (or a charged scalar). $D_\mu = \partial_\mu + ie'V_\mu$ is the usual covariant derivative in terms of the $U(1)'$ gauge coupling e', and λ is the strength of the dark sector Yukawa interaction. After spontaneous breaking of $U(1)_S$, H' develops a vacuum expectation value, and the complex scalar ϕ (or, respectively, Dirac fermion ψ) will be split into two real (or Majorana) components which we will denote collectively as χ_1 and χ_2 with masses $m_i = m_{\chi_i}$. The low-energy Lagrangian for real scalar WIMPs is then

$$\mathcal{L}_{\rm DM}^b = \sum_{i=1,2}\left(\frac{1}{2}(\partial_\mu\chi_i)^2 - \frac{1}{2}m_i^2\chi_i^2\right) + e'V_\mu(\chi_1\partial_\mu\chi_2 - \chi_2\partial_\mu\chi_1), \tag{5}$$

while in the fermionic case, where χ_i is a two-component spinor,

$$\mathcal{L}_{\rm DM}^f = \sum_{i=1,2}\left(i\chi_i^\dagger\bar{\sigma}_\mu\partial_\mu\chi_i - \frac{1}{2}(m_i\chi_i\chi_i + \text{H.c.})\right) - ie'V_\mu(\bar{\chi}_1\bar{\sigma}_\mu\chi_2 - \bar{\chi}_2\bar{\sigma}_\mu\chi_1). \tag{6}$$

In (2), (5), and (6), we have suppressed the details of interactions with the physical Higgs, as it will not be important for the present discussion. Of more relevance here is that the WIMP masses are split by an amount that scales with the Higgs vacuum expectation value v' and the mass of the mediator:

$$\Delta m \equiv m_2 - m_1 \simeq \lambda v' \sim \frac{\lambda}{e'}m_V \ll m_{1,2}, \tag{7}$$

where the final inequality reflects the assumed hierarchy, with the mediator much lighter than the WIMP(s). For the analysis that follows, this is more relevant than the concrete mechanism for producing $\Delta m \sim v'$, as the latter can be achieved within a number of different model-dependent realizations. Equations (2), (5), and (6) constitute the starting point for the calculation of χ_1 and χ_2 scattering off nuclei. However, before proceeding in this direction, we will first discuss the requirements on the splitting imposed by the need for Sommerfeld-enhanced annihilation in the galaxy, and also the lifetime of the excited WIMP states.

A. Sommerfeld-enhanced annihilation

Sommerfeld enhancement for the scattering and annihilation of WIMPs in the galactic halo, with $(\pi\alpha'/v) \gg 1$ when $v \ll \alpha'$ and small m_V, originates from the modification of free-particle wave functions due to the impact of the Coulomb interaction when the particle separation falls below the de Broglie wavelength $(mv)^{-1}$. It is then clear that this enhancement will only arise (away from possible resonances in the interparticle potential) if the mass splitting between the states is smaller than the Coulomb (or rather Yukawa) potential energy,

$$\Delta m \lesssim V(r \simeq \lambda_{dB}) \Rightarrow \Delta m \lesssim \alpha' m v \sim E_{\rm kin}\left(\frac{\pi\alpha'}{v}\right), \tag{8}$$

where in the latter relations we also assumed that $m_V \lesssim mv$. Therefore, we find that in order to preserve a large enhancement the mass splitting must be smaller than the WIMP kinetic energy times the enhancement factor. Thus, for a TeV-scale WIMP in the galactic halo with $E_{\rm kin} \sim 1$ MeV, requiring an $\mathcal{O}(10^3)$ enhancement to "explain" the PAMELA positron excess [14] implies the splitting should not exceed 1 GeV. In this paper, we will consider a range of mass splittings but generally well below this threshold.

B. Excited state lifetime

Once produced, for example, in the early Universe, the longevity of χ_2 is very important due to the possibility for exothermic inelastic scattering in direct-detection experiments. The decay $\chi_2 \to \chi_1$ depends sensitively on the size of the splitting. For $\Delta m > 2m_e$, the photon-mediated decay to charged particles dominates, and χ_2 decays on time scales much shorter than the age of the Universe for the range in κ that we consider. Although this may still lead to observable consequences in big bang nucleosynthesis, or during later cosmological epochs [23], such effects fall outside the scope of the present paper. The other possibil-

ity, $\Delta m < 2m_e$, only allows decays to photons and neutrinos within the minimal $U(1)_S$ model.

The $\chi_2 \to \chi_1 \nu\bar{\nu}$ decay is mediated by $V - Z$ mixing, and its rate summed over three neutrino families is calculated to be

$$\Gamma_\nu = \frac{4\sin^2\theta_W^4}{315\pi^3}\frac{G_F^2\Delta m^9}{m_V^4}\frac{\alpha'\kappa^2}{\alpha} \simeq 3\times 10^{-53}\ \text{GeV}\times\frac{\alpha'}{\alpha}\left(\frac{\kappa}{10^{-3}}\right)^2\left(\frac{\Delta m}{100\ \text{keV}}\right)^9 \times\left(\frac{100\ \text{MeV}}{m_V}\right)^4, \quad (9)$$

which is much smaller than the inverse age of the Universe, $\tau_U^{-1} \simeq 1.5\times 10^{-42}$ GeV, unless $\Delta m > 1$ MeV, and/or $m_V < 10$ MeV. Note that the rate is suppressed by the ninth power of the energy released, as compared to the usual Δm^5 scaling for weak decays. This extra suppression can be traced back to the fact that $V - Z$ mixing produces a propagator proportional to q^2, which is saturated by Δm. Similar extra suppression factors are to be expected in the decay rate of any SM-singlet WIMPs with derivative interactions to the SM.

The decay to photons is mediated by the virtual production of an e^+e^- pair, leading to a loop-suppressed $\chi_2 \to \chi_1 + 3\gamma$ decay channel. Unlike (9), the decay to photons involves $V^* \to 3\gamma$, which is phase-space suppressed but does not require a weak transition. Assuming a small mass splitting of order 100 keV, we can estimate the rate using the calculated $V \to 3\gamma$ decay width for on-shell V bosons with $m_V \ll 2m_e$ [24],

$$\Gamma_V(m_V \ll 2m_e) = \frac{17\alpha^3\alpha'\kappa^2}{2^7 3^6 5^3\pi^3}\frac{m_V^9}{m_e^8}, \quad (10)$$

and replacing m_V by the momentum $q \sim \Delta m$ of the virtual V. Since the phase space for the 3-photon decay is maximized for momentum $q \sim \Delta m$, we obtain the estimate

$$\Gamma_{\chi_2\to\chi_1+3\gamma} \lesssim \Gamma_{V^*\to 3\gamma}(m_V^* \simeq \Delta m)\times\alpha'\left(\frac{\Delta m}{m_V}\right)^4 \simeq 4\times 10^{-47}\ \text{GeV}\times\left(\frac{\kappa}{10^{-3}}\right)^2\left(\frac{\Delta m}{100\ \text{keV}}\right)^{13} \times\left(\frac{100\ \text{MeV}}{m_V}\right)^4, \quad (11)$$

which is again much smaller than τ_U^{-1}, unless Δm approaches 1 MeV. Although this calculation needs to be modified in this limit, we would expect decays above the electron threshold to be fairly rapid in any case, as noted above.

We conclude that for the parameters of interest here, the excited state lifetime can be well in excess of the age of the Universe. Thus, a primordial excited state population can survive from the big bang, while scattering in the halo may also repopulate excited states in the late Universe. Of course, it is important to emphasize that the longevity of the excited states in this model follows from the highly suppressed decay rate to neutrinos and photons. The presence of additional fields, with a mass below Δm, would allow other decay channels and thus a significantly enhanced decay rate. For example, one natural possibility within the current scenario would be the presence of a very light Higgs particle, so that $\chi_2 \to \chi_1 + h'$ would be kinematically allowed. However, in this case the longevity of the Higgs particle may itself have other cosmological and astrophysical implications that lead to independent constraints.

III. ELASTIC AND INELASTIC SCATTERING OF MULTICOMPONENT WIMPS

The basic quantity which determines the mass-normalized counting rate in a given detector is the differential event rate per unit energy,

$$\frac{dR}{dE_R} = N_T\frac{\rho_\chi}{m_\chi}\int_{v_{\min}}^{v_{\max}} d^3v\, v f(v, v_E)\frac{d\sigma}{dE_R}. \quad (12)$$

N_T is the number of scattering centers per unit detector mass, $vn_\chi = v\rho_\chi/m_\chi$ is the incident WIMP flux, and $f(v, v_E)$ is the WIMP velocity distribution in the galactic halo that we take to be Maxwellian with escape velocity in the range $500\ \text{km/s} \le v_E \le 600\ \text{km/s}$ [25]. The total number of counts then follows from integrating over the energy bins and multiplying by the effective exposure, e.g. in kg-days, for a given detector [26].

The form of the counting rate (12) isolates the physics of the WIMP-nucleon interaction in the differential cross section $d\sigma/dE_R$. It has become customary to express the spin-independent differential cross section in the form [1]

$$\frac{d\sigma}{dE_R} = \frac{m_N}{2v^2}\frac{\sigma_{\text{nucl}}^{\text{eff}}}{\mu_n^2}\left[\frac{f_p Z + f_n(A-Z)}{f_{\text{nucl}}}\right]^2 F^2(E_R), \quad (13)$$

where m_N is the nuclear mass, μ_n is the WIMP-nucleon reduced mass (which is close to m_p for electroweak-scale WIMPs), and $F^2(E_R)$ denotes a possible recoil-energy-dependent form factor. This formula presumes that the force mediating the interaction is short range, so that the scattering amplitudes on individual nucleons f_{nucl} can be considered constant. $\sigma_{\text{nucl}}^{\text{eff}}$ is then the effective cross section per nucleon, which is now the standard figure of merit for comparing the sensitivities of different experiments.

We now turn specifically to the scattering of quasidegenerate multicomponent WIMPs. There are three basic processes that are possible (see Fig. 1):

(a) elastic scattering: $\chi_{(1,2)}N \to \chi_{(1,2)}N$, (14)

(b) endothermic scattering ($Q = -\Delta m$): $\chi_1 N \to \chi_2 N$, (15)

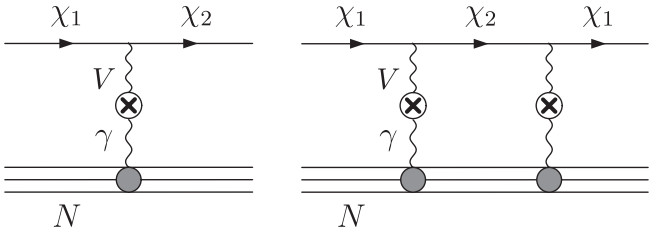

FIG. 1. First and second Born amplitudes for χ_1-nucleus scattering.

(c) exothermic scattering ($Q = \Delta m$): $\chi_2 N \to \chi_1 N$. (16)

The V-boson vertex in our model always connects χ_1 and χ_2, and thus elastic scattering can only occur starting at second order in the Born approximation, while the inelastic processes (15) and (16) can occur at leading order, and we will consider these first.

In order to reduce the parameter space of the model when calculating direct-detection constraints, we make the well-motivated assumption that the ratio α'/m_χ is restricted to (constant) values which yield the correct relic abundance, i.e. for fermionic WIMPs [17,19],

$$\alpha' = 10^{-2} \times \left(\frac{m_\chi}{270\ \text{GeV}}\right). \tag{17}$$

Further details regarding the WIMP relic abundance are discussed in Sec. IV.

A. First-order inelastic scattering

As discussed in the Introduction, the relation between Δm and the typical kinetic energy of the WIMP-nucleus pair is essential. The first-order Born approximation gives the following differential WIMP-nucleus scattering cross section:

$$\frac{d\sigma_{\chi_{1(2)}\to\chi_{2(1)}}}{d\Omega_f} = 4\mu_N^2 \alpha\alpha'\kappa^2 \frac{|\mathbf{k}_f|}{|\mathbf{k}_i|} \frac{F_N^2(q)}{(\mathbf{q}^2 + m_V^2)^2}, \tag{18}$$

where μ_N is the reduced mass of the WIMP-nucleus system, $\mathbf{k}_{i(f)}$ is the initial (final) momentum in the center-of-mass (c.o.m.) frame, $\mathbf{q}$ is the momentum transfer, and $F_N(q)$ is the nuclear charge form factor, $F_N \to Z$ for $q^{-1} \gg r_N$. The magnitude of the momentum transfer depends on the kinetic energy in the c.o.m. frame $E_{\text{kin}} = \mathbf{k}_i^2/(2\mu_N) = k^2/(2\mu_N)$, the scattering angle θ, and the mass splitting Δm,

$$\frac{\mathbf{q}^2}{2\mu_N} = 2E_{\text{kin}}\left(1 - \sqrt{1 \mp \frac{\Delta m}{E_{\text{kin}}}}\cos\theta\right) \mp \Delta m, \tag{19}$$

where the minus sign corresponds to $\chi_1 \to \chi_2$ and the plus sign to $\chi_2 \to \chi_1$ scattering. The recoil energy of the nucleus in (12) is then $E_R = \mathbf{q}^2/2m_N$. The relation (19) illustrates that for the endothermic process (15), $E_{\text{kin}}^{\text{min}} = \Delta m$, while for the exothermic scattering (16), $E_{\text{kin}}^{\text{min}} = 0$ where $\mathbf{q}^2(E_{\text{kin}} = 0) = 2\Delta m\mu_N$. In the limit $\Delta m \ll E_{\text{kin}}$, Eq. (19) reduces to the standard elastic scattering relation, $q = \sqrt{2k^2(1 - \cos\theta)}$. In the limit $E_{\text{kin}}\mu_N \ll m_V$ and neglecting the form factor dependence, $r_N \lesssim (E_{\text{kin}}\mu_N)^{1/2}$, the total cross section reduces to the elementary formula [19]

$$\sigma_{\chi_{1(2)}\to\chi_{2(1)}} = \frac{16\pi Z^2\alpha\alpha'\kappa^2\mu_N^2}{m_V^4}, \tag{20}$$

which can be interpreted as the scattering of a WIMP due to its finite electromagnetic charge radius [27]. In the opposite limit of small m_V, $E_{\text{kin}}\mu_N \gg m_V$, the scattering becomes Rutherford-like, and the total cross section saturates at the minimal momentum transfer, $\sigma \sim 1/\mathbf{q}_{\text{min}}^2$, which in an experimental setting would be determined by the lowest detectable recoil energy E_R. The generalization of (20) to a finite mass splitting in the limit $E_{\text{kin}}\mu_N$, $\Delta m\mu_N \ll m_V$ is straightforward:

$$\sigma_{\chi_{1(2)}\to\chi_{2(1)}} = \frac{16\pi Z^2\alpha\alpha'\kappa^2\mu_N^2}{m_V^4}\sqrt{1 \mp \frac{\Delta m}{E_{\text{kin}}}}. \tag{21}$$

For slow incoming particles with $\Delta m \gg E_{\text{kin}}$, we see that for exothermic scattering the factor $\sqrt{\cdots} \propto 1/v$ so that the rate is enhanced by the familiar $\sigma \sim 1/v$ dependence of the cross section.

To set constraints on the parameters of the U(1)$_S$ model, we insert the differential cross section (18) into Eq. (12) and calculate the 90% Poisson confidence limits (CL) from direct-detection data. In order to facilitate as simple a comparison as possible between constraints from endothermic, exothermic, and elastic scattering, in this paper we will restrict ourselves to data from the CDMS experiment [3]. Conservative limits result from the assumption that the galactic dark matter component is dominated by χ_1, and so we focus for the moment on the endothermic first-order process. We shall return to constraints following from exothermic $\chi_2 \to \chi_1$ scattering in the next section. As can be seen in Fig. 2, the strong sensitivity to κ/m_V^2 at $\Delta m = 0$ noted in [16,17] diminishes as Δm increases. Moreover, for a given recoil energy, there exists a certain critical value of mass splitting above which kinematics forbids the endothermic inelastic scattering. Thus if Δm is large enough the sensitivity is completely lost, and this occurs for $\Delta m \gtrsim 190$ keV for the parameters used in Fig. 2.

B. Second-order elastic scattering

With a larger mass splitting $\Delta m \gtrsim$ MeV between the WIMP χ_1 and its excited partner χ_2, kinematics forbids the first-order scattering process $\chi_1 N \to \chi_2 N$. However, elastic scattering (14) can still proceed via double $V - \gamma$ exchange, as depicted in Fig. 1. For a light U(1)$_S$ vector mediator with $m_V \lesssim 50$ MeV or so, the WIMP-nucleus

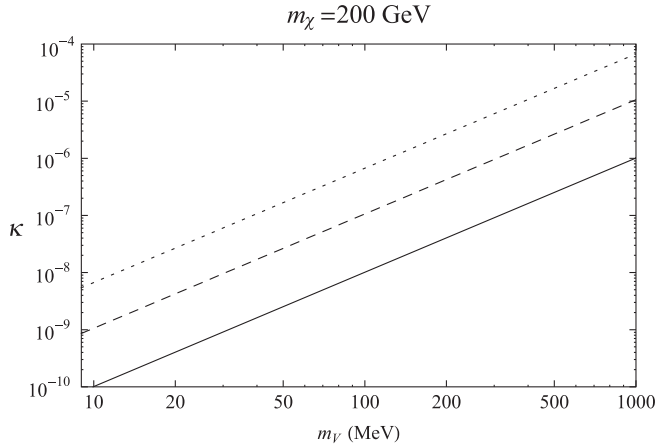

FIG. 2. *Endothermic inelastic scattering constraints*: 90% CDMS confidence limits on κ as a function of vector mass m_V for a 200 GeV WIMP, with α' chosen to yield the correct thermal relic abundance. We show constraints from the endothermic scattering $\chi_1 N \to \chi_2 N$ for $\Delta m = 0$ (solid line), 100 keV (dashed line), and 150 keV (dotted line). No endothermic events are expected in this case for splittings above $\Delta m \sim 190$ keV.

interaction range is comparable to or larger than the typical nuclear radius r_N, leading to saturation of the virtual momenta at $|\mathbf{q}_{\rm virt}| \sim m_V \lesssim r_N^{-1}$ and resulting in an $\mathcal{O}(Z^2)$ scaling of the WIMP-nucleus scattering amplitude. This implies that such an amplitude cannot be decomposed into a sum of scattering processes on individual nucleons. Consequently, in this regime Eq. (13) is applicable only as a definition of the effective nucleon amplitudes that would lead to the equivalent WIMP-nucleus cross section. However, once the range of the mediating force falls below 1 fm and m_V starts approaching a GeV, WIMP scattering starts probing the internal nuclear structure, and ultimately the scattering amplitude does become a sum of amplitudes for scattering on individual nucleons, as in (13). A further increase in m_V above the GeV scale allows the $\mathrm{U}(1)_S$ sector to be integrated out, leading to a WIMP-quark effective Lagrangian that can be used to calculate $f_{p(n)}$.

In what follows, we calculate the second-order elastic scattering amplitude, keeping m_V under a GeV. For light and intermediate-mass $\mathrm{U}(1)_S$ mediators, the WIMP-nucleus cross section can be straightforwardly calculated using a nonrelativistic potential scattering treatment in the presence of a nuclear charge form factor. For m_V^{-1} well below 1 fm, analogous formulas can easily be derived for WIMP-nucleon scattering. The wave function of the WIMP-nucleus system under the influence of a central potential is $\psi(\mathbf{x}) = e^{i\mathbf{k}_i\cdot\mathbf{x}} + f(\theta)e^{i\mathbf{k}_f\cdot\mathbf{x}}/r$, where $f(\theta)$ is the scattering amplitude. The second-order Born amplitude for scattering from a potential $V(\mathbf{x})$ can be written, in general, as

$$f(\theta) = -\frac{\mu_N}{2\pi}\int d^3\mathbf{x} d^3\mathbf{x}' e^{-i\mathbf{k}_f\cdot\mathbf{x}} V(\mathbf{x}) G(\mathbf{x}-\mathbf{x}') V(\mathbf{x}') e^{i\mathbf{k}_i\cdot\mathbf{x}'}, \tag{22}$$

where $G = (E_\mathbf{k} - E'_\mathbf{p})^{-1}$ is the nonrelativistic propagator for the χ_2-nucleus system, which in momentum space becomes

$$\begin{aligned} G(\mathbf{p}) &= [(\sqrt{m_{\chi_1}^2 + \mathbf{k}^2} + \sqrt{m_N^2 + \mathbf{k}^2}) \\ &\quad - (\sqrt{m_{\chi_2}^2 + \mathbf{p}^2} + \sqrt{m_N^2 + \mathbf{p}^2})]^{-1} \\ &\simeq \left[-\Delta m + \frac{\mathbf{k}^2}{2\mu_N} - \frac{\mathbf{p}^2}{2\mu_N}\right]^{-1}, \end{aligned} \tag{23}$$

where $\mathbf{p}$ is the intermediate momentum. For a large enough splitting we can neglect the kinetic energy terms in the propagator, and this occurs for $\Delta m \gtrsim E_{\rm kin} \sim 100$ keV for a typical nucleus with mass of order $m_N = 50$ GeV. Retaining only the constant $(\Delta m)^{-1}$ term in the propagator allows us to use the completeness relation and remove the integration over intermediate states. Thus the amplitude greatly simplifies and is expressed via the square of the central potential $V(\mathbf{x}) = V(r)$:

$$f(q) = \frac{2\mu_N}{q\Delta m}\int_0^\infty dr r V^2(r)\sin qr, \tag{24}$$

where $q = \sqrt{2k^2(1-\cos\theta)}$ and θ are the momentum transfer and the scattering angle in the c.o.m. frame, as before. The expression (24) is equivalent to the first-order Born formula for the effective potential $V_{\rm eff} = V^2(r)/\Delta m$.

The potential can be written, in general, as

$$\begin{aligned} V(r) &= \kappa\sqrt{\alpha\alpha'}\int \frac{d^3\mathbf{p}}{(2\pi)^3}\frac{F(\mathbf{p})}{\mathbf{p}^2 + m_V^2} e^{i\mathbf{p}\cdot\mathbf{x}} \\ &= \frac{\kappa Z\sqrt{\alpha\alpha'}}{2\pi^2 r}\int_0^\infty dp \frac{(F(p)/Z) p \sin pr}{p^2 + m_V^2} \end{aligned} \tag{25}$$

where $F(p)/Z$ is the charge-normalized nuclear form factor, with $F(0)/Z = 1$. To compute this potential, we must specify $F(p)$. Using a model with a uniform charge distribution within a sphere of finite radius r_N gives

$$F(p) = \frac{3j_1(pr_N)}{pr_N}, \tag{26}$$

where j_1 is a spherical Bessel function. This simple model of the nucleus allows us to analytically compute the potential in Eq. (25)[2]:

[2]Note that for the first-order endothermic and exothermic inelastic processes, we utilize the more common Helm form factor [28], which contains an additional exponential factor $\exp(-s^2p^2)$ relative to Eq. (26). The parameter s is the nuclear skin depth with a typical value of 1 fm, corresponding to energy scales of order 200 MeV. For the case of coherent elastic WIMP-nucleus scattering considered here, the interaction range is approximately $m_V^{-1} \sim r_N > s$, so that the simple solid sphere model of the nucleus is a reasonable approximation.

$$V(r) = \frac{3Z\kappa\sqrt{\alpha\alpha'}}{m_V^2 r_N^3} f(m_V r_N, m_V r)$$

$$\text{where } f(x,\hat{r})|_{r<r_N} = 1 - (1+x)e^{-x}\frac{\sinh\hat{r}}{\hat{r}},$$

$$f(x,\hat{r})|_{r>r_N} = (x\cosh x - \sinh x)\frac{e^{-\hat{r}}}{\hat{r}}. \quad (27)$$

This is essentially the "Yukawa potential" of a hard sphere, and it is clear from (27) that, in the limit $m_V \to 0$ at finite r_N, the potential reduces to the usual electrostatic potential created by a spherical charge distribution, while for $r_N \to 0$, $V(r) \sim \exp(-m_V r)/r$. Taking both limits simultaneously reduces (27) to an ordinary Coulomb potential, in which case the scattering amplitude (24) simplifies dramatically to give

$$f(q) = \frac{\pi Z^2 \alpha\alpha'\kappa^2\mu_N}{q\Delta m} \Rightarrow \sigma_{\rm tot} = \frac{2\pi^3\mu_N^2 (Z\alpha)^4\kappa^4(\frac{\alpha'}{\alpha})^2}{k^2(\Delta m)^2}\ln\left(\frac{1}{r_N m_V}\right), \quad (28)$$

where the final integral over momentum transfer is approximated by a logarithm, given the assumption $m_V \ll q \ll r_N^{-1}$. Because the range of validity for (28) is almost nonexistent, and the nuclear form factor has been neglected, we will not make use of it for deriving constraints on the model. However, it has the merit of illustrating the Z^4 scaling of the cross section that translates into the following Z-independent prediction for the effective cross section per nucleon,

$$\sigma_{\rm nucl} \approx 10^{-42}\ {\rm cm}^2 \times \left(\frac{\alpha'}{\alpha}\right)^2\left(\frac{\kappa}{10^{-4}}\right)^4\left(\frac{1\ {\rm MeV}}{\Delta m}\right)^2 \quad \text{for small } m_V \quad (29)$$

where we have inserted typical values for various parameters such as $Z/A \simeq 0.5$, the characteristic recoil energy, etc. We can see that direct-detection experiments can potentially have strong sensitivity to κ for light mediators, due to the resulting long-range interaction. Equation (29) represents the sensitivity in the "best case," which erodes rather rapidly with increasing m_V.

To derive actual limits on the parameters of the model, we write the differential cross section as

$$\frac{d\sigma}{dE_R} = \frac{2\pi m_N}{\mu_N^2 v^2}|f(E_R)|^2. \quad (30)$$

Inserting Eq. (30) with full form factor dependence into (12), we calculate the constraints on the parameter space using the results from the CDMS experiment. These constraints are shown in Fig. 3 for two cases of $m_V = 10$ and 100 MeV. It is easy to see the reduced sensitivity to κ as M_V^{-1} becomes smaller than the nuclear radius.

For larger mediator masses, one can easily observe this change in the Z scaling. Once $m_V r_N \gg 1$, then inside the nucleus the potential (27) becomes $V(r) \sim 3Z\kappa\sqrt{\alpha\alpha'}m_V^{-2}r_N^{-3}$, and the amplitude at small q has the following scaling with Z: $f(q=0) \sim Z^2 r_N^{-3} \sim Z$. The switch from Z^2 to Z in the amplitude signifies the loss of coherent nuclear response, and at that point the process is best described by scattering on individual nucleons. Taking $m_V \sim$ GeV, we calculate the WIMP-nucleon elastic scattering amplitude, following the same procedure as developed for the nucleus. We use a proton form factor $F_p(p) = (1 + a^2p^2)^{-2}$ with $a \simeq 0.84$ fm related to the charge radius of the proton. The amplitude in Eq. (24), the cross section, and the rate can all be straightforwardly computed, and we obtain the estimate

$$\sigma_{\rm nucl} \approx 10^{-42}\ {\rm cm}^2 \times \left(\frac{\alpha'}{\alpha}\right)^2\left(\frac{\kappa}{10^{-1}}\right)^4\left(\frac{10\ {\rm MeV}}{\Delta m}\right)^2 \quad \text{for } m_V \sim 1\ {\rm GeV}. \quad (31)$$

It is apparent that the loss of coherence in the elastic scattering process means that the sensitivity to κ in this case is no better than particle physics probes with SM particles. The dependence on m_V of the sensitivity is exhibited in Fig. 4, showing the transition around $m_V \sim 1/r_N$.

Finally, for a fermionic WIMP, a loop-induced effective coupling of the nucleon spin to an off-shell photon becomes possible. This can be interpreted as an effective anapole moment of the WIMP [9,27]. Although such a coupling will be proportional to the first power of κ, it has no implications for the direct detection for two reasons. First, the scattering amplitude is spin dependent and thus not coherently enhanced. Moreover, the amplitude is proportional to the small relative velocity of the WIMP-nucleus system, leading to a further suppression. Note that this velocity dependence adds an additional suppres-

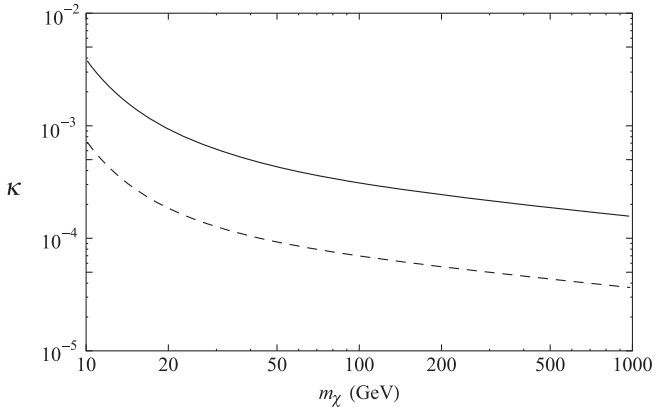

FIG. 3. *Elastic scattering constraints*: 90% CDMS confidence limits on κ as a function of WIMP mass m_χ for a mass splitting $\Delta m = 10$ MeV, with α' chosen to yield the correct thermal relic abundance. We show constraints from the elastic scattering $\chi_1 N \to \chi_1 N$ for $m_V = 100$ MeV (solid line) and 10 MeV (dashed line).

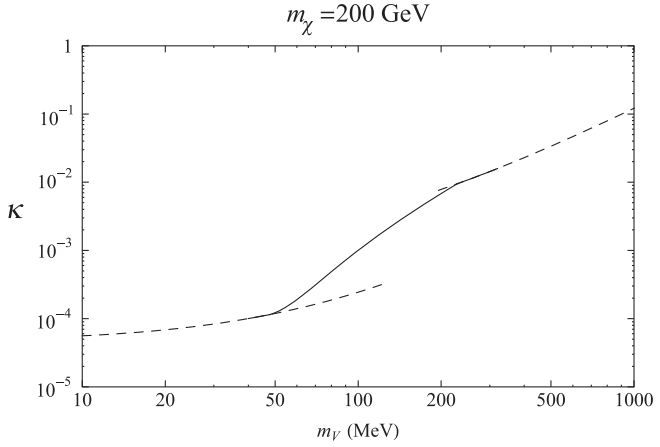

FIG. 4. *Elastic scattering constraints*: 90% CDMS confidence limits on κ as a function of vector mass m_V for a 200 GeV WIMP with mass splitting $\Delta m = 10$ MeV, where α' is chosen to yield the correct thermal relic abundance. We show constraints from elastic scattering $\chi_1 N \to \chi_1 N$, where the solid line interpolates between the nuclear and nucleon scattering descriptions.

sion compared to the usual spin-dependent couplings encountered in, e.g., supersymmetric WIMP models.

C. Nonperturbative scattering and WIMP-nucleus binding

In the previous section we assumed that the Born approximation is applicable, which may not always be the case if m_V is small and κ is taken to be large. For small Δm and m_V, the criterion for the applicability of the Born approximation is $Z\sqrt{\alpha\alpha'}\kappa/v \ll 1$. Since $Z\alpha$ could maximally be of order one, we see that for galactic WIMPs with $v \sim 10^{-3}$, this is satisfied for any $\kappa < 10^{-3}$. We note that for small m_V the exclusion contours in Fig. 4 are well below this threshold and thus well within the range of validity of the Born approximation. Nonetheless, it is intriguing to consider whether $V - \gamma$ exchange may lead to the formation of a bound state, as suggested in Ref. [17].

For a suitable choice of parameters, it is indeed possible for the Yukawa-like potential between the WIMP and a large nucleus to support a bound state, or have a quasistationary state just above the continuum threshold. The potential arises from $V - \gamma$ exchange, and in the limit $m_V \ll 1/r_N$ takes the form

$$V(r) = -\kappa Z \frac{\sqrt{\alpha\alpha'}}{r} e^{-m_v r}. \tag{32}$$

In the limit of small Δm, the existence of at least one bound state then requires (for $m_N \ll m_\chi$)

$$\kappa \geq 1.4 \frac{m_V}{Z\sqrt{\alpha\alpha' m_N}}, \tag{33}$$

which can reach interesting parts of the parameter space for light mediators and large nuclei; e.g. for the Germanium target at CDMS, a bound state is possible for $\kappa > 10^{-3}$ for $\alpha' = \alpha$ and $m_V \sim 10$ MeV.

The binding energy is rather small, varying from the Coulomb limit $E_b = (\kappa Z)^2 \alpha\alpha' m_\chi/4$, which could be up to $\mathcal{O}(100$ keV) for large $(\kappa Z) \sim 0.1$, down to $E_b = 0$ at the threshold (33). For the situation where the Dirac WIMP is split by Δm into Majorana components, this potential will inhabit the off-diagonal terms in a 2×2 potential matrix; thus for a mass splitting $\Delta m \sim 100$ keV, it is apparent that only large nuclei will allow for recombination.

If kinematically accessible, since the recombination rate for $\chi N \to (\chi N) + \gamma$ is large, we can conservatively regard the threshold (33) as a constraint on the parameters of the model for small splittings. This would most likely arise through the formation of too large a relative abundance of anomalous heavy elements, but it would also produce a distinctive signature in underground detectors. Nonetheless, it is worth remarking that the strong dependence on the atomic mass means that the most stringent constraints on the abundance of anomalous heavy isotopes of light elements will, as was the case for the pseudodegenerate scenarios of [9], be less important. Finally, if Δm is increased or κ decreased, the bound state can be pushed into the continuum, and may lead to a resonant feature in WIMP-nucleus scattering. However, this again will likely be well inside the region excluded by Fig. 4.

IV. INELASTIC SCATTERING AND DEEXCITATION CONSTRAINTS

The presence of an ambient population of excited χ_2 states in the galactic halo would allow for exothermic inelastic $\chi_2 \to \chi_1$ scattering (16) in detectors. Given the enhanced rate, this can potentially act as a novel and powerful probe of multicomponent WIMP scenarios in which a sizable excited state population is present. However, these constraints will depend sensitively on Δm, as a large O(MeV) splitting would result in too great an energy deposition within the detector, $E \sim \Delta m m_\chi/(m_\chi + m_N)$, falling outside the fiducial recoil-energy window for most searches. Moreover, larger values for Δm may considerably shorten the lifetime of χ_2, as in Eq. (11), reducing the local number density. Therefore, in this section, we will concentrate on a relatively small splitting: $\Delta m \lesssim 500$ keV.

While the overall abundance of χ_2 and χ_1 is fixed as soon as chemical equilibrium is lost in the early Universe, namely, after the temperature drops below $0.05 m_\chi$, the relative abundance of χ_2 and χ_1 can vary. The possibility of exothermic scattering in detectors depends on the ambient number density n_2 of the excited χ_2 population in the galactic halo, which arises from three main sources, cosmological, galactic, and local:

$$n_2 = n_2^{(c)} + n_2^{(g)} + n_2^{(l)}. \tag{34}$$

The cosmological abundance is regulated by the χ_2 lifetime and freeze-out of the $\chi_2 \to \chi_1$ rate in the early Universe, while the galactic source is related to the inverse possibility for $\chi_1 \to \chi_2$ up-scattering in the galaxy, with energy supplied by the WIMP kinetic energy [12,13]. The local source could originate from the scattering of χ_1 on heavy nuclei in the Earth's interior, e.g. $\chi_1 \mathrm{Pb} \to \chi_2 \mathrm{Pb}$. It is relatively insensitive to the χ_2 lifetime, provided that it is longer than the WIMP time-of-flight through the Earth. However, it is easy to see that the weak-scale WIMP-nucleus cross section cannot lead to more than an $\mathcal{O}(10^{-8})$ excitation probability. Thus the cosmological and galactic sources are, in general, far more important, but depend crucially on the size of the $\chi_2 \leftrightarrow \chi_1$ interconversion rate, which may, in general, have a number of contributions:

(i) *Double (de)excitation*:

$$\chi_2 + \chi_2 \leftrightarrow \chi_1 + \chi_1. \tag{35}$$

This is the most generic possibility, and the only one which will be important in the minimal $\mathrm{U}(1)_S$ model for a generic range of parameters. Note that it depends sensitively on the number density of χ_2 states, and thus will tend to freeze out—at the latest—when the temperature drops below Δm. This will allow us to determine the minimal fractional abundance of χ_2 relative to χ_1 in the $\mathrm{U}(1)_S$ scenario.

(ii) *SM thermalization*:

$$\chi_2 + \mathrm{SM} \leftrightarrow \chi_1 + \mathrm{SM}. \tag{36}$$

For this process to compete with (35) requires a significant interaction rate between the WIMP and SM sectors,[3] and amounts to maintaining thermal contact of the WIMPs with the SM bath down to low temperatures, $T_\gamma \lesssim \Delta m$. Since we take $\Delta m < m_e$, this may only happen if the rate for scattering off electrons is larger than the weak rate, $4\pi\kappa\sqrt{\alpha\alpha'}/m_V^2 > G_F$, because for $T_\gamma < m_e$ the number density of electrons decreases exponentially. However, for the fiducial range of values that we consider, $\kappa\sqrt{\alpha\alpha'}/m_V^2$ is broadly comparable to G_F, and the scattering off electrons will have the effect of maintaining thermal equilibrium between the WIMP and SM sectors [1,29], possibly all the way to $T_\gamma \sim$ MeV, but will not in itself lead to depletion of the χ_2 abundance.

(iii) *Dark thermalization*:

$$\chi_2 + X \leftrightarrow \chi_1 + X. \tag{37}$$

This process assumes a more complex dark sector with additional light degrees of freedom. Given that the dark sector decouples relatively early from the SM bath, and subsequently cools more quickly, if this additional deexcitation rate can remain in equilibrium well below this decoupling scale, then it could allow for a significant depletion of the excited state fraction. Such a process then becomes intertwined with the possibility for $\chi_2 \to \chi_1 + X$ decay. This is not relevant for the minimal $\mathrm{U}(1)_S$ scenario for $m_V > \mathcal{O}(\mathrm{MeV})$ as considered here, and so we will not consider it in detail. However, it is worth noting that the presence of additional light states may have further implications for big bang nucleosynthesis and cosmology, particularly if this interaction rate is large, as would be needed for this process to be important.

In the minimal $\mathrm{U}(1)_S$ scenario, for $\Delta m < m_e$, the double deexcitation process (35) is generally the most relevant, and we will now consider the freeze-out in more detail, in order to estimate the *minimal* fractional abundance of χ_2 relevant for direct detection.

A. Estimate of the fractional abundance of excited WIMPs

In the early Universe, chemical freeze-out occurs at relatively high scales, where the distinction between χ_1 and χ_2 is unimportant. Since we have $m_\chi \gg m_V$, the primary annihilation process involves $\bar{\chi}\chi \to VV$, where the two V bosons are on shell and subsequently decay to the SM via kinetic mixing with the photon. For $m_V \ll m_\chi$ we have [17,19]

$$\langle \sigma v \rangle_{\mathrm{ann}} = \frac{\pi(\alpha')^2}{2m_\chi^2} \to 2.4 \times 10^{-26}\ \mathrm{cm}^3\ \mathrm{s}^{-1}, \tag{38}$$

where the latter relation follows from ensuring that the WIMPs have a relic density that saturates the measured value of Ω_{DM}. For our purposes, chemical freeze-out for χ_1 and χ_2 will occur at a high temperature scale $T_f \sim m_\chi/20$, and so we can simply take (38) as a constraint relating the coupling α' to m_χ as in (17).

If the χ_2 lifetime exceeds τ_{U}, as discussed in Sec. II B, and the rate for interconversion (35) is slow—dropping below the Hubble expansion rate before the average WIMP energy falls below Δm—the inevitable prediction is $n_2/n_1 \simeq 1$. For example, if in the $\mathrm{U}(1)_S$ model the range of the mediator m_V^{-1} is very short (e.g. weak scale), χ_2 and χ_1 will be equally abundant. However, given a relatively long interaction range for small m_V, the reaction (35) could remain in equilibrium down to WIMP energies of $E \lesssim \Delta m$, resulting in a significant depletion of χ_2 states.

After chemical decoupling, the WIMPs remain in thermal equilibrium down to lower temperatures, scattering off $\mathrm{U}(1)_S$ vectors in the dark sector via the Thomson-like process $\chi V \to \chi V$ and/or off SM charged particles. Scattering off V's becomes inefficient once the temperature drops below m_V, and most of the vector particles

[3] We are grateful to D. Morrissey for pointing out the relevance of SM thermalization.

decay. The κ-dependent scattering of WIMPs on SM charged particles can be straightforwardly computed e.g. in the temperature interval $1\ \mathrm{MeV} \lesssim T \lesssim 100\ \mathrm{MeV}$, where scattering off electrons is dominant, with a rate scaling as $\kappa^2 \alpha \alpha' m_V^{-4} T^5$. After these processes fall below the Hubble rate, the WIMPs are no longer in kinetic equilibrium with the SM thermal bath and begin to cool more rapidly, maintaining a quasithermal spectrum with temperature T_χ:

$$T_\chi = T_\gamma^2/T_* \quad \text{for } T_\gamma < T_*. \tag{39}$$

For very small κ, scattering on V dominates, and $T_* \sim m_V$. However, if the rate of rescattering on electrons is comparable to the weak rate, i.e. $4\pi\kappa(\alpha\alpha')^{1/2} m_V^{-2} \sim 10^{-5} m_p^{-2}$, decoupling is postponed to a temperature $T_* \sim$ MeV.

Once kinetic decoupling from the SM bath is complete, the interconversion process (35) is capable of driving down the fractional abundance of χ_2, provided that it is faster than the Hubble rate, and the energy of the emerging nonthermal χ_1's is quickly redistributed in the WIMP sector. If such a quasithermal state is maintained below $T_\chi < \Delta m$, the fractional abundance of excited states will be exponentially suppressed,

$$\frac{n_2}{n_1} \sim \exp\left(-\frac{\Delta m}{T_\chi}\right). \tag{40}$$

Because of the finite rate for the $\chi_2\chi_2 \to \chi_1\chi_1$ deexcitation process, the approximate freeze-out condition is

$$H(T_\gamma^f) = [n_2 \langle \sigma_{22\to 11} v \rangle]_{T_\chi^f}, \tag{41}$$

where the SM and dark sector freeze-out temperatures are related via (39). This results in the following estimate for the fractional freeze-out abundance:

$$\frac{n_2}{n_2 + n_1} = \eta_b^{-1} \frac{\Omega_b}{\Omega_{\mathrm{DM}}} \frac{m_\chi}{m_p} \frac{H(T_\gamma^f)}{[n_\gamma \langle \sigma_{22\to 11} v \rangle]_{T_\chi^f}}, \tag{42}$$

where $\eta_b = 6.2 \times 10^{-10}$ is the baryon-to-photon ratio, $\Omega_b/\Omega_{\mathrm{DM}} \simeq 0.2$ is the ratio of baryonic to DM energy densities, m_p is the proton mass, and $n_\gamma(T) = 0.24 T^3$ is the photon number density. The Hubble rate is given by $H(T) \simeq 1.7 g_* M_{\mathrm{Pl}}^{-1} T^2$, while $M_{\mathrm{Pl}} = 1.2 \times 10^{19}$ GeV and $g_* \sim 10$.

The minimal value for n_2/n_1 is achieved when the $\chi_1 \leftrightarrow \chi_2$ interconversion cross section is maximized. Since we are working in the regime $\alpha'/v \gtrsim 1$ with small m_V, perturbation theory is not applicable, and the Schrödinger equation should be solved numerically as in the recent paper [30]. However, for our discussion it suffices to saturate $\sigma_{22\to 11}$ by the s-wave unitarity limit or by the range of the force carrier, m_V^{-2}, whichever is smaller:

$$\sigma_{22\to 11}^{\max} \sim \frac{\pi}{k^2} \ \text{for } k \gtrsim m_V \Rightarrow \langle \sigma_{22\to 11} v \rangle \lesssim \frac{\pi (T_*)^{1/2}}{m_\chi^{3/2} T_\gamma}. \tag{43}$$

Combining (42) and (43), we obtain the minimal fractional χ_2 abundance surviving from the early cosmological epoch:

$$\left[\frac{n_2}{n_1}\right]_{\min}^{(c)} \simeq 10^{-2} \times \left(\frac{m_\chi}{300\ \mathrm{GeV}}\right)^{5/2} \left(\frac{10\ \mathrm{MeV}}{T_*}\right)^{1/2}, \tag{44}$$

where the abundance is also implicitly bounded from above by $[n_2/n_1]_{\max} = 1$. Note that if thermalization via electron scattering is negligible, then in order to maximize the rate one can choose $m_V \sim k \sim (m_\chi \Delta m)^{1/2}$ at WIMP energies comparable to Δm, so that the optimal value for T_* in (44) is $T_* \sim m_V \sim (m_\chi \Delta m)^{1/2}$. More generally, the estimate (44) indicates that for a TeV-scale WIMP, the fractional abundance of χ_2 does not drop below 5%, independent of the strength of the deexcitation reaction. We should note that more accurate computations of the excited fraction n_2/n_1 are certainly possible, but the estimate (44) will suffice to obtain an estimate of the direct-detection rate.

In addition to this cosmological source, there is also the probability of endothermic up-scattering in the galaxy, adding to the fractional abundance in (44). Since the rates for the forward and backward reactions are related, we can estimate the excited state abundance from galactic scattering to be

$$\left[\frac{n_2}{n_1}\right]^{(g)} \sim \tau_{\mathrm{int}} \frac{\rho_\chi}{m_\chi} \langle \sigma_{11\to 22} v \rangle_{\mathrm{gal}}, \tag{45}$$

where $\rho_\chi \sim 0.3\ \mathrm{GeV\,cm^{-3}}$ is the local dark matter energy density, $v \sim 10^{-3}$, and τ_{int} is the "integration time," equal at least to the age of the Milky Way, 13 bn yr, or the lifetime of χ_2, whichever is smaller. Assuming a lifetime in excess of 10 bn yr, we obtain a simple estimate,

$$\left[\frac{n_2}{n_1}\right]^{(g)} \sim 10^{-4} \times \left(\frac{300\ \mathrm{GeV}}{m_\chi}\right)^3 \times \exp(-\Delta m/T_{\mathrm{eff}}), \tag{46}$$

where T_{eff} is the effective WIMP temperature in the galactic halo. This suggests that the galactic source is somewhat subdominant, but can, in principle, compete with the cosmological source if the WIMP mass is $\sim$100 GeV.

B. Constraints from direct detection

Making use of Eq. (44) to determine the minimal fractional χ_2 abundance, and the results of the previous section, we are now able to calculate the exothermic scattering rate as a function of the parameters in the $\mathrm{U}(1)_S$ model. To be specific, we choose $m_V = 1$ GeV, and vary Δm and κ, recalling that elastic scattering processes do not place significant constraints on κ for such large values of m_V. The CDMS constraints for a 100 GeV and a 1 TeV WIMP are presented in Fig. 5. It is clear that for small values of Δm, exothermic scattering places very stringent bounds on values of the mixing parameter κ. Also, as one expects, the constraints start to deteriorate for splittings $\Delta m \gtrsim 100$ keV, as these exothermic events will generally have

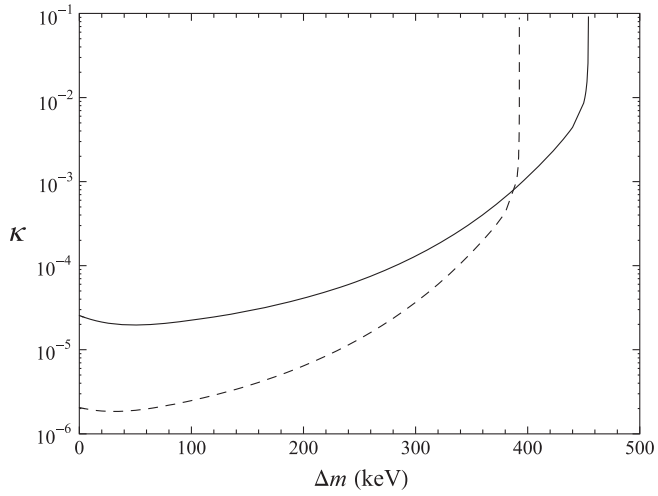

FIG. 5. *Exothermic inelastic scattering constraints*: 90% CDMS confidence limits on κ as a function of mass splitting Δm for a vector mass $m_V = 1$ GeV, where α' is chosen to yield the correct thermal relic abundance. We show constraints from exothermic inelastic scattering $\chi_2 N \to \chi_1 N$ for WIMPs with masses $m_\chi = 100$ GeV (solid line) and 1 TeV (dashed line). The constraints rapidly deteriorate for large Δm, as in this case most scatterings will have a large nuclear recoil well above the maximum detector sensitivity of $E_R = 100$ keV.

a large recoil energy, $E_R \gtrsim 100$ keV, outside the window of most direct-detection experiments.

Another intriguing feature of Fig. 5 is the increased sensitivity of direct-detection experiments for heavy WIMPs. Naively the differential rate (12) falls as m_χ^{-1}, but in our case there is a hidden m_χ dependence in the sensitivity coming from both the necessity of larger $U(1)_S$ couplings α' to obtain the appropriate cosmological DM abundance, and the WIMP mass dependence contained in our estimate of the minimal fractional abundance in Eq. (44).

As an application, we consider the feasibility of the minimal $U(1)_S$ scenario to account for the DAMA anomaly [5] in light of constraints coming from both endothermic and exothermic scattering. The inelastic dark matter scenario [8] was proposed to reconcile the null results of direct-detection experiments with the annual modulation signal observed by the DAMA experiment, and multicomponent WIMPs in the $U(1)_S$ scenario provide one realization of this idea. To calculate the DAMA preferred regions, we utilize a χ^2 goodness-of-fit test [31], including data from the first 12 bins between 2 keVee and 8 keVee [5]. We show the combined constraints from endothermic and exothermic processes from CDMS in Fig. 6 for two values of the WIMP mass, 100 GeV and 1 TeV, along with the DAMA preferred regions at the 90% and 99% CL. Since the relevant range of κ, m_V, and α' considered in this paper implies a rate for WIMP scattering off the thermal electron-positron bath which is close to the weak scale even for $\kappa \sim 10^{-4}$ as discussed above, the relevant choice of decoupling temperature is $T_* \sim \mathcal{O}(1\text{–}10$ MeV).

One observes that in various parameter ranges exothermic scattering can, in principle, provide more stringent bounds on the parameters of the model than endothermic scattering or second-order elastic scattering for small Δm, as shown in Figs. 2 and 4. One should keep in mind, however, the model dependence of such a conclusion, as discussed earlier in this section. Indeed, any extension of the minimal $U(1)_S$ model with additional light particles that could facilitate the decay/deexcitation of χ_2 may result

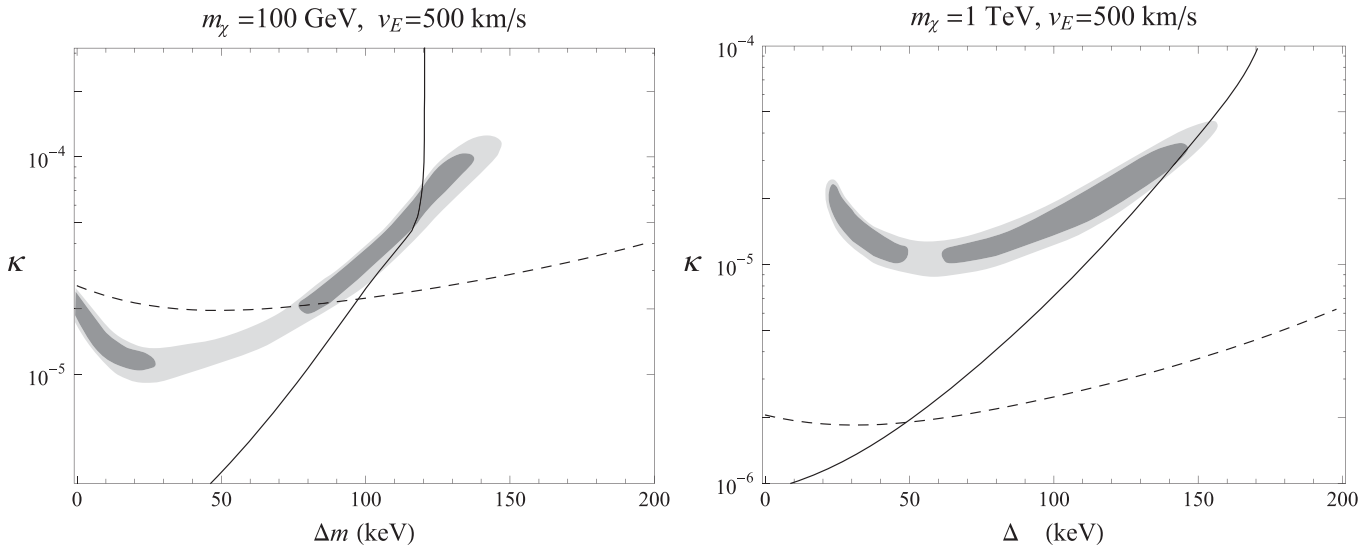

FIG. 6. Constraints from endothermic (solid line) and exothermic (dashed line) scattering in the CDMS detector at 90% confidence level for a 100 GeV WIMP and 1 GeV $U(1)_S$ vector, with α' chosen to yield the correct thermal relic abundance. We show constraints for a 100 GeV WIMP (left panel) and 1 TeV WIMP (right panel). The complementarity of both constraints is clearly seen. The light (dark) shaded region corresponds to the DAMA 99% (90%) CL preferred region, which is excluded by both constraints.

in the absence of any significant χ_2 population. The elastic and endothermic scattering constraints are, in contrast, completely independent of details concerning the χ_2 lifetime and/or deexcitation rate.

Finally, it is also worth noting that another possible signature is via ionization due to exothermic WIMP scattering on atomic electrons in the detector, which will be essentially monochromatic, as it is practically independent of the WIMP velocity, with a cross section that scales linearly with Z.

V. DISCUSSION AND CONCLUSIONS

In this paper, we have carried out a systematic study of the nuclear scattering of multicomponent WIMPs forming part of a SM-singlet dark sector, focusing on its minimal implementation via a $U(1)_S$ mediator. The mass splitting of the components by Δm leads to rich structure for scattering phenomenology. While forbidding first-order elastic scattering, the second-order Born cross section still implies significant sensitivity for direct-detection experiments such as CDMS and XENON for a wide range of mass splittings. Moreover, for small splittings of $\mathcal{O}(100)$ keV, first-order inelastic scattering provides far more stringent constraints, particularly since in this regime the excited states may have a lifetime exceeding the age of the Universe and thus a residual population in the halo can allow exothermic down-scattering in the detector. This exothermic process leads to additional constraints on specific ''inelastic DM'' scenarios, via scattering on relatively light target nuclei that would otherwise have no sensitivity for larger Δm.

In rather general terms, it seems there is a clear tension between the constraints discussed here, associated with enhanced scattering rates, and the presence of relic populations of excited states that may be relevant for other aspects of dark matter phenomenology. We will conclude by discussing some specific implications in this vein, but going beyond the specific $U(1)_S$ scenario:

(i) *Inelastic DM and DAMA*.—It was noticed a few years ago [8] that a multicomponent WIMP scenario with a small 100 keV splitting could resolve the tension between the annual modulation signal observed by DAMA [5], with the fact that the required cross sections were seemingly ruled out by the null results of other experiments with lighter materials. Despite further null results from XENON [4], recent analyses suggest there is seemingly a small parameter range where all experiments can be reconciled [32]. Our simple observation here is that with such a small splitting, the lifetime of a SM-singlet WIMP excited state will generically be long, and thus the possibility of exothermic scattering of a residual excited state population will be an important constraint, as in Fig. 6 for the $U(1)_S$ scenario. Ameliorating this constraint would require consideration of models with further decay channels, e.g. via adding a SM charge to the WIMP (as for a sneutrino), or via additional light states.

(ii) *Multicomponent states and the INTEGRAL 511 keV line*.—Attempts to attribute the well-measured 511 keV line from the Galactic center to WIMP interactions generically involve metastable states with decays to positrons. This may be via delayed decays of a relic [13], or via excitation and decay as in the ''exciting dark matter'' proposal [12]. In the latter case, utilizing the WIMP kinetic energy in the halo requires a scattering cross section well in excess of s-wave unitarity [13,30], unless optimistic assumptions are made concerning the galactic halo profile. Recently, it was suggested that introducing a three-component WIMP sector $\{\chi_1, \chi_2, \chi_3\}$ [30] would resolve this issue by utilizing a large relic population of the middle state χ_2, with a small 100 keV splitting from the highest state χ_3. Scattering could then excite this transition with a reasonable cross section, with the subsequent rapid decay to the ground state χ_1 liberating $\mathcal{O}(1$ MeV) and sourcing the 511 keV line. In this case it is clear that the lifetime and fractional abundance for χ_2 must be large, suggesting an analogous problem with the large cross section for exothermic scattering in direct detection. However, it is worth noting that in this scenario the energy release would be of $\mathcal{O}(1$ MeV), which may in fact be too large to be easily identifiable given the current fiducial recoil-energy range.

(iii) *Direct detection of exothermic scattering*.—These issues raise the important question of the detectability of exothermic scattering in direct-detection experiments. It is clear that for larger splittings, $\Delta m > 300$ keV, the peak of the recoil-energy spectrum tends to fall outside the conventional fiducial range. It would clearly be of interest to know whether existing experiments are able to expand the conventional 10–100 keV fiducial region, in order to provide more stringent constraints on these scenarios.

ACKNOWLEDGMENTS

We would like to thank J. Cline, D. Morrissey, and I. Yavin for useful correspondence and helpful discussions. The work of A. R. and M. P. is supported in part by NSERC, Canada, and research at the Perimeter Institute is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MEDT.

Note added.—While this work was being completed, the preprint [33] appeared on the arXiv, which also considers exothermic scattering through deexcitation, and thus has some overlap with the discussion in Sec. IV.

[1] See e.g. G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rep. **267**, 195 (1996); G. Bertone, D. Hooper, and J. Silk, Phys. Rep. **405**, 279 (2005).

[2] B. W. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977); M. I. Vysotsky, A. D. Dolgov, and Y. B. Zeldovich, Pis'ma Zh. Eksp. Teor. Fiz. **26**, 200 (1977) [JETP Lett. **26**, 188 (1977)].

[3] D. S. Akerib *et al.* (CDMS Collaboration), Phys. Rev. Lett. **93**, 211301 (2004); **96**, 011302 (2006); Z. Ahmed *et al.* (CDMS Collaboration), Phys. Rev. Lett. **102**, 011301 (2009).

[4] J. Angle *et al.* (XENON Collaboration), Phys. Rev. Lett. **100**, 021303 (2008).

[5] R. Bernabei *et al.* (DAMA Collaboration), Phys. Lett. B **480**, 23 (2000); Eur. Phys. J. C **56**, 333 (2008).

[6] T. Han and R. Hempfling, Phys. Lett. B **415**, 161 (1997).

[7] L. J. Hall, T. Moroi, and H. Murayama, Phys. Lett. B **424**, 305 (1998).

[8] D. Tucker-Smith and N. Weiner, Phys. Rev. D **64**, 043502 (2001).

[9] M. Pospelov and A. Ritz, Phys. Rev. D **78**, 055003 (2008).

[10] K. Griest and D. Seckel, Phys. Rev. D **43**, 3191 (1991).

[11] M. Pospelov, Phys. Rev. Lett. **98**, 231301 (2007); C. Bird, K. Koopmans, and M. Pospelov, Phys. Rev. D **78**, 083010 (2008); T. Jittoh, K. Kohri, M. Koike, J. Sato, T. Shimomura, and M. Yamanaka, Phys. Rev. D **76**, 125023 (2007).

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

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

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

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

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

[17] M. Pospelov and A. Ritz, Phys. Lett. B **671**, 391 (2009).

[18] I. Cholis, L. Goodenough, and N. Weiner, arXiv:0802.2922 [Phys. Rev. D (to be published)].

[19] M. Pospelov, A. Ritz, and M. B. Voloshin, Phys. Lett. B **662**, 53 (2008).

[20] B. Holdom, Phys. Lett. **166B**, 196 (1986).

[21] N. Arkani-Hamed and N. Weiner, J. High Energy Phys. 12 (2008) 104; M. Pospelov, arXiv:0811.1030; M. Baumgart, C. Cheung, J. T. Ruderman, L. T. Wang, and I. Yavin, J. High Energy Phys. 04 (2009) 014; Y. Bai and Z. Han, arXiv:0902.0006; C. Cheung, J. T. Ruderman, L. T. Wang, and I. Yavin, arXiv:0902.3246; A. Katz and R. Sundrum, arXiv:0902.3271; B. Batell, M. Pospelov, and A. Ritz, arXiv:0903.0363 [Phys. Rev. D (to be published)].

[22] K. R. Dienes, C. F. Kolda, and J. March-Russell, Nucl. Phys. **B492**, 104 (1997); D. Hooper and K. M. Zurek, Phys. Rev. D **77**, 087302 (2008).

[23] D. P. Finkbeiner, N. Padmanabhan, and N. Weiner, Phys. Rev. D **78**, 063530 (2008).

[24] M. Pospelov, A. Ritz, and M. B. Voloshin, Phys. Rev. D **78**, 115012 (2008).

[25] M. C. Smith *et al.*, Mon. Not. R. Astron. Soc. **379**, 755 (2007).

[26] J. D. Lewin and P. F. Smith, Astropart. Phys. **6**, 87 (1996).

[27] M. Pospelov and T. ter Veldhuis, Phys. Lett. B **480**, 181 (2000).

[28] R. H. Helm, Phys. Rev. **104**, 1466 (1956).

[29] For some recent analyses, see e.g. X. l. Chen, M. Kamionkowski, and X. m. Zhang, Phys. Rev. D **64**, 021302 (2001); S. Hofmann, D. J. Schwarz, and H. Stoecker, Phys. Rev. D **64**, 083507 (2001); S. Profumo, K. Sigurdson, and M. Kamionkowski, Phys. Rev. Lett. **97**, 031301 (2006).

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

[31] C. Amsler *et al.* (Particle Data Group), Phys. Lett. B **667**, 1 (2008).

[32] F. Petriello and K. M. Zurek, J. High Energy Phys. 09 (2008) 047; S. Chang, G. D. Kribs, D. Tucker-Smith, and N. Weiner, Phys. Rev. D **79**, 043513 (2009); M. Fairbairn and T. Schwetz, J. Cosmol. Astropart. Phys. 01 (2009) 037; C. Savage, G. Gelmini, P. Gondolo, and K. Freese, J. Cosmol. Astropart. Phys. 04 (2009) 010; Y. Cui, D. E. Morrissey, D. Poland, and L. Randall, arXiv:0901.0557.

[33] D. P. Finkbeiner, T. Slatyer, N. Weiner, and I. Yavin, arXiv:0903.1037.