Enabling forbidden dark matter

PHYSICAL REVIEW D **96,** 083521 (2017)

Enabling forbidden dark matter

James M. Cline,[1,2,*] Hongwan Liu,[3,†] Tracy R. Slatyer,[3,‡] and Wei Xue[3,§]

[1]*CERN, Theoretical Physics Department, CH-1211 Geneva 23, Switzerland*

[2]*Department of Physics, McGill University, 3600 Rue University, Montreal, Quebec, Canada H3A 2T8*

[3]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

(Received 17 March 2017; published 20 October 2017)

The thermal relic density of dark matter is conventionally set by two-body annihilations. We point out that in many simple models, $3 \to 2$ annihilations can play an important role in determining the relic density over a broad range of model parameters. This occurs when the two-body annihilation is kinematically forbidden, but the $3 \to 2$ process is allowed; we call this scenario *not-forbidden dark matter*. We illustrate this mechanism for a vector-portal dark matter model, showing that for a dark matter mass of $m_\chi \sim$ MeV-10 GeV, $3 \to 2$ processes not only lead to the observed relic density, but also imply a self-interaction cross section that can solve the cusp/core problem. This can be accomplished while remaining consistent with stringent CMB constraints on light dark matter, and can potentially be discovered at future direct detection experiments.

DOI: 10.1103/PhysRevD.96.083521

I. INTRODUCTION

The particle physics nature of dark matter (DM) is still a mystery despite undeniable evidence of its gravitational interactions. The observed relic abundance of DM may provide a clue to its nongravitational interactions, as in the classic weakly interacting massive particle scenario, where the freeze-out of $2 \to 2$ annihilation of DM particles to the standard model (SM) particles sets the late-time abundance of DM. Many variations on the standard thermal freezeout scenario have recently been considered (e.g., [1–25]); in this article, we point out that even for simple and weakly-coupled dark sectors, $3 \to 2$ annihilations—as illustrated in Fig. 1—can play a critical role.

For weakly-coupled DM, $3 \to 2$ processes are usually considered to be subdominant to their $2 \to 2$ counterparts at the time of freeze-out, but if the latter are kinematically suppressed while $3 \to 2$ is unsuppressed, the situation is more complex. This can occur when the DM couples to a "mediator" particle with a mass somewhat larger than that of the DM itself, as might arise in a hidden sector characterized by a single scale.

Kinematic suppression of $2 \to 2$ annihilation, leading to a novel cosmological history during freeze-out, was previously invoked in the "Forbidden DM" [9] and "Impeded DM" [25] scenarios; the new feature in our study is the presence of a kinematically allowed dark-sector $3 \to 2$ annihilation channel. We refer to this scenario as *not-forbidden dark matter* (NFDM). The $3 \to 2$ channel is also important in the strongly interacting massive particle (SIMP) scenario [3], but work on SIMPs has focused on strongly coupled theories with scalar DM [4,26], whereas NFDM is a more generic mechanism: it is potentially important in any situation where $2 \to 2$ annihilations within the dark sector are kinematically suppressed, and has no obvious dependence on whether the DM is

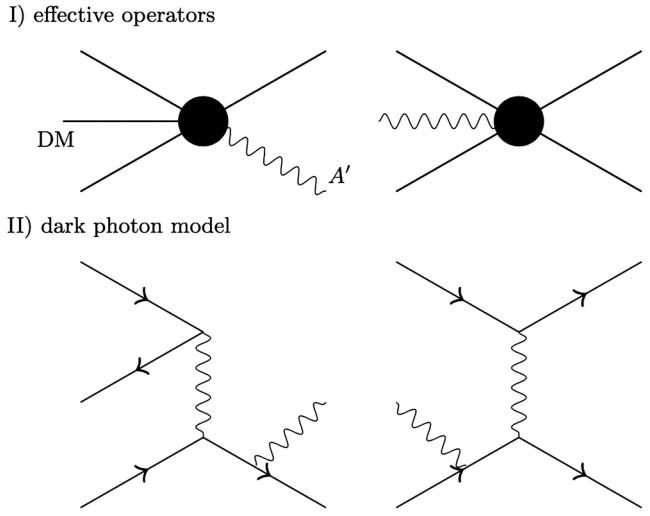

FIG. 1. Schematic description of *not-forbidden dark matter* (NFDM) paradigm. (I) effective operators for the $3 \to 2$ scattering processes; (II) explicit model described in the text: vector-portal dark matter model.

*jcline@physics.mcgill.ca
†hongwan@mit.edu
‡tslatyer@mit.edu
§weixue@mit.edu

fermionic or bosonic or whether the dark sector coupling is strong or weak. Hidden sector or multicomponent DM models may have regions of parameter space where NFDM is an important mechanism to consider.

We illustrate our paradigm with a Dirac fermion DM charged under a hidden $U(1)$ symmetry, with dark gauge boson A'. This mediator can provide a portal to the SM by having a small coupling to the electromagnetic current $J^\mu_{\rm EM}$ through a kinetic mixing term $(\epsilon/2)F'_{\mu\nu}F^{\mu\nu}$. In the mass basis, the Lagrangian becomes

$$\mathcal{L} \supset -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} + \frac{1}{2}m_{A'}^2 A'_\mu A'^\mu + \bar{\chi}(i\not{D} - m_\chi)\chi + eJ^\mu_{\rm EM}(A_\mu + \epsilon A'_\mu). \quad (1)$$

The gauge coupling is $\alpha' = g'^2/4\pi$, and $\not{D} \equiv \not{\partial} - ig'\not{A}'$. It is clear in this basis that there is no tree-level coupling between χ and the SM photon. We can also consistently assume that the dark Higgs boson giving mass to A' is very heavy and can be neglected in the effective description [27]. Depending upon the size of the kinetic mixing parameter ϵ, there are two possible regimes of interest:

(1) ϵ is relatively large, such that the hidden sector and the SM sector have the same temperature before DM freeze-out, while ϵ is still small enough so that $3 \to 2$ and $2 \to 2$ reactions involving only hidden sector particles dominate over annihilation of χ to SM particles;
(2) For sufficiently small $\epsilon \lesssim 10^{-8}$, the hidden sector will have its own temperature and in the limit $\epsilon \to 0$, it becomes secluded: both χ and A' contribute to the ultimate DM density.

In Sec. II, we discuss the freeze-out history of the NFDM model, and by solving the Boltzmann equations, we determine the dark sector parameter values $\{m_\chi, m_{A'}, \epsilon\}$ that are consistent with the observed relic density. In Sec. III we incorporate constraints from a variety of astrophysical and laboratory searches, showing that a significant parameter region is allowed while realizing the NFDM mechanism. Conclusions are given in Sec. IV. In the Appendix, we present a more detailed account of how the order of freeze-out of the various reactions determines the relic abundance; the dependence of our results on the temperature of the dark sector; how the constraints change with $m_{A'}/m_\chi$, and cross sections for the relevant scattering processes.

II. COSMOLOGY

Previous studies of the vector-portal DM model, shown in Eq. (1), have divided the parameter space into two broad regions: $m_\chi < m_{A'}$ or $m_\chi > m_{A'}$. In the latter case, the dominant process at the epoch of thermal freeze-out is $\chi\bar{\chi} \to A'A'$ followed by A' decays to SM particles, whereas when $m_\chi < m_{A'}$, the s-channel annihilation $\chi\bar{\chi} \to f\bar{f}$ to SM particles f via off-shell A' is dominant. This regime is ruled out for $m_\chi \sim$ MeV-GeV by CMB constraints [28–32].

In the present work, however, we are interested in the intermediate region $m_\chi \lesssim m_{A'}$, where it is possible for the $3 \to 2$ scatterings $\chi\chi\bar{\chi} \to \chi A'$ or $\chi\bar{\chi}A' \to \chi\bar{\chi}$ to have an important effect on the dark matter abundance. The system is governed by the coupled Boltzmann equations for the χ and A' densities. For $m_\chi \lesssim m_{A'}$, the relevant terms in these equations are

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -\frac{1}{4}\langle\sigma v^2\rangle_{\substack{\chi\chi\bar{\chi}\\ \to\chi A'}}\left(n_\chi^3 - n_{\chi,0}^2 n_\chi \frac{n_{A'}}{n_{A',0}}\right) + \langle\sigma v\rangle_{\substack{A'A'\\ \to\bar{\chi}\chi}}\left(n_{A'}^2 - n_{A',0}^2\frac{n_\chi^2}{n_{\chi,0}^2}\right), \quad (2)$$

$$\frac{dn_{A'}}{dt} + 3Hn_{A'} = \frac{1}{8}\langle\sigma v^2\rangle_{\substack{\chi\chi\bar{\chi}\\ \to\chi A'}}\left(n_\chi^3 - n_{\chi,0}^2 n_\chi \frac{n_{A'}}{n_{A',0}}\right) - \langle\sigma v\rangle_{\substack{A'A'\\ \to\bar{\chi}\chi}}\left(n_{A'}^2 - n_{A',0}^2\frac{n_\chi^2}{n_{\chi,0}^2}\right) - \Gamma_{A'\to f\bar{f}}(n_{A'} - n_{A',0}), \quad (3)$$

where $n_\chi(n_{\chi,0})$ denotes the (equilibrium) density of $\chi + \bar{\chi}$, and similarly $n_{A'}(n_{A',0})$ for the dark photon. Throughout this paper, we have assumed zero chemical potential for χ and $\bar{\chi}$, and take the densities of χ and $\bar{\chi}$ to be equal. The $1/4$ in the first term of Eq. (2) is the symmetry factor for Dirac DM, taking into account the two identical particles in the initial state and the fact that each annihilation process removes a $\chi\bar{\chi}$ pair. The conjugate process $\chi\bar{\chi}\bar{\chi} \to \bar{\chi}A'$ is also accounted for in this factor. The relative numerical factors between the two equations are consistent with the way each process changes the number density of χ and A'; for example, the factor of $1/4$ and $1/8$ in the first terms of Eqs. (2) and (3) respectively are consistent with the fact that the $3 \to 2$ process has a net effect of removing a $\chi\bar{\chi}$ pair and producing a single A'. A detailed discussion of the derivation of the Boltzmann equation for $3 \to 2$ processes can be found in [33].

Other $3 \to 2$ processes such as $\chi\bar{\chi}A' \to \chi\bar{\chi}$, $3A' \to \chi\bar{\chi}$ etc. are important only in the case of $m_{A'}/m_\chi < 1$ and $\epsilon = 0$ in Sec. II B. The complete Boltzmann equations containing all of these processes are shown in Eq. (D1) and (D2) in Appendix D. All numerical results in this paper across the full range of $m_{A'}/m_\chi$ considered were obtained using the complete equations. Expressions for the cross sections are given in Appendix E.

We will focus on the two regimes where (1) the hidden sector and the SM remain in thermal equilibrium, requiring values of the kinetic mixing $\epsilon \gtrsim 10^{-7}$ (but still small enough to avoid dominance of the $\chi\bar{\chi} \to e^+e^-$ process); (2) the hidden sector is secluded from the SM, $\epsilon \to 0$.

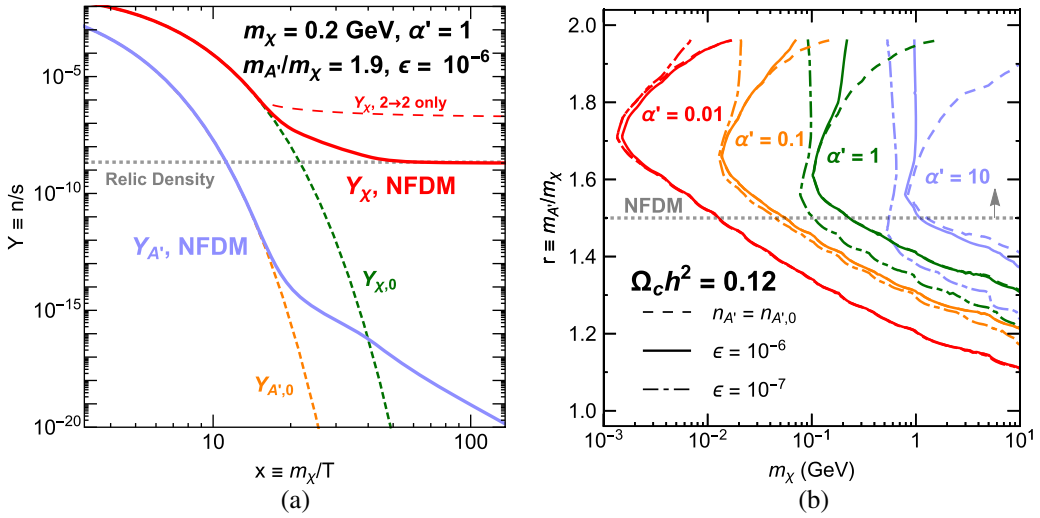

FIG. 2. Relic density in the NFDM scenario, assuming kinetic equilibrium of the dark sector with the SM. (a) The evolution of the energy density of χ (red) and A' (blue) for all processes (bold) and the corresponding energy density of χ excluding the $3 \to 2$ process (red, dotted). The equilibrium distribution of χ (green) and A' (orange) are also shown for reference; (b) contours of the observed present-day relic density in the m_χ-r plane for different values of the coupling constant α'.

A. Kinetic equilibrium with the SM

For sufficiently large ϵ, the scattering process $\chi e^\pm \to \chi e^\pm$ is fast enough to keep the dark and visible sectors in kinetic equilibrium, $T_d = T_{\rm SM}$. By comparing the rate inferred from the $\chi e^\pm \to \chi e^\pm$ cross section to the Hubble rate H at DM freeze-out, we estimate the condition to be

$$\epsilon \gtrsim 10^{-8}\left(\frac{0.1}{\alpha'}\right)^{1/2}\left(\frac{m_\chi}{1\ \text{GeV}}\right)^{1/2}, \tag{4}$$

taking $x_f \sim 20$, the $e^\pm$ to be relativistic, and $m_{A'} \simeq m_\chi$. This leaves a significant range of $\epsilon \lesssim 10^{-8}$–$10^{-4}$, depending upon m_χ, such that A'-mediated annihilations $\chi\bar{\chi} \to e^+e^-$ are out of equilibrium (a requirement of our scenario), as will be shown below.

We take the dark sector masses to be in the ranges $m_\chi \lesssim 10$ GeV and $m_\chi \lesssim m_{A'} < 2m_\chi$. The lower bound on $m_{A'}$ makes $\chi\bar{\chi} \to A'A'$ kinematically inaccessible, while the upper bound forbids the $A' \to \chi\bar{\chi}$ decay channel. If $m_{A'} > 2m_\chi$, the number-changing process $\chi\chi\bar{\chi} \to A'\chi$ effectively becomes number-conserving, $\chi\chi\bar{\chi} \to \chi\chi\bar{\chi}$. In terms of the parameter $r \equiv m_{A'}/m_\chi$, the relevant range is thus $1 \lesssim r \lesssim 2$.

It is enlightening to compare the equilibrium rates (per χ particle) of the $3 \to 2$ and $2 \to 2$ reactions in Eqs. (2)–(3), $\Gamma_{\substack{\chi\chi\bar{\chi} \\ \to \chi A'}} \sim \langle\sigma v^2\rangle_{\substack{\chi\chi\bar{\chi} \\ \to \chi A'}} n_{\chi,0}^2$ and $\Gamma_{\substack{A'A' \\ \to \bar{\chi}\chi}} \sim \langle\sigma v\rangle_{\substack{A'A' \\ \to \bar{\chi}\chi}} n_{A',0}^2/n_{\chi,0}$. From the exponential dependences in the equilibrium number densities, $n_{i,0} \sim \exp(-m_i/T)$, we find that if $m_{A'} \gtrsim \frac{3}{2}m_\chi$, the $3 \to 2$ reaction will be exponentially enhanced with respect to the $2 \to 2$ reaction at low temperatures.

The Boltzmann equations are solved numerically, and the results shown in Fig 2. As an example, Fig. 2(a) illustrates the evolution of the χ and A' abundances as a function of $x \equiv m_\chi/T$ with $m_\chi = 0.2$ GeV, gauge coupling $\alpha' = 1$, kinetic mixing $\epsilon = 10^{-6}$ and the ratio $r = 1.9$. This example has been chosen to emphasize the importance of $3 \to 2$ scatterings, but similar results are obtained for $r \gtrsim 1.5$. Here, in the case with only $2 \to 2$ annihilation, the DM abundance would reach its relic value at $x_f \sim 20$; in our NFDM case, in contrast, the $3 \to 2$ processes and decay of the A' control the freeze-out, and their interplay leads to an extended freeze-out continuing to $x_f \sim 60$. If we neglect the $3 \to 2$ process the resulting abundance is overestimated by several orders of magnitude. It is noteworthy that $Y_{A'}$ departs from the equilibrium abundance at late times, even though the rate for $A' \to e^+e^-$ exceeds the Hubble rate, because the $3 \to 2$ or $2 \to 2$ processes can also strongly affect the evolution of $n_{A'}$.

In Fig. 2(b) we plot the contours in the m_χ-r plane matching the observed relic density [31], for several values of α' and ϵ. We consider values of $\alpha' \leq 4\pi$, since every loop integral introduced in a Feynman diagram typically introduces an additional factor of $\alpha/4\pi$, and so perturbativity is naively maintained for this range of α'. $n_{A'} = n_{A',0}$ corresponds to large ϵ, where the rate for $A' \to e^+e^-$ dominates the rates for either of the two annihilation processes that generate A' s. The region $r \lesssim 1.5$ corresponds to the forbidden DM regime, and Ref. [9] studied this regime with the assumption of $n_{A'} = n_{A',0}$: smaller values of ϵ show increasing deviation from the relic density contours obtained from this assumption, even for $r < 1.5$. For the rest of the paper, we will focus on the NFDM region $1.5 \lesssim r < 2$, where the $3 \to 2$ process leads to a strong

transition in the behavior of the relic density contour, with the exact value of r for the transition depending on the coupling constant α'.

Normally the DM relic density is set by the strongest annihilation channel, which is also the last to freeze out, since only a single Boltzmann equation for DM is considered. This applies when ϵ is large, forcing $n_{A'} \simeq n_{A',0}$ (dashed contours). These contours turn to the right as $r \to 2$ because the $3 \to 2$ cross section diverges, $\langle \sigma v^2 \rangle_{\substack{\chi\chi\bar{\chi} \\ \to \chi A'}} \propto \alpha'^3 m_\chi^{-5} (r-2)^{-7/2}$, and $Y_\chi \sim x_f^2/[m_{pl} m_\chi^2 \langle \sigma v^2 \rangle_{\substack{\chi\chi\bar{\chi} \\ \to \chi A'}}]^{1/2}$. Thus obtaining the correct relic density as $r \to 2$ requires a larger value of m_χ.

In contrast, for moderate values of ϵ, the NFDM mechanism applies, where the two coupled Boltzmann equations must be solved together. In general, we find that typically the *two strongest processes* (either annihilations or decays) keep the coupled system in equilibrium until the rate for one process (per χ particle) becomes comparable to the Hubble rate, and thus any weaker processes are not relevant for determining the relic abundance. In this regime, typically the decay of $A' \to e^+e^-$ and either the $3 \to 2$ or $2 \to 2$ annihilation are the relevant processes. In particular, for $r \gtrsim 1.5$–1.8, the $3 \to 2$ scatterings are faster than $2 \to 2$, and so they dominate the freeze-out, as shown in Fig. 2(a). The combination of $3 \to 2$ scatterings and A' decays can lead to a nonequilibrium density for the A' particles during the freeze-out of the $3 \to 2$ process if ϵ is sufficiently small (e.g., $\epsilon \sim 10^{-6}$–10^{-7}), resulting in a lengthy freeze-out and an ϵ-dependent relic density. This behavior corresponds to the divergence of the dashed and solid contours in Fig. 2(b) at large r.

B. Secluded hidden sector

Next we consider the limit of $\epsilon \to 0$, so that the dark photon is effectively stable, and the hidden sector is secluded. This analysis can be easily applied to multicomponent DM models. Even though secluded hidden sectors are in general difficult to probe due to the lack of any interaction with the SM, they are not entirely impossible to study. Secluded hidden sectors can, for example, be constrained by the number of relativistic degrees of freedom during big bang nucleosynthesis (BBN). Furthermore, in the U(1) theory considered here, the relic abundance is set by the coupling strength α', which in turn determines the self-interaction cross section in the dark sector. This cross section is a prediction of the model, and has observable consequences for structure formation, which can in principle be highly constraining.

Moreover, the secluded case is a useful limit that gives insight into the region of parameter space where ϵ is small but non-zero, so that kinetic equilibrium cannot be maintained with the SM. Despite the small couplings to the SM, this regime can still be effectively probed by observations of the cooling of SN1987a [34]. The secluded limit is also highly instructive as an illustration of the rich behavior that can occur in the $U(1)$ vector portal DM model when the $2 \to 2$ and $3 \to 2$ annihilations are the dominant processes at freeze-out.

To avoid warm or hot dark matter [12], we assume that χ couples additionally to some relativistic degree of freedom ϕ until freeze-out, strongly enough to maintain thermal equilibrium in the dark sector so that the DM temperature redshifts with the Hubble expansion in the conventional manner, $T \sim 1/a$. However, the coupling of ϕ to χ should be sufficiently weak that annihilation of $\chi\bar{\chi} \to \phi\phi$ is negligible during freeze-out, to make the NFDM freeze-out mechanism dominate over conventional $2 \to 2$ annihilation.

For a concrete model of how this can be achieved, we take ϕ to be a light scalar charged under some additional $U(1)$ symmetry, interacting with the dark sector through the dimension-5 operator $(1/\Lambda)\bar{\chi}\chi\phi^*\phi$. The $T \sim 1/a$ dependence is maintained by $\chi\phi \to \chi\phi$ scatterings, which has a rate that scales as $n_\phi \langle \sigma v \rangle_{\chi\phi \to \chi\phi}$, while the $\chi\bar{\chi} \to \phi^*\phi$ rate scales as $n_\chi \langle \sigma v \rangle_{\chi\bar{\chi} \to \phi^*\phi}$. To obtain a parametric estimate for a value of Λ that would maintain both $T \sim 1/a$ and subdominance to the $2 \to 2$ and $3 \to 2$ processes considered in Eq. (2) and (3), we take $\langle \sigma v \rangle_{\chi\phi \to \chi\phi} \sim \langle \sigma v \rangle_{\chi\bar{\chi} \to \phi^*\phi} \sim 1/\Lambda^2$, and look for values of Λ which ensure that the $\chi\bar{\chi} \to \phi^*\phi$ annihilation rate is subdominant up to the point of freeze-out of the two main processes. This condition is most difficult to satisfy in the case where $r = 2$, and the $2 \to 2$ rate becomes highly suppressed. Nevertheless, we find that in this limit, a suitable range of Λ is $m_\chi^{4/3} m_{\rm pl}^{2/3} \lesssim \Lambda^2 \lesssim m_\chi m_{\rm pl}$, which for GeV dark matter corresponds to $10^6 \lesssim \Lambda/\mathrm{GeV} \lesssim 10^9$.

More generally, the dark sector has its own temperature T_d which need not be the same as that of the visible sector, $T_{\rm SM}$; it is determined by details of the thermal cosmological history such as the efficiency of reheating into the dark sector after inflation. The relic abundance in this case depends upon the unknown parameter $\gamma \equiv T_d/T_{\rm SM}$, but in a simple way: $Y_\chi \propto \gamma^{p(r)}$ where $p(r) \sim 1.6$–1.8 depends upon the mass ratio $r = m_{A'}/m_\chi$. Here we illustrate the case of $\gamma = 1$.

The evolution of n_χ and $n_{A'}$ for the secluded dark sector is shown in Fig. 3(a), taking $m_\chi = 35$ MeV, $\alpha' = 1$, $r = 1.95$ as an example to illustrate the important interplay between the $2 \to 2$ and $3 \to 2$ interactions. Keeping only the $3 \to 2$ reaction would predict that A' becomes the dominant DM component, whereas in reality it remains highly subdominant. Again the freeze-out process is prolonged, starting with the decoupling of $2 \to 2$ scatterings at $x \sim 20$, while the $3 \to 2$ reactions decouple at $x \sim 150$. Interestingly, $n_{A'}$ temporarily grows between these two times, allowing the $2 \to 2$ rate to come back above H just before freeze-out completes.

In Fig. 3(b), we plot contours corresponding to the observed thermal relic density in the m_χ-r plane for different values of α'. In the following we give a brief explanation of

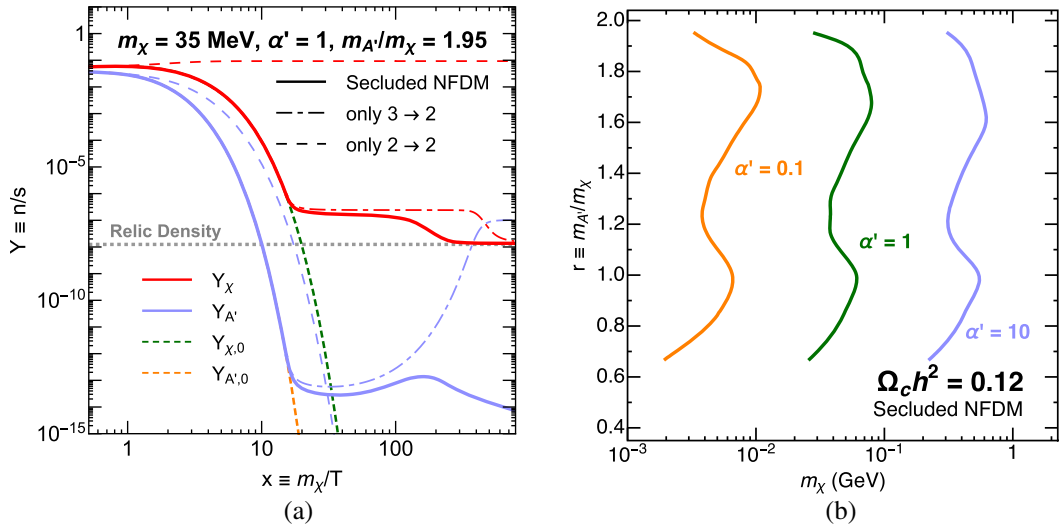

FIG. 3. NFDM, secluded hidden sector: (a) The evolution of energy density of χ (red) and A' (blue) with (solid) all relevant processes; (dot-dashed) only $3 \to 2$ processes; (dashed) only $2 \to 2$ processes. The equilibrium distribution of χ (green) and A' (orange) are also shown for reference; (b) contours of the observed present-day relic density in the $m_\chi - r$ parameter space for different coupling constants α'.

the contour shapes in the regions $r \lesssim 1$, $1 \lesssim r \lesssim 1.5$ and $1.5 \lesssim r \lesssim 2$, which each show a distinct qualitative behavior:

(1) $r \lesssim 1$. Being lighter than χ, A' is the dominant DM constituent. The fastest process in this mass range is the $2 \to 2$ process $\chi\bar{\chi} \to A'A'$. Significantly below $r = 1$, the second fastest process is $3A' \to \chi\bar{\chi}$, since $n_{A',0} > n_{\chi,0}$. Near the threshold, with $n_{A',0} \sim n_{\chi,0}$, all of the other possible $3 \to 2$ processes ($\chi\chi A' \to \chi\chi$, $\chi\bar{\chi}A' \to \chi\bar{\chi}$, $\chi\bar{\chi}A' \to A'A'$, $\chi A'A' \to \chi A'$, as well as $\chi\chi\bar{\chi} \to \chi A'$ plus any conjugate processes) become important. The relic abundance curves in fig. 3 are computed with all of these processes taken into account in the complete Boltzmann equations shown in eq. (D1) and (D2).

(2) $1 \lesssim r \lesssim 1.5$. χ is the dominant DM component. The fastest reaction is $A'A' \leftrightarrow \chi\bar{\chi}$, and it enforces $n_{A'} = n_{A',0} n_\chi / n_{\chi,0}$ during the freeze-out, and the second fastest reaction is now $\chi\chi\bar{\chi} \to \chi A'$, which determines the DM abundance. The $3 \to 2$ rate goes as $n_\chi^2 \langle \sigma v^2 \rangle$, which depends only weakly on r through the phase space. Therefore there is no strong correlation between the abundance and r in this region.

(3) $1.5 \lesssim r \lesssim 2$. χ is the dominant DM component, but now its abundance is determined by the two freeze-out events $A'A' \to \chi\bar{\chi}$ (whose rate becomes comparable to Hubble at later times) followed by $\chi\chi\bar{\chi} \to \chi A'$. At large $r \lesssim 2$, just before freeze-out completes, both reactions are faster than H, allowing one to estimate the freeze-out times. Taking the $2 \to 2$ and $3 \to 2$ rates $\sim H$, and $n_{A'} \simeq n_{A',0} n_\chi / n_{\chi,0}$ enforced by fast $3 \to 2$ scatterings, we can analytically derive contours consistent with the numerical results.

III. CONSTRAINTS

The parameter space of NFDM is constrained by a variety of experimental observations: (i) dark photon limits coming from the cooling of SN1987a [34]; (ii) similar bounds from beam dump experiments [35,36]; (iii) limits on the thermally-averaged cross section of $\chi\bar{\chi} \to e^+e^-$ deduced from the CMB power spectrum measured by Planck [31,32,37–39], and (iv) direct detection constraints on the dark matter-nucleon scattering cross section from PandaX-II [40], LUX [41], and CDMSLite [42]. Although we have only assumed a coupling to electrons in much of this analysis for simplicity, these direct detection limits are relevant to the vector-portal DM model considered here.

Future direct detection experiments including SuperCDMS SNOLAB [43], as well as electron scattering off germanium [44–47] and graphene [48] are also shown in the same plot. Other current limits from XENON10 [49], indirect detection [50] and lower bounds on m_χ from $N_{\rm eff}$ [51] are subdominant to the current constraints presented here and are not shown.

Figure 4 summarizes these constraints in the m_χ-ϵ plane for the illustrative value of $r = 1.8$, with α' fixed to give the correct present-day relic density, subject to the perturbativity constraint $\alpha' \leq 4\pi$. At a large ϵ and small m_χ, conventional freeze-out from $\chi\bar{\chi} \to e^+e^-$ annihilations dominates over the NFDM mechanism, but this is ruled out by the CMB constraint. The approximately horizontal red dashed contour shows the minimum value of ϵ for which the visible and dark sectors are in kinetic equilibrium, estimated in Eq. (4).

Self-interactions between dark matter particles with a cross section $\sigma_{\rm SI} \sim 0.1 \lesssim \sigma_{\rm SI}/m_\chi \lesssim 1$ cm^2 g^{-1} can potentially resolve the core-cusp and the too-big-to-fail problems

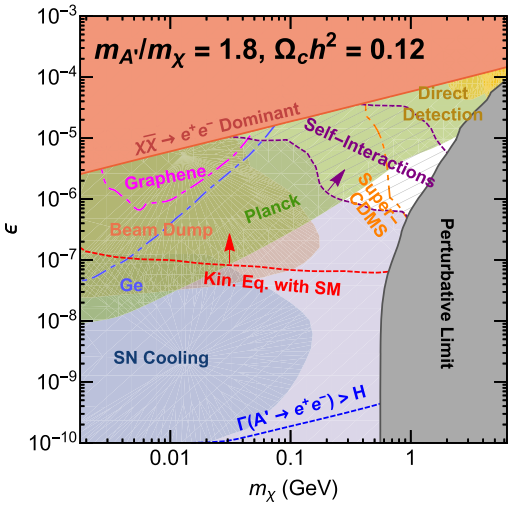

FIG. 4. Constraints in the m_χ-ϵ plane for the case of $m_{A'}/m_\chi = 1.8$, with α' chosen to produce the observed relic density. The allowed region is shown in white. The upper-left shaded region (red) indicates where freeze-out is dominated by the conventional $\chi\bar{\chi} \to e^+e^-$ annihilations. Limits are derived from the CMB power spectrum [31] (green), beam dump experiments [35,36] (pale orange), SN1987a cooling [34] (blue), direct detection [40–42] (yellow) and perturbativity, $\alpha' \geq 4\pi$ (gray). The projected reach of SuperCDMS [43] (orange dot-dashed line), electron ionization of graphene [48] (magenta dot-dashed line) and germanium in a low-threshold experiment [44] (blue dot-dashed line) are also shown. The curve near $\epsilon \sim 10^{-7}$ indicates where kinetic equilibrium with SM is established (red dashed line). The region of parameter space where the self-interaction cross section exceeds current limits ($\sigma/m_\chi > 1$ cm^2/g) (purple), and the region where the self-interaction cross section can potentially solve the small-scale structure problems (0.1 cm^2/g $< \sigma/m_\chi <$ 1 cm^2/g) (purple dashed lines) are displayed. The purple arrow points into the region allowed by self-interaction bounds, above and to the right of the line. The A' decay rate is faster than H at freeze-out above the lowest (blue) curve.

of small-structure formation with cold DM [52–54] while remaining consistent with experimental constraints, which set an upper bound of between 1–2 cm^2 g^{-1} [55–57]. A DM mass of $m_\chi \sim (0.1\text{–}1)$ GeV with $\epsilon \sim 10^{-7}$–10^{-6} in our model leads to a velocity-independent self-interaction cross section that lies within this range, and can provide a possible solution to both puzzles (though recent analysis of clusters indicates some preference for a velocity-dependent cross section [58]). The preferred region is between the purple dashed lines in Fig. 4, while the cosmologically constrained region is shown in purple.

IV. SUMMARY AND OUTLOOK

We have demonstrated a novel scenario called *not-forbidden dark matter*, where an allowed $3 \to 2$ annihilation process compensates for its conventional $2 \to 2$ counterpart being kinematically forbidden during thermal freeze-out. This mechanism can be potentially important in a variety of hidden sector models, including vector-portal, scalar-portal and composite DM. The DM mass and the mediator (or second DM) mass are of the same order, which would naturally arise in a hidden sector characterized by a single scale.

Taking the vector-portal DM model as an example, we found that in some parts of the NFDM parameter space, the combined effect of $3 \to 2$, $2 \to 2$ and A' decay channels is to significantly prolong the period of freeze-out. The commonly-neglected $3 \to 2$ annihilation channel can change the predicted relic density by orders of magnitude. Although this model is restricted by an abundance of experimental constraints, viable examples remain in the mass range $\sim(0.1\text{–}1)$ GeV, with a self-interaction cross section that is coincidentally of the right order for solving the small scale structure problems of ΛCDM cosmological simulations. This is a well-motivated target for future direct detection [43,45,46] and dark photon searches [59–63].

While we were completing this work, [64] appeared, presenting a related idea. Their work focuses on keV-MeV scalar DM and requires additional scalar "assister" particles.

ACKNOWLEDGMENTS

We thank Yonit Hochberg, Lina Necib, Nicholas Rodd, Joshua Ruderman, and Yotam Soreq for useful discussions. We also thank A. D. Dolgov for translating his early work on DM $3 \to 2$ annihilations [65,66] and bringing it to our attention. J. C. is supported by NSERC (Canada) and FRQNT (Québec). H. L., T. S. and W. X. are supported by the U.S. Department of Energy under Contracts No. DE-SC-00012567 and No. DE-SC-0013999.

APPENDIX A: COUPLED BOLTZMANN EQUATIONS AND PROLONGED FREEZE-OUT

As mentioned above, an essential difference between NFDM and conventional DM freeze-out is the importance of tracking the evolution of both the DM χ and the mediator particle (in our model, A'), by solving the coupled Boltzmann equations [Eqs. (2) and (3) for relevant terms when $r \gtrsim 1$, Eqs. (D1) and (D2) for the complete equations] for their respective densities. The presence of two equations implies that more than one scattering (or decay) process can be important for determining the final abundance; hence both the fastest and second fastest reactions are typically relevant.

This is in contrast to conventional DM freeze-out based upon a single Boltzmann equation, where the abundance depends upon the strongest channel. In the large ϵ limit of our model, A' decay is the fastest process, and enforces equilibrium of A', $n_{A'} = n_{A',0}$. Hence smaller values of ϵ are necessary to realize the rich cosmology that comes from the interplay of the coupled Boltzmann equations of χ and A'.

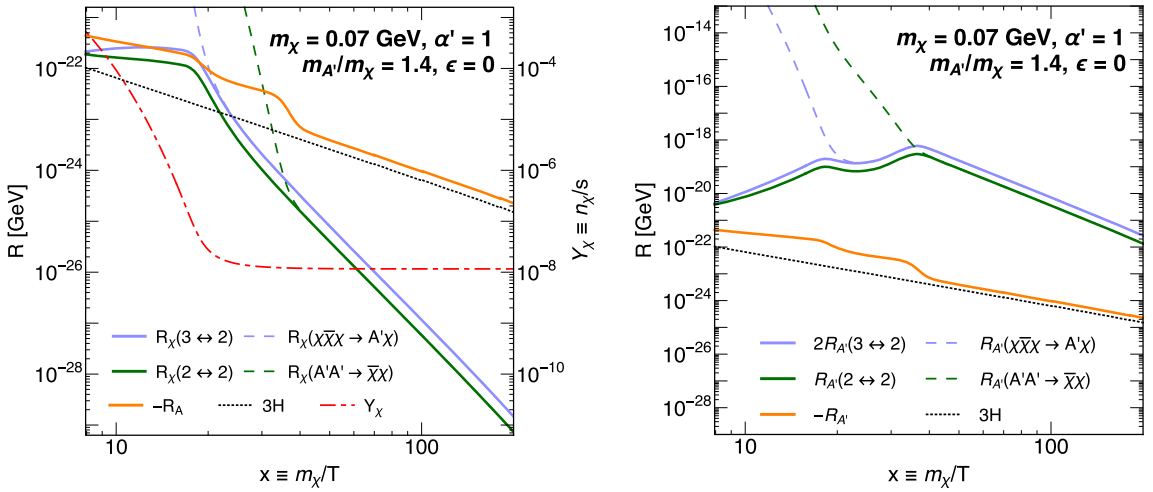

FIG. 5. Rates of different processes during freeze-out for $m_{A'}/m_\chi = 1.4$: (a) evolution of $R_\chi(3 \leftrightarrow 2)$ (light blue), $R_\chi(2 \leftrightarrow 2)$ (green) and the A' total rate $R_{A'}$ (orange) as a function of $x \equiv m_\chi/T$. Rates for processes in one direction, $R_\chi(\chi\chi\bar{\chi} \to A'\chi)$ (light blue, dashed) and $R_\chi(A'A' \to \chi\bar{\chi})$ (green, dashed) are also shown. The dark matter abundance Y_χ (red, dot-dashed) is plotted, with the appropriate (dimensionless) units given on the right-hand axis; (b) evolution of $R_{A'}(3 \leftrightarrow 2)$ (light blue), $R_{A'}(2 \leftrightarrow 2)$ (green) and the A' total rate $R_{A'}$ (orange) as a function of x. Rates for processes in one direction, $R_{A'}(\chi\chi\bar{\chi} \to A'\chi)$ (light blue, dashed) and $R_{A'}(A'A' \to \chi\bar{\chi})$ (green, dashed) are also shown. The evolution of the Hubble rate H (black, dotted) is shown in both plots for reference.

To simplify the subsequent discussion, we assume that these ϵ-suppressed reactions are negligibly slow, i.e. we work in the secluded dark sector regime of the NFDM model.

It is useful to define the net rate of $3 \leftrightarrow 2$ or $2 \leftrightarrow 2$ interactions per χ or A' particle by considering the collision terms in the Boltzmann equations, written in the form $R_\chi \equiv d\log n_\chi/dt = -3H - R_\chi(3 \leftrightarrow 2) + R_\chi(2 \leftrightarrow 2)$ and $R_{A'} \equiv d\log n_{A'}/dt = -3H + R_{A'}(3 \leftrightarrow 2) - R_{A'}(2 \leftrightarrow 2)$, where

$$\begin{aligned} R_\chi(3 \leftrightarrow 2) &\equiv 2\frac{n_{A'}}{n_\chi} R_{A'}(3 \leftrightarrow 2) \\ &= \frac{1}{4}\langle\sigma v^2\rangle_{\chi\chi\bar{\chi}\to\chi A'}\left(n_\chi^2 - n_{\chi,0}^2\frac{n_{A'}}{n_{A',0}}\right) \\ &\equiv R_\chi(\chi\chi\bar{\chi} \to \chi A') - R_\chi(\chi A' \to \chi\chi\bar{\chi}), \end{aligned} \tag{A1}$$

$$\begin{aligned} R_\chi(2 \leftrightarrow 2) &\equiv \frac{n_{A'}}{n_\chi} R_{A'}(2 \leftrightarrow 2) \\ &= \langle\sigma v\rangle_{A'A'\to\bar{\chi}\chi}\left(\frac{n_{A'}^2}{n_\chi} - \frac{n_{A',0}^2 n_\chi}{n_{\chi,0}^2}\right) \\ &\equiv R_\chi(A'A' \to \bar{\chi}\chi) - R_\chi(\bar{\chi}\chi \to A'A'). \end{aligned} \tag{A2}$$

Likewise, we define $2(n_{A'}/n_\chi)R_{A'}(\chi\chi\bar{\chi} \to \chi A') \equiv R_\chi(\chi\chi\bar{\chi} \to \chi A')$ and so on for the unidirectional rates. In this way, the signs for these definitions have been chosen so that all of the rates of individual subprocesses are now positive, although the overall rates R_χ and $R_{A'}$ can have any sign. When $m_{A'} > m_\chi$ and $T < m_\chi, m_{A'}$, the lower density of A' relative to χ implies that $R_\chi(3 \leftrightarrow 2)$ ($R_\chi(2 \leftrightarrow 2)$) is generally smaller in magnitude than $R_{A'}(3 \leftrightarrow 2)$ ($R_{A'}(2 \leftrightarrow 2)$). Thus the rates R_χ tend to fall below H earlier than the corresponding rates $R_{A'}$. This separation between freeze-out of χ and A' is the origin of the prolonged duration of the overall freeze-out process.

Suppose that only one channel, for example $2 \to 2$, occurs fast enough such that $R_\chi(A'A' \to \bar{\chi}\chi) \gg H$; then this rate tends to be nearly equal to that of the reverse reaction, $R_\chi(\bar{\chi}\chi \to A'A')$, enforcing the condition $n_{A'}^2 \simeq n_{A',0}^2 n_\chi^2/n_{\chi,0}^2$ (though the cancellation is imperfect, so that the total rate $R_\chi(2 \leftrightarrow 2)$ is also typically greater than H). This by itself is not sufficient to force both the χ and A' densities to track their equilibrium values. For that, one generically needs both $R_\chi(3 \leftrightarrow 2) > H$ and $R_\chi(2 \leftrightarrow 2) > H$ so that both independent combinations $n_\chi - n_{\chi,0}$ and $n_{A'} - n_{A',0}$ are driven to zero.[1] This is always true at sufficiently early times, allowing us to use equilibrium initial conditions for the Boltzmann equations. The DM density n_χ starts to deviate from equilibrium when the rate of the weaker annihilation channel becomes comparable to H; hence the second-strongest channel initiates the freeze-out process.

To illustrate this behavior, we show some examples of the evolution of the rates in Figs. 5(a), 5(b), and 6. Each example has the same DM mass $m_\chi = 70$ MeV, coupling $\alpha' = 1$, and kinetic mixing $\epsilon = 0$, but different values of $r = 1.4,\ 1.7,\ 1.9$. In these figures, the dot-dashed lines

[1]The typical behavior is that the strongest process is such that both the forward and backward rates exceed H, as well as their difference. For the second-strongest, only one of these need be greater than H.

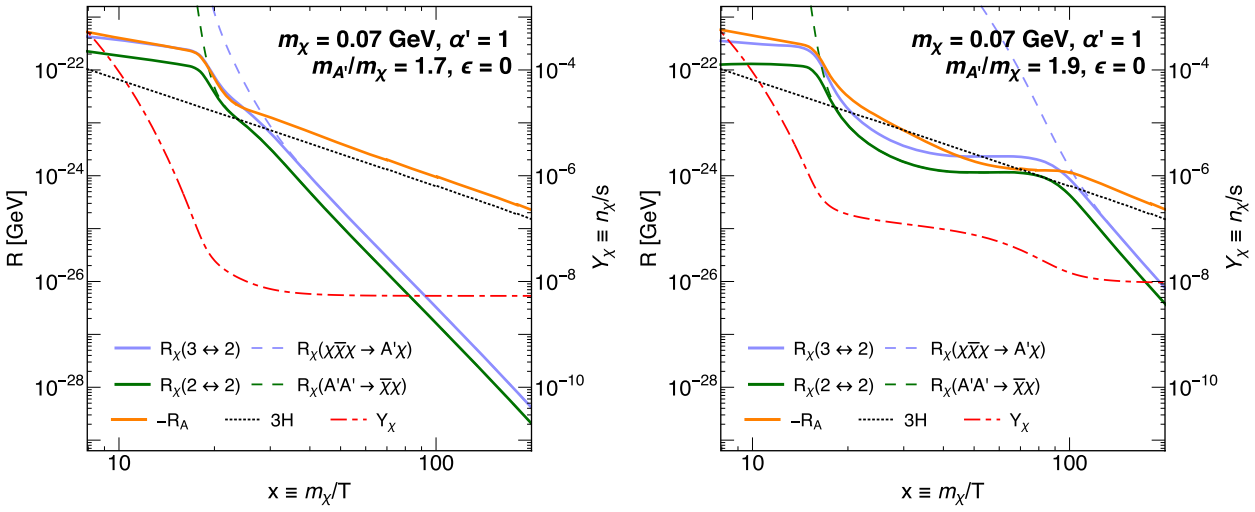

FIG. 6. Same as Fig. 5(a), but with (a) $m_{A'}/m_\chi = 1.7$, and (b) $m_{A'}/m_\chi = 1.9$.

corresponding to the evolution of DM number density are shown to highlight the time of DM freeze-out. For $r = 1.4$, the freeze-out period is relatively short; for $r = 1.7$, freeze-out is prolonged; and for $r = 1.9$, the freeze-out is prolonged further and may indeed be thought of as two separated freeze-outs.

In Fig. 5(b), we show the two rates $R_{A'}(3 \leftrightarrow 2)$ and $R_{A'}(2 \leftrightarrow 2)$, which are much larger than H; these cancel each other to order H. The behavior is similar for other values of r. Since $R_{A'}(3 \leftrightarrow 2) \simeq R_A(2 \leftrightarrow 2)$, Eqs. (A1), (A2) imply that $R_\chi(3 \leftrightarrow 2) \simeq 2R_\chi(2 \leftrightarrow 2)$. This relation is demonstrated in Fig. 5(a) and 6.

Comparing these net rates however does not tell us which process controls freeze-out. Instead, we should look at the dashed lines, which indicate the unidirectional rates, $R_\chi(A'A' \to \bar{\chi}\chi)$ and $R_\chi(\chi\chi\bar{\chi} \to \chi A')$. Processes are out of equilibrium when these dashed lines overlap with the solid lines. From the unidirectional rates, we can identify the weaker annihilation channel and thus which process initiates the freeze-out. For $r = 1.4$, the weaker process is $3 \to 2$, and fig. 5a confirms that the freeze-out is indeed triggered by $3 \to 2$. For $r = 1.7$ and $r = 1.9$, the dashed line for $R_\chi(A'A' \to \bar{\chi}\chi)$ in fig. 6 merges with the solid line, $R_\chi(2 \leftrightarrow 2)$, when the rate is about $3H$. It is the weaker channel $2 \to 2$ that initiates freeze-out.

One difference between $r < 1.5$ and $r > 1.5$ in figs. 5(a) versus 6 is that $r > 1.5$ normally has a longer freeze-out. The duration depends upon whether the rate of the weaker annihilation channel is sensitive to $n_{A'}$. For $r > 1.5$, the weaker process $2 \to 2$ has the rate $R_\chi(A'A' \to \chi\chi) \sim \langle\sigma v\rangle_{A'A' \to \bar{\chi}\chi} n_{A'}^2/n_\chi$. Prior to the final freeze-out, the larger $3 \to 2$ rate imposes the constraint that $n_{A'} \simeq n_{A',0} n_\chi^2/n_{\chi,0}^2 \sim r^{3/2} x^{3/2} m^{-3} \exp((2-r)x) n_\chi^2$. Since $n_{A'}$ increases with time, this means that the $2 \to 2$ rate $R_\chi(A'A' \to \bar{\chi}\chi)$ can be kept at the same order as H for a long period. For this reason, the freeze-out is prolonged. For $r < 1.5$, the duration is relatively short, because the rate of the weaker $3 \to 2$ process goes as $R_\chi(\chi\chi\bar{\chi} \to \chi A') \propto n_\chi^2$, where n_χ is decreasing with time.

Armed with our insight that the second-strongest channel matters critically for freeze-out, and having understood the reason that the freeze-out process is longer for $r \gtrsim 1.5$, we can also explain the shape of the contours in Fig. 2(b) in the main text. As discussed briefly in the main text, there are several important regimes:

(1) For $n_{A'} = n_{A',0}$, corresponding to large ϵ, the two Boltzmann equations are reduced to one, and the shape of the contours can be understood using the usual parametrics of DM freeze-out. This behavior occurs for the contours overlapping the dashed contours in Fig. 2(b).

(2) For $\epsilon = 0$, we recover the secluded case discussed above, where the interplay of the $3 \leftrightarrow 2$ and $2 \leftrightarrow 2$ processes controls the freeze-out. This behavior also occurs in the region where $r \sim 1.5$ and $\alpha' = 10$, because the $3 \to 2$ and $2 \to 2$ rates for DM are significantly larger than $R_\chi(A' \to e^+e^-) \equiv \Gamma(A' \to e^+e^-) n_{A'}/n_\chi$, so that $\Gamma_{A' \to e^+e^-}$ can be neglected.[2]

(3) For moderate ϵ, the rates of the three processes should be compared in order to determine which is weakest, and hence irrelevant to the DM freeze-out. The relevant rates to compare are $R_\chi(\chi\bar{\chi}\chi \to A'\chi)$, $R_\chi(A'A' \to \chi\bar{\chi})$ and $R_\chi(A' \to e^+e^-)$, evaluated at the Hubble crossing time of the annihilation processes. Consider the case where $2 \to 2$ has a

[2]We define the A' decay rate $R_\chi(A' \to e^+e^-)$ with respect to the DM density; even though this quantity does not appear in Boltzmann equation of DM, the coupling to the A' Boltzmann equation will cause $A' \to e^+e^-$ to play an important role in determining the rate for DM processes in some cases.

lower rate than $3 \to 2$, such that it falls below H first. Whether the DM density freezes out or not at this time depends on the relative sizes of $R_\chi(A' \to e^+e^-)$ and $R_\chi(A'A' \to \chi\bar{\chi})$. When $R_\chi(A'A' \to \chi\bar{\chi}) < R_\chi(A' \to e^+e^-)$, the larger $R_\chi(A' \to e^+e^-)$ rate in the coupled Boltzmann equations provides enough constraints to keep n_χ and $n_{A'}$ near their equilibrium values. An example of this more complex case is shown in Fig. 2(a) of the main text, where the freeze-out starts when $R_\chi(3 \leftrightarrow 2) \sim H$. Using the Boltzmann equation of A', this rate is determined by the A' decay, $R_\chi(3 \leftrightarrow 2) \sim 2R_\chi(A' \to e^+e^-)$; freeze-out then terminates when $R_\chi(\chi\bar{\chi}\chi \to A'\chi) \sim H$.

More broadly, this case is realized when $\epsilon = 10^{-6}$, and either r is close to 2, or α' is large and $r > 1.5$. Figure 2(b) shows that in this region the $\epsilon = 10^{-6}$ contours (solid lines) do not overlap with the dashed contours, for which the constraint $n_{A'} = n_{A',0}$ is imposed. In this regime the $3 \to 2$ rate is the largest, and when $R_\chi(A'A' \to \chi\bar{\chi}) \sim H$, $R_\chi(A' \to e^+e^-) > R_\chi(A'A' \to \chi\bar{\chi})$. The freeze-out is thus controlled by $3 \to 2$ processes and the decay of A'. In this case the A' decay rate is not fast enough to keep the A' abundance in equilibrium, and both $n_{A'}$ and n_χ are increased during freeze-out relative to their values when the A' s remain in equilibrium. The χ annihilation rate thus needs to be increased to maintain the correct relic density, requiring lower χ masses; this is the reason that the contours in Fig. 2(b) bend toward lower masses as ϵ is decreased, for large r.

APPENDIX B: DEPENDENCE ON T_d

If the dark sector is secluded, its temperature T_d may differ from that of the visible sector, $T_{\rm SM}$. This difference affects the evolution of the χ and A' densities and ultimately the DM relic abundance. In the Boltzmann equations [Eqs. (2, 3) in the main text], taking $T_d \neq T_{\rm SM}$ changes the equilibrium densities, so that $n_{\chi,0} \sim \exp(-m_\chi/T_d) = \exp(-x/\gamma)$, where we have defined $\gamma \equiv T_d/T_{\rm SM}$, and x is still given by $x \equiv m_\chi/T_{\rm SM}$. Likewise, $n_{A',0} \sim \exp(-rx/\gamma)$. Keeping in mind that H is determined by $T_{\rm SM}$, we can solve the Boltzmann equations for $n_\chi(x)$ and $n_{A'}(x)$ with the γ-dependence coming from the equilibrium densities.

Fig. 7 shows the behavior of the ratio of relic abundances $\Omega_c(T_d)/\Omega_c(T_d = T_{\rm SM})$ as a function of T_d for $0.1 \le \gamma \le 1$. Having $T_d < T_{\rm SM}$ leads to an earlier freeze-out, since the exponential decrease in the equilibrium densities occurs more rapidly. For values of r where the backward and forward $3 \to 2$ processes fall out of equilibrium at freeze-out, we expect that $n_\chi^2 \sim H/\langle\sigma v^2\rangle_{\chi\chi\bar{\chi}\to\chi A'} \sim 1/x_f^2$, and therefore that the relic abundance scales as $\Omega_c \sim x_f^2$. For

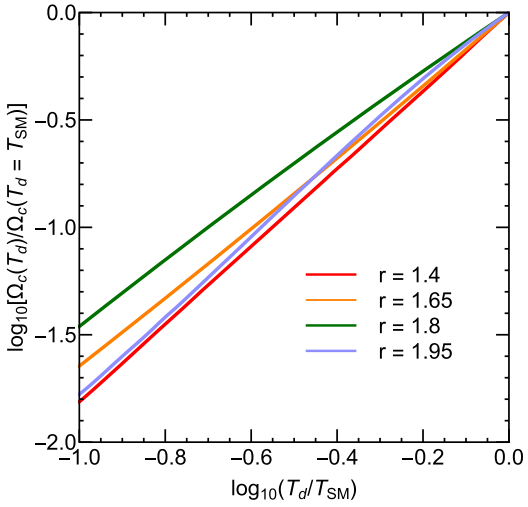

FIG. 7. Ratio of the relic abundance when $T_d < T_{\rm SM}$ to the relic abundance with $T_d = T_{\rm SM}$ as a function of $\gamma \equiv T_d/T_{\rm SM}$ for $r \equiv m_{A'}/m_\chi = 1.4$ (red), 1.65 (orange), 1.8 (green) and 1.95 (light blue).

$1 < r \lesssim 1.5$ where the $3 \to 2$ process determines the DM abundance, the exponential dependence of $n_{\chi,0}$ with x/γ results in $\Omega_c \sim \gamma^2$. On the other hand, in the case of $r \lesssim 2$, the second freeze-out occurs at $n_\chi \propto n_{\chi,0}^4/n_{A',0}^2$, and a similar argument leads again to $\Omega_c \sim \gamma^2$. At intermediate values of r, both the $3 \to 2$ and $2 \to 2$ processes freeze out at similar times. For a $2 \to 2$ freeze-out, $n_\chi \sim H/\langle\sigma v\rangle_{\bar{\chi}\chi\to A'A'}$, and as a result $\Omega_c \sim \gamma$. Qualitatively, we expect the γ dependence to lie between these two regimes for intermediate values of r.

APPENDIX C: CONSTRAINTS AT DIFFERENT r

Figure 8 shows the constraints in the m_χ-ϵ plane for two representative values of r, with α' fixed to give the correct present-day relic density. These have the same general features as in Fig. 4 in the main text, but also exhibit several distinct characteristics that we explain here.

At $r = 1.4$, the transition from the secluded limit ($\epsilon \to 0$) to the kinetic equilibrium limit occurs in the range $\epsilon \sim 10^{-9}$–10^{-6}, which leads to a rapid decrease in α' between these two phases at a fixed value of m_χ. This explains the rapid change in the behavior of the region with suitable self-interaction for $m_\chi \lesssim 100$ MeV.

At $r = 1.6$, the most distinctive feature is the change in behavior of the region where annihilations to e^+e^- dominates at $m_\chi \sim 100$ MeV. At masses smaller than this point, the $2 \to 2$ process is freezes out last, while at larger masses, it is the $3 \to 2$ process which does so. This difference accounts for the change in the slope of the boundary. There is no such transition for the other two cases, since at $r = 1.8$, the $3 \to 2$ process always freezes out last, while for $r = 1.4$ it is the $2 \to 2$ process instead.

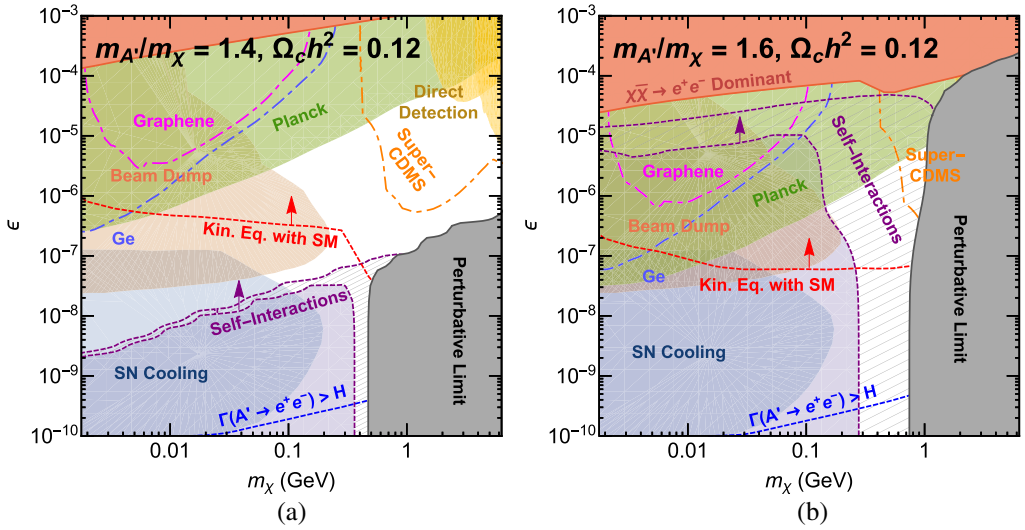

FIG. 8. Same as Fig. 4 in the main text, but with (a) $m_{A'}/m_\chi = 1.4$, and (b) $m_{A'}/m_\chi = 1.6$.

For all values of r, a significant part of the $m_\chi - \epsilon$ parameter space is still consistent with the present-day relic density while evading experimental constraints, showing that the NFDM scenario is robust against taking different values of $r \gtrsim 1.5$.

APPENDIX D: COMPLETE BOLTZMANN EQUATIONS

The complete Boltzmann equations, including all relevant $2 \to 2$ and $3 \to 2$ processes for the full range of r considered is

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -\frac{1}{4}\langle\sigma v^2\rangle_{\substack{\chi\chi\bar{\chi} \\ \to \chi A'}}\left(n_\chi^3 - n_{\chi,0}^2 n_\chi \frac{n_{A'}}{n_{A',0}}\right) + \langle\sigma v\rangle_{\substack{A'A' \\ \to \bar{\chi}\chi}}\left(n_{A'}^2 - n_{A',0}^2\frac{n_\chi^2}{n_{\chi,0}^2}\right) - \frac{1}{2}\langle\sigma v^2\rangle_{\substack{\chi\bar{\chi}A' \\ \to A'A'}}\left(n^2 n_{A'} - n_{\chi,0}^2\frac{n_{A'}^2}{n_{A',0}}\right)$$
$$+\frac{1}{3}\langle\sigma v^2\rangle_{\substack{A'A'A' \\ \to \chi\bar{\chi}}}\left(n_{A'}^3 - n_{A',0}^3\frac{n_\chi^2}{n_{\chi,0}^2}\right) - \frac{1}{2}\langle\sigma v\rangle_{\chi\bar{\chi}\to e^+e^-}(n_\chi^2 - n_{\chi,0}^2), \tag{D1}$$

$$\frac{dn_{A'}}{dt} + 3Hn_{A'} = \frac{1}{8}\langle\sigma v^2\rangle_{\substack{\chi\chi\bar{\chi} \\ \to \chi A'}}\left(n_\chi^3 - n_{\chi,0}^2 n_\chi \frac{n_{A'}}{n_{A',0}}\right) - \langle\sigma v\rangle_{\substack{A'A' \\ \to \bar{\chi}\chi}}\left(n_{A'}^2 - n_{A',0}^2\frac{n_\chi^2}{n_{\chi,0}^2}\right) - \Gamma_{A'\to f\bar{f}}(n_{A'} - n_{A',0})$$
$$-\frac{1}{4}(\langle\sigma v^2\rangle_{\substack{\chi\bar{\chi}A' \\ \to \chi\bar{\chi}}} + \langle\sigma v^2\rangle_{\substack{\chi\chi A' \\ \to \chi\chi}})(n_\chi^2 n_{A'} - n_\chi^2 n_{A',0}) + \frac{1}{4}\langle\sigma v^2\rangle_{\substack{\chi\bar{\chi}A' \\ \to A'A'}}\left(n_\chi^2 n_{A'} - n_{\chi,0}^2\frac{n_{A'}^2}{n_{A',0}}\right)$$
$$-\frac{1}{2}\langle\sigma v^2\rangle_{\substack{\chi A'A' \\ \to \chi A'}}(n_\chi n_{A'}^2 - n_\chi n_{A'} n_{A',0}) - \frac{1}{2}\langle\sigma v^2\rangle_{\substack{A'A'A' \\ \to \chi\bar{\chi}}}\left(n_{A'}^3 - n_{A',0}^3\frac{n_\chi^2}{n_{\chi,0}^2}\right). \tag{D2}$$

The symmetry factors preceding each term properly account for the number of identical particles in the initial state, the net number of particles created or destroyed in each annihilation process, as well as conjugate processes. These equations are used in all numerical calculations shown in the paper.

APPENDIX E: CROSS SECTIONS AND DECAY RATES

The decay rate for $A' \to e^+e^-$ is

$$\Gamma(A' \to e^+e^-) = \frac{\epsilon^2\alpha_{em}}{3} m_{A'}\left(1 + 2\frac{m_e^2}{m_{A'}^2}\right)\sqrt{1 - 4\frac{m_e^2}{m_{A'}^2}}. \tag{E1}$$

TABLE I. List of initial- and final-state averaged squared matrix element $\overline{|\mathcal{M}|^2}$ of each process, as well as the phase space factor P such that $\langle\sigma v\rangle$ or $\langle\sigma v^2\rangle = P\overline{|\mathcal{M}|^2}$. All values are evaluated at the kinematic threshold. For the last two processes, we use the expression for $\chi\bar{\chi} \to A'A'$ for $r < 1$ and $A'A' \to \chi\bar{\chi}$ for $r > 1$.

Process	$\overline{\lvert\mathcal{M}\rvert^2}$	Phase Space
$A'A'A' \to \chi\bar{\chi}$	$\frac{g'^6(153r^6-47r^4-60r^2+24)}{9m_\chi^2 r^8}$	$\frac{\sqrt{9r^2-4}}{48\pi m_\chi^3 r^4}$
$\chi A'A' \to \chi A'$	$\frac{2g'^6(195r^8+1156r^7+4670r^6+9444r^5+12214r^4+11192r^3+6732r^2+2272r+320)}{9m_\chi^2(r+1)^2(r+2)^4(2r+1)(r^2-2r-2)^2}$	$\frac{3\sqrt{3}\sqrt{3r^2+8r+4}}{32\pi m_\chi^3 r(2r+1)^2}$
$\chi\chi A' \to \chi\chi$	$\frac{2g'^6 r(r+4)}{3m_\chi^2(r+1)^2(r+2)^2}$	$\frac{\sqrt{r(r+4)}}{32\pi m_\chi^3 r(r+2)}$
$\chi\bar{\chi}A' \to \chi\bar{\chi}$	$\frac{g'^6(r+4)(9r^6+24r^5+4r^4-40r^3+168r^2-224r+128)}{6m_\chi^2 r^3(r-2)^2(r+1)^2(r+2)^2}$	$\frac{\sqrt{r(r+4)}}{16\pi m_\chi^3 r(r+2)}$
$\chi\bar{\chi}A' \to A'A'$	$\frac{16g'^6(21r^6-4r^5-17r^4+24r^3+216r^2+288r+112)}{27m_\chi^2(r-2)^4(r+1)^4(r+2)^2}$	$\frac{9\sqrt{-3r^2+4r+4}}{128\pi m_\chi^3 r(r+2)}$
$\chi\bar{\chi}\chi \to A'\chi$	$\frac{g'^6(r-4)(r+4)(-32r^8+167r^6-534r^4+668r^2-512)}{36m_\chi^2(r^2-4)^4(r^2+2)^2}$	$\frac{\sqrt{r^4-20r^2+64}}{96\pi m_\chi^3}$
$\chi\bar{\chi} \to A'A'$	$\frac{16g'^4(1-r^2)}{9(r^2-2)^2}$	$\frac{9\sqrt{1-r^2}}{64\pi m_\chi^2}$
$A'A' \to \chi\bar{\chi}$	$\frac{32g'^4(r^4-1)}{9r^4}$	$\frac{\sqrt{r^2-1}}{8\pi m_\chi^2 r^3}$
$\chi\bar{\chi} \to e^+e^-$	$\frac{4e^2e^2g'^2(2+m_e^2/m_\chi^2)}{(r^2-4)^2}$	$\frac{\sqrt{1-m_e^2/m_\chi^2}}{8\pi m_\chi^2}$

For scattering cross sections, the thermally-averaged $2 \to 2$ cross section for the process $1 + 2 \to 3 + 4$ is given by

$$\langle\sigma v\rangle_{12\to34} = \frac{1}{S_f}\frac{1}{n_1 n_2}\int\prod_{i=1}^{4}\frac{g_i d^3\vec{p}_i}{(2\pi)^3 2E_i} \times (2\pi)^4\delta^4(p_1+p_2-p_3-p_4) f_1 f_2 \overline{|\mathcal{M}|^2}, \tag{E2}$$

where g_i is the number of degrees of freedom and f_i is the phase space distribution of species i. The averaged squared matrix element $\overline{|\mathcal{M}|^2}$ is averaged over both the initial and final state degrees of freedom. $S_f = \prod_i n_i!$ is a symmetry factor, where n_i is the number of identical particles of species i in the final state. Initial state symmetry factors are included explicitly in the Boltzmann equation, Eq. (2) and (3). This convention may differ from other sources in the literature.

Similarly, the thermally-averaged $3 \to 2$ cross section for the process $1 + 2 + 3 \to 4 + 5$ is

$$\langle\sigma v^2\rangle_{123\to45} = \frac{1}{S_f}\frac{1}{n_1 n_2 n_3}\int\prod_{i=1}^{5}\frac{g_i d^3 p_i}{(2\pi)^3 2E_i} \times (2\pi)^4\delta^4(p_1+p_2+p_3-p_4-p_5) f_1 f_2 f_3 \overline{|\mathcal{M}|^2}. \tag{E3}$$

For simplicity and unless otherwise stated, we give cross sections at the kinematic threshold of the respective processes. In this limit, thermally averaged cross sections are

$$\langle\sigma v\rangle_{12\to34} = \frac{g_3 g_4}{32\pi S_f m_1 m_2}\lambda^{1/2}(m_1+m_2, m_3, m_4)\overline{|\mathcal{M}|^2}, \tag{E4}$$

and

$$\langle\sigma v^2\rangle_{123\to45} = \frac{g_4 g_5}{64\pi S_f m_1 m_2 m_3} \times \lambda^{1/2}(m_1+m_2+m_3, m_4, m_5)\overline{|\mathcal{M}|^2}, \tag{E5}$$

where $\lambda(x,y,z) \equiv (1-(z+y)^2/x^2)(1-(z-y)^2/x^2)$. This expression agrees with the result for the specific process of $3\chi \to 2\chi$ computed in [67].[3]

In Table I, we list all of the number changing processes that are included in the Boltzmann equations Eq. (D1) and (D2), the initial- and final-state averaged squared matrix element $\overline{|\mathcal{M}|^2}$ of each process as well as the phase space factor P such that $\langle\sigma v\rangle$ or $\langle\sigma v^2\rangle = P\overline{|\mathcal{M}|^2}$. We define $r \equiv m_{A'}/m_\chi$ throughout.

Two other processes that are important to our analysis are $\chi e^\pm \to \chi e^\pm$ which maintains kinetic equilibrium between the dark sector and the SM, and dark matter-dark matter scattering.

(i) $\chi e^\pm \to \chi e^\pm$: this cross section is important in determining if the DM is in kinetic equilibrium with the SM. In the limit where $T < \mu_{e\chi}$, where $\mu_{e\chi}$ is the electron-DM reduced mass, we have

[3]Note that different conventions are used between this paper and [67].

$$\langle\sigma v\rangle = \frac{2(g'\epsilon e)^2\mu_{e\chi}^2}{\pi m_{A'}^4}\left(\frac{2T}{\pi\mu_{e\chi}}\right)^{1/2}. \tag{E6}$$

At high temperatures, it approaches the limit

$$\langle\sigma v\rangle \to \frac{(g'\epsilon e)^2}{4\pi m_{A'}^2}. \tag{E7}$$

To get accurate results, however, we must use the exact thermal average over the cross section for $\chi e^\pm \to \chi e^\pm$, which is given by:

$$\sigma = \frac{(g'\epsilon e)^2}{8\pi}\left[\frac{1}{s} + \frac{2}{r^2 m_\chi^2} + \frac{8m_e^2 + r^4 m_\chi^2}{r^2[h(m_\chi,s) + r^2 m_\chi^2 s]} - \frac{2(r^2 m_\chi^2 + s)}{h(m_\chi,s)}\log\left(1 + \frac{h(m_\chi,s)}{r^2 m_\chi^2 s}\right)\right], \tag{E8}$$

where $h(m_\chi,s) = [s-(m_\chi+m_e)^2][s-(m_\chi-m_e)^2]$. The thermal average is then given by

$$\langle\sigma v\rangle = \int_{M^2}^\infty \frac{ds}{\sqrt{s}}\cdot\frac{h(m_\chi,s)K_1(\sqrt{s}/T)\sigma}{8Tm_\chi^2 m_e^2 K_2(m_\chi/T)K_2(m_e/T)}, \tag{E9}$$

where $M = m_e + m_\chi$.

(ii) $\chi\chi \to \chi\chi$: the self-interaction cross section, averaged over particle-particle and particle-antiparticle scattering, is [9]

$$\frac{\sigma_{\rm SI}}{m_\chi} = \frac{3g'^4}{16\pi m_\chi^3}\frac{16 - 16r^2 + 5r^4}{r^4(r^2-4)^2}. \tag{E10}$$

[1] M. Pospelov, A. Ritz, and M. B. Voloshin, Phys. Lett. B **662**, 53 (2008).
[2] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, Phys. Rev. D **79**, 015014 (2009).
[3] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, Phys. Rev. Lett. **113**, 171301 (2014).
[4] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, Phys. Rev. Lett. **115**, 021301 (2015).
[5] H. M. Lee and M.-S. Seo, Phys. Lett. B **748**, 316 (2015).
[6] Y. Hochberg, E. Kuflik, and H. Murayama, J. High Energy Phys. 05 (2016) 090.
[7] N. Bernal, X. Chu, and J. Pradler, Phys. Rev. D **95**, 115023 (2017).
[8] R. T. D'Agnolo and A. Hook, Phys. Rev. D **91**, 115020 (2015).
[9] R. T. D'Agnolo and J. T. Ruderman, Phys. Rev. Lett. **115**, 061301 (2015).
[10] A. Delgado, A. Martin, and N. Raj, Phys. Rev. D **95**, 035002 (2017).
[11] E. D. Carlson, M. E. Machacek, and L. J. Hall, Astrophys. J. **398**, 43 (1992).
[12] D. Pappadopulo, J. T. Ruderman, and G. Trevisan, Phys. Rev. D **94**, 035005 (2016).
[13] N. Bernal, X. Chu, C. Garcia-Cely, T. Hambye, and B. Zaldivar, J. Cosmol. Astropart. Phys. 03 (2016) 018.
[14] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, Phys. Rev. Lett. **116**, 221302 (2016).
[15] N. Bernal and X. Chu, J. Cosmol. Astropart. Phys. 01 (2016) 006.
[16] M. Farina, D. Pappadopulo, J. T. Ruderman, and G. Trevisan, J. High Energy Phys. 12 (2016) 039.
[17] J. A. Dror, E. Kuflik, and W. H. Ng, Phys. Rev. Lett. **117**, 211801 (2016).
[18] S. Okawa, M. Tanabashi, and M. Yamanaka, Phys. Rev. D **95**, 023006 (2017).
[19] P. Bandyopadhyay, E. J. Chun, and J.-C. Park, J. High Energy Phys. 06 (2011) 129.
[20] F. D'Eramo and J. Thaler, J. High Energy Phys. 06 (2010) 109.
[21] K. Agashe, Y. Cui, L. Necib, and J. Thaler, J. Cosmol. Astropart. Phys. 10 (2014) 062.
[22] J. Berger, Y. Cui, and Y. Zhao, J. Cosmol. Astropart. Phys. 02 (2015) 005.
[23] J. Kopp, J. Liu, and X.-P. Wang, J. High Energy Phys. 04 (2015) 105.
[24] A. Berlin, D. Hooper, and G. Krnjaic, Phys. Lett. B **760**, 106 (2016).
[25] J. Kopp, J. Liu, T. R. Slatyer, X.-P. Wang, and W. Xue, J. High Energy Phys. 12 (2016) 033.
[26] S.-M. Choi, Y.-J. Kang, and H. M. Lee, J. High Energy Phys. 12 (2016) 099.
[27] F. Kahlhoefer, K. Schmidt-Hoberg, T. Schwetz, and S. Vogl, J. High Energy Phys. 02 (2016) 016.
[28] J. A. Adams, S. Sarkar, and D. Sciama, Mon. Not. R. Astron. Soc. **301**, 210 (1998).
[29] X.-L. Chen and M. Kamionkowski, Phys. Rev. D **70**, 043502 (2004).
[30] N. Padmanabhan and D. P. Finkbeiner, Phys. Rev. D **72**, 023508 (2005).
[31] P. A. R. Ade *et al.* (Planck), Astron. Astrophys. **594**, A13 (2016).
[32] T. R. Slatyer, Phys. Rev. D **93**, 023527 (2016).
[33] N. Bernal, C. Garcia-Cely, and R. Rosenfeld, J. Cosmol. Astropart. Phys. 04 (2015) 012.
[34] J. B. Dent, F. Ferrer, and L. M. Krauss, arXiv:1201.2683.
[35] J. D. Bjorken, R. Essig, P. Schuster, and N. Toro, Phys. Rev. D **80**, 075018 (2009).

[36] S. Andreas, C. Niebuhr, and A. Ringwald, Phys. Rev. D **86**, 095019 (2012).
[37] D. P. Finkbeiner, S. Galli, T. Lin, and T. R. Slatyer, Phys. Rev. D **85**, 043522 (2012).
[38] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, Phys. Rev. D **80**, 043526 (2009).
[39] H. Liu, T. R. Slatyer, and J. Zavala, Phys. Rev. D **94**, 063507 (2016).
[40] A. Tan *et al.* (PandaX-II), Phys. Rev. Lett. **117**, 121303 (2016).
[41] D. S. Akerib *et al.* (LUX), Phys. Rev. Lett. **116**, 161301 (2016).
[42] R. Agnese *et al.* (SuperCDMS), Phys. Rev. Lett. **116**, 071301 (2016).
[43] R. Agnese *et al.* (SuperCDMS), Phys. Rev. D **95**, 082002 (2017).
[44] S. K. Lee, M. Lisanti, S. Mishra-Sharma, and B. R. Safdi, Phys. Rev. D **92**, 083517 (2015).
[45] S. Derenzo, R. Essig, A. Massari, A. Soto, and T.-T. Yu, Phys. Rev. D **96**, 016026 (2017).
[46] R. Essig, J. Mardon, and T. Volansky, Phys. Rev. D **85**, 076007 (2012).
[47] R. Essig, M. Fernandez-Serra, J. Mardon, A. Soto, T. Volansky, and T.-T. Yu, J. High Energy Phys. 05 (2016) 046.
[48] Y. Hochberg, Y. Kahn, M. Lisanti, C. G. Tully, and K. M. Zurek, Phys. Lett. B **772**, 239 (2017).
[49] R. Essig, T. Volansky, and T.-T. Yu, Phys. Rev. D **96**, 043017 (2017).
[50] R. Essig, E. Kuflik, S. D. McDermott, T. Volansky, and K. M. Zurek, J. High Energy Phys. 11 (2013) 193.
[51] C. Boehm, M. J. Dolan, and C. McCabe, J. Cosmol. Astropart. Phys. 08 (2013) 041.
[52] M. Rocha, A. H. G. Peter, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, J. Onorbe, and L. A. Moustakas, Mon. Not. R. Astron. Soc. **430**, 81 (2013).
[53] D. N. Spergel and P. J. Steinhardt, Phys. Rev. Lett. **84**, 3760 (2000).
[54] J. Zavala, M. Vogelsberger, and M. G. Walker, Mon. Not. R. Astron. Soc. Lett. **431**, L20 (2013).
[55] D. Harvey, R. Massey, T. Kitching, A. Taylor, and E. Tittley, Science **347**, 1462 (2015).
[56] A. Robertson, R. Massey, and V. Eke, Mon. Not. R. Astron. Soc. **465**, 569 (2017).
[57] S. Tulin and H.-B. Yu, arXiv:1705.02358.
[58] M. Kaplinghat, S. Tulin, and H.-B. Yu, Phys. Rev. Lett. **116**, 041302 (2016).
[59] P. Ilten, J. Thaler, M. Williams, and W. Xue, Phys. Rev. D **92**, 115017 (2015).
[60] P. Ilten, Y. Soreq, J. Thaler, M. Williams, and W. Xue, Phys. Rev. Lett. **116**, 251803 (2016).
[61] S. Alekhin *et al.*, Rep. Prog. Phys. **79**, 124201 (2016).
[62] S. Gardner, R. J. Holt, and A. S. Tadepalli, Phys. Rev. D **93**, 115015 (2016).
[63] O. Moreno, arXiv:1310.2060.
[64] U. K. Dey, T. N. Maity, and T. S. Ray, J. Cosmol. Astropart. Phys. 03 (2017) 045.
[65] A. Dolgov, Yad. Fiz. **31**, 1522 (1980) [Sov. J. Nucl. Phys. **31**, 790 (1980)].
[66] A. D. Dolgov, arXiv:1705.03689.
[67] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, J. High Energy Phys. 08 (2017) 078.