PHYSICAL REVIEW D **94,** 103529 (2016)

Suppressing structure formation at dwarf galaxy scales and below: Late kinetic decoupling as a compelling alternative to warm dark matter

Torsten Bringmann[*]
Department of Physics, University of Oslo, Box 1048, N-0371 Oslo, Norway

Håvard Tveit Ihle[†]
Institute of Theoretical Astrophysics, University of Oslo, N-0315 Oslo, Norway

Jörn Kersten[‡]
University of Bergen, Institute for Physics and Technology, Postboks 7803, N-5020 Bergen, Norway

Parampreet Walia[§]
Department of Physics, University of Oslo, Box 1048, N-0371 Oslo, Norway
(Received 28 June 2016; published 29 November 2016)

Warm dark matter cosmologies have been widely studied as an alternative to the cold dark matter paradigm, the characteristic feature being a suppression of structure formation on small cosmological scales. A very similar situation occurs if standard cold dark matter particles are kept in local thermal equilibrium with a, possibly dark, relativistic species until the Universe has cooled down to keV temperatures. We perform a systematic phenomenological study of this possibility, and classify all minimal models containing dark matter and an arbitrary radiation component that allows such a late kinetic decoupling. We recover explicit cases recently discussed in the literature and identify new classes of examples that are very interesting from a model-building point of view. In some of these models dark matter is inevitably self-interacting, which is remarkable in view of recent observational support for this possibility. Hence, dark matter models featuring late kinetic decoupling have the potential not only to alleviate the missing satellites problem but also to address other problems of the cosmological concordance model on small scales, in particular the cusp-core and too-big-too-fail problems, in some cases without invoking any additional input.

DOI: 10.1103/PhysRevD.94.103529

I. INTRODUCTION

Dark matter (DM) is about five times as abundant as ordinary matter [1] and known to be the dominant driver of cosmological structure formation. The ΛCDM cosmological concordance model, which treats DM as a completely cold and collisionless component in the cosmic energy budget, is remarkably successful in describing the large-scale structure of the Universe [2,3]. At galactic scales and below, on the other hand, the observational situation is less clear and leaves considerable room for various new physics effects leading to deviations from the standard scenario. Several observations at such scales have even been claimed to be in tension with the expectations within the ΛCDM paradigm [4–16]; see Ref. [17] for a recent discussion. One of the most often discussed and most long-standing of these issues is the problem of "missing satellites" of the Milky Way [5,6], as compared to the typical number expected in ΛCDM cosmology, which subsequently was complemented by an observed underabundance also of small galaxies in the field [7,9,13].

According to the leading hypothesis, DM consists of a new type of elementary particles [18]. The most often studied class of models postulates weakly interacting massive particles (WIMPs) to form the DM and connects the observed DM abundance in a theoretically compelling way to extensions of the standard model of particle physics at energies beyond the electroweak scale. Standard WIMPs, like the supersymmetric neutralino [19] or the first Kaluza-Klein excitation of the photon [20], are prototype examples of cold dark matter (CDM) as required by the ΛCDM paradigm. Null searches for such WIMPs at the CERN LHC [21,22] or in direct detection experiments located deep underground [23,24], however, start to severely limit this possibility. Furthermore, from a theoretical perspective WIMPs are by far not the only possible option for a good DM candidate [25–28]. DM particles may instead have significantly stronger nongravitational interactions, either within a yet-to-be-explored dark sector or with ordinary standard model particles. This may visibly affect the distribution of the observed structure in the Universe [29–62]. Recently, a framework for an effective theory of structure formation (ETHOS) has been developed [63] that will eventually allow us to directly map the particle physics parameters in such models to cosmological

[*]torsten.bringmann@fys.uio.no
[†]h.t.ihle@astro.uio.no
[‡]joern.kersten@uib.no
[§]p.s.walia@fys.uio.no

 © 2016 American Physical Society

observables at low redshift. This is particularly relevant for those types of models that would evade any of the more traditional ways to search for DM at colliders, in direct or indirect detection experiments. In this case, detailed observations of the distribution of matter at small scales, for example in terms of the power spectrum, may be the only way to test the DM particle hypothesis.

One of the most prominent, potentially observable features of this type would be an exponential suppression of power in the spectrum of matter density fluctuations at sub-Mpc scales. The classical way to achieve this is through the free streaming of warm DM (WDM) particles, in which keV sterile neutrinos provide the prototype example for a well-motivated DM candidate of this type [28]. Such a cutoff in the power spectrum is strongly constrained by observations of the Lyman-α forest, typically translated into a lower bound on the WDM mass. Recent analyses report limits as stringent as $m_{\rm WDM} \gtrsim 4.35$ keV [64,65], which however has been argued to be overly restrictive when taking into account that the warm intergalactic medium could mimic such a cutoff [66]. Completely independent bounds of very roughly $m_{\rm WDM} \gtrsim 1$ keV arise from the observed phase-space densities of Milky Way satellites and from subhalo number counts in N-body simulations [67] as well as from weak lensing observations [68]. This range for the WDM mass, and hence the location of the cutoff, is interesting because already a value of $m_{\rm WDM} \sim 2$ keV would provide a solution to the missing satellites problem [69,70], with slightly larger values at least alleviating it. Historically, this was indeed one of the prime motivations to focus on WDM [31]. A drawback of WDM models in this mass range, however, is that they cannot address the other, at least as pressing, small-scale problems briefly mentioned in the beginning. In particular, the cuspy inner density profile of DM halos expected in CDM cosmology is not affected in any significant way [71,72], leaving the cusp-core problem [4,8,11] unexplained.

An alternative and much less explored way of creating a cutoff in the power spectrum at these scales arises if *cold* DM is kept in local thermal equilibrium with a relativistic species until the Universe has cooled down to sub-keV temperatures [33,34,37,43,59,73]. In this case, DM thus decouples kinetically much later than in the case of standard WIMPs [74]. The remaining viscous coupling between the two fluids then typically leads to a characteristic "dark" oscillation pattern in the power spectrum, with a strong suppression at small scales [75,76] as confirmed by explicit numerical simulations [17,77]. Interestingly, this is a possibility that arises rather naturally in self-interacting DM models, allowing us to address not only the missing satellites problem but at the same time also *all* other shortcomings of the ΛCDM paradigm [43]. This observation has already triggered significant interest and led to a number of specific model-building attempts [47,48,51,54,55,57,78] as well as the first fully self-consistent numerical simulations of structure formation for this class of models [17].

Here, we take a much broader perspective and aim at classifying, in a systematic way, the minimal possibilities that can lead to late kinetic decoupling with an observationally relevant cutoff in the power spectrum. Such a cutoff may or may not be related to a solution of the missing satellites problem, but would in any case provide a fascinating observational signature that helps to narrow down the identity of DM. We use the language of simplified models to describe the main ingredients that are necessary for any model building in this direction, depending on the spin of the CDM particle and its relativistic scattering partner. This relativistic particle may be some form of dark radiation (DR), the photon or one of the active neutrinos (though we will see that the first option is favored). We note that the existence of such a DR component is cosmologically very interesting in its own right [79,80] and can even be invoked to improve the consistency of different cosmological data sets [81–85]. As a result of our encompassing approach, we recover all previously identified configurations with scalars, fermions and vectors that lead to late kinetic decoupling, and also find further solutions that open new avenues for future model building. In our analysis, we fully include recent developments in the theoretical description of the decoupling process (see the discussion in the Appendix for more details).

This article is organized as follows. We start by discussing in Sec. II the generic requirements and limits for any DM model to feature sufficiently late kinetic decoupling such as to leave an observable imprint on the power spectrum or to alleviate the missing satellites problem. In Sec. III we restrict ourselves to simplified models containing only the CDM particle and its interaction with a relativistic particle, and provide a classification of all such models with the sought-after properties. We extend this classification in Sec. IV by allowing for a further, independent virtual particle mediating the interaction. We present a summary of our results and conclude in Sec. V. In two appendixes, we provide a concise review of the kinetic decoupling of DM particles from a thermal bath (Appendix A) and list the elastic scattering matrix elements, as well as the ETHOS parameters, for all models relevant for our discussion (Appendix B).

II. DARK MATTER SCATTERING WITH (DARK) RADIATION

As motivated in the Introduction, we are interested in scenarios where highly nonrelativistic DM can be kept in local thermal equilibrium with a relativistic species until late times, via the elastic scattering processes schematically shown in Fig. 1. The kinetic decoupling of DM from this radiation component (see Appendix A for details) then leads to a small-scale cutoff in the power spectrum of density fluctuations, corresponding to a minimal halo mass of [17]

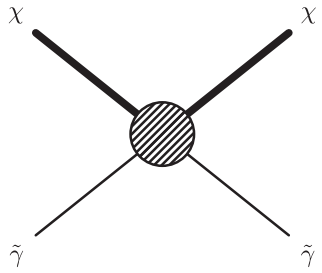

FIG. 1. Schematic illustration of the elastic scattering of a DM particle χ with a (possibly dark) relativistic particle $\tilde{\gamma}$. Throughout this article, we use thick lines to denote heavy (nonrelativistic) particles and thin lines to denotes light (relativistic) particles of any spin.

$$M_{\mathrm{cut,kd}} = 5 \times 10^{10} \left(\frac{T_{\mathrm{kd}}}{100\ \mathrm{eV}}\right)^{-3} h^{-1}\, M_{\odot}, \qquad (1)$$

where T_{kd} is the (photon) temperature at which decoupling occurs and $h \simeq 0.68$ [1] is the Hubble constant in units of 100 km s^{-1} Mpc^{-1} [note that this relation critically depends on how T_{kd} is defined; see the discussion after Eq. (A12)]. This should be compared to the corresponding cutoff in the halo mass function [17]

$$M_{\mathrm{cut,WDM}} = 10^{11} \left(\frac{m_{\mathrm{WDM}}}{\mathrm{keV}}\right)^{-4} h^{-1}\, M_{\odot} \qquad (2)$$

that is expected for a standard *warm* DM candidate (which decouples at temperatures much higher than keV). In order for such a cutoff to be observable for *cold* DM, and to potentially address the missing satellites problem, we thus need kinetic decoupling temperatures somewhat smaller than 1 keV, i.e. much smaller than the MeV to GeV temperatures one encounters for standard WIMPs [74].

Let us stress again that we focus here on situations where the dominant suppression mechanism arises from acoustic oscillations of a CDM component [75,76], and this is the assumption under which Eq. (1) is valid. Free streaming of the DM particles, which is the dominant effect for WDM, in principle leads to an independent suppression of the power spectrum [86]. Following Ref. [74], we estimate that this effect is subdominant for DM masses above 100 keV, for $T_{\mathrm{kd}} \sim 0.1$ keV and for a DR temperature equal to that of the photons, while for significantly colder DR free streaming becomes important only for even smaller DM masses. For simplicity, we restrict ourselves to DM particles that are heavy enough to lie outside this intermediate regime between CDM and WDM.

Our goal is to systematically classify all (minimal) possibilities that could give rise to such a late kinetic decoupling of CDM. To this end, we choose to be completely agnostic about the nature of DM and the radiation component, so the latter could either be a form of dark radiation (e.g. sterile neutrinos) or be given by the standard cosmological photon or neutrino background. We simply assume that there is one DM species, denoted by χ, and one radiation species scattering with χ, denoted by $\tilde{\gamma}$. We allow arbitrary spins for both species, and scrutinize all relevant simplified model Lagrangians (which obviously could be embedded in more complete frameworks) to see whether they allow for kinetic decoupling temperatures in the keV range or not.

A. Generic requirements for late kinetic decoupling

Before we start this endeavor, let us illustrate the general challenges for model building that we should expect to encounter. Consider for simplicity the case where the scattering amplitude close to kinetic decoupling can be approximated by a power law in the energy ω of the relativistic scattering partner

$$|\mathcal{M}|^2 \simeq c_n \eta_{\chi} (\omega/m_{\chi})^n, \qquad (3)$$

where the matrix element squared here and in the following is understood to be *summed* over the internal degrees of freedom (d.o.f.) η of *both* initial and final states. For later convenience, we have extracted the d.o.f. of the initial DM particle, η_{χ}, from the definition of c_n. For $n > -1$, we can analytically solve the Boltzmann equation to determine T_{kd} [see Eq. (A16) in Appendix A], plug the result into Eq. (1) and find

$$M_{\mathrm{cut}} \equiv M_{\mathrm{cut,kd}} \simeq M_n \xi^{3\frac{n+4}{n+2}} \left(\frac{c_n}{0.001}\right)^{\frac{3}{n+2}} \times \left(\frac{g_{\mathrm{eff}}}{3.36}\right)^{-\frac{3}{4+2n}} \left(\frac{m_{\chi}}{100\ \mathrm{GeV}}\right)^{-3\frac{n+3}{n+2}}. \qquad (4)$$

Here, we have introduced

$$\xi \equiv T_{\tilde{\gamma}}/T. \qquad (5)$$

$c_n \sim 10^{-3}$ very roughly corresponds to the case where the electroweak coupling mediates the scattering process, and g_{eff} is the usual effective number of relativistic degrees of freedom around kinetic decoupling. M_n is a numerical constant that is independent of the couplings or masses of the theory, and plotted in Fig. 2 as a function of n and in units of $M_{\odot}$ (assuming a fermionic $\tilde{\gamma}$; for a bosonic scattering partner, M_n would increase by an amount not visible at the resolution of the figure). For reference, we also indicate the cutoff mass induced by a 2 keV thermal WDM candidate; as discussed in the Introduction, this provides a rough distinction between what is ruled out by Ly-α data and what would help to alleviate the missing satellites problem.

Typical WIMP DM candidates are well described by the $n = 2$ case, for which we have $M_2 = 4.4 \times 10^{-7} M_{\odot}$; for 100 GeV neutralinos, for example, one finds roughly $10^{-7} M_{\odot} \lesssim M_{\mathrm{cut}} \lesssim 10^{-4} M_{\odot}$ [74]. Observable values of M_{cut}, close to what is still allowed by Ly-α data, thus clearly require a significant deviation from the standard scenario. According to Eq. (4), there exist only a handful of basic possibilities to increase M_{cut} in such a way. Let us briefly discuss them in turn.

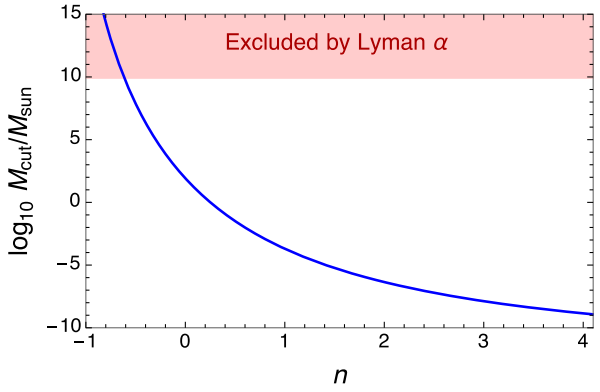

FIG. 2. Cutoff in the halo mass function resulting from DM scattering with a radiation component $\tilde{\gamma}$. This assumes $M_{\rm cut} = M_n$ as introduced in Eqs. (3)–(4), i.e. an amplitude that scales with the energy ω of $\tilde{\gamma}$ as $|\mathcal{M}|^2 \propto \omega^n$, a coupling strength roughly corresponding to the electroweak coupling and a DM mass of 100 GeV. For comparison, we also indicate the value that is roughly excluded by Ly-α data, with slightly smaller values allowing a potential solution to the missing satellites problem.

(i) *Maximizing the radiation temperature*. For DM scattering with photons or active neutrinos, we have by definition $\xi = 1$ and $\xi = (4/11)^{1/3} = 0.71$, respectively. If $\tilde{\gamma}$ constitutes a form of DR, on the other hand, ξ is in principle a free parameter.

Observations of the cosmic microwave background (CMB), however, exclude the existence of an additional radiation component corresponding to the contribution of one more massless neutrino (i.e. $\Delta N_{\rm eff} = 1$) by more than 3σ [1]. For a fermionic (bosonic) $\tilde{\gamma}$, this translates to $\xi < 0.85(0.82)/\eta_{\tilde{\gamma}}^{1/4}$, inevitably implying a further *suppression* of $M_{\rm cut}$ with respect to what is shown in Fig. 2.[1] Let us also mention that there exists an independent, weaker constraint on ξ from big bang nucleosynthesis [87]. Interestingly, this constraint actually *favors* a small DR component (unlike the one from CMB observations).

Note, finally, that there often exists a *lower* bound on the value of ξ that can be achieved in a given model-building framework. If $\tilde{\gamma}$ has been in thermal equilibrium with photons down to some temperature $T_{\rm eq}$, for example, and there was no additional entropy production in the visible sector afterwards, we have $\xi \gtrsim 0.34[g_{\rm eff}(T_{\rm eq})/100]^{-1/3}\eta_{\tilde{\gamma}}^{1/3}$.

(ii) *Minimizing the energy dependence of the scattering matrix*. It will obviously help if n is as small as possible. The simplest way to achieve a *constant* scattering amplitude ($n = 0$), in particular, is a contact interaction, i.e. an (effective) four-point vertex. We will study this option further in Sec. III A. In situations with propagators almost on shell (see the next point), the scattering rate can in extreme cases even *increase* with decreasing energy ω (corresponding to $n < 0$). This happens in particular when the DR particle appears in the t-channel, a scenario which we discuss in detail in Secs. III C and III D.

(iii) *Increasing the effective coupling strength*. In Fig. 3 we show which value of c_n in Eq. (3) is needed to produce $M_{\rm cut} = 10^{10} M_\odot$ (for $\xi = 0.5$ and $g_{\rm eff} = 3.36$). As can be seen, very large c_n, and hence efficient enhancement mechanisms for the amplitude, are needed for $n > 0$ and $m_\chi \gtrsim 1$ GeV. For large DM masses, the required amplitude even becomes so large that unitarity violation starts to become a possible concern; in critical cases, this needs to be checked on a model-by-model basis (see, e.g., Ref. [88]). Perturbativity restricts couplings to satisfy $\alpha \lesssim 1$, so the only option to achieve such large amplitudes is a virtual particle that is almost on shell in the particular kinematical situation we are interested in ($\omega \ll m_\chi$ and hence $-t \ll m_\chi^2$). This can be arranged both in the s/u-channel and in the t-channel. As these possibilities typically correspond to unrelated interaction terms in the Lagrangian, we will discuss them separately in Secs. III and IV.

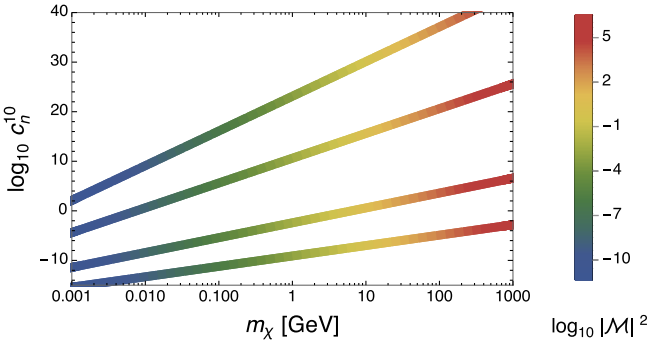

FIG. 3. For a scattering amplitude $|\mathcal{M}|^2 \simeq c_n(\omega/m_\chi)^n$ close to kinetic decoupling, this figure shows the value of c_n that results in $M_{\rm cut} = 10^{10} M_\odot$, as a function of the DM mass m_χ. From top to bottom, the lines correspond to $n = (4, 2, 0, -0.9)$, and we have throughout fixed $\xi = T_{\tilde{\gamma}}/T = 0.5$ and $g_{\rm eff} = 3.36$. The color scale indicates the size of the matrix element for $\omega \to \langle\omega\rangle_{T_{\rm kd}}$.

[1]The CMB bound can in principle be evaded if $\tilde{\gamma}$ becomes nonrelativistic right after kinetic decoupling, i.e. at $100\ {\rm eV} \gtrsim T \gtrsim 10\ {\rm eV}$. Assuming that there is no entropy production in the dark sector afterwards, however, this implies a warm (or even hot) DM density today of $\rho_{\tilde{\gamma}}^0 = (\zeta(3)/\pi^2) N_{\tilde{\gamma}} m_{\tilde{\gamma}} \eta_{\tilde{\gamma}} \xi^3 T_0^3$, where $N_{\tilde{\gamma}} = 1$ ($N_{\tilde{\gamma}} = 3/4$) for a bosonic (fermionic) $\tilde{\gamma}$. If we demand that this contribution make up at most a fraction f of the total observed DM density, we obtain a bound on ξ which turns out to be comparable to the CMB bound, $\xi < 0.8(f/N_{\tilde{\gamma}}\eta_{\tilde{\gamma}})^{1/3}(m_{\tilde{\gamma}}/10\ {\rm eV})^{-1/3}$. Even if $\tilde{\gamma}$ is kept in thermal equilibrium with another, relativistic species this conclusion does not change qualitatively; we discuss this case in Sec. III C.

(iv) *Decreasing the DM mass*. From the discussion so far, a DM mass significantly smaller than our reference value, 100 keV $\lesssim m_\chi \ll 100$ GeV, appears as maybe the most straightforward way to achieve a larger $M_{\rm cut}$. Indeed, DM particles much lighter than typical WIMP DM candidates are by no means a problem *per se*—though, as we will see, they might be disfavored in given model frameworks. Recall that the lower bound here simply results from the range of validity of Eq. (1); for lower DM masses free-streaming effects would have to be taken into account (unless $\xi \ll 1$), which is beyond the scope of this work. Note that *if* the DM relic density is set by chemical decoupling from any of the standard model particles, the lower bound tightens to $m_\chi \gtrsim 1$ MeV [89].

B. Bounds from inherently related processes

Models with a large scattering rate $\chi\tilde{\gamma} \leftrightarrow \chi\tilde{\gamma}$ typically imply large annihilation rates, $\chi\chi \to \tilde{\gamma}\tilde{\gamma}$, and in some cases significant DM self-interaction rates, $\chi\chi \to \chi\chi$, as well. In the following, we will discuss the *generic* constraints for model building that result from these processes (while we leave more model-specific considerations for later).

1. Dark matter annihilation $\chi\chi \to \tilde{\gamma}\tilde{\gamma}$

So far, we have avoided making any assumptions about how DM was produced in the first place. However, the fact that we consider interactions between χ and the thermally distributed $\tilde{\gamma}$ in order to achieve late kinetic decoupling indeed strongly suggests that DM was thermally produced in the early Universe. For cold DM, the relic density then typically scales roughly as $\Omega_\chi h^2 \propto m_\chi^2/g'^4$, with g' being the effective coupling to drive chemical decoupling. More concretely, if we assume $\mathcal{M}_{\chi\chi\to\tilde{\gamma}\tilde{\gamma}} \simeq g'^2$ at chemical freeze-out and only take into account the leading, velocity-independent part of the cross section, $\sigma v = g'^4/32\pi m_\chi^2$, the relic density is given by [90]

$$\Omega_\chi h^2 \simeq 8.81 \times 10^{-5} \frac{x_f}{g_{\rm eff}^{1/2}(T_{\rm cd})} g'^{-4} \left(\frac{m_\chi}{100\ \text{GeV}}\right)^2, \tag{6}$$

where v is the relative velocity between the two annihilating DM particles (if the leading contribution to the cross section in the zero-velocity limit is instead given by $\sigma v = g'^4/32\pi m_\chi^2 v^2$, this expression must be multiplied by $x_f/3$). Here, $x_f \equiv m_\chi/T_{\rm cd}$ and the effective number of degrees of freedom $g_{\rm eff}$ are evaluated at the temperature $T_{\rm cd}$ of *chemical* decoupling. A relic density in agreement with the observed value of $\Omega_\chi h^2 = 0.1188 \pm 0.0010$ [1] can thus not only be obtained for the specific combination of weak-scale masses $m_\chi \sim 100$ GeV and couplings $g'^2 \sim 0.04$, but also for any other combination of m_χ and α' that leaves the ratio of these quantities constant (this is sometimes referred to as WIMPless DM [91]).

Let us now denote with g the effective coupling for the scattering process, so we expect $c_n \sim g^4$ in Eq. (3). Rotating the corresponding diagrams, schematically shown in Fig. 1, we see that $\chi\chi \to \tilde{\gamma}\tilde{\gamma}$ must at least *contribute* to the total DM annihilation rate. Demanding that these annihilation processes do not deplete the DM abundance below the observed value implies a rough *upper bound* of $g^2 \lesssim g'^2 \sim 0.04(m_\chi/100\ \text{GeV})$ for $\xi \sim 1$ (note that this argument applies even if the initial DM abundance was *not* produced thermally). Using Eq. (4), this in turn restricts the cutoff mass approximately to

$$M_{\rm cut} \lesssim M_n \left(\frac{m_\chi}{100\ \text{GeV}}\right)^{-3\frac{n+1}{n+2}}. \tag{7}$$

Even when taking into account the impact on the DM abundance, considering lighter DM particles will thus in general help significantly to achieve a larger value of $M_{\rm cut}$.

Due to the different kinematics of scattering and annihilation processes, an intermediate particle nearly on shell in the former case is not on shell in the latter. This means that the value of the matrix element can be much larger for scattering than for annihilation, even though the "same" diagrams are involved. As can easily be checked, the argument of the preceding paragraph still runs through, in exactly the same way, when taking into account that c_n/g^4 may in fact be much larger than unity because of this effect—a possibility which we will make excessive use of. In this case, the right-hand side of Eq. (7) should be multiplied by a factor of $(c_n/g^4)^{3/(n+2)}$.

2. Dark matter self-scattering $\chi\chi \to \chi\chi$

Coupling DM to (dark) radiation inevitably implies that DM will be self-interacting, too. If $\tilde{\gamma}$ is bosonic and there is a direct coupling χ-χ-$\tilde{\gamma}$, for example, this will induce a Yukawa-type potential with strength $\alpha_\chi = g_\chi^2/4\pi$ and range $1/m_{\tilde{\gamma}}$ between the DM particles, resulting in a characteristic velocity-dependent self-scattering cross section.[2] We are typically interested in the classical limit ($m_\chi v \gg m_{\tilde{\gamma}}$), where the velocity-weighted transfer cross section peaks at around $v_{\rm max} \sim 0.1 g_\chi \sqrt{m_{\tilde{\gamma}}/m_\chi}$, resulting in $\sigma_T(v_{\rm max}) \sim 10/m_{\tilde{\gamma}}^2$; for larger velocities $v \gg v_{\rm max}$, the transfer cross section then drops sharply as $\sigma_T \propto v^{-4}$ (see e.g. Ref. [96]). Observations of dwarf-scale systems result in upper bounds on the self-interaction rate of roughly $\sigma_T/m_\chi \lesssim 10\ \text{cm}^2/\text{g} \simeq 4.6 \times 10^4/\text{GeV}^3$ [97–101], while constraints on cluster scales are up to two orders of magnitude more stringent [102,103]. Given that $v_{\rm max}$ is smaller than the typical velocities encountered in dwarf

[2]For pseudoscalar mediators, the situation is much more involved [92–95] and a full analysis is beyond the scope of this work.

galaxies, however, the former limit constrains the coupling g_χ much more severely in our case.

Following Ref. [63], we define $\langle\sigma_T\rangle_{30}$ as the transfer cross section σ_T averaged over a Maxwellian velocity distribution with a most probable velocity of $v_M = 30$ km/s, a value representative for dwarf galaxies. For $\sigma_T(v)$ we use the perturbative result [96] in the Born limit ($\alpha_\chi m_\chi \ll m_{\tilde\gamma}$), the parametrization obtained in Ref. [63] for the classical limit ($m_\chi v \gg m_{\tilde\gamma}$), and an analytical result outside these two regimes which results from approximating the Yukawa potential by a Hulthén potential [49]. In Fig. 4, we plot $\langle\sigma_T\rangle_{30}/m_\chi$ as a function of the coupling g_χ for various masses m_χ. We choose a reference value of $m_{\tilde\gamma} = 100$ eV, noting that the cross sections would become even larger for lighter $\tilde\gamma$ particles. In the figure, one can clearly identify the different regimes for σ_T (as well as the imperfect matching conditions, which are an artifact of our parametrization and bear no physical significance). For $\alpha_\chi \lesssim 10^{-7}(m_\chi/\mathrm{GeV})^{-1}$ we are thus in the Born regime (for the choice of $m_{\tilde\gamma}$ adopted in this figure), while for larger coupling we are in the classical regime (apart from very small DM masses, where the characteristic resonances from the intermediate regime start to appear). Besides the bound on the self-interaction rate mentioned above, we also include in the figure the generic upper bound on g_χ that results from DM annihilation processes $\chi\chi \to \tilde\gamma\,\tilde\gamma$; see the discussion in the previous subsection.

As one can clearly see, the self-interaction bounds are typically much stronger than those from the relic density. This implies in particular that the annihilation process $\chi\chi \to \tilde\gamma\,\tilde\gamma$ cannot be responsible for setting the correct DM density in the first place: in this case one would need coupling strengths close to the transition between solid and dashed lines in Fig. 4. For DM masses below the TeV scale, the self-interaction limits become in any case so severe that extremely tiny couplings are needed to evade them. In the classical regime, in particular, we find a simple scaling

$$\frac{\langle\sigma_T\rangle_{30}}{m_\chi} \sim 5.3\left(\frac{\alpha_\chi}{10^{-5}}\right)^{1.5}\left(\frac{m_\chi}{100\ \mathrm{GeV}}\right)^{-2.5}\left(\frac{m_{\tilde\gamma}}{\mathrm{keV}}\right)^{-0.5}\frac{\mathrm{cm}^2}{\mathrm{g}}, \tag{8}$$

which is valid for DM masses of

$$m_\chi \gtrsim \frac{m_{\tilde\gamma}}{100\ \mathrm{eV}}\max\left[1, \frac{1}{10^7\alpha_\chi}\right]\ \mathrm{GeV}. \tag{9}$$

For a direct coupling of DM to bosonic $\tilde\gamma$, the self-interaction bound thus makes it generically very hard to obtain as large cutoff masses as desired. Plugging the resulting constraint on g_χ into the expression for the cutoff mass, Eq. (4), we obtain the rough estimate of

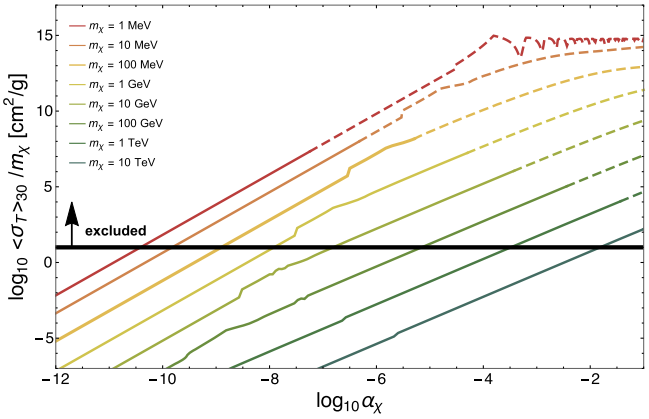

FIG. 4. DM self-interaction in dwarf-scale systems, induced by a direct coupling between light bosonic $\tilde\gamma$ and DM, as a function of $\alpha_\chi = g_\chi^2/4\pi$. From top to bottom, the curves show the case for a DM mass increasing from $m_\chi = 1$ MeV to $m_\chi = 10$ TeV. We have throughout assumed $m_{\tilde\gamma} = 100$ eV; lighter masses would give higher cross sections. The dashed parts of the curves indicate where the coupling strength g_χ is so large that $\chi\chi \to \tilde\gamma\,\tilde\gamma$ would deplete the DM abundance below the observed value. Values above the horizontal thick line are excluded by dwarf galaxy observations.

$$M_{\mathrm{cut}} \lesssim M_n\left(\frac{c_n}{10^6 g_\chi^4}\right)^{\frac{3}{n+2}}\left(\frac{m_\chi}{100\ \mathrm{GeV}}\right)^{\frac{1-3n}{n+2}}, \tag{10}$$

where we have conservatively assumed $m_{\tilde\gamma} = 1$ keV and M_n is shown in Fig. 2. This implies that, compared to the generic expectation of $c_n \sim g_\chi^4$, very large enhancements of the amplitude from almost on-shell virtual particles are necessary to achieve sufficiently large values of M_{cut}. Interestingly, it is also no longer favorable to consider small DM masses when taking into account the self-interaction bound.

Let us however stress that DM self-interactions do not only provide useful *constraints* on the type of models that we want to consider. Rather, as briefly mentioned in the Introduction, they have been invoked as *solutions* to a number of shortcomings of ΛCDM cosmology at small scales. In particular, it has been argued [96,104–106] that a velocity-dependent transfer cross section σ_T resulting from a Yukawa potential may successfully address both the *cusp-core* problem [4,8,11] and the *too-big-to-fail* problem [10,14] that appear at the scale of dwarf galaxies, without violating the stringent constraints at cluster scales. For this solution to work, self-interaction cross sections not more than one order of magnitude below the constraints on $\langle\sigma_T\rangle_{30}/m_\chi$ shown in Fig. 4 are needed. Another recent observation that has sparked significant interest in DM self-interactions is the cluster Abell 3827, where one of the member galaxies falling towards the center of the cluster appears to be displaced from its own gravitational well [16,107].

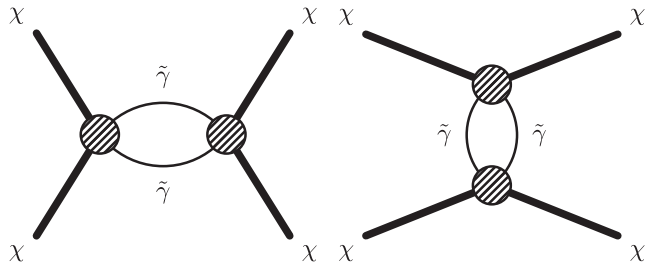

FIG. 5. *Minimal* effective DM self-interaction resulting from the χ–$\tilde{\gamma}$ interactions shown in Fig. 1.

Light fermions $\tilde{\gamma}$, on the other hand, obviously do not generate a Yukawa potential. A lower bound on the DM self-interaction rate can then be obtained by considering the diagrams displayed in Fig. 5, i.e. by forming a loop of the scattering process shown in Fig. 1. Assuming again that the shaded blobs in Fig. 5 are of the order of g^2, we roughly estimate $|\mathcal{M}|^2 \sim g^8/16\pi^2$. The transfer cross section will thus be of the order of $\sigma_T \sim 10^{-5} g^8/m_\chi^2$. Even when using the stronger constraint $\sigma_T/m_\chi \lesssim 0.1\ \mathrm{cm}^2/\mathrm{g}$ from cluster scales (as σ_T is to a good approximation velocity independent in this case), this only results in the essentially insignificant bound $g^2 \lesssim (m_\chi/\mathrm{MeV})^{3/4}$, which is always much weaker than the relic density bound considered above for the range of DM masses relevant for our discussion, $m_\chi \gtrsim 0.1$ MeV. We can in fact turn this argument around: if the effective couplings represented by the blobs in Fig. 5 were large enough to result in a significant self-interaction rate, this would imply a DM self-annihilation rate too large to be consistent with the DM abundance observed today. Stronger bounds from self-interactions can arise in specific models, however, as we will see in the following.

III. TWO-PARTICLE MODELS

We first consider models with the minimal possible particle content; i.e. we assume that there is no additional particle mediating the scattering process $\chi\tilde{\gamma} \to \chi\tilde{\gamma}$. This leaves three basic topologies that we will study in more detail in the following: (A) contact interactions, (B) s/u-channel mediated scattering processes and (C) dominantly t-channel mediated scattering processes. For each of these cases, we will discuss all possible spin combinations that potentially could lead to late kinetic decoupling; i.e. we allow in principle both χ and $\tilde{\gamma}$ to be a scalar (Dirac) fermion

$\tilde{\gamma} \setminus \chi$		**Scalar**			**Fermion**			**Vector**
	TOP	LKD	TP	σ_T	LKD	TP	σ_T	
Scalar	$4p$	$m_\chi \lesssim$ MeV	Yes	Constant	(only dim > 4)			$\langle\sigma_T\rangle_{30}$
	t	$m_{\tilde{\gamma}} \sim 1\,\mathrm{keV}$ $m_\chi \gtrsim 100\alpha_\chi^{3/5}$ TeV	$\langle\sigma_T\rangle_{30}$ (for $m_\chi \gtrsim 1$ MeV)	Yukawa	$m_{\tilde{\gamma}} \sim 1\,\mathrm{keV}$ $m_\chi \gtrsim 100\alpha_\chi^{3/5}$ TeV	$\langle\sigma_T\rangle_{30}$ (for $m_\chi \gtrsim 1$ MeV)	Yukawa	
	s/u	$\langle\sigma_T\rangle_{30}$			$\langle\sigma_T\rangle_{30}$			
Fermion		(only dim > 4 due to Z_2)			(only dim > 4)			Z_2
Vector	$4p$	(only dim > 4)			(only dim > 4)			Z_2
	s/u	$\langle\sigma_T\rangle_{30}$			$\langle\sigma_T\rangle_{30}$			
	$SU(N)$	$m_{\tilde{\gamma}} \sim 1\,\mathrm{keV}$ $m_\chi \gtrsim 10\alpha_\chi^{3/5}$ TeV	$\langle\sigma_T\rangle_{30}$ (for $m_\chi \gtrsim 1$ MeV)	Yukawa	$m_{\tilde{\gamma}} \sim 1\,\mathrm{keV}$ $m_\chi \gtrsim 10\alpha_\chi^{3/5}$ TeV	$\langle\sigma_T\rangle_{30}$ (for $m_\chi \gtrsim 1$ MeV)	Yukawa	(only broken $SU(M) \to SU(N)$)

FIG. 6. Overview of results for the two-particle models we have considered, where the DM particle χ and the (dark) radiation particle $\tilde{\gamma}$ can be scalars, Dirac fermions and vectors, respectively. Here, TOP denotes the topology of the (dominant) DM-DR scattering amplitude. Late kinetic decoupling (LKD) indicates whether in these types of models a small-scale cutoff as large as $M_{\mathrm{cut}} \sim 10^{10} M_\odot$ can be arranged. Thermal production (TP) indicates whether the observed DM density can be explained by thermal production via $\chi\chi \to \tilde{\gamma}\tilde{\gamma}$, and σ_T indicates the type of the DM self-interaction rate (only for viable models). A white cell indicates that LKD is possible and that models of this type additionally satisfy the indicated property (e.g. TP). Dark gray indicates that the model is either ruled out, or that it is not possible to achieve LKD, for the reason stated. Here, $\langle\sigma_T\rangle_{30}$ indicates the DM self-interaction strength at the scale of dwarf galaxy scales and Z_2 the assumed symmetry to stabilize DM. Operators with dimension (dim) larger than four map to the scalar/scalar four-point case if they lead to an approximately constant scattering amplitude; otherwise they are too small to lead to LKD. Note that t-channel scattering with vector DR is only possible in a non-Abelian gauge theory and hence is covered in the $SU(N)$ part.

or vector particle, respectively (for the sake of brevity, we will however not explicitly consider pseudoscalars and axial vectors in our analysis). For simplicity, and to avoid unphysical results, we will assume that a vector boson is always associated with a gauge symmetry—which may however be spontaneously broken to allow for $m_{\tilde{\gamma}} \neq 0$. We further require that DM be stabilized by a Z_2 symmetry; i.e. we do not allow for vertices with an odd number of χ particles. For a quick overview of our results for this type of models, we refer the reader to Fig. 6.

A. Pointlike interactions

Let us start with the case of a *single*, pointlike contact interaction, as depicted in Fig. 7. The simplest possibility to obtain this is with a dimension-4 operator. Due to gauge invariance, the only such operator that we need to study separately is in fact the case of a "portal interaction" between a scalar χ and a scalar $\tilde{\gamma}$, leading to a constant scattering amplitude. This is because a four-point coupling involving (broken or unbroken) gauge fields would imply the existence of further three-point couplings that unavoidably lead to additional diagrams of the form studied in the subsequent Secs. III B and III C.

Higher dimensional operators that lead to an (almost) constant scattering rate will have the same phenomenology, albeit with a suppressed amplitude, and hence do not have to be studied separately. Alternatively, a higher-dimensional operator containing derivatives or fermionic DR could add an energy dependence to the scattering rate. As any such operator is irrelevant in the language of effective field theory, i.e. suppressed at low energies, it will necessarily yield $n > 0$ in Eq. (3). Given that c_n is suppressed by a large mass scale, Fig. 3 then tells us that this possibility will not succeed in producing sufficiently large values of $M_{\rm cut}$.

The only pointlike interaction we have to consider in more detail at this point is thus a portal interaction of the form $\mathcal{L} \supset \frac{\lambda}{4}\chi^2\tilde{\gamma}^2$. This implies $|\mathcal{M}|^2 = \lambda^2 = \langle|\mathcal{M}|^2\rangle_t$ and, cf. Eqs. (3)–(4),

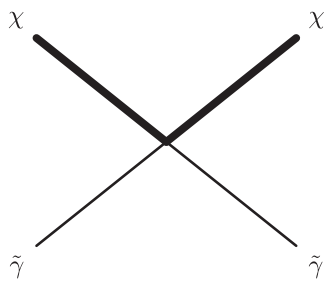

FIG. 7. Diagram illustrating a *pointlike interaction* of a DM particle χ with a (possibly dark) relativistic particle $\tilde{\gamma}$. Because we focus on unsuppressed interactions, only dimension-4 operators are considered, which restricts the analysis of this topology to bosonic particles.

$$M_{\rm cut}^{4S} \simeq 8.4 \times 10^{10} \xi^6 \lambda^3 \left(\frac{m_\chi}{10\ {\rm GeV}}\right)^{-9/2} M_\odot, \quad (11)$$

seemingly implying that a cutoff in the desired range can be obtained for any DM mass smaller than a few GeV.

As stressed in Sec. II B 1, however, an additional upper bound on λ results from the requirement that the DM annihilation rate $\chi\chi \to \tilde{\gamma}\tilde{\gamma}$ should not become so large that it would deplete the initial DM abundance (thermally produced or not) below the currently observed value. In Eq. (6) we should thus simply replace $g' \to \sqrt{\lambda}$, and require the resulting value for Ω_χ not to be smaller than the observed one. This leads to[3]

$$M_{\rm cut}^{4S} \lesssim 3 \times 10^{10} \xi^{15/2} \left(\frac{m_\chi}{\rm MeV}\right)^{-3/2} M_\odot, \quad (12)$$

where the maximal value for λ, and hence $M_{\rm cut}$, is achieved if DM is actually produced thermally (and this process is dominated by the same portal coupling between DM and DR). Given that $\xi \gtrsim 1$ is strongly constrained by CMB observations (see also footnote 1), DM in this simplest scenario must thus be lighter than about 1 MeV in order to produce a cutoff in an observationally interesting range. As discussed, free-streaming effects start to further increase the cutoff mass for $m_\chi \lesssim 0.1$ MeV (or even lighter DM masses if $\xi \ll 1$). The resulting additional suppression of structure implies that the same value of $M_{\rm cut}$ can be achieved for smaller values of ξ, which allows us to satisfy the strong CMB constraints on this quantity by an even larger margin (while $m_\chi \ll 0.1$ MeV would simply result in the standard WDM case). We leave a full exploration of this interesting regime for future work.

For this mass range, the DM annihilation bound becomes $\lambda \lesssim 7 \times 10^{-7} \xi^{0.5} m_\chi/{\rm MeV}$. Even though it is parametrically suppressed by a factor of λ^4, however, the induced DM self-coupling (see Fig. 5) for this model is actually log divergent. To be able to remove this divergence by renormalization, we thus must add an interaction term $\Delta\mathcal{L} = (\lambda'/4!)\chi^4$. Its finite part can thus be tuned to any desired value, independent of the above discussion, leading to a velocity-independent DM self-interaction cross section. Let us conclude the discussion of this case by remarking that the required small value of λ might most naturally be realized by a dim > 4 operator (as long as it leads to an approximately constant scattering rate, with

[3]We note that x_f depends logarithmically on the DM mass, and we used here the approximation given by Kolb and Turner [108]. Furthermore, we took into account the impact of $T_{\tilde{\gamma}} \neq T$ during freeze-out by assuming $x_f \propto \xi$ in Eq. (6). We checked this latter assumption explicitly by solving the full Boltzmann equation provided in Ref. [90], finding that the actual scaling is more accurately given by $x_f \propto \xi^r$, with $1.1 \lesssim r \lesssim 1.2$ (where r is larger for smaller values of m_χ and/or ξ). Note that we assume that ξ remains constant between chemical and kinetic decoupling.

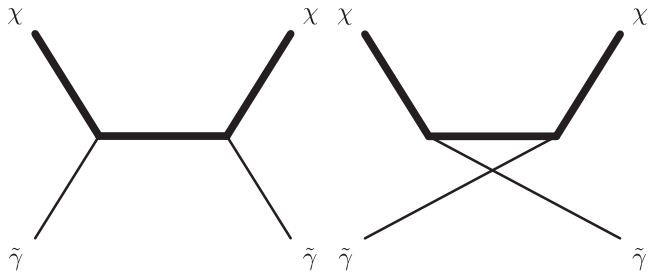

FIG. 8. As in Fig. 7, but in the presence of a χ-χ-$\tilde{\gamma}$ coupling, which leads to a resonance in the s- (left) and u-channel (right).

$n = 0$). See Refs. [37,51] for examples of large cutoff masses resulting from such effective operators (and sub-GeV DM masses).

B. Scattering exclusively via s/u-channel

Let us next consider situations where the scattering proceeds exclusively via the s-channel (and hence also the u-channel). The requirement to stabilize DM via a Z_2 symmetry then implies, as illustrated in Fig. 8, that the virtual particle must be χ (recall that by assumption no further particle beyond χ and $\tilde{\gamma}$ can be involved in our simplified two-particle model). This means that $\tilde{\gamma}$ must be bosonic, and we fully need to take into account the stringent constraints on DM self-interactions discussed in Sec. II B 2. Here, we only consider the situations that arise when $\tilde{\gamma}$ is a scalar or an Abelian gauge boson, and χ is either a scalar or a fermion. Otherwise—i.e. for non-Abelian DR or vector DM—there are necessarily four-point, s/u- and t-channel diagrams involved in the scattering process; we defer the treatment of these cases to Sec. III D. We calculate all relevant matrix elements in Appendix B, and list the results in Table I.

While both diagrams in Fig. 8 individually contain a resonance, those leading contributions cancel exactly in the $t \to 0$ limit in all cases. The result is an effective scattering amplitude that is to a very good approximation independent of the energy ω of the relativistic scattering partners. We thus obtain the *same* result as in the contact interaction case, Eq. (11), with the understanding that we should replace λ^2 by the corresponding expression for $\langle|\mathcal{M}|^2\rangle/\eta_\chi$ stated in Table I, where η_χ denotes the number of internal degrees of freedom of the DM particle. The essential difference, however, is that now we have a three-point coupling giving rise to a strong Yukawa potential between the DM particles. We can thus combine the result for the cutoff mass, Eq. (4), with the constraint $\langle\sigma_T\rangle_{30}/m_\chi \lesssim 10\ \mathrm{cm}^2/\mathrm{g}$ on the transfer cross section, where $\langle\sigma_T\rangle_{30}$ is supplied in Eq. (8). This results in

$$M_{\mathrm{cut}}^{s/u} \lesssim 2 \times 10^{-7} \xi^6 r^{\frac{3}{2}} \left(\frac{m_{\tilde{\gamma}}}{\mathrm{keV}}\right) \left(\frac{m_\chi}{100\ \mathrm{GeV}}\right)^{\frac{1}{2}} M_\odot. \quad (13)$$

Here, we have introduced $r \equiv \langle|\mathcal{M}|^2\rangle/(\eta_\chi g_\chi^4)$, where g_χ is the dimensionless coupling constant that enters the Yukawa potential—for the scalar/scalar (fermion/vector) case, e.g., we have $r = 1/2$ ($r = 16/3$).

Equation (13) clearly demonstrates that the strong constraints on DM self-interactions make it impossible to achieve late kinetic decoupling if the scattering is only mediated through s- and u-channel diagrams. We note that we arrived at this conclusion completely independently of the DM production mechanism. We *have* assumed in this argument, however, that the DM self-scattering takes place in the classical regime. For very light, (sub-)MeV DM [e.g. scalar DM scattering with hidden $U(1)$ vectors [38]] it may thus be possible to achieve large cutoff values and evade the self-scattering constraints.

C. Scattering dominantly via t-channel

Due to the Z_2 symmetry for the χ particles, any scattering diagram involving a t-channel exchange is of the form displayed in Fig. 9. Just as for the s/u case, this topology thus only allows scalar or non-Abelian $\tilde{\gamma}$. Here, we only consider the former case, deferring a dedicated discussion of the latter case to the next subsection. Such models have two independent coupling constants for the $\chi - \chi - \tilde{\gamma}$ and $\tilde{\gamma} - \tilde{\gamma} - \tilde{\gamma}$ vertices. The presence of the former induces s- and u-channel diagrams of the type discussed above. Here, we will thus require that those couplings be small enough to satisfy the self-interaction constraints of Fig. 4, cf. Eq. (8), and that the t-channel diagram dominate the scattering process. We note that this is indeed the generic situation, even for a $\tilde{\gamma} - \tilde{\gamma} - \tilde{\gamma}$ coupling much smaller than the $\chi - \chi - \tilde{\gamma}$ coupling, because of the strong kinematic enhancement of the t-channel diagram. We calculate the two relevant matrix elements, i.e. those for scalar and fermionic DM, in Appendix B and list the results in Table I.

As expected from the familiar Coulomb case, the scattering amplitude from the t-channel exchange of a massless particle diverges, and has to be regulated by introducing a nonvanishing DR mass term. In fact, such a mass term can be argued to arise from requiring the potential to be bounded from below: in our simplified model, this can only be achieved by adding a four-point

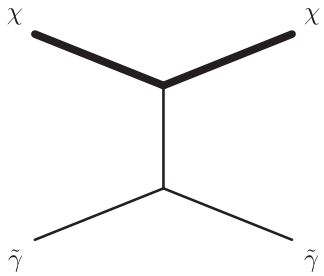

FIG. 9. As in Fig. 8, but in the presence of an additional 3-$\tilde{\gamma}$ coupling, which leads to a resonance in the t-channel *in addition* to the s/u resonances shown in Fig. 8.

interaction term $(\lambda/4!)\tilde{\gamma}^4$ to the scalar potential $V(\tilde{\gamma}) = (m_0^2/2)\tilde{\gamma}^2 + (\mu_{\tilde{\gamma}}/3!)\tilde{\gamma}^3$; considering the global minimum of this potential then leads to the conclusion that $m_{\tilde{\gamma}} \sim \max(m_0, \mu_{\tilde{\gamma}}/\sqrt{\lambda}) \gtrsim \mu_{\tilde{\gamma}}$, largely independent of the value of m_0. In general, Debye screening will furthermore generate a thermal mass of the order of $m_{\tilde{\gamma}}^{\rm Debye} \sim \sqrt{\lambda} T_{\tilde{\gamma}}$. At temperatures $T_{\tilde{\gamma}} \gg \mu_{\tilde{\gamma}}$, the combination of this effect and the requirement of vacuum stability thus even lead to $m_{\tilde{\gamma}} \gg \mu_{\tilde{\gamma}}$, essentially independent of the size of λ.

In the case of scalar DM, we have another dimensionful constant μ_χ which denotes the $\chi - \chi - \tilde{\gamma}$ coupling. Perturbativity and the absence of a global minimum in the scalar potential with $\chi \neq 0$, which would break the DM-stabilizing Z_2 symmetry, restrict μ_χ to be sufficiently smaller than m_χ. However, generically we still expect $\mu_\chi \gg m_{\tilde{\gamma}}$. Lastly, we would like to mention that for scalar DM and DR one will generally also have a portal interaction term in the Lagrangian as discussed in Sec. III A. Due to the strong kinematic enhancement of the t-channel diagram, however, this term will not have any significant effect unless $\mu_\chi \ll m_{\tilde{\gamma}}$.

In the limit where DR is highly relativistic, we find for fermionic DM an average scattering amplitude of

$$\langle |\mathcal{M}|^2 \rangle_t = \frac{g_\chi^2 \mu_{\tilde{\gamma}}^2 m_\chi^2}{\omega^4} \ln \frac{4\omega^2}{m_{\tilde{\gamma}}^2}. \tag{14}$$

For scalar DM we find the same expression after replacing the dimensionless DM-DM-DR coupling g_χ with the corresponding dimensionful coupling μ_χ as $g_\chi \to \mu_\chi/(\sqrt{8} m_\chi)$. We remind the reader that for such an energy dependence, the general analytic solution referred to in Eq. (4) is no longer valid. We can still immediately see that in this case the momentum transfer rate γ, cf. Eqs. (A7)–(A8), would fall with temperature *less* rapidly than the Hubble rate. In such a situation, DM and DR would initially *not* be in local thermal equilibrium. Once they enter it, however, they would not leave it anymore—leading to a depletion of structure on large scales that is unacceptable from an observational point of view (unless the couplings are chosen to be so small that thermal equilibrium would only be reached late during matter domination).

In view of this rather unexpected behavior, let us lift our general assumption of ultrarelativistic DR and investigate which effect an increased DR mass $m_{\tilde{\gamma}}$ would have on the cosmological behavior of this class of models. We will assume that $\tilde{\gamma}$ still follows a thermal distribution,[4] so we should expect that at some point the Boltzmann suppression of the $\tilde{\gamma}$ number density will dominate over the $T_{\tilde{\gamma}}^{-4}$ scaling from Eq. (14), leading to a suppression of the momentum transfer rate and hence kinetic decoupling relatively shortly after the DR has become nonrelativistic.

To investigate this in more detail, we solve the full Boltzmann Eq. (A13) numerically, noting that the solution close to kinetic decoupling, and *for a given value of* $T_{\rm kd}$, only depends on two parameter ratios, $m_{\tilde{\gamma}}/\xi$ and $g_\chi^2 \mu_{\tilde{\gamma}}^2/m_\chi$. We find that the former quantity is essentially fixed by the requirement to obtain a cutoff mass of $M_{\rm cut} = 10^{10} M_\odot$, varying only from $m_{\tilde{\gamma}}/\xi = 0.53$ keV for $g_\chi^2 \mu_{\tilde{\gamma}}^2/m_\chi = 10^{-20}$ keV to $m_{\tilde{\gamma}}/\xi = 2.0$ keV for $g_\chi^2 \mu_{\tilde{\gamma}}^2/m_\chi = 10^{-10}$ keV. For very small couplings leading to $g_\chi^2 \mu_{\tilde{\gamma}}^2/m_\chi \ll 10^{-20}$ keV, on the other hand, it is no longer possible to achieve a cutoff mass of $M_{\rm cut} = 10^{10} M_\odot$ because χ and $\tilde{\gamma}$ would not reach local thermal equilibrium early enough in the first place.

In Fig. 10, we show the DR mass that is required in this scenario to obtain a cutoff mass of $M_{\rm cut} \simeq 10^{10} M_\odot$, as a function of m_χ/α_χ and for several values of ξ (solid lines). We also show, for various values of m_χ, the constraints that arise from conservatively requiring that the DM self-interaction not become too strong, namely $\langle \sigma_T \rangle_{30} < 30\ {\rm cm}^2/g$; everything to the left of the dashed lines is thus excluded. For $m_\chi \gtrsim 1$ GeV, these constraints scale as expected for the classical regime, cf. Eq. (8), i.e. $(m_\chi/\alpha_\chi)_{\rm min} \propto m_\chi^{-2/3}$. Decreasing the DM mass below about 1 GeV, the limits do not tighten significantly anymore. As shown exemplarily for $m_\chi = 1$ MeV, they feature instead a much stronger dependence on $m_{\tilde{\gamma}}$ in this regime. This implies, as expected, that for DM masses even closer to the DR mass of order keV (required

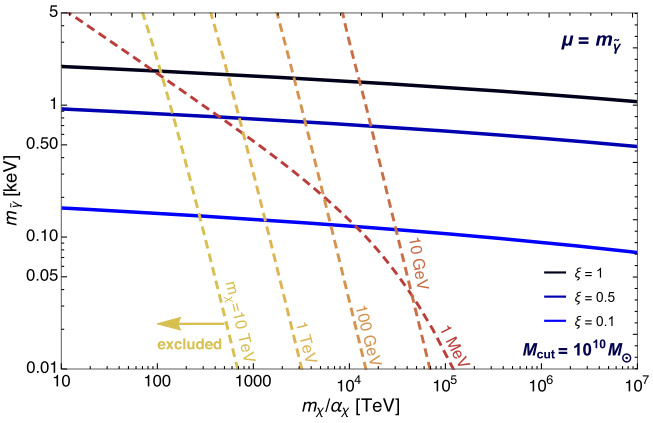

FIG. 10. Fermionic DM of mass m_χ scattering with scalar dark radiation of mass $m_{\tilde{\gamma}}$, with $\alpha_\chi \equiv g_\chi^2/(4\pi)$ and the cubic DR self-coupling fixed to $\mu = m_{\tilde{\gamma}}$. Solid lines show the parameter combinations that lead to $M_{\rm cut} \simeq 10^{10} M_\odot$, for a DR to photon temperature ratio of $\xi = 1, 0.5, 0.1$ (from top to bottom). Dashed lines give the constraints that result from DM self-scattering, for a given DM mass; everything left of the respective dashed curve is excluded.

[4]In contrast to the effectively massless case, this requires thermal equilibrium of $\tilde{\gamma}$ with at least one further relativistic species φ of temperature $T_{\tilde{\gamma}}$. Since we consider here by construction a situation in which $\tilde{\gamma}$ is nonrelativistic around kinetic decoupling of the DM particles, the CMB bound on the energy density of additional degrees of freedom (i.e. on ξ) thus becomes independent of $\eta_{\tilde{\gamma}}$ and only depends on η_φ.

for sufficiently late kinetic decoupling) those limits start to become *less* stringent again. In the plot, we have fixed $\mu = m_{\tilde{\gamma}}$. Smaller values will shift the solid lines to the left, by a factor of $m_{\tilde{\gamma}}^2/\mu^2$. As long as μ is still large enough to bring DM and DR into local thermal equilibrium, this has hardly any effect on the allowed range of parameters in this model.

Models with fermionic DM that couple to a keV-scale scalar with a cubic self-coupling thus allow us to have both large cutoff masses and DM self-interaction strengths relevant at the scale of dwarf galaxies, for a broad range of DM masses. For $m_\chi \gtrsim 1$ GeV, this is very roughly achieved for a coupling strength of $\alpha_\chi \sim 10^{-6}(m_\chi/10\ \mathrm{GeV})^{5/3}$, while smaller DM masses require a stronger coupling than expected from this simple scaling law. Interestingly, for DM masses smaller than around 1 MeV, the constraints on the DM self-interaction rate are no longer stronger than those from the DM annihilation rate (as would be the case for $m_{\tilde{\gamma}} \ll 1$ keV). This implies that in this setup one may in fact have thermally produced DM, with both M_{cut} and $\langle\sigma_T\rangle_{30}$ in a range that is interesting from the point of view of small-scale ΛCDM problems. Let us finally stress that the above discussion applies in full analogy to the case of scalar DM, with the already mentioned replacement $g_\chi \to \mu_\chi/(\sqrt{8}m_\chi)$.

D. DM-DR interactions through all channels

Lastly, we consider those cases where treating s/u- and t-channel (as well as four-point) diagrams separately is no longer possible because of gauge invariance. We first note that a vector DM particle is generally not allowed if fermions exist that are charged under the same gauge group, because the assumed Z_2 symmetry would be incompatible with covariant derivatives. In a dark sector with a minimal field content without fermions, on the other hand, the spontaneous breaking of a $U(1)$ symmetry necessarily leads to a massive vector that obeys a Z_2 symmetry and hence constitutes a very natural DM candidate, which has been discussed e.g. in the context of Higgs portal models [109]. In a similar fashion, breaking a non-Abelian group leads to *two* independent Z_2 symmetries and hence two different DM particles [110]—a situation which we will not study further because at this point we are only interested in scenarios with a single DM particle.

As shown in Appendix B, the scattering amplitude for Abelian vector DM and scalar DR is independent of the DR energy in the limit that we are considering, and thus leads to the same phenomenology as discussed in Sec. III B for interactions that proceed exclusively through s/u-channel exchange. This implies in particular that late kinetic decoupling cannot be achieved for Abelian vector DM because the required coupling strength is ruled out by the resulting strong DM self-interaction.

In the context of the two-particle models that we consider here, the only case that we have left out from our discussion so far is non-Abelian DR. The DM particle can then be either a scalar or a fermion, which leads to identical results for the scattering rates [up to a constant factor of order unity (see Table I) and a subdominant contribution from the four-point coupling in the scalar case].[5] We note that the case of fermionic DM scattering with non-Abelian DR has been previously studied in Ref. [60], where it was also pointed out that the necessarily small gauge couplings imply that confinement is irrelevant. DR can hence be described as a perfect fluid just like in all the other model types we study here.

Also in this case, a DR mass has to be introduced in order to regularize the scattering amplitude. Such a mass arises inevitably from screening effects in the thermal plasma and can be estimated as $m_{\tilde{\gamma}}^{\mathrm{Debye}} \sim g_\chi T_{\tilde{\gamma}}$ [112]. On top of this thermal mass, there can of course also be a temperature-independent mass if the gauge symmetry is spontaneously broken. In the limit where the DR is still ultrarelativistic, the squared scattering amplitude is in any case of the same form as Eq. (14), but with the leading ω^{-4} dependence replaced by a ω^{-2} dependence. Such a dependence implies that the momentum transfer rate scales as $\gamma \propto T_{\tilde{\gamma}}^2$, i.e. (for constant ξ) in the same way as the Hubble rate during radiation domination. During matter domination, on the other hand, γ will quickly fall behind $H \propto T^{3/2}$.

If the leading contribution to the DR mass is thermal, this can result in a very interesting phenomenology, where all density perturbation modes that enter the horizon before matter-radiation equality are suppressed in a smooth way (while those that enter after equality are essentially unaffected). Tuning the $SU(N)$ coupling strength to $\alpha_\chi \sim 10^{-9}$ (for $N = 2$ and $m_\chi \sim$ TeV), in particular, would help to alleviate a certain tension in the normalization of the power spectrum of density fluctuations as inferred from different types of observations [60,61]. Measurements of the CMB [1], in particular, predict a value of the observable σ_8 that is about 2σ larger than what is obtained from large-scale structure data [85,113]. Adopting the above parameter values, we find that the resulting DM self-interaction becomes $\langle\sigma_T\rangle_{30}/m_\chi \lesssim 1\ \mathrm{cm}^2/\mathrm{g}$ for $m_{\tilde{\gamma}} \gtrsim 10^{-8}$ eV, thus evading the observational constraints on this quantity (see the discussion in Sec. II B 2). We note that the necessarily small value of α_χ implies that the process $\chi\chi \to \tilde{\gamma}\tilde{\gamma}$ cannot be responsible for the thermal production of DM in this scenario.

Let us instead entertain the possibility, as in the preceding section, that the non-Abelian gauge bosons also have a

[5]It was only recently pointed out [111] that it is also possible to break a non-Abelian group *partially* in such a way that the gauge bosons of a residual non-Abelian subgroup would constitute DR, and DM would consist of vector particles stabilized by a Z_2 symmetry. The phenomenology of such a setup depends on the exact breaking pattern, and contains anyway more than one DM particle for the concrete situation considered in [111]. Hence, we do not further consider this possibility among the minimal scenarios we focus on here.

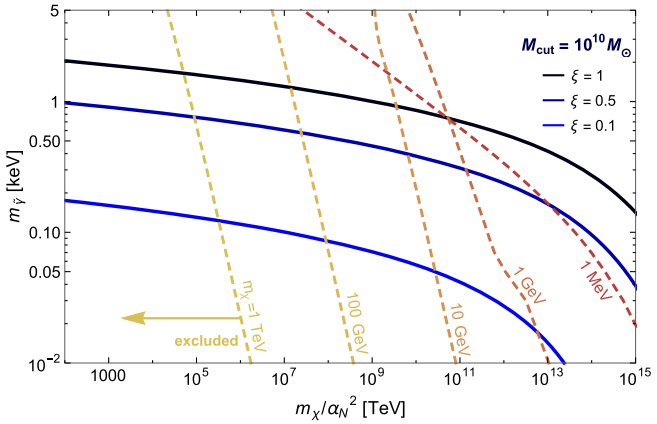

FIG. 11. Fermionic DM of mass m_χ scattering with non-Abelian DR of mass $m_{\tilde{\gamma}}$. For DR consisting of $SU(N)$ gauge bosons, we define $\alpha_N \equiv g_\chi^2\sqrt{(N^2-1)}/(4\pi)$. Solid lines show the parameter combinations that lead to $M_{\rm cut} = 10^{10} M_\odot$, for a DR to photon temperature ratio of $\xi = 1, 0.5, 0.1$ (from top to bottom). Dashed lines show the constraints that result from DM self-scattering, for $N = 2$ and a given DM mass; everything left of the respective curve is excluded. Larger values of N result in weaker constraints.

constant mass term which starts to dominate around keV temperatures. Requiring again that $\tilde{\gamma}$ be somehow kept in chemical equilibrium even after it becomes nonrelativistic, this would lead to the characteristic exponential cutoff in the power spectrum that is the main focus of this article. We thus solve the full Boltzmann Eq. (A13) numerically, requiring again that $M_{\rm cut} = 10^{10} M_\odot$. Similar to the t-channel case discussed above, this fixes the ratio $m_{\tilde{\gamma}}/\xi$ as a function of the ratio $\alpha_N^2 m_{\tilde{\gamma}}^2/m_\chi$, where we have defined $\alpha_N \equiv g_\chi^2\sqrt{(N^2-1)}/(4\pi)$. Again, the allowed value of $m_{\tilde{\gamma}}$ shows very little variation, from $m_{\tilde{\gamma}} = 2.2$ keV for $\alpha_N^2 m_{\tilde{\gamma}}^2/m_\chi = 10^{-10}$ keV to $m_{\tilde{\gamma}} = 0.18$ keV for $\alpha_N^2 m_{\tilde{\gamma}}^2/m_\chi = 10^{-25}$ keV. For $\alpha_N^2 m_{\tilde{\gamma}}^2/m_\chi \lll 10^{-25}$ keV, χ and $\tilde{\gamma}$ are not in local thermal equilibrium at high temperatures.

In analogy to the case of scalar DR in the t-channel discussed in the previous subsection, we plot in Fig. 11 the value of the dark gluon mass $m_{\tilde{\gamma}}$ that is needed for a cutoff mass $M_{\rm cut} \simeq 10^{10} M_\odot$, as a function of m_χ/α_N^2 and for several values of ξ (solid lines). We also show, as dashed lines, the constraints connected to DM self-interactions; here, we simply rescaled the available parametrizations for σ_T by the difference between $SU(N)$ mediators and $U(1)$ mediators expected at tree level.[6] For $m_\chi \gtrsim 10$ GeV, these constraints scale as expected for the classical regime, cf. Eq. (8), i.e. $(m_\chi/\alpha_\chi^2)_{\rm min} \propto m_\chi^{-7/3}$. For smaller DM masses, the limits weaken with respect to this scaling for the largest DR masses shown in the figure. Also in this case, this implies that there is a small region in parameter space where relatively light *thermal* DM, with $m_\chi \lesssim 1$ MeV, can produce an observable cutoff in the power spectrum *and* feature an observationally relevant, but not yet excluded self-interaction rate.

[6]Concretely, we find $|\mathcal{M}|^2_{SU(N)}/|\mathcal{M}|^2_{U(1)} = (N^2-1)/4 \equiv \alpha_\chi'^2/\alpha_\chi^2$, after *summing* over all colors, and then use $\sigma_{T,SU(N)}(\alpha_\chi) = \sigma_{T,U(1)}(\alpha_\chi')/N^2$ to account for the color average in the initial state. We stress that this prescription is only an approximation to the full higher-order σ_T, in the nonperturbative regime, but note that it reproduces the exact result in the Born regime.

IV. THREE-PARTICLE MODELS

We now extend our discussion to simplified models where the scattering between χ and $\tilde{\gamma}$ is mediated by a *different* particle. As before, we will require that DM be stabilized by a Z_2 symmetry; this time, however, we will allow for further, heavier particles to carry the same parity (which corresponds to the standard situation in typical scenarios with WIMP DM candidates, like supersymmetry or universal extra dimensions). This restricts the logical possibilities to the *same topologies* as considered in the previous section, i.e. scattering exclusively via a heavy particle χ' in the s/u-channel (as depicted in Fig. 8) or scattering via a light particle $\tilde{\gamma}'$ in the t-channel (as depicted in Fig. 9). For simplicity, we also do not explicitly study the possibility of vector DM (see Sec. III D for some general considerations concerning this option), and restrict the discussion to couplings described by dimension-4 operators (though we comment in Appendix B 2 on some opportunities that arise when lifting this assumption).

The fact that χ and χ' (as well as $\tilde{\gamma}$ and $\tilde{\gamma}'$) may differ in both spin and mass opens several new avenues for model building and the phenomenology of these models. Most strikingly, more combinations of particle spins are now possible (including fermionic $\tilde{\gamma}$) and the scattering can proceed *exclusively* through the t-channel. The new mass scale, furthermore, can help to avoid bounds on the self-interaction of DM, and qualitatively change the resonance structure of the s/u-channel diagrams. As before, we will discuss the two fundamental topologies separately, focusing on those models and aspects that result in a qualitative difference to the two-particle models.

A. Scattering via s/u-channel

Let us first consider models with a mediator particle χ' that is slightly heavier than χ and shares the same Z_2 parity. Defining $\Delta m \equiv m_{\chi'} - m_\chi$, we restrict our analysis to masses for which we have $m_\chi \gg \Delta m \gg \omega \gg m_{\tilde{\gamma}}$. Larger values of Δm would simply result in suppressed scattering rates; very small values, on the other hand, would typically involve serious fine-tuning in concrete models (and, furthermore, in many cases just lead to situations that are fully analogous to the s/u-channel two-particle models discussed in the previous section).

A complete list of the relevant models, along with results for the scattering matrix elements, is given in Table II. Note that this time there appear no vector particles in this classification. This is because non-Abelian gauge bosons are not compatible with the imposed Z_2 symmetry, for the topology considered here, while Abelian gauge bosons only couple to a pair of *identical* particles (appearing e.g. in the situation studied in Sec. III). To leading order, the squared amplitudes are all of the form

$$\langle|\mathcal{M}|^2\rangle_t = \frac{r\eta_\chi g_\chi^4}{\delta^2}\left(\frac{\omega}{m_\chi}\right)^n, \tag{15}$$

where $n=0$ for scalar DR, and $n=2$ if $\tilde{\gamma}$ is a fermion. Here, g_χ denotes the χ-χ'-$\tilde{\gamma}$ coupling (divided by m_χ in the one case it is dimensionful, namely when all particles are scalars); δ is given by $\delta \equiv \Delta m/m_\chi$; and r is a model-dependent constant with $1 \leq r \leq 16$. Defining $M_{10} \equiv M_{\rm cut}/10^{10}M_\odot$, we can use the analytic expression (4) for the cutoff mass and find in terms of the parameters introduced above[7]

$$M_{10}^{n=0} \simeq 8.4(6.8)\xi^6\left(\frac{rg_\chi^4}{\delta^2}\right)^{\frac{3}{2}}\left(\frac{m_\chi}{10\ \text{GeV}}\right)^{-\frac{9}{2}}, \tag{16}$$

$$M_{10}^{n=2} \simeq 7.9(7.7)\xi^{\frac{9}{2}}\left(\frac{rg_\chi^4}{\delta^2}\right)^{\frac{3}{4}}\left(\frac{m_\chi}{10\ \text{MeV}}\right)^{-\frac{15}{4}}. \tag{17}$$

The main constraint on this types of models typically results from the requirement that the pair-annihilation rate of χ should not be so large that it would deplete the number density of χ below the cosmological abundance of DM. Following the discussion in Sec. II B 1, we thus have to demand that $r^{1/4}g_\chi$ be smaller than the value of g' in Eq. (6) that is needed for $\Omega_\chi h^2 \simeq 0.119$.[8] Using furthermore $x_f \propto \xi$, this leads to the following upper bounds on the cutoff mass:

$$M_{10}^{n=0} \lesssim 0.9\xi^{\frac{15}{2}}\left(\frac{\delta}{0.01}\right)^{-3}\left(\frac{m_\chi}{10\ \text{GeV}}\right)^{-\frac{3}{2}}, \tag{18}$$

$$M_{10}^{n=2} \lesssim 4\xi^{\frac{21}{4}}\left(\frac{\delta}{0.01}\right)^{-\frac{3}{2}}\left(\frac{m_\chi}{100\ \text{keV}}\right)^{-\frac{9}{4}}. \tag{19}$$

We stress that while the actual bounds are model dependent, because the DM annihilation rate may be dominated by processes other than $\chi\chi \to \tilde{\gamma}\,\tilde{\gamma}$, the above expressions provide very useful order-of-magnitude estimates that allow us to classify in which models cutoff masses $M_{\rm cut} \sim \mathcal{O}(10^{10}M_\odot)$ can in principle be achieved.

For cases where the scattering rate is almost constant ($n=0$), such large cutoffs can relatively easily be obtained for DM masses up to around 10 GeV (or even larger DM masses if one is willing to accept a fine-tuning between m_χ and $m_{\chi'}$ beyond the percent level). We note that Eq. (16) reproduces as expected the result for a scalar four-point coupling that we earlier derived in Eq. (11), after replacing $r\eta_\chi g_\chi^4/\delta^2 \to \lambda^2$. The different conclusions about the maximal mass scale of the DM particles in these cases ($\sim$10 GeV vs $\sim$1 MeV) arise thus exclusively due to the on-shell enhancement resulting from $\delta \ll 1$. For an example similar to this type of model, where a fermionic DM particle interacts with a fermionic mediator and *pseudo-scalar* DR particles, see Ref. [54] (but note that in this case the model contains a further scalar t-channel mediator, as discussed in Sec. IV B).

For cases with $n=2$, i.e. for a fermionic $\tilde{\gamma}$, viable models in the above sense are restricted to a small range of sub-MeV DM masses (for an example of such a model, where fermionic DM couples to neutrinos via a scalar, see Ref. [78]). Similar to the situation discussed in Sec. III A, the mass range of interest extends to $m_\chi \lesssim 100$ keV, where free-streaming effects have to be taken into account. In this regime, ξ can hence be chosen small enough to avoid any tension with CMB data and yet suppress the power spectrum as desired.

It is very interesting to note that *all* models discussed in this section are in principle viable, if only for a relatively small range of DM masses and mediator particles that are highly degenerate in mass with the DM particles. None of these models, on the other hand, naturally gives rise to large DM self-interaction rates. Similar to the case of the simple scalar four-point interaction discussed above, those would have to be added by hand.

B. Scattering via t-channel

In this section, we consider models where we add a light bosonic particle $\tilde{\gamma}'$ to mediate the interaction between χ and $\tilde{\gamma}$ via a t-channel diagram. For simplicity we only consider models with the following hierarchy of energy scales: $m_\chi \gg m_{\tilde{\gamma}'} \gg \omega \gg m_{\tilde{\gamma}}$. This ensures that we are sufficiently far away from the situation discussed in the two-particle case, while retaining the possibility of large scattering rate enhancements through an almost on-shell mediator particle $\tilde{\gamma}'$.

We provide a complete list of the relevant models, as well as results for the scattering matrix elements, in Table II. For a dimensionful $\tilde{\gamma}$-$\tilde{\gamma}$-$\tilde{\gamma}'$ coupling $\mu_{\tilde{\gamma}}$, the scattering amplitudes are always constant, to leading order, and given by

[7]The leading number refers to a bosonic $\tilde{\gamma}$ and the one in parentheses to a fermionic $\tilde{\gamma}$.

[8]Here, we do not include the δ dependence in the comparison of the effective coupling constants because this derives from an on-shell enhancement that is absent for the annihilation process. Note also that in models where χ' is close in mass to χ, coannihilations [114] become important. This will *increase* the effective annihilation rate during freeze-out, hence leading to a *stronger* constraint on g. The actual limits on $M_{\rm cut}$ are thus slightly more stringent than stated in Eqs. (18)–(19)—apart from models with p-wave rather than s-wave annihilation, where the additional factor of $x_f/3$ in Eq. (6) has the opposite effect.

$$\langle|\mathcal{M}|^2\rangle_t = r\eta_\chi g_\chi^2 \left(\frac{\mu_{\tilde\gamma}}{m_{\tilde\gamma'}}\right)^2 \left(\frac{m_\chi}{m_{\tilde\gamma'}}\right)^2. \quad (20)$$

Otherwise, they take the form

$$\langle|\mathcal{M}|^2\rangle_t = r\eta_\chi g_\chi^2 g_{\tilde\gamma}^2 \left(\frac{m_\chi}{m_{\tilde\gamma'}}\right)^4 \left(\frac{\omega}{m_\chi}\right)^2. \quad (21)$$

Here, g_χ denotes the χ-χ-$\tilde\gamma'$ coupling (divided by m_χ in cases where it is dimensionful), and $g_{\tilde\gamma}$ denotes the $\tilde\gamma$-$\tilde\gamma$-$\tilde\gamma'$ coupling; r is a model-dependent constant in the range $1 \le r \le 128/3$. In all these cases, the form of the amplitude allows us to use the analytic expression (4) for the cutoff mass $M_{\rm cut}$. For the constant amplitude, Eq. (20), this leads to

$$M_{10}^{n=0} \simeq 8.4(6.8)\xi^6 r^{\frac{3}{2}} \left(\frac{g_\chi \mu_{\tilde\gamma}}{m_{\tilde\gamma'}}\right)^3 \left(\frac{m_{\tilde\gamma'}}{\rm GeV}\right)^{-3} \left(\frac{m_\chi}{\rm TeV}\right)^{-\frac{3}{2}}, \quad (22)$$

while for the $n = 2$ case, Eq. (21), the resulting cutoff mass becomes

$$M_{10}^{n=2} \simeq 1.4(1.4)\xi^{\frac{9}{2}} (r g_\chi^2 g_{\tilde\gamma}^2)^{\frac{3}{4}} \left(\frac{m_{\tilde\gamma'}}{\rm MeV}\right)^{-3} \left(\frac{m_\chi}{\rm TeV}\right)^{-\frac{3}{4}}. \quad (23)$$

An important phenomenological difference of these models, as compared to the three-particle models in the s/u-channel, is that the light mediator $\tilde\gamma'$ will mediate a significant velocity-dependent DM self-interaction. Because we now have the freedom to choose $m_{\tilde\gamma'} \gg m_{\tilde\gamma}$, the DM self-interaction rate can be much more easily arranged to be in an observationally relevant range (e.g. such as to mitigate the ΛCDM small-scale problems). In fact, this can be done while at the same time allowing for thermally produced DM. In Fig. 12 we illustrate this point by plotting the value of $\alpha_\chi = g_\chi^2/4\pi$ as a function of m_χ that is required to obtain $\langle\sigma_T\rangle_{30}/m_\chi = 1\ {\rm cm}^2/{\rm g}$, for various values of $m_{\tilde\gamma'}$. For large DM masses, $m_\chi \gtrsim 100$ GeV in the plot, we are in the classical regime for σ_T; for DM masses below about 10 GeV and small values of α_χ, we are instead in the Born regime. The small jumps that are visible in between are not physical but result from the fact that the parametrizations that we adopt here do not connect the various regimes described in Sec. II B 2 in a perfectly smooth way. Once the ratio of mediator to DM mass becomes large enough, strong resonances develop in σ_T. As indicated by shaded areas in Fig. 12, this allows multiple solutions to $\langle\sigma_T\rangle_{30}/m_\chi = 1\ {\rm cm}^2/{\rm g}$. Here, the steps in the upper envelopes of these shaded areas reflect the number of resonances where this condition can be met.

In the same figure we show, for comparison, the value of α_χ that follows from Eq. (6) when assuming that the process $\chi\chi \to \tilde\gamma'\tilde\gamma'$ proceeds with a rate of $\sigma v = \pi\alpha_\chi^2/2m_\chi^2$ and is fully responsible for setting the correct relic density (labeled "s-wave"). We also show the case of

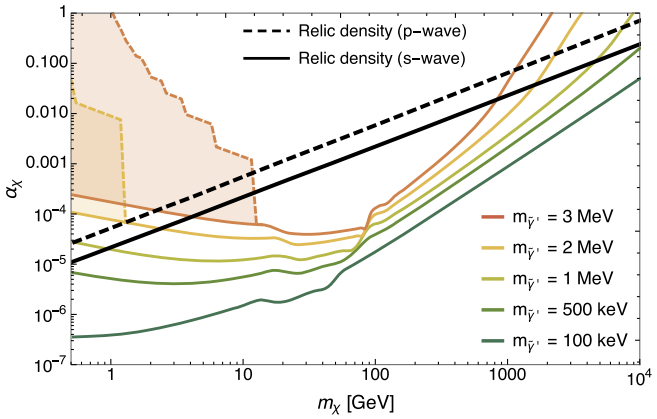

FIG. 12. Combinations of α_χ and m_χ that lead to $\langle\sigma_T\rangle_{30}/m_\chi = 1\ {\rm cm}^2/{\rm g}$, for various mediator masses $m_{\tilde\gamma'}$. The shaded areas show the range of parameters where the appearance of resonances in σ_T allows multiple solutions to this condition. For comparison, we also show the value of α_χ that results in the correct relic density when considering only the process $\chi\chi \to \tilde\gamma'\tilde\gamma'$, assuming that $\xi = 0.5$ (see text for further details).

$\sigma v = \pi\alpha_\chi^2 v^2/2m_\chi^2$ (labeled "p-wave"). We note that a more accurate treatment would depend on the concrete model. For a fermionic DM particle χ annihilating to a vector $\tilde\gamma'$ (s-wave) or scalar $\tilde\gamma'$ (p-wave), for example, the actual annihilation rate at lowest order is larger by a factor of 2 and 3/2, respectively (this corresponds to models (B23) to (B27) in Appendix B). Accordingly, the s-wave (p-wave) line would move downwards by about 30% (20%). In general, the annihilation rate also receives an enhancement due to the Sommerfeld effect. Compared to what is shown in the figure, this will result in a slightly smaller value of α_χ that is necessary to achieve the correct relic density; even in the vicinity of resonances, however, this is only an $\mathcal{O}(1)$ effect [115] which again would hardly be visible at the resolution given here.

Even though the details are somewhat model dependent, Fig. 12 clearly illustrates that for mediator masses $m_{\tilde\gamma'} \gtrsim 1$ MeV and DM masses in the TeV range it is in general possible to accommodate thermal DM production and a DM self-interaction rate that is sufficiently large to visibly affect the inner structure of subhalos at the scale of dwarf galaxies. Let us now investigate the consequences for the cutoff in the power spectrum in this regime, assuming again that the above s-wave annihilation cross section is responsible for setting the relic density. This fixes g_χ in Eqs. (22)–(23), which thus become

$$M_{10}^{n=0} \simeq \xi^{\frac{27}{4}} r^{\frac{3}{2}} \left(\frac{\mu_{\tilde\gamma}}{m_{\tilde\gamma'}}\right)^3 \left(\frac{m_{\tilde\gamma'}}{\rm GeV}\right)^{-3} \left(\frac{x_f}{g_{\rm eff}(T_{\rm cd})}\right)^{\frac{3}{4}}, \quad (24)$$

$$M_{10}^{n=2} \simeq \frac{1}{2}\xi^{\frac{39}{8}} r^{\frac{3}{4}} g_{\tilde\gamma}^{\frac{3}{2}} \left(\frac{m_{\tilde\gamma'}}{\rm MeV}\right)^{-3} \left(\frac{x_f}{g_{\rm eff}(T_{\rm cd})}\right)^{\frac{3}{8}}. \quad (25)$$

Note that now there is only a very weak dependence of the cutoff on the DM mass, through x_f and $g_{\rm eff}$ (as well as the Sommerfeld effect; see [43,47] for examples). For MeV mediators and couplings $\alpha_{\tilde{\gamma}} \sim \alpha_\chi \sim 10^{-2}$, thermally produced DM can thus lead to $M_{10} \sim 1$ in the $n = 2$ case, implying in particular that for $m_\chi \sim$ TeV a simultaneous solution of *all* small-scale problems of ΛCDM is possible. A similar phenomenology is obtained for $n = 0$, i.e. for models with dimensionful $\tilde{\gamma}$-$\tilde{\gamma}$-$\tilde{\gamma}'$ couplings, if one adopts $\mu_{\tilde{\gamma}} \sim 10^{-3} m_{\tilde{\gamma}}$. Indeed, the $n = 2$ case corresponds exactly to the situation first described in Ref. [43], and is followed up by several concrete examples for model building with vector mediators and fermionic DM and fermionic DR [47,48,55,57] as well as scalar DR [54]. As one can see from this discussion, however, there exists a rather large variety of models that fall into this class, including the possibility of scalar mediators. The possibilities for future model building that we have pointed out here thus go clearly beyond the specific settings considered so far.

The very large values of the tree-level scattering amplitude we need for TeV-scale DM particles (see also Fig. 3) may lead to worries about the reliability of the calculation, since higher-order corrections could be important. Therefore, we calculated the full one-loop correction arising from the exchange of one additional scalar mediator for fermionic DM and scalar DR using `LoopTools` [116]. It turned out that this correction can safely be neglected. This can be traced back to the fact that the DR particle in the loop is highly virtual since the change of four-momentum upon entering the loop is of order $\sqrt{|t|} \sim T_{\rm kd} \gg m_{\tilde{\gamma}}$. The kinematical situation is thus different from the scattering or annihilation of nonrelativistic particles, which remain nearly on shell when exchanging one or more light mediators and thus experience Sommerfeld enhancement [117–119].

To conclude this section, let us point out that there is yet another class of thermally produced DM models, visible in the *low-mass* part of Fig. 12, where the self-interaction rate is at the right level to potentially address the cusp-core or the too-big-to-fail problem. In contrast to the class of solutions discussed in the previous paragraph, here the transfer cross section σ_T is either in the Born or the resonant regime. Given that the cutoff is almost independent of m_χ for all thermally produced models considered in this section, however, the same conditions on $m_{\tilde{\gamma}}$ and $g_{\tilde{\gamma}}$ (or $\mu_{\tilde{\gamma}}$) as just discussed above will lead to $M_{10} \sim 1$, though of course the different values of α_χ and m_χ will lead to different requirements for concrete model building. We have thus identified a whole new class of GeV DM models that could potentially address all ΛCDM small-scale problems simultaneously. We leave a more detailed investigation of the expected rich phenomenology as an interesting direction for future work.

V. CONCLUSIONS

If cold DM is kept in local thermal equilibrium with a relativistic species (dark radiation) until the Universe has cooled down to temperatures below ∼1 keV, this results in a characteristic suppression of the power spectrum of matter density fluctuations for scales below what corresponds roughly to the size of the smallest dwarf galaxies. Such a cutoff may help to alleviate the problem of missing satellites in the cosmological concordance model. More importantly, it provides quite in general a fascinating way of probing new particle physics in the dark sector by using astrophysical observables connected to the distribution of cosmological structure. This type of probe is thus highly complementary to traditional attempts to identify the particle nature of DM.

In this article, we have provided a systematic classification of the minimal model-building options that allow for such a scenario. The simplest solution turns out to be a contact interaction between a DM particle with $m_\chi \lesssim$ 1 MeV and a relativistic DR particle, either in the form of a four-point portal interaction between two scalars or via a suppressed, higher-dimensional operator (Sec. III A). Scenarios where DM couples via a three-point coupling to DR, on the other hand, are severely constrained by observational bounds on the strength of DM self-interactions, leaving no room for a sufficiently late kinetic decoupling (Sec. III B). This problem may be circumvented by allowing for a mediator particle that is slightly heavier than DM (Sec. IV A) or lighter than DM but significantly heavier than DR (Sec. IV B). In the first class of models, DM cannot be too heavy [typically $m_\chi \lesssim \mathcal{O}(10\ {\rm GeV})$]; in the second class, the possibility to get an observable cutoff for thermally produced DM turns out to be almost independent of the DM mass. In Appendix B, we provide the corresponding ETHOS [17,63] parameters of our simplified particle physics models; similar parameters will result in almost identical results when performing full numerical simulations of structure formation for such DM candidates.

Within the classes of models considered, we do not find examples where photons could play the role of the dark radiation component to achieve sufficiently late kinetic decoupling. As discussed in Appendix B 2, this may change to some extent if higher-dimensional operators are included in the discussion of possible interactions between the mediator and the radiation component. The fact that the left-handed leptons of the standard model are contained in $SU(2)$ doublets makes it furthermore challenging to construct models where late kinetic decoupling can be achieved with (active) neutrinos; see again Appendix B 2 for a discussion.

The main phenomenological difference to WDM scenarios, which lead to a similar cutoff in the power spectrum, is that in particular the class of models featuring t-channel mediators much lighter than DM (and much heavier than

DR) naturally gives rise to relatively large DM self-interaction rates. For a few concrete models with TeV-scale DM particles and MeV-scale mediators, it has been noticed before that this fact can be used to simultaneously alleviate *all* small-scale problems of ΛCDM cosmology for thermally produced DM. We have not only demonstrated that the models studied so far fall into a much broader class of viable solutions with this property, but also identified a new class of GeV-scale DM models with similar properties (Sec. IV B). This opens promising and largely unexplored model-building avenues.

We have furthermore shown that the cubic self-interaction of a scalar DR particle makes the DM-DR interaction increasingly efficient for small energies (Sec. III C). In this case, DM and DR would inevitably be in local thermal equilibrium at late times, though not necessarily at early times. If the dark radiation particles are instead massive, with $m_{\tilde{\gamma}} \sim 1$ keV, they will however decouple around the same time as in the other cases discussed here. For (sub-)MeV DM, such a scenario would in fact also allow for thermally produced DM with self-interaction rates in the observationally relevant rate (similar to the case of non-Abelian dark radiation; see Sec. III D). We leave a more detailed investigation of this interesting observation for future work.

ACKNOWLEDGMENTS

We would like to thank Tobias Binder, Thomas Hahn, Jasper Hasenkamp, Andrzej Hryczuk, Jörg Jäckel, Felix Kahlhoefer, Oleg Lebedev and Kai Schmidt-Hoberg for very fruitful discussions.

Note added.—Recently, two studies appeared on the arXiv that independently identified some of the new scenarios for very late kinetic decoupling that we have described and classified here. In particular, Binder *et al.* [120] pointed out that the fermion/scalar/fermion combination shown as entry (B24) in Table II provides such a new solution, while Tang [121] identified interacting scalar DR and fermionic DM as a further possibility (see the fermion/scalar case in Table I, and the discussion in Sec. III C).

APPENDIX A: KINETIC DECOUPLING

The kinetic decoupling of DM particles from a thermal bath can be described from first principles by solving the underlying Boltzmann equation [76,122]—just as the standard way of calculating the relic density of thermally produced DM particles [90] is based on solving the Boltzmann equation during an earlier epoch of chemical decoupling. The original formalism [122] was later extended to nonrelativistic scattering partners [74] that may have a temperature differing from that of photons [43,47], situations in which the DM number density or the effective number of relativistic degrees of freedom can change during or after decoupling [42,74] and, most recently, to the case where the scattering amplitude is not Taylor expandable around small momentum transfer [123,124]. Here, we provide a brief summary taking into account these more recent developments.

Consider a particle $\tilde{\gamma}$ with a thermal distribution $g^{\pm}$ of temperature $T_{\tilde{\gamma}}$, and a nonrelativistic DM particle χ that can interact with $\tilde{\gamma}$. The Boltzmann equation that governs the evolution of the DM phase-space distribution f in an expanding Friedmann-Robertson-Walker universe is then given by $L[f] = C[f]$, with the Liouville operator[9]

$$L[f] = E(\partial_t - H\mathbf{p}\cdot\partial_{\mathbf{p}})f(\mathbf{p}), \tag{A1}$$

and a collision term

$$\begin{aligned} C[f] = \frac{1}{2\eta_\chi}\int\frac{d^3k}{(2\pi)^3 2\omega}\int\frac{d^3\tilde{k}}{(2\pi)^3 2\tilde{\omega}}\int\frac{d^3\tilde{p}}{(2\pi)^3 2\tilde{E}} \\ \times (2\pi)^4\delta^{(4)}(\tilde{p}+\tilde{k}-p-k)|\mathcal{M}|^2_{\chi\tilde{\gamma}\leftrightarrow\chi\tilde{\gamma}} \\ \times [(1\mp g^{\pm})(\omega)g^{\pm}(\tilde{\omega})f(\tilde{\mathbf{p}}) \\ - (1\mp g^{\pm})(\tilde{\omega})g^{\pm}(\omega)f(\mathbf{p})]. \end{aligned} \tag{A2}$$

Here, $H = \dot{a}/a$ is the Hubble parameter, and $(E, \mathbf{p})$ and $(\omega, \mathbf{k})$ are the 4-momenta of the incoming particles χ and $\tilde{\gamma}$, respectively (outgoing momenta are denoted with a tilde). The scattering amplitude $|\mathcal{M}|^2$ is *summed* over all internal degrees of freedom, and the phase-space densities are normalized such that, e.g., the number density of the particle χ is given by $n_\chi = \eta_\chi \int d^3p f(\mathbf{p})/(2\pi)^3$.

Even when the DM particle is no longer in local thermal equilibrium, one can now *define* a parameter

$$T_\chi \equiv \frac{\eta_\chi}{3m_\chi n_\chi}\int\frac{d^3p}{(2\pi)^3}\mathbf{p}^2 f(\mathbf{p}). \tag{A3}$$

Introducing further the dimensionless parameters

$$x \equiv m_\chi/T, \tag{A4}$$

$$y \equiv m_\chi T_\chi s^{-2/3}, \tag{A5}$$

the second moment of the full Boltzmann equation, keeping only leading terms in $\mathbf{p}^2/m_\chi^2$, reduces to[10]

[9] All momenta that appear in these expressions are physical, as opposed to comoving, and the time dependence of f is understood to arise, to leading order, exclusively from the expansion of the Universe via $\mathbf{p} \propto 1/a$ (where a is the scale factor). Note further that we are throughout using conventions for the normalization of quantum fields and their interactions that are consistent with those of Peskin and Schroeder [125].

[10] Note that this assumes a constant comoving DM particle number density during and after kinetic decoupling. Otherwise an additional term must be added to this equation that couples the evolution of T_χ to that of n_χ [42]. This can be relevant, e.g., in the presence of resonances or a strong Sommerfeld enhancement of the DM annihilation rate.

$$\frac{d\log y}{d\log x} = \left(1 - \frac{1}{3}\frac{d\log g_{*S}}{d\log x}\right)\frac{\gamma(T_{\tilde{\gamma}})}{H(T)}\left(\frac{y_{\rm eq}}{y} - 1\right). \quad \text{(A6)}$$

Here, s is the total entropy density, g_{*S} are the effective entropy degrees of freedom of *all* relativistic particles in the Universe and $y_{\rm eq}$ is given by Eq. (A5) with $T_\chi \to T_{\tilde{\gamma}}$. The momentum transfer rate γ, finally, is given by

$$\gamma(T_{\tilde{\gamma}}) = \frac{1}{48\pi^3\eta_\chi T_{\tilde{\gamma}} m_\chi^3} \times \int d\omega k^4 (1 \mp g^\pm) g^\pm(\omega) |\mathcal{M}|^2_{\substack{t=0 \\ s=m_\chi^2+2m_\chi\omega+m_{\tilde{\gamma}}^2}}, \quad \text{(A7)}$$

where $k \equiv |\mathbf{k}|$. The above expression only holds if $|\mathcal{M}|^2$ is Taylor expandable around $t = 0$ [in the sense that $|\mathcal{M}|^2 = |\mathcal{M}|^2_{t=0}[1 + \mathcal{O}(\omega^2/m_\chi^2)]$, taking into account that t is of the same order as ω^2]. While this is typically a good assumption, it fails for example if the denominator is suppressed by ω or t (because the propagator is almost on shell). In such situations, we have to make the replacement [123]

$$|\mathcal{M}|^2_{\substack{t=0 \\ s=m_\chi^2+2m_\chi\omega+m_{\tilde{\gamma}}^2}} \to \langle|\mathcal{M}|^2\rangle_t \equiv \frac{1}{8k^4}\int_{-4k^2}^0 dt(-t)|\mathcal{M}|^2. \quad \text{(A8)}$$

This allows us to rewrite $\gamma(T_{\tilde{\gamma}})$ in terms of the total transfer cross section, $\sigma_T \equiv \int d\Omega(1 - \cos\theta)d\sigma/d\Omega$, as

$$\gamma(T_{\tilde{\gamma}}) = \frac{1}{3\pi^2\eta_\chi m_\chi}\int d\omega g^\pm(\omega)\partial_\omega(k^4\sigma_T), \quad \text{(A9)}$$

where we have used that $g^\pm(1 \mp g^\pm)(\omega) = -T_{\tilde{\gamma}}\partial_\omega g^\pm(\omega)$.

The solution to Eq. (A6) before and after DM leaves local thermal equilibrium with the $\tilde{\gamma}$ particles is of the form

$$T_\chi(T) = \begin{cases} T_{\tilde{\gamma}} & \text{for } T \gtrsim T_{\rm kd} \\ \mathcal{C}/a^2 & \text{for } T \lesssim T_{\rm kd} \end{cases} \quad \text{(A10)}$$

with a constant $\mathcal{C}$ that is uniquely determined by the solution of the differential equation. Given that the transition between the two regimes typically happens rather fast, it is natural to *define* the kinetic decoupling temperature as the point where the two asymptotics meet (see Fig. 1 in Ref. [74]). This is equivalent to rewriting $\mathcal{C}$, and hence Eq. (A10), as

$$T_\chi(T) = \begin{cases} T_{\tilde{\gamma}}(T) & \text{for } T \gtrsim T_{\rm kd} \\ \xi T_{\rm kd}[a(T_{\rm kd})/a(T)]^2 & \text{for } T \lesssim T_{\rm kd} \end{cases} \quad \text{(A11)}$$

where we have introduced

$$\xi \equiv T_{\tilde{\gamma}}/T. \quad \text{(A12)}$$

Other definitions of the kinetic decoupling temperature exist in the literature (see e.g. [50,51,57,126,127]), for example requiring that $\gamma = H$ at the time of kinetic decoupling, which are all related to the definition advocated here by multiplying the small temperature regime of Eq. (A11) with a constant different from unity.

The most prominent observable connected to kinetic decoupling is that of a cutoff in the power spectrum of matter density perturbations. For very late kinetic decoupling, the dominant mechanism of suppressing the growth of DM perturbations is dark acoustic oscillations [75,76] (unless DM is very light, in which case free-streaming effects [86] start to dominate). As recently confirmed numerically [17], the resulting minimal halo mass is then given by Eq. (1), which is in rather good agreement with earlier analytic estimates. Note that $T_{\rm kd}$ in this expression is calculated by using the definition given by Eq. (A11); for an alternative definition, the expected magnitude of $M_{\rm cut}$ has to be correspondingly rescaled.

Let us conclude this section by making explicit how the above general analysis simplifies for the purpose of the specific application we are interested in for most of this article: DM scattering with a highly relativistic species, resulting in kinetic decoupling in the keV range. The latter implies that we are still well in the radiation dominated era, $H^2 = (4\pi^3 G/45)g_{\rm eff}T^4$, with a constant number of effective relativistic degrees of freedom $g_{\rm eff} = 3.36$.[11] Equation (A6) then becomes

$$\frac{dT_\chi}{dT} - 2\frac{T_\chi}{T} = (T_\chi - \xi T)\frac{f(T_{\tilde{\gamma}})}{T^3}, \quad \text{(A13)}$$

where

$$f(T_{\tilde{\gamma}}) = \frac{3\sqrt{5/\pi}}{2\pi g_{\rm eff}^{1/2}} M_{\rm Pl}\gamma(T_{\tilde{\gamma}}) = \frac{\sqrt{5/\pi}}{2(2\pi)^4}\frac{M_{\rm Pl}}{g_{\rm eff}^{1/2}\eta_\chi m_\chi^3}\int d\omega g^\pm \partial_\omega(\omega^4\langle|\mathcal{M}|^2\rangle_t). \quad \text{(A14)}$$

As in Eq. (A8), we can evaluate the amplitude at $t = 0$ instead of taking the average *if* the Taylor series around this point locally provides a good approximation. In general, the above two equations need to be solved numerically to determine $T_{\rm kd}$ according to Eq. (A11), as implemented in `DarkSUSY` [74,128]. In many cases of practical interest, the amplitude is furthermore well approximated by a power law for small energies,[12]

[11] If $\tilde{\gamma}$ constitutes some form of *dark* radiation, this would in principle contribute additional degrees of freedom on top of those from the standard model neutrinos and photons taken into account here. Such an additional contribution is observationally strongly constrained [1], and would anyway change the prediction for $T_{\rm kd}$ only by a factor $\lesssim \frac{1}{4}\Delta g_{\rm eff}/g_{\rm eff}$; cf. Eq. (A16).

[12] Note that this introduces the coefficient c_n with the correct prescription of summing and averaging the amplitude squared over initial and final states, in the sense that this is how it enters in the momentum transfer rate; see Eqs. (A7) and (A14).

$$\frac{1}{\eta_\chi}\langle|\mathcal{M}|^2\rangle_t = c_n\frac{\omega^n}{m_\chi^n} + \mathcal{O}\left(\frac{\omega^{n+1}}{m_\chi^{n+1}}\right). \tag{A15}$$

In this case, Eqs. (A13)–(A14) can be solved analytically even for noninteger $n > -1$ [122], and the kinetic decoupling temperature as defined in Eq. (A11) is given by

$$\frac{T_{\rm kd}}{m_\chi} = \left(\frac{\xi T^2}{m_\chi T_\chi}\right)_{T\lesssim T_{\rm kd}} = \left[\left(\frac{a}{n+2}\right)^{1/(n+2)}\Gamma\left(\frac{n+1}{n+2}\right)\right]^{-1}, \tag{A16}$$

with

$$a = \sqrt{\frac{5}{2(2\pi)^9 g_{\rm eff}}}(n+4)!\zeta(n+4)\xi^{n+4}c_n\frac{M_{\rm Pl}}{m_\chi} \tag{A17}$$

for a bosonic $\tilde{\gamma}$. If $\tilde{\gamma}$ is a fermion, the above expression has to be multiplied by a factor of $1-2^{-(n+3)}$.

APPENDIX B: SCATTERING MATRIX ELEMENTS

In this appendix, we provide the Lagrangians and scattering amplitudes for all models included in our analysis. While we use the full expressions to calculate the kinetic decoupling temperature, we state here only the leading terms for $|\mathcal{M}|^2$ in the limit $m_\chi \gg \omega \gg m_{\tilde{\gamma}}$. In each case, we also check explicitly whether keeping only these leading order terms provides a good estimate for the calculation of $T_{\rm kd}$, and whether simply evaluating $|\mathcal{M}|^2$ for $t=0$ leads to a reliable estimate of $T_{\rm kd}$ or whether one has instead to use the t-averaging prescription given in Eq. (A8). See Appendix A for more details about how to calculate $T_{\rm kd}$.

1. Two-particle models

Let us first consider those simplified models that only contain the (cold) DM particle χ and the (relativistic) scattering partner $\tilde{\gamma}$. As motivated in Sec. III, we are then interested in the following interaction terms (to indicate the spin of the involved particles, we denote scalars always with ϕ, vectors with V and fermions with ψ).

A. *Scalar four-point interaction*

$$\Delta\mathcal{L} = \frac{\lambda}{4}\phi_\chi^2\phi_{\tilde{\gamma}}^2. \tag{B1}$$

B. *DM-DR interactions through s/u-channel*

(i) Scalar-scalar

$$\Delta\mathcal{L} = \frac{\mu_\chi}{2}\phi_\chi^2\phi_{\tilde{\gamma}}. \tag{B2}$$

(ii) Fermion-scalar

$$\Delta\mathcal{L} = g_\chi\bar{\psi}_\chi\psi_\chi\phi_{\tilde{\gamma}}. \tag{B3}$$

(iii) Scalar–$U(1)$ vector

$$\Delta\mathcal{L} = ig_\chi[(\partial_\mu\phi_\chi^\dagger)\phi_\chi - \phi_\chi^\dagger(\partial_\mu\phi_\chi)]V_{\tilde{\gamma}}^\mu - g_\chi^2(V_{\tilde{\gamma}}^\mu)^2|\phi_\chi|^2. \tag{B4}$$

(iv) Fermion–$U(1)$ vector

$$\Delta\mathcal{L} = g_\chi\bar{\psi}_\chi \not{V}_{\tilde{\gamma}}\psi_\chi. \tag{B5}$$

C. *DM-DR interactions through t-channel*

(i) Scalar-scalar

$$\Delta\mathcal{L} = \frac{\mu_\chi}{2}\phi_\chi^2\phi_{\tilde{\gamma}} + \frac{\mu_{\tilde{\gamma}}}{6}\phi_{\tilde{\gamma}}^3. \tag{B6}$$

(ii) Fermion-scalar

$$\Delta\mathcal{L} = g_\chi\bar{\psi}_\chi\psi_\chi\phi_{\tilde{\gamma}} + \frac{\mu_{\tilde{\gamma}}}{6}\phi_{\tilde{\gamma}}^3. \tag{B7}$$

D. *DM-DR interactions through all channels*

(i) Vector-scalar

$$\Delta\mathcal{L} = g_\chi m_\chi V_{\chi\mu}V_\chi^\mu\phi_{\tilde{\gamma}} + \frac{1}{2}g_\chi^2 V_{\chi\mu}V_\chi^\mu\phi_{\tilde{\gamma}}^2 - \frac{1}{2}g_\chi\frac{m_{\tilde{\gamma}}^2}{m_\chi}\phi_{\tilde{\gamma}}^3. \tag{B8}$$

[This results from a spontaneously broken $U(1)$ symmetry; the imaginary part of the original scalar field Φ thus gives the mass to V_χ^μ.]

(ii) Scalar–$SU(N)$ vector

$$\Delta\mathcal{L} = -\frac{1}{2}\,\mathrm{Tr}[(F_{\mu\nu}^a)^2] - g_\chi^2 V_{\tilde{\gamma}}^{a\mu}V_{\tilde{\gamma}\mu}^b\Phi_\chi^\dagger t^a t^b\Phi_\chi + ig_\chi[(\partial_\mu\Phi_\chi^\dagger)t^a\Phi_\chi - \Phi_\chi^\dagger t^a(\partial_\mu\Phi_\chi)]V_{\tilde{\gamma}}^{a\mu}. \tag{B9}$$

(iii) Fermion–$SU(N)$ vector

$$\Delta\mathcal{L} = -\frac{1}{2}\,\mathrm{Tr}[(F_{\mu\nu}^a)^2] + g_\chi\bar{\Psi}_\chi\not{V}_{\tilde{\gamma}}^a t^a\Psi_\chi. \tag{B10}$$

We list the squared amplitudes for all those models, with and without averaging over t, in Table I. In all cases, the amplitudes squared are summed (not averaged) over all external spins, polarization states and "colors," as well as over particles and antiparticles. We also provide the ratio of kinetic decoupling temperatures that results when using the t-averaging prescription and the simplified $t=0$ prescription, respectively. To obtain this ratio, we calculated the kinetic decoupling temperature by solving the full process

TABLE I. Full list of relevant two-particle models. For the scattering matrix elements, only the leading terms in ω/m_χ are given, assuming $m_\chi \gg \omega \gg m_{\tilde{\gamma}}$ (for the t-channel results, we have further assumed that $\mu_{\tilde{\gamma}}$ is sufficiently large that the t-channel amplitude always dominates over the s/u-channel amplitudes). In the fifth column, we state the ratio of the kinetic decoupling temperature resulting from the t-averaging prescription to that from the $t \to 0$ prescription. The next-to-last column indicates whether keeping only the leading order result for the amplitude (after averaging or setting $t = 0$) provides a good estimate for $T_{\rm kd}$. In this case the analytical solution, Eq. (A16), can be used; otherwise, Eq. (A13) must be solved numerically. The last column, finally, states the full set of ETHOS parameters [63] that describe the respective model, as defined in Eqs. (B11)–(B13).

DM/DR	$\Delta\mathcal{L}$	$\langle\lvert\mathcal{M}\rvert^2\rangle_t$	$\lvert\mathcal{M}\rvert^2_{t=0}$	$\frac{T_{\rm kd}(\langle\lvert\mathcal{M}\rvert^2\rangle_t)}{T_{\rm kd}(\lvert\mathcal{M}\rvert^2_{t=0})}$	$\lvert\mathcal{M}\rvert^2 \propto (\frac{\omega}{m_\chi})^n$	ETHOS parameters
			Four-point (contact interaction only)			
Scalar/scalar	(B1)	λ^2	λ^2	1.0	✓	$\{a_2, \alpha_{l\geq 2} = 1\}$
			s/u-channel			
Scalar/scalar	(B2)	$\frac{\mu_\chi^4}{2m_\chi^4}$	$\mathcal{O}(m_{\tilde{\gamma}}^4)$	Not applicable	✓/✗	$\{a_2, \alpha_2 = \frac{3}{5}, \alpha_{l\geq 3} = \frac{2}{3}\}$
Fermion/scalar	(B3)	$\frac{16 g_\chi^4}{3}$	$16 g_\chi^4$	1.7	✓	$\{a_2, \alpha_2 = \frac{3}{5}, \alpha_{l\geq 3} = 1\}$
Scalar/vector	(B4)	$\frac{32 g_\chi^4}{3}$	$16 g_\chi^4$	1.2	✓	$\{a_2, \alpha_2 = \frac{9}{10}, \alpha_{l\geq 3} = 1\}$
Fermion/vector	(B5)	$\frac{64 g_\chi^4}{3}$	$32 g_\chi^4$	1.2	✓	$\{a_2, \alpha_2 = \frac{9}{10}, \alpha_{l\geq 3} = 1\}$
			t-channel			
Scalar/scalar	(B6)	$\frac{\mu_\chi^2 \mu_{\tilde{\gamma}}^2}{8\omega^4} \ln \frac{4\omega^2}{m_{\tilde{\gamma}}^2}$	$\frac{\mu_\chi^2 \mu_{\tilde{\gamma}}^2}{m_{\tilde{\gamma}}^4}$	Not applicable	✗	($m_{\tilde{\gamma}} \to 0$ undefined)
Fermion/scalar	(B7)	$\frac{2 g_\chi^2 \mu_{\tilde{\gamma}}^2 m_\chi^2 \ln(4\omega^2/m_{\tilde{\gamma}}^2)}{\omega^4}$	$\frac{16 g_\chi^2 \mu_{\tilde{\gamma}}^2 m_\chi^2}{m_{\tilde{\gamma}}^4}$	Not applicable	✗	($m_{\tilde{\gamma}} \to 0$ undefined)
			DM-DR interactions through *all* channels			
Vector/scalar	(B8)	$4 g_\chi^4$	$48 g_\chi^4$	3.5	✓	$\{a_2, \alpha_2 = \frac{3}{5}, \alpha_{l\geq 3} = 1\}$
Scalar/vector [$SU(N)$]	(B9)	$\frac{9 g_\chi^4 C_F C_A^2 m_\chi^2 \ln \frac{4\omega^2}{m_{\tilde{\gamma}}^2}}{\omega^2}$	$\frac{72 g_\chi^4 C_F C_A^2 m_\chi^2 \omega^2}{m_{\tilde{\gamma}}^4}$	Not applicable	✗	($m_{\tilde{\gamma}} \to 0$ undefined)
Fermion/vector [$SU(N)$]	(B10)	$\frac{18 g_\chi^4 C_F C_A^2 m_\chi^2 \ln \frac{4\omega^2}{m_{\tilde{\gamma}}^2}}{\omega^2}$	$\frac{144 g_\chi^4 C_F C_A^2 m_\chi^2 \omega^2}{m_{\tilde{\gamma}}^4}$	Not applicable	✗	($m_{\tilde{\gamma}} \to 0$ undefined)

Eq. (A13) numerically. In the next-to-last column of Table I, we indicate whether expanding the amplitude as a power law in the energy of the relativistic scattering partner provides an accurate estimate of the correct decoupling temperature (i.e. whether the analytic result given in Eq. (A16) agrees with the full numerical result at the percent level).

We are studying here situations with intrinsically large kinematic enhancements, i.e. where the presence of small quantities in propagators (namely t, ω and $m_{\tilde{\gamma}}$) can have significant effects on the amplitude. It should therefore not be a surprise that we identify cases where the simple $t = 0$ prescription breaks down completely. One of those examples is the case of scalar/scalar scattering via the s/u channel, where the amplitude evaluated at $t = 0$ is proportional to $m_{\tilde{\gamma}}^4/m_\chi^4$ while the averaged amplitude is not suppressed by the small DR mass. A similar issue appears, certainly not unexpectedly, in all cases where $\tilde{\gamma}$ appears in the t-channel. Apart from that, we confirmed that if the squared amplitude takes the form of a power law in the energy *close to kinetic decoupling*, as well as at slightly higher temperatures, the analytic solution (A16) for the kinetic decoupling temperature provides a very reliable estimate for the full numerical result. For DR in the t-channel, however, the amplitude close to kinetic decoupling is not of the form given in Eq. (A15) and, consequently, the analytic solution cannot be expected to apply.

As mentioned in the introduction, finally, ETHOS [63] provides an efficient way of classifying the impact of DM models on structure formation by means of a handful of phenomenological parameters—in the sense that every DM model with similar ETHOS parameters leads to almost identical results in full numerical simulations. An important input here is the DR opacity to DM scattering, $\dot{\kappa}_{\tilde{\gamma}-\chi}$, which for relativistic DR typically can be parametrized as

$$\frac{\dot{\kappa}_{\tilde{\gamma}-\chi}}{\Omega_\chi h^2} \equiv \frac{-n_\chi}{16\pi m_\chi^2 (1+z)\Omega_\chi h^2} \frac{\int d\omega \omega^3 g^{\pm}(\omega)[A_0 - A_1]}{\int d\omega \omega^3 g^{\pm}(\omega)} = -\sum_m a_m \left(\frac{1+z}{1+z_{\rm kd}}\right)^m, \tag{B11}$$

where z denotes the cosmological redshift and

$$A_l(\omega) \equiv \frac{1}{2}\int_{-1}^{1} d\cos\theta P_l(\cos\theta)\left.\frac{|\mathcal{M}|^2}{\eta_\chi \eta_{\tilde{\gamma}}}\right|_{\substack{t=2\omega^2(\cos\theta-1)\\ s=m_\chi^2+2\omega m_\chi}}. \quad \text{(B12)}$$

In the above expression, θ is the angle between the incoming and outgoing DR particle, and P_l denotes the lth Legendre polynomial. This means that $A_0 - A_1$ is, up to a constant, essentially just the transfer cross section σ_T for DM-DR scattering, and $\dot{\kappa}_{\tilde{\gamma}-\chi}$ is thus closely related to the momentum transfer rate γ; see Eq. (A9). The only other relevant parameters for the models studied here are a set of angular coefficients α_l defined by

$$\alpha_l \equiv \frac{\int d\omega \omega^3 g^{\pm}(\omega)[A_0(\omega) - A_l(\omega)]}{\int d\omega \omega^3 g^{\pm}(\omega)[A_0(\omega) - A_1(\omega)]}, \qquad l \geq 2. \quad \text{(B13)}$$

In Table I, we provide for each model the full set of nonvanishing parameters $\{a_n, \alpha_l\}$ in the limit considered here, namely $T_{\tilde{\gamma}} \ll m_\chi$. We also indicate those cases where the above expressions do not apply because the limit $m_{\tilde{\gamma}} \to 0$ cannot be taken, a situation for which the ETHOS parametrization has not been worked out yet. For models with the same set of parameters, the effect of the cutoff in the primordial power spectrum on nonlinear structure formation will be identical. We note that the value of α_l has a rather limited impact in this respect, as it leaves shape and location of the *first* peak in the linear power spectrum mostly unaffected [63].

2. Three-particle models

We now consider simplified models where the scattering between nonrelativistic DM particles χ and relativistic particles $\tilde{\gamma}$ is mediated by a *different* particle—either a "DM-like" particle χ' which is slightly heavier than χ (leading to s/u-channel exchange) or a "DR-like" particle $\tilde{\gamma}'$ which is much heavier than $\tilde{\gamma}$ (leading to t-channel exchange). As motivated in Sec. IV, we are then interested in the following interaction terms of dimension 4 (to indicate the spin of χ and $\tilde{\gamma}$, we denote again scalars always with ϕ, vectors with V and fermions with ψ).

A. *DM-DR interactions through a mediator χ' in the s/u-channels*

(i) Scalar-scalar-scalar

$$\Delta\mathcal{L} = \mu_\chi \phi_\chi \phi_{\chi'} \phi_{\tilde{\gamma}}. \quad \text{(B14)}$$

(ii) Scalar-fermion-fermion

$$\Delta\mathcal{L} = g_\chi \phi_\chi \bar{\psi}_{\chi'} \psi_{\tilde{\gamma}} + \text{H.c.} \quad \text{(B15)}$$

(iii) Fermion-scalar-fermion

$$\Delta\mathcal{L} = g_\chi \phi_{\chi'} \bar{\psi}_{\tilde{\gamma}} \psi_\chi + \text{H.c.} \quad \text{(B16)}$$

(iii) Fermion-fermion-scalar

$$\Delta\mathcal{L} = g_\chi \phi_{\tilde{\gamma}} \bar{\psi}_{\chi'} \psi_\chi + \text{H.c.} \quad \text{(B17)}$$

B. *DM-DR interactions through a mediator $\tilde{\gamma}'$ in the t-channel*

(i) Scalar-scalar-scalar

$$\Delta\mathcal{L} = \frac{\mu_\chi}{2}\phi_{\tilde{\gamma}'}\phi_\chi^2 + \frac{\mu_{\tilde{\gamma}}}{2}\phi_{\tilde{\gamma}'}\phi_{\tilde{\gamma}}^2. \quad \text{(B18)}$$

(ii) Scalar-scalar-fermion

$$\Delta\mathcal{L} = \frac{\mu_\chi}{2}\phi_{\tilde{\gamma}'}\phi_\chi^2 + g_{\tilde{\gamma}}\bar{\psi}_{\tilde{\gamma}}\psi_{\tilde{\gamma}}\phi_{\tilde{\gamma}'}. \quad \text{(B19)}$$

(iii) Scalar-scalar-vector

$$\Delta\mathcal{L} = \frac{\mu_\chi}{2}\phi_{\tilde{\gamma}'}\phi_\chi^2 + \mu_{\tilde{\gamma}}\phi_{\tilde{\gamma}'}V_{\mu\tilde{\gamma}}V_{\tilde{\gamma}}^{\mu}. \quad \text{(B20)}$$

(We assume that the gauge symmetry is broken spontaneously by $\langle\phi_{\tilde{\gamma}'}\rangle \propto \mu_{\tilde{\gamma}}$.)

(iv) Scalar-vector-scalar

$$\Delta\mathcal{L} = ig_\chi[(\partial_\mu\phi_\chi^\dagger)\phi_\chi - \phi_\chi^\dagger(\partial_\mu\phi_\chi)]V_{\tilde{\gamma}'}^{\mu} + ig_{\tilde{\gamma}}[(\partial_\mu\phi_{\tilde{\gamma}}^\dagger)\phi_{\tilde{\gamma}} - \phi_{\tilde{\gamma}}^\dagger(\partial_\mu\phi_{\tilde{\gamma}})]V_{\tilde{\gamma}'}^{\mu}. \quad \text{(B21)}$$

(v) Scalar-vector-fermion

$$\Delta\mathcal{L} = ig_\chi[(\partial_\mu\phi_\chi^\dagger)\phi_\chi - \phi_\chi^\dagger(\partial_\mu\phi_\chi)]V_{\tilde{\gamma}'}^{\mu} + g_{\tilde{\gamma}}\bar{\psi}_{\tilde{\gamma}}\not{V}_{\tilde{\gamma}'}\psi_{\tilde{\gamma}}. \quad \text{(B22)}$$

(vi) Fermion-scalar-scalar

$$\Delta\mathcal{L} = g_\chi\bar{\psi}_\chi\psi_\chi\phi_{\tilde{\gamma}'} + \frac{\mu_{\tilde{\gamma}}}{2}\phi_{\tilde{\gamma}'}\phi_{\tilde{\gamma}}^2. \quad \text{(B23)}$$

(vii) Fermion-scalar-fermion

$$\Delta\mathcal{L} = g_\chi\bar{\psi}_\chi\psi_\chi\phi_{\tilde{\gamma}'} + g_{\tilde{\gamma}}\bar{\psi}_{\tilde{\gamma}}\psi_{\tilde{\gamma}}\phi_{\tilde{\gamma}'}. \quad \text{(B24)}$$

(viii) Fermion-scalar-vector

$$\Delta\mathcal{L} = g_\chi\bar{\psi}_\chi\psi_\chi\phi_{\tilde{\gamma}'} + \mu_{\tilde{\gamma}}\phi_{\tilde{\gamma}'}V_{\mu\tilde{\gamma}}V_{\tilde{\gamma}}^{\mu}. \quad \text{(B25)}$$

(We assume again that the gauge symmetry is broken spontaneously by $\langle\phi_{\tilde{\gamma}'}\rangle \propto \mu_{\tilde{\gamma}}$.)

(ix) Fermion-vector-scalar

$$\Delta\mathcal{L} = g_\chi\bar{\psi}_\chi\not{V}_{\tilde{\gamma}'}\psi_\chi + ig_{\tilde{\gamma}}[(\partial_\mu\phi_{\tilde{\gamma}}^\dagger)\phi_{\tilde{\gamma}} - \phi_{\tilde{\gamma}}^\dagger(\partial_\mu\phi_{\tilde{\gamma}})]V_{\tilde{\gamma}'}^{\mu}. \quad \text{(B26)}$$

TABLE II. Full list of relevant three-particle models—including DM particles χ, (dark) radiation particles $\tilde{\gamma}$, and mediator particles χ' or $\tilde{\gamma}'$. For s/u-channel and t-channel processes, we have assumed $m_\chi \gg \Delta m \equiv m_{\chi'} - m_\chi \gg \omega \gg m_{\tilde{\gamma}}$ and $m_\chi \gg m_{\tilde{\gamma}'} \gg \omega \gg m_{\tilde{\gamma}}$, respectively; in all cases, we state only the leading terms for the squared amplitude. The last three columns are defined as in Table I.

DM/mediator/DR	$\Delta\mathcal{L}$	$\langle\lvert\mathcal{M}\rvert^2\rangle_t$	$\lvert\mathcal{M}\rvert^2_{t=0}$	$\frac{T_{\rm kd}(\langle\lvert\mathcal{M}\rvert^2\rangle_t)}{T_{\rm kd}(\lvert\mathcal{M}\rvert^2_{t=0})}$	$\lvert\mathcal{M}\rvert^2 \propto (\frac{\omega}{m_\chi})^n$	ETHOS parameters
			s/u-channel			
Scalar/scalar/scalar	(B14)	$\frac{\mu_\chi^4}{m_\chi^2\Delta m^2}$	$\frac{\mu_\chi^4}{m_\chi^2\Delta m^2}$	1.0	✓	$\{a_2, \alpha_{l\geq 2} = 1\}$
Scalar/fermion/fermion	(B15)	$\frac{32 g_\chi^4\omega^2}{3\Delta m^2}$	$\frac{16 g_\chi^4\omega^4}{\Delta m^4}$	Not applicable	✓	$\{a_4, \alpha_{l\geq 2} = \frac{3}{4}\}$
Fermion/scalar/fermion	(B16)	$\frac{88 g_\chi^4\omega^2}{3\Delta m^2}$	$\frac{40 g_\chi^4\omega^2}{\Delta m^2}$	1.1	✓	$\{a_4, \alpha_{l\geq 2} = \frac{12}{11}\}$
Fermion/fermion/scalar	(B17)	$\frac{64 g_\chi^4 m_\chi^2}{\Delta m^2}$	$\frac{64 g_\chi^4 m_\chi^2}{\Delta m^2}$	1.0	✓	$\{a_2, \alpha_{l\geq 2} = 1\}$
			t-channel			
Scalar/scalar/scalar	(B18)	$\frac{\mu_\chi^2\mu_{\tilde{\gamma}}^2}{m_{\tilde{\gamma}'}^4}$	$\frac{\mu_\chi^2\mu_{\tilde{\gamma}}^2}{m_{\tilde{\gamma}'}^4}$	1.0	✓	$\{a_2, \alpha_{l\geq 2} = 1\}$
Scalar/scalar/fermion	(B19)	$\frac{32\mu_\chi^2 g_{\tilde{\gamma}}^2\omega^2}{3m_{\tilde{\gamma}'}^4}$	$\mathcal{O}(m_{\tilde{\gamma}}^2)$	Not applicable	✓/✗	$\{a_4, \alpha_{l\geq 2} = \frac{3}{4}\}$
Scalar/scalar/vector	(B20)	$\frac{4\mu_\chi^2\mu_{\tilde{\gamma}}^2}{m_{\tilde{\gamma}'}^4}$	$\frac{4\mu_\chi^2\mu_{\tilde{\gamma}}^2}{m_{\tilde{\gamma}'}^4}$	1.0	✓	$\{a_2, \alpha_{l\geq 2} = 1\}$
Scalar/vector/scalar	(B21)	$\frac{64 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	$\frac{64 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	1.0	✓	$\{a_4, \alpha_{l\geq 2} = 1\}$
Scalar/vector/fermion	(B22)	$\frac{128 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{3m_{\tilde{\gamma}'}^4}$	$\frac{128 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	1.3	✓	$\{a_4, \alpha_{l\geq 2} = \frac{3}{2}\}$
Fermion/scalar/scalar	(B23)	$\frac{16 g_\chi^2\mu_{\tilde{\gamma}}^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	$\frac{16 g_\chi^2\mu_{\tilde{\gamma}}^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	1.0	✓	$\{a_2, \alpha_{l\geq 2} = 1\}$
Fermion/scalar/fermion	(B24)	$\frac{512 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{3m_{\tilde{\gamma}'}^4}$	$\mathcal{O}(m_{\tilde{\gamma}}^2)$	Not applicable	✓/✗	$\{a_4, \alpha_{l\geq 2} = \frac{3}{4}\}$
Fermion/scalar/vector	(B25)	$\frac{32 g_\chi^2\mu_{\tilde{\gamma}}^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	$\frac{32 g_\chi^2\mu_{\tilde{\gamma}}^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	1.0	✓	$\{a_2, \alpha_{l\geq 2} = 1\}$
Fermion/vector/scalar	(B26)	$\frac{128 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	$\frac{128 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	1.0	✓	$\{a_4, \alpha_{l\geq 2} = 1\}$
Fermion/vector/fermion	(B27)	$\frac{256 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{3m_{\tilde{\gamma}'}^4}$	$\frac{256 g_\chi^2 g_{\tilde{\gamma}}^2\omega^2 m_\chi^2}{m_{\tilde{\gamma}'}^4}$	1.3	✓	$\{a_4, \alpha_{l\geq 2} = \frac{3}{2}\}$

(x) Fermion-vector-fermion

$$\Delta\mathcal{L} = g_\chi\bar{\psi}_\chi \not{V}_{\tilde{\gamma}'}\psi_\chi + g_{\tilde{\gamma}}\bar{\psi}_{\tilde{\gamma}}\not{V}_{\tilde{\gamma}'}\psi_{\tilde{\gamma}}. \quad \text{(B27)}$$

We note that the dimensionful coupling $\mu_{\tilde{\gamma}}$ cannot be chosen completely independently of the mediator mass $m_{\tilde{\gamma}'}$; in Eqs. (B25) and (B20), e.g., it is the vacuum expectation value of the same field that gives rise to both quantities. We list the resulting squared amplitudes in Table II, following the same format as in Table I for the two-particle models. Note that we focus on parameter choices where we can expect qualitative differences with respect to the two-particle models discussed above. Sufficiently close to kinetic decoupling, we therefore require $m_\chi \gg \Delta m \equiv m_{\chi'} - m_\chi \gg \omega \gg m_{\tilde{\gamma}}$ (for s/u-channel mediated processes) and $m_\chi \gg m_{\tilde{\gamma}'} \gg \omega \gg m_{\tilde{\gamma}}$ (for t-channel mediated processes), respectively.

Also in this case, we identify situations where the simple $t = 0$ prescription leads to a qualitatively wrong result for the inferred decoupling temperature. In the s/u-channel, this happens for the combination of scalar DM and fermionic χ' and $\tilde{\gamma}$, where $\lvert\mathcal{M}\rvert^2_{t=0} \propto \omega^4$ while $\langle\lvert\mathcal{M}\rvert^2\rangle_t \propto \omega^2$. In the t-channel the "critical" combinations concern fermionic $\tilde{\gamma}$ and scalar $\tilde{\gamma}'$: unlike that suggested by the result for $\lvert\mathcal{M}\rvert^2_{t=0}$, those combinations do not lead to an insignificant scattering rate—but rather to a scattering rate that is (up to a constant factor) the same as in the case of exchanging a *vector* particle $\tilde{\gamma}'$. As long as one uses the correct description for calculating the matrix element, on the other hand, the analytic solution (A16) for the kinetic decoupling temperature always provides a very reliable estimate for the full numerical result.

For convenience, we provide the ETHOS parameters also for all three-particle models. For the case of $\{a_m, \alpha_l\} = \{a_4, \alpha_{l\geq 2} = 3/2\}$, which appears e.g. for fermion-fermion scattering via vector exchange [see Eq. (B27)] as studied in Refs. [47,48,55,57], detailed numerical simulations have already been performed [17].

Those simulations included the effect of DM self-interactions, for which we do not explicitly list the relevant ETHOS parameters here. We note, finally, that a recent computation for the models (B24) and (B27) resulted in identical linear power spectra for these two cases when neglecting the impact of perturbations in the DR fluid [120]. Including this effect, which is encoded in the parameters α_l, we thus expect (small) differences between the power spectra generated by DM-DR scattering mediated by scalar and vector mediators, respectively (see the discussion in Ref. [63]).

Let us, finally, point out that none of the models studied in this section (nor in the previous section where we considered two-particle models) allows for the possibility that $\tilde{\gamma}$ is the SM photon. For the Lagrangians in Eqs. (B4)–(B5), for example, this is excluded because it would lead to too large DM self-interactions (as discussed in Sec. III B); the Lagrangians (B20) and (B25), on the other hand, are not compatible with an unbroken $U(1)$ gauge symmetry. We note that this conclusion may change when including higher-dimensional operators in the discussion, an interesting candidate being e.g. a scalar $\tilde{\gamma}'$ in the t-channel that couples via $\tilde{\gamma}' F_{\mu\nu} F^{\mu\nu}$ to photons. As the effective coupling $g_{\tilde{\gamma}'}$ of such higher-dimensional operators would necessarily be suppressed, however, we expect from Fig. 12 that such solutions would require even smaller mediator masses $m_{\tilde{\gamma}'}$, and hence even lighter DM.

Similarly, it appears to be challenging for any concrete model building to identify $\tilde{\gamma}$ with the SM neutrino. The main reason is that due to $SU(2)$ gauge invariance any new state coupling to the neutrino should couple with equal strength to the (left-handed) SM electron. This argument basically excludes the possibility of achieving late kinetic decoupling with $\tilde{\gamma} = \nu_L$ in the Lagrangians stated in Eqs. (B19), (B22), (B24), and (B27), because the coupling of electrons to new light states is generally strongly constrained. In the case of Eqs. (B15)–(B16), on the other hand, the leptons would couple to *heavy* new states—which is not excluded if the mass scale is high enough (in supersymmetry, e.g., this would correspond to a coupling among the neutralino, lepton and slepton). Inspecting Eqs. (18)–(19), however, tells us that a large mass scale can only be reconciled with late kinetic decoupling for extremely small mass splittings δ. Again, higher-dimensional operators may potentially allow for qualitatively different options.

[1] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016).
[2] V. Springel, Mon. Not. R. Astron. Soc. **364**, 1105 (2005).
[3] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. F. Snyder, S. Bird, D. Nelson, and L. Hernquist, Nature (London) **509**, 177 (2014).
[4] W. J. G. de Blok and S. S. McGaugh, Mon. Not. R. Astron. Soc. **290**, 533 (1997).
[5] A. A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, Astrophys. J. **522**, 82 (1999).
[6] B. Moore, S. Ghigna, F. Governato, G. Lake, T. R. Quinn, J. Stadel, and P. Tozzi, Astrophys. J. **524**, L19 (1999).
[7] J. Zavala, Y. P. Jing, A. Faltenbacher, G. Yepes, Y. Hoffman, S. Gottlober, and B. Catinella, Astrophys. J. **700**, 1779 (2009).
[8] S.-H. Oh, W. J. G. de Blok, E. Brinks, F. Walter, and R. C. Kennicutt, Jr., Astron. J. **141**, 193 (2011).
[9] E. Papastergis, A. M. Martin, R. Giovanelli, and M. P. Haynes, Astrophys. J. **739**, 38 (2011).
[10] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, Mon. Not. R. Astron. Soc. **415**, L40 (2011).
[11] M. G. Walker and J. Penarrubia, Astrophys. J. **742**, 20 (2011).
[12] M. S. Pawlowski, P. Kroupa, and H. Jerjen, Mon. Not. R. Astron. Soc. **435**, 1928 (2013).
[13] A. Klypin, I. Karachentsev, D. Makarov, and O. Nasonova, Mon. Not. R. Astron. Soc. **454**, 1798 (2015).
[14] E. Papastergis, R. Giovanelli, M. P. Haynes, and F. Shankar, Astron. Astrophys. **574**, A113 (2015).
[15] K. A. Oman *et al.*, Mon. Not. R. Astron. Soc. **452**, 3650 (2015).
[16] R. Massey *et al.*, Mon. Not. R. Astron. Soc. **449**, 3393 (2015).
[17] M. Vogelsberger, J. Zavala, F.-Y. Cyr-Racine, C. Pfrommer, T. Bringmann, and K. Sigurdson, Mon. Not. R. Astron. Soc. **460**, 1399 (2016).
[18] J. Silk *et al.*, *Particle Dark Matter: Observations, Models and Searches*, edited by G. Bertone (University Press, Cambridge, 2010).
[19] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rep. **267**, 195 (1996).
[20] D. Hooper and S. Profumo, Phys. Rep. **453**, 29 (2007).
[21] A. Cakir (ATLAS and CMS Collaborations), *Proc. Sci.*, FPCP2015 (2015) 024 [arXiv:1507.08427].
[22] G. Aad *et al.* (ATLAS Collaboration), J. High Energy Phys. 10 (2015) 134.
[23] E. Aprile *et al.* (XENON100 Collaboration), Phys. Rev. Lett. **109**, 181301 (2012).
[24] D. S. Akerib *et al.* (LUX Collaboration), Phys. Rev. Lett. **112**, 091303 (2014).
[25] F. D. Steffen, Eur. Phys. J. C **59**, 557 (2009).
[26] J. L. Feng, Annu. Rev. Astron. Astrophys. **48**, 495 (2010).
[27] H. Baer, K.-Y. Choi, J. E. Kim, and L. Roszkowski, Phys. Rep. **555**, 1 (2015).
[28] R. Adhikari *et al.*, arXiv:1602.04816.
[29] H. Goldberg and L. J. Hall, Phys. Lett. B **174**, 151 (1986).

[30] B.-A. Gradwohl and J. A. Frieman, Astrophys. J. **398**, 407 (1992).
[31] P. Bode, J. P. Ostriker, and N. Turok, Astrophys. J. **556**, 93 (2001).
[32] S. Hannestad and R. J. Scherrer, Phys. Rev. D **62**, 043522 (2000).
[33] C. Boehm, P. Fayet, and R. Schaeffer, Phys. Lett. B **518**, 8 (2001).
[34] C. Boehm, A. Riazuelo, S. H. Hansen, and R. Schaeffer, Phys. Rev. D **66**, 083505 (2002).
[35] C. Boehm and R. Schaeffer, Astron. Astrophys. **438**, 419 (2005).
[36] J. A. R. Cembranos, J. L. Feng, A. Rajaraman, and F. Takayama, Phys. Rev. Lett. **95**, 181301 (2005).
[37] D. Hooper, M. Kaplinghat, L. E. Strigari, and K. M. Zurek, Phys. Rev. D **76**, 103515 (2007).
[38] J. L. Feng, M. Kaplinghat, H. Tu, and H.-B. Yu, J. Cosmol. Astropart. Phys. 07 (2009) 004.
[39] D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, and C. M. Wells, J. Cosmol. Astropart. Phys. 05 (2010) 021.
[40] L. Ackerman, M. R. Buckley, S. M. Carroll, and M. Kamionkowski, Phys. Rev. D **79**, 023519 (2009).
[41] D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, and C. M. Wells, J. Cosmol. Astropart. Phys. 10 (2011) 011.
[42] L. G. van den Aarssen, T. Bringmann, and Y. C. Goedecke, Phys. Rev. D **85**, 123512 (2012).
[43] L. G. van den Aarssen, T. Bringmann, and C. Pfrommer, Phys. Rev. Lett. **109**, 231301 (2012).
[44] J. M. Cline, Z. Liu, and W. Xue, Phys. Rev. D **85**, 101302 (2012).
[45] F.-Y. Cyr-Racine and K. Sigurdson, Phys. Rev. D **87**, 103515 (2013).
[46] S. Tulin, H.-B. Yu, and K. M. Zurek, Phys. Rev. Lett. **110**, 111301 (2013).
[47] T. Bringmann, J. Hasenkamp, and J. Kersten, J. Cosmol. Astropart. Phys. 07 (2014) 042.
[48] B. Dasgupta and J. Kopp, Phys. Rev. Lett. **112**, 031803 (2014).
[49] S. Tulin, H.-B. Yu, and K. M. Zurek, Phys. Rev. D **87**, 115007 (2013).
[50] F.-Y. Cyr-Racine, R. de Putter, A. Raccanelli, and K. Sigurdson, Phys. Rev. D **89**, 063517 (2014).
[51] I. M. Shoemaker, Phys. Dark Univ. **2**, 157 (2013).
[52] J. M. Cline, Z. Liu, G. Moore, and W. Xue, Phys. Rev. D **90**, 015023 (2014).
[53] J. Fan, A. Katz, L. Randall, and M. Reece, Phys. Dark Univ. **2**, 139 (2013).
[54] X. Chu and B. Dasgupta, Phys. Rev. Lett. **113**, 161301 (2014).
[55] P. Ko and Y. Tang, Phys. Lett. B **739**, 62 (2014).
[56] R. Foot and S. Vagnozzi, Phys. Rev. D **91**, 023512 (2015).
[57] J. F. Cherry, A. Friedland, and I. M. Shoemaker, arXiv: 1411.1071.
[58] M. R. Buckley, J. Zavala, F.-Y. Cyr-Racine, K. Sigurdson, and M. Vogelsberger, Phys. Rev. D **90**, 043524 (2014).
[59] C. Boehm, J. A. Schewtschenko, R. J. Wilkinson, C. M. Baugh, and S. Pascoli, Mon. Not. R. Astron. Soc. **445**, L31 (2014).
[60] M. A. Buen-Abad, G. Marques-Tavares, and M. Schmaltz, Phys. Rev. D **92**, 023531 (2015).
[61] J. Lesgourgues, G. Marques-Tavares, and M. Schmaltz, J. Cosmol. Astropart. Phys. 02 (2016) 037.
[62] R. Foot and S. Vagnozzi, J. Cosmol. Astropart. Phys. 07 (2016) 013.
[63] F.-Y. Cyr-Racine, K. Sigurdson, J. Zavala, T. Bringmann, M. Vogelsberger, and C. Pfrommer, Phys. Rev. D **93**, 123527 (2016).
[64] M. Viel, G. D. Becker, J. S. Bolton, and M. G. Haehnelt, Phys. Rev. D **88**, 043502 (2013).
[65] J. Baur, N. Palanque-Delabrouille, C. Yèche, C. Magneville, and M. Viel, J. Cosmol. Astropart. Phys. 08 (2016) 012.
[66] A. Garzilli, A. Boyarsky, and O. Ruchayskiy, arXiv: 1510.07006.
[67] S. Horiuchi, P. J. Humphrey, J. Onorbe, K. N. Abazajian, M. Kaplinghat, and S. Garrison-Kimmel, Phys. Rev. D **89**, 025017 (2014).
[68] K. T. Inoue, R. Takahashi, T. Takahashi, and T. Ishiyama, Mon. Not. R. Astron. Soc. **448**, 2704 (2015).
[69] A. Schneider, D. Anderhalden, A. Maccio, and J. Diemand, Mon. Not. R. Astron. Soc. **441**, L6 (2014).
[70] A. V. Macciò and F. Fontanot, Mon. Not. R. Astron. Soc. **404**, L16 (2010).
[71] A. V. Macciò, S. Paduroiu, D. Anderhalden, A. Schneider, and B. Moore, Mon. Not. R. Astron. Soc. **424**, 1105 (2012); **428**, 3715(E) (2013).
[72] S. Shao, L. Gao, T. Theuns, and C. S. Frenk, Mon. Not. R. Astron. Soc. **430**, 2346 (2013).
[73] X.-l. Chen, M. Kamionkowski, and X.-m. Zhang, Phys. Rev. D **64**, 021302 (2001).
[74] T. Bringmann, New J. Phys. **11**, 105027 (2009).
[75] A. Loeb and M. Zaldarriaga, Phys. Rev. D **71**, 103520 (2005).
[76] E. Bertschinger, Phys. Rev. D **74**, 063509 (2006).
[77] J. A. Schewtschenko, C. M. Baugh, R. J. Wilkinson, C. Boehm, S. Pascoli, and T. Sawala, Mon. Not. R. Astron. Soc. **461**, 2282 (2016).
[78] B. Bertoni, S. Ipek, D. McKeen, and A. E. Nelson, J. High Energy Phys. 04 (2015) 170.
[79] C. Brust, D. E. Kaplan, and M. T. Walters, J. High Energy Phys. 12 (2013) 058.
[80] M. Archidiacono, E. Giusarma, S. Hannestad, and O. Mena, Adv. High Energy Phys. **2013**, 191047 (2013).
[81] M. Wyman, D. H. Rudd, R. A. Vanderveld, and W. Hu, Phys. Rev. Lett. **112**, 051302 (2014).
[82] J. Hamann and J. Hasenkamp, J. Cosmol. Astropart. Phys. 10 (2013) 044.
[83] R. A. Battye and A. Moss, Phys. Rev. Lett. **112**, 051303 (2014).
[84] S. Gariazzo, C. Giunti, and M. Laveder, J. High Energy Phys. 11 (2013) 211.
[85] R. A. Battye, T. Charnock, and A. Moss, Phys. Rev. D **91**, 103508 (2015).
[86] A. M. Green, S. Hofmann, and D. J. Schwarz, J. Cosmol. Astropart. Phys. 08 (2005) 003.
[87] K. M. Nollett and G. Steigman, Phys. Rev. D **91**, 083505 (2015).

[88] F. Kahlhoefer, K. Schmidt-Hoberg, T. Schwetz, and S. Vogl, J. High Energy Phys. 02 (2016) 016.
[89] C. Boehm, M. J. Dolan, and C. McCabe, J. Cosmol. Astropart. Phys. 08 (2013) 041.
[90] P. Gondolo and G. Gelmini, Nucl. Phys. **B360**, 145 (1991).
[91] J. L. Feng and J. Kumar, Phys. Rev. Lett. **101**, 231301 (2008).
[92] B. Bellazzini, M. Cliche, and P. Tanedo, Phys. Rev. D **88**, 083506 (2013).
[93] M. Archidiacono, S. Hannestad, R. S. Hansen, and T. Tram, Phys. Rev. D **91**, 065021 (2015).
[94] P. F. Bedaque, M. I. Buchoff, and R. K. Mishra, J. High Energy Phys. 11 (2009) 046.
[95] M. J. Dolan, F. Kahlhoefer, C. McCabe, and K. Schmidt-Hoberg, J. High Energy Phys. 03 (2015) 171; 07 (2015) 103(E).
[96] J. L. Feng, M. Kaplinghat, and H.-B. Yu, Phys. Rev. Lett. **104**, 151301 (2010).
[97] B. D. Wandelt, R. Dave, G. R. Farrar, P. C. McGuire, D. N. Spergel, and P. J. Steinhardt, in *Proceedings, 4th International Symposium, DM 2000, Marina del Rey, 2000*, pp. 263–274.
[98] D. N. Spergel and P. J. Steinhardt, Phys. Rev. Lett. **84**, 3760 (2000).
[99] 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).
[100] J. Zavala, M. Vogelsberger, and M. G. Walker, Mon. Not. R. Astron. Soc. **431**, L20 (2013).
[101] M. Kaplinghat, S. Tulin, and H.-B. Yu, Phys. Rev. Lett. **116**, 041302 (2016).
[102] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray, and W. Tucker, Astrophys. J. **606**, 819 (2004).
[103] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradac, Astrophys. J. **679**, 1173 (2008).
[104] M. R. Buckley and P. J. Fox, Phys. Rev. D **81**, 083522 (2010).
[105] A. Loeb and N. Weiner, Phys. Rev. Lett. **106**, 171302 (2011).
[106] M. Vogelsberger, J. Zavala, and A. Loeb, Mon. Not. R. Astron. Soc. **423**, 3740 (2012).
[107] F. Kahlhoefer, K. Schmidt-Hoberg, J. Kummer, and S. Sarkar, Mon. Not. R. Astron. Soc. **452**, L54 (2015).
[108] E. W. Kolb and M. S. Turner, Front. Phys. **69**, 1 (1990).
[109] O. Lebedev, H. M. Lee, and Y. Mambrini, Phys. Lett. B **707**, 570 (2012).
[110] C. Gross, O. Lebedev, and Y. Mambrini, J. High Energy Phys. 08 (2015) 158.
[111] P. Ko and Y. Tang, arXiv:1609.02307.
[112] P. B. Arnold and L. G. Yaffe, Phys. Rev. D **52**, 7208 (1995).
[113] F. Beutler *et al.* (BOSS Collaboration), Mon. Not. R. Astron. Soc. **444**, 3501 (2014).
[114] J. Edsjö and P. Gondolo, Phys. Rev. D **56**, 1879 (1997).
[115] J. L. Feng, M. Kaplinghat, and H.-B. Yu, Phys. Rev. D **82**, 083525 (2010).
[116] T. Hahn and M. Perez-Victoria, Comput. Phys. Commun. **118**, 153 (1999).
[117] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, Phys. Rev. D **71**, 063528 (2005).
[118] R. Iengo, J. High Energy Phys. 05 (2009) 024.
[119] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, Phys. Rev. D **79**, 015014 (2009).
[120] T. Binder, L. Covi, A. Kamada, H. Murayama, T. Takahashi, and N. Yoshida, arXiv:1602.07624.
[121] Y. Tang, Phys. Lett. B **757**, 387 (2016).
[122] T. Bringmann and S. Hofmann, J. Cosmol. Astropart. Phys. 04 (2007) 016.
[123] J. Kasahara, Ph.D. thesis, University of Utah, 2009.
[124] P. Gondolo, J. Hisano, and K. Kadota, Phys. Rev. D **86**, 083523 (2012).
[125] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Westview Press, Boulder, 1995).
[126] S. Profumo, K. Sigurdson, and M. Kamionkowski, Phys. Rev. Lett. **97**, 031301 (2006).
[127] L. Visinelli and P. Gondolo, Phys. Rev. D **91**, 083526 (2015).
[128] P. Gondolo, J. Edsjö, P. Ullio, L. Bergström, M. Schelke, E. Baltz, T. Bringmann, and G. Duda, http://www.darksusy.org.