

## Reannihilation of self-interacting dark matter

Tobias Binder,<sup>1,\*</sup> Michael Gustafsson,<sup>1,†</sup> Ayuki Kamada,<sup>2,‡</sup> Stefan Marinus Rodrigues Sandner,<sup>3,§</sup> and Max Wiesner<sup>4,||</sup>

<sup>1</sup>*Institute for Theoretical Physics, Georg-August University Göttingen,  
Friedrich-Hund-Platz 1, Göttingen D-37077, Germany*

<sup>2</sup>*Center for Theoretical Physics of the Universe, Institute for Basic Science (IBS), Daejeon 34126, Korea*

<sup>3</sup>*DESY, Notkestraße 85, 22607 Hamburg, Germany*

<sup>4</sup>*Bethe Center for Theoretical Physics and Physikalisches Institut der Universität Bonn,  
Nussallee 12, 53115 Bonn, Germany*



(Received 16 March 2018; published 13 June 2018)

We explore the phenomenology of having a second epoch of dark matter annihilation into dark radiation long after the standard thermal freeze-out. Such a hidden *reannihilation* process could affect visible sectors only gravitationally. As a concrete realization we consider self-interacting dark matter (SIDM) with a light force mediator coupled to dark radiation. We demonstrate how resonantly Sommerfeld enhanced cross sections emerge to induce the reannihilation epoch. The effect is a temporally local modification of the Hubble expansion rate, and we show that the cosmic microwave background (CMB) measurements—as well as other observations—have a high sensitivity to observe this phenomenon. Special attention is given to the model region where late kinetic decoupling and strong self-interactions can alleviate several small-scale problems in the cold dark matter paradigm at the same time. Interestingly, we find that reannihilation might here also simultaneously lower the tension between CMB and low-redshift astronomical observations of  $H_0$  and  $\sigma_8$ . Moreover, we identify reannihilation as a clear signature to discriminate between the phenomenologically otherwise almost identical vector and scalar mediator realizations of SIDM.

DOI: 10.1103/PhysRevD.97.123004

### I. INTRODUCTION

The cosmological  $\Lambda$ CDM model has been very successful in describing the large-scale structures of the Universe. Its cold dark matter (CDM) ingredient consists of a collisionless matter component that enables one to fit the observed anisotropies in the cosmic microwave background (CMB) [1] and to explain their evolution to form structures such as galaxies. Despite these successes, there are potential tensions between some of its predictions and observations (see, e.g., Ref. [2]). On dwarf galactic scales, there is the “missing satellite problem” of too few discovered satellite galaxies around the Milky Way [3,4] and the “core-cusp problem” of too cored, rather than cuspy, dark matter (DM) density profiles [5,6] when compared to predictions in CDM setups. The “too-big-to-fail problem” [7] tries to sharpen these arguments by pointing out that in particular the biggest satellites in simulations, which should not fail to form stars and not escape detection, have DM density profiles more concentrated than those observed

[8,9]. The situation is, however, complicated by the fact that uncertain feedbacks from baryonic processes can be expected to play a role on subgalactic scales [10–12]. Nevertheless, it has been claimed in Refs. [13–15] that current state-of-the-art hydrodynamic simulations [16]—taking into account baryonic feedbacks in CDM setups—have not been able to predict the observed diversity of rotation curves in dwarf galaxies (see, however, e.g., Ref. [17] for a possible explanation in CDM setups).

A possible way to address the missing satellite problem is to keep the DM particles in kinetic equilibrium with a relativistic species until the Universe cooled down to keV temperatures. This would lead to dark acoustic oscillations of density fluctuations below sub-Mpc scales and consequently to the suppression of the abundance of satellite galaxies [18–27]. A possibility for addressing the core-cusp problem is to have a large DM self-scattering cross section of the order of  $\sigma/m_{\text{DM}} \sim 1 \text{ cm}^2 \text{ g}^{-1}$  on subgalactic scales (with a typical rotation velocity  $v_0 \sim 30 \text{ km/s}$ ) [28]. This cross section must presumably decrease with velocity to allow for a smaller impact on galaxy cluster scales (with  $v_0 \sim 10^3 \text{ km/s}$ ) where no deviations from CDM predictions are observed [29]. Such strongly self-interacting dark matter (SIDM) particles would also be more sensitive to gravitational feedback from baryons, enabling them in

\*tobias.binder@theorie.physik.uni-goettingen.de

†michael.gustafsson@theorie.physik.uni-goettingen.de

‡akamada@ibs.re.kr

§stefan.marinus.rodrigues.sandner@desy.de

||mwiesner@th.physik.uni-bonn.de

addition to explain the diversity in rotation curves of dwarf galaxies [13,14].

It is interesting that all this phenomenology can arise naturally in simple three-particle models, where a light force mediator ( $\phi$ ) induces both the desired strong DM ( $\chi$ ) self-interaction and late DM kinetic decoupling from a thermal background particle ( $l$ ). The new force mediator could be either a vector or a scalar boson, both giving Yukawa potentials with proper velocity dependent  $\chi - \chi$  scattering. These SIDM setups can then produce the observed DM abundance via standard thermal freeze-out.

It has recently been shown that these particular types of SIDM models are strongly constrained if the mediator dominantly decays to visible standard model (SM) particles: the vector mediator setup is in tension with indirect detection experiments and CMB observations [30–32]; and most of the parameter space for scalar mediator setups are ruled out by direct detection bounds and big bang nucleosynthesis (BBN) data [33].

A still perfectly valid setup exists if the dark sector is essentially closed. That is, with the mediator being a singlet under the SM gauge groups, the above constraints are clearly avoided. Meanwhile the DM self-interaction properties all remain and the thermal DM freeze-out would occur from a dark radiation background.

In this work, we investigate a novel cosmological signature of these SIDM models to discriminate between vector and scalar boson mediator setups—in the case of a fully closed dark sector. With a light mediator present, Sommerfeld enhanced cross sections are expected with particularly strong enhancements possible for small DM particle velocities. For  $s$ -wave annihilation, exclusively present in the vector mediator case, a *second epoch of annihilation* [34–37] can occur after DM kinetically decouples from its thermal background.<sup>1</sup> Here, we explore for the first time the cosmological consequences of such a reannihilation period in more detail by setting up the required coupled Boltzmann and Hubble expansion equations. We show how this impacts the Hubble parameter significantly enough to be constrained by existing CMB data. In the SIDM region, we interestingly find that reannihilation allows us to ameliorate discrepancies in the CDM paradigm between CMB and low-redshift astronomical measurements of the Hubble rate and  $\sigma_8$  [39–46].

The article is structured as follows. In Sec. II we review the closely degenerate phenomenology of the vector and scalar mediator SIDM setups. In Sec. III, analytic estimates and the formalism to describe reannihilation are set up. A scan over model parameters are performed in Sec. IV and the regions where reannihilation can happen are discussed.

<sup>1</sup>DM reannihilation can also occur if the Hubble expansion rate rapidly drops soon after the DM freeze-out, e.g., in scalar-tensor quintessence models [38], while in our case reannihilation is driven purely by the strong enhancement in annihilation rate with decreasing DM velocity.

We investigate the impact on cosmology and constraints from CMB observation in Sec. V for two different scenarios: reannihilation taking place before and after recombination.

## II. LIGHT MEDIATOR SETUPS AND THEIR PHENOMENOLOGY

In this work, we will consider two effective three-particle models, each of them having four free parameters in their Lagrangians,

$$\mathcal{L}_V \supset g_\chi \bar{\chi} \gamma^\mu \chi \phi_\mu + g_l \bar{l} \gamma^\mu l \phi_\mu, \quad (1)$$

$$\mathcal{L}_S \supset g_\chi \bar{\chi} \chi \phi + g_l \bar{l} l \phi, \quad (2)$$

where  $g_\chi$  and  $g_l$  are the coupling constants. The DM particle  $\chi$  is a spin 1/2 Dirac fermion with mass  $m_\chi$ , and the vector  $\phi_\mu$  or scalar  $\phi$  mediator has a mass of  $m_\phi \ll m_\chi$ . The dark-radiation background particle  $l$  has spin 1/2 and is considered to be massless.

Assuming that  $l$ ,  $\phi$ , and  $\chi$  form a dark sector, which effectively decouples from the SM plasma, leads to an additional free parameter, namely, the temperature ratio,

$$r \equiv \frac{T_l}{T_\gamma}, \quad (3)$$

where  $T_l$  is the dark radiation temperature and  $T_\gamma$  is the SM photon temperature. Fixing this temperature ratio at a particular time, e.g., at the temperature  $T_\gamma^{\text{BBN}} = 1$  MeV, its further temperature dependence is given from entropy conservation as

$$r(T_\gamma) = r_{\text{BBN}} \left( \frac{g_s(T_\gamma)}{g_s(T_\gamma^{\text{BBN}})} \right)^{1/3}, \quad (4)$$

where  $g_s$  is the SM entropy degrees of freedom and we assume that entropy production in the dark sector can be ignored after DM chemically decouples. In this work we choose, unless quoted differently,  $r_{\text{BBN}} = 0.5$ , which turns into  $r \simeq 0.35$  after electron-positron decoupling and is compatible with current BBN constraints [47,48]. Such ratios are achieved if the dark sector kinetically decouples from the SM plasma above a temperature of  $T_\gamma \simeq 40$  GeV. For this work it is, however, not required to specify the coupling to the SM leading to kinetic equilibration between the two sectors. Temperature ratios of this order could also be achieved by some inflationary models.

In the rest of this section we will highlight the similarities and differences between the two models and present the phenomenological results to be used in subsequent sections.

### A. Velocity dependent self-interactions

DM self-interactions lead to an isothermal DM velocity distribution in the inner region of halos. If the self-scattering cross section is of the order of  $\sigma/m_{\text{DM}} \sim 1 \text{ cm}^2 \text{ g}^{-1}$  the DM density distribution in dwarf galaxies is characterized by a kpc-sized core. This mechanism enables one to mitigate the core-cusp and the too-big-to-fail problems [49]. SIDM alone, however, cannot explain the observed diversity of dwarf galaxy rotation curves since it changes the density profile universally among similar-size halos. Rotation velocities in the inner region are observed to differ by up to a factor of  $\simeq 4$  among halos with a similar rotation velocity in the outer region. This diversity was not predicted by state-of-the-art hydrodynamic simulations taking into account baryonic feedbacks such as galaxy formation and supernova explosions [16]. A key observation is, however, that the SIDM profile is quite sensitive to the presence of the baryonic bulge and disk in the inner part of a galaxy. An isothermal DM velocity distribution is determined by the total gravitational potential, which in the inner region can be dominated by the galaxy's baryonic content. Together with the measured baryon distribution, SIDM is able to address the observed diversity in dwarf galaxy rotation curves [13,14].

A thermalized DM halo, on the other hand, may be incompatible with observations of galaxy clusters. Its distribution is virtually spherical, but a strong lens system prefers a sizable ellipticity of the lens galaxy cluster [50]. While the projection effect in the lens analysis is subject to caveat, the constraint would be as severe as  $\sigma/m_{\text{DM}} \lesssim 0.1 \text{ cm}^2 \text{ g}^{-1}$  [51]. A merging cluster system like a bullet cluster also provides a good test for SIDM. The reported tight constraint is  $\sigma/m_{\text{DM}} \lesssim 0.7 \text{ cm}^2 \text{ g}^{-1}$  [52]; otherwise a sizable amount of DM mass evaporates from the subclusters during the collision and the resultant system is incompatible with the observed mass-to-light ratios. One may have to be careful about the uncertainty in the unconstrained initial condition of the system. Although it is too early to conclude (see, e.g., Ref. [53] for a comprehensive summary), the velocity dependence may have to be introduced into the self-scattering cross section to reduce the effects of SIDM in galaxy clusters while keeping a sizable cross section in dwarf galaxies. The desired velocity dependence can naturally be realized in both the light vector and scalar mediator setups of Eqs. (1) and (2) [54,55].

The averaged self-scattering cross section in a thermal DM halo with a characteristic velocity  $v_0$  can be computed from

$$(\sigma_T)_{v_0} = \frac{4\pi}{(\sqrt{2\pi}v_0)^3} \int \sigma_T e^{-v_{\text{rel}}^2/(2v_0^2)} v_{\text{rel}}^2 dv_{\text{rel}}, \quad (5)$$

where  $\sigma_T$  is the transfer cross section:

$$\sigma_T \equiv \int d\Omega (1 - \cos \theta) \frac{d\sigma_{XX \rightarrow XX}}{d\Omega}. \quad (6)$$

For the Yukawa-potential scattering, induced by our light mediators, we will use the ETHOS fitting functions for  $\sigma_T$  in the classical regime ( $m_\chi v_{\text{rel}} \gtrsim m_\phi$ )—as they are provided in Eqs. (45) and (46) of Ref. [24] and originally proposed in Ref. [54]. In the parameter region where  $s$ -wave scattering is dominant (quantum-resonant regime), we will use the analytic expression provided in Eq. (A5) of Ref. [54]. We assume DM to be symmetric and average over particle and antiparticle scattering contributions as suggested in Ref. [24]. In Sec. IV we use these expressions to find the model parameter region where  $(\sigma_T)_{30 \text{ km/s}}/m_\chi \in [0.1, 10] \text{ cm}^2 \text{ g}^{-1}$  (sizeable self-scattering cross section on dwarf galactic scales).

We remark that neither the ETHOS fitting functions nor the numerical solution of the scattering amplitude in Ref. [54] accounts for the correct quantum statistics in their computation of  $\sigma_T$ . They rely on classical assumptions such as the distinguishability of DM particles. It is hard to realize proper quantum corrections in SIDM  $N$ -body simulations, but when adopting a classical treatment it is at least important to keep track of the expected theoretical uncertainties. In Appendix A we compare the commonly adopted classical approximation of  $\sigma_T$  to the proper quantum mechanical treatment [33] for all the scattering possibilities of our mediator setups. We find that there are up to a factor of 2 corrections on dwarf galactic scales between these two approaches; however, they have no relevant impact on our results.

### B. Dark acoustic oscillations

The coupling between nonrelativistic DM particles and radiation leads to competition between gravity and radiative pressure. The pressure effect is strong during kinetic equilibrium between DM and relativistic particles, leading to *dark acoustic oscillations* of the DM density perturbations inside the causal horizon. Therefore, matter density fluctuations can only grow significantly after DM kinetically decouples. This leads to the fact that the resultant matter power spectrum is suppressed on length scales shorter than the Hubble horizon distance at kinetic decoupling. The minimal protohalo mass (or cutoff mass) below which the halo mass function is suppressed can be estimated by the mass inside a sphere with the radius of the Hubble horizon at the time of DM kinetic decoupling:

$$\begin{aligned} M^{\text{cut}} &\equiv \rho_m \frac{4\pi}{3} \left( \frac{1}{H} \right)^3 \Big|_{\text{DM kinetic decoupling}} \\ &= 2.2 \times 10^8 r_{\text{kd}}^3 \left( \frac{1 \text{ keV}}{T_1^{\text{kd}}} \right)^3 M_\odot. \end{aligned} \quad (7)$$

Here,  $\rho_m$  is the total matter density and  $H$  is the Hubble expansion rate. We see that a kinetic decoupling

temperature  $T_l^{\text{kd}}$  of the order of 1 keV leads to the suppression of the halo mass function on dwarf-galaxy masses and hence addresses the missing satellite problem. Especially in our case of late kinetic decoupling and nonrelativistic DM, this damping dominates over the free-streaming effect. The damping mechanism of dark acoustic oscillations has been extensively investigated by many authors [18–27, 56–68]—in part also in the context of SIDM. For a classification of DM models leading to late kinetic decoupling we refer readers to Ref. [69].

In our setups both the mediator  $\phi$  and the fermionic particle  $l$  can act as pressure sources leading to dark acoustic oscillations. In the parameter space we will consider, the scattering between DM  $\chi$  and  $l$  dominates over that between  $\chi$  and  $\phi$ . The kinetic decoupling temperature  $T_l^{\text{kd}}$  can be defined as the temperature when the Hubble expansion rate  $H$  equals the momentum transfer rate  $\gamma$ .<sup>2</sup> In Refs. [24, 27] the momentum transfer rate is derived to be

$$\gamma = \frac{1}{6m_\chi T_l} \sum_{s_l} \int \frac{d^3 \mathbf{p}_l}{(2\pi)^3} f_l^{\text{eq}} (1 - f_l^{\text{eq}}) \int_{-4\mathbf{p}_l^2}^0 dt(-t) \frac{d\sigma}{dt} v_{\text{rel}}, \quad (8)$$

and explicit expressions of the elastic  $l - \chi$  scattering cross section  $d\sigma/dt v_{\text{rel}}$  and the kinetic decoupling temperature can be found in Refs. [27, 69]. This momentum transfer rate of  $l - \chi$  scattering scales as  $\gamma \propto T_l^6$  for both scalar and vector mediators. Furthermore, both scenarios acquire the same minimal halo mass  $M^{\text{cut}}$  for similar coupling constants and particle masses [27]. The suppression of the halo mass function mainly depends on the mediator mass  $m_\phi$  and for both models a cutoff mass around the dwarf galactic scale can be achieved for  $m_\phi$  of the order of 1 MeV.

The predicted matter power spectra for scalar and vector mediators differ slightly in shape due to differences in the angular dependence of their  $\chi - l$  scattering amplitudes [24, 25]. As a consequence, the two models are in principle distinguishable, but in a recent study [72] it was shown that the differences are too small to be seen in current CMB observations. Future observations of CMB spectral distortions, however, might be sensitive enough to discriminate models where DM is kept in kinetic equilibrium via SM photon scattering and where the DM kinetic equilibrium is kept via SM neutrino scattering [73].

Recently, a combined Ly- $\alpha$  forest data analysis [74] constrained the damping of the matter power spectrum due to the free-streaming effect of warm dark matter (WDM). The lower limit on the WDM mass can be approximately translated into a lower limit of the kinetic decoupling

<sup>2</sup>For a more precise definition of a kinetic decoupling temperature and its matching to the nonlinear cutoff in the matter-power spectrum, see Refs. [24, 25, 70, 71].

temperature by equating the suppressed matter power spectra in a certain range of wavelengths. The authors of Ref. [75] found with this estimate a lower limit of  $T_l^{\text{kd}}/r \gtrsim 1$  keV (0.6 keV),<sup>3</sup> which according to Eq. (7) results in an upper limit on the cutoff mass of

$$M^{\text{cut}} \lesssim 2 \times 10^8 M_\odot (10^9 M_\odot). \quad (9)$$

Cutoff masses in the range  $10^7$  to  $10^9 M_\odot$  are indicated in Fig. 4 of Sec. IV, where  $M^{\text{cut}}$  is a function of our model parameters as determined from Eqs. (3.14) in Ref. [27].

### C. Sommerfeld enhanced annihilation

The two models differ in the leading-order cross section results of their DM annihilation channels:  $\chi\bar{\chi} \rightarrow \phi\phi$  and  $\chi\bar{\chi} \rightarrow \bar{l}l$ . In the vector mediator case, both processes are  $s$ -wave dominated and in the scalar mediator case they are  $p$ -wave dominated. Generically for both models, the annihilation rate is Sommerfeld enhanced in the DM nonrelativistic regime, and for each particular annihilation channel the cross section factorizes into a short and a long-range contribution:

$$(\sigma v_{\text{rel}})_V \simeq S_s(v_{\text{rel}}) \sum_i (\sigma v_{\text{rel}})_{0,i}^s, \quad (10)$$

$$(\sigma v_{\text{rel}})_S \simeq S_p(v_{\text{rel}}) \sum_i (\sigma v_{\text{rel}})_{0,i}^p, \quad (11)$$

The long-range force corrections are encoded in the velocity dependent Sommerfeld factor  $S(v_{\text{rel}})$  multiplying *universally* the tree-level cross section  $(\sigma v_{\text{rel}})_{0,i}$  for each particular annihilation channel  $i$ . In particular, the vector mediator model has

$$(\sigma v_{\text{rel}})_{0,\phi\phi}^s \simeq \frac{\pi \alpha_\chi^2}{m_\chi^2}, \quad (12)$$

$$(\sigma v_{\text{rel}})_{0,\bar{l}l}^s \simeq \frac{\pi \alpha_\chi \alpha_l}{m_\chi^2}, \quad (13)$$

where  $\alpha_{\chi(l)} \equiv g_{\chi(l)}^2/4\pi$  and in the scalar mediator case the corresponding tree-level cross sections are instead  $(\sigma v_{\text{rel}})_{0,i}^p \propto v_{\text{rel}}^2$ .

$S(v_{\text{rel}})$  can be obtained from the two DM particles' wave-function value at the interaction point by numerically solving their Schrödinger equation with the potential resulting from  $t$ -channel exchanges of the light mediator [76–79]. In the static limit, both mediator types induce a Yukawa potential. It was shown in Ref. [80] that the

<sup>3</sup>The strong (the weak) limit is derived from the Lyman- $\alpha$  measurements in Ref. [74], where a power-law assumption (a free floating value) is used to describe the redshift evolution of the intergalactic medium temperature.

Sommerfeld factor resulting from the Hulthén potential describes very accurately the numerical solution of the Schrödinger equation with a Yukawa potential. The advantage of the Hulthén potential is that analytic solutions of  $S$  for  $s$ - and  $p$ -waves are known. The expression of  $S$  for  $s$ -wave annihilation is given by (see, e.g., Ref. [80])

$$S_s(v_{\text{rel}}) = \frac{\pi}{\epsilon_v} \frac{\sinh\left(\frac{12\epsilon_v}{\epsilon_\phi\pi}\right)}{\cosh\left(\frac{12\epsilon_v}{\epsilon_\phi\pi}\right) - \cos\left[2\pi\sqrt{\frac{6}{\epsilon_\phi\pi^2} - \left(\frac{6\epsilon_v}{\epsilon_\phi\pi^2}\right)^2}\right]}, \quad (14)$$

where the two dimensionless parameters are defined as

$$\epsilon_v \equiv \frac{v_{\text{rel}}}{2\alpha_\chi}, \quad (15)$$

$$\epsilon_\phi \equiv \frac{m_\phi}{\alpha_\chi m_\chi}. \quad (16)$$

The Sommerfeld factor in Eq. (14) is resonantly enhanced if the parametric condition,

$$\epsilon_\phi = \frac{6}{n^2\pi^2} \quad \text{with } n \in \mathbb{Z}^+, \quad (17)$$

is fulfilled. The position of the  $n$ th ‘‘Sommerfeld resonance’’ is the same for  $s$ - and  $p$ -wave annihilation (except for  $n = 1$  where the resonance is absent in  $S_p$ ). We will refer to this equation as the ‘‘parametric resonance condition.’’ For this work, the most important difference is that only in the case of  $s$ -wave annihilation, the total cross section scales as

$$(\sigma v_{\text{rel}})_V \propto v_{\text{rel}}^{-2} \quad \text{for } v_{\text{rel}} \lesssim m_\phi/m_\chi, \quad (18)$$

when  $\epsilon_\phi$  is close to a resonance condition. For  $p$ -waves, the cross section is constant in this regime and never scales stronger than  $v^{-1}$ . The  $v^{-2}$  feature of  $(\sigma v_{\text{rel}})_V$  is thus only available in the vector mediator model.

The implications of the strongly velocity dependent enhancement in Eq. (18) are the main part of this work. As we will see, it can lead to a reannihilation period where the comoving DM number density significantly decreases a second time. What is important to note is that the analytic formula of the Sommerfeld factor as given in Eq. (14) can violate the  $s$ -wave unitarity bound for low velocities when the parametric resonance condition is exactly (or almost) fulfilled. In the numerical analyses in the subsequent sections we will therefore always use the improved analytic solution provided in Ref. [81], accounting for a physical behavior on top of a resonance and correcting the approximate expression in Eq. (14) for extremely low relative velocities. In Appendix B, we provide the details of this improved analytic formula and discuss the important role of saturation of  $(\sigma v_{\text{rel}})_V$  below the unitarity limit.

### III. AN EPOCH OF REANNIHILATION

In the previous section we put emphasis on the very similar phenomenology of the two light mediator models in Eqs. (1) and (2): they are practically identical candidates for alleviating multiple small-scale structure formation issues in a comparable model parameter space. In the following, we point out that even in the case of not including any couplings at all to SM particles and therefore ‘‘hiding’’ the dark sector, the *impact on cosmology at late times can be significantly different*. More precisely, we show that only in the vector mediator case can a strong Sommerfeld enhancement, such as in Eq. (18), lead to a second period of annihilation.

In Fig. 1 an example of a reannihilation epoch is shown. After kinetic decoupling the DM abundance decreases by 1 order of magnitude before the time of matter-radiation equality. The final  $\chi$  relic abundance coincides with the observed CDM value,  $\Omega_c h^2 = 0.1197$  [central value of ‘‘Planck 2015 (TT + lowP)’’ analysis [1]]. In Sec. III A, we provide analytic estimates and an intuitive understanding of when and in which region of the parameter space of the vector mediator model reannihilation can happen.

The reannihilation process necessarily starts after kinetic decoupling, as in the example of Fig. 1. During this process, the evolution of the DM temperature  $T_\chi$  does not follow the typical  $T_\chi \propto T_l^2$  scaling for kinetically decoupled nonrelativistic particles, since the Sommerfeld enhancement leads to a strongly velocity dependent annihilation cross section. The DM particles preferably annihilate at low momenta,

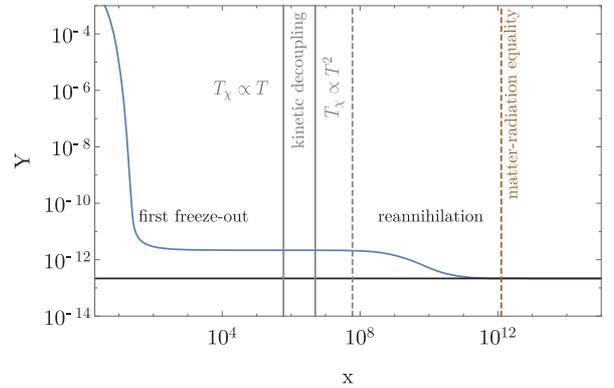


FIG. 1. Reannihilation process shown as a function of  $x \equiv m_\chi/T_\gamma$ , changing the DM comoving number density  $Y \equiv n_\chi/s$  by 1 order of magnitude. The final abundance coincides with the correct value (black horizontal line). Here, we have chosen the parameters  $m_\chi = 1$  TeV,  $\alpha_\chi = 0.007$ ,  $m_\phi \simeq 1$  GeV, and the resonance number  $n \simeq 2$  (where  $m_\phi$  is tuned to get the correct relic density). Between the gray solid lines kinetic decoupling happens and the scaling of the DM temperature changes from  $T_\chi \propto T$  to  $T_\chi \propto T^2$ . The dashed gray line indicates the start of reannihilation, where velocity-dependent annihilation leads to deviation from the  $T_\chi \propto T^2$  scaling.

which leads to an increase of the DM temperature and in turn influences the annihilation rate. It requires one to go beyond the standard way of describing the DM number density evolution [82] to cover kinetic decoupling and the impact of DM velocity dependent annihilation on the DM temperature. In Sec. III B we adopt the method developed in Ref. [37] (and first estimated in Ref. [35]) of how to deal with the number density computation in such a case correctly. We further extend the set of equations by including the impact of the injected dark radiation on the expansion rate. Moreover, we provide a reliability check of the method proposed in Ref. [37] by solving the Boltzmann equation at phase-space density level.

### A. Estimates

To analytically quantify if and when DM reannihilation happens, we study the ratio between the annihilation and expansion rates,

$$\Gamma \equiv \frac{\langle \sigma v_{\text{rel}} \rangle_{x'} Y}{H/s}, \quad (19)$$

where the dimensionless form of the DM number density  $n_\chi$  is defined as

$$Y \equiv \frac{n_\chi}{s}. \quad (20)$$

Here, the SM entropy  $s = g_s(2\pi^2/45)T_\gamma^3$  and the Hubble expansion rate,  $H \propto \sqrt{g_{\text{eff}}}T_\gamma^2$  during radiation domination and  $\propto T_\gamma^{3/2}$  during matter domination, are both dynamical functions of

$$x \equiv \frac{m_\chi}{T_\gamma}. \quad (21)$$

We follow the evolution of  $\Gamma$  after the first freeze-out, so that we can assume  $Y$  to be constant until the start of reannihilation. The thermally averaged cross section is a function of the DM temperature  $T_\chi$  and can be written in the nonrelativistic limit as

$$\langle \sigma v_{\text{rel}} \rangle_{x'} = \frac{(x')^{3/2}}{2\sqrt{\pi}} \int_0^\infty (\sigma v_{\text{rel}}) e^{-x'v_{\text{rel}}^2/4} v_{\text{rel}}^2 dv_{\text{rel}}, \quad (22)$$

where it is a function of

$$x' \equiv \frac{m_\chi}{T_\chi}. \quad (23)$$

We note that for a cross section of the form of  $(\sigma v_{\text{rel}}) \propto v_{\text{rel}}^{-n}$ , where we consider here only  $n = 0, 1, 2$ , the thermally averaged cross section can be computed analytically and scales as  $\langle \sigma v_{\text{rel}} \rangle_{x'} \propto x'^{-n/2}$ .

To now estimate the scaling of  $\Gamma$  as a function of  $x$ , we approximate the kinetic decoupling as an instantaneous process such that we can write

$$x' = \begin{cases} \frac{m_\chi}{T_i} = \frac{x}{r} & \text{before kinetic decoupling,} \\ \frac{m_\chi T_i^{\text{kd}}}{T_\chi^2} = \frac{x^2}{r^2 x_i^{\text{kd}}} & \text{after kinetic decoupling.} \end{cases} \quad (24)$$

Here,  $x_i^{\text{kd}} \equiv m_\chi/T_i^{\text{kd}}$  and the dynamical temperature ratio  $r$  is defined in Eq. (4). The exact evolution of  $x'$ , beyond the approximation of instantaneous kinetic decoupling, is a part of Sec. III B.

We provide in Table I the scaling of  $\Gamma(x)$  for different types of velocity dependent cross sections in the instantaneous kinetic decoupling approximation. Let us discuss some of its entries in the temporal order of the example scenario shown in Fig. 1. After chemical decoupling, where  $\Gamma$  drops below 1,  $\Gamma$  scales as  $x^{-1}$  until the Sommerfeld factor (or the total  $s$ -wave annihilation cross section) starts to dominantly scale as  $S \propto 1/v_{\text{rel}}$ . From this point to kinetic decoupling  $\Gamma$  further decreases in the phase of the  $S \propto 1/v_{\text{rel}}$  scaling, followed by a period where  $\Gamma$  stays constant. After kinetic decoupling, when  $T_\chi$  starts to drop quickly, the  $S \propto 1/v_{\text{rel}}^2$  scaling dominates and leads to an increase of  $\Gamma$  as is highlighted by boldface  $x$  in the table. When  $\Gamma$  starts to approach 1 again, the DM abundance significantly decreases a second time as seen in Fig. 1. The reannihilation process stops when the Sommerfeld enhancement is saturated, finally leading to  $(\sigma v_{\text{rel}}) \propto \text{const}$  and to a fast

TABLE I. Evolution of  $\Gamma$  at different cosmological epochs and for different DM annihilation cross sections  $(\sigma v_{\text{rel}})$ . Only after kinetic decoupling and with a cross section scaling as  $1/v^2$  can the ratio of annihilation over expansion rate,  $\Gamma$ , grow to enter a period of reannihilation. Sommerfeld enhanced  $s$ -wave annihilation features such a  $1/v^2$  scaling.  $r$  is the ratio of dark radiation to photon temperature, as defined in Eq. (4).

$(\sigma v_{\text{rel}}) \propto$	$(\sqrt{g_{\text{eff}}}/g_s) \times \Gamma \propto$			
	Radiation-dominated epoch		Matter-dominated epoch	
	<i>Before</i> kinetic decoupling	<i>After</i> kinetic decoupling	<i>Before</i> kinetic decoupling	<i>After</i> kinetic decoupling
Constant	$x^{-1}$	$x^{-1}$	$x^{-3/2}$	$x^{-3/2}$
$1/v_{\text{rel}}$	$r^{-1/2} x^{-1/2}$	$r^{-1}$	$r^{-1/2} x^{-1}$	$r^{-1} x^{-1/2}$
$1/v_{\text{rel}}^2$	$r^{-1}$	$r^{-2} \mathbf{x}$	$r^{-1} x^{-1/2}$	$r^{-2} \mathbf{x}^{1/2}$

decrease of  $\Gamma$ —as can be read off from Table I. The saturation velocity depends on how parameters combine to the resonance condition, given in Eq. (17). For Fig. 1 we have chosen a point slightly next to a resonance, such that the saturation effect gives the correct value of the relic density. If exactly on top of the resonance, the reannihilation process would have been longer, further reducing the abundance of DM. In general, the evolution pattern of  $\Gamma$  can vary depending on the model parameters.

An important quantity, used in the following sections, is the redshift  $z_{\text{rea}}$  defined by when the DM comoving number density changes first by 1% due to reannihilation.  $z_{\text{rea}}$  can be determined from the value of  $\Gamma$  and a detailed derivation is provided in Appendix C. For our vector mediator model, we find that the onset of the reannihilation process is roughly given by

$$z_{\text{rea}} \simeq 100 \times \left(\frac{r_{\text{kd}}}{0.36}\right)^{-5} \left(\frac{\alpha_\chi}{0.02}\right) \left(\frac{m_\chi}{\text{TeV}}\right)^{-3/2} \left(\frac{m_\phi}{1.2 \text{ MeV}}\right)^4, \quad (25)$$

where we have assumed  $\alpha_\chi = \alpha_l$  and that  $r$  does not change after kinetic decoupling. From this equation it can be recognized that the onset of reannihilation in the matter dominated epoch has a strong dependence on the temperature ratio and the mediator mass. Note that in the parameter region around the reference values in Eq. (25), cutoff masses of the order of  $10^8 M_\odot$  and sizable self-interactions on dwarf galactic scales can be achieved simultaneously. Strictly speaking, the simple power-law scaling in Eq. (25) is only valid for  $z_{\text{rea}} \ll z_{\text{eq}}$ , where  $z_{\text{eq}} \simeq 3400$  is the matter-radiation equality redshift [1], and when the first freeze-out is not significantly affected by Sommerfeld corrections. We discuss a more general expression for  $z_{\text{rea}}$  in Appendix C 3 that will later be used in Sec. IV to identify the parameter region where reannihilation happens after recombination.

Another region of interest to identify is where reannihilation stops in the radiation dominated era, because here a change in the DM abundance has in general less impact on, e.g., the Hubble expansion rate. This situation occurs if the saturation temperature  $T_\gamma^{\text{sat}}$  is higher than the matter-radiation equality temperature  $T_\gamma^{\text{eq}} = 0.80 \text{ eV}$ . The saturation temperature  $T_\gamma^{\text{sat}}$  as a function of the free parameters is derived in Eq. (C17). From this equation it can be read off that the minimum value of  $T_\gamma^{\text{sat}}$  is given by the minimum  $\alpha_\chi$  value that can give a resonance. This occurs when  $n = 1$  in Eq. (17) and is given by

$$\alpha_\chi^{\text{min}} = \frac{\pi^2 m_\phi}{6 m_\chi}. \quad (26)$$

Inserting this into the result of the saturation temperature [Eq. (C17)], we find

$$\begin{aligned} T_\gamma^{\text{sat}} &\gtrsim T_\gamma^{\text{sat, min}} \\ &= 0.6 \text{ eV} \left(\frac{r_{\text{sat}}}{0.36}\right)^{-1} \left(\frac{m_\phi}{2 \text{ GeV}}\right)^4 \left(\frac{m_\chi}{\text{TeV}}\right)^{-7/2} \left(\frac{T_l^{\text{kd}}}{\text{MeV}}\right)^{1/2}, \end{aligned} \quad (27)$$

and in the case of  $\alpha_\chi = \alpha_l$  the kinetic decoupling temperature in terms of the minimum coupling is given by

$$T_l^{\text{kd}} = 1 \text{ MeV} \left(\frac{r_{\text{kd}}}{0.5}\right)^{-1/2} \left(\frac{m_\chi}{\text{TeV}}\right)^{3/4} \left(\frac{m_\phi}{2 \text{ GeV}}\right)^{1/2}. \quad (28)$$

Note that the result in Eq. (27) is quite general and can be used to estimate the parameter region where one does not expect to have reannihilation below a certain temperature. It is independent of the physics happening before kinetic decoupling and only assumes that the maximum enhancement is given by the  $s$ -wave unitarity bound and that the saturation temperature is lower than the kinetic decoupling temperature.

Even though we focus on a vector mediator model here, any DM setup where  $s$ -wave annihilation is Sommerfeld enhanced via a Yukawa potential can lead to an epoch of reannihilation. Or, more general, any DM model where the total cross section scales as  $(\sigma v_{\text{rel}}) \propto v_{\text{rel}}^{-1-\epsilon}$ , with  $\epsilon > 0$ , can lead to an epoch of reannihilation. This excludes, in particular,  $p$ -wave annihilation or Coulomb potentials to have the feature of a reannihilation epoch.

## B. Numerical methods

In the previous sections, we established when reannihilation can start and for how long it can last. We now turn to investigate its exact impact on the DM relic density and the Hubble expansion. To track the DM number density [Eq. (20)] and the injected energy density evolution during reannihilation, we set up the following coupled differential equations:

$$\frac{dY}{dx} = -\frac{s}{\tilde{H}x} \langle \sigma v_{\text{rel}} \rangle_y Y^2, \quad (29)$$

$$\frac{dy}{dx} = -\frac{2\gamma}{\tilde{H}x} [y - y^{\text{eq}}] + \frac{s}{\tilde{H}x} y Y [\langle \sigma v_{\text{rel}} \rangle_y - \langle \sigma v_{\text{rel}} \rangle_{y,2}], \quad (30)$$

$$\frac{dY_l}{dx} = -\frac{H}{\tilde{H}x} Y_l + \frac{s}{\tilde{H}x} \langle \sigma v_{\text{rel}} \rangle_y Y^2, \quad (31)$$

where we have defined the dimensionless temperatures as

$$y \equiv \frac{m_\chi T_\chi}{s^{2/3}}, \quad (32)$$

$$y^{\text{eq}} \equiv \frac{m_\chi T_l}{s^{2/3}} = \frac{m_\chi r T_\gamma}{s^{2/3}}. \quad (33)$$

The energy density of the injected dark radiation, given by

$$Y_l \equiv \frac{\rho_l}{m_\chi s}, \quad (34)$$

is fully included in the Hubble expansion rate

$$H^2 = \frac{8\pi G}{3} [\rho_\gamma + \rho_\nu + \rho_b + \rho_{\text{dark}} + \rho_\Lambda], \quad (35)$$

where the total dark sector energy density is given by

$$\rho_{\text{dark}} = 2m_\chi s(Y + Y_l), \quad (36)$$

with the factor of 2 originating from the sum of DM particle and antiparticle contributions.  $\tilde{H}$  is defined as

$$\tilde{H} = \frac{H}{1 + \frac{1}{3} \frac{T}{g_s} \frac{dg_s}{dT}}, \quad (37)$$

where the evolution of the SM's entropy degrees of freedom  $g_s$  and the effective number of relativistic degrees of freedom  $g_{\text{eff}}$  are taken from Ref. [83].

Equations (29) and (30) can be derived from the Boltzmann equation in the limit of nonrelativistic DM particles

$$m_\chi (\partial_t - H\mathbf{p} \cdot \nabla_{\mathbf{p}}) f_\chi = C_{\chi\chi \rightarrow \phi\phi, ll}^{\text{non-rel}} [f_\chi] + C_{\chi l \leftrightarrow \chi l}^{\text{non-rel}} [f_\chi, f_l^{\text{eq}}], \quad (38)$$

by taking the zeroth

$$n_\chi = g \int \frac{d^3 p}{(2\pi)^3} f_\chi, \quad (39)$$

and the second moment with respect to momentum

$$T_\chi = \frac{g}{3n_\chi} \int \frac{d^3 p}{(2\pi)^3} \frac{p^2}{m_\chi} f_\chi, \quad (40)$$

respectively.  $g$  is the DM internal spin degrees of freedom. Kinetic decoupling from dark radiation is taken into account by the first term in Eq. (30) and the impact of annihilation on the DM temperature by the last term in the same equation. Equations (29) and (30) were derived for the first time in Ref. [37] and can also be obtained by taking the nonrelativistic limit of the more general equations as fully derived in Ref. [84]. The equations of the latter work include relativistic corrections and also the production of DM, where both Eqs. (29) and (30) get correction terms. In this work, for late kinetic decoupling, it is evident that both corrections can be neglected. Due to different conventions, the momentum transfer rate  $\gamma$  is here defined to be a factor of 2 smaller than in Ref. [84].

In this work we include for the first time the evolution of the dark radiation governed by Eq. (31) and the impact of reannihilation on the Hubble expansion rate as in Eq. (35). Note that both the direct production of  $l$  and the instantaneous decay of the produced vector mediators  $\phi$  into fermions  $l$  are included in the equations via the *total* averaged cross section  $\langle \sigma v_{\text{rel}} \rangle_y$ . The first term on the right-hand side of Eq. (31) captures redshifting of the injected dark radiation, while the second term covers that all DM annihilations instantaneously transform nonrelativistic DM particles into dark radiation.

In the rest of this work we will for simplicity assume that the two couplings  $g_\chi$  and  $g_l$  of the vector mediator model, as given in Eq. (1), are equal. Order one deviations from this assumption do not influence our analysis, since most of the quantities, such as kinetic decoupling temperature, have a minor dependence on  $g_l$ . Furthermore, in some part of the parameter space bound state formation processes might be relevant [31,85–88], but it is beyond the scope of this work to investigate it further.

In order to be able to evaluate the phase-space averaged cross sections,  $\langle \sigma v_{\text{rel}} \rangle$  and  $\langle \sigma v_{\text{rel}} \rangle_2$  defined as

$$\langle \sigma v_{\text{rel}} \rangle \equiv \frac{g^2}{n_\chi^2} \int \frac{d^3 p d^3 \tilde{p}}{(2\pi)^6} (\sigma v_{\text{rel}}) f_\chi(p) f_\chi(\tilde{p}), \quad (41)$$

$$\langle \sigma v_{\text{rel}} \rangle_2 \equiv \frac{g^2}{T n_\chi^2} \int \frac{d^3 p d^3 \tilde{p}}{(2\pi)^6} \frac{p^2}{3m_\chi} (\sigma v_{\text{rel}}) f_\chi(p) f_\chi(\tilde{p}), \quad (42)$$

one has to make an assumption on the form of the DM phase-space distribution. In the limit of a larger self-scattering rate than the annihilation rate the following form is motivated:

$$f_\chi = \frac{n_\chi(T)}{n_{\chi, \text{eq}}(T_\chi)} \exp\left(-\frac{m_\chi + p^2/(2m_\chi)}{T_\chi}\right) \Big|_{T_\chi = y s^{2/3}/m_\chi}, \quad (43)$$

where the  $T_\chi$  evolution is governed via Eq. (30). This ansatz leads to the final form of  $\langle \sigma v_{\text{rel}} \rangle_y$  given in Eq. (22), and for simplifying the momentum square weighted annihilation cross section  $\langle \sigma v_{\text{rel}} \rangle_{y,2}$  we refer to the result presented in Ref. [37]. Let us point out that in the DM temperature evolution equation [Eq. (30)] the two averaged cross sections appear as the difference  $[\langle \sigma v_{\text{rel}} \rangle_y - \langle \sigma v_{\text{rel}} \rangle_{y,2}]$ . For Sommerfeld enhanced cross sections, this difference is always positive since  $\langle \sigma v_{\text{rel}} \rangle_{y,2}$  has more integral support at higher momenta, where the annihilation cross section is smaller. If Sommerfeld enhanced annihilation is still significant, we therefore expect that  $y$  should increase (DM self-heating) after kinetic decoupling [37]. In Fig. 2, we see that this is indeed the case. Without reannihilation,  $y$  would otherwise remain almost constant after the kinetic decoupling ended just above  $x \sim 10^6$ .

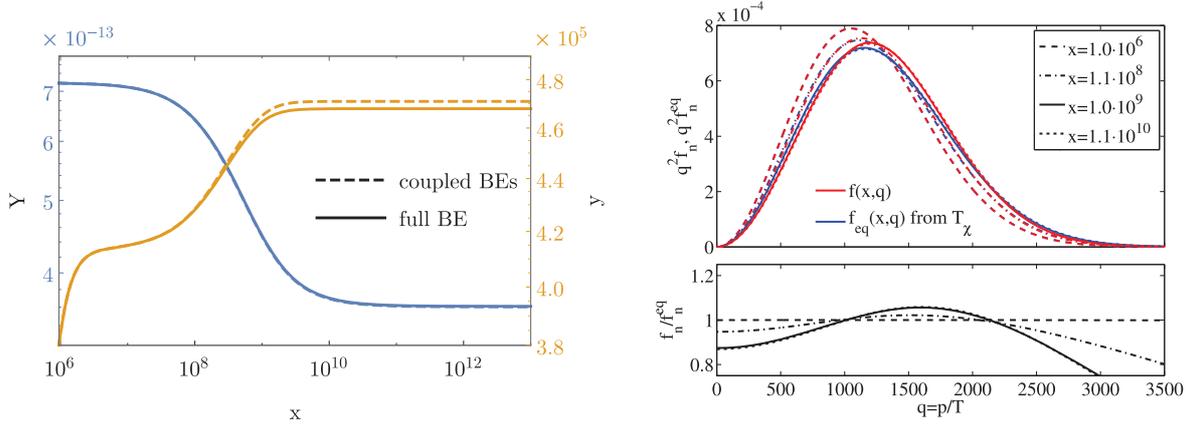


FIG. 2. Time evolution of the DM abundance  $Y \equiv n_\chi/s$ , its temperature  $y \equiv \frac{m_\chi}{s^{2/3}} T_\chi$ , and its phase-space density  $f(q)$  with  $q = p/T_\chi$ . *Left panel:* The evolution of  $Y$  (blue line) and  $y$  (yellow line) in the case of strongly self-interacting DM (dotted lines) and in the case of no DM self-interactions (solid lines). *Right panel:* Unit normalized phase-space distributions  $f_n(q)$  from our full numerical solution of the Boltzmann equation (red lines) compared to thermal equilibrium distributions  $f_n^{\text{eq}}(q)$  with the same “temperature”  $T_\chi$  (blue lines). The phase-space distributions are shown at four different  $x \simeq 10^6$  (solid line),  $10^8$  (dashed line),  $10^9$  (dot-dashed line), and  $10^{10}$  (dotted line). The bottom panel shows the ratio  $f_n(q)/f_n^{\text{eq}}(q)$ . The DM model is  $m_\chi = 600$  GeV,  $m_\phi \simeq 1$  GeV, and  $\alpha_\chi$  chosen such that the relic abundance retains the observed DM abundance after the reannihilation period. Both plot styles are chosen to resemble those in Ref. [84].

It was argued in Refs. [35,37] that the self-scattering rate can potentially drop below the annihilation rate at the time of reannihilation. In this case it is possible that the ansatz in Eq. (43) is not justified and the momentum moment approach might differ from an exact solution of the full Boltzmann equation. In the following, however, we confirm for the first time that the momentum moment approach describes *remarkably well* our reannihilation process of Sommerfeld enhanced annihilation, *even in the limit of zero self-scattering*.

Only in the rest of this section, to compare the momentum moment approach in Eqs. (29) and (30) to a full phase-space density solution of the Boltzmann equation, we set for simplicity  $r = 1$ , the number of relativistic degrees of freedom to be constant and neglect the impact of reannihilation on the Hubble expansion rate. We then follow the approach of Ref. [84], using the dimensionless coordinates

$$x(t, p) \equiv \frac{m_\chi}{T_\gamma}, \quad (44)$$

$$q(t, p) \equiv \frac{p}{T_\gamma}, \quad (45)$$

to rewrite Eq. (38) for the DM phase-space distribution  $f_\chi(x, q)$  as

$$\begin{aligned} \partial_x f_\chi(x, q) = & -\frac{m_\chi^3 g}{H x^4 4\pi^2} \int d\tilde{q} \tilde{q}^2 \int d\cos\theta (\sigma v_{\text{rel}}) f_\chi(q) f_\chi(\tilde{q}) \\ & + \frac{\gamma(x)}{Hx} \left[ x \partial_q^2 + \left( q + \frac{2x}{q} \right) \partial_q + 3 \right] f_\chi, \end{aligned} \quad (46)$$

where  $\theta$  is the angle between the annihilating DM particles’ comoving momenta  $\mathbf{q}$  and  $\tilde{\mathbf{q}}$ . The Fokker-Planck scattering term has an attractor solution, the nonrelativistic Maxwell distribution. This matches the ansatz in Eq. (43) for  $T_\chi = T_\gamma$ .

By adapting the code developed in Ref. [84] (to become public [89]), we solve Eq. (46) and compare its solution to that of Eqs. (29) and (30). The result around the reannihilation period is presented in Fig. 2 for one example model. In the left panel, the solid and dashed blue curves show the DM abundance  $Y$  from solving Eq. (46) and Eqs. (29) and (30), respectively. After a period of reannihilation starting at  $x \sim 10^8$ , where the effective cross section scales as  $1/v^2$ , the reannihilation stops around  $x \sim 10^{10}$  when the Sommerfeld enhancement is saturated and  $(\sigma v_{\text{rel}})$  is effectively constant. In this example, the DM abundance depletes further by 50% during the reannihilation process and converges to the observed CDM relic density. The difference in  $Y$  between the two approaches is less than 1%, and the blue curves are virtually overlapping. So, while it is true that the velocity dependent annihilation cross section acts to heat up DM—as shown by the rise of the yellow lines of  $y$  around  $x \sim 10^8$ – $10^{10}$  after DM kinetically decoupled at  $x \simeq 2 \times 10^6$ —the distortion of  $f_\chi(q)$  from a thermal shape is not large enough to significantly alter the relic abundance result. In the right panel of Fig. 2, we show the resulting shape of  $f_\chi(q)$  (red curves) from the full Boltzmann equation, assuming zero DM self-scattering. If we compare those (red curves) to reference thermal distributions  $f_\chi^{\text{eq}}(q)$  (blue curves) that have the same  $T_\chi$ , we see that there is a distortion at the 10% level from thermal equilibrium distributions for

$q \lesssim 2500$ . However, this has little effect on the relic abundance because during most of the reannihilation period the effective cross section is close to saturation and varies little with  $q$ . The fact that  $f_\chi$  falls below the corresponding thermal distribution  $f_\chi^{\text{eq}}$  at larger  $q$  does not have any practical implications—as the number density in the high momentum tail is negligible. The reason for this  $f_\chi/f_\chi^{\text{eq}}$  suppression is a spurious effect from annihilation at low momenta. The alteration of  $f_\chi$  at low momenta leads to a best-fit thermal distribution  $f_\chi^{\text{eq}}$  with a higher temperature, which in turn is a distribution that has a tail of more large momentum particles. In the following, we will only investigate small changes in  $Y$  and can therefore safely use our system of coupled Eqs. (29)–(31) and (35), which assumes a thermal shape of  $f_\chi^{\text{eq}}$ .

#### IV. PARAMETER SCAN

Reannihilation leads to the fact that multiple values of  $\alpha_\chi$  can give the observed DM abundance for fixed model parameters  $m_\chi$  and  $m_\phi$ . An initial DM overabundance from the first freeze-out due to lower  $\alpha_\chi$  values can be compensated by a second period of annihilation. More precisely this is possible if the three free parameters<sup>4</sup> combine to be close to the parametric resonance condition in Eq. (17). To see this explicitly, we show in Fig. 3 the relic abundance as a function of the DM coupling  $\alpha_\chi$  for fixed values of  $m_\chi$  and  $m_\phi$ . The dashed curve shows the relic abundance relative to the correct value if one ignores reannihilation and computes numerically the evolution of the number density in the standard approach [82]. Clearly, the abundance roughly scales as  $\Omega_\chi h^2 \propto \alpha_\chi^{-2}$ , and there is a unique solution leading to the correct relic abundance at  $\alpha_\chi \simeq 0.021$  in this example. It is also demonstrated how the final DM abundance changes by solving Eqs. (29)–(31) numerically for discrete points where the parametric resonance condition is *exactly* fulfilled (red points). This indicates the *maximal effect* of reannihilation that can be achieved for these  $m_\chi$  and  $m_\phi$  values. The ability to lower  $\Omega_\chi$  by reannihilation is limited either by the saturation of the Sommerfeld enhancement or by the finite age of the Universe (where DM halo formation and dark energy domination eventually also halt the reannihilation period).

The fifth red point from the left is the first resonance that can give the correct relic density. For this point, and the other resonances shown farther to the right, there has to be an  $\alpha_\chi$  in the vicinity of the exact resonance point that reproduces the measured relic density. In fact, there are two  $\alpha_\chi$  possibilities for each of these resonances since  $\Omega_\chi h^2$  is a smooth function of  $\alpha_\chi$  that coincides with the off-resonant result between the resonances. Larger values of  $\alpha_\chi$  than

<sup>4</sup>We remind the reader that we fix  $\alpha_\chi = \alpha_l$  and  $r_{\text{BBN}} = 0.5$  with a temperature dependence of  $r$  as in Eq. (4).

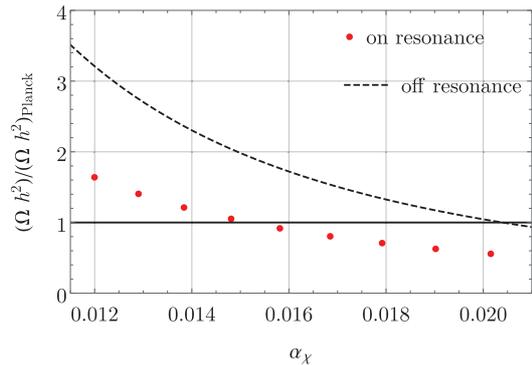


FIG. 3. Relic abundance ratio shown vs the coupling  $\alpha_\chi$  for fixed  $m_\chi = 1$  TeV and  $m_\phi = 10$  MeV. Dashed black curve is the result for only taking the standard thermal freeze-out into account (labeled as off resonance). The red dots present points where the parametric resonance condition is exactly fulfilled and reannihilation thus lowers the relic abundance maximally. Moving left or right from an exact resonance point by changing  $\alpha_\chi$  slightly can lead to  $(\Omega_\chi h^2)/(\Omega_c h^2)_{\text{Planck}} = 1$  but only for the red points that cross the horizontal black line. The relic abundance is therefore degenerate in these (almost) on resonance  $\alpha_\chi$  values.

those shown in the figure do not lead to the correct abundance. To conclude, for given  $m_\chi$  and  $m_\phi$  there is a finite number of resonant points that can lead to the correct relic density. In the example of Fig. 3, in particular, there are five resonances that go below the correct value of the DM abundance and therefore  $2 \times 5 + 1 = 11$  viable options for  $\alpha_\chi$ .

Having explained above the prescription of counting resonances that result in the correct relic density, we proceed to analyze on-resonance models in a wide parameter range by solving Eqs. (29)–(31) numerically. We apply the counting prescription to every point on a discrete grid of the order of 0.1 megapixels in the  $(m_\chi, m_\phi)$ -plane, and the result is shown in Fig. 4. In our counting algorithm we further require that a resonance should have a sizable impact; i.e., we request the DM relic abundance to change by at least 1% in order to be counted. It can be recognized that multiple options of  $\alpha_\chi$  values exist in a huge parameter space region and in the “hot spot” around  $m_\chi \simeq 2$  TeV and  $m_\phi \simeq 100$  MeV we can have up to  $2 \times 8 + 1 = 17$  viable  $\alpha_\chi$  values.

The region of multiple  $\alpha_\chi$  values is bordered from below by the grey shaded region, where no resonances leading to the correct relic density exist because  $\epsilon_\phi^{\text{min}} \equiv \frac{m_\phi}{\alpha_\chi^{\text{max}} m_\chi} > \frac{6}{1^2 \pi^2}$ , where  $\alpha_\chi^{\text{max}}$  represents the maximum (nearly-on-resonant or off-resonant) coupling value leading to the correct relic abundance. From above and to the left, the region is bounded by the requirement of enabling at least a 1% change in the DM relic abundance due to reannihilation, as implemented in the counting algorithm. In Appendix C 2, we provide analytical estimates to explain this “max 1%” boundary where reannihilation can no longer significantly

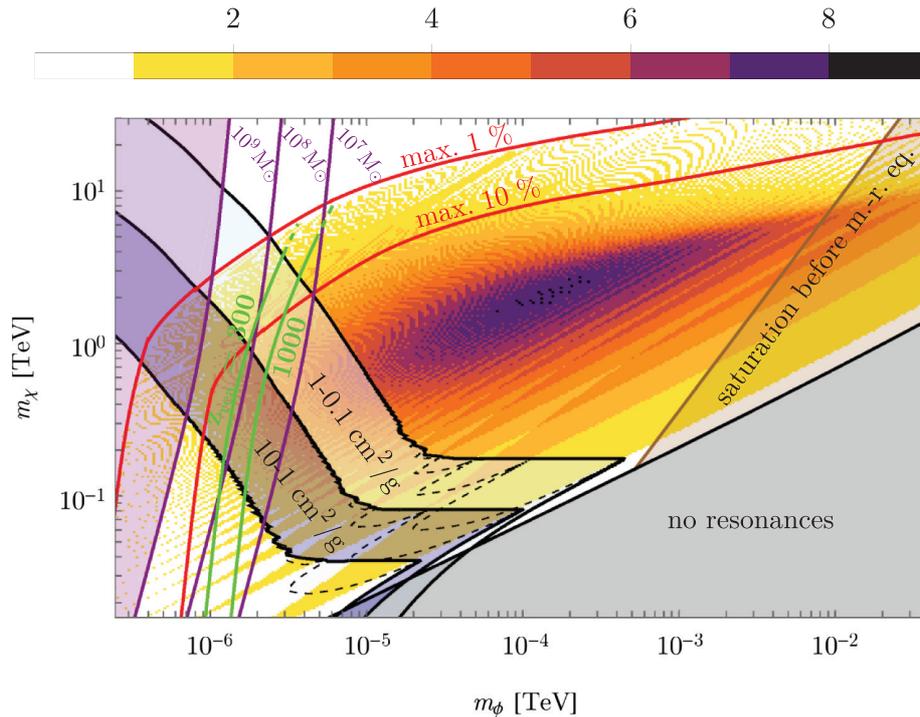


FIG. 4. Number of Sommerfeld resonances, color coded as given in the top panel, leading to the correct relic density today and changing the comoving DM number density by at least 1% during the epoch of reannihilation. The red solid lines show our analytic estimates (see Appendix C 2) of the border where reannihilation can change the relic abundance at most by 1% and 10%. In the shaded grey area in the bottom right part of the figure, no resonances are available leading to the correct relic density. The brown shaded area represents the estimated region where reannihilation cannot proceed after matter-radiation equality. Blue and light blue shaded areas cover the parameter space where DM has a sizable self-scattering cross section on dwarf galactic scales:  $(\sigma_T)_{30 \text{ km/s}}/m_\chi \in [0.1, 10] \text{ cm}^2 \text{ g}^{-1}$ . The “proper” SIDM region, in both the quantum-resonant and the classical self-scattering regimes, overlaps with the parameter space where sizable reannihilation can occur. In the quantum-resonant regime,  $\alpha_\chi$  is adjusted in the computation of  $\sigma_T$  such that for given  $m_\chi$  and  $m_\phi$  the resonance condition,  $\epsilon_\phi = 6/(n^2\pi^2)$ , is fulfilled for a given integer  $n$  (see last subsection of Sec. IV for a detailed explanation). For comparison, the black dashed self-scattering band is for  $\alpha_\chi$  satisfying the relic density constraint without taking reannihilation or resonances into account. Cutoff masses of the order of  $10^7$ ,  $10^8$ , and  $10^9 M_\odot$  in the halo-mass function are represented by the purple lines. In the stripe between the green lines, reannihilation induces the first decrease of the DM comoving number density by 1% between redshifts of  $z = 300$  and  $z = 1000$ —while the maximal change in the DM abundance can be read off from the red lines. In the parameter space where the blue region, the green lines, and the purple lines all overlap, SIDM could at the same time alleviate several small-scale structure formation problems and tensions between cosmological parameters derived from CMB and low-redshift astronomical observations (see Sec. V B and Fig. 7).

change the DM abundance. The analytical estimates are presented in Fig. 4 in terms of the red curves representing the boundaries where reannihilation can maximally change the abundance by 1% and 10%, respectively. For points close to the bottom left part of the red lines, reannihilation starts too late in order to change the relic abundance by more than 1% and 10% until today, respectively. For points close to the red lines where they start to bend for the first time in the left of the plot ( $m_\phi \simeq 1 \text{ MeV}$ ,  $m_\chi \simeq \text{TeV}$ ), saturation of the Sommerfeld enhancement before today prevents the abundance to change. In the region where the red curves bend a second time ( $m_\phi \simeq 10 \text{ MeV}$ ,  $m_\chi \simeq 10 \text{ TeV}$ ), saturation happens around matter-radiation equality, while in the upper right region saturation takes place in the radiation dominated epoch for counted

resonances. One can clearly see that our estimates match very well the numerical results (1% line). Let us remark that these analytical estimates can be applied also to other cases, e.g., where  $\alpha_\chi \neq \alpha_l$  or  $r_{\text{BBN}} \neq 0.5$  and one does not have to necessarily run a numerical differential equation solver to find these borders.

In the brown shaded region reannihilation stops before matter-radiation equality for every resonance. This statement is true even for  $\alpha_\chi$  values that do not give the correct relic density. The dashed brown line is the border at which the minimum saturation temperature equals the matter-radiation equality temperature, i.e.,  $T_\gamma^{\text{sat.min}} = 0.80 \text{ eV}$  [see Eq. (27)]. In Sec. V we will see that the observational consequences of reannihilation are expected to be negligible for models deep inside this brown region.

### A. SIDM region

It can be recognized in Fig. 4 that the SIDM region  $(\sigma_T)_{30 \text{ km/s}}/m_\chi \in [0.1, 10] \text{ cm}^2 \text{ g}^{-1}$  significantly overlaps in the classical and quantum-resonant regime with the region potentially having a sizable reannihilation process. We also show the reannihilation redshift  $z_{\text{rea}}$  in terms of the green lines. In the *most interesting SIDM region*, where also a sizable cutoff mass around  $10^8 M_\odot$  can be achieved, we conclude from the green lines that reannihilation typically happens in the *matter dominated epoch*.

The computation of the self-scattering cross section  $(\sigma_T)_{30 \text{ km/s}}$  needs some further consideration in the parameter region where reannihilation happens. The multiple  $\alpha_\chi$  values leading to the correct relic abundance would also lead to multiple values of  $(\sigma_T)_{30 \text{ km/s}}$  for fixed  $m_\phi$  and  $m_\chi$ . For the blue self-scattering band in Fig. 4 we take the nearly-on-resonant  $\alpha_\chi$  value that is closest to the off-resonant  $\alpha_\chi$  leading to the correct relic density. This is a conservative choice since resonances with lower  $n$  would give more sizable reannihilation and thus more often be constrained by, e.g., CMB observations. In the classical scattering region, this choice has, however, virtually no impact on the self-scattering band since the resonances are very close to each other and therefore  $(\sigma_T)_{30 \text{ km/s}}$  does not change significantly when choosing an off-resonant or the closest on-resonant value of  $\alpha_\chi$ . However, in the quantum-resonant regime it makes a significant difference from using an off-resonant value when computing  $(\sigma_T)_{30 \text{ km/s}}$  as in, e.g., Refs. [30,33,54,55], which results in the dashed black curves (where  $\alpha_\chi$  is uniquely set by the standard relic density constraint, taking no reannihilation into account). In the Born regime ( $\epsilon_\phi \gg 1$ ) we chose  $\alpha_\chi$  as in the traditional computation since no resonances are available and therefore the coupling is unique.

Let us comment on the choices of fixed  $z_{\text{rea}}$  contours given by the green lines in Fig. 4. Our calculations show that a reannihilation process at  $z_{\text{rea}} \simeq 300$  with 5% change in the DM abundance starts to saturate between a redshift of  $z \sim 30$ –50, when most of DM is already confined in virialized halos. Our homogeneous and isotropic treatment of the Boltzmann equation is expected to break down in this nonlinear regime due to the increase of DM particle velocities in gravitationally bounded structures. Therefore we regard  $z_{\text{rea}} \simeq 300$  as a lower critical value above which  $\sim 5\%$  change in the abundance can be achieved. The reannihilation process starting at  $z_{\text{rea}} \simeq 1000$  with  $\sim 10\%$  change in the abundance saturates much earlier than the time when most of the structures become nonlinear and should therefore be safe from this caveat. For redshifts just above  $z_{\text{rea}} \simeq 1000$  many CMB quantities might be affected strongly since reannihilation happens around recombination. A simple approximation of the green lines can be obtained by solving Eq. (25) for fixed  $z_{\text{rea}}$ . However, this equation is strictly speaking only valid in the regime where

$z_{\text{rea}} \ll z_{\text{eq}}$  and therefore a not good approximation in the case of  $z_{\text{rea}} \simeq 1000$ . The green lines in Fig. 4 are the solution of an improved equation discussed in detail in Appendix C 3.

### V. COSMOLOGICAL IMPACT

The change in the DM number density and the redshifting of injected dark radiation during reannihilation modifies the expansion rate of the Universe when compared to the  $\Lambda$ CDM cosmology. Since this process is time dependent, the naive constraints on extra relativistic degrees of freedom  $\Delta N_{\text{eff}}$  cannot be applied in general. Instead, we suggest that the following basic quantities derived from time integration of the modified Hubble expansion rate should not be strongly affected; otherwise reannihilation would hardly reproduce the measured CMB anisotropies or the baryon acoustic oscillation observed in galaxy clustering.

The angular size of the sound horizon  $\theta_*$  at  $z = z_*$ , where  $z_*$  is defined as the redshift where the optical depth  $\tau$  equals unity [90], is a geometrical quantity directly related to the peak positions in the CMB power spectrum and thus precisely measured. We will work with the value reported by the Planck 2015 (TT + lowP) analysis [1]:

$$100\theta_* = 1.04105 \pm 0.00046, \quad (47)$$

along with

$$z_* = 1090.09 \pm 0.42. \quad (48)$$

From Ref. [90] we have

$$100\theta_* = 100 \times r_s(z_*)/D_A(z_*). \quad (49)$$

The sound horizon  $r_s$  and angular diameter distance  $D_A$  are given by

$$r_s(z) = \int_0^{1/(1+z)} \frac{da}{a^2 H \sqrt{3(1+R)}}, \quad (50)$$

$$D_A(z) = \int_{1/(1+z)}^1 \frac{da}{a^2 H}, \quad (51)$$

where

$$R = \frac{3\rho_b}{4\rho_\gamma} = \frac{3a\Omega_b h^2}{4\Omega_\gamma h^2}, \quad (52)$$

and  $a$  is the cosmological scale factor.  $r_s(z_*)$  captures the information of the Hubble expansion rate before recombination while  $D_A(z_*)$  is sensitive to that between recombination and today. The definition and further explanation of the introduced quantities can be found in Ref. [90]. The standard Hubble expansion rate is given by

$$H^2 = \frac{8\pi G}{3} [\rho_\gamma + \rho_\nu + \rho_c + \rho_b + \rho_\Lambda]. \quad (53)$$

In Appendix D, we provide the details of the cosmological parameters we use to render the above quantities compatible with the Planck 2015 (TT + lowP) measurements [1]. This set of parameters defines our standard Hubble expansion rate of the  $\Lambda$ CDM cosmology. When including reannihilation we will replace the standard CDM energy density  $\rho_c$  with the quantity given in Eq. (36). Note that there might exist a compensation between the reannihilation effect and, e.g., the choice of the SM neutrino masses  $m_\nu$  entering the parametrization of energy density  $\rho_\nu$  in Eq. (53). However, we do not consider this possibility here and fix  $m_\nu$  as in the Planck 2015 (TT + lowP) analysis [1]. Next we show how the basic quantities given above are sensitive to reannihilation.

### A. Reannihilation before recombination

We here consider reannihilations starting in the radiation dominated epoch and explore the impact on  $100\theta_*$ . In particular, we investigate the case where the DM abundance is initially overabundant by a few percent and reannihilation leads to the correct observed value. The evolution of the DM number density and the modified Hubble expansion rate are shown in Fig. 5 for such a few scenarios. It can be seen that the modified Hubble expansion rate starts to increase relative to standard  $\Lambda$ CDM around the transition from radiation to matter dominated epoch, which is due to the initial overabundance of DM. It can be recognized that although reannihilation has already saturated around recombination  $z_*$ , the Hubble expansion rate is still modified afterwards. This can be explained by the gradual

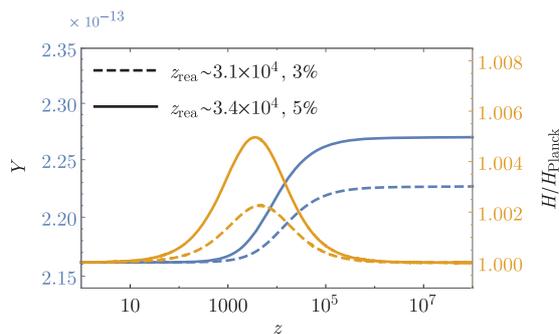


FIG. 5. Evolution of the DM number density  $Y = n_\chi/s$  (blue line) and the corresponding expansion rate  $H$  (yellow line) shown as a function of the redshift. The onset (1% change in  $Y$ ) of reannihilation for the dashed and solid curves is around  $z \simeq 3 \times 10^4$  and the DM abundance is initially enhanced by 3% and 5%, respectively. The final relic abundances coincide with  $(\Omega_c h^2)_{\text{Planck}} = 0.1197$ , and the ratio  $H/H_{\text{Planck}}$  therefore reaches 1 at low redshifts. Both scenarios would be in strong tension with the observed value of  $100\theta_*$ ; see Fig. 6.

redshifting of the injected dark radiation, which delays the return to the standard Hubble expansion rate.

We consider now the impact of the modified expansion rate on  $100\theta_*$  by investigating the integrations over  $H$  as they appear in Eq. (49). The naive number of standard deviations away from the reported  $100\theta_*$  value in Eq. (47) are calculated, and the results as a function of  $z_{\text{rea}}$  for a fixed amount of DM depletion are shown in Fig. 6. It can be seen that both scenarios presented in Fig. 5, where the DM abundance was initially enhanced by only a few percent, are in strong tension with the value of  $100\theta_*$  constrained by the Planck data. Furthermore, it can be recognized that the angular size of the sound horizon is sensitive even to percentage changes in the DM abundance in the radiation dominated epoch. However, the deeper in the radiation dominated epoch the reannihilation process takes place the less impact it has on the sound horizon and the more DM would be allowed to annihilate into dark radiation. This can be simply understood by the fact that changes in the DM abundance in the radiation dominated epoch have no significant impact on the expansion rate as long as the correct abundance is achieved sufficiently before matter-radiation equality. The process of reannihilation *necessarily* takes place in the radiation dominated epoch for parameters in the brown shaded region of Fig. 4. Note that points on the left side of the brown line can still have saturation either before or after matter-radiation equality.

To produce Figs. 5 and 6, we used  $m_\chi = 1$  TeV and varied  $m_\phi$  around  $\mathcal{O}(10)$  MeV and adjusted  $\alpha_\chi$  to have the desired DM relic abundance—but the same result would also be found with other DM model parameters that have the same  $z_{\text{rea}}$  and amount of DM depletion during the reannihilation process. From our background considerations, we therefore expect that a full Boltzmann code analysis of the CMB would lead to tight constraints on the change in the DM abundance in most of the parameter space in Fig. 4 and hence lower the viable number of  $\alpha_\chi$  values.

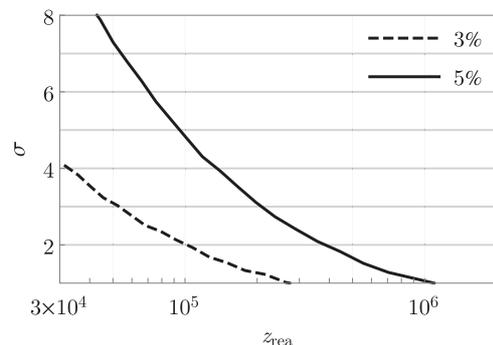


FIG. 6. Number of standard deviations from the  $(100\theta_*)_{\text{Planck}}$  measurement vs the redshift of reannihilation onset  $z_{\text{rea}}$  (i.e., when the comoving DM abundance first changes by more than 1%). The two curves refer to 3% and 5% total change in the relic abundance where the final value reaches  $(\Omega_c h^2)_{\text{Planck}} = 0.1197$ .

## B. Reannihilation after recombination

We now turn to explore the impact on cosmology from reannihilation at late times. The region of interest is now where reannihilation happens after recombination,  $z_{\text{rea}} \lesssim z_*$ , and especially the area between the green lines in Fig. 4. It is interesting to note that this area has overlap with both the relevant SIDM region of sizable self-scattering and where the DM halo abundance is suppressed below the mass around  $10^8 M_\odot$ .

The main difference compared to the previous section is that we will here impose compatibility with the basic CMB quantities [Eqs. (49)–(51)] constrained by Planck, while at the same time demonstrate that allowed modifications of the Hubble expansion can alleviate tensions between different cosmological measurements within the  $\Lambda$ CDM model. Several works have pointed out the so-called  $H_0$  tension; a discrepancy within the  $\Lambda$ CDM model between the measured value of the Hubble constant using CMB data [1],  $H_0 = 67.31 \pm 0.96 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (68% C.L.), and local measurements using only low redshift data,  $H_0 = 73.24 \pm 1.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (68% C.L.) [39]. Another tension concerns large-scale structure data and the value of the matter fluctuation amplitude on scales of  $8h^{-1} \text{ Mpc}$ ,  $\sigma_8$ . This issue is related to the  $H_0$  tension, as the Hubble parameter correlates with the matter density  $\Omega_m$  and  $\sigma_8$ . Constraints in the  $\sigma_8 - \Omega_m$  plane have been widely discussed in the literature [40–46], since current CMB data provide significantly different constraints than the thermal Sunyaev-Zel’dovich cluster counts [91] and galaxy weak lensing results [92,93], which both prefer lower values of  $\sigma_8$ .

Our approach will be to require the initial DM abundance to coincide, until recombination, with the reported central value of Planck. This leaves the sound horizon at recombination unaltered,  $r_s(z_*) = r_s^{\text{Planck}}(z_*)$ , since it is a distance derived from integrating  $H(a)$  from  $a = 0$  to the redshift of recombination. Depending on  $z_{\text{rea}}$ , reannihilation can then lower the DM abundance after recombination and injects energy in the form of dark radiation until the process saturates. The loss of DM particles and the redshifting of the dark radiation lowers the Hubble expansion rate  $H$  at later times when compared to the  $\Lambda$ CDM setup, which thus modifies  $D_A$ . On the one hand, we require that the tightly constrained quantity  $100\theta_*$  is not affected, but, on the other hand, we allow some amount of reannihilation to happen. This can be achieved by increasing the dark energy content  $\rho_\Lambda$  in Eq. (35), such that the period of lower  $H$  in the matter dominated epoch is compensated by a period of enhanced  $H$  in the dark energy dominated epoch. In practice, we iteratively change  $\rho_\Lambda$  to find the desirable  $H$  evolution such that  $100\theta_*$  does not change when reannihilation lowers the DM abundance.

The modified expansion rate and the angular diameter distance, computed as explained above, are shown in Fig. 7 together with low-redshift astronomical data: Hubble Space Telescope (HST) at  $z = 0$  [39], SDSS/BOSS at  $z = 0.35$

[94,95],  $z = 0.57$  [96,97], and  $z = 2.34$  [98]. We demonstrate the modification for 5% and 10% changes in the DM abundance for  $z_{\text{rea}} = 300$  and 700. To be in the special SIDM region we have used the parameters  $m_\chi = 700 \text{ GeV}$ ,  $m_\phi \in [1.6, 2.0] \text{ MeV}$ , and  $\alpha_\chi$  tuned to get the 5% and 10% changes in the DM abundance, but the same result would be found for every DM model that has the same  $z_{\text{rea}}$  and change in the DM abundance (see Fig. 4 for further possible options). In Fig. 7, one can see that low-redshift data prefer a 6%–11% larger value of  $H_0$  than that inferred by the  $\Lambda$ CDM interpretation of CMB data. Interestingly, it can be seen that this tension is mitigated by the reannihilation process when changing the DM abundance by 5%–10% after recombination. The  $H/H_{\text{Planck}}$  ratio increases at low redshifts below  $z \lesssim 1$  because  $\Omega_\Lambda$  needs to be larger to keep the highly constrained quantity  $100\theta_*$  unchanged. The reannihilation scenario is also in better agreement with several measurements of the angular diameter distance at low redshifts, while the point reported by Ref. [97] still favors a pure  $\Lambda$ CDM cosmology.

The reduction of  $\Omega_m$  at low redshifts due to reannihilation leads to a suppressed growth of the matter density perturbations, which might solve the discrepancy in the  $\sigma_8 - \Omega_m$  plane in  $\Lambda$ CDM [1,99]. The conversion of DM mass density into radiation energy lowers the growth factor since radiation can escape from the gravitational potential and does not contribute to the gravitational growth. As a consequence, the resultant matter power spectra would be suppressed compared to the  $\Lambda$ CDM cosmology, and thus reannihilation can potentially solve the  $\sigma_8 - \Omega_m$  tension.

The solution of the  $\sigma_8 - \Omega_m$  tension was discussed for a similar scenario where a part of DM decays into dark radiation after recombination [43]. In Fig. 7 we also show our results from a decaying DM scenario, while it was similarly investigated in Ref. [42]. In this setup, the dark sector consists of a DM component of stable  $\chi$  particles and mother particles ( $M$ ) that can decay into effectively massless daughter particles ( $D$ ). The energy density evolution of the latter two components can be obtained by solving

$$\dot{\rho}_M + 3H\rho_M = -\Gamma\rho_M, \quad (54)$$

$$\dot{\rho}_D + 4H\rho_D = \Gamma\rho_M, \quad (55)$$

numerically, with initial condition  $\rho_M(t_i) = f_{\text{dec}}\rho_\chi$  and  $\rho_D(t_i) = 0$ . The total dark matter sector’s energy density then evolves as

$$\rho_{\text{dark}}^{\text{decay}} = \rho_M + \rho_D + (1 - f_{\text{dec}})\rho_\chi. \quad (56)$$

For comparison, we fix the fraction  $f_{\text{dec}}$  of decaying DM (mother particles) with respect to the stable component  $\chi$  to 5% and 10%, i.e.,  $f_{\text{dec}} = 0.05$  and  $f_{\text{dec}} = 0.1$ . We then match the decay rate  $\Gamma$  such that  $H_0$  coincides with the reannihilation result, while again adjusting the dark energy

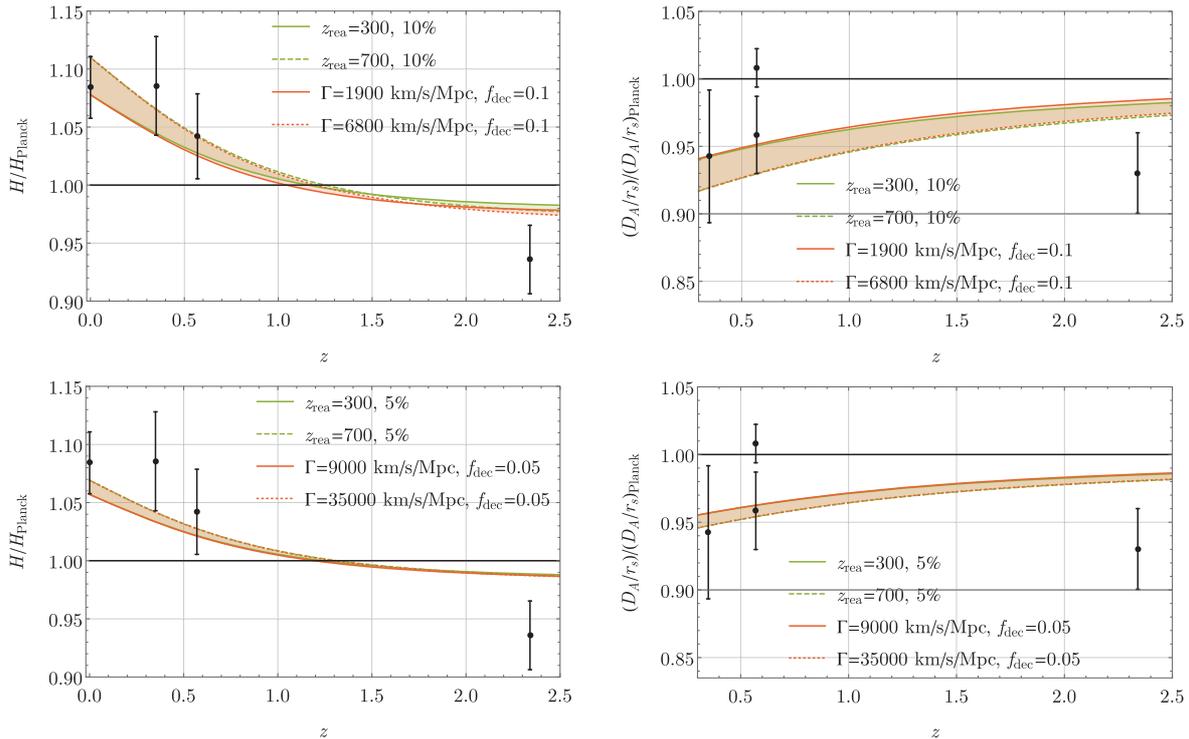


FIG. 7. Hubble expansion rate (left) and angular diameter distance (right) ratio vs redshift for reannihilating (green lines) and decaying DM (red lines). In both DM scenarios the abundance is changed by 10% (top) and 5% (bottom) after recombination, and the results are almost identical in this redshift interval. Data points are from HST, SDSS, and BOSS data [39,94–98] and are shown relative to the CMB derived value from Planck data [1] in the  $\Lambda$ CDM setup.

density to leave the CMB observable  $100\theta_*$  unchanged. In Fig. 7 you clearly see that at low redshifts,  $z \lesssim 2.5$ , the reannihilation and decaying DM models can mimic each other. They are not distinguishable from these existing astronomical data.

However, let us in the following explain why we believe that these two scenarios impact differently on the evolution of linear perturbations and thus are potentially distinguishable in a CMB power spectrum analysis. In particular, it was shown in a detailed analysis of Ref. [46] that the CMB observation is still sensitive to decaying DM even long after recombination through the late integrated Sachs-Wolfe effect as it happens in the parameter range as shown in Fig. 7. The conclusion was that the impact on the CMB power spectrum for the decaying DM scenarios shown in our figures are too strong and essentially ruled out. However, reannihilation has several different features and the results presented in Ref. [46] for decaying DM cannot be trivially mapped one-to-one to annihilating DM. First of all, it is clear that the energy density of the unstable mother particles  $\rho_M$  decays exponentially fast in time. In the reannihilation case the DM density changes slower and its duration is longer compared to decaying DM. Therefore, the evolution of the injected dark radiation (or daughter particles) and the modification of the Hubble expansion rate are different. Second, the right-hand side of the

Boltzmann equation is proportional to DM density squared for reannihilation, while for decaying DM it is instead linear in the density. This might lead to further differences in the evolution of cosmological perturbations. Third, the annihilation process is velocity dependent, and one does not expect reannihilation to happen at wavelength modes that have already formed sizable structure. Finally, for large cutoff masses of the order  $10^8 M_\odot$  it has been found that the reionization history is different compared to  $\Lambda$ CDM predictions [100,101]. We leave a development of a Boltzmann code and a more detailed investigation of these issues to future work.

## VI. SUMMARY AND CONCLUSION

In this work we have explored the observational imprints of a second period of DM annihilation into dark radiation. We have shown that such an epoch of reannihilation can arise in DM models where the annihilation cross section is  $s$ -wave dominated and resonantly Sommerfeld enhanced. As a concrete realization we have considered a simple model where sizable self-interactions are induced by a light vector mediator, interacting with a dark matter particle and a massless background particle in a fully closed dark sector. We have extensively analyzed the reannihilation phenomenology of this model and found that this process can

change the initial DM number density set by the standard thermal freeze-out, by up to a factor of several in a wide range of the model parameter space. Furthermore, the onset of reannihilation can range from being deep in the radiation dominated epoch to the beginning of halo formation.

In the most interesting parameter region of our considered particle physics model—where several small-scale structure formation issues can be addressed—we have shown that the reannihilation process starts during the matter dominated epoch. Existing CMB data, which are sensitive to even only a few percentage changes in the DM abundance during this epoch, might confirm the existence of such scenarios. We have interestingly found that in the same parameter region the reannihilation process might reduce the tension between CMB and low-redshift astronomical observations of  $H_0$  and  $\sigma_8$ —although our discussion is limited at the background level. We have also demonstrated that reannihilation can be used as a clear signature to break the otherwise close degeneracy between scalar and vector mediator realizations of self-interacting dark matter models.

The effects on cosmological perturbations, especially on the CMB power spectrum, might be nontrivial even in the cases where reannihilation happens much later than recombination or occurs deep in the radiation dominated epoch. A dedicated Boltzmann code deserves to be developed to identify the detailed signatures of reannihilation and to clarify how well tensions between CMB and low-redshift astronomical observations can be alleviated.

### ACKNOWLEDGMENTS

M. G. and T. B. thank Torsten Bringmann, Laura Covi, Andrzej Hryczuk, Sebastian Wild, and Hai-bo Yu for reading and commenting on our draft, as well as the participants at SIDM workshop in Copenhagen for stimulating discussions. A. K. thanks Ryusuke Jinno and Toyokazu Sekiguchi for useful discussions. M. G. and T. B. have received funding from the European Union’s Horizon 2020 research and innovation program under Grant Agreements No. 690575 and No. 674896. T. B. gratefully acknowledges financial support from the German

Science Foundation (DFG RTG 1493). The work of A. K. was supported by IBS under the Project Code IBS-R018-D1. S. R. S. and M. W. gratefully thank the ITP Göttingen for the nice hospitality during the early stage of this project, which is partially based on our Bachelor’s theses [102,103]. We all thank Marcel Langenberg for his support with the GWDG computer cluster.

### APPENDIX A: THEORETICAL UNCERTAINTIES IN THE COMPUTATION OF SELF-SCATTERING CROSS SECTIONS

In Fig. 8 we show a comparison between  $\sigma_T/m_\chi$  in a classical approximation and in the quantum treatment discussed in detail in the appendix of Ref. [33]. From this figure we conclude that there are quantum corrections in both the vector and the scalar mediator setups, but for our work they are small enough to be neglected. We note that we see a tendency of an increase of the corrections for higher velocities. This might imply larger corrections on Galactic cluster scales, with  $v_0 \sim 1000$  km/s. To perform a precise calculation on such velocity scales it would require the summation of many more scattering phases  $\delta_\ell$  which is beyond the scope of this paper.

### APPENDIX B: SELF-CONSISTENT DESCRIPTION OF THE SOMMERFELD ENHANCEMENT

It was pointed out in Ref. [81] that close to a resonance it is required to calculate the Sommerfeld enhancement self-consistently in order not to violate the partial wave unitarity limit. This means that in the derivation of the DM non-relativistic effective theory both the long *and* short range contributions have to be taken into account in the effective potential of the two-body Schrödinger equation. In our scenario the long range part is the Yukawa potential, while the short range contribution consists of the hard annihilation and scattering processes. The *regulated* formula for the total *s*-wave Sommerfeld enhanced annihilation cross section derived from a self-consistent solution of the Schrödinger equation is given for attractive forces by [81]

$$(\sigma v_{\text{rel}})_{\text{ann}} \simeq (\sigma v_{\text{rel}})_{\text{ann},0} \times \frac{S(v_{\text{rel}})}{\left| 1 + v_{\text{rel}} \left( -\sqrt{\frac{\mu^2 \sigma_{\text{sc},0}}{4\pi}} - \left( \frac{\mu^2 (\sigma v_{\text{rel}})_{\text{ann},0}}{4\pi} \right)^2 - i \frac{\mu^2 (\sigma v_{\text{rel}})_{\text{ann},0}}{4\pi} \right) (T(v_{\text{rel}}) + iS(v_{\text{rel}})) \right|^2}. \quad (\text{B1})$$

In our work we approximate the Yukawa potential as the Hulthén potential for which  $S(v_{\text{rel}})$  is given in Eq. (14) and  $T$  takes the form [81]

$$T(v_{\text{rel}}) \simeq -\frac{1}{2\epsilon_v} (H(\alpha_+) + H(\alpha_-) + H(-\alpha_+) + H(-\alpha_-) - \{p \rightarrow p_0\}), \quad (\text{B2})$$

$$\alpha_\pm = i \frac{\epsilon_v}{\epsilon_\phi \pi^2 / 6} \pm \sqrt{\frac{1}{\epsilon_\phi \pi^2 / 6} - \left( \frac{\epsilon_v}{\epsilon_\phi \pi^2 / 6} \right)^2}. \quad (\text{B3})$$

Here,  $H(z)$  is the analytic continuation of the  $z$ th harmonic number. For the tree-level annihilation cross section  $(\sigma v_{\text{rel}})_{\text{ann},0}$  in Eq. (B1) we take the sum over all tree-level channels,

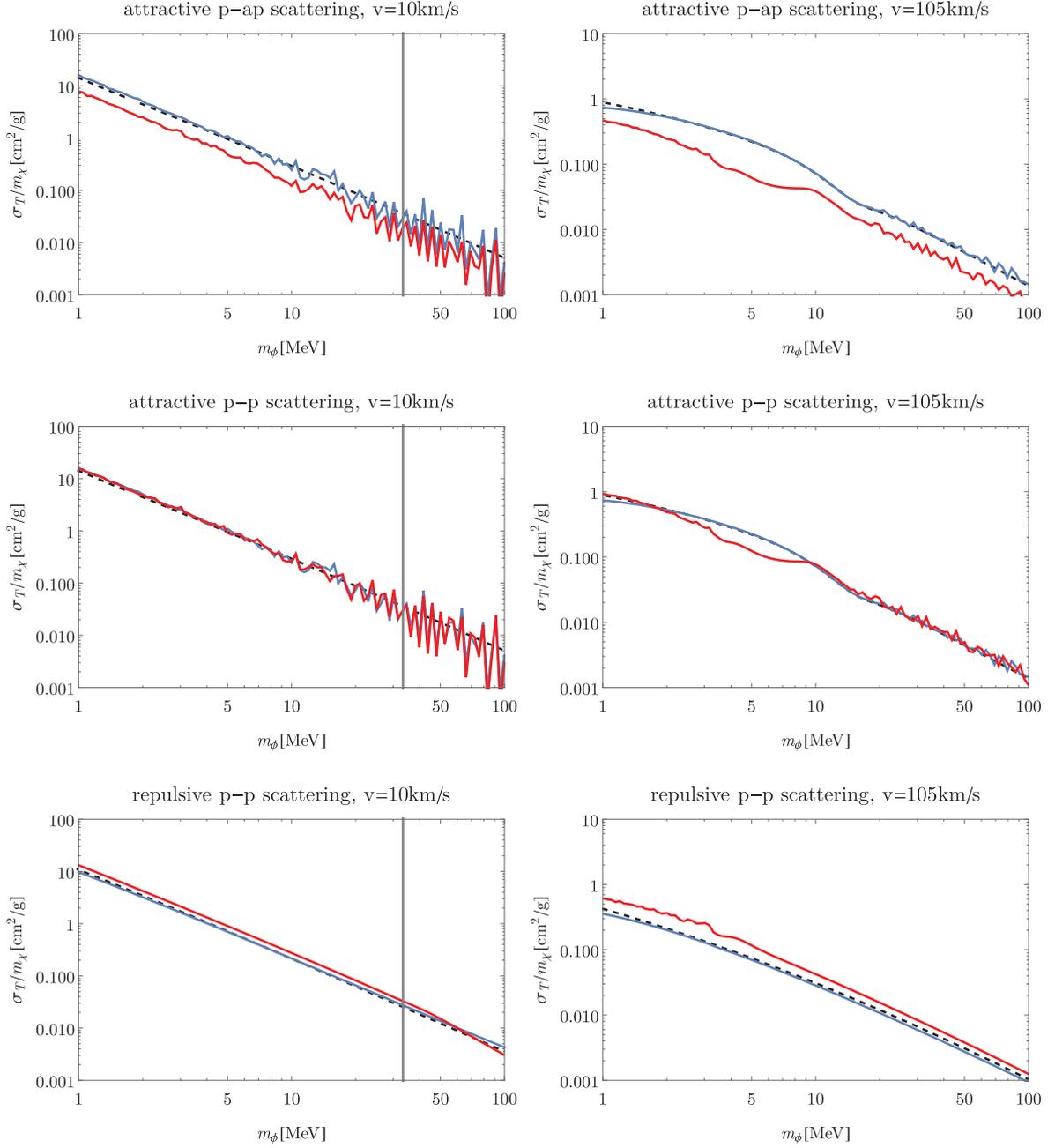


FIG. 8. Self-scattering transfer cross sections vs mediator mass  $m_\phi$  from numerical solutions of the Schrödinger equation [33,54,55] (red and blue lines) compared to the ETHOS [24] fitting functions (black dashed lines). The particle model is fixed to  $m_\chi = 1$  TeV and  $\alpha_\chi = 0.033$ . *Left*: A relative DM velocity  $v_0 = 10$  km/s. *Right*:  $v_0 = 105$  km/s. *Top*: Scattering with an attractive Yukawa potential between particle-antiparticle (p-ap) (applies to vector and scalar mediators). *Middle*: Attractive particle-particle scattering (scalar mediators). *Bottom*: Repulsive particle-particle scattering (vector mediators). The numerical solutions include the computation and summation of phase shifts  $\delta_\ell$  up to  $\ell = 150$  (left) and  $\ell = 225$  (right). The red curve corresponds to the computation of  $\sigma_T$  when including quantum statistics and averaging  $d\sigma/d\Omega$  over  $1 - |\cos\theta|$ , as suggested in Ref. [33]. Points to the left of the gray line are in the regime  $m_\chi v_0 \gtrsim m_\phi$  where the blue and dashed black curves are expected to coincide. Deviation from the red line indicates the theoretical bias of taking a classical approximation and averaging the scattering amplitude over  $1 - \cos\theta$  [see Eq. (6)] instead of  $1 - |\cos\theta|$ .

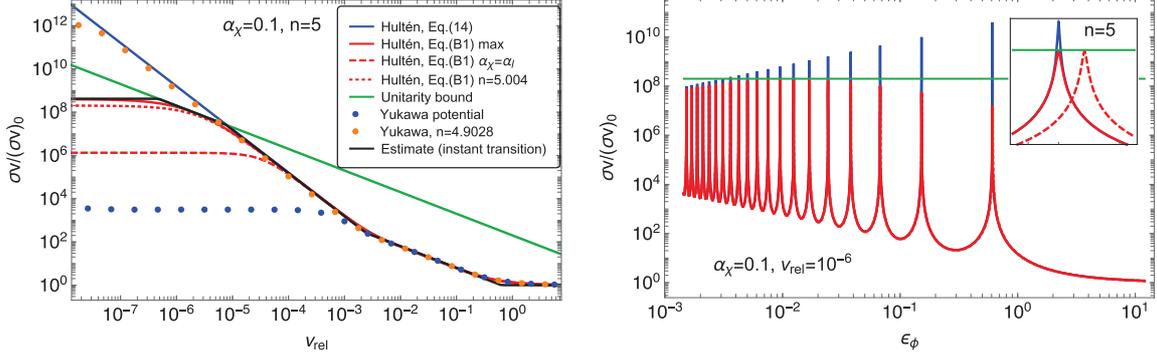


FIG. 9. *Left*: Comparison between various approximations of the  $s$ -wave Sommerfeld factor  $(\sigma v_{\text{rel}})_{\text{ann}}/(\sigma v_{\text{rel}})_{\text{ann},0}$  for a model with  $\alpha_\chi = \alpha_l = 0.1$ . The regulated Hulthén potential solution (red line) with  $\sigma_{\text{sc},0} = (\mu/4\pi)(\sigma v_{\text{rel}})_{\text{ann},0}^2$  is compared to the unregulated solution (blue line) on the parametric resonance point  $n = 5$ . The black curve shows our instantaneous transitions estimate to the red solid line, which respects the  $s$ -wave unitarity bound shown by the green line. For the remaining cases the exact resonance point is slightly shifted from  $\epsilon_\phi = 6/(n\pi)^2$  and  $n$  needs to be tuned to find maximal Sommerfeld enhancement: the regulated Hulthén potential solution with  $\sigma_{\text{sc},0} = 3\alpha_\chi^2\pi/(4m_\chi^2)$  for  $n = 5$  (dashed red line) and  $n = 5.004$  (dotted red line); the numerical solution of the Schrödinger equation with a Yukawa potential [104] for  $n = 5$  (blue dots) and  $n = 4.9028$  (orange dots). *Right*: Sommerfeld enhancement as a function of  $\epsilon_\phi$ . Same color coding as in the left panel, but only the first three entries and the unitarity bound are plotted. The inset panel is a zoom-in around the fifth resonance, covering a 1% range in  $\epsilon_\phi$  and 4 orders of magnitude range in  $(\sigma v_{\text{rel}})_{\text{ann}}$ , showing the choice of  $\sigma_{\text{sc},0}$  only shifts the resonance slightly while leaving the maximal enhancement unaffected.

$$(\sigma v_{\text{rel}})_{\text{ann},0} = \sum_i (\sigma v_{\text{rel}})_{0,i}, \quad (\text{B4})$$

as given in Eqs. (12) and (13). For the hard self-scattering cross section  $\sigma_{\text{sc},0}$  in Eq. (B1) we take

$$\sigma_{\text{sc},0} = \frac{3\alpha_\chi^2\pi}{4m_\chi^2}, \quad (\text{B5})$$

which can be obtained from the  $s$ -channel diagram of nonrelativistic particle-antiparticle scattering. In our computation of  $T(v_{\text{rel}})$  we drop the matching term  $\{p \rightarrow p_0\}$ , since it is only relevant close to the high energy scale  $p_0$  [81].

The regulated Hulthén potential solution as described above comes with a subtlety discussed in the following. The short range quantities  $(\sigma v_{\text{rel}})_{\text{ann},0}$  and  $\sigma_{\text{sc},0}$  affect the parametric resonance condition slightly when compared to the unregulated solution  $S(v)$ . To avoid having to repeatedly find the precise numerical resonance condition of a regulated solution when studying each single resonance in, e.g., the parameter scan of Fig. 4, we decided throughout this work to approximate  $\sigma_{\text{sc},0} = \frac{\mu}{4\pi}(\sigma v_{\text{rel}})_{\text{ann},0}^2$  such that the square root in the denominator of Eq. (B1) vanishes. In the right panel of Fig. 9 it is demonstrated that this choice of  $\sigma_{\text{sc},0}$  only shifts the parametric resonance condition back to the known expression  $\epsilon_\phi = 6/(n^2\pi^2)$ ; however, the height of the enhancement peak is practically unaffected. By numerical evidence, we have further checked that  $(\sigma v_{\text{rel}})_{\text{ann}}$  is modified by at most about 10% for all velocities in all the parameter regions we study. Furthermore, we looked at the numerical solution with

the Yukawa potential and demonstrate in Fig. 9 that also in this case only the resonance condition slightly deviates from  $\epsilon_\phi = 6/(n^2\pi^2)$ . It can be seen in all cases of the regulated Hulthén potential solution that the maximal enhancement respects the unitarity bound of  $s$ -wave annihilation cross sections, given by (see, e.g., Ref. [81])

$$\sigma_{\text{max}} = \frac{\pi}{\mu^2 v_{\text{rel}}^2}, \quad (\text{B6})$$

where the reduced mass is here given by  $\mu = m_\chi/2$ . Naively, the value of the Sommerfeld factor where it saturates can now be obtained from  $\sigma_{\text{max}} = (\sigma v_{\text{rel}})_{\text{ann},0} S(v_{\text{rel}})/v_{\text{rel}}$ , namely,

$$S^{\text{sat}}(v_{\text{rel}}) = \frac{\pi}{\mu^2 v_{\text{rel}} (\sigma v_{\text{rel}})_{\text{ann},0}}. \quad (\text{B7})$$

This expression will be used in Appendix C to estimate the saturation velocity of the Sommerfeld factor.

## APPENDIX C: DERIVATION OF ANALYTIC ESTIMATES

In Sec. C1 we find approximate expressions of the Sommerfeld enhancement that we then use to estimate  $\langle \sigma v_{\text{rel}} \rangle_{\chi'}$ . The latter quantity is used in Sec. C2 to find analytical expressions for  $\Gamma$ , which finally allows us to estimate the change in DM abundance due to reannihilation. Based on these results we estimate the onset time of reannihilation in Sec. C3.

### 1. Sommerfeld enhanced annihilation on a resonance

In the following, we first estimate  $S(v)$  and then  $\langle\sigma v_{\text{rel}}\rangle_{x'}$  in the case where the parameter  $\epsilon_\phi$  fulfills the resonance condition. It can be recognized from the left panel of Fig. 9 that the Sommerfeld factor has regions where  $S(v) \propto 1$ ,  $1/v$ , and  $1/v^2$  and a region where it starts to saturate and finally reaches a maximal value at low velocities. We approximate the transitions between these different regions as instantaneous transitions at the following transition velocities:

$$v_{1/v} = 2\pi\alpha_\chi, \quad (\text{C1})$$

$$v_{1/v^2} = \frac{m_\phi}{m_\chi}, \quad (\text{C2})$$

$$v_{\text{sat}} = \frac{\pi m_\phi}{2 m_\chi} \alpha_\chi^2 (\alpha_\chi + \alpha_l), \quad (\text{C3})$$

$$v_{\text{max}} = \frac{\pi m_\phi}{32 m_\chi} \alpha_\chi^2 (\alpha_\chi + \alpha_l). \quad (\text{C4})$$

We summarize this instantaneous description of  $S$  in Table II and demonstrate in Fig. 9 that this approximation (black line) matches well the numerical (red solid line) solution within each definite scaling regime. We have tested several on-resonant values of  $\epsilon_\phi$  and found in all the cases a similarly good result. The value of  $v_{\text{sat}}$  can be obtained by equating the values of  $S$  from the regime of  $1/v^2$  scaling in Table II with Eq. (B7). For lower velocities than  $v_{\text{sat}}$  we consider that  $S$  follows the scaling of the partial-wave unitarity bound, i.e.,  $1/v$  scaling.  $v_{\text{max}}$  is the velocity where  $S$  reaches the maximal value:

$$S^{\text{max}} = \frac{m_\chi}{2\pi\alpha_\chi m_\phi} \left( \frac{4\pi}{\mu^2 (\sigma v_{\text{rel}})_0} \right)^2, \quad (\text{C5})$$

which can directly be obtained from Eq. (B1) with  $\sigma_{\text{sc},0} = \frac{\mu}{4\pi} (\sigma v_{\text{rel}})_{\text{ann},0}^2$ , ignoring contributions from  $T$  and taking the limit of  $v \rightarrow 0$ .

A similar instantaneous transitions description will next be used to estimate the temperature evolution of the

TABLE II. Instantaneous approximation of the Sommerfeld factor and  $\langle\sigma v_{\text{rel}}\rangle$  for an on-resonance  $s$ -wave annihilation.

$v \in$	$S(v)$	$x' \in$	$\langle\sigma v_{\text{rel}}\rangle/(\sigma v_{\text{rel}})_0$
$[\infty, v_{1/v}]$	1	$[0, x'_{1/v}]$	1
$[v_{1/v}, v_{1/v^2}]$	$(\frac{v_{1/v}}{v})$	$[x'_{1/v}, x'_{1/v^2}]$	$(\frac{x'}{x'_{1/v}})^{1/2}$
$[v_{1/v^2}, v_{\text{sat}}]$	$(\frac{v_{1/v}}{v_{1/v^2}})(\frac{v_{1/v^2}}{v})^2$	$[x'_{1/v^2}, x'_{\text{sat}}]$	$(\frac{x'}{x'_{1/v^2}})^{1/2} (\frac{x'}{x'_{1/v^2}})$
$[v_{\text{sat}}, v_{\text{max}}]$	$(\frac{v_{1/v}}{v_{1/v^2}})(\frac{v_{1/v^2}}{v_{\text{sat}}})^2 (\frac{v_{\text{sat}}}{v})$	$[x'_{\text{sat}}, x'_{\text{max}}]$	$(\frac{x'}{x'_{1/v^2}})^{1/2} (\frac{x'_{\text{sat}}}{x'_{1/v^2}}) (\frac{x'}{x'_{\text{sat}}})^{1/2}$
$[v_{\text{max}}, 0]$	$S^{\text{max}}$	$[x'_{\text{max}}, \infty]$	$S^{\text{max}}$

thermally averaged cross section  $\langle\sigma v_{\text{rel}}\rangle_{x'}$  defined in Eq. (22). In the following we will drop the index  $x'$  to shorten the notation.  $\langle\sigma v_{\text{rel}}\rangle$  has a definite power-law dependence on  $x' = m_\chi/T_\chi$  at temperatures where one particular scaling of  $S(v)$  dominates. In analogy to the transition velocities,  $v_i$ , in Table II, we define instantaneous transition temperatures as

$$x'_{1/v} = \frac{c_{1/v}}{v_{1/v}^2}, \quad (\text{C6})$$

$$x'_{1/v^2} = \frac{c_{1/v^2}}{v_{1/v^2}^2}, \quad (\text{C7})$$

$$x'_{\text{sat}} = \frac{c_{\text{sat}}}{v_{\text{sat}}^2}, \quad (\text{C8})$$

$$x'_{\text{max}} = \frac{c_{\text{max}}}{v_{\text{max}}^2}, \quad (\text{C9})$$

and adjust the coefficients  $c_i$  such that the approximation coincides with the numerically obtained values of  $\langle\sigma v_{\text{rel}}\rangle$  within each definite scaling regime. For  $x'$  larger than  $x'_{\text{max}}$  we require that  $\langle\sigma v_{\text{rel}}\rangle/(\sigma v_{\text{rel}})_0 = S^{\text{max}}$ , which automatically determines the last matching coefficient:

$$c_{\text{max}} = \frac{c_{1/v} c_{1/v^2}}{c_{\text{sat}}}. \quad (\text{C10})$$

In Table II we summarize the instantaneous approximation of  $\langle\sigma v_{\text{rel}}\rangle$ . In particular, we find that the instantaneous approximation with  $c_{1/v} = 3$ ,  $c_{1/v^2} = 3/2$ , and  $c_{\text{sat}} = 1$  matches well the numerical result of  $\langle\sigma v_{\text{rel}}\rangle$  within each definite scaling regime. Next, we use this result to estimate the size of  $\Gamma$ .

### 2. Estimating the maximal change in the number density

In this appendix, we estimate the change in the relic abundance due to reannihilation. The ratio between the comoving DM abundances at kinetic decoupling ( $x_{\text{kd}}$ ) and today ( $x_0$ ) can be obtained from the standard solution of the Boltzmann equation [Eq. (29)], given by

$$\frac{Y(x_{\text{kd}})}{Y(x_0)} = 1 + \int_{x_{\text{kd}}}^{x_0} dx \frac{\Gamma}{x}. \quad (\text{C11})$$

Here,  $\Gamma$  is defined as in Eq. (19), but with the replacement  $Y(x) \rightarrow Y(x_{\text{kd}})$ . Note that the right-hand side is thus independent of the evolution of  $Y(x)$ . The aim is now to further simplify this formal solution by approximating the time integral. The dominant contribution is from the  $x$  range where  $\Gamma$  is maximal. In the following, we first derive simple power-law expressions of the maximal value of  $\Gamma$  and second show how to approximate the time integration in various cases.

The usual order of chemical before kinetic decoupling ( $x_{\text{cd}} \lesssim x_{\text{kd}}$ ) and kinetic decoupling before matter radiation equality ( $x_{\text{kd}} \lesssim x_0$ ) to have an adequate structure formation history implies the following time order:  $x_{\text{cd}} \lesssim x_{\text{kd}} \lesssim x_{\text{eq}} \lesssim x_0$ , where the subscripts label the SM photon temperature at chemical decoupling, kinetic decoupling, matter-radiation equality, and today, respectively. There are now several options to align the times  $x_{1/v}$ ,  $x_{1/v^2}$ , and  $x_{\text{sat}}$  [set by Eqs. (C6), (C7), and (C8) after converting  $x'$  into  $x$  via Eq. (24)] in between the fixed time order  $x_{\text{cd}} \lesssim x_{\text{kd}} \lesssim x_{\text{eq}} \lesssim x_0$ . It turns out that only five different cases (time alignments) are relevant for us, and those are summarized as follows:

Case	$\lesssim x_{\text{cd}} \lesssim$	$\lesssim x_{\text{kd}} \lesssim$	$\lesssim x_{\text{eq}} \lesssim$	$\lesssim x_0 \lesssim$
1.		$x_{1/v}$	$x_{1/v^2}$ $x_{1/v} \lesssim x_{1/v^2}$	$x_{\text{sat}}$ $x_{\text{sat}}$
2.		$x_{1/v}$	$x_{1/v^2}$ $x_{1/v} \lesssim x_{1/v^2}$	$x_{\text{sat}}$ $x_{\text{sat}}$
3.		$x_{1/v}$	$x_{1/v^2} \lesssim x_{\text{sat}}$ $x_{1/v} \lesssim x_{1/v^2} \lesssim x_{\text{sat}}$	
4.	$x_{1/v}$ $x_{1/v}$	$x_{1/v^2}$	$x_{1/v^2}$	$x_{\text{sat}}$ $x_{\text{sat}}$
5.	$x_{1/v}$ $x_{1/v}$	$x_{1/v^2}$	$x_{1/v^2} \lesssim x_{\text{sat}}$ $x_{\text{sat}}$	

The two options given in each case lead to the same result in the final form of  $\Gamma$  as can be shown explicitly

(without proof here). In the first case, the Sommerfeld enhancement saturates at later times than the age of the Universe:  $x_0 \lesssim x_{\text{sat}}$ . This implies that  $\Gamma$  reaches its maximal value today. In all other cases shown in the table, the maximal value is given at the time of saturation of the Sommerfeld enhancement. In the second case, saturation happens between matter-radiation equality and today while in the third case saturation is before matter-radiation equality. In the fourth and fifth cases, the Sommerfeld enhancement becomes sizable at the first freeze-out as we have  $x_{1/v} \lesssim x_{\text{cd}}$ .

From here on we are always assuming that we are exactly on a Sommerfeld resonance point. The maximal value of  $\Gamma$  as a function of the free parameters in these five different cases can be obtained as follows. We define  $x_{\text{cd}}$  as the time when  $\Gamma = 1$ . Requiring  $Y(x_{\text{kd}})$  to coincide with the value of  $Y$  corresponding to the correct relic density, we can determine  $x_{\text{cd}}$  as a function of  $m_\chi$  only. For the  $m_\chi$  range between 10 GeV and 40 TeV we find that  $x_{\text{cd}}$  varies approximately between 7 and 22. This variation is a consequence of the fixed temperature ratio  $r$  at BBN and the impact of the Sommerfeld effect on the first freeze-out temperature for DM masses above the TeV scale. At times later than  $x_{\text{cd}}$ , the evolution of  $\Gamma$  in all five cases directly follows from the entries of Table I and the results of the previous section. For example, the estimate of  $\Gamma_1$  is found to be

$$\Gamma_1 = \frac{(g_s/\sqrt{g_{\text{eff}}})_0}{(g_s/\sqrt{g_{\text{eff}}})_{\text{cd}}} \frac{x_{\text{cd}}}{x_{1/v}} \underbrace{\left(\frac{r_{1/v} x_{1/v}}{r_{\text{kd}} x_{\text{kd}}}\right)^{1/2}}_{\substack{x_{\text{cd}} \lesssim x_{1/v}, \\ S(v)=1}} \underbrace{\left(\frac{r_{1/v^2}^{-1}}{r_{\text{kd}}^{-1}}\right)}_{\substack{x_{\text{kd}} \lesssim x_{1/v^2}, \\ S(v) \propto 1/v}} \underbrace{\left(\frac{r_{\text{eq}}^{-2} x_{\text{eq}}}{r_{1/v^2}^{-2} x_{1/v^2}}\right)}_{\substack{x_{1/v^2} \lesssim x_{\text{eq}}, \\ S(v) \propto 1/v^2}} \underbrace{\left(\frac{r_0^{-2} x_0^{1/2}}{r_{\text{eq}}^{-2} x_{\text{eq}}^{1/2}}\right)}_{\substack{x_{\text{eq}} \lesssim x_0, \\ S(v) \propto 1/v^2}}. \quad (\text{C12})$$

By inserting  $x_{1/v}$  and  $x_{1/v^2}$  into this expression and applying the same procedure to the second and third cases, we find the maximum value of  $\Gamma$  is given by

$$\Gamma_{1,2,3} = \frac{(g_s/\sqrt{g_{\text{eff}}})_{\text{sat}}}{(g_s/\sqrt{g_{\text{eff}}})_{\text{cd}}} \frac{x_{\text{cd}}}{r_{\text{sat}}^2} \frac{2\pi}{\sqrt{c_{1/v} c_{1/v^2}}} \frac{\alpha_\chi m_\phi}{m_\chi} \begin{cases} \left(\frac{T^{\text{kd2}}}{T_{\text{eq}}^{\text{eq}} T_\gamma^0}\right)^{1/2} & \text{for } T_\gamma^0 \gtrsim T_\gamma^{\text{sat}}, \\ \left(\frac{T^{\text{kd2}}}{T_{\text{eq}}^{\text{eq}} T_\gamma^{\text{sat}}}\right)^{1/2} & \text{for } T_\gamma^{\text{eq}} \gtrsim T_\gamma^{\text{sat}} \gtrsim T_\gamma^0, \\ \left(\frac{T^{\text{kd}}}{T_\gamma^{\text{sat}}}\right) & \text{for } T_\gamma^{\text{sat}} \gtrsim T_\gamma^{\text{eq}}, \end{cases} \quad (\text{C13})$$

and in the last two cases where  $x_{\text{cd}} \gtrsim x_{1/v}$  we find

$$\Gamma_{4,5} = \frac{(g_s/\sqrt{g_{\text{eff}}})_{\text{sat}}}{(g_s/\sqrt{g_{\text{eff}}})_{\text{cd}}} \frac{(x_{\text{cd}} r_{\text{cd}})^{1/2}}{r_{\text{sat}}^2} \frac{1}{\sqrt{c_{1/v^2}}} \frac{m_\phi}{m_\chi} \begin{cases} \left(\frac{T^{\text{kd2}}}{T_{\text{eq}}^{\text{eq}} T_\gamma^{\text{sat}}}\right)^{1/2} & \text{for } T_\gamma^{\text{eq}} \gtrsim T_\gamma^{\text{sat}} \gtrsim T_\gamma^0, \\ \left(\frac{T^{\text{kd}}}{T_\gamma^{\text{sat}}}\right) & \text{for } T_\gamma^{\text{sat}} \gtrsim T_\gamma^{\text{eq}}. \end{cases} \quad (\text{C14})$$

The kinetic decoupling temperature in the equal charge case ( $g_\chi = g_l$ ) and two species of  $l$  (particle and antiparticles) is given by [27]

$$\frac{T_l^{\text{kd}}}{1 \text{ keV}} = 0.25 \times \left(\frac{r_{\text{kd}}}{0.36}\right)^{-1/2} \left(\frac{\alpha_\chi}{0.025}\right)^{-1/2} \times \left(\frac{m_\chi}{1 \text{ TeV}}\right)^{1/4} \left(\frac{m_\phi}{1 \text{ MeV}}\right). \quad (\text{C15})$$

We find the saturation temperature from Eq. (C8):

$$T_\gamma^{\text{sat}} = \frac{\pi}{r_{\text{sat}} \sqrt{c_{\text{sat}}}} \alpha_\chi^3 m_\phi (x_{\text{kd}}^l)^{-1/2} \quad (\text{C16})$$

$$= 2.96 \times 10^{-3} \text{ eV} \left(\frac{r_{\text{sat}}}{0.36}\right)^{-1} \left(\frac{\alpha_\chi}{0.025}\right)^3 \times \left(\frac{m_\chi}{1 \text{ TeV}}\right)^{-1/2} \left(\frac{m_\phi}{1 \text{ MeV}}\right) \left(\frac{T_l^{\text{kd}}}{1 \text{ keV}}\right)^{1/2}. \quad (\text{C17})$$

We now simplify the time integration of  $\Gamma$  to obtain the change in DM abundance due to reannihilation. A simple case is case 1 where  $x_0 \lesssim x_{\text{sat}}$ .  $\Gamma$  takes the maximal value of  $\Gamma_1$  today, and hence the integration can be simplified as

$$\int_{x_{\text{kd}}}^{x_0} dx \frac{\Gamma}{x} \approx \int_{x_{\text{eq}}}^{x_0} dx \frac{\Gamma}{x} \simeq 2 \times \Gamma_1. \quad (\text{C18})$$

Inserting this result into Eq. (C11) and solving for  $m_\chi$  for given  $m_\phi$  we find the maximal DM changes that reannihilation can cause. By setting the left-hand side of Eq. (C11) to 1.01 and 1.1 (correspond to “max 1%” and “max 10%” DM changes, respectively) we obtain the most left parts of the red lines in Fig. 4. In practice, this equation is solved numerically since we use tabulated values for  $x_{\text{cd}}$  and  $\alpha_\chi$ , where the latter quantity is chosen such that  $Y(x_{\text{kd}})$  gives the correct relic density. Another simple case is case 5 where  $x_{\text{sat}} \lesssim x_{\text{eq}}$ . Here, the maximum value of  $\Gamma$  is given by the saturation temperature in the radiation dominated epoch, leading to the simplification

$$\int_{x_{\text{kd}}}^{x_0} dx \frac{\Gamma}{x} \approx \int_{x_{1/v^2}}^{x_{\text{sat}}} + \int_{x_{\text{sat}}}^{x_{\text{max}}} + \int_{x_{\text{max}}}^{x_{\text{eq}}} dx \frac{\Gamma}{x} \approx \Gamma_4 [2 + \log(T_\gamma^{\text{sat}}/T_\gamma^{\text{max}})]. \quad (\text{C19})$$

In the last approximation we assumed that  $T_\gamma^{\text{sat}} \ll T_\gamma^{1/v^2}$  and  $T_\gamma^{\text{max}} \gg T_\gamma^{\text{eq}}$ . The temperature ratio appearing in the latter equation is a constant and is given by  $T_\gamma^{\text{sat}}/T_\gamma^{\text{max}} \simeq 34$ , which can be seen by applying the definitions. The abundance ratio has a power-law dependence on the parameters and corresponds to the segments of the red lines in the top right part of Fig. 4. For the intermediate regimes where saturation happens close to today or to matter-radiation equality, a simple power-law scaling cannot be found for capturing accurately the transitions. These regimes are the regions in Fig. 4 where the red curves start to bend in the  $\log(m_\chi) - \log(m_\phi)$  plane. The procedure to obtain the solution in these regimes are still the same as in the simplest cases described

above; however, the expressions become lengthy and for simplicity we do not show these cases here. Note that in all the estimates of  $\Gamma$  shown here, we have neglected the minor impact of the dark energy as well as the effect of nonlinear structure formation, assuming our homogeneous DM density treatment is valid until today. To evaluate our estimates the following values are used:

$$r_0 = 0.36, \quad (\text{C20})$$

$$T_\gamma^{\text{eq}} = 0.80 \text{ eV}, \quad (\text{C21})$$

$$T_\gamma^0 = 2.34 \times 10^{-4} \text{ eV}, \quad (\text{C22})$$

$$(g_s/\sqrt{g_{\text{eff}}})_0 = 2.12. \quad (\text{C23})$$

### 3. Redshift of reannihilation onset

In Sec. III A, we have defined the onset of reannihilation as the redshift  $z_{\text{rea}}$  where the comoving number density changes first by 1% after kinetic decoupling. Using Eq. (C11),  $z_{\text{rea}}$  can be found by solving the integral equation

$$0.01 = \int_{x_{\text{kd}}}^{x_{\text{rea}}} dx \frac{\Gamma}{x}, \quad (\text{C24})$$

where  $x_{\text{rea}} = \frac{m_\chi}{T_\gamma^0(1+z_{\text{rea}})}$ . We are mainly interested in the case where reannihilation happens between recombination and today. It turns out that the relevant parameter region is where  $x_{\text{cd}} \lesssim x_{1/v}$  and  $\Gamma$  as a function of temperature can be obtained from  $\Gamma_{1/2/3}$ ,

$$\Gamma(x) = 0.033 \times \frac{9.7}{(g_s/\sqrt{g_{\text{eff}}})_{\text{cd}}} \frac{x_{\text{cd}}}{18} \left(\frac{\alpha_\chi}{0.02}\right) \left(\frac{m_\chi}{\text{TeV}}\right)^{-1} \left(\frac{m_\phi}{\text{MeV}}\right) \times \left(\frac{T_l^{\text{kd}}}{0.25 \text{ keV}}\right) \begin{cases} \left(\frac{x}{x_0}\right)^{1/2} & x \gtrsim x_{\text{eq}}, \\ \left(\frac{x_{\text{eq}}}{x_0}\right)^{1/2} \frac{x}{x_{\text{eq}}} & x_{\text{eq}} \gtrsim x. \end{cases} \quad (\text{C25})$$

The time integration can be approximated as

$$\int_{x_{\text{kd}}}^{x_{\text{rea}}} dx \frac{\Gamma}{x} \approx \int_{x_{1/v^2}}^{x_{\text{eq}}} + \int_{x_{\text{eq}}}^{x_{\text{rea}}} dx \frac{\Gamma}{x} \approx \Gamma(x_{\text{rea}}) [2 - (T_\gamma^0/T_\gamma^{\text{eq}})^{1/2} (1 + z_{\text{rea}})^{1/2}]. \quad (\text{C26})$$

Taking this approximation in Eq. (C24) and solving for fixed  $z_{\text{rea}}$  we obtain the green lines in Fig. 4. In the case where  $x_{\text{rea}} \gg x_{\text{eq}}$  we can approximate the integral as

$$\int_{x_{\text{kd}}}^{x_{\text{rea}}} dx \frac{\Gamma}{x} \approx 2\Gamma(x_{\text{rea}}). \quad (\text{C27})$$

Taking this approximation in Eq. (C24) and solving for  $z_{\text{rea}}$  we finally obtain Eq. (25).

In regions where reannihilation can only change the DM abundances by less than 1%,  $z_{\text{rea}}$  is no longer defined—in Fig. 4 this is where the green lines stop. Note that we have

implicitly assumed that the saturation temperature is much lower than the reannihilation temperature. In the critical region, where the saturation redshift approaches  $z_{\text{rea}}$  this approximation is no longer valid, and we indicate this by the solid green curves changing into dashed green curves in Fig. 4. Since the dashed region is outside the SIDM blue band we do not investigate this case further, but we have confirmed that our numerical code exactly reproduces our estimates in its valid regime but starts to deviate when the green lines become dashed.

#### APPENDIX D: STANDARD HUBBLE EXPANSION RATE

The Hubble expansion rate as a function of the standard energy densities is given by Eq. (53). When including reannihilation we replace  $\rho_c$  by  $\rho_{\text{dark}}$  via Eq. (36), and when studying decaying DM we replace  $\rho_c$  via Eq. (56). In all cases, we take an effective neutrino mass  $m_\nu$  into account in the time evolution of  $\rho_\nu$ . We introduce a single massive eigenstate (minimum-mass normal hierarchy) such that the SM neutrino energy density evolves according to

$$\frac{\rho_\nu}{\rho_\gamma} = \frac{N_{\text{eff}}}{3} \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} \left[ 2 + \frac{I_\nu\left(\frac{m_\nu}{T_\nu^0(1+z)}\right)}{I_\nu(0)} \right], \quad (\text{D1})$$

where  $T_\nu^0 = (N_{\text{eff}}/3)^{1/4}(4/11)^{1/3}T_\gamma^0$  and

$$I_\nu(x) = \frac{1}{\pi^2} \int_0^\infty dy \sqrt{x^2 + y^2} \frac{y^2}{e^y + 1}, \quad (\text{D2})$$

with  $I_\nu(0) = 7\pi^2/120$  and the default value of the CMB temperature of today is  $T_\gamma^0 = 2.7255 \pm 0.0006$  K [105]. We derive the photon energy density from the temperature of today to be

$$\Omega_\gamma h^2 = 2.4728 \times 10^{-5}, \quad (\text{D3})$$

and other default parameters that we use, from Planck 2015 [1], are

$$m_\nu = 0.06 \text{ eV}, \quad (\text{D4})$$

$$N_{\text{eff}} = 3.046. \quad (\text{D5})$$

Furthermore, we use the results of the Planck 2015 (TT + lowP) analysis [1] where the relevant base parameters are constrained to be

$$\Omega_c h^2 = 0.1197 \pm 0.0022, \quad (\text{D6})$$

$$\Omega_b h^2 = 0.02222 \pm 0.00023, \quad (\text{D7})$$

and the derived parameters from the same analysis are given by

$$\Omega_\Lambda = 0.685 \pm 0.013, \quad (\text{D8})$$

$$h = 0.6731 \pm 0.0096, \quad (\text{D9})$$

$$z_* = 1090.09 \pm 0.42, \quad (\text{D10})$$

$$z_{\text{drag}} = 1059.57 \pm 0.46. \quad (\text{D11})$$

Using Eqs. (D3)–(D11) in Eqs. (49)–(52), we reproduce the Planck 2015 reported values (given within the parentheses below) of  $100\theta_*$ ,  $r_s(z_*)$  and  $r_s(z_{\text{drag}})$ ,

$$100\theta_* = 1.04103(1.04105 \pm 0.00046), \quad (\text{D12})$$

$$r_s(z_*) = 144.625(144.61 \pm 0.49), \quad (\text{D13})$$

$$r_s(z_{\text{drag}}) = 147.34(147.33 \pm 0.49). \quad (\text{D14})$$

- 
- [1] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A13 (2016), arXiv:1502.01589.  
[2] J. S. Bullock and M. Boylan-Kolchin, *Annu. Rev. Astron. Astrophys.* **55**, 343 (2017), arXiv:1707.04256.  
[3] B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi, *Astrophys. J.* **524**, L19 (1999), arXiv:astro-ph/9907411.  
[4] A. V. Kravtsov, *Adv. Astron.* **2010**, 281913 (2010), arXiv:0906.3295.  
[5] B. Moore, T. Quinn, F. Governato, J. Stadel, and G. Lake, *Mon. Not. R. Astron. Soc.* **310**, 1147 (1999), arXiv:astro-ph/9903164.  
[6] W. J. G. de Blok, *Adv. Astron.* **2010**, 789293 (2010), arXiv:0910.3538.  
[7] M. Vogelsberger, J. Zavala, and A. Loeb, *Mon. Not. R. Astron. Soc.* **423**, 3740 (2012), arXiv:1201.5892.  
[8] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, *Mon. Not. R. Astron. Soc.* **415**, L40 (2011), arXiv:1103.0007.  
[9] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, *Mon. Not. R. Astron. Soc.* **422**, 1203 (2012), arXiv:1111.2048.  
[10] T. Sawala *et al.*, *Mon. Not. R. Astron. Soc.* **457**, 1931 (2016), arXiv:1511.01098.

- [11] A. A. Dutton, A. V. Macciò, J. Frings, L. Wang, G. S. Stinson, C. Penzo, and X. Kang, *Mon. Not. R. Astron. Soc.* **457**, L74 (2016), arXiv:1512.00453.
- [12] A. R. Wetzel, P. F. Hopkins, J.-h. Kim, C.-A. Faucher-Giguère, D. Kereš, and E. Quataert, *Astrophys. J.* **827**, L23 (2016), arXiv:1602.05957.
- [13] A. Kamada, M. Kaplinghat, A. B. Pace, and H.-B. Yu, *Phys. Rev. Lett.* **119**, 111102 (2017), arXiv:1611.02716.
- [14] P. Creasey, O. Sameie, L. V. Sales, H.-B. Yu, M. Vogelsberger, and J. Zavala, *Mon. Not. R. Astron. Soc.* **468**, 2283 (2017), arXiv:1612.03903.
- [15] A. Robertson, R. Massey, V. Eke, S. Tulin, H.-B. Yu, Y. Bahé, D. J. Barnes, R. G. Bower, R. A. Crain, C. D. Vecchia, S. T. Kay, M. Schaller, and J. Schaye, *Mon. Not. R. Astron. Soc.* **476**, L20 (2018), arXiv:1711.09096.
- [16] K. A. Oman *et al.*, *Mon. Not. R. Astron. Soc.* **452**, 3650 (2015), arXiv:1504.01437.
- [17] I. M. Santos-Santos, A. Di Cintio, C. B. Brook, A. Macciò, A. Dutton, and R. Domínguez-Tenreiro, *Mon. Not. R. Astron. Soc.* **473**, 4392 (2018), arXiv:1706.04202.
- [18] C. Boehm, P. Fayet, and R. Schaeffer, *Phys. Lett. B* **518**, 8 (2001), arXiv:astro-ph/0012504.
- [19] L. G. van den Aarssen, T. Bringmann, and C. Pfrommer, *Phys. Rev. Lett.* **109**, 231301 (2012), arXiv:1205.5809.
- [20] F.-Y. Cyr-Racine and K. Sigurdson, *Phys. Rev. D* **87**, 103515 (2013), arXiv:1209.5752.
- [21] A. Kamada, N. Yoshida, K. Kohri, and T. Takahashi, *J. Cosmol. Astropart. Phys.* **03** (2013) 008, arXiv:1301.2744.
- [22] C. Boehm, J. A. Schewtschenko, R. J. Wilkinson, C. M. Baugh, and S. Pascoli, *Mon. Not. R. Astron. Soc.* **445**, L31 (2014), arXiv:1404.7012.
- [23] J. A. Schewtschenko, R. J. Wilkinson, C. M. Baugh, C. Boehm, and S. Pascoli, *Mon. Not. R. Astron. Soc.* **449**, 3587 (2015), arXiv:1412.4905.
- [24] F.-Y. Cyr-Racine, K. Sigurdson, J. Zavala, T. Bringmann, M. Vogelsberger, and C. Pfrommer, *Phys. Rev. D* **93**, 123527 (2016), arXiv:1512.05344.
- [25] M. Vogelsberger, J. Zavala, F.-Y. Cyr-Racine, C. Pfrommer, T. Bringmann, and K. Sigurdson, *Mon. Not. R. Astron. Soc.* **460**, 1399 (2016), arXiv:1512.05349.
- [26] J. A. Schewtschenko, C. M. Baugh, R. J. Wilkinson, C. Boehm, S. Pascoli, and T. Sawala, *Mon. Not. R. Astron. Soc.* **461**, 2282 (2016), arXiv:1512.06774.
- [27] T. Binder, L. Covi, A. Kamada, H. Murayama, T. Takahashi, and N. Yoshida, *J. Cosmol. Astropart. Phys.* **11** (2016) 043, arXiv:1602.07624.
- [28] D. N. Spergel and P. J. Steinhardt, *Phys. Rev. Lett.* **84**, 3760 (2000), arXiv:astro-ph/9909386.
- [29] M. Kaplinghat, S. Tulin, and H.-B. Yu, *Phys. Rev. Lett.* **116**, 041302 (2016), arXiv:1508.03339.
- [30] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg, and P. Walia, *Phys. Rev. Lett.* **118**, 141802 (2017), arXiv:1612.00845.
- [31] M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, *J. Cosmol. Astropart. Phys.* **05** (2017) 036, arXiv:1612.07295.
- [32] D. S. Robertson and I. F. M. Albuquerque, *J. Cosmol. Astropart. Phys.* **02** (2018) 056, arXiv:1711.02052.
- [33] F. Kahlhoefer, K. Schmidt-Hoberg, and S. Wild, *J. Cosmol. Astropart. Phys.* **08** (2017) 003, arXiv:1704.02149.
- [34] J. B. Dent, S. Dutta, and R. J. Scherrer, *Phys. Lett. B* **687**, 275 (2010), arXiv:0909.4128.
- [35] J. L. Feng, M. Kaplinghat, and H.-B. Yu, *Phys. Rev. D* **82**, 083525 (2010), arXiv:1005.4678.
- [36] J. Zavala, M. Vogelsberger, and S. D. M. White, *Phys. Rev. D* **81**, 083502 (2010), arXiv:0910.5221.
- [37] L. G. van den Aarssen, T. Bringmann, and Y. C. Goedecke, *Phys. Rev. D* **85**, 123512 (2012), arXiv:1202.5456.
- [38] R. Catena, N. Fornengo, A. Masiero, M. Pietroni, and F. Rosati, *Phys. Rev. D* **70**, 063519 (2004), arXiv:astro-ph/0403614.
- [39] A. G. Riess *et al.*, *Astrophys. J.* **826**, 56 (2016), arXiv:1604.01424.
- [40] R. A. Battye, T. Charnock, and A. Moss, *Phys. Rev. D* **91**, 103508 (2015), arXiv:1409.2769.
- [41] L. Salvati, M. Douspis, and N. Aghanim, *Astron. Astrophys.* (2018), arXiv:1708.00697.
- [42] Z. Berezhiani, A. D. Dolgov, and I. I. Tkachev, *Phys. Rev. D* **92**, 061303 (2015), arXiv:1505.03644.
- [43] K. Enqvist, S. Nadathur, T. Sekiguchi, and T. Takahashi, *J. Cosmol. Astropart. Phys.* **19** (2015) 067, arXiv:1505.05511.
- [44] J. Lesgourgues, G. Marques-Tavares, and M. Schmaltz, *J. Cosmol. Astropart. Phys.* **02** (2016) 037, arXiv:1507.04351.
- [45] A. Chudaykin, D. Gorbunov, and I. Tkachev, *Phys. Rev. D* **94**, 023528 (2016), arXiv:1602.08121.
- [46] V. Poulin, P. D. Serpico, and J. Lesgourgues, *J. Cosmol. Astropart. Phys.* **08** (2016) 036, arXiv:1606.02073.
- [47] K. M. Nollett and G. Steigman, *Phys. Rev. D* **91**, 083505 (2015), arXiv:1411.6005.
- [48] M. Hufnagel, K. Schmidt-Hoberg, and S. Wild, *J. Cosmol. Astropart. Phys.* **02** (2018) 044, arXiv:1712.03972.
- [49] O. D. Elbert, J. S. Bullock, S. Garrison-Kimmel, M. Rocha, J. Oñorbe, and A. H. G. Peter, *Mon. Not. R. Astron. Soc.* **453**, 29 (2015), arXiv:1412.1477.
- [50] J. Miralda-Escudé, *Astrophys. J.* **564**, 60 (2002), arXiv:astro-ph/0002050.
- [51] A. H. G. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat, *Mon. Not. R. Astron. Soc.* **430**, 105 (2013), arXiv:1208.3026.
- [52] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradač, *Astrophys. J.* **679**, 1173 (2008), arXiv:0704.0261.
- [53] S. Tulin and H.-B. Yu, *Phys. Rep.* **730**, 1 (2018), arXiv:1705.02358.
- [54] S. Tulin, H.-B. Yu, and K. M. Zurek, *Phys. Rev. D* **87**, 115007 (2013), arXiv:1302.3898.
- [55] S. Tulin, H.-B. Yu, and K. M. Zurek, *Phys. Rev. Lett.* **110**, 111301 (2013), arXiv:1210.0900.
- [56] C. Boehm, A. Riazuelo, S. H. Hansen, and R. Schaeffer, *Phys. Rev. D* **66**, 083505 (2002), arXiv:astro-ph/0112522.
- [57] X. Chen, S. Hannestad, and R. J. Scherrer, *Phys. Rev. D* **65**, 123515 (2002), arXiv:astro-ph/0202496.
- [58] K. Sigurdson, M. Doran, A. Kurylov, R. R. Caldwell, and M. Kamionkowski, *Phys. Rev. D* **70**, 083501 (2004); **73**, 089903(E) (2006), arXiv:astro-ph/0406355.

- [59] C. Boehm and R. Schaeffer, *Astron. Astrophys.* **438**, 419 (2005), arXiv:astro-ph/0410591.
- [60] G. Mangano, A. Melchiorri, P. Serra, A. Cooray, and M. Kamionkowski, *Phys. Rev. D* **74**, 043517 (2006), arXiv:astro-ph/0606190.
- [61] P. Serra, F. Zalamea, A. Cooray, G. Mangano, and A. Melchiorri, *Phys. Rev. D* **81**, 043507 (2010), arXiv:0911.4411.
- [62] R. J. Wilkinson, J. Lesgourgues, and C. Boehm, *J. Cosmol. Astropart. Phys.* **04** (2014) 026, arXiv:1309.7588.
- [63] F.-Y. Cyr-Racine, R. de Putter, A. Raccanelli, and K. Sigurdson, *Phys. Rev. D* **89**, 063517 (2014), arXiv:1310.3278.
- [64] C. Dvorkin, K. Blum, and M. Kamionkowski, *Phys. Rev. D* **89**, 023519 (2014), arXiv:1311.2937.
- [65] R. J. Wilkinson, C. Boehm, and J. Lesgourgues, *J. Cosmol. Astropart. Phys.* **05** (2014) 011, arXiv:1401.7597.
- [66] M. Escudero, O. Mena, A. C. Vincent, R. J. Wilkinson, and C. Boehm, *J. Cosmol. Astropart. Phys.* **09** (2015) 034, arXiv:1505.06735.
- [67] Y. Ali-Haïmoud, J. Chluba, and M. Kamionkowski, *Phys. Rev. Lett.* **115**, 071304 (2015), arXiv:1506.04745.
- [68] A. Kamada and T. Takahashi, *J. Cosmol. Astropart. Phys.* **01** (2018) 047, arXiv:1703.02338.
- [69] T. Bringmann, H. T. Ihle, J. Kersten, and P. Walia, *Phys. Rev. D* **94**, 103529 (2016), arXiv:1603.04884.
- [70] T. Bringmann and S. Hofmann, *J. Cosmol. Astropart. Phys.* **04** (2007) 016; **03** (2016) E02, arXiv:hep-ph/0612238.
- [71] T. Bringmann, *New J. Phys.* **11**, 105027 (2009), arXiv:0903.0189.
- [72] M. Archidiacono, S. Bohr, S. Hannestad, J. H. Jørgensen, and J. Lesgourgues, *J. Cosmol. Astropart. Phys.* **11** (2017) 010.
- [73] J. A. D. Diacoumis and Y. Y. Y. Wong, *J. Cosmol. Astropart. Phys.* **09** (2017) 011, .
- [74] V. Iršič *et al.*, *Phys. Rev. D* **96**, 023522 (2017), arXiv:1702.01764.
- [75] R. Huo, M. Kaplinghat, Z. Pan, and H.-B. Yu, arXiv:1709.09717.
- [76] J. Hisano, S. Matsumoto, and M. M. Nojiri, *Phys. Rev. Lett.* **92**, 031303 (2004), arXiv:hep-ph/0307216.
- [77] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, *Phys. Rev. D* **71**, 063528 (2005), arXiv:hep-ph/0412403.
- [78] J. Hisano, S. Matsumoto, M. Nagai, O. Saito, and M. Senami, *Phys. Lett. B* **646**, 34 (2007), arXiv:hepph/0610249.
- [79] R. Iengo, arXiv:0903.0317.
- [80] S. Cassel, *J. Phys. G* **37**, 105009 (2010), arXiv:0903.5307.
- [81] K. Blum, R. Sato, and T. R. Slatyer, *J. Cosmol. Astropart. Phys.* **06** (2016) 021, .
- [82] P. Gondolo and G. Gelmini, *Nucl. Phys.* **B360**, 145 (1991).
- [83] M. Drees, F. Hajkarim, and E. R. Schmitz, *J. Cosmol. Astropart. Phys.* **06** (2015) 025, arXiv:1503.03513.
- [84] T. Binder, T. Bringmann, M. Gustafsson, and A. Hryczuk, *Phys. Rev. D* **96**, 115010 (2017), arXiv:1706.07433.
- [85] K. Petraki, M. Postma, and M. Wiechers, *J. High Energy Phys.* **06** (2015) 128, arXiv:1505.00109.
- [86] K. Petraki, M. Postma, and J. de Vries, *J. High Energy Phys.* **04** (2017) 077, arXiv:1611.01394.
- [87] H. An, M. B. Wise, and Y. Zhang, *Phys. Rev. D* **93**, 115020 (2016), arXiv:1604.01776.
- [88] I. Baldes and K. Petraki, *J. Cosmol. Astropart. Phys.* **09** (2017) 028, .
- [89] T. Binder, T. Bringmann, M. Gustafsson, and A. Hryczuk (to be published).
- [90] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **571**, A16 (2014), arXiv:1303.5076.
- [91] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A24 (2016), arXiv:1502.01597.
- [92] N. MacCrann, J. Zuntz, S. Bridle, B. Jain, and M. R. Becker, *Mon. Not. R. Astron. Soc.* **451**, 2877 (2015), arXiv:1408.4742.
- [93] H. Hildebrandt *et al.*, *Mon. Not. R. Astron. Soc.* **465**, 1454 (2017), arXiv:1606.05338.
- [94] C.-H. Chuang and Y. Wang, *Mon. Not. R. Astron. Soc.* **426**, 226 (2012), arXiv:1102.2251.
- [95] X. Xu, A. J. Cuesta, N. Padmanabhan, D. J. Eisenstein, and C. K. McBride, *Mon. Not. R. Astron. Soc.* **431**, 2834 (2013), arXiv:1206.6732.
- [96] E. A. Kazin *et al.*, *Mon. Not. R. Astron. Soc.* **435**, 64 (2013), arXiv:1303.4391.
- [97] L. Anderson *et al.*, *Mon. Not. R. Astron. Soc.* **441**, 24 (2014), arXiv:1312.4877.
- [98] T. Delubac *et al.*, *Astron. Astrophys.* **574**, A59 (2015), arXiv:1404.1801.
- [99] C. Heymans *et al.*, *Mon. Not. R. Astron. Soc.* **432**, 2433 (2013), arXiv:1303.1808.
- [100] M. R. Lovell, J. Zavala, M. Vogelsberger, X. Shen, F.-Y. Cyr-Racine, C. Pfrommer, K. Sigurdson, M. Boylan-Kolchin, and A. Pillepich, *Mon. Not. R. Astron. Soc.* **477**, 2886 (2018), arXiv:1711.10497.
- [101] S. Das, R. Mondal, V. Rentala, and S. Suresh, arXiv:1712.03976.
- [102] S. R. Sandner, Bachelor's thesis, Göttingen University, Göttingen, Germany, 2016.
- [103] M. Wiesner, Bachelor's thesis, Göttingen University, Göttingen, Germany, 2016.
- [104] R. Iengo, *J. High Energy Phys.* **05** (2009) 024, arXiv:0902.0688.
- [105] D. J. Fixsen, *Astrophys. J.* **707**, 916 (2009), arXiv:0911.1955.